Bayesianism and Explanatory Unification: A Compatibilist Account

Thomas Blanchard*†

Proponents of IBE claim that the ability of a hypothesis to explain a range of phenomena in a unifying way contributes to the hypothesis's credibility in light of these phenomena. I propose a Bayesian justification of this claim that reveals a hitherto unnoticed role for explanatory unification in evaluating the plausibility of a hypothesis: considerations of explanatory unification enter into the determination of a hypothesis's prior by affecting its 'explanatory coherence', that is, the extent to which the hypothesis offers mutually cohesive explanations of various phenomena.

1. Introduction. According to Bayesianism, scientific inference is governed by the rule of conditionalization, which says that upon learning a piece of evidence e one should update one's credence in a hypothesis h by replacing one's initial probability for h $C(h)$ with its posterior probability $C(h/e)$. According to explanationism—another popular theory of scientific inference—the central principle governing scientific inference is inference to the best explanation (IBE). In its probabilistic version, IBE says that upon learning e one should assign substantially greater degrees of confidence to those hypotheses that best explain e, that is, those hypotheses that optimally combine explanatory virtues (simplicity, precision, fruitfulness, etc.).

At least on the surface, these two accounts of scientific inference look different. Since they are both plausible in their own right, it would be desirable to show that they are mutually compatible, that is, that the explanatory virtues that bear on hypothesis confirmation according to explanationism

Received August 2016; revised July 2017.

*To contact the author, please write to: Illinois Wesleyan University, Department of Philosophy, 1312 Park Street, Bloomington, IL 61701; e-mail: tblancha@iwu.edu.

†I would like to thank Heather Demarest, Michael Hicks, and two anonymous reviewers for very helpful comments. Thanks are also due to the John Templeton Foundation's Varieties of Understanding project for funding this research.

Philosophy of Science, 85 (October 2018) pp. 682–703. 0031-8248/2018/8504-0007$10.00

should also influence the credence assignments of rational Bayesian agents. This article contributes to this project by offering a compatibilist account of the explanatory virtue of unification. Considerations of explanatory unification figure prominently in many central episodes of scientific history: for instance, Copernicanism's ability to unify various celestial phenomena treated as independent by Ptolemaism played a major role in the acceptance of the heliocentric model by the scientific community. And the principle that when choosing between competing explanatory hypotheses, we should, ceteris paribus, pick those hypotheses that explain the data in a unifying way rather than those that do not is arguably one of the few substantive inferential rules common to all domains of scientific inquiry (Janssen 2002). Thus the task of finding a place for considerations of explanatory unification within the Bayesian framework is a particularly important aspect of the compatibilist project.

Following Janssen (2002), I take the defining feature of explanatory unification to be this: a hypothesis explains a range of phenomena in a unifying way when it traces them back to a 'common origin', that is, a common explanatory basis.[1] To illustrate, suppose that Jones has both pleuritis and a malar rash.[2] One unifying explanation of these data is that Jones has lupus, a common cause of both symptoms. By contrast, the hypothesis that Jones's pleuritis was caused by the flu while his rash was caused by some other, independent disease (Bloom's syndrome, say) also provides an explanation of both symptoms, but not one that unifies them: instead, the hypothesis is merely the conjunction of two separate explanations, one for each symptom. In this example, the common origin posited by the unifying hypothesis is a common cause. In other cases, the postulated common origin may instead be some mechanism, law, or theoretical principle that we may hesitate to call a 'cause' but nevertheless constitutes a common explanatory basis for the data. For instance, Darwinism unifies various phenomena such as atrophied organs, fossils, and homologies by explaining each of them as a consequence of the same biological mechanism, namely, that of evolution by natural selection.

A distinctive feature of unifying explanations is that they reveal the cooccurrence of their explananda to be no coincidence and thereby contribute to our sense of the world as an understandable and coherent place. In that respect, such explanations are especially 'lovely', which explains why unification figures prominently on the explanationist's list of considerations that bear on hypothesis confirmation. Yet it is worth pointing out explicitly that

1. This minimal definition captures the common theme running through the various accounts of explanatory unification that have been proposed in the philosophy of science, including, e.g., Whewell's (1847) notion of 'consilience of inductions', Friedman's (1974) idea that unifying explanations reduce the number of phenomena we need to posit as brute facts, and Kitcher's (1981) account of explanatory unification as repeated application of the same stringent argument pattern.

2. I borrow this example from Lange (2004), with some modifications.

no reasonable explanationist would regard unifying explanations as always more plausible than nonunifying ones. Other explanatory considerations also come into play when evaluating the credibility of an explanatory hypothesis, so that a nonunifying explanation may well be overall more credible than a unifying one. For instance, in a circumstance in which the proposition that Jones has the flu explains his pleuritis better than the lupus hypothesis (perhaps because patients like Jones rarely get pleuritis when they have lupus but typically do when they have the flu), IBE's verdict may well be that the nonunifying explanation of Jones's symptoms should receive greater credence.

Henderson (2014) helpfully distinguishes two strategies to coordinate considerations of explanatory unification with Bayesianism, each one corresponding to a specific compatibilist stance one may adopt on the relationships between explanationism and Bayesianism generally. According to what she calls 'constraint-based compatibilism' (692–95), explanatory considerations should be regarded as an indispensable external supplement to the raw Bayesian machinery: epistemic agents must take into account considerations of explanatory power when assigning their priors and/or likelihoods to hypotheses, so that explanatory virtues come into play in Bayesian updating as external constraints on 'correct' credences.[3] Regarding unification, the most explicit proposal along these lines is Lipton's (2004, 115) suggestion that "considerations of unification, simplicity and their ilk" constrain rational priors on hypotheses, so that simpler and more unifying hypotheses should receive higher prior probability. Yet Lipton does not flesh out this suggestion in any detail, and at any rate his proposal faces difficulties that reflect more general issues with constraint-based compatibilism (see Henderson 2014, 696–98). In particular, Lipton does not explain why Bayesianism needs to be supplemented with considerations of unification in the first place, leading one to suspect that the proposal amounts to a mere ad hoc accommodation of explanatory unification within the Bayesian framework rather than a real integration of the two.

A more attractive compatibilist strategy would be to try to show that an inferential privilege for unifying explanations need not be superimposed on the Bayesian machinery from the outside but falls out of Bayesian inference by itself, so that a Bayesian agent will naturally come to regard hypotheses that unify the data as, ceteris paribus, more plausible. If this 'emergent compatibilist' (Henderson 2014, 698) strategy could be made to work, it would yield a genuine integration of considerations of explanatory unification within the Bayesian framework rather than the mere accommodation supplied by constraint-based approaches, thereby providing a very satisfying picture of the

3. This form of compatibilism is advocated, for instance, by Okasha (2000), Lipton (2004), and Weisberg (2009).

relationships between explanationism about unification and Bayesianism. It is this strategy that I will follow here.

This article can be regarded as an extension of Myrvold's (2003, 2017) account, whose central notion is that of 'mutual information unification' or '*MI*-unification' for short. (As Myrvold explicitly notes, explanatory unification and *MI*-unification are not the same thing—a point to which I will come back at the end of this article. But the two notions bear interesting relations to each other, as we will see.) Myrvold defines the mutual information of a set of propositions $\{p_1, \ldots, p_n\}$ (for an agent at time t) as[4]

$$I(p_1, \ldots, p_n) = \log_2 \left[\frac{C(p_1 \,\&\, \ldots \,\&\, p_n)}{C(p_1) \ldots C(p_n)} \right], \tag{1}$$

where C is the agent's initial credence function conditional on her background knowledge at t; I measures the extent to which the members of the set are evidentially relevant to one another, that is, the extent to which they probabilistically strengthen or 'cohere with' one another. (Note that I is the logarithmic version of Shogenji's [1999] popular measure of the coherence of a set of propositions.) A hypothesis is *MI*-unifying in Myrvold's sense when it "render[s] what, on prior grounds, appear to be independent phenomena informationally relevant to each other" (Myrvold 2003, 400). And on Bayesianism, Myrvold shows, the ability of a hypothesis to *MI*-unify the data contributes to the incremental support bestowed on the hypothesis by those data—a result that makes it possible to rationalize various paradigmatic instances of inference to the most unifying explanation in Bayesian terms. Yet in a wide range of cases in which a hypothesis provides a unified explanation of the data, the hypothesis does not *MI*-unify the evidence any more than a nonunifying explanation does. I will show that in such cases, the notion of mutual information can still help us reconcile the explanationist claim about the confirmatory role of explanatory unification with Bayesianism. As I will explain, the key to such a reconciliation lies in the fact that explanatory unification contributes to a hypothesis's Bayesian prior by affecting its 'explanatory coherence', that is, the extent to which the hypothesis offers mutually cohesive explanations of various phenomena.

2. MI-Unification. Let me start with a brief summary of Myrvold's account of *MI*-unification. As mentioned, a hypothesis is *MI*-unifying when it renders various phenomena evidentially relevant to one another. More precisely, Myrvold defines the degree to which a hypothesis h *MI*-unifies a set of pieces of evidence $\{e_1, \ldots, e_n\}$ as follows:[5]

4. I borrow the phrase 'mutual information' from Myrvold (2017). Mutual information is called 'informational relevance' in Myrvold (2003).

5. *MIU* is called '*U*' in Myrvold (2003).

$$MIU(e_1, \ldots, e_n; h) = I(e_1, \ldots, e_n/h) - I(e_1, \ldots, e_n)$$
$$= \log_2\left[\frac{C(e_1 \,\&\, \ldots \,\&\, e_n/h)}{C(e_1/h)\ldots C(e_n/h)}\right] - \log_2\left[\frac{C(e_1 \,\&\, \ldots \,\&\, e_n)}{C(e_1)\ldots C(e_n)}\right]. \quad (2)$$

This quantity is positive when the e_i's provide more evidence for one another on h than unconditionally, and negative otherwise. Thus *MIU* effectively measures the amount by which h increases or decreases the mutual informational relevance of the evidence set. Note also that for two hypotheses h_1 and h_2,

$$MIU(e_1, \ldots, e_n; h_1) - MIU(e_1, \ldots, e_n; h_2) = I(e_1, \ldots, e_n/h_1) - I(e_1, \ldots, e_n/h_2). \quad (3)$$

That is, h_1 *MI*-unifies the evidence more than h_2 does just in case the mutual information of the evidence is higher given h_1 than given h_2. Myrvold (2003) shows that the relative incremental support that two hypotheses h_1 and h_2 receive from an evidence set—a standard measure of which is the log likelihood ratio of the hypotheses—can be decomposed as follows:[6]

$$\log_2\left[\frac{C(e_1 \,\&\, \ldots \,\&\, e_n/h_1)}{C(e_1 \,\&\, \ldots \,\&\, e_n/h_2)}\right] = \log_2\left[\frac{C(e_1/h_1)}{C(e_1/h_2)}\right] + \cdots + \log_2\left[\frac{C(e_n/h_1)}{C(e_n/h_2)}\right] + MIU(e_1, \ldots, e_n; h_1) - MIU(e_1, \ldots, e_n; h_2). \quad (4)$$

For future purposes it will be useful to focus on the comparative absolute support that h_1 and h_2 receive from the data, as measured in their log posterior ratio. Equation (4) entails that[7]

6. More accurately, what Myrvold (2003) shows is that the degree of support bestowed on a hypothesis h by an evidence set $\{e_1, \ldots, e_n\}$—measured by the quantity $\log_2[C(h/e_1 \,\&\, \ldots \,\&\, e_n)/C(h)]$—is the sum of the degree of support bestowed on h by each e_i considered individually and the extent to which h *MI*-unifies the evidence set:

$$\log_2\left[\frac{C(h/e_1 \,\&\, \ldots \,\&\, e_n)}{C(h)}\right] = \log_2\left[\frac{C(h/e_1)}{C(h)}\right] + \cdots + \log_2\left[\frac{C(h/e_n)}{C(h)}\right] + MIU(e_1, \ldots, e_n; h).$$

Equation (4) is a direct consequence of this equation.

7. To see this, note that

$$\log_2\left[\frac{C(h_1/e_1 \,\&\, \ldots \,\&\, e_n)}{C(h_2/e_1 \,\&\, \ldots \,\&\, e_n)}\right] = \log_2\left[\frac{C(e_1 \,\&\, \ldots \,\&\, e_n/h_1)C(h_1)}{C(e_1 \,\&\, \ldots \,\&\, e_n/h_2)C(h_2)}\right].$$

Then (5) follows from (4) by simple manipulation of the logarithms.

$$\log_2\left[\frac{C(h_1/e_1\ \&\ \ldots\ \&\ e_n)}{C(h_2/e_1\ \&\ \ldots\ \&\ e_n)}\right] = \log_2\left[\frac{C(e_1/h_1)}{C(e_1/h_2)}\right] + \cdots + \log_2\left[\frac{C(e_n/h_1)}{C(e_n/h_2)}\right] + \log_2\left[\frac{C(h_1)}{C(h_2)}\right] + MIU(e_1, \ldots, e_n; h_1) - MIU(e_1, \ldots, e_n; h_2). \quad (5)$$

Thus the relative credibility of h_1 and h_2 given the data is the sum of three kinds of terms. The first is the log likelihood ratio of h_1 and h_2 on each e_i, which measures the extent to which h_1 makes each e_i considered on its own more probable than h_2 does. The second is the log prior ratio of h_1 and h_2, that is, the extent to which h_1 is more initially plausible than h_2. The third and final term is the extent to which h_1 *MI*-unifies the data compared to h_2. This means that a greater degree of *MI*-unification increases the relative credibility of the relevant hypothesis.

These results are good news for the compatibilist. To see why, consider a simple example, which I borrow from McGrew (2003, 563–64).[8] In 1979, two quasar images were discovered to have the exact same spectral characteristics. It was suggested that these two images were produced by a single quasar whose radiation was bent by a gravitational lens—a massive gravitational object situated between Earth and the quasar.[9] The advantage of this hypothesis is that by tracing back the two images to a common origin, it virtually guarantees the match between their spectral characteristics—a match that would be an extraordinary coincidence if the images came from two different quasars. This also means that the 'one quasar' hypothesis *MI*-unifies the spectra in question, while the 'two quasars' hypothesis does not. Let s_1 and s_2 be propositions describing the spectral characteristics of the first and second quasar images, respectively; h_t is the 'two quasars' hypothesis, that is, the proposal that each image is the result of radiation emitted by a separate quasar, while h_o is the 'one quasar' hypothesis, according to which there is a single quasar in the vicinity of the two images that emits radiation bent by gravitational lensing. On h_t, the spectral characteristics of one image provide virtually no information about the spectral characteristics of the other image, so that $C(s_1\ \&\ s_2/h_t) = C(s_1/h_t)C(s_2/h_t)$ and hence $I(s_1, s_2/h_t) = 0$. By contrast, on h_o, s_1 is virtually guaranteed to hold iff s_2 holds, so that

8. See Myrvold's (2003) discussion of Copernicanism and Newtonian mechanics for more involved examples. The example originally comes from Salmon (2001). McGrew (2003) independently offers an account of *MI*-unification under the name of 'consilience'. His formal explication of that notion is superficially different from but formally equivalent to Myrvold's (see Schupbach 2005).

9. Indeed, subsequent observations revealed the presence of a cluster of elliptical galaxies in the vicinity of the two images that is responsible for the lensing.

$C(s_1 \,\&\, s_2/h_o) \approx C(s_1/h_o)$. On the other hand, h_o makes no specific prediction about the precise spectral characteristics we should observe each image to have, so that $C(s_1/h_o)$ and $C(s_2/h_o)$ are both very low. Consequently, $C(s_1/h_o)C(s_2/h_o)$ is much lower than $C(s_1/h_o)$ and hence also much lower than $C(s_1 \,\&\, s_2/h_o)$. Thus $I(s_1, s_2/h_o)$ is very high. By (3), it follows that $MIU(s_1, s_2; h_o) > MIU(s_1, s_2; h_t)$. This means that if the hypotheses under consideration are otherwise on a par with respect to the first two kinds of terms in (5), a Bayesian agent will assign more posterior credence to h_o than h_t. For instance, imagine a Bayesian agent who regards h_t and h_o as equally a priori plausible and as assigning the same probabilities to s_1 and s_2 considered on their own, so that $C(h_t) = C(h_o)$, $C(s_1/h_t) = C(s_1/h_o)$, and $C(s_1/h_t) = C(s_2/h_t)$. Then (5) entails that the agent should assign greater credence to h_o than h_t given s_1 and s_2, thereby providing a natural Bayesian reconstruction of this paradigmatic example of inference to the most unifying explanation.

Let me point out a wrinkle here. Arguably h_o does not really fall under the scope of this article, since it does not really explain either s_1 or s_2. One may reasonably think that an explanation of an item of evidence should, at a minimum, raise the probability of its occurrence. But h_o remains entirely unspecific about the precise features of the common source it postulates and hence does not make any specific prediction about the spectral characteristics of either image. So as McGrew (2003, 563) points out, it does nothing to raise the probability of either s_1 or s_2 ($C(s_1/h_o) = C(s_1)$ and $C(s_2/h_o) = C(s_2)$). (Similar remarks apply to the equally unspecific h_t.) It would perhaps be more appropriate to say that what h_o explains is not s_1 or s_2 in themselves, but the striking correlation between the two.

It may therefore be worth illustrating the results above by considering a case in which the competing hypotheses do clearly provide some explanation of each datum in the evidence set. Thus suppose that we precisify the 'bare-bones' hypothesis h_o by adding to it a hypothesis about the specific characteristics of the postulated quasar and gravitational lens that are responsible for the specific spectra of the two images (the composition of the gas cloud surrounding the black hole, the mass of the lens, etc.) in a way that provides at least some explanation of the specific spectral characteristics of the two images. Call the resulting hypothesis H_o. Hypothesis H_o, we may suppose, raises the probabilities of both s_1 and s_2. Yet if H_o is to be at all a realistic hypothesis that an actual astrophysicist may be in a position to entertain, it will still assign to each of these data a probability less than 1. (To get a probability of 1 for both s_1 and s_2, one would have to describe the postulated quasar in an excruciating amount of detail.) But H_o still virtually guarantees the match between s_1 and s_2, so that $I(s_1, s_2; H_o)$ remains higher than 0. Now contrast H_o with a similar precisification of h_t that not only posits the existence of two quasars but also makes certain precise assumptions

about the physical characteristics of the two quasars—call it H_t. On H_t, s_1 and s_2 are still mutually evidentially irrelevant since they have different and unrelated origins, so that $I(s_1, s_2; H_t) = 0$. Hence H_o still *MI*-unifies the data more than H_t. Thus, ceteris paribus, a Bayesian agent will regard the more unifying explanation of the data H_o as better supported than the nonunifying explanation H_t.

Yet not all instances of the inferential advantage claimed by explanationists for unifying explanations can be accounted for by way of *MI*-unification: a unifying explanation need not *MI*-unify the data any more than a competing nonunifying explanation does. To illustrate, return to Lange's (2004) medical diagnosis example introduced at the outset of the article.[10] Suppose we are comparing two alternative explanations of Jones's pleuritis (p) and malar rash (m): the hypothesis h_l that Jones has lupus, a common cause of both pleuritis and malar rashes, and the hypothesis h_{fb} that Jones happens to have two unrelated diseases, the flu—which causes pleuritis but not malar rashes—and Bloom's syndrome—which causes malar rashes but not pleuritis. (Then h_{fb} is nothing more than the conjunction of two separate, unrelated explanations of the symptoms: the hypothesis h_f that Jones has the flu and the hypothesis h_b that Jones has Bloom's syndrome.) Although h_l provides a unified explanation of the symptoms while h_{fb} does not, both hypotheses are on a par with respect to *MI*-unification. On h_{fb}, both symptoms are mutually evidentially irrelevant, because on the assumption that p and m were produced by unrelated diseases, neither one provides evidence for the other:

$$C(p \,\&\, m/h_{fb}) = C(p/h_{fb})C(m/h_{fb}). \tag{6}$$

But the same is true on h_l:

$$C(p \,\&\, m/h_l) = C(p/h_l)C(m/h_l). \tag{7}$$

The reason is that since p and m provide evidence for each other only because they are both signs of lupus, once the presence of lupus is held fixed, the occurrence of one symptom provides no additional information regarding the other. And (6) and (7) entail that $I(p, m; h_l) = I(p, m; h_{fb}) = 0$ and hence also that $MIU(p, m; h_l) = MIU(p, m; h_{fb})$. Thus the explanationist claim that the explanatory unification afforded by h_l contributes to its plausibility (so that, ceteris paribus, h_l should be preferred to h_{fb}) cannot be coordinated with Bayesianism by appealing to considerations of *MI*-unification. The challenge for the compatibilist, then, is to find some other way to ratio-

10. Lange offered the example precisely to make the point that common origin explanations do not necessarily *MI*-unify their explananda. My presentation here differs slightly from Lange's, which does not explicitly compare the lupus hypothesis with an alternative separate-causes hypothesis.

nalize this claim in Bayesian terms. The main purpose of the account offered in the next section is primarily intended to address this challenge. As we will see, in cases in which the unifying explanation does not *MI*-unify the data more than a nonunifying explanation, the explanationist claim that explanatory unification contributes to a hypothesis's credibility can be explained in Bayesian terms via the fact that Bayesianism favors hypotheses that offer mutually informationally relevant explanations of the data. So on Bayesian updating there are two ways in which considerations of explanatory unification can affect the credibility of a hypothesis: by increasing the mutual information of the explananda or by offering mutually informationally relevant accounts of these explananda. The upshot is a general mutual information account of explanatory unification.

3. A General Mutual Information Account of Explanatory Unification. As a first step, it will be useful to remind ourselves of a point already made in the introduction: from an explanationist point of view, other explanatory considerations besides unification come into play when evaluating the credibility of a hypothesis in light of the data. In particular, one needs to consider not only whether the hypothesis unifies the phenomena but also how well it explains each individual item of evidence considered in itself. For instance, when evaluating whether h_l or h_{fb} is the better explanation of Jones's symptoms, we should also take into account how well each hypothesis explains each symptom considered on its own. If Jones's malar rash is better explained by Bloom's syndrome than by lupus, and likewise his pleuritis is better explained by the flu than by lupus, IBE's verdict may well be that h_{fb} is more credible than h_l in light of the symptoms. To reconcile explanationism about unification and Bayesianism in cases such as Lange's medical example, it will therefore be useful to examine which factors come into play in determining how well a hypothesis explains a piece of evidence considered on its own and how to understand these factors in Bayesian terms.

To do so, let us briefly consider an example of competing explanations of a single item of evidence due to Okasha (2000, 702–3). While examining a distressed 5-year-old child, a doctor decides that the best explanation for his distress (d) is that he tore a ligament, the alternative explanation being that he pulled a muscle. The doctor is asked to explain her reasoning. Her answer is that "preadolescent children very rarely pull muscles, but often tear ligaments" and that "the symptoms, though compatible with either diagnosis, are exactly what we would expect if the child has torn a ligament, though not if he has pulled a muscle" (Okasha 2000, 703). Here two explanatory virtues come into play in evaluating the quality of each competing explanation of d. The first is the extent to which the explanation coheres with the doctor's background beliefs about the patient, reflected in her initial credence in the explanation. The second is the extent to which d is to be ex-

pected in light of the explanation—that is, how well the explanation fits the symptom, which is reflected in the likelihood of the explanation on *d*. Indeed, from a compatibilist perspective, these two factors ultimately and entirely determine the quality of the relevant explanations, as they also fix which explanation has a higher probability in light of the data.[11] Hence, from a compatibilist point of view, how well a hypothesis explains an individual piece of evidence *e* considered on its own is determined by the likelihood of the explanation of *e* that *h* provides and the prior probability of this explanation. In the case of a hypothesis *h* that explains multiple items of evidence $e_1, \ldots, e_n$ at once, how well *h* explains a given e_i on its own is determined by the prior probability and likelihood on e_i of the explanation that *h* provides for e_i.

At this point we need to bring into the foreground an intuitive distinction that will be crucial in what follows. When dealing with a hypothesis that explains multiple items of evidence $e_1, \ldots, e_n$, we should take care to distinguish between the *hypothesis* itself and the *explanation* that the hypothesis provides for some piece of evidence e_i considered on its own. Those two need not be the same thing: specifically, a hypothesis may contain an explanation of e_i while also containing additional content that comes into play in explaining other items of evidence but is itself explanatorily irrelevant to e_i. By way of illustration, consider h_{fb}. This hypothesis, remember, offers separate explanations of each of Jones's symptoms: the hypothesis h_f that *p* is due to the flu (which causes pleuritis but not malar rashes) and the hypothesis h_b that *m* is due to Bloom's syndrome (which causes malar rashes but not pleuritis). Thus, the explanation that h_{fb} provides of *p* is h_f: the additional content h_b is explanatorily irrelevant to the pleuritis. And the explanation that it provides of *m* is simply h_b, as the additional hypothesis h_f is explanatorily irrelevant to the rash. We can also apply the distinction to the competing hypothesis h_l. But here it is a distinction without a difference: h_l provides the same explanation for each symptom considered in isolation, namely, that it was caused by lupus. So the entire content of the hypothesis is explanatorily relevant to each symptom, and the explanation that the hypothesis provides for each such symptom is equivalent to the hypothesis itself.

With these considerations in mind, it becomes easy to provide a reconciliation of explanationism about unification and Bayesianism that includes

11. The qualification 'ultimately' is important here. The point is not that explanatory virtues other than initial plausibility and probability assigned to the evidence are irrelevant to the quality of the explanation, but that all other virtues will influence how good the explanation is by way of influencing how well the explanation fares with respect to those virtues. This leaves room for other explanatory considerations to have an influence. For instance, the initial plausibility of the explanation will be affected by considerations such as (e.g.) its coherence with the agent's higher-level theories of the world.

those cases in which the more unifying explanation does not *MI*-unify the data more. Let me show this by first examining Lange's medical diagnosis example in more detail and considering a situation in which both hypotheses offer equally good explanations of each symptom considered on its own. That is, let us suppose that the flu and lupus hypotheses equally well explain Jones's pleuritis considered by itself and that likewise the Bloom's syndrome and lupus hypotheses equally well explain Jones's rash considered on its own. Specifically, let us assume that the flu and the lupus hypotheses assign the same probability to p and are equally initially plausible, so that

$$C(p/h_f) = C(p/h_l), \tag{8}$$

$$C(h_f) = C(h_l). \tag{9}$$

And likewise, let us assume that the Bloom's syndrome and lupus hypotheses assign the same probability to m and are equally initially plausible, so that

$$C(m/h_b) = C(m/h_l), \tag{10}$$

$$C(h_b) = C(h_l). \tag{11}$$

In such a situation an explanationist would contend that we should ascribe more credence to the more unifying hypothesis h_l than to h_{fb}, since the two hypotheses are otherwise explanatorily on a par. The good news is that Bayesianism agrees with this verdict, as can be seen by looking in turn at the likelihoods and priors of both hypotheses.

With respect to their likelihoods on p & m, both hypotheses are on a par. One way to see this is to note first that since Bloom's syndrome is causally irrelevant to pleuritis, learning that Jones has Bloom's syndrome in addition to the flu should not change one's estimate of the probability that Jones has pleuritis. Likewise, since the flu is causally irrelevant to malar rash, learning that Jones has the flu in addition to Bloom's syndrome should not change one's estimate of the probability of m. Hence

$$C(p/h_f) = C(p/h_{fb}), \tag{12}$$

$$C(m/h_b) = C(m/h_{fb}). \tag{13}$$

Thus by (8) and (10), h_l and h_{fb} assign the same probability to each symptom considered in isolation. And since neither hypothesis *MI*-unifies the symptoms more than the other, (4) entails that they have equal likelihoods on

p & m. So as far as likelihood is concerned, there is no advantage for the unifying hypothesis h_l.

But in the present situation h_l has a higher prior probability than h_{fb}, and for reasons that directly reflect the fact that, by contrast to h_{fb}, it supplies a common origin for Jones's symptoms. To see this, note that because the flu and Bloom's syndrome are by hypothesis causally independent, they should also be statistically independent of each other. As a result, h_f and h_b are mutually evidentially irrelevant, so that

$$C(h_{fb}) = C(h_f)C(h_b). \tag{14}$$

And given (9) and (11), this entails that $C(h_l) > C(h_{fb})$. Hence although both hypotheses provide equally good explanations of each symptom considered in isolation of the other, the more unifying hypothesis h_l is more credible than h_{fb} in light of the data, because it is more plausible to begin with. Moreover, the prior disadvantage of h_{fb} compared to h_l can be analyzed in a way that is very illuminating for our purposes. Note that (14) is just a way to express the fact that the explanations of each symptom provided by h_{fb} are mutually evidentially irrelevant to one another, so that

$$I(h_f, h_b) = 0. \tag{15}$$

By contrast, because h_l traces back the symptoms to the same source, the explanations that it provides for each symptom are one and the same and hence mutually evidentially relevant. Specifically, the mutual information of these explanations is the mutual information of the set $\{h_l, h_l\}$:

$$I(h_l, h_l) = \log_2\left[\frac{C(h_l \,\&\, h_l)}{C(h_l)C(h_l)}\right] = \log_2\left[\frac{1}{C(h_l)}\right], \tag{16}$$

which is positive as long as the agent is not already certain of h_l. Now it is easy to verify that the priors of h_{fb} and h_l can be rewritten on a logarithmic scale as the sum of the prior probabilities of the explanations they offer for each symptom considered on its own and the mutual information of these explanations:

$$\log_2(C(h_{fb})) = \log_2(C(h_f)) + \log_2(C(h_b)) + I(h_f, h_b), \tag{17}$$

$$\log_2(C(h_l)) = \log_2(C(h_l)) + \log_2(C(h_l)) + I(h_l, h_l). \tag{18}$$

Given that each of the first two terms on the right-hand side of (17) equals the corresponding term on the right-hand side of (18), (17) and (18) show that the prior advantage of h_l is a direct consequence of the fact that by contrast to h_{fb}, it offers mutually cohesive explanations of each individual symptom considered in isolation. This suggests that in addition to *MI*-unification,

considerations of explanatory unification can also affect the credibility of a hypothesis by affecting the mutual informational relevance of the explanations that the hypothesis provides for each item of evidence considered on its own.

Let me now turn these considerations into a general account.[12] First, we need some formal notation to keep track of the distinction between a hypothesis and the explanation that the hypothesis provides for a piece of evidence. When a hypothesis h explains multiple items of evidence $e_1, \ldots, e_n$, I will use $h^{\rightarrow e_i}$ to denote the explanation that h provides for e_i considered on its own: $h^{\rightarrow e_i}$, then, contains all and only those parts of h that are explanatorily relevant to e_i. Specifying what 'explanatory relevance' precisely amounts to is a task for a theory of explanation; for my purposes it is enough to note that in many cases we have clear intuitions on whether some proposition or fact is explanatorily relevant to an explanandum. For instance, in the case of h_{fb}, it is intuitively obvious that $h_{fb}^{\rightarrow p} = h_f$ and $h_{fb}^{\rightarrow m} = h_b$. And in the case of h_l, it is clear that $h_l^{\rightarrow p}$ and $h_l^{\rightarrow m}$ are both identical to h_l itself, reflecting the fact that the hypothesis provides the same explanation for both symptoms.[13] In this example the explanations provided by the competing hypotheses are either entirely different propositions (in the case of h_{fb}) or propositions identical to one another and to the hypothesis itself (in the case of h_l). There are of course intermediate situations in which a hypothesis provides partially overlapping explanations of various pieces of evidence—explanations that share some but not all of their content. For example, consider two ideal gases enclosed in separate containers; let g_1 describe the pressure of the first gas and g_2 describe the pressure of the second gas. And consider the hypothesis $h_g = B \,\&\, vt_1 \,\&\, vt_2$, where B is Boyle's law, vt_1 describes the volume and temperature of the first gas, and vt_2 describes the volume and temperature of the second gas. While B is explanatorily relevant to both g_1 and g_2, vt_1 is explanatorily relevant to g_1 only, while vt_2 is explanatorily relevant to g_2 only. So $h_g^{\rightarrow g_1}$ is $B \,\&\, vt_1$, and $h_g^{\rightarrow g_2}$ is $B \,\&\, vt_2$. Hypotheses $h_g^{\rightarrow g_1}$ and $h_g^{\rightarrow g_2}$ partially overlap in that they both contain B.

For simplicity, we will assume that when h explains an evidence set $\{e_1, \ldots, e_n\}$, h can be entirely rewritten as the conjunction of the explanations that it provides for each e_i considered in isolation, so that

$$h = h^{\rightarrow e_1} \,\&\, \ldots \,\&\, h^{\rightarrow e_n}. \tag{19}$$

12. I am heavily indebted to an anonymous reviewer for very helpful suggestions regarding how to present the account.

13. Note that in some cases it may be unclear whether and how the formalism can be applied to a hypothesis h, either because it is unclear whether the hypothesis really constitutes an explanation of the evidence set or because it is unclear which parts of h are explanatorily relevant to a given e_i and which are not.

This is to assume h contains no 'explanatorily idle' content, so that every part of h is explanatorily relevant to at least one e_i. This assumption could be relaxed without changing anything central to the account proposed in this section, but it substantially simplifies some of the equations to be presented below.

Let us now consider the relative credibility of two competing explanations h_1 and h_2 of an evidence set $\{e_1 \ldots e_n\}$, as expressed in the log ratio of their posterior probabilities. From Bayes' theorem, it follows that

$$\log_2\left[\frac{C(h_1/e_1 \,\&\, \ldots \,\&\, e_n)}{C(h_2/e_1 \,\&\, \ldots \,\&\, e_n)}\right] = \log_2\left[\frac{C(e_1 \,\&\, \ldots \,\&\, e_n/h_1)}{C(e_1 \,\&\, \ldots \,\&\, e_n/h_2)}\right] + \log_2\left[\frac{C(h_1)}{C(h_2)}\right]. \tag{20}$$

If we consider the two terms on the right-hand side of (20), we will find that each is determined not only by the likelihood and/or prior of the $h_1^{\rightarrow e_i}$'s and $h_2^{\rightarrow e_i}$'s (i.e., the factors that determine how well each hypothesis explains each item of evidence considered in isolation), but also by an additional mutual information term that represents the impact of considerations of explanatory unification.

Consider the log likelihood ratio of h_1 and h_2 first. We will assume that when a hypothesis h explains an evidence set (whether in a unifying or nonunifying way), then for each e_i in the set

$$C(e_i/h^{\rightarrow e_i}) = C(e_i/h). \tag{21}$$

That is, $h^{\rightarrow e_i}$ screens off the additional content (if any) of h from e_i. We have already seen that this assumption is satisfied in the case of h_{fb} (see [12] and [13] above). And since $h_l^{\rightarrow p}$ and $h_l^{\rightarrow m}$ are each identical to h_l, (21) is trivially satisfied in the case of h_l. (Similar considerations apply in the other examples discussed in this article.) From (4) and (21), we get

$$\begin{aligned}\log_2\left[\frac{C(e_1 \,\&\, \ldots \,\&\, e_n/h_1)}{C(e_1 \,\&\, \ldots \,\&\, e_n/h_2)}\right] = {} & \log_2\left[\frac{C(e_1/h_1^{\rightarrow e_1})}{C(e_1/h_2^{\rightarrow e_1})}\right] + \cdots \\ & + \log_2\left[\frac{C(e_n/h_1^{\rightarrow e_n})}{C(e_n/h_2^{\rightarrow e_n})}\right] \\ & + MIU(e_1, \ldots, e_n; h_1) - MIU(e_1, \ldots, e_n; h_2).\end{aligned} \tag{22}$$

This says that the log likelihood ratio of the two hypotheses is the sum of two kinds of terms: the log likelihood ratio of $h_1^{\rightarrow e_i}$ and $h_2^{\rightarrow e_i}$ for each e_i and the extent to which h_1 *MI*-unifies the collective evidence compared to h_2. In other words, the likelihood of h_1 on the data as a whole compared to that of h_2 is determined by the degrees to which the individual explanations provided by h_1 and h_2 make the relevant explanandum expectable and the extent to which the explananda are mutually informationally relevant on each hypoth-

esis. Since a hypothesis that traces its various explananda back to a common origin may thereby render them mutually informationally relevant (as in the case of H_o vs. H_t), this is a first way in which from a Bayesian point of view considerations of explanatory unification can favor a unifying hypothesis.

Let us turn now to the log prior ratio of h_1 and h_2. For our purposes, it is illuminating to note that it can be parsed as follows:[14]

$$\log_2\left[\frac{C(h_1)}{C(h_2)}\right] = \log_2\left[\frac{C(h_1^{\rightarrow e_1})}{C(h_2^{\rightarrow e_1})}\right] + \cdots + \log_2\left[\frac{C(h_1^{\rightarrow e_n})}{C(h_2^{\rightarrow e_n})}\right] + I(h_1^{\rightarrow e_1}, \ldots, h_1^{\rightarrow e_n}) - I(h_2^{\rightarrow e_1}, \ldots, h_2^{\rightarrow e_n}). \tag{23}$$

Equation (23) tells us that the log prior ratio of h_1 and h_2 is the sum of two kinds of terms. The first is the log prior ratio of $h_1^{\rightarrow e_i}$ and $h_2^{\rightarrow e_i}$ for each e_i, which measures the relative initial plausibility of the explanations of e_i provided by h_1 and h_2 respectively. The second term is the extent to which the explanations provided by h_1 are more or less mutually informationally relevant than the explanations provided by h_2, that is, the extent to which the former probabilistically strengthen or 'cohere with' one another compared to the latter. Let us call the quantity $I(h^{\rightarrow e_1}, \ldots, h^{\rightarrow e_n})$ the *explanatory coherence* of h. (This label is appropriate in light of the fact that I is simply the logarithmic version of Shogenji's [1999] standard measure of the coherence of a set of propositions.) The second term in (23), then, measures the extent to which h_1 is explanatorily coherent compared to h_2.

The key point for our purposes is that whether a hypothesis is explanatorily coherent is intimately tied to whether it offers a unified explanation of the data: when h_1 traces back the various e_i's to a common origin while h_2 does not, we can expect that h_1 will be more explanatorily coherent than h_2. The reason is that when a hypothesis h offers a unified explanation of the evidence, each of the explanations of the individual phenomena provided by h will appeal to the same explanatory fact—the 'common origin' of the phenomena in question. As a result, if the agent is not already certain of the truth of these various explanations, each of them will provide positive evidence for the truth of the others, and the explanatory coherence of h will be positive. For instance, the fact that h_l traces back each symptom to the same source is reflected in the fact that its explanatory coherence is positive (see [16] above). By contrast,

14. To see why, note that from (19), it follows that

$$\begin{aligned}\log_2(C(h)) &= \log_2(C(h^{\rightarrow e_1} \;\&\; \ldots \;\&\; h^{\rightarrow e_n}))\\ &= \log_2(C(h^{\rightarrow e_1}) \ldots C(h^{\rightarrow e_n})) + \log_2\left[\frac{C(h^{\rightarrow e_1} \;\&\; \ldots \;\&\; h^{\rightarrow e_n})}{C(h^{\rightarrow e_1}) \ldots C(h^{\rightarrow e_n})}\right]\\ &= \log_2(C(h^{\rightarrow e_1})) + \cdots + \log_2(C(h^{\rightarrow e_n})) + I(h^{\rightarrow e_1}, \ldots, h^{\rightarrow e_n}),\end{aligned}$$

from which (23) follows immediately.

if h explains the phenomena in a nonunifying way, h will consist merely in the conjunction of separate explanations for each phenomenon. Because these explanations appeal to independent explanatory factors, they will be mutually evidentially neutral, so that the explanatory coherence of the hypothesis will be null, as in the case of h_{fb} (see [15]).

While in this example h_l has positive explanatory coherence because it supplies a common cause of its explananda, it should be clear that a hypothesis will also receive a positive degree of explanatory coherence as long as it provides some common explanatory basis for the phenomena, including one we may hesitate to call a 'cause'. Consider, for example, the aforementioned hypothesis h_g, which traces back the pressures of two gases g_1 and g_2 to a 'common origin' in a broad sense of the phrase, namely, the fact that both gases conform to Boyle's law B. Remember that $h_g^{\rightarrow g_1}$ is B & vt_1 and $h_g^{\rightarrow g_2}$ is B & vt_2, where vt_1 describes the volume and temperature of the first gas and vt_2 describes the volume and temperature of the second gas. Let us make the natural assumption that B, vt_1, and vt_2 are all evidentially independent of one another and also that B is evidentially independent of vt_1 & vt_2. Then the explanatory coherence of h_g is

$$I(h_g^{\rightarrow g_1}, h_g^{\rightarrow g_2}) = \frac{C(B \,\&\, vt_1 \,\&\, vt_2)}{C(B \,\&\, vt_1)C(B \,\&\, vt_2)} = \frac{1}{C(B)}, \tag{24}$$

and hence it is positive as long as $C(B) < 1$.

To summarize, the upshot of (23) is that from a Bayesian point of view, considerations of explanatory unification affect the credibility of a hypothesis not only by way of *MI*-unification but also in the form of a prior bias in favor of hypotheses that offer mutually cohesive explanations of the data. The prior probability of the hypothesis is determined not only by the extent to which each of the individual explanations it offers is initially plausible but also by the extent to which these explanations cohere with one another. And when these explanations all trace back their relevant explanandum to a single factor, they will thereby be mutually relevant to one another.

As a final step, note that by plugging our formula for the log likelihood ratio (22) and our formula for the log prior ratio (23) into (20), we obtain the following pleasing way of parsing the relative credibility of h_1 and h_2:

$$\begin{aligned}\log_2\left[\frac{C(h_1/e_1 \,\&\, \ldots \,\&\, e_n)}{C(h_2/e_1 \,\&\, \ldots \,\&\, e_n)}\right] &= \log_2\left[\frac{C(e_1/h_1^{\rightarrow e_1})}{C(e_1/h_2^{\rightarrow e_1})}\right] + \log_2\left[\frac{C(h_1^{\rightarrow e_1})}{C(h_2^{\rightarrow e_1})}\right] \\ &\quad + \cdots + \log_2\left[\frac{C(e_n/h_1^{\rightarrow e_n})}{C(e_n/h_2^{\rightarrow e_n})}\right] + \log_2\left[\frac{C(h_1^{\rightarrow e_n})}{C(h_2^{\rightarrow e_n})}\right] \\ &\quad + MIU(e_1, \ldots, e_n; h_1) - MIU(e_1, \ldots, e_n; h_2) \\ &\quad + I(h_1^{\rightarrow e_1}, \ldots, h_1^{\rightarrow e_n}) - I(h_2^{\rightarrow e_1}, \ldots, h_2^{\rightarrow e_n}).\end{aligned} \tag{25}$$

Equation (25) tells us that the relative credibility of h_1 and h_2 in light of the data is the sum of three kinds of terms. The first is the log likelihood and log prior ratios of $h_1^{\rightarrow e_i}$ and $h_2^{\rightarrow e_i}$ for each e_i. This first term effectively measures the relative quality of the explanations that h_1 and h_2 provide for each item of evidence taken by itself. The next two terms correspond to two ways in which considerations of explanatory unification can affect the relative credibility of the two hypotheses: by creating mutual informational relevance between the explananda or by increasing the mutual informational relevance of the various explanations that the hypothesis provides.[15] And while an explanatory unifying hypothesis may not *MI*-unify the evidence more than a nonunifying one, we can nevertheless expect the explanatory coherence of the unifying hypothesis to be higher. Thus (25) provides a general Bayesian vindication of the explanationist contention that considerations of explanatory unification contribute to the plausibility of a hypothesis.

Let me close this section with two additional remarks on this account. The first remark is that *MI*-unification and explanatory coherence are only two of many factors that enter into the prior and/or posterior credibility of a hypothesis. So the account of course does not entail that a unifying explanation should always be assigned more credence than a nonunifying one. But this is perfectly consistent with explanationism. To illustrate, let us return to the medical diagnosis example. As an inspection of (23) reveals, for h_{fb} to be at least as initially plausible as h_l, either h_f or h_b must be more initially plausible than h_l. Hence the prior bias for h_l encoded in (23) reveals itself in the fact that h_{fb} will be initially no less plausible than h_l only if for at least one of Jones's symptoms, the explanation of that symptom that h_{fb} provides is more initially plausible than and hence in a certain respect explanatorily superior to the competing explanation provided by h_l. Now, even if h_l is more initially plausible than h_{fb}, its posterior probability may well be lower than that of h_{fb} if the latter assigns a much higher probability to the data than h_l does. But as an inspection of (22) reveals, in any such case either the flu hypothesis must make Jones's pleuritis more likely than the lupus hypothesis does or the Bloom's syndrome hypothesis must make the malar rash more likely than the lupus hypothesis itself. Again, this is a situation in which for at least one of the two symptoms, h_{fb} provides an explanation of this symptom that has an important explanatory advantage over the explanation provided by h_l. In all such circumstances, the claim that h_{fb} may be overall more plausible than h_l is one that an explanationist can perfectly agree with. From an

15. Of course these are not mutually exclusive. Thus, in the quasar example, it is easy to see that H_o not only is more *MI*-unifying than H_t but also has higher explanatory coherence.

explanationist point of view, what happens is that the explanatory advantage bestowed on h_l by its unifying power is counterbalanced by other explanatory considerations that count in favor of h_{fb}.

Second, insofar as the account proposed here gives a role to considerations of explanatory unification in influencing Bayesian priors (by way of explanatory coherence), it provides a partial vindication of Lipton's (2004, 115) contention that when assigning prior credences to hypotheses, a Bayesian agent should privilege those that offer unified explanations of the data. But whereas Lipton's proposal remains rather sketchy, the present account allows us to envision precisely which form this privilege should take: in assigning prior credence to a conjunction of explanations, a Bayesian agent should take into account not only the prior probability of each explanation but also the degree to which they cohere or 'agree' with one another. And while Lipton's constraint on priors is superimposed from the outside, the prior preference for unifying explanations encoded in (23) emerges naturally from the Bayesian machinery considered by itself.

4. Mutual Information, Common Origin, and Brute Coincidences. In closing, let me address an objection one may raise against the account offered in the previous section.[16] To bring the objection out it is useful to first look at a complaint registered by Lange (2004) against the notion of *MI*-unification. Consider Einstein's quantum light hypothesis h_q, which says that light is quantized (as a matter of physical law). Einstein showed that on h_q various apparently unrelated phenomena involving light should in fact be expected to occur together.[17] Hence h_q *MI*-unifies the relevant phenomena. But now consider an alternative hypothesis $h_{q'}$ according to which by sheer coincidence light happens to behave as if it were quantized. As Lange points out, this hypothesis *MI*-unifies the phenomena in question just as well, since it makes their co-occurrence as likely as h_q does. This example shows that the notion of *MI*-unification is insensitive to the distinction between hypotheses that render the data mutually informationally relevant by positing a common origin and hypotheses that create mutual informational relevance between the data by positing a brute coincidence. Yet, Lange claims, the distinction has confirmatory significance: a hypothesis such as h_q should "receive more support from the phenomena" (2004, 212) than a hypothesis such as $h_{q'}$, in virtue of the fact that the former but not the latter unifies the phenomena in a genuinely explanatory way. (Call this *Lange's*

16. I am grateful to an anonymous reviewer for pressing me to address the issues discussed in this section.

17. Examples include the fact that light obeys Stokes's law of photoluminescence and the fact that X-rays produce secondary cathode rays (Lange 2004, 209).

thesis.) Lange's complaint, then, is that the notion of *MI*-unification cannot help us reconcile this thesis with Bayesianism.

Now, as Myrvold (2017, 98–101) notes, if read as a claim about incremental support, Lange's thesis seems simply incompatible with Bayesianism. As equation (4) in section 2 makes clear, for Bayesianism incremental support is determined entirely by the likelihood of the hypothesis on each individual datum and the extent to which it *MI*-unifies these data. There is simply no space for an extra incremental boost to a hypothesis's credibility due to the fact that it posits a common origin of the phenomena rather than a brute coincidence. Yet one might wonder whether the general mutual information account of unification presented in section 3 could help provide a Bayesian justification of Lange's thesis read as a claim about absolute support, by revealing that genuinely explanatory unification increases a hypothesis's prior probability more than unification by brute coincidence does. Indeed, a proponent of Lange's thesis would likely insist that only if the answer to this question is positive can the account be deemed to genuinely reconcile Bayesianism and explanationism.

But the answer is negative. This can easily be seen by returning to our earlier example involving the explanation of two gas pressures g_1 and g_2 in terms of a hypothesis h_g (which conjoins two propositions vt_1 and vt_2 describing the volumes and temperatures of the two gases with the hypothesis B that as a matter of law all gases conform to Boyle's formula). On the account proposed in section 3, considerations of explanatory unification contribute to h_g's prior plausibility by making the hypothesis positively explanatorily coherent. More specifically, the equation in footnote 14 entails that h_g's log prior is the sum of the log priors of the explanations of each individual datum that h_g offers and of its explanatory coherence:

$$\begin{aligned}\log_2(C(h_g)) = {} & \log_2(C(vt_1 \,\&\, B)) + \log_2(C(vt_2 \,\&\, B)) \\ & + I(vt_1 \,\&\, B, vt_2 \,\&\, B).\end{aligned} \tag{26}$$

And as we have seen, $I(vt_1 \,\&\, B, vt_2 \,\&\, B)$ is positive and equal to $1/C(B)$, reflecting the fact that by tracing both phenomena to a common explanatory basis (namely, Boyle's law), the two explanations $vt_1 \,\&\, B$ and $vt_2 \,\&\, B$ probabilistically strengthen one another.

But now consider an alternative hypothesis $h_{g'}$ conjoining two subhypotheses $vt_1 \,\&\, B'$ and $vt_2 \,\&\, B'$, where B' is the proposition that by sheer coincidences gases happen to behave in the way described by Boyle's formula. Although $h_{g'}$ 'unifies' g_1 and g_2 in a nonexplanatory way (i.e., by tracing both back to a single brute coincidence rather to some common explanatory basis), the account of section 3 entails that this form of unification nevertheless positively contributes to the hypothesis's prior plausibility. Specifically, the log prior of $h_{g'}$ can be decomposed as the sum of the log priors of its two subhypotheses and the mutual information of these subhypotheses:

$$\log_2(C(h_{g'})) = \log_2(C(vt_1 \,\&\, B')) + \log_2(C(vt_2 \,\&\, B')) + I(vt_1 \,\&\, B', vt_2 \,\&\, B'). \tag{27}$$

And assuming as before that vt_1 and vt_2 are evidentially irrelevant to one another, and also that B' is evidentially independent of vt_1, vt_2, and their conjunction, we have

$$I(vt_1 \,\&\, B', vt_2 \,\&\, B') = \frac{1}{C(B')}, \tag{28}$$

which is positive as long as $C(B') < 1$. True, it would be improper to call this quantity the 'explanatory coherence' of $h_{g'}$, since neither vt_1 & B' nor vt_2 & B' can be deemed to really explain g_1 and g_2. Nevertheless, because these two hypotheses 'account for' these phenomena in terms of the same brute coincidence and hence cohere with each other, considerations of mutual information contribute to $h_{g'}$'s prior plausibility just as much as they do in the case of h_g. In fact, their contribution is even greater in the case of $h_{g'}$. The reason is that B' is considerably less plausible than B and should thus be assigned lower credence than B, so that $1/C(B') > 1/C(B)$.[18] So considerations of mutual information cannot help us capture in Bayesian terms the contention that h_g should be regarded as more a priori plausible than $h_{g'}$ on the ground that only the former unifies the data in a genuinely explanatory way, as the contribution to $H_{g'}$'s prior due to (nonexplanatory) informational coherence is no less than the contribution to h_g's prior due to explanatory coherence.[19]

The upshot is that the account proposed in this article provides no way to justify Lange's thesis in Bayesian terms, as it gives a role to considerations of explanatory unification in Bayesian updating only by way of the *I* terms and *MIU* terms in (25), none of which are sensitive to the distinction between unification by common origin and unification by brute coincidence. Indeed, the account strongly suggests that however we understand it exactly, the thesis is simply irreconcilable with Bayesianism. Thus the account allows explanationists to reconcile their claims about unification with Bayesianism only at the cost of rejecting Lange's thesis and hence endorsing a more modest version of explanationism than some may like.

But in my view there are two considerations that make this a fair price to pay for the explanationist. First, note that the examples discussed in this

18. This reflects a general feature of Shogenji's (1999) measure of coherence (of which mutual information is the logarithmic version), namely, that this measure is sensitive to priors (see Fitelson [2003] and Glass [2003] for discussion).

19. Note that with respect to *MI*-unification h_g and $h_{g'}$ are on a par. The reason is that both hypotheses assign a probability of 1 to both g_1 and g_2, so that these two pieces of evidence are evidentially irrelevant to one another conditional on each hypothesis.

section are rather artificial cases insofar as the 'brute coincidence' hypotheses they involve are particularly contrived and uninteresting. While in those cases the account proposed here yields a verdict that some explanationists may find unpalatable, this is more than counterbalanced by the fact that in more realistic and interesting cases in which considerations of explanatory unification are deployed in inference (such as the ones discussed in previous sections), the account yields a perfectly good Bayesian explanation of the confirmatory role of this explanatory virtue—a very desirable outcome in its own right.

Second, it should be noted that while considerations of mutual information in themselves do not favor genuinely explanatory over brute coincidence hypotheses, Bayesianism nevertheless yields the sensible verdict that hypotheses of the former kind should generally be assigned more credence than hypotheses of the latter kind (see Myrvold 2017, 104). To illustrate, consider h_g versus $h_{g'}$ again. Because the brute coincidence that $h_{g'}$ posits is wildly implausible, a Bayesian agent will naturally assign a considerably lower prior to $h_{g'}$ than h_g and hence also a much lower posterior given that both hypotheses are on a par with respect to likelihood. This is perfectly compatible with the fact that considerations of mutual information do not favor h_g over $h_{g'}$. There are of course other considerations that come into judging the prior plausibility of these hypotheses: namely, the priors of the subhypotheses of which they are composed, which are considerably higher in the case of h_g. The point generalizes: because hypotheses that posit brute coincidences have in general much less initial plausibility than competing common origin hypotheses,[20] the Bayesian will agree with the explanationist that the latter should usually receive greater posterior credence than the former.

Overall, then, it seems to me that the consequence of my account discussed in this section is one that a reasonable explanationist can perfectly live with.

REFERENCES

Fitelson, B. 2003. "A Probabilistic Theory of Coherence." *Analysis* 63:194–99.
Friedman, M. 1974. "Explanation and Scientific Understanding." *Journal of Philosophy* 71:5–19.
Glass, D. 2003. "Problems with Priors in Probabilistic Measures of Coherence." *Erkenntnis* 63: 375–85.
Henderson, L. 2014. "Bayesianism and Inference to the Best Explanation." *British Journal for the Philosophy of Science* 65:687–715.
Janssen, M. 2002. "*COI* Stories: Explanation and Evidence in the History of Science." *Perspectives on Science* 10 (4): 457–522.
Kitcher, P. 1981. "Explanatory Unification." *Philosophy of Science* 48 (4): 507–31.

20. Although not always, as Myrvold (2017, 104–5) points out.

Lange, M. 2004. "Bayesianism and Unification: A Reply to Wayne Myrvold." *Philosophy of Science* 71 (2): 205–15.
Lipton, P. 2004. *Inference to the Best Explanation*. 2nd ed. London: Routledge.
McGrew, T. 2003. "Confirmation, Heuristics and Explanatory Reasoning." *British Journal for the Philosophy of Science* 54:553–67.
Myrvold, W. 2003. "A Bayesian Account of the Virtue of Unification." *Philosophy of Science* 70:399–423.
———. 2017. "On the Evidential Import of Unification." *Philosophy of Science* 84:92–114.
Okasha, S. 2000. "Van Fraassen's Critique of Inference to the Best Explanation." *Studies in History and Philosophy of Science* 31:691–710.
Salmon, W. 2001. "Reflections of a Bashful Bayesian: A Reply to Peter Lipton." In *Explanation: Theoretical Approaches and Applications*, ed. G. Hon and S. Rakover, 119–34. Dordrecht: Kluwer.
Schupbach, J. 2005. "On a Bayesian Analysis of the Virtue of Unification." *Philosophy of Science* 72:594–607.
Shogenji, T. 1999. "Is Coherence Truth-Conducive?" *Analysis* 59 (264): 338–45.
Weisberg, J. 2009. "Locating IBE in the Bayesian Framework." *Synthese* 167:125–43.
Whewell, W. 1847. *The Philosophy of the Inductive Sciences, Founded upon Their History*. 2nd ed. London: John W. Parker.