EVIDENTIAL DIVERSITY AND THE NEGATION OF H: A PROBABILISTIC ACCOUNT OF THE VALUE OF VARIED EVIDENCE

LYDIA MCGREW

The value of varied evidence, I propose, lies in the fact that more varied evidence is less coherent on the assumption of the negation of the hypothesis under consideration than less varied evidence. I contrast my own analysis with several other Bayesian analyses of the value of evidential diversity and show how my account explains cases where it seems intuitively that evidential variety is valuable for confirmation.

1. Introduction

Philosophers of science and probability theorists have long wrestled with a principle that scientists rely on routinely—namely, that more varied evidence is better for confirmation, all else being equal, than less varied evidence. While the evidential value of diversity seems intuitively obvious, giving a precise and convincing probabilistic explication of its worth has proven surprisingly difficult.

William Whewell calls this feature of evidence “the consilience of inductions” and gives this classic statement of its value:

[T]he evidence in favour of our induction is of a much higher and forcible character when it enables us to explain and determine cases of a *kind different* from those which were contemplated in the formation of our hypothesis. The instances in which this has occurred, indeed, impress us with a conviction that the truth of our hypothesis is certain. No accident could give rise to such an extraordinary coincidence That rules springing from remote and unconnected quarters should thus leap to the same point, can only arise from *that* being the point where truth resides.

Contact: Lydia McGrew <lydiamcgrew@gmail.com>

Accordingly the cases in which inductions from classes of facts altogether different have thus *jumped together*, belong only to the best established theories which the history of science contains. (Whewell 1847: 65, emphasis in original)

Nearly a hundred years later Rudolf Carnap makes a similar claim:

A generally accepted and applied rule of scientific method says that for testing a given law we should choose a variety of specimens as great as possible. For instance, in order to test the law that all metals expand by heat, we should examine not only specimens of iron, but of many different metals. It seems clear that a greater variety of instances allows a more effective examination of the law Generally speaking, the degree of confirmation of a law on the evidence of a number of confirming experiments should depend not only on the total number of (positive) instances found but also on their variety, i.e. on the way they are distributed among various kinds. (Carnap 1945: 93)

Bayesian probability theorists have suspected that the correct analysis of the value of diverse evidence has something to do with evidential independence, and various relations to dependence and independence have been proposed. I suggest, in contrast with previous proposals, that the confirmational value of diverse evidence for some hypothesis lies specifically in its ability to render evidence less positively dependent than less diverse evidence would, *ceteris paribus*, on the assumption of the *negation* of the hypothesis under consideration.¹

2. Earlier Independence Analyses

Colin Howson and Peter Urbach's analysis of the value of varied evidence, sometimes known as the correlation approach, proposes that a set of items of evidence is of greater confirmational value if the items of evidence are *unconditionally* independent of one another than if they are unconditionally positively relevant to each other and that the more unconditionally positively relevant the items of evidence are to one another the less confirmationally valuable they are, *ceteris paribus*.

1. In this paper H and ~H designate propositions which form a probabilistic partition; probability for purposes of this paper is defined over a propositional language. As used here, H represents not a variable but a proposition; hence, a statement that H screens off the items of evidence from each other *does not* entail that ~H also screens.

Evidence that is varied is often regarded as offering better support to a hypothesis than an equally extensive volume of homogeneous evidence According to the Bayesian, if two data sets are entailed by a hypothesis (or have similar probabilities relative to it), and one of them confirms more strongly than the other, this must be due to a corresponding difference between the data in their probabilities The idea of similarity between items of evidence is expressed naturally in probabilistic terms by saying that e_1 and e_2 are similar if $P(e_2|e_1)$ is higher than $P(e_2)$, and one might add that the more the first probability exceeds the second, the greater the similarity. This means that e_2 would provide less support if e_1 had already been cited as evidence than if it was cited by itself. (Howson & Urbach 1993: 113–114)

Howson and Urbach's idea seems to be that the confirmation of H by the conjunction ($E_1 \& E_2$) is lessened if the confirmational force of E_2 is less when E_1 has already been taken into account. Since, they claim, this is the case when E_1 and E_2 are unconditionally positively relevant to each other (given some other conditions), absence of unconditional positive relevance among individual items of evidence for H is, they propose, a good probabilistic model for the confirmational value of varied evidence.

Branden Fitelson (2001: S132–134) has shown that there are severe problems with Howson and Urbach's analysis. To begin with, their probabilistic claim concerning the lessened value of the second item of evidence holds only when E_1 and E_2 are both entailed by H and only when confirmation is measured by the r measure of confirmation—i.e.,

$$r(H,E) = P(H|E)/P(H) = P(E|H)/P(E)^2$$

These points are problematic for taking the correlation approach to be an accurate analysis of the value of varied evidence, if for no other reason than that varied evidence seems to be valuable regardless of whether any of the evidence under consideration is entailed by H . And even in cases where the evidence is entailed by the theory, Howson and Urbach's point about the lessened value of the second item of evidence when the first has been taken into account is an artifact of the role played by the prior probability of each item of evidence in the r measure of confirmation. This is an undesirable restriction of their point to a single (and disputed) concept of confirmation.

Moreover, intuitive cases of valuable varied evidence crop up frequently

2. In this paper I will use the simple ratios to represent the confirmation measures r and l rather than the logarithms of the ratios.

when two items of evidence are unconditionally positively relevant to each other, which is a good indication that the Howson and Urbach analysis does not account for the value of varied evidence. Fitelson (2001), following Elliot Sober (1989), points out that we have an intuitive case of the value of varied evidence when we contrast the value of two newspaper reports of the result of a football game, taken from the same newspaper, with the value of one newspaper report and one report from a different news source (such as a radio report from a different news outlet). Yet these items of evidence are not *unconditionally* independent of each other. They are positively relevant to each other. The radio report gives us some reason to expect the newspaper report, and vice versa.^{3, 4}

Wayne Myrvold (1996) suggests a more nuanced and plausible account than Howson and Urbach's. As Myrvold points out,

It turns out that the quantity which plays a key role in the degree of confirmation of h by $e_1 \& e_2$ is not $S(e_1, e_2)$ [a measure of unconditional coherence], but rather the ratio of $S(e_1, e_2 | h)$ to $S(e_1, e_2)$. The degree of confirmation of h by $e_1 \& e_2$ is a product of three factors: the degree of confirmation of h by e_1 alone, the degree of confirmation of h by e_2 alone, and an "interaction term" which is the ratio of the degree of similarity of the body of evidence, conditionalized upon h , to its prior degree of similarity. (Myrvold 1996: 662–3).

What Myrvold here calls S is identical to the measure of probabilistic coherence suggested by Tomoji Shogenji (1999). It is the ratio of the probability of the conjunction of items in a set of propositions to the product of their individual probabilities. In the case of a set of evidence $\{E_1, \dots, E_n\}$, it is

3. The fact that Howson and Urbach's account does not deal well with such cases is a result of their failure to set a *ceteris paribus* condition that, in comparing confirmation by sets of evidence, the individual confirmations provided by the items, or even their product, must be held constant. Intuitively varied evidence may have high unconditional positive relevance merely as a result of the fact that different items all confirm H and confirm one another indirectly in that way. Unconditional positive relevance of evidence can thus rise without any negative effect upon the confirmational value of the conjunction of the items of evidence, if it reflects a rise in the confirmation that the items give to H .

4. John Earman's account (1992: 77–79, 240 n. 15) of the value of varied evidence is very similar to Howson and Urbach's, though Earman focuses not on r but on the posterior probability of H . He points out that, if H entails all of the evidence and if we hold constant the prior of H , by Bayes's Theorem the posterior of H will be higher insofar as the prior probability of the conjunction of the evidence is lower. The unconditional coherence of the evidence is a factor affecting the prior probability of the conjunction of evidence. Hence, Earman relates valuable diversity of evidence to decreased unconditional coherence among items of evidence. This account suffers from two of the same problems as Howson and Urbach's—namely, it addresses only cases where H entails all of the evidence, and it cannot explain the fact that there are intuitive cases of valuable diversity of evidence despite high unconditional coherence of evidence.

$$\frac{P(E_1 \& \dots \& E_n)}{P(E_1) \times \dots \times P(E_n)}.$$

The conditionalized version that Myrvold calls $S(e_1, e_2 | h)$ has gone by various names in the literature. It has been called a measure of theoretical consilience, a conditionalized measure of probabilistic similarity, and a conditional degree of coherence. (See McGrew 2003: 562; Myrvold 1996: 662; Schupbach 2005: 597; and Shogenji 2013: 2532–3). It is the probability of the conjunction of the E_i 's conditional on H to the product of their individual probabilities conditional on H, i.e.,

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 | H) \times \dots \times P(E_n | H)}.$$

I will usually be referring to this ratio as the “conditional coherence of the evidence on H” or the “coherence of the evidence conditional on H,” and *mutatis mutandis* for $\sim H$. Myrvold (1996), using the r measure of confirmation, shows that the confirmation of H by $(E_1 \& \dots \& E_n)$ can be analyzed in terms of the individual confirmations provided by each of the e_i 's and the ratio of the coherence of the evidence conditional on H to its unconditional coherence (Myrvold 1996: 663).

$$P(H|E_1 \& \dots \& E_n)/P(H) = P(H|E_1)/P(H) \times \dots \times P(H|E_n)/P(H) \times \frac{P(E_1 \& \dots \& E_n | H)/P(E_1 | H) \times \dots \times P(E_n | H)}{P(E_1 \& \dots \& E_n)/P(E_1) \times \dots \times P(E_n)}.$$

The difficulty in understanding and evaluating Myrvold's proposal as an account of the value of varied evidence arises from the fact that he does not explicitly state *ceteris paribus* conditions for applying his analysis to cases of varied evidence. He says that “a diverse body of evidence confirms an hypothesis more strongly if the hypothesis renders the evidence less diverse” (1996: 663), which draws our attention to the final ratio in the above equation; but this does not make it clear whether, in comparing a less diverse to a more diverse body of evidence, the unconditional coherence of the evidence (the denominator) is to be held constant while the conditional coherence on H (the numerator) is raised, whether the conditional coherence on H is to be held constant while the unconditional coherence is lowered, or whether neither needs to be held constant. If neither needs to be held constant in his analysis, Myrvold is saying that the value of diverse evidence lies in an increase in the “interaction term” taken as a whole.

He also does not specify (though one would assume that he intends this) that the individual confirmations of *H* by the items of evidence are to be held constant (or that their product is to be held constant) or that the prior probability of *H* is to be held constant when we envisage comparing a less diverse to a more diverse body of evidence.⁵ (See appendix.)

Greg Novack (2007: 709) suggests that Myrvold (1996) may be understood as proposing that, if we hold constant the product of the individual confirmations offered by the items of evidence to *H*, and if we hold constant the conditional coherence of the evidence on *H*, then the evidence confirms *H* more strongly if it is less unconditionally coherent. (See appendix for the relationship between a similar set of *ceteris paribus* conditions and my own analysis.) That statement is *true* given Myrvold's decomposition of confirmation by the conjunction, but it is not clear that those *ceteris paribus* conditions capture the analysis of the value of diverse evidence that Myrvold intends to offer.

There is some reason to think otherwise from Myrvold's (2003) discussion of the William Whewell quotation given at the beginning of this paper. Whewell is clearly talking about evidential diversity, and, although Myrvold does not there use an explicit phrase such as "diverse evidence," he states that, if Whewell envisages the items of evidence as independent of each other conditional on *H*, his own analysis of consilience is not pertinent to Whewell's discussion:

On a common reading of Whewell's account of consilience, the unifying hypothesis is a fully specified hypothesis that entails evidence from disparate domains. If this were the whole story according to Whewell, there would [be] little room for a close parallel between our account [of consilience] and Whewell's, as two evidence statements that are both entailed by a hypothesis *h* are informationally irrelevant to each other, conditional upon *h*. (Myrvold 2003: 418)

He then argues that, in fact, that need not be the whole story according to Whewell and offers a different interpretation, concluding,

A hypothesis that entails such a law-like connection will render the facts in the disparate domains connected by the law informationally relevant to each other. (Myrvold 2003: 419)

This seems to indicate that facts from disparate domains that confirm *H* are particularly valuable because *H* renders them *positively relevant* to each other. If that

5. If these are held constant, it is impossible to hold constant the unconditional coherence of the items while raising the conditional coherence on *H*. Proof omitted.

is intended as an analysis of the value of diverse evidence, one would not (contra Novack's interpretation of Myrvold 1996) be holding constant the conditional coherence of the evidence on H but rather *increasing* it when moving from less diverse to more diverse bodies of evidence. I will return later, when I apply my own analysis to a number of intuitive cases, to the question of whether the value of diverse evidence should be understood in terms of increasing coherence conditional on H.

Finally, there is Branden Fitelson's proposal (2001: S130ff) of what he calls a partial analysis of the confirmational significance of evidential diversity. Fitelson suggests that an important part of the value of evidential diversity lies in its ability to render items of evidence "confirmationally independent regarding" an hypothesis (2001: S130-134). The concept of confirmational independence as Fitelson defines it can be stated informally by saying that the members of a set of items of evidence $\{E_1, \dots, E_n\}$ are confirmationally independent of each other concerning H, according to some measure of confirmation c , just in case each item has the same degree of confirmational impact on H (as measured by c) regardless of whether the item is taken into account by itself or after all of the other items have already been taken into account. Under confirmational independence, later items confirm H neither more nor less strongly given that the other items have already been conditionalized on (2001: S125). Fitelson points out that, when one uses the likelihood measure of confirmation (which he favors), a sufficient condition for confirmational independence is that both H and $\sim H$ screen off the items of evidence from each other, making them probabilistically independent of each other conditional on either H or $\sim H$ (2001: S130). That is, for two items, $P(E_1|H \& E_2) = P(E_1|H)$ and *mutatis mutandis* for $\sim H$. For larger bodies of evidence, $P(E_1|H \& E_2 \& \dots \& E_n) = P(E_1|H)$, and the same *mutatis mutandis* for $\sim H$.⁶

My own analysis of the value of evidential diversity bears a resemblance to Fitelson's, but it is much more specific. Fitelson's analysis seems incomplete, since moving from confirmational dependence to confirmational independence is not confirmationally valuable *per se*. Whether it is valuable or not depends upon what it is being contrasted with. For example, if all of the items of evidence are confirmationally *positively* relevant to each other concerning H, so that each item has *more* confirmational force for H when the other items are taken into account than it would have alone, then all else being equal it would have a negative effect on the confirmation of H if this were changed so that the items were confirmationally independent of each other concerning H instead. To argue that varied evidence confirms more strongly because of confirmational in-

6. Items of evidence may also be confirmationally independent of each other concerning H on the likelihood measure if the degree of dependence of the items as measured by the conditionalized measure of coherence described above is equal for H and $\sim H$. (This is my point, not Fitelson's.)

dependence, one would have to argue that, when more varied evidence seems intuitively to confirm more strongly than less varied evidence, the less varied evidence is confirmationally dependent in a way that is *harmful* rather than *helpful* to the confirmation of H. In that case, the more varied set is better, all else being equal, if it is more confirmationally independent than the less varied set. Fitelson does not provide such an argument, though my own account indicates how and when this is the case.

Second, it is possible for one set of evidence to be confirmationally dependent concerning H while another is confirmationally independent because of a difference in the dependence of the evidence conditional on H rather than on $\sim H$. For example, we can envisage a scenario in which the items of evidence are negatively probabilistically correlated conditional on H, making their confirmational dependence negative (each item confirms more weakly when other items are taken into account). In that case, a different set of evidence would be more confirmatory (using the *l* measure of confirmation), all else being equal, if the items in that set of evidence were screened off from each other by H. And in that case, the change from confirmational dependence to independence would be helpful to H. Is this how varied evidence works? Fitelson directs our attention to confirmational independence but does not specify whether the value of diversity in making evidence confirmationally independent lies on what one might call the H side, the $\sim H$ side, or both.⁷

My own argument is that the significant value of more varied evidence lies not in altering the dependence of evidence conditional on H but rather in rendering evidence less coherent (a concept which extends to negative dependence) conditional on $\sim H$. That is to say, my thesis is that, whenever it seems that we have valuable varied evidence, this can be understood by seeing that the varied evidence decreases the conditional coherence of the evidence on $\sim H$,

$$\frac{P(E_1 \& \dots \& E_n | \sim H)}{P(E_1 | \sim H) \times \dots \times P(E_n | \sim H)}.$$

as compared to less varied evidence.

3. The Confirmational Value of Reduced Coherence on $\sim H$

Suppose that we are using the likelihood ratio

7. An analysis relating variety of evidence to independence conditional on each of H and $\sim H$ and hence similar to Fitelson's is mentioned briefly in Ahmed (2015: 22, fn. 22) based on the suggestion of an unnamed referee for that paper.

$$l(H,E) = P(E|H)/P(E|\sim H)$$

as a measure of confirmation. Then, the extent to which $(E_1 \& \dots \& E_n)$ confirms H equals (by definition)

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 \& \dots \& E_n | \sim H)}.$$

I stipulate the following *ceteris paribus* conditions for two sets of evidence $\{E_1, \dots, E_n\}$ and $\{E'_1, \dots, E'_n\}$ and some hypothesis H.

1) The products of the Bayes factors for all items of evidence in each of the two sets are the same. That is,

$$\frac{P(E_1|H)}{P(E_1|\sim H)} \times \dots \times \frac{P(E_n|H)}{P(E_n|\sim H)} = \frac{P(E'_1|H)}{P(E'_1|\sim H)} \times \dots \times \frac{P(E'_n|H)}{P(E'_n|\sim H)}$$

2) The coherence of the items of evidence in each of the two sets conditional on H is the same.

That is,

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 | H) \times \dots \times P(E_n | H)} = \frac{P(E'_1 \& \dots \& E'_n | H)}{P(E'_1 | H) \times \dots \times P(E'_n | H)}.$$

8

As shown in Myrvold (2003: 412–413) the Bayes factor for the conjunction $(E_1 \& \dots \& E_n)$, which expresses the confirmation the conjunction gives to H by the *l* measure, can be expressed as

8. I assume that all conditional probabilities are greater than zero. My wording in what follows also implies that all items of evidence individually confirm H and that the conjunctions of the items confirm H in each case. Strictly speaking, these two conditions concerning confirmation are not necessary to show that decreased coherence on $\sim H$ raises the “confirmation” of H by the *l* measure—that is to say, raises that quantity that is the confirmation by the *l* measure. And the same for confirmation by the *r* measure discussed in the appendix. Therefore, I have not listed them as *ceteris paribus* conditions. However, I will be using language throughout this paper that implies that both sets of evidence *do confirm* H and discussing cases in which one set confirms H *more* than the other. Generally in discussions of the value of varied evidence it is assumed that the (real or hypothetical) less varied evidence *does* confirm H but that varied evidence confirms more strongly. Similarly, it could be difficult to evaluate whether a case involved valuable varied evidence if the individual Bayes factors for the items of evidence varied, even if their product remained the same. However, for the proof only the weaker assumption that the product of the individual Bayes factors remains the same is needed. I owe this last point to an anonymous reviewer. I mention these additional conditions not as a requirement of the proof but to relate the result to intuitive cases of valuable varied evidence.

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 \& \dots \& E_n | \sim H)} = \frac{P(E_1 | H)}{P(E_1 | \sim H)} \times \dots \times \frac{P(E_n | H)}{P(E_n | \sim H)} \times \frac{P(E_1 \& \dots \& E_n | H) / P(E_1 | H) \times \dots \times P(E_n | H)}{P(E_1 \& \dots \& E_n | \sim H) / P(E_1 | \sim H) \times \dots \times P(E_n | \sim H)} \quad 9$$

It follows, given the *ceteris paribus* conditions, that $(E_1' \& \dots \& E_n')$ provides greater confirmation to H than $(E_1 \& \dots \& E_n)$ by the *l* measure just in case

$$P(E_1 \& \dots \& E_n | \sim H) / P(E_1 | \sim H) \times \dots \times P(E_n | \sim H) > P(E_1' \& \dots \& E_n' | \sim H) / P(E_1' | \sim H) \times \dots \times P(E_n' | \sim H)$$

That is to say, if we hold constant the product of the individual confirmations that the items of evidence give to H and hold constant the extent to which H itself unifies the items of evidence, then the second set of evidence provides more confirmation to H by the *l* measure than the first set just in case the items in the first set of evidence are more coherent on the assumption of $\sim H$ than the items in the second set of evidence. So if some feature of a set of evidence makes it the case that, all else being equal, the items in the set are less coherent given $\sim H$ than the items in another set, that feature increases the confirmation of H by the evidence. An analogous result can be shown for the *r* measure of confirmation. (See appendix.)¹⁰

It is therefore not hard to see that reduced coherence conditional on the negation of an hypothesis is confirmationally valuable to the hypothesis. However, it does not follow from that point alone that a decrease in conditional coherence on the negation is the proper analysis of the value of diverse evidence. I could have chosen different *ceteris paribus* conditions and shown that other things are also helpful to confirmation. But some of those other factors will clearly not work as candidates for the locus of the value of diverse evidence. For example, it is implausible that the value of diverse evidence is that items of evidence that seem intuitively more diverse are generally *individually* more confirmatory of H. The point of diverse evidence seems to have something to do with the relation of the items of evidence *to one another*, not with the extent to which each item individually confirms H, aside from the consideration of any other items. So it seems that on any plausible analysis of the value of diversity, we are justified in holding constant the individual confirmations that the items of evidence give to H and *ipso facto* the product of the individual confirmations. Also, since diversity is a

9. Since Myrvold (2003) is using logs, his equation takes a different form. I had independently derived this equivalence (based on the argument in Lydia McGrew and Timothy McGrew 2009: 633–34) before becoming aware of it in Myrvold's work.

10. Note that the concept of making items of evidence less coherent given an hypothesis does *not* have probabilistic independence given that hypothesis as a limit. It can include making those items *negatively* relevant to each other. A group of evidence items that are screened off from each other by $\sim H$ may be more or less coherent conditional on $\sim H$ than a group of items that are not screened off (that are probabilistically dependent conditional on $\sim H$), depending on whether the conditional measure of coherence for the dependent group on $\sim H$ is greater or less than 1.

feature of the evidence rather than of the hypothesis under consideration, the prior probability of H should not be changed when we compare more diverse with less diverse bodies of evidence.¹¹ And, as pointed out above, Fitelson has argued persuasively contra Howson and Urbach that lessened unconditional positive relevance among items of evidence is not, all by itself, a correct locus for the value of diverse evidence.

However, the possibility remains (at least for some philosophers) that the value of diverse evidence defies precise probabilistic analysis. (See Novack, 1978, and Carnap, 1945: 93–94, for a discussion of this position as held by Ernest Nagel, 1939, 68–71.) Even if we reject that perspective, there is still the question of whether the value of diverse evidence lies to any significant extent in the coherence of the evidence conditional on H —a quantity that I have held fixed in the result in this section. An application of my analysis to specific examples in which it seems intuitively that varied evidence is valuable will show the adequacy of the account proposed here.

4. Application to Intuitive Cases

4.1. Newton and Celestial and Terrestrial Confirmations

Consider the Newtonian proposition that the attractive force between any two masses in the universe is directly proportional to their mass and inversely proportional to the square of the distance between them. Intuitively, it seems that the Newtonian theory (taken as a universal law) is especially well confirmed if we find that not only terrestrial objects like apples and cannon balls but also celestial entities like the moon and planets obey these rules. And the more different “kinds” of things we find that obey Newton’s law in relation to each other, the better for the confirmation of the theory—a comet and the earth, the sun and the planets, the tides and the moon, and so forth (see Thagard 1978: 81). This intuition persists despite the fact that it is famously difficult to articulate a precise definition of the concept of a “kind” of evidence or of a “kind” of event.

For purposes of applying my analysis of the value of diverse evidence to specific cases, it will be useful to have in hand the concept of a subhypothesis. A subhypothesis of some hypothesis H is any proposition with a probability between 0 and 1 that entails H , and most often I will be using the term to refer to what would more precisely be called a *proper* subhypothesis of H (or more often of $\sim H$)—that is, a proposition that has a probability between 0 and 1, that

11. This is an added *ceteris paribus* condition for the proof for the r measure in the appendix.

entails H, but that is not entailed by H, and *mutatis mutandis* for $\sim H$.¹² It is possible for a subhypothesis, under particular evidential circumstances, to have an equal probability to the probability of what we might think of informally as the original (or “umbrella”) hypothesis and for the two to be mutually entailing. When H and one of its subhypotheses are mutually entailing and hence the same hypothesis under different descriptions, it is a purely pragmatic matter which one is spoken of as a subhypothesis of the other, and nothing in my argument turns on that designation in that case.

For example, suppose that H is, “There is no dagger before Macbeth’s eyes.” Then a subhypothesis of H, H_1 , is, “There is no dagger before Macbeth’s eyes, but Macbeth is hallucinating that there is a dagger.” Under normal evidential circumstances, H_1 entails H but not *vice versa*. It is entirely possible that there is no dagger before Macbeth and that he is also not hallucinating a dagger. But if we conditionalize on the evidence, “Macbeth seems to see a dagger before him which he cannot touch,” and if we rule out as impossible the idea that any sort of hologram hoax is being played upon Macbeth, then it seems that the only way for there to be no dagger in front of Macbeth is if he is, in fact, hallucinating a dagger. Under those evidential conditions it makes sense to say that H and H_1 are equivalent and that $P(H) = P(H_1)$.

The fact that proper subhypotheses are not entailed by the hypothesis of interest is useful in generating intermediate-valued likelihoods for Bayesian inference. So, if H is “Jones is a murderer” we may partition the probability space of H according to whether or not Jones is a clumsy murderer. Neither of those subhypotheses (“Jones is a clumsy murderer” or “Jones is a non-clumsy murderer”) is entailed by H, but each entails H. Their respective probabilities conditional on H will influence the conditional probability of particular items of evidence. If the detective has reason to believe that Jones, if he were the murderer at all,

12. An anonymous reviewer points out that the account of an hypothesis and a subhypothesis given here would suggest that “All metals expand when heated” is a subhypothesis of “Iron expands when heated,” which might seem backwards. The idea of a subhypothesis of H given here is meant to capture the concept of “a way in which H as opposed to $\sim H$ obtains.” Subhypotheses of H are supposed to constitute parts of the probability partition of H (if they are proper subhypotheses) or the whole of H (if they are both entailed by H and entail H). And *mutatis mutandis* for $\sim H$. Since “Iron expands when heated” is compatible with “It is not the case that all metals expand when heated,” the instance of iron’s expansion should *not* be considered a subhypothesis of the general claim that all metals expand when heated. Appearances to the contrary notwithstanding, then, it is quite legitimate for “All metals expand when heated” to be a subhypothesis of “Iron expands when heated,” since the former describes one possible state of affairs in which the latter, as opposed to its negation, holds. Of course, in some epistemic context, the hypothesis of interest might be “All metals expand when heated,” but in that case, “Iron expands when heated” should not be a subhypothesis thereof, though an entailed proposition. It does not capture the concepts I am trying to use here to regard the relationship of hypothesis to subhypothesis as the relationship of a general proposition to the specific instances entailed by it.

would be highly intelligent and careful, the discovery that the murderer appears to have been a complete blunderer could lead the detective to look for a different suspect.

To return to the case of Newton, why is it that Newton's universal law seems better confirmed by a set of evidence consisting of a terrestrial and a celestial example of its apparent application (say, the fall of an apple, on the one hand, and the motion of a planet around the sun, on the other) than by a set consisting only of two terrestrial examples (say, the fall of the apple and the fall of a cannon ball)? Consider the hypothesis $\sim N$ which is just this negation:

$\sim N$ Newton's universal law is false.

Now consider a subhypothesis

$(\sim N)_1$ Newton's universal law is false, but there is a special property of terrestrial bodies that causes them to appear to obey Newton's law in relation to the earth.

Both of the terrestrial instances confirm N , but they also can be explained by $(\sim N)_1$.¹³ While, as Wesley Salmon has noted (1990: 191), it is difficult to give probabilities to evidence conditional on the full negation of an hypothesis, which includes the infamous "catchall," it is clear that $(\sim N)_1$ gives much higher probability to each of the terrestrial instances than the average probability of each of those instances conditional on $\sim N$ as a whole. The average probability of an item of evidence E_i conditional on $\sim N$ as a whole is the likelihood $P(E_i | \sim N)$. In this particular case, where E_1 and E_2 concern terrestrial bodies, $P(E_1 | (\sim N)_1)$ is 1, and the same for E_2 , since $(\sim N)_1$ expressly states that terrestrial bodies appear to obey Newton's law.¹⁴ The probability of the individual items given $\sim N$ is intuitively far less than 1. Hence, assuming that $(\sim N)_1$ has a probability greater than zero conditional on $\sim N$, each item confirms $(\sim N)_1$ conditional on $\sim N$.

The result is that to some degree the two terrestrial items of evidence are positively dependent conditional on $\sim N$. In many cases,

$$P(E_1 | \sim H \ \& \ E_2) > P(E_1 | \sim H)$$

13. In this particular case the individual items of evidence actually confirm $(\sim N)_1$ overall, which is an additional reason to be concerned about $(\sim N)_1$. However, it is not generally the case that, in all situations where varied evidence is valuable, the individual items in a less varied set *must* confirm the unifying subhypothesis overall. Examples and proof omitted.

14. The subhypotheses of $\sim N$ discussed here and in Section 5 have all the marks of *ad hoc* hypotheses. See Lydia McGrew (2014).

if

$$P((\sim H)_1 | \sim H) > 0$$

and if

$$P(E_1 | (\sim H)_1) > P(E_1 | \sim H) \text{ and}$$

$$P(E_2 | (\sim H)_1) > P(E_2 | \sim H),^{15}$$

and such a situation holds for $(\sim N)_1$ and $\sim N$ and explains the dependence of the terrestrial instances. Informally, we can say that, if one knew with certainty that Newton's universal law were *false*, and if one had the first terrestrial instance in hand (the fall of an apple), this would give one some reason to expect the second terrestrial instance (the fall of a cannon ball). The first terrestrial instance, conditional on $\sim N$, would give one some reason to expect the second because of its confirmation of $(\sim N)_1$, given $\sim N$. (It should be noted that, if $P((\sim N)_1 | \sim N)$ is very, very low, the dependence induced will also be negligible. See the discussion in Section 5 concerning subhypotheses with extremely low probability conditional on $\sim N$.)

But this is not the case when it comes to the set of observations consisting of a terrestrial and a celestial instance of the apparent satisfaction of Newton's law. $(\sim N)_1$ gives no help to $\sim N$ when it comes to a piece of evidence that appears to illustrate Newton's law in the *celestial* realm—say, the path of another planet around the sun. In order to try to explain both a terrestrial and a celestial instance under $\sim N$, one would have to reach for a different subhypothesis of $\sim N$. For example, one might add,

$(\sim N)_2$ Newton's universal law is false, but there is a special property of planets in the solar system that causes them to appear to obey Newton's law in relation to the sun.

But $(\sim N)_1$ and $(\sim N)_2$ as stated seem to be at least somewhat, if not entirely, independent of each other conditional on $\sim N$. (See Section 5 of this paper for a discussion of other potential subhypotheses of $\sim N$.) Hence, the intuitively more varied set of evidence can be described as more varied in the probabilistic sense that it is less positively dependent conditional on $\sim N$. And this is helpful, all else being equal, to the confirmation of N , as the previous section has shown.

15. I am indebted to an anonymous reviewer for drawing my attention to the incompleteness of my conditions here, which I have now indicated by "in many cases." The evidence could fail to be positively dependent given $\sim H$ if these conditions held but if there were also some countervailing epistemic cause of *negative* dependence among the items of evidence conditional on $\sim H$.

Note that there is no argument to be made that the more varied evidence is more positively dependent conditional on N than the less varied evidence. While that would make the varied evidence helpful to the confirmation of N if it were true, there appears to be no case for such a claim. Suppose that we knew with certainty that N was true and that we had in hand a terrestrial instance in which N was satisfied—say, the fact that the acceleration of an apple’s fall satisfies Newton’s inverse square law. Described in these terms, this observation is entailed by the truth of N. So, too, is the proposition that Jupiter and the sun interact in accordance with Newton’s law. Here it is useful to remember that we are not talking about the exact, empirical descriptions of the motions of the planet Jupiter or the fall of the apple. These must be filled in by observation. But the law entails that both terrestrial and celestial bodies will *act in accordance with the law*. So the terrestrial observation of the satisfaction of the law and the celestial observation of it are not positively dependent on the assumption of N, and neither are two terrestrial observations.¹⁶ They are all independent of each other, conditional on N. Therefore, the set consisting of the terrestrial and celestial observations are not more dependent, conditional on N, than the set consisting only of terrestrial observations. So the value of varied evidence here is not a result of increased coherence conditional on N. It does, however, appear to be well explained by the decreased coherence of the more varied evidence conditional on $\sim N$.

4.2. Medical testing

As Fitelson notes (2001: S133), medical testing is an area in which varied evidence seems particularly valuable. Suppose that we have an imaging result (e.g., an MRI) that shows a suspicious mass that might be a cancerous tumor. If the hypothesis under test is

16. Note that when I talk about “a terrestrial observation” or “a celestial observation” in this example, I am combining what are strictly speaking multiple *specific* empirical observations into one “observation” as an item of evidence—an instance in which Newton’s law seems to apply to the attraction between two bodies. One observes, say, the specific mass and starting point of the apple, its acceleration toward earth, the mass of the earth, and so forth, and that constitutes an observation of the apparent application of Newton’s law to the case of the apple’s fall. It is certainly true that one can break down these observations further in such a way that, taking some specific information together with N, one can predict what the remaining information will be. E.g., given N, one can predict the acceleration of the apple if one knows the relevant masses and distance. In that sense, N does render some specific empirical information positively relevant to some other specific empirical information. But one observation that N *seems to obtain* in a given case does not, conditional on N, make another such observation more likely. Moreover, the way in which N renders some highly specific items of empirical information relevant to others does not differentiate in any way between more varied and less varied sets of evidence. It is just as much the case for specific terrestrial observations as for the specific observations within the terrestrial and celestial realms. Hence, it cannot be the proper analysis of the special value of varied evidence.

C The patient has cancer,

it seems intuitively more useful to get the result of a different, at least equally accurate, type of test (such as a biopsy) than to take another image of the same kind of the same part of the body, even if the imaging test has a high rate of accuracy. Contrast a scenario in which the same kind of test (e.g., the imaging test) is done twice, imaging the same part of the body, with one in which two different kinds of tests are done.¹⁷ Suppose that both tests come back positive in both scenarios. What is the special confirmational virtue in the set of tests with greater intuitive variety?

Here matters are a bit complicated by the fact that, in real life, factors other than *variety* of evidence intrude. For example, doctors will often start with an individually less accurate but also less invasive test and do a more invasive, but more accurate, test only if the first test comes back positive. This means that the individual Bayes factors of the tests are not, in such real-life cases, actually equal. Repeating the first type of test would mean repeating a test with less accuracy as opposed to doing a second test that, if positive, has a stronger Bayes factor for C all by itself.

But since it is safe to say that in those real-life cases something other than varied evidence is causing the results to vary, let us imagine that the two kinds of medical tests really do have equal confirmational value for C *individually* if they come back positive. Why is the case from two different kinds of tests for the disease still intuitively stronger than the case from repeating the same test twice?

As in the analysis of varied evidence for Newton's law, the answer seems to lie in subhypotheses of $\sim C$ —in other words, the concern is about false positives. If there is a condition that mimics or “looks like” cancer, it is plausible that it will mimic it in a way that will repeatedly affect one type of test more than it would affect different types of tests. For example, if the first type of test is imaging, then we can think of

($\sim C$)₁ The patient does not have cancer, but the patient has a benign mass that looks like cancer on an MRI.

A positive MRI result confirms C, but it also confirms ($\sim C$)₁ conditional on $\sim C$. By the same reasoning given in the Newton example, this means that two positive

17. For purposes of this analysis, I would treat tests as being of “different kinds” if one took images of different parts of the body—e.g., checking for a tumor in some other place where it might appear. I owe this point to an anonymous reviewer. See Section 5 on telling what counts as a different kind of evidence. Images showing masses in different parts of the body would have less coherence on the assumption of $\sim H$ than images giving the same result for the same part of the body.

MRI results are somewhat positively relevant to each other conditional on $\sim C$. But this is not the case for an MRI result and a result from a test other than imaging, such as a biopsy. Absent some other, highly unusual background evidence, $(\sim C)_1$ does not account for a positive biopsy. So the positive results from the intuitively varied tests are more independent, or even entirely independent, conditional on $\sim C$ and hence confirm C more strongly than the less varied results.

The question arises at this point whether there is also value in the more varied evidence by way of increased coherence on the assumption of C . It is possible that varied evidence could have more than one positive effect upon confirmation, and showing that it decreases coherence conditional on $\sim C$ does not amount by itself to an argument that there are no other positive effects. But it seems that there are not. One way to see this is to imagine holding constant the degree of dependence conditional on C ; it seems that when we do that, the value of varying the tests remains. Suppose an idealized case, for example, in which the medical tests have a false *negative* rate of 0. That is to say, if there is cancer, the test will infallibly report it. There is, let us imagine, no such thing as cancer that is undetectable by these tests. That does not mean, however, that the tests are infallible across the board. They are still subject to false positives—positive results that do not really indicate cancer. By envisaging a false negative rate of 0, we make the probability of a positive result given C equal to 1, which makes the results entirely independent conditional on C . This makes it impossible to increase dependence conditional on C , whether by increasing the variety of the tests or in any other way. Nonetheless, even if the probability of each of the items of evidence conditional on C were 1, the value of varied tests seems to remain just as strong, because varied evidence helps to allay worries about ways in which there could be a false positive—e.g., the concern that the patient has a benign mass that causes imaging technology to register a positive result. The thought experiment of considering the items—both for non-varied and varied evidence—to be independent conditional on C suggests strongly that the entire value of varying the evidence lies in the decrease in coherence conditional on $\sim C$.

4.3. *Witness testimony*

C.I. Lewis suggests that witness testimony that agrees can strongly support the conclusion attested even when the individual testimonies offer only small support to that proposition:

The principle in question may be illustrated by the example of a number of witnesses, each of them not especially trustworthy as individual reporters, who independently tell the same circumstantial story. In case of such concurrence, one must quickly be convinced that what they tell

is practically certain. In similar fashion, the probability of an objective belief . . . may come to have very high probability, even on the basis of confirmations which, taken separately, might not warrant a particularly high degree of assurance. (1946: 239)

Our previous example of the relatively unreliable witnesses who independently tell the same circumstantial story, is another illustration of the logic of congruence; . . . For any one of these reports, taken singly, the extent to which it confirms what is reported may be slight. And antecedently, the probability of what is reported may also be small. But congruence of the reports establishes a high probability of what they agree upon, by principles of probability determination which are familiar: . . . (1946: 346)

Lewis's work has been extremely important in the field known as Bayesian coherence theory, which has sought to formalize these ideas about the power of independent witnesses to the same event.

The concept of independence was central to Lewis's idea, as shown both in the quotations above and here:

In general, that is the relationship of empirical beliefs we hold, so far as these are believed initially on grounds which are independent, and are not based on the same evidence, or the one of them believed merely because the others have already been accepted. Congruence of such beliefs may be a potent and valid ground of their credibility. (1946: 347)

Lewis was very explicit as to what he meant by independent testimony:

Speaking in terms of this familiar method, the considerations needing to be weighed in determining any particular relationship of congruence, would include the following . . . (3) the independence which the consequences have of one another (that is; whether, supposing ' H_1 ' false, the probability of a consequence . . . depends on that of another . . .) . . . (1946: 349, emphasis in original)

In other words, Lewis requires that the evidence must be independent conditional on the *negation* of the hypothesis under consideration.

It is especially easy to see why this is important for testimonial evidence. In the case of testimony, investigators and listeners are concerned about the possibility that witnesses have either colluded with each other deliberately or else have influenced one another's stories unconsciously. This is of concern for con-

firmation because, if such collusion or influence has occurred, the agreement among the stories may be explained by *something other than the truth of what is related*. In other words, there is a subhypothesis (or more than one subhypothesis) of $\sim H$ that would make the contents of the testimony mutually dependent. Even without the intent to deceive, if one witness hears another witness say that the perpetrator of the crime wore a red hat, the second witness may unconsciously manufacture an apparent memory of a red hat and may attest to it, even though there was no red hat. This makes it difficult to sort out fact from fiction and to find out what portion of the witness's story was actually seen by that witness. In the extreme case, multiple witnesses may deliberately agree upon a story in order to deceive, even when the story is false in its entirety.

It is interesting to note that in these scenarios the relevant subhypotheses of $\sim H$ actually make the items of testimony directly relevant to each other. This is a different way of producing coherence conditional on $\sim H$ from the scenarios discussed in earlier sections. In those sections, positive dependence on $\sim H$ arose from the existence of a subhypothesis of $\sim H$ that gave higher-than-average probability, conditional on $\sim H$, to the individual items of evidence. In that case, it was not necessary for the subhypothesis to be a "dependence subhypothesis." That is to say, it was not required that the items of evidence be dependent conditional on the subhypothesis itself. Dependence conditional on $\sim H$ arose from other probabilistic factors. In the case of mutual influence or collusion among witnesses, however, those subhypotheses *actually state* that the items of evidence are dependent on one another: One person has been influenced by another's story, or witnesses are deliberately agreeing upon a story. Suppose that the hypothesis of interest is

J Jones robbed First Federal Bank

and that the subhypothesis of collusion is

($\sim J$)₁ Jones did not rob First Federal Bank, and Smith and Robinson have colluded to testify falsely that they saw him do so.

In that case, where E_1 and E_2 are the testimonies of Smith and Robinson to having seen Jones rob the bank,

$$P(E_1 | E_2 \ \& \ (\sim J)_1) > P(E_1 | (\sim J)_1).$$

This is what makes it the case that

$$P(E_1 | E_2 \ \& \ \sim J) > P(E_1 | \sim J),$$

though if $P((\sim J)_1 | \sim J)$ is very small, the positive dependence of the evidence conditional on $\sim J$ will not be very great. (See discussion of this point in Section 5 of this paper.)

As there is in general more than one way for items of evidence to be dependent conditional on the negation of the hypothesis, so it is for testimony. Shogenji (2013: 2529) has pointed out that even if collusion and other forms of direct dependence between testimonies are treated as having probability 0, items of testimony may still be dependent conditional on the negation of the hypothesis in question. For example, suppose that the hypothesis of interest is

M Jones murdered Johnson

and that the two witnesses, Smith and Robinson, state that

W The murder weapon belongs to Jones.

Let us take it that W is unproblematically positively relevant to M and that E_1 and E_2 , Smith's and Robinson's testimonies to W, have confirmatory Bayes factors not only for W but also for M, because they confirm W. Even if we somehow rule out all collusion and influence between Smith and Robinson concerning W—that is, the content of their own testimonies—this does not *ipso facto* make it the case that E_1 and E_2 are independent conditional on $\sim M$. For it is quite possible that

$(\sim M)_1$ Jones did not murder Johnson, the murder weapon belongs to Jones, but someone else has framed Jones by using his weapon to commit the murder.

E_1 and E_2 both confirm not only M but also confirm $(\sim M)_1$ conditional on $\sim M$, creating some degree of dependence conditional on $\sim M$, though the degree of that dependence will depend upon $P((\sim M)_1 | \sim M)$. This case, the reader will notice, is exactly parallel to the scenarios envisaged for Newton's law and medical testing above.

Witness testimony provides us with a helpful field of examples concerning varied evidence, because witness testimony has semantic content. We can hypothetically vary the testimony more or make it more similar fairly readily, which is useful for testing an account of evidential variety. Consider the concern about collusion, modeled by $(\sim J)_1$. It is quite easy to see that variation in the details of Smith's and Robinson's testimonies can help to assure the investigators that $(\sim J)_1$ is not the true explanation of the similarities (in general outline) between their claims. These variations need not be actual contradictions, though they may be

(see below). If Smith emphasizes that the perpetrator was tall, heavy, and wore a red shirt, while Robinson makes no mention of any of these details but emphasizes instead that the perpetrator walked with a limp, had a grating voice, and wore a brown jacket, and if they both pick Jones out of a lineup as the person whom they saw robbing the bank, this variation of details together with agreement on the central point (that Jones robbed the bank) is not well explained by $(\sim J)_1$. It is possible in theory that Smith and Robinson carefully chose varying details to emphasize in order to make their collusion more believable, but that is not what we would most expect if they were colluding. Indeed, it would seem at least somewhat more likely that they would be careful to coordinate their details rather than varying them.

The case is even more interesting if the witnesses actually appear to contradict one another concerning some minor detail. If one says that the robber's shirt was red while the other says that it was blue, but both agree in picking out Jones as the perpetrator, this is *extremely* unlikely given $(\sim J)_1$. While counsel for the defense may make much of the apparent contradiction concerning shirt color, the confirmational value of disconfirming collusion (or even significant influence)—the value of making it unlikely that something *other than J* is the true explanation of the witnesses' agreement—is much more important than the possibility that a witness has made an error about a minor detail like shirt color. On this point, the comments of the 19th-century English jurist Thomas Starkie are particularly relevant:

It is here to be observed, that partial variances in the testimony of different witnesses, on minute and collateral points, although they frequently afford the adverse advocate a topic for copious observation, are of little importance, unless they be of too prominent and striking a nature to be ascribed to mere inadvertence, inattention, or defect of memory.

It has been well remarked by a great observer, that "the usual character of human testimony is substantial truth under circumstantial variety." It so rarely happens that witnesses of the same transaction perfectly and entirely agree in all points connected with it, that an entire and complete coincidence in every particular, so far from strengthening their credit, not unfrequently engenders a suspicion of practice and concert. (1888: 488-89)

In many instances one can plausibly argue that, if the witnesses actually appear to contradict one another concerning some minor detail while agreeing about the central point represented by *J*, their testimonies (taken in their details) are actually *negatively* relevant to one another given $(\sim J)_1$. That is to say, if they are colluding, the details of one witness's evidence lead one to expect that the

other witness's evidence, whatever else it contains, will *not* contain an outright contradiction of the first witness's testimony. In that case, depending on what other subhypotheses of $\sim J$ are under consideration, it may well be the case that the variations in the witnesses' testimonies render them not merely independent on $\sim J$ but actually *negatively* dependent conditional on $\sim J$.

Here it is worth emphasizing that such evaluations are highly sensitive to background knowledge and hence partake of the messiness of empirical investigation. Starkie hints at this point when he adds the qualification "unless they be of too prominent and striking a nature to be ascribed to mere inadvertence, inattention, or defect of memory." Consider the well-known probabilistic example of four students who ask their professor to let them make up a recently missed exam because, they allege, they were traveling together over the weekend and were delayed by a flat tire which they had to change. The professor, suspecting that they have a less defensible reason for their joint absence, cannily puts them in separate rooms and gives them a single question to answer: "Which tire?" Tacitly, the professor is assuming (rightly or wrongly) that, if they colluded, they probably did *not* think to coordinate their stories that carefully prior to his isolating them but that, if they are telling the truth, the event has been recent enough and vivid enough that they should all be able to remember which tire went flat. In that case, the professor is taking it that a contradiction among them on the question of the tire actually *supports* collusion—specifically, a type of sloppy attempted collusion which he, knowing the students, considers quite plausible.

Whether or not outright contradictions among witness testimonies are such that they make us question the witnesses' reliability or veracity rather than decreasing (or only decreasing) the probability of collusion among them is not something that there are simple rules for, because it is in the nature of the case an empirical matter. The proof given above indicates that, when that happens, either coherence conditional on $\sim H$ has increased rather than decreasing or else one of the *ceteris paribus* conditions has been violated, and possibly both.¹⁸ My contention is not that anything that one might call "variety among witness testimonies" (including any contradiction) is epistemically valuable. Rather, I intend to offer an account of intuitively valuable variety in evidence. So I contend that when it *does* seem that there is a factor present in the epistemic situation of witness testimony that we are inclined to consider the "value of varied evidence," it is properly analyzed in terms of lessened coherence of the evidence conditional on the negation of the hypothesis of interest. Moreover, it can be useful to recognize the *possibility* that variation or minor contradiction among testimonies is

18. One plausible idea is that outright contradiction between testimonies decreases, at least in some interesting cases, the coherence of the evidence conditional on $\sim H$ and also conditional on H and that the helpfulness or harmfulness of the contradictions for the confirmation of H depends on whether the decrease in coherence induced by the contradictions is greater for H or for $\sim H$.

epistemically valuable, as Starkie emphasizes, in order not to get “hung up” over minor variations or contradictions that ought rather to be, in that situation, regarded as epistemically helpful to the confirmation of the central point at issue.

Given that many conceivable instances of the value of varied evidence in witness testimony can be understood in terms of reduced coherence on $\sim H$, the question arises whether there is any other way in which varied testimonial evidence is of significant confirmational value. Testimony is different from the earlier cases we have examined because, in most realistic scenarios, testimonial evidence (especially in its details) does not have particularly high probability conditional on H . Even if the event attested to actually took place, it would be highly artificial to assign a given witness’s testimony a probability of 1 conditional on the event. A given individual may not have had an opportunity to see the event clearly or may not wish to speak out about it. And even if a witness does have access to the event and is forthcoming about it, the specific details of his testimony are not by any means predictable merely given the general truth of the story. Hence, while individual witness testimony will typically have confirmational value for the truth of what is attested, it may do so by comparison with the individual probability of the individual testimony conditional on $\sim H$, not because of an absolutely high probability conditional on H . When we are using the l measure of confirmation, the individual Bayes factor is a ratio of $P(E_1|H)$ to $P(E_1|\sim H)$, so of course absolutely high $P(E_1|H)$ is not required. The possibility of low individual conditional probability for a witness’s testimony on H leaves room for the possibility of positive dependence among testimonies on the assumption that H is true. This leads in turn to the question as to whether such consilience constitutes a significant part of the intuitive value of *diverse* testimonial evidence in addition to the value of independence conditional on $\sim H$. That is to say, when we intuitively say that testimonial evidence is confirmationally valuable *because* it is diverse, does any significant part of the analysis of that value lie in coherence of the testimonies on the assumption of H ?

There seems to be one type of case in which we can contrast less varied with more varied evidence and see *some* of the value of the increased variety in coherence on H : Consider a situation in which one single witness, who affirms the event described in H , repeats the same story multiple times in approximately the same words. This seems, intuitively, to be an *extremely* unvaried set of evidence. Contrast that with a scenario in which multiple different witnesses (the same number as the set of repetitions in the first scenario) affirm H , also using approximately the same words as each other. This is not what we would intuitively call an extremely varied evidence set. But at least it involves different witnesses rather than repetitions by the same witness, and to that extent it is more varied than the scenario in which one witness repeats his story. One can argue that *some* part of the confirmational value of moving from one witness to multiple

witnesses (depending as always on the specifics) could be this: We might be unsure whether the event described in H was accessible to witnesses at all and whether people would be inclined to talk about it if it were. Conditionalizing on the first witness's testimony confirms not H alone, but conditional on H confirms H_1 , something like, "The event [described in H] occurred, was accessible to witnesses, and witnesses are willing to talk about it." H_1 , in turn, gives us more reason than H itself to expect another testimony to the same event. Hence, by confirming H_1 conditional on H , the relatively greater variety included in multiple witness testimonies produces *some* dependence conditional on H . And this is somewhat helpful to the confirmation of H by (E_1 & E_2).

But a moment's thought shows that a *far* larger portion of the value of multiple witnesses as compared to repetition by one witness should be attributed to our old friend, reduced coherence conditional on $\sim H$. For if a single witness is telling a lie (is insane, has been deceived, etc.), a repetition of the story in approximately the same words is what we would expect given that the story is false. Repetitions of a single witness's story have a *very* high level of dependence given $\sim H$, intuitively far more than the testimony of multiple witnesses. For multiple witnesses one must at least raise other concerns, such as those about mutual influence or collusion, in order to produce dependence of their testimonies on $\sim H$. So even though some slight confirmational value of greater variety in this *specific* shift from one witness to multiple witnesses lies "on the H side," the significant value of variety for confirmation in this case lies, as in the other examples, "on the $\sim H$ side."¹⁹

The analysis of the value of evidential variety in witness testimony confirms the conclusion that, when it seems intuitively that varied evidence is valuable to confirmation, the virtue of variety lies in rendering evidence less coherent conditional on $\sim H$.

The analysis I have given of varied evidence in the Newtonian case, the case of medical testing, and the case of witness testimony can readily be extended to other cases that I have not discussed, such as the confirmation of Lavoisier's oxygen theory by varied evidence (Thagard 1978: 77–78, 81), Sober's contrast (1989) between two reports from the same newspaper and a newspaper and radio account, and even Paul Horwich's curve-fitting example (Horwich 1982: 119–122).

19. It is possible to show using probabilistic modeling that a little dependence goes farther in influencing the confirmational value of a set of evidence if the items of evidence are dependent conditional on the hypothesis that gives lower probability to the *individual* items of evidence. By "dependence" *just here* I mean the extent to which the items of evidence confirm one another conditional on the hypothesis, which affects the numerator of the conditional coherence ratio for that hypothesis. The conditional coherence ratio for the hypothesis giving lower individual confirmations is especially sensitive to an increase in its own numerator. I have shown some of this modeling in Lydia McGrew (forthcoming).

5. The Difficulty of Defining Varied Kinds of Evidence

Not only does this analysis cover a wide variety of canonical cases of the value of varied evidence, it also explains the difficulty in defining “classes” or “kinds” of evidence and hence in defining the concept of variety itself (see Thagard 1978: 80), while offering a flexible solution to that problem. If evidential variety is, as this analysis implies, both a probabilistic and a comparative concept, we would expect that what counts as variety would vary along a continuum, with different evidential backgrounds, and between different comparison classes. Multiple witnesses who all say the same thing are more varied than one witness repeating himself, but multiple witnesses whose testimonies vary in incidental details while agreeing on the central point may be more varied still, and more valuable. This is explicable in terms of decreasing positive relevance (and even negative relevance) conditional on $\sim H$. In some evidential contexts, the fact that a scientific theory appears to apply to different metals will be of very little special help in confirming the theory; it will not seem like significantly varied evidence. But when testing another scientific theory the apparent applicability of the theory to a variety of metals will be extremely important. Thagard (1978: 80) suggests that the division of facts into relevantly varied classes must be simply accepted by the philosopher of science based upon the pragmatic agreement of working scientists. But while that may do as a rough approximation of probabilistic relevance, the philosopher will prefer a more principled account. My account indicates that the division of evidence into significantly varied kinds or classes depends upon the subhypotheses of $\sim H$ that arise in testing each theory.

Wesley Salmon notes the difficulty of deciding what sort of variety is relevant, and his analysis of the value of variety bears an interesting similarity to mine, though he does not state his conclusions in terms of conditional evidential dependence.

[T]here is a fundamental difficulty in the very concept of variety. Any observation is different from any other in an unlimited number of ways, and any observation is similar to any other in an unlimited number of ways. It is therefore necessary to characterize similarities and differences that are relevant to confirmation. I suggest the following approach. A general hypothesis has a certain domain of applicability, and the basic idea behind the variety of instances is to test the hypothesis in different parts of its domain. It is always possible to make arbitrary partitions of the domain, but a splitting of the domain is significant only if it is not too implausible to suppose that the hypothesis holds in one part of the domain but not in another. Now we could strongly insist upon having observations of Mars on Tuesdays and Sundays as well as the other days

of the week, in months whose names contain the letter “r,” . . . etc. However, we do not find it plausible to suppose that Newton’s law holds for Mars in some of these subdomains but not in others. By contrast, it is not completely absurd to suppose that Newton’s law would be suitable for bodies of astronomic dimensions located at astronomic distances from one another, but that it does not hold for smaller masses and shorter distances The variety of instances helps us to eliminate other hypotheses, but such elimination has a point only if the alternative hypotheses being tested have nonnegligible prior probabilities. (Salmon 1966: 130–31)

Salmon relates this point to the importance of prior probabilities in Bayesian probability theory. But the analysis given here allows us to go farther than this. The unimportance for confirmation of merely *conceivable* subhypotheses of $\sim H$ that have extremely low probability is only indirectly related to their absolutely low prior probability. More important is their improbability conditional on $\sim H$.

My analysis suggests that the reason that such subhypotheses of $\sim H$ (e.g., that Newton’s law is false but appears to apply on certain days of the week) are unproblematic is that their extremely low probability conditional on $\sim H$ makes them insignificant when it comes to rendering items of evidence dependent conditional on $\sim H$. Hence, we need not bother to make special attempts to rule them out by testing. If we happen to have made all of our Newtonian observations on Monday–Friday, we can invent a contrived subhypothesis of $\sim N$ according to which Newton’s law is false, taken universally, but appears to apply on weekdays.²⁰ Technically, our observations thus far are rendered dependent on $\sim N$ to a negligible degree by this proper subhypothesis of $\sim N$. But for purposes of inference this can be disregarded because of the overwhelmingly low probability of this subhypothesis conditional on $\sim N$. If N were false, there is no reason whatsoever to think that some cause would make it appear to be true on weekdays.

Similarly, consider the subhypothesis

($\sim N$)₃ Newton’s universal law is false, but all naturally occurring events, not directly caused by human action, appear to obey Newton’s law.²¹

This subhypothesis would mean that the fall of the apple and the movements of the planets are expected to appear to obey Newton’s law, but it would not explain the fall of the cannon ball—thus splitting the terrestrial instances apart

20. This type of example involving a highly contrived subhypothesis of $\sim H$ is closely connected to the analysis of ad hocness. See Lydia McGrew (2014).

21. I owe this suggestion to an anonymous reviewer.

from one another and uniting one of them to a celestial instance. But our background evidence is strongly against such an hypothesis; we regularly notice that the motions induced by human action (the movements of dropped objects, for example) do not appear to behave in a notably different way, given the initial conditions, from events that occur spontaneously in the natural realm (the fall of an apple from a tree). This effect of background knowledge in making the probability of $(\sim N)_3$ conditional on $\sim N$ negligible explains the fact that we intuitively consider the celestial and terrestrial instances to be more varied than the two terrestrial instances.

The relevance of the probability of subhypotheses conditional on $\sim H$ also shows why, even given evidence that is *somewhat* varied in comparison to other evidence, real or hypothetical, we would prefer evidence that is even *more* varied. In the case of celestial and terrestrial instances of Newton's law, there is also the possibility of

- $(\sim N)_4$ Newton's universal law is false, but there is a property of all objects in the solar system that causes them to appear to obey Newton's law.²²

This subhypothesis does not seem overwhelmingly improbable given $\sim N$. It would, conditional on $\sim N$, deal with both the celestial and terrestrial instances within the solar system, but it would not explain a finding that distant stars also appear to obey Newton's law. So, while celestial and terrestrial instances within the solar system are understandably regarded as more valuable in virtue of their variety than an otherwise equal set consisting only of terrestrial instances, a set of evidence including instances from greater distances would plausibly be more varied still and hence more valuable.

The difficulty in defining variety arises, then, from the conditions of empirical inquiry which make variety relative both to the hypothesis under consideration and to the evidential background relevant to alternative hypotheses. What counts as significant variety will depend upon which alternative hypotheses have non-negligible probabilities conditional on the negation of the hypothesis in question and what work those subhypotheses do to make evidence dependent conditional on the negation.

The analysis given here explains both canonical examples of the value of evidential diversity and puzzles that arise in applying the concept to particular cases. It gives us tools, such as the concept of a subhypothesis and the focus on coherence conditional on $\sim H$, with which to evaluate those puzzles and to make our thoughts about them more precise. It has a good claim, therefore, to being a

22. This subhypothesis was suggested by an anonymous reviewer.

correct and perhaps even a comprehensive probabilistic account of the intuitive value of evidential diversity.

Appendix: The Confirmational Value of Reduced Coherence on $\sim H$ Using the r Measure

Suppose that we are using the r measure of confirmation in its simple ratio form. Then the confirmation that a conjunction of evidence ($E_1 \& \dots \& E_n$) gives to H is, by definition,

$$\frac{P(H|E_1 \& \dots \& E_n)}{P(H)}.$$

I assume the *ceteris paribus* conditions 1 and 2 specified for the proof for measure l in the body of this paper for two bodies of evidence $\{E_1, \dots, E_n\}$ and $\{E_1', \dots, E_n'\}$ and an hypothesis H .²³ For this result I add

3) The prior probability of H is held constant when comparing the confirmational value of the two bodies of evidence.

By condition 3,

$$\frac{P(H|E_1 \& \dots \& E_n)}{P(H)} < \frac{P(H|E_1' \& \dots \& E_n')}{P(H)}$$

just in case

$$P(H|E_1 \& \dots \& E_n) < P(H|E_1' \& \dots \& E_n').$$

By the odds form of Bayes's Theorem,

$$\frac{P(H|E)}{P(\sim H|E)} = \frac{P(E|H)}{P(E|\sim H)} \times \frac{P(H)}{P(\sim H)}$$

and condition 3, it follows that

$$P(H|E_1 \& \dots \& E_n) < P(H|E_1' \& \dots \& E_n')$$

just in case

23. Assume all probabilities in this proof to be greater than zero.

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 \& \dots \& E_n | \sim H)} < \frac{P(E_1' \& \dots \& E_n' | H)}{P(E_1' \& \dots \& E_n' | \sim H)}.$$

As shown in the body of the paper, *ceteris paribus*,

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 \& \dots \& E_n | \sim H)} < \frac{P(E_1' \& \dots \& E_n' | H)}{P(E_1' \& \dots \& E_n' | \sim H)}$$

just in case

$$P(E_1 \& \dots \& E_n | \sim H) / P(E_1 | \sim H) \times \dots \times P(E_n | \sim H) > P(E_1' \& \dots \& E_n' | \sim H) / P(E_1' | \sim H) \times \dots \times P(E_n' | \sim H)$$

Therefore, *ceteris paribus*,

$$\frac{P(E_1 \& \dots \& E_n | H)}{P(E_1 \& \dots \& E_n | \sim H)} < \frac{P(E_1' \& \dots \& E_n' | H)}{P(E_1' \& \dots \& E_n' | \sim H)}$$

just in case

$$P(E_1 \& \dots \& E_n | \sim H) / P(E_1 | \sim H) \times \dots \times P(E_n | \sim H) > P(E_1' \& \dots \& E_n' | \sim H) / P(E_1' | \sim H) \times \dots \times P(E_n' | \sim H)$$

Under the given *ceteris paribus* conditions, the confirmation of H by $(E_1' \& \dots \& E_n')$, according to the *r* measure, is greater than the confirmation of H by $(E_1 \& \dots \& E_n)$ just in case the items of evidence in the set $\{E_1, \dots, E_n\}$ are more coherent conditional on $\sim H$ than the items of evidence in the set $\{E_1', \dots, E_n'\}$. Therefore, decreased coherence of evidence conditional on $\sim H$ is valuable to confirmation, all else being equal, according to the *r* measure of confirmation.²⁴

References

- Ahmed, Arif (2015). Hume and the Independent Witnesses. *Mind*, 124(496), 1013–1044. <http://dx.doi.org/10.1093/mind/fzv076>
- Carnap, Rudolf (1945). On Inductive Logic. *Philosophy of Science*, 12(2), 72–97. <http://dx.doi.org/10.1086/286851>
- Earman, John (1992). *Bayes or Bust*. MIT Press.
- Fitelson, Branden (2001). A Bayesian Account of Independent Evidence With Applications. *Philosophy of Science*, 68(3), Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association, S123–S140. <http://dx.doi.org/10.1086/392903>

24. I thank Timothy McGrew for pointing out that this result could be shown more concisely and with fewer *ceteris paribus* conditions than in an earlier version I proposed.

- Horwich, Paul (1982). *Probability and Evidence*. Cambridge University Press.
- Howson, Colin and Peter Urbach (1993). *Scientific Reasoning: The Bayesian Approach*. Open Court.
- Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. Open Court.
- McGrew, Lydia (forthcoming). Accounting for Dependence: Relative Consilience as a Correction Factor in Cumulative Case Arguments, *Australasian Journal of Philosophy*.
- McGrew, Lydia (2014). On Not Counting the Cost: Ad Hocness and Disconfirmation. *Acta Analytica*, 29(4), 491–505. <http://dx.doi.org/10.1007/s12136-014-0225-9>
- McGrew, Lydia and Timothy McGrew (2009). The Argument from Miracles: A Cumulative Case for the Resurrection of Jesus of Nazareth. In William Lane Craig and James Porter Moreland (Eds.), *The Blackwell Companion to Natural Theology* (593–662). Wiley-Blackwell. <http://dx.doi.org/10.1002/9781444308334.ch11>
- McGrew, Timothy (2003). Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science*, 54(4), 553–67. <http://dx.doi.org/10.1093/bjps/54.4.553>
- Myrvold, Wayne (1996). Bayesianism and Diverse Evidence. *Philosophy of Science*, 63(4), 661–665. <http://dx.doi.org/10.1086/289983>
- Myrvold, Wayne (2003). A Bayesian Account of the Virtue of Unification. *Philosophy of Science*, 70(2), 399–423. <http://dx.doi.org/10.1086/375475>
- Nagel, Ernest (1939). *Principles of the Theory of Probability*. University of Chicago Press.
- Novack, Greg (2007). Does Evidential Variety Depend On How the Evidence Is Described? *Philosophy of Science*, 74(5), Proceedings of the 2006 Biennial Meeting of the Philosophy of Science Association, 701–711. <http://dx.doi.org/10.1086/525615>
- Salmon, Wesley C. (1966). *The Foundations of Scientific Inference*. University of Pittsburgh Press.
- Salmon, Wesley C. (1990). Rationality and Objectivity in Science or Tom Kuhn meets Tom Bayes. In C. Wade Savage (Ed.), *Scientific Theories: Minnesota Studies in the Philosophy of Science* (Vol. 14, 175–204). University of Minnesota Press.
- Schupbach, Jonah (2005). On a Bayesian Analysis of the Virtue of Unification. *Philosophy of Science*, 72(4), 594–607. <http://dx.doi.org/10.1086/505186>
- Shogenji, Tomoji (1999). Is Coherence Truth-conducive? *Analysis*, 59(4), 338–345. <http://dx.doi.org/10.1093/analys/59.4.338>
- Shogenji, Tomoji (2013). Coherence of the Contents and the Transmission of Probabilistic Support. *Synthese*, 190(13), 2525–2545. <http://dx.doi.org/10.1007/s11229-011-0003-9>
- Sober, Elliot (1989). Independent Evidence about a Common Cause. *Philosophy of Science*, 56(2), 275–87. <http://dx.doi.org/10.1086/289487>
- Starkie, Thomas (1876). *A Practical Treatise of the Law of Evidence*. T. & J. W. Johnson.
- Thagard, Paul R. (1978). The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy*, 75(2), 76–92. <http://dx.doi.org/10.2307/2025686>
- Whewell, William (1847). *The Philosophy of the Inductive Science, Founded Upon Their History*. John W. Parker.