THE VARIETY-OF-EVIDENCE THESIS: A BAYESIAN EXPLORATION OF ITS SURPRISING FAILURES

FRANÇOIS CLAVEAU AND OLIVIER GRENIER

ABSTRACT. Diversity of evidence is widely claimed to be crucial for evidence amalgamation to have distinctive epistemic merits. Bayesian epistemologists capture this idea in the variety-of-evidence thesis: *ceteris paribus*, the strength of confirmation of a hypothesis by an evidential set increases with the diversity of the evidential elements in that set. Yet, formal exploration of this thesis has shown that it fails to be generally true. This article demonstrates that the thesis fails in even more circumstances than recent results would lead us to expect. Most importantly, it can fail whatever the chance that the evidential sources are unreliable. Our results hold for two types of degrees of variety: reliability independence and testable aspect independence. We conclude that the variety-of-evidence thesis can, at best, be interpreted as an exception-prone rule of thumb.

Keywords. evidence amalgamation; Bayesian epistemology; evidence variety; evidence independence; robustness; triangulation

1. INTRODUCTION

Amalgamating evidential elements is something worth pursuing in a wide array of circumstances. Scientists – from physics (Istituto Nazionale di Fisica Nucleare, 2011; Claveau, 2013, p. 94) to economics (Downward and Mearman, 2007) and nursing science (Thurmond, 2001) – praise this practice of combining answers from various sources.

One condition generally believed to be necessary for evidence amalgamation to generate significant epistemic benefits is that the elements brought together be 'varied' or, in other words, that they be "in some relevant sense or senses 'independent' of each other" (Woodward, 2006, p. 234). The scare quotes in the previous sentence indicate that what variety or independence are in this context and how they relate to the success of evidence amalgamation is not altogether clear. These issues are the subject of active philosophical investigation.[1]

Among Bayesian epistemologists, the alleged virtuous relationship between evidential variety and evidence amalgamation is summarized in the variety-of-evidence thesis.

> **Variety-of-evidence thesis:** *Ceteris paribus*, the strength of confirmation of a hypothesis by an evidential set increases with the diversity of the evidential elements in that set (Claveau, 2013, p. 95).

[1] The most recent research includes Schupbach (2015); Kuorikoski and Marchionni (2016); Stegenga and Menon (2017).

Although some scholars have claimed to be able to formally prove this thesis (Earman, 1992; Howson and Urbach, 1993), the wind changed course with the publications of results by Bovens and Hartmann (2002, 2003): in their Bayesian model of scientific inference, 'variety' does not uniformly lead to more confirmation. Some assumptions of Bovens and Hartmann's model have been challenged by one of us (Claveau, 2013), but the variety-of-evidence thesis ultimately does not universally hold in his own model.

At the end of Claveau's article , it is suggested that, even if one accepts his assumptions, "the conclusion to draw about the variety-of-evidence thesis is not straightforward." (Claveau, 2013, p. 103) Indeed, Claveau finds that, in his model, the variety-of-evidence thesis is false only if it is highly likely that the evidential sources are *unreliable* (at least 82 % likely). He thus proposes a strategy to save the variety-of-evidence thesis from refutation: it "could be interpreted as implicitly assuming that the evidential sources are sufficiently trustworthy to begin with."

This proposition cries out for further exploration. Since Claveau's model is far from being fully general, a research priority should be to investigate whether, in different evidential structures, a high probability of unreliable sources is necessary for the variety-of-evidence thesis to break down. The primary goal of this article is to report on such an investigation. We show that Claveau's proposal to save the variety-of-evidence thesis from refutation by stipulating that reliability must be minimally probable is a dead end: in extensions of his model, we find that the variety-of-evidence thesis can sometimes break down even if evidential sources are most probably reliable.

More generally, this article contributes to a better understanding of how (if at all) variety contributes to confirmation by studying, like Bovens and Hartmann (2003) and Claveau (2013), different evidential structures (i.e., how evidential elements and a hypothesis relate to each other). Like them we represent evidential structures with the help of Bayesian networks (Pearl, 2009). This strategy allows us to go beyond arguments based on intuitive considerations by literally computing how much confirmation a given evidential set should grant to a hypothesis. Although this strategy has great benefits according to us, the usual caveat about modeling applies. Models are idealizations. Their conclusions might fail to be real-world relevant if they are driven by distortions introduced in the idealization process.

The next section describes the assumptions of our model and offer two interpretations of the notion of evidential variety: variety as unreliability independence and variety as testable aspect independence. In section 3, we assess the variety-of-evidence thesis by means of pairwise comparisons among evidential structures. Section 4 pushes further a method introduced by Claveau (2013, sec. 5) to model *degrees* of variety instead of only comparing the extremes of minimal and maximal variety. Our results offer little hope that the variety-of-evidence thesis can be justified if it is interpreted as an exceptionless generalization. Notably, testable aspect independence is decisive to the disconcerting conclusion we draw here: when added to Claveau's (2013) model, cases of the variety-of-evidence thesis breaking down occur even if the sources are considered most probably reliable. But the thesis can still have action-guiding force – at least, that is what we suggest in the conclusion.

2. Setting up the model

The possible relationships between a hypothesis and its evidence are diverse. The different structures can be represented by Bayesian networks that include, among others, a hypothesis variable H and a set of evidential variables E_i. In this article, we study a class of structures in which, if we interpret the Bayesian network causally (Pearl, 2009), the evidential variables are effects of the hypothesis. This is the class on which the Bayesian literature assessing the variety-of-evidence thesis has focused (Bovens and Hartmann, 2002, 2003; Claveau, 2013).[2]

Table 1 represents different structures that are studied by Bovens and Hartmann (2003, ch. 4). As one can see, each evidential variable E_i is an effect of H, but only indirectly through a variable C, which can be interpreted as a testable consequence of the hypothesis. Given his specific goal, Claveau (2013) studied even simpler structures in which testable consequences do not enter. In his models, the hypothesis variable is thus a direct cause of the evidential variables. In this section, we use structures with testable consequences. The analyzed structures are thus the ones used by Bovens and Hartmann. Yet, as we will describe below, we accept Claveau's argument with respect to the problematic assumption used by Bovens and Hartmann to model unreliable sources of evidence. Our contribution is thus to extend Claveau's analysis using the more complex evidential structures initially proposed by Bovens and Hartmann (but without their problematic assumption regarding unreliable sources).

The evidential structures use four types of propositional variables :

- The hypothesis variable $H = \{h, \neg h\}$, where h stands for the proposition that the hypothesis is true and $\neg h$ stands for its negation.
- The testable consequence variable $C_i = \{c_i, \neg c_i\}$, where c_i stands for the proposition that a testable consequence of h holds and $\neg c_i$ stands for its negation.
- The evidential variable $E_i = \{e_i, \neg e_i\}$, where e_i stands for a positive report regarding the hypothesis, that is, a report to the effect that a testable consequence of h holds, and $\neg e_i$ stands for a negative report.
- The reliability variable $R_i = \{r_i, \neg r_i\}$, where r_i stands for the proposition that the evidential source i (the one having as output E_i) is reliable and $\neg r_i$ stands for the proposition that the source is unreliable.
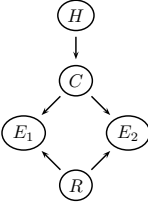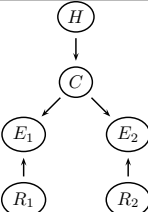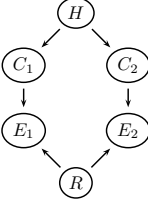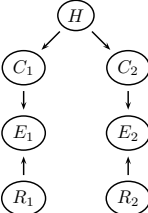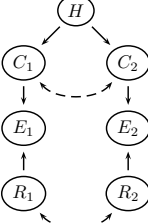
All structures in table 1 are associated to a joint probability distribution over the set of variables $\{H, C_1, C_2, E_1, E_2, R_1, R_2\}$, where the subscript on C and R are omitted when the two possible instances of each type of variable are collapsed into a single variable. The probabilistic independencies among the variables can be read off the Bayesian networks using the d-separate criterion (Pearl, 1988, pp. 117-18). For instance, all five structures share:

- $R_i \perp\!\!\!\perp H, C_j$ for all i, j in $\{1, 2\}$, which means that, before learning the evidential report E_i, learning that the associated evidential source is reliable

---

[2] Wheeler and Scheines (2011, 2013) propose different evidential structures, but do not assess the variety-of-evidence thesis with them. Stegenga and Menon (2017) provide a partial assessment limited to noting cases of what they call "dyssynergystic evidence". We had results for evidential structures in which the hypothesis is caused by other variables rather than being a root node, but we decided not to present them in this article due to space constraints. We can however report that the variety-of-evidence thesis is not better supported in this alternative class of evidential structures.

Table 1. *Diverse evidential structures with Cs as testable consequences of the hypothesis. The last column gives the likelihood ratios for each structure according to the assumptions spelled out in section 2.*

	Type of diversity	Network representation	Likelihood ratio
A	Shared consequence Shared reliability	$H \to C \to E_1, E_2 \to R$	$L_A = \frac{qr + \alpha\bar{r}}{pr + \alpha\bar{r}}$
B	Shared consequence Independent reliability	$H \to C \to E_1 \leftarrow R_1,\ E_2 \leftarrow R_2$	$L_B = \frac{qr^2 + 2qr\alpha\bar{r} + (\alpha\bar{r})^2}{pr^2 + 2pr\alpha\bar{r} + (\alpha\bar{r})^2}$
C	Independent consequences Shared reliability	$H \to C_1 \to E_1,\ C_2 \to E_2 \to R$	$L_C = \frac{q^2 r + \alpha\bar{r}}{p^2 r + \alpha\bar{r}}$
D	Independent consequences Independent reliability	$H \to C_1 \to E_1 \leftarrow R_1,\ C_2 \to E_2 \leftarrow R_2$	$L_D = \frac{(qr + \alpha\bar{r})^2}{(pr + \alpha\bar{r})^2}$
G	General structure	$H \to C_1 \to E_1 \leftarrow R_1,\ C_2 \to E_2 \leftarrow R_2$	$L_G = \frac{\omega_{rr}(q^{1+\gamma}) + \omega_{rh}(q^{1+\gamma} - \bar{q}^{(1+\gamma)} + 1) + \omega_{hh}}{\omega_{rr}(p^{1+\gamma}) + \omega_{rh}(p^{1+\gamma} - \bar{p}^{(1+\gamma)} + 1) + \omega_{hh}}$

or not has no effect on the strength of belief in the hypothesis nor on the strength of the belief that the testable consequences hold.

- $E_1, E_2 \perp\!\!\!\perp H | C_1, C_2$, which means that, once the truth or falsity of the testable consequences is known, learning the evidential reports has no effect on the strength of belief in the hypothesis.

The four structures differ on two dimensions: to what extent reliability and testable consequences are shared by evidential elements. Each dimension captures according to us an acceptable, yet partial specification of the notion of 'variety' in the variety-of-evidence thesis.[3] Firstly, there are various reasons why one evidential report might fail to track the state of a testable consequence (e.g., auxiliary hypotheses might be flawed, some manipulations in the experiment might introduce errors, the researchers might simply fabricate results). If two reports share little of these reasons (e.g., they are provided by different research teams, they use different techniques), learning that one report is unreliable should not make us substantially revise our belief that the other report is also unreliable. The two reports thus have high variety in the following sense:

Variety as (un)reliability independence: Evidential elements are varied if most reasons why one element would be unreliable are not shared by the other element(s).

In table 1, structures A and C are less varied in this sense than structures B and D because reports in the former have *fully shared reliability* ($R_1 = R_2 = R$) while the latter have *fully independent reliability* ($R_1 \perp\!\!\!\perp R_2$).

Secondly, evidential elements might be more or less varied depending on whether they report on different aspects related to the hypothesis (the C) or on exactly the same aspect. In this article, the 'aspects' are testable consequences of the hypothesis, but they need not be consequences.[4] For instance, other structures could feature testable causes of the hypothesis. In any case, the rationale for being concerned that the aspects are varied is that each aspect may fail to perfectly track the state of the hypothesis. For example, the hypothesis that climate change is proceeding at a faster pace than anticipated increases the likelihood of various consequences at regional levels, but the hypothesis neither entails these consequences nor is entailed by these aspects occurring. We thus have another interpretation of variety:

Variety as testable aspect independence: Evidential elements are varied if the testable aspects of the hypothesis on which they report are different.

In table 1, structures A and B have a *fully shared testable consequence*, so they are less varied in this sense than structures C and structure D, which have *fully independent consequences*.

There is a common trait to these two notions of variety: in the language of Bayesian networks, they are about the extent to which the evidential elements share *ancestors* (simply parents here). The only difference between the two notions is that the reliability dimension is about a property that is not related to the hypothesis under consideration ($R_i \perp\!\!\!\perp H$), while the dimension regarding testable consequences of the hypothesis (or, more generally, 'aspects' of the hypothesis) is about properties that stand in-between the hypothesis and the evidence. We leave to another time the investigation of whether this notion of 'ancestor independence' can constitute the long sought-for general notion of evidential variety.

---

[3] For recent attempts to spell out a complete characterization of evidential variety in arguments relying on evidence amalgamation, see Schupbach (2015); Kuorikoski and Marchionni (2016). Assessing these propositions falls outside the scope of our article.

[4] 'Aspect' is used for lack of a better word. For comments and the relevant literature on aspects that are not consequences, see footnote 3.

Now that we have two notions of variety which fall under the same genus, we need to complete the specification of our model in order to be able to assess the variety-of-evidence thesis. We start with prior probabilities for the root variables H and R_i :

$$(1) \qquad P(h) = h_0 \quad \text{and} \quad P(r_i) = r_i$$

where h_0 and r_i are parameters strictly between 0 and 1.[5] Therefore, the prior degree of belief that the hypothesis is true is h_0 and the prior degree of belief that the evidential source i is reliable is r_i. Inversely, $\bar{h}_0$ is the prior probability that the hypothesis is false and $\bar{r}_i$ is the prior probability that the source i is unreliable.

How unreliability is modeled must be carefully thought through. Bovens and Hartmann (2003) model an unreliable source as a randomizer: it has a certain probability α of producing a positive report irrespective of whether the consequence holds. There are two aspects to this specification. Following Claveau (2013), we accept one aspect and reject the other.[6] What we accept is the irrelevance of an unreliable source:

$$(2) \qquad P(C_i|e_i, \neg r_i) = P(C_i|\neg e_i, \neg r_i) = P(C_i|\neg r_i)$$

In words, if it is known that source i is unreliable, learning a positive or a negative report changes nothing to the probability of the testable consequence. No revision of belief in the truth of the consequence (and the hypothesis) is called for. Condition (2) implies that:

$$(3) \qquad P(e_i|c_i, \neg r_i) = P(e_i|\neg c_i, \neg r_i) = P(e_i|\neg r_i) = \alpha_i,$$

which is our third parameter.

The second aspect of Bovens and Hartmann's specification of an unreliable source, the randomization aspect, is brought to the fore when we consider structures with fully shared reliability (structures A and C). In these structures, the following expression holds:

$$P(e_2|e_1, \neg r) = P(e_2|\neg e_1, \neg r) = \alpha_2,$$

which means that, once it is known that the common source is unreliable, learning that one report is positive or negative changes nothing to the belief that the other report will go in one direction. It is a strange assumption to make: while unreliability is shared, the direction of the unreliable report is not. We follow Claveau (2013, sec. 4) in modeling an unreliable source as being *systematically* rather than randomly biased. To do so, the unreliable state of the reliability variable R_i must be subdivided:

- $\neg r_i = \{b_i^h, b_i^{\neg h}\}$, where b_i^h stands for the proposition that the source is biased toward a positive report for the hypothesis regardless of its truth, and $b_i^{\neg h}$ stands for the proposition that the source is biased toward a negative report. The reliability variable thus has three possible states: $R_i = \{r_i, b_i^h, b_i^{\neg h}\}$.

---

[5] Although we use r_i both for the state of variable R_i (its other state being $\neg r_i$) and for a value of the parameter, namely the prior probability of this state, $P(r_i)$, this conflation is unlikely to generate confusion.

[6] We also urge the reader to remember that our modeling choices are not covering all possible notions of an unreliable source, see Claveau (2013, p. 98).

Table 2. *Probability of a positive report given the values of C_i and R_i*

$P(e_i\|C_i, R_i)$	r_i	b_i^h	$b_i^{\neg h}$
C_i	1	1	0
$\neg C_i$	0	1	0

Probabilities for the two types of bias conditional on the source being unreliable come straight out of expression (3):

$$(4) \qquad P(b_i^h|\neg r_i) = \alpha_i \quad \text{and} \quad P(b_i^{\neg h}|\neg r_i) = \bar{\alpha}_i$$

Combining 1 and 4, we have that the prior probability of a positive bias $P(b_i^h)$ is $\alpha_i \bar{r}_i$ and the prior probability of a negative bias $P(b_i^{\neg h})$ is $\bar{\alpha}_i \bar{r}_i$, which are in fact also what Bovens and Hartmann have. The difference between the two versions comes from the specification of how likely the evidential reports are conditional on their parents R_i and C_i, see table 2. The idea here is that, if the source is reliable, the report will perfectly track whether the consequence holds (first column of the table). If the source is unreliable, the report will be fully determined by the direction of the bias. In consequence, if it is only known that a fully shared source is unreliable but the actual direction of the bias is unknown, learning a first positive report informs us that the bias is positive and thus requires a revision to full certainty that the other report will also be positive.[7]

What remains to be specified is how the consequence variable C_i varies with its parent, H:

$$(5) \qquad P(c_i|h) = p_i \quad \text{and} \quad P(\neg c_i|h) = \bar{p}_i$$

$$(6) \qquad P(c_i|\neg h) = q_i \quad \text{and} \quad P(\neg c_i|\neg h) = \bar{q}_i$$

The best testable consequences are the ones that are fully sensitive and fully specific. A fully sensitive consequence always holds if the hypothesis is true: $p_i = 1$. A fully specific consequence never holds if the hypothesis is false: $q_i = 0$. But testable consequences rarely have these ideal properties. The only thing we have to require in order for a positive report to c_i to be evidence for h is that the consequence is more likely to hold if the hypothesis is true than if it is false:

$$(7) \qquad p_i > q_i.$$

Now that all parameters are specified, we need a plausible interpretation of the ceteris paribus clause in the variety-of-evidence thesis. If we are comparing structures X and Y, the clause that 'all other things are equal' implies various symmetry requirements. First, the initial degree of belief in the hypotheses must be equal across structures, that is $P_X(h) = P_Y(h) = h_0$. Second, all positive reports e_i in each structure should have the same confirmatory strength for h, i.e. $P_X(h|e_i) = P_Y(h|e_j)$ for all $i, j = \{1, 2\}$. *Sufficient* conditions for this constraint to hold is that the different α_i, r_i, p_i and q_i are reduced to single α, r, p, q across the structures (Bovens and Hartmann, 2003, p. 104).

We now have the ingredients to assess the variety-of-evidence thesis in the context of the structures represented in table 1.

---

[7] This revision to full certainty is a stringent assumption that is dropped in section 4 where we introduce degrees of variety.

3. The variety-of-evidence thesis and structures at extremes on each dimension of variety

In this section, we focus on structures A to D in table 1, leaving structure G for the next section. With these four structures, we can interpret the variety-of-evidence thesis in the following way:

Variety-of-evidence thesis (pairwise comparison): *Ceteris paribus*, if a structure X produces more diverse evidential elements than structure Y, then

$$P_X(h|e_1, e_2) > P_Y(h|e_1, e_2).$$

The probability $P(h|e_1, e_2)$, the posterior belief in the hypothesis given two positive reports, will also be referred to by $P^*(h)$.

Given our two specifications of the notion of evidential variety in the previous section, the variety-of-evidence thesis pronounces unambiguously about relative strength of confirmation for almost all comparisons of the structures A to D. Only the comparison between structures B and C does not give a clear result because each structure has more variety than the other on one dimension. We thus focus on the 5 other comparisons. The results can be summarized in three propositions (proofs are in appendix A.2)

Proposition 1 (Maximal difference in variety)**.** *The strength of confirmation of h with two positive reports e_1 and e_2 is,* ceteris paribus, *stronger in structure D than in structure A. That is,*

$$P_D^*(h) > P_A^*(h), \text{ for all admissible parameter values.}$$

This proposition is implied by the variety-of-evidence thesis because evidential sources that test for the same consequence and share the same reliability state (such as in structure A) clearly produce evidence less varied than evidential sources having the opposite characteristics on both dimensions (such as in structure D).

What about comparisons where structures are differently varied on only one of these two dimensions? We start by looking at variety on the reliability dimension.

Proposition 2 (Shared vs independent reliability)**.** *The strength of confirmation of h with two positive reports e_1 and e_2 is,* ceteris paribus, *stronger in structure B than in structure A and, similarly, stronger in structure D than in structure C. That is,*

$$P_B^*(h) > P_A^*(h) \text{ and}$$
$$P_D^*(h) > P_C^*(h), \text{ for all admissible parameter values.}$$

This proposition confirms that the result in Claveau (2013, p. 106) cannot be attributed to the omission of the C_i as intermediate causes of the E_i. By taking the evidential structures in Bovens and Hartmann (2002, 2003) and only replacing their dubious assumption about the behavior of unreliable evidential sources, we save the variety-of-evidence thesis from refutation in these pairwise comparisons.

What remains to be compared are structures that differ only on the number of consequences tested. It is here that the support for the variety-of-evidence thesis breaks down:

Proposition 3 (Shared vs independent consequences)**.** Ceteris paribus, *the strength of confirmation of h with two positive reports e_1 and e_2 is not always stronger in*

structure C than in structure A and, similarly, is not always stronger in structure D than in structure B. More specifically:

$$(8) \qquad P_C^*(h) > P_A^*(h) \Leftrightarrow rpq + \alpha\bar{r}(p + q - 1) > 0.$$

$$(9) \qquad P_D^*(h) > P_B^*(h) \Leftrightarrow (r^2 + 2r\alpha\bar{r})pq + (\alpha\bar{r})^2(p + q - 1) > 0$$

This time, the proposition is consistent with the result in Bovens and Hartmann (2003, p. 100). In other words, their assumption about the behavior of unreliable sources does not matter here: like them, we reach the conclusion that it is sometimes more confirmatory to test again the same consequence than to test a different consequence.

Inequalities on the right-hand-side of (8) and (9) have the same form, the second inequality being only more complex with respect to the parameters capturing reliability because it has to cover diverging reliability states (i.e., when $R1 \neq R2$). These inequalities can be reverse only if $p < (1 - q)$, which is simply:
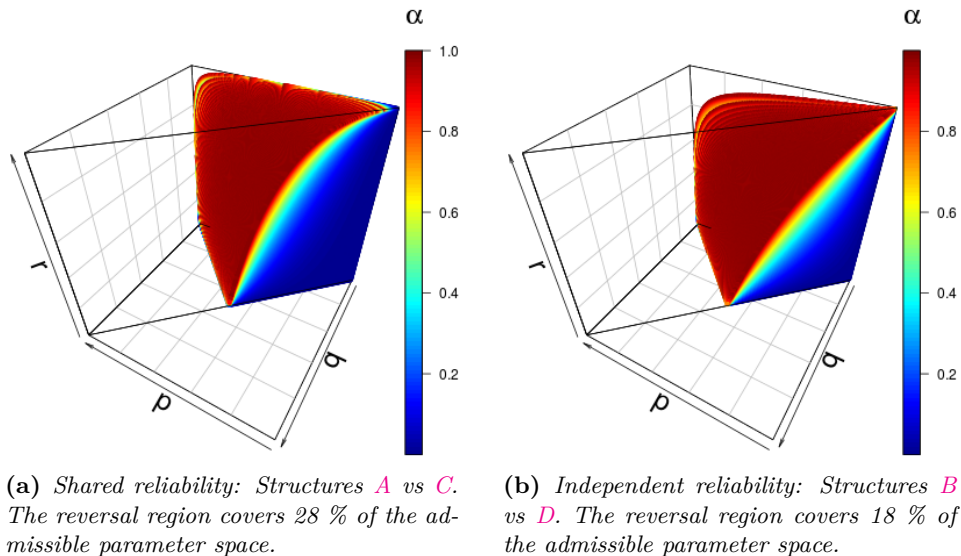
$$(10) \qquad P(c_i|h) < P(\neg c_i|\neg h).$$

In statistical language, this inequality is equivalent to saying that the specificity of the consequence, $P(\neg c_i|\neg h)$, is higher than its sensitivity, $P(c_i|h)$. In causal language, the condition can be interpreted as saying that h as a cause of c_i is closer to being a necessary cause (if it fails to hold, c_i is unlikely to result from something else) than to being a sufficient one (even if it holds, whether c_i occurs hinges on other factors).

It must be stressed that consequences meeting condition (10) do not form an odd base to assess a hypothesis. For instance, a recent review of screening tests in medicine indicates that tests that are more specific than sensitive are more frequent than the reverse (Maxim et al., 2014, fig. 1 and table A1). A consequence which is more specific than sensitive can be an extremely good base to *validate* a hypothesis: in the extreme case of a perfectly specific consequence of h, i.e., $P(\neg c_i|\neg h) = 1$, establishing that the consequence holds would simultaneously establish that the hypothesis is true because the consequence cannot hold if the hypothesis is false, i.e., $P(c_i|\neg h) = 0$.

Meeting condition (10) is not sufficient to overturn the variety-of-evidence thesis, other constraints on the four parameters must hold. Figure 1 represents the parameter regions where the less varied structure in terms of number of consequences tested dominates the more varied structure. The typical combination where less variety dominates more variety is when:

-i- There is a low probability that the consequence occurs given that the hypothesis is the case (low p)
-ii- There is a low probability that the sources are reliable (low r)
-iii- There is a high probability that the bias is positive given that the source is unreliable (high α)

At this point, one might worry that the situations we are discussing characterize extremely weak evidential sources, so weak that the commonsensical (in contrast to formal) understanding of what evidential sources are would not count those as genuine providers of evidence. For instance, if the probability of a positive bias ($\alpha\bar{r}$) from a source is almost one, a reasoner who does not have the cognitive resources to work with extremely fine-grained degrees of belief could use the following heuristic:

(a) *Shared reliability: Structures A vs C. The reversal region covers 28 % of the admissible parameter space.*

(b) *Independent reliability: Structures B vs D. The reversal region covers 18 % of the admissible parameter space.*

Figure 1. *Parameter region where confirmation is stronger for structures where only one consequence is tested instead of two. Each colored point gives the coordinates in the 4-dimensional space $\{p, q, r, \alpha\}$ where the strength of confirmation from both structures is equal. Starting from this point, any decrease in p, q or r, or any increase in α leads to a strict reversal of what the variety-of-evidence thesis claims. Given constraint $p > q$, the admissible region of the parameter space is the half of the hypercube behind the diagonal line.*

Heuristic for weak evidential sources: If a positive report for the hypothesis comes from a source that is highly likely to be positively biased ($\alpha \bar{r} \lesssim 1$), do not consider this report as evidence for the hypothesis (although all assumptions stated in section 2 hold).

This heuristic is justified by the fact that a positive report from such a source ought to have only a tiny influence on the degree of belief in h. Indeed, the Bayesian reasoner ought to take into account that the report is most probably attributable to a positive bias, not to the truth of the hypothesis. Consequently, a real reasoner who can hardly track minute changes in rational degrees of belief might has well follow the heuristic in labelling the reports as 'not evidence'.

If reversals of the variety-of-evidence thesis identified in proposition 3 involved only extremely weak evidential sources, we might be justified in concluding that these reversals are not a genuine refutation of the thesis. After all, the thesis is meant as a guide for practitioners. Its scope of validity could thus be restricted to what commonsense takes to be actual evidence. This suggestion would follow Claveau (2013, p. 113) in proposing to "read the result as highlighting the danger of using extremely weak evidential sources, rather than as a direct refutation of the variety-of-evidence thesis."

However, the situations we are considering here are not limited to weak evidential sources. Although the combination of the three characteristics listed above is typical, the reversal of the variety-of-evidence thesis can occur with parameter

Table 3. *Examples of how a less diverse structure on the dimensions of consequences produces stronger confirmation of h. For h, b_i^h and c_i, we show the prior probability ($\emptyset$), the probability when the first evidential element is acknowledged (e_1) and the probability when all evidence is in (e_1, e_2).*

(a) *Comparison of structures with shared reliability. Parameters are: $h_0 = .5$, $r = .75$, $\alpha = .33$, $p = .33$ and $q = 0.01$.*

		$P(h|\cdot)$			$P(b^h|\cdot)$			$P(c_i|\cdot)$		
		$\emptyset$	e_1	e_1, e_2	$\emptyset$	e_1	e_1, e_2	$\emptyset$	e_1	e_1, e_2
Structure A	$i=1$	.50	.79	**.79**	.08	.39	.39	.17	.67	.67
	$i=2$									
Structure C	$i=1$	.50	.79	**.67**	.08	.39	.67	.17	.67	.44
	$i=2$								.26	

(b) *Comparison of structures with independent reliability. Parameters are: $h_0 = .5$, $r = .5$, $\alpha = .25$, $p = .25$ and $q = 0.01$.*

		$P(h|\cdot)$			$P(b_i^h|\cdot)$			$P(c_i|\cdot)$		
		$\emptyset$	e_1	e_1, e_2	$\emptyset$	e_1	e_1, e_2	$\emptyset$	e_1	e_1, e_2
Structure B	$i=1$	.50	.66	**.85**	.13	.66	.37	.13	.43	.79
	$i=2$					.13				
Structure D	$i=1$	.50	.66	**.79**	.13	.66	.60	.13	.43	.50
	$i=2$					.13			.17	

combinations that do not characterize weak evidential sources.[8] Table 3 illustrates two such situations.[9]

In panel 3a, we have a credible evidential source ($r = .75$), a tendency toward negative bias ($\alpha = .33$), a low sensitivity of the consequence ($p = .33$) but an extremely high specificity ($1 - q = .99$). This situation involves a low probability of receiving a positive report, $P(e_1) = .21$, so the first positive report comes as a surprise. In consequence, there are dramatic adjustments to the probability of a positive bias (from .08 to .39) and to the probability that the relevant consequence holds (from .17 to .67). Since the relationship from h to c_i is extremely specific, the jump in the probability of c_1 transfers into a jump into the probability of the hypothesis (from .50 to .79), but since the relationship is weakly sensitive, this jump translates into only a slight increase in the probability of the other consequence for structure C (from .17 to .26). There is no surprise when the second evidential element comes in. If we are in structure A, E_2 must be identical to E_1 because

---

[8] For instance, our numerical analysis reveals that any combination of the parameters $\{\alpha, r\}$ is associated to a reversal of the strength of confirmation between structures A and C (shared reliability) under some values of the other paramaters (p, q). This is a pretty strong result: in this comparison, the variety-of-evidence thesis can lead us astray whatever the probability that the source is positively biased ($\alpha \bar{r}$). In the comparison between structures B and D (independent reliability), 98 % of the space $\alpha \times r$ is susceptible to reversals; only if we have the combination of a highly reliable source ($r \lesssim 1$) with a tendency to bias extremely skewed toward a negative result ($\alpha \gtrsim 0$) is the variety-of-evidence thesis not subject to counterexamples.

[9] The conditional probabilities in table 3 have been computed using the gRain package in R (Højsgaard, 2012).

both are fully determined by the C and R that they share. This is why there is no probability changed when e_2 is registered in structure A. For structure C, it is possible that $c_2 \neq c_1$, so there is genuinely new information in learning e_2. But e_2 is met with suspicion compared to e_1. The prior probability of positive bias being quite higher than before, a bias affecting both reports becomes an explanation more likely than the possibility that the two consequences, although weakly sensitive, happen both to hold: the probability of positive bias jumps again (from .39 to .67). These increasing doubts in the reliability of the source impact the probability of the hypothesis through the probability of the consequences: it is cut below what it was based on only one report, which means that e_1, e_2 disconfirms h compared to e_1. In the end the gap in $P(h|e_1 e_2)$ between the two structures is quite large: the less diverse structure A dominates by 12 percentage points the more diverse structure C.

The example in panel 3b, which compares structures with independent reliability, has strong similarities with the previous example. Note that some parameters are somewhat more extreme, with a lower credibility of the sources ($r = .5$), lower probability of positive bias ($\alpha = .25$) and even weaker sensitivity of the consequences ($p = .25$). Because of low sensitivity of the consequences, the consequence c_2 not related to the first positive report sees its probability increase only slightly upon learning this report (from .13 to .17). When the second positive report comes in, it is directly related to a consequence that is quite likely (.43) in the less varied structure B in comparison to the other structure D (.17). Consequently, the probability propagates more strongly toward the consequence and, ultimately, to the hypothesis in the less varied structure than in the other structure.

In both examples of table 3, the heuristic for weak evidential sources does not apply: the sources are not weak. In fact, the evidential elements confirm quite strongly the hypothesis, changing its probability from .5 to as much as .85. It would thus be a grave mistake to disregard this evidence on the alleged ground that it does not correspond to what commonsense judges to be proper evidence.

This result is pretty damaging to the variety-of-evidence thesis. Yet, one must keep in mind in interpreting this result that more diversity with respect to the number of testable consequences increases the strength of confirmation for the majority of admissible parameter combinations, as Figure 1 illustrates. As soon as sensitivity (p) is greater or equal to specificity ($1 - q$) – a condition met in half of the parameter space – the variety-of-evidence thesis is on safe ground. If the reverse holds, there is still 44 % of the remaining parameter space where more variety increases the strength of confirmation in the comparison between structures with shared reliability (subfigure 1a). In the comparison between structures with independent reliability (subfigure 1b), this fraction is as high as 64 %.

4. Degrees of variety

So far, we have only looked at structures A to D, which differ only in terms of the number of consequences and the number of reliability variables implicated. As Claveau (2013, p. 107) notes, the variety-of-evidence thesis does not seem to be meant to cover only these extreme structures where, on each dimension, we have either full independence or full dependence. In between these extremes, there is a continuum of degrees of dependence between the consequences or between the reliability of the sources. Structure G represents the general case for which the

four evidential structures discussed so far are special cases. The dotted arrows between the C_i nodes and the R_i nodes indicate that there can be a probabilistic association between these variables that cannot be accounted for by the solid arrows in the network. We remain unspecific about the reason for this probabilistic association. In general causal terms, it might well be that one of the two variables is causing the other or that they share a cause that is not present in the graph. Regardless of the reason, the relevance of structure G comes from the fact that it can capture situations where the reliability of the sources is anywhere in between being fully independent and being fully shared, and that the same holds for testable consequences.

The reader might wonder whether this extension to degrees of independence is really relevant. Does independence of reliability and independence of testable consequences come in degree? Our answer is a clear 'Yes'. Start with the dimension of reliability. One can have various reasons to doubt the reliability of two reports. Some of these reasons will be shared by the two reports, others not. For instance, our first evidential element e_1 might be a positive report from research team i using apparatus k. If our second evidential element e_2 is a positive report from the same research team i, but using a different apparatus l, the evidential set $\{e_1, e_2\}$ will come from less reliability independent sources than if we have evidential element e_2' (instead of e_2) that is produced from another research team j with apparatus l. But it would clearly be too bold to say that this alternative set $\{e_1, e_2'\}$ comes from sources that are fully reliability independent; for all we know, they might share the same flawed auxiliary assumptions, the same publication bias, etc.

A similar point can be made about the testable consequences. Structures in table 1 already specify that C_1 and C_2 share a cause, namely H. But why should this be the only causal relationship between C_1 and C_2 while not going as far as lumping them into a single C such as in structures A and B? Examples abound. Higher average temperatures in a given year in the Canadian Prairies (c_1) and in Eastern Europe (c_2) are both relatively sensitive consequences of the existence of climate change (h), but they can also arise from the El Niño Southern Oscillation. It is thus neither the case that the climate data for the two regions correlate perfectly (that would be a case of fully shared consequence), nor that if we condition on the existence of climate change, the variability in their climate data do not correlate at all (a case of fully independent consequences).

A serious analysis of the variety-of-evidence thesis must thus take into account that variety comes in degree. Our plan for the rest of this section is as follows: first, we describe how we model the degree of reliability independence; second, we describe how we model the degree of consequence independence; third, we analyze the variety-of-evidence thesis by interpreting variety as higher degree of reliability independence; finally, we do the same with the interpretation in terms of higher degree of consequence independence.

We follow Claveau (2013, sec. 5) in how we model the probabilistic association between the reliability variables. This association is fully captured by specifying the joint probabilities for the nine possible combinations of their values (whether each is reliable, positively biased or negatively biased). The first panel of Table 4 (the one for structure G) gives a general notation for these nine possibilities. Symmetry between the two sources – a constraint imposed by our interpretation of the ceteris paribus clause – is assumed. The last two panels of table 4 give the specific values

Table 4. *Joint probabilities for the reliability variables (from Claveau, 2013).*

$P(R_1, R_2)$	r_2	b_2^h	$b_2^{\neg h}$
	General structure G		
r_1	ω_{rr}	ω_{rh}	$\omega_{r\neg h}$
b_1^h	ω_{rh}	ω_{hh}	$\omega_{h\neg h}$
$b_1^{\neg h}$	$\omega_{r\neg h}$	$\omega_{h\neg h}$	$\omega_{\neg h\neg h}$
	Fully shared (structures A and C)		
r_1	r	0	0
b_1^h	0	$\alpha\bar{r}$	0
$b_1^{\neg h}$	0	0	$\bar{\alpha}\bar{r}$
	Independent (structures B and D)		
r_1	r^2	$r\alpha\bar{r}$	$r\bar{\alpha}\bar{r}$
b_1^h	$r\alpha\bar{r}$	$(\alpha\bar{r})^2$	$\alpha\bar{\alpha}\bar{r}^2$
$b_1^{\neg h}$	$r\bar{\alpha}\bar{r}$	$\alpha\bar{\alpha}\bar{r}^2$	$(\bar{\alpha}\bar{r})^2$

taken by the ω for, respectively, structures with fully shared reliability and fully independent reliability. Note that each element on the main diagonal of the third panel is the square of the same element in the second panel. A metric for degrees of reliability independence can be characterize using this fact.

Define a variable $\delta \in [0,1]$ which is meant to be a measure of the distance of the evidential set to the extreme of fully shared reliability, i.e., when $\delta = 0$ we are effectively in a situation with only one R, when $\delta = 1$ we have two probabilistically independent reliability variables (such as in structures B and D), and when δ is strictly between 0 and 1, independence is only partial. From these considerations, it can easily be seen that the elements on the main diagonal of the first matrix in table 4 are:

$$(11) \qquad \omega_{rr} = r^{1+\delta}, \quad \omega_{hh} = (\alpha\bar{r})^{1+\delta}, \quad \omega_{\neg h\neg h} = (\bar{\alpha}\bar{r})^{1+\delta}.$$

Specifying the off-diagonal elements is more intricate. Relying on the further assumption that the marginal (instead of the joint) probabilities of the reliability variables are not a function of δ, Claveau (2013, p. 109) finds that the off-diagonal element needed here is

$$(12) \qquad \omega_{rh} = r + \alpha\bar{r} - .5(1 + \omega_{rr} + \omega_{hh} - \omega_{\neg h\neg h}).$$

The above elements – the metric δ and the expressions in (11-12) – describe how we model degree of reliability independence. Now, we propose to do something very similar to model degree of independence between the consequences. Being what they are, consequences are not root nodes like the R_i and H are in our evidential structures. What must be specified is the joint probability of the consequences given their parent, i.e., $P(C_1, C_2|H)$.

Table 5 gives the different parameter combinations for this joint probability. For fully shared consequences (first block of columns), there is no probability mass for the options in the middle ($\neg c_1, c_2$ and $c_1, \neg c_2$) because it is not possible that one consequence obtains and not the other – this is why the corresponding graphs A and B depict only one C. For fully independent consequences (second block of

Table 5. *Joint probabilities for the consequence variables. The columns for fully shared consequences correspond to structures A and B; those for fully independent consequences correspond to structures C and D; and the general structure is in Figure G.*

$P(C_1, C_2\|H)$	Fully shared		Independence		General structure	
	h	$\neg h$	h	$\neg h$	h	$\neg h$
c_1, c_2	p	q	p^2	q^2	$p^{(1+\gamma)}$	$q^{(1+\gamma)}$
$\neg c_1, c_2$	0	0	$p\bar{p}$	$q\bar{q}$	θ_h	$\theta_{\neg h}$
$c_1, \neg c_2$	0	0	$p\bar{p}$	$q\bar{q}$	θ_h	$\theta_{\neg h}$
$\neg c_1, \neg c_2$	$\bar{p}$	$\bar{q}$	$\bar{p}^2$	$\bar{q}^2$	$\bar{p}^{(1+\gamma)}$	$\bar{q}^{(1+\gamma)}$

columns), the joint probability $P(C_1, C_2|H)$ can be factorized as $P(C_1|H)P(C_2|H)$, as can be read off the corresponding graphs C and D. A comparison of these two extremes gives a measure similar to the one we have for reliability independence. We thus use variable $\gamma \in [0, 1]$ for independence of consequences, where we have a fully shared consequence when $\gamma = 0$ and fully independent consequences when $\gamma = 1$. Assuming symmetry as always – i.e., $P(\neg c_1, c_2|H) = P(c_1, \neg c_2|H)$ – and using the fact that the rows of each column must sum to one, we can solve for the remaining terms in the last block of columns of table 5:

$$(13) \qquad \theta_h = \frac{1 - p^{1+\gamma} - \bar{p}^{(1+\gamma)}}{2} \qquad \theta_{\neg h} = \frac{1 - q^{1+\gamma} - \bar{q}^{(1+\gamma)}}{2}$$

Equipped with characterizations for degrees of variety on our two dimensions, we can derive the likelihood ratio for general structure G as presented in table 1 (see appendix A.1 for the derivation). And we are now in a position to assess the variety-of-evidence thesis under its formulation in terms of degrees.

The variety-of-evidence thesis can now be restated in two ways. The first one is the reliability independence version :

Degree-of-reliability-independence thesis: Ceteris paribus, $\frac{\partial P_G^*(h)}{\partial \delta} > 0$ for all admissible values of p, q, γ, r, α and δ.

Claveau (2013, p. 109) assessed a less general version of this formulation because his parameter space only spanned r, α and δ. Our first result establishes an extension of his result in two directions:

Proposition 4. (Range of validity of Claveau's (2013) result) *If the situation is such that either*
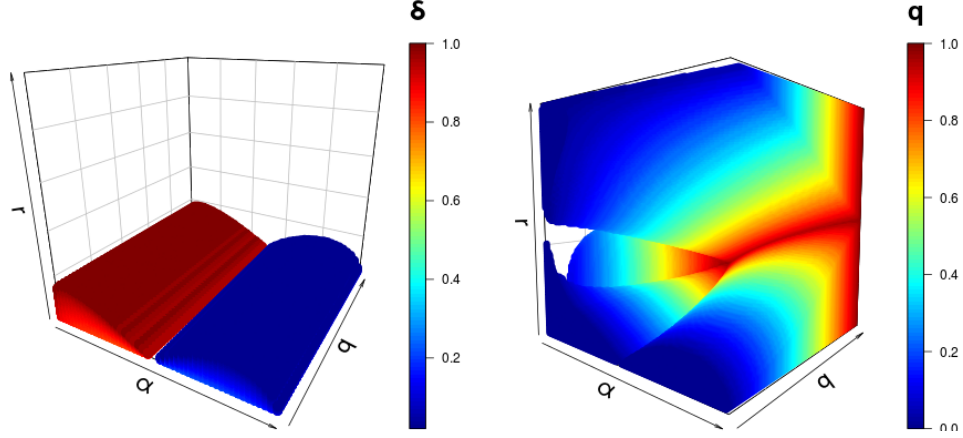
(1) *the testable consequences C_i are both fully sensitive ($p = 1$) and fully specific ($q = 0$), or*
(2) *the testable consequence is fully shared by the two evidential elements ($\gamma = 0$, so $C_1 = C_2 = C$),*

then $\frac{\partial P_G^*(h)}{\partial \delta} > 0$ if and only if:

$$(1 - 2\bar{\alpha}\bar{r})ln(\alpha\bar{r}) + (\bar{\alpha}\bar{r})^{1+\delta}ln(\frac{\alpha}{\bar{\alpha}}) < 0.$$

This inequality is what Claveau (2013, p. 109) found in his analysis of degrees of reliability independence. The proposition extends the result by showing how it can hold under specific parametrizations in structure G. The first possibility –

(a) *Reversal regions when conditions in proposition 4 hold (e.g., γ = 0). Figure 4A in Claveau (2013) corresponds to a (r × α)-slice of this graph where q = 0.*

(b) *Reversal regions when conditions in conjecture 1 hold (γ = 1). Color tracks parameter q to show depth.*

Figure 2. *Regions of the parameter space where confirmation decreases for some upward movement in reliability independence and for some value of the other free parameter (either p or γ).*

perfectly sensitive and specific consequences – is, in fact, a case where the level of testable consequences in our model becomes superfluous; no surprise that we reach the result derived from a model without this level. The second possibility is when consequences are collapsed into a single variable like in structures A and B. The proposition shows that, under this specification, the values for sensitivity (p) and specificity ($1 - q$) do not influence whether the reversal of the variety-of-evidence thesis occurs. Figure 2a depicts the regions where reversal occurs under the second condition in proposition 4. We see that the probability of a reliable source (r, the vertical dimension), the probability for the direction of bias (α, one horizontal dimension) and the degree of reliability independence where the reversal occurs (the color dimension, where blue is the region close to fully shared reliability and red is the region close to fully independent reliability) are not influenced by the specificity of the testable consequences ($1 - q$, the other horizontal dimension).

As Claveau (2013) emphasizes, the region in figure 2a where confirmation does not increase monotonically with reliability independence characterizes especially weak evidential sources. Indeed, no source with a probability of being reliable above 18 % generates a reversal. At least for the region where α is high, one could rely on what has been called above the heuristic for weak evidential sources to claim that the alleged 'evidence' should not be counted as evidence to begin with. It might be a fair argument to make in defense of the relevance of the variety-of-evidence thesis, but its importance is dwarfed by the next two conjectures. Note that we call the results below 'conjectures' because we only establish them through numerical analysis. No analytic proofs of these results have been attempted.

Conjecture 1 (Degree of reliability independence under fully independent consequences)**.** *If the consequences are fully independent ($\gamma = 1$), we have a reversal of*

the degree-of-reliability-independence thesis, that is

$$\frac{\partial P_G^*(h)}{\partial \delta} \leq 0$$

in the following situations:

(1) For any value of r given that the values of the other parameters (α, p, q, δ) are selected appropriately.
(2) In at least 40 % of the parameter space $p \times q \times \alpha \times r$.
(3) In at least 12 % of the parameter space $p \times q \times \alpha \times r \times \delta$.[10]

This conjecture is particularly damaging for the degree-of-reliability-independence thesis. Figure 2b depicts the region in the $(q \times \alpha \times r)$-space where the inequality claimed by the thesis can be reversed (given appropriate values for the other parameters). The contrast with the figure next to it is striking. The only difference in the models that generate these two figures are with respect to the degree of consequence independence γ: figure 2a shows the reversal region when consequences are fully shared, while figure 2b covers situations where consequences are fully independent.
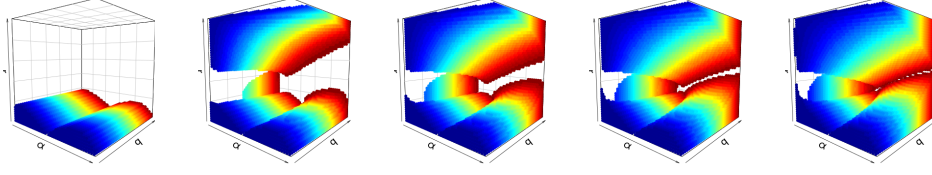
The first clause in conjecture 1 captures a property which comes across clearly in figure 2b: whatever the probability that the source is reliable (vertical axis), it is possible that more reliability independence generates less confirmation. In contrast, as soon as sources have more than 1 chance out of 5 to be unbiased, there was no worries to have about the validity of the degree-of-reliability-independence thesis in situations depicted in figure 2a (and in the structure explored by Claveau, 2013). The most striking reversals correspond to the front-upper-left portion of figure 2b: they occur even though reliability is high (r close to 1), a bias is unlikely to be toward a positive report (α close to 0) and the testable consequences are highly specific (low value of q). This region of the parameter space is where any evidence seeker wants to be. It is thus troubling that the variety-of-evidence thesis fails to hold even in this region.

The last two clauses in conjecture 1 put this result in perspective. They indicate proportions of the parameter spaces where the reversal occur. It is important not to interpret these fractions as probabilities: they would be probabilities only if each point in the parameter space was as likely. The second clause tells us that, among the parameters coding for probability of bias (r and α) and the ones for consequence sensitivity (p) and specificity (q), 2 combinations out of 5 do not give a monotonic relationship between degree of reliability independence and strength of confirmation.

Although this proportion is impressively high (it is barely 10 % in figure 2a), one has to keep in mind that the relationship between degree of reliability independence and strength of confirmation is never decreasing over the full spectrum of degrees of reliability independence. Indeed, our result in proposition 2 must be kept in mind: if one travels all the way from fully shared to fully independent reliability (from $\delta = 0$ to $\delta = 1$), strength of confirmation is guaranteed to increase. Here comes the last clause of conjecture 1: if we also add degree of reliability independence (δ) to the space, then 12 % of the parameter combinations are in the reversal region (barely 2 % in figure 2a). So the region which respects the relationship postulated

---

[10] Our conjectures are based on extensive numerical analysis and data representation (e.g., figure 2b). Our R script, using mainly packages data.table (Dowle et al., 2013) and plot3D (Soetaert, 2016), can be provided upon request.

Figure 3. *Regions of the $(\alpha \times q \times r)$-space where reversals of the degree-of-reliability-independence thesis occur. The graphs have the same axes than the ones in figure 2. They differ from each other by their degree of consequence independence (γ), which are from left to right $\{0, .2, .4, .6, .8\}$. The first graph thus depicts the same situation than figure 2a (but colors are based on q rather than δ).*

by the degree-of-reliability-independence thesis spans by far the greatest share of the probability space. Yet, the regions where the relationship does not hold are neither minute nor necessarily characteristic of extremely weak evidence.

To complete our analysis of the degree-of-reliability-independence thesis, we need to characterize what happens between the two extremes covered by proposition 4 and conjecture 1 and depicted in the two representations in Figure 2:

Conjecture 2 (Degree of reliability independence under intermediate degrees of consequence independence)**.** *The parameter region where the degree-of-reliability-independence thesis is reversed increases monotonically with the degree of consequence independence.*

This conjecture means that proposition 4 corresponds to the best case scenario for the degree-of-reliability-independence thesis in structure G and conjecture 1 is the worst case scenario. Figure 3 illustrate the steady progression from best to worst as degree of consequence independence is increased: as consequences become more independent, the parameter combinations generating counterintuitive results with respect to evidential variety increases (from 2 % to 12 %).
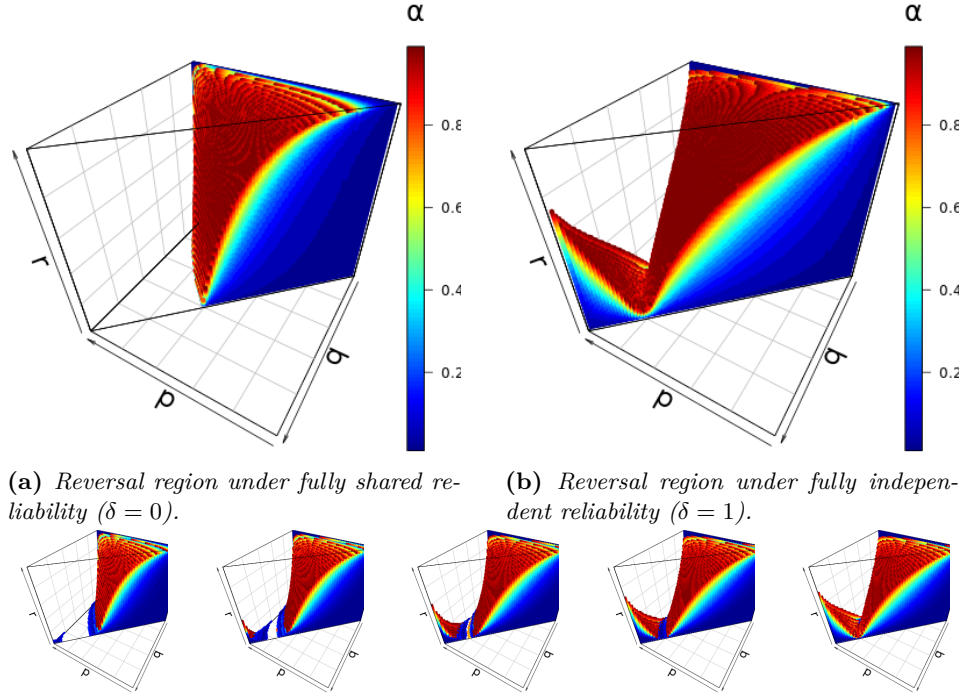
Now that we have explored the variety-of-independence thesis as interpreted in terms of degrees of independence on the dimension of reliability, we are left with the interpretation on the dimension of testable consequences :

> **Degree-of-consequence-variety thesis:** Ceteris paribus, $\frac{\partial P_{\mathrm{G}}^*(h)}{\partial \gamma} > 0$ for all admissible values of p, q, γ, r, α and δ.

When we were comparing structures A to D, the dimension of consequences already gave a blow to the variety-of-evidence thesis: having fully independent consequences was not always better for confirmation than having fully shared consequences (see proposition 3). We thus already know that the degree-of-consequence-variety thesis will not come out unscathed. Without keeping the suspense any longer, we state the result of our analysis in one main conjecture :

Conjecture 3 (Reversal of the degree-of-consequence-variety thesis)**.** *The regions of the parameter space where an increase in the degree of consequence independence γ is not accompanied by an increase in the strength of confirmation $P_G(h|e_1, e_2)$ have the following characteristics:*

(1) For any degree of reliability independence δ:
   (a) there are reversals even in the region $p > 1 - q$,

(a) *Reversal region under fully shared re-*
liability (δ = 0).

(b) *Reversal region under fully indepen-*
dent reliability (δ = 1).

Figure 4. *Regions of the $(\alpha \times q \times r)$-space where reversals of the degree-*
of-consequence-independence thesis occur. The top two graphs give the
two extreme cases with respect to reliability. The line of smaller graphs
below give intermediate cases corresponding to $\delta = \{.1, .3, .5, .7, .9\}$.

> (b) *all combinations of $r \times \alpha$ produce a reversals provided that the values*
> *of the other parameters (p, q, γ) are selected appropriately.*
>
> (2) *As the degree of reliability independence approaches full independence ($\delta =$*
> *1):*
> > (a) *all admissible combinations of $\{p, q\}$ except the single point $\{1, 0.5\}$*
> > *come to produce reversals provided that the values of the other param-*
> > *eters (α, r, γ) are selected appropriately.*
> > (b) *the region of the parameter space $p \times q \times \alpha \times r$ where reversals occur*
> > *grows, going from .44 to .50 as a fraction of the space.*
> > (c) *the region of the parameter space $p \times q \times \alpha \times r \times \gamma$ where reversals*
> > *occur shrinks, going from .30 to .25 as a fraction of the space.*

These results show that, as with the dimension of reliability, introducing de-
grees in our measure of consequence variety makes the variety-of-evidence thesis
even less supported. When we were making dichotomous comparisons between
fully shared and fully independent consequences, we found that the thesis held for
cases where consequence sensitivity was higher than consequence specificity, see
expression (10). The first result of conjecture 3 is that this safe space is no more.
The second result is that, for any combination of the probability that sources are
reliable (r) and for any strength of directionality in the potential bias (α), the

degree-of-consequence-independence thesis is subject to counterexamples.[11]   The
last three results are comparable to what we state in conjecture 2 about the degree-
of-reliability-independence thesis: as independence increases on the other dimension
(here reliability), counterexamples to the version of the thesis under consideration
(here the degree-of-consequence-independence thesis) become more pervasive. This
phenomenon is depicted in figure 4. The phenomenon is less dramatic in this case,
as can be grasped by comparing how much more the sequence of graphs in fig-
ure 3 represents more important changes than the sequence in figure 4. In fact,
the fraction of the full parameter space subject to reversal shrinks somewhat, while
the fraction in the subspace not including degree of consequence independence (γ)
grows. Beyond these peculiarities, a key result is the sheer size of the fractions
involved (the last two results of conjecture 3 to be compared with conjecture 1): to
summarize it in a single number, around 28 % of points in the space made of our
six parameters are such that a slight increase in the diversity of consequences is
not associated to more strength of confirmation. The degree-of-consequence-variety
thesis pronounces wrongly for more than one out of four admissible parameter com-
binations.

5. Conclusion

Variety of evidence matters to confirmation. Yet, contrary to what the variety-
of-evidence thesis claims, the relationship is not always positive. In this article, we
have attempted (following Bovens and Hartmann, 2002, 2003) to formally capture
two notions of variety. First, two sources are varied to the extent that the aspects
of the hypothesis they are testing are independent (variety as testable aspect inde-
pendence). Second, two sources are varied to the extent that the reasons why they
would be unreliable informants with respect to their testable aspect are indepen-
dent (variety as reliability independence). By extending a method introduced by
Claveau (2013, sec. 5) to consider these two dimensions of variety as continua, we
have shown that there are diverse situations where more variety decreases rather
than increases confirmation.

What should we do about the variety-of-evidence thesis given these results? One
option is to question the relevance of our results by arguing that the idealizations on
which they are based are unwarranted. Throughout the article, we have provided
arguments to defend our assumptions when we saw the need to. We leave it to the
skeptics to formulate objections.

If our idealizations and, consequently, our results are accepted, the variety-of-
evidence thesis as a universally quantified generalization must be rejected. The
thesis is however still acceptable under a different interpretation: it can serve as a
rule of thumb for practitioners. We saw that the region of the full parameter space
where confirmation increases with variety is significantly larger than the region
where it does not. In the worst case (i.e., the interpretation in terms of degree of
consequence variety), the regions are in a ratio of 3 to 1. We would need to be
in a world where the probabilities over the parameter space are strangely skewed
toward the problematic parameter combinations for this rule of thumb to be a

---

[11] To compare this result with what we found in our comparison of fully shared and fully
independent consequences, see footnote 8. One can also compare this result to the first result
in conjecture 1 (about the degree-of-reliability-independence thesis), which could only generalize
over r, not α.

counterproductive guide for practical epistemology. Although this interpretation rescues the variety-of-evidence thesis, it should also incite us to do better: we ought to strive for better guides in our epistemic choices than a thesis that directs us most of the time in an appropriate direction, but frequently fails us.

Appendix A. Proofs

A.1. **Likelihood Ratios with Cs as Consequences : General and Specific Evidential Structures**

$P(h|e_1, e_2)$ can be represented by the likelihood-ratio form :

$$P^*(h) = \frac{h_0}{h_0 + \bar{h}_0 L}, \quad \text{where} \quad L = \frac{P(e_1, e_2|\neg h)}{P(e_1, e_2|h)}$$

Given that the ceteris paribus clause imposes a common h_0 across structures, we can make pairwise comparisons of structures in terms of likelihood ratios directly, noting that:

$$P_X^*(h) > P_Y^*(h) \Leftrightarrow L_Y > L_X.$$

Likelihood ratio : proof for general structure G. The likelihood ratio for the general structure G can be calculated with the first panel of table 4 and the two last columns of table 5:

$$
\begin{aligned}
L_{\mathrm{G}} &= \frac{P_{\mathrm{G}}(e_1, e_2|\neg h)}{P_{\mathrm{G}}(e_1, e_2|h)} \\
&= \frac{\sum_{C_1,C_2,R_1,R_2}(\prod_{i=1,2} P_{\mathrm{G}}(e_i|C_i, R_i))P_{\mathrm{G}}(C_1, C_2|\neg h)P_{\mathrm{G}}(R_1, R_2)}{\sum_{C_1,C_2,R_1,R_2}(\prod_{i=1,2} P_{\mathrm{G}}(e_i|C_i, R_i))P_{\mathrm{G}}(C_1, C_2|h)P_{\mathrm{G}}(R_1, R_2)} \\
&= \frac{\omega_{rr}(q^{1+\gamma}) + \omega_{rh}(2q^{1+\gamma} + 2\theta_{\neg h}) + \omega_{hh}(q^{1+\gamma} + 2\theta_{\neg h} + \bar{q}^{1+\gamma})}{\omega_{rr}(p^{1+\gamma}) + \omega_{rh}(2p^{1+\gamma} + 2\theta_{h}) + \omega_{hh}(p^{1+\gamma} + 2\theta_{h} + \bar{p}^{1+\gamma})} \\
&= \frac{\omega_{rr}(q^{1+\gamma}) + \omega_{rh}(q^{1+\gamma} - \bar{q}^{(1+\gamma)} + 1) + \omega_{hh}}{\omega_{rr}(p^{1+\gamma}) + \omega_{rh}(p^{1+\gamma} - \bar{p}^{(1+\gamma)} + 1) + \omega_{hh}}, \qquad \text{using (13)}
\end{aligned}
$$

which is the expression in table 1. □

We can expand this likelihood by substituting the original parameters for the ω. Using equations (11) and (12), we have:

$$
\begin{aligned}
L_{\mathrm{G}} &= \frac{r^{1+\delta}(q^{1+\gamma}) + (r + \alpha\bar{r} - .5(1 + \omega_{rr} + \omega_{hh} - \omega_{\neg h \neg h}))(q^{1+\gamma} - \bar{q}^{1+\gamma} + 1) + (\alpha\bar{r})^{1+\delta}}{r^{1+\delta}(p^{1+\gamma}) + (r + \alpha\bar{r} - .5(1 + \omega_{rr} + \omega_{hh} - \omega_{\neg h \neg h}))(p^{1+\gamma} - \bar{p}^{1+\gamma} + 1) + (\alpha\bar{r})^{1+\delta}} \\
&= \frac{r^{1+\delta}(q^{1+\gamma}) + (r + \alpha\bar{r} - .5(1 + r^{1+\delta} + (\alpha\bar{r})^{1+\delta} - (\bar{\alpha}\bar{r})^{1+\delta}))(q^{1+\gamma} - \bar{q}^{1+\gamma} + 1) + (\alpha\bar{r})^{1+\delta}}{r^{1+\delta}(p^{1+\gamma}) + (r + \alpha\bar{r} - .5(1 + r^{1+\delta} + (\alpha\bar{r})^{1+\delta} - (\bar{\alpha}\bar{r})^{1+\delta}))(p^{1+\gamma} - \bar{p}^{1+\gamma} + 1) + (\alpha\bar{r})^{1+\delta}}
\end{aligned}
$$

Likelihood ratios for specific evidential structures can then be calculated by attributing a value of 0 to δ and γ for shared reliability and shared consequence, and of 1 for independent reliability and independent consequences :

Likelihood ratio: proof for structure A (shared reliability, shared consequence).

$$L_{\mathrm{A}} = \frac{r^1 q^1 + (r + \alpha\bar{r} - .5(1 + r^1 + (\alpha\bar{r})^1 - (\bar{\alpha}\bar{r})^1))(q^1 - \bar{q}^1 + 1) + (\alpha\bar{r})^1}{r^1(p^1) + (r + \alpha\bar{r} - .5(1 + r^1 + (\alpha\bar{r})^1 - (\bar{\alpha}\bar{r})^1))(p^1 - \bar{p}^1 + 1) + (\alpha\bar{r})^1}$$

$$= \frac{qr + 0 \times (q - \bar{q} + 1) + \alpha\bar{r}}{pr + 0 \times (p - \bar{p} + 1) + \alpha\bar{r}}$$

$$L_{\mathrm{A}} = \frac{qr + \alpha\bar{r}}{pr + \alpha\bar{r}}$$

$\square$

Likelihood ratio: proof for structure B (independent reliability, shared consequence).

$$L_{\mathrm{B}} = \frac{r^2 q^1 + (r + \alpha\bar{r} - .5(1 + r^2 + (\alpha\bar{r})^2 - (\bar{\alpha}\bar{r})^2))(q^1 - \bar{q}^1 + 1) + (\alpha\bar{r})^2}{r^2(p^1) + (r + \alpha\bar{r} - .5(1 + r^2 + (\alpha\bar{r})^2 - (\bar{\alpha}\bar{r})^2))(p^1 - \bar{p}^1 + 1) + (\alpha\bar{r})^2}$$

$$= \frac{qr^2 + (r\alpha\bar{r})(2q) + (\alpha\bar{r})^2}{pr^2 + (r\alpha\bar{r})(2p) + (\alpha\bar{r})^2}$$

$$L_{\mathrm{B}} = \frac{qr^2 + 2qr\alpha\bar{r} + (\alpha\bar{r})^2}{pr^2 + 2pr\alpha\bar{r} + (\alpha\bar{r})^2}$$

$\square$

Likelihood ratio: proof for structure C (shared reliability, independent consequences).

$$L_{\mathrm{C}} = \frac{r^1 q^2 + (r + \alpha\bar{r} - .5(1 + r^1 + (\alpha\bar{r})^1 - (\bar{\alpha}\bar{r})^1))(q^2 - \bar{q}^2 + 1) + (\alpha\bar{r})^1}{r^1(p^2) + (r + \alpha\bar{r} - .5(1 + r^1 + (\alpha\bar{r})^1 - (\bar{\alpha}\bar{r})^1))(p^2 - \bar{p}^2 + 1) + (\alpha\bar{r})^1}$$

$$= \frac{q^2 r + 0 \times (q^2 - \bar{q}^2 + 1) + \alpha\bar{r}}{p^2 r + 0 \times (p^2 - \bar{p}^2 + 1) + \alpha\bar{r}}$$

$$L_{\mathrm{C}} = \frac{q^2 r + \alpha\bar{r}}{p^2 r + \alpha\bar{r}}$$

$\square$

Likelihood ratio: proof for structure D (independent reliability and consequences).

$$L_D = \frac{r^2 q^2 + (r + \alpha\bar{r} - .5(1 + r^2 + (\alpha\bar{r})^2 - (\bar{\alpha}\bar{r})^2))(q^2 - \bar{q}^2 + 1) + (\alpha\bar{r})^2}{r^2(p^2) + (r + \alpha\bar{r} - .5(1 + r^2 + (\alpha\bar{r})^2 - (\bar{\alpha}\bar{r})^2))(p^2 - \bar{p}^2 + 1) + (\alpha\bar{r})^2}$$

$$= \frac{(qr)^2 + (r\alpha\bar{r})(2q) + (\alpha\bar{r})^2}{(pr)^2 + (r\alpha\bar{r})(2p) + (\alpha\bar{r})^2}$$

$$L_D = \frac{(qr + \alpha\bar{r})^2}{(pr + \alpha\bar{r})^2}$$

$\square$

A.2. **Comparison of Confirmatory Strengths with Cs as Consequences**

Proof of proposition 1. In terms of likelihood-ratio the proposition is: $L_A > L_D$, for all admissible values of p, q, r, α.

$$L_A > L_D$$
$$\frac{qr + \alpha\bar{r}}{pr + \alpha\bar{r}} > \frac{(qr + \alpha\bar{r})^2}{(pr + \alpha\bar{r})^2}$$
$$pr + \alpha\bar{r} > qr + \alpha\bar{r}$$
$$p > q,$$

which is true by assumption 7. $\square$

Proof of proposition 2. $L_A > L_B$ and $L_C > L_D$, for all admissible values of p, q, r, α.
We start by the first inequality:

$$L_A > L_B$$
$$\frac{qr + \alpha\bar{r}}{pr + \alpha\bar{r}} > \frac{qr^2 + 2qr\alpha\bar{r} + (\alpha\bar{r})^2}{pr^2 + 2pr\alpha\bar{r} + (\alpha\bar{r})^2}$$
$$qr(\alpha\bar{r})^2 + pr^2\alpha\bar{r} + 2pr(\alpha\bar{r})^2 > pr(\alpha\bar{r})^2 + qr^2\alpha\bar{r} + 2qr(\alpha\bar{r})^2$$
$$r^2\alpha\bar{r}(p - q) + 2r(\alpha\bar{r})^2(p - q) > r(\alpha\bar{r})^2(p - q)$$
$$r + \alpha\bar{r} > 0$$

The division by $p - q$ to get to the last line leaves the direction of inequality unchanged because of assumption 7 $p > q$. Since r and α are assumed to be always strictly positive, the inequality holds.
Now the second inequality:

$$L_C > L_D$$
$$\frac{q^2r + \alpha\bar{r}}{p^2r + \alpha\bar{r}} > \frac{(qr + \alpha\bar{r})^2}{(pr + \alpha\bar{r})^2}$$
$$p^2r + 2pq^2r + 2p\alpha\bar{r} + q^2\alpha\bar{r} > p^2r + 2pq^2r + 2p\alpha\bar{r} + q^2\alpha\bar{r}$$
$$r(p + q)(p - q) + 2\alpha\bar{r}(p - q) > 2pqr(p - q) + \alpha\bar{r}(p + q)(p - q)$$
$$r(p + q - 2pq) + \alpha\bar{r}(2 - (p + q)) > 0$$

Since p < 1 and q < 1, 2 - (p + q) > 0 is always true. Thus, we need to prove that p + q - 2pq > 0 is always true :

$$p + q - 2pq > 0$$
$$\frac{p + q}{pq} > 2$$
$$\frac{1}{q} + \frac{1}{p} > 2$$

Since $0 \le q \le 1$ and $0 \le p \le 1$, then $\frac{1}{q} \ge 1$ and $\frac{1}{p} \ge 1$, which entail that the sum of both fractions is superior to 2.

$\square$

Proof of proposition 3.

$$L_A > L_C \text{ iff } rpq + \alpha\bar{r}(p + q - 1) > 0.$$
$$L_B > L_D \text{ iff } (r^2 + 2r\alpha\bar{r})pq + (\alpha\bar{r})^2(p + q - 1) > 0$$

Starting with the first inequality:

$$L_A > L_C$$
$$\frac{qr + \alpha\bar{r}}{pr + \alpha\bar{r}} > \frac{q^2r + \alpha\bar{r}}{p^2r + \alpha\bar{r}}$$
$$r^2p^2q + qr\alpha\bar{r} + p^2r\alpha\bar{r} > r^2q^2p + pr\alpha\bar{r} + q^2r\alpha\bar{r}$$
$$r^2pq(p - q) + r\alpha\bar{r}(p + q)(p - q) > r\alpha\bar{r}(p - q)$$

(14)
$$rpq + \alpha\bar{r}(p + q - 1) > 0$$

Expression 14 being as claimed in the proposition.

To ease the proof for the second inequality, define the following expressions:

$$A = r^2 \qquad\qquad B = 2r\alpha\bar{r} \qquad\qquad C = (\alpha\bar{r})^2$$

Starting from the inequality implied by the variety-of-evidence thesis, we have:

$$L_B > L_D$$
$$\frac{qr^2 + q2r\alpha\bar{r} + (\alpha\bar{r})^2}{pr^2 + p2r\alpha\bar{r} + (\alpha\bar{r})^2} > \frac{q^2r^2 + q2r\alpha\bar{r} + (\alpha\bar{r})^2}{p^2r^2 + p2r\alpha\bar{r} + (\alpha\bar{r})^2}$$
$$\frac{qA + qB + C}{pA + pB + C} > \frac{q^2A + qB + C}{p^2A + pB + C}$$
$$pq(pA + B)(A + B) + C(p^2A + pB + qA + qB)$$
$$> pq(qA + B)(A + B) + C(q^2A + qB + pA + pB)$$
$$A(A + B)pq(p - q) + AC(p^2 + q - q^2 - p) > 0$$
$$(A + B)pq(p - q) + C(p + q - 1)(p - q) > 0$$
$$(r^2 + 2r\alpha\bar{r})pq + (\alpha\bar{r})^2(p + q - 1) > 0$$

$\square$

Proof of proposition 4. To prove that the inequality holds if the first disjunct holds, we start with the likelihood ratio for the general structure and impose $p = 1$ and $q = 0$:

$$L_G = \frac{\omega_{rr}(q^{1+\gamma}) + \omega_{rh}(q^{1+\gamma} - \bar{q}^{(1+\gamma)} + 1) + \omega_{hh}}{\omega_{rr}(p^{1+\gamma}) + \omega_{rh}(p^{1+\gamma} - \bar{p}^{(1+\gamma)} + 1) + \omega_{hh}}$$
$$= \frac{\omega_{hh}}{\omega_{rr} + 2\omega_{rh} + \omega_{hh}}$$

This likelihood ratio is identical to the expression in Claveau (2013, eq. 11), which implies that the results in expression (13) and in figure 4 of this article also hold here.

Proving the second part of the proposition requires more work. Starting from the likelihood ratio of the general structure and imposing $\gamma = 0$:

$$L_{\mathrm{G}} = \frac{\omega_{rr}(q^{1+\gamma}) + \omega_{rh}(q^{1+\gamma} - \bar{q}^{(1+\gamma)} + 1) + \omega_{hh}}{\omega_{rr}(p^{1+\gamma}) + \omega_{rh}(p^{1+\gamma} - \bar{p}^{(1+\gamma)} + 1) + \omega_{hh}}$$
$$= \frac{\omega_{rr}q + 2\omega_{rh}q + \omega_{hh}}{\omega_{rr}p + 2\omega_{rh}p + \omega_{hh}}$$

Taking the derivative of this expression with respect to δ:

$$\frac{\partial L_{\mathrm{G}}}{\partial \delta} = \frac{\omega'_{rr}q + 2\omega'_{rh}q + \omega'_{hh}}{\omega_{rr}p + 2\omega_{rh}p + \omega_{hh}} - \frac{(\omega_{rr}q + 2\omega_{rh}q + \omega_{hh})(\omega'_{rr}p + 2\omega'_{rh}p + \omega'_{hh})}{(\omega_{rr}p + 2\omega_{rh}p + \omega_{hh})^2}$$

Using the inequality claimed by the degree-of-reliability-independence thesis:

$$\frac{\partial L_{\mathrm{G}}}{\partial \delta} < 0$$
$$(\omega'_{rr}q + 2\omega'_{rh}q + \omega'_{hh})(\omega_{rr}p + 2\omega_{rh}p + \omega_{hh})$$
$$< (\omega_{rr}q + 2\omega_{rh}q + \omega_{hh})(\omega'_{rr}p + 2\omega'_{rh}p + \omega'_{hh})$$
$$\omega'_{rr}\omega_{hh}q + 2\omega'_{rh}\omega_{hh}q + \omega'_{hh}\omega_{rr}p + 2\omega'_{hh}\omega_{rh}p$$
$$< \omega'_{rr}\omega_{hh}p + 2\omega'_{rh}\omega_{hh}p + \omega'_{hh}\omega_{rr}q + 2\omega'_{hh}\omega_{rh}q$$
$$\omega'_{hh}\omega_{rr} - \omega'_{rr}\omega_{hh} + 2\omega'_{hh}\omega_{rh} - 2\omega'_{rh}\omega_{hh} < 0$$

The derivatives of the ω with respect to δ are:

$$\omega'_{rr} = \frac{\partial \omega_{rr}}{\partial \delta} = r^{1+\delta}\ln(r) = \omega_{rr}\ln(r)$$
$$\omega'_{hh} = \frac{\partial \omega_{hh}}{\partial \delta} = (\alpha\bar{r})^{1+\delta}\ln(\alpha\bar{r}) = \omega_{hh}\ln(\alpha\bar{r})$$
$$\omega'_{rh} = \frac{\partial \omega_{rh}}{\partial \delta} = -0.5(\omega_{rr}\ln(r) + \omega_{hh}\ln(\alpha\bar{r}) - \omega_{\neg h\neg h}\ln(\bar{\alpha}\bar{r}))$$

Substituting these derivatives in the inequality, we have:

$$\omega_{hh}\omega_{rr}\ln(\alpha\bar{r}) - \omega_{rr}\omega_{hh}\ln(r) + 2\omega_{hh}\omega_{rh}\ln(\alpha\bar{r}) - 2\omega'_{rh}\omega_{hh} < 0$$
$$\omega_{rr}\ln(\alpha\bar{r}) - \omega_{rr}\ln(r) + 2\omega_{rh}\ln(\alpha\bar{r}) + \omega_{rr}\ln(r) + \omega_{hh}\ln(\alpha\bar{r}) - \omega_{\neg h\neg h}\ln(\bar{\alpha}\bar{r}) < 0$$
$$\omega_{rr}\ln(\alpha\bar{r}) + 2\omega_{rh}\ln(\alpha\bar{r}) + \omega_{hh}\ln(\alpha\bar{r}) - \omega_{\neg h\neg h}\ln(\bar{\alpha}\bar{r}) < 0$$
$$(\omega_{rr} + 2r + 2\alpha\bar{r} - 1 - \omega_{rr} - \omega_{hh} + \omega_{\neg h\neg h} + \omega_{hh})\ln(\alpha\bar{r}) - \omega_{\neg h\neg h}\ln(\bar{\alpha}\bar{r}) < 0$$
$$(2(r + \alpha\bar{r}) - 1)\ln(\alpha\bar{r}) + \omega_{\neg h\neg h}(\ln(\alpha\bar{r} - \ln(\bar{\alpha}\bar{r})) < 0$$
$$(1 - 2\bar{\alpha}\bar{r})ln(\alpha\bar{r}) + (\bar{\alpha}\bar{r})^{1+\delta}ln(\frac{\alpha}{\bar{\alpha}}) < 0$$

The last line being exactly the inequality claimed in proposition 4.

□

References

Bovens, L. and Hartmann, S. (2002). Bayesian Networks and the Problem of Unreliable Instruments. *Philosophy of Science*, 69(1):29–72.

Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology.* Oxford University Press, Oxford.

Claveau, F. (2013). The Independence Condition in the Variety-of-Evidence Thesis. *Philosophy of Science*, 80(1):94–118.

Dowle, M., Short, T., Srinivasan, S. L. w. c. f. A., and Saporta, R. (2013). *data.table: Extension of data.frame for fast indexing, fast ordered joins, fast assignment, fast grouping and list columns.* R package version 1.8.10.

Downward, P. and Mearman, A. (2007). Retroduction as Mixed-Methods Triangulation in Economic Research: Reorienting Economics into Social Science. *Cambridge Journal of Economics*, 31(1):77 –99.

Earman, J. (1992). *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory.* MIT Press, Cambridge, MA.

Howson, C. and Urbach, P. (1993). *Scientific Reasoning: The Bayesian Approach (Second Edition).* Open Court, Chicago.

Højsgaard, S. (2012). Graphical Independence Networks with the gRain Package for R. *Journal of Statistical Software*, 46(1):1–26.

Istituto Nazionale di Fisica Nucleare (2011). Particles Appear to Travel Faster Than Light: OPERA Experiment Reports Anomaly in Flight Time of Neutrinos. *ScienceDaily.* Retrieved October 26, 2011, from www.sciencedaily.com/releases/2011/09/110923084425.htm.

Kuorikoski, J. and Marchionni, C. (2016). Evidential Diversity and the Triangulation of Phenomena. *Philosophy of Science*, 83(2):227–247.

Maxim, L. D., Niebo, R., and Utell, M. J. (2014). Screening tests: a review with examples. *Inhalation Toxicology*, 26(13):811–828.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Francisco.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference, Second Edition.* Cambridge University Press, Cambridge, UK.

Schupbach, J. N. (2015). Robustness, Diversity of Evidence, and Probabilistic Independence. In Mäki, U., Votsis, I., Ruphy, S., and Schurz, G., editors, *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, number 1 in European Studies in Philosophy of Science, pages 305–316. Springer International Publishing. DOI: 10.1007/978-3-319-23015-3_23.

Soetaert, K. (2016). *plot3D: Plotting Multi-Dimensional Data.*

Stegenga, J. and Menon, T. (2017). Robustness and Independent Evidence. *Philosophy of Science.*

Thurmond, V. A. (2001). The Point of Triangulation. *Journal of Nursing Scholarship*, 33(3):253–258.

Wheeler, G. and Scheines, R. (2011). Causation, Association and Confirmation. In Dieks, D., Gonzalez, W. J., Hartmann, S., Uebel, T., and Weber, M., editors, *Explanation, Prediction, and Confirmation*, pages 37–51. Springer Netherlands, Dordrecht. http://www.springerlink.com/content/t7j6u41256j85028/.

Wheeler, G. and Scheines, R. (2013). Coherence and Confirmation Through Causation. *Mind*, 122(435):135–170.

Woodward, J. (2006). Some Varieties of Robustness. *Journal of Economic Methodology*, 13(2):219–240.