Variety of Evidence

Abstract

Varied evidence confirms more strongly than less varied evidence, ceteris paribus. This epistemological Variety of Evidence Thesis enjoys widespread intuitive support. We put forward a novel explication of one notion of varied evidence and the Variety of Evidence Thesis within Bayesian models of scientific inference by appealing to measures of entropy. Our explication of the Variety of Evidence Thesis holds in many of our models which also pronounce on disconfirmatory and discordant evidence. We argue that our models pronounce rightly. Against a backdrop of failures of the Variety of Evidence Thesis, the intuitive case for the Variety of Evidence Thesis emerges strengthened. Our models do however not support the general case for the thesis since our explication of it fails to hold in certain cases. The parameter space of this failure is explored and an explanation for the failure is offered.

1 Introduction

Varied evidence for a hypothesis confirms it more strongly than less varied evidence, ceteris paribus. In this explicit form, this thought can be traced back in the philosophy of science at least to (Carnap, 1962, p. 230). It also appears in (Keynes, 1921, p. 253, emphasis original) "*The* variety *of the circumstances, in which the Newtonian generalisation is fulfilled, rather than the number of them, is what seems to impress our reasonable faculties*", (Horwich, 1982, p. 118) "*It is an undeniable element of scientific methodology that theories are better confirmed by a broad variety of different sorts of evidence than by a narrow and repetitive set of data*", in (Earman, 1992, p. 77) "*It is a truism of scientific methodology that variety of evidence can be as important or even more important that the sheer amount of evidence*" and in (Claveau, 2013, p. 94) "*Seeking a variety of evidence for a hypothesis is standard practice in science, as well as in normal life*", as well as in the very recent Claveau and Grenier (2018); Kuorikoski and Marchionni (2016); Stegenga and Menon (2017).

Variety of evidence is not only discussed by philosophers but also, for example, in social psychology in (Hüffmeier et al., 2016, p. 82) "*The more often a new finding is replicated, and the more different types of replications there are in particular, the more trust and confidence there should be in the validity of this finding*" and in epidemiology, by (Borm et al., 2009, p. 711) *"The results of a single trial should be interpreted with caution. When it is difficult to predict or determine how trial-specific factors influence the results, the best way to evaluate the performance of a treatment is to use multiple, possibly smaller, trials."*

(Hempel, 1966, p. 224) points out that: *"For that the confirmation of a hypothesis depends not only on the quantity of the favorable evidence available, but also on its variety: the greater the variety, the stronger the resulting support."* Hempel's thinking was given a logical formulation in (Meehl, 1990, pp. 109).

Given the predominance of Bayesian reasoning and modelling and the widespread intuitive support for this Variety of Evidence Thesis (VET), one would think that a general Bayesian analysis of the VET comprising of an explication of the notion of varied evidence and the VET has been developed. This, however, is not the case.

2 The Story of the Variety of Evidence Thesis

2.1 The Story – So Far

While the intuitive appeal of the VET is widely spread, its status within the Bayesian paradigm has been contested. There are three main approaches, the latter two overlap.[1]

The eliminative approach put forward in Horwich (1982) aims to capture the intuition that varied evidence for a hypothesis rules out competing hypotheses better than non-varied evidence, see also (Earman, 1992, p. 79), (Glymour, 1980, pp. 139-142), (Horwich, 1998, pp. 611-613), (Howson and Urbach, 2006, pp. 124-126) and (Schupbach, 2015, p. 314). Schupbach (2017) recently argued that varied evidence does better at ruling out competing explanations.

[1] The notion of evidential variety has received a range of labels, the reader is referred to (Claveau, 2011, p. 249, Footnote 18). Refer to Lloyd (2015); Vezér (2017) for recent Variety of Evidence reasoning in (the rational reconstruction of) climate science.

Wayne (1995) extends Horwich's approach from deterministic hypotheses to non-deterministic, i.e., statistical hypotheses and unearths cases in which his version of the VET fails. (Steel, 1996, p. 671) defends the eliminative approach by pointing out that Wayne has not taken a pertinent ceteris paribus condition into account. Fitelson (1996) criticises Wayne (1995) for an uncharitable reconstruction of Horwich (1982). However, Fitelson (1996) finds cases in which the VET fails even in a more charitable reconstruction of the eliminative approach. See also the very recent *blinded for review*.

In the correlation approach of (Earman, 1992, Section 3.5), see also Franklin and Howson (1984), one considers a body of evidence $\mathcal{E} = \{E_1, \ldots, E_n\}$ which is entailed by the agent's background knowledge K and the hypothesis of interest H. The variety of the body of evidence $\mathcal{E}$ is greater, the slower $P(E_n|E_1 \ldots E_{n-1}K)$ approaches one. It is then shown that the posterior probability of the hypothesis of interest is greater the greater this variety of the body of evidence. Wayne (1995) later pointed out – correctly we think – that this approach fails to incorporate the prior probabilities of the items of evidence themselves. Adding a ceteris paribus clause requiring that the prior probabilities of the items of evidence to be equal resolves this issue, as argued in Wayne (1995); Myrvold (1996).

Wayne (1995) also alleges that this approach is incomplete as it omits a discussion of how judgements of similarity depend on theoretical context. Steel (1996); Myrvold (1996) provide such discussions within the Bayesian paradigm.

A major blow to the correlation approach was dealt by Bovens and Hartmann (2002, 2003). Their starting point is different, (Bovens and Hartmann, 2003, p. 93-94) "*interpret more varied evidence as evidence that stems from multiple instruments (rather than a single instrument) and that tests multiple testable consequences (rather than a single testable consequence) of the hypothesis*". Their analysis unearths cases in which their VET fails within their models of scientific inference.

Furthermore, they also show that if varied evidence is conceptualised in the correlational sense, then the principle that more varied evidence leads, ceteris paribus, to more confirmatory support of the hypothesis fails in particular circumstances (Bovens and Hartmann, 2003, p. 104-106). The counter-intuitive results obtained by Bovens and Hartmann do not contradict Earman's results since they relax the assumption that the hypothesis has to entail the evidence and they hence consider a much larger set of situations. Their VET fails for some of these situations which were not considered by Earman (Bovens and Hartmann, 2003, Section 4.2 and 4.3).

The Bayesian quest for a vindication of the VET in full generality has failed.

While the quest for a vindication of the VET is over, a case could be made for the VET in particular epistemological cases or applications. The only such attempt is that of Claveau (2013); Claveau and Grenier (2018) in the philosophy of science.

The starting point is a different construal of the unreliability of scientific instruments. Rather than construing unreliable instruments as complete randomisers – the approach taken by Bovens and Hartmann – unreliable instruments are modelled as systematically biased. The models satisfy a version of the VET on first pass, see (Claveau, 2013, Section 4).[2] On second pass, Claveau drops the assumption that the systematic biases of instruments are fully independent. In this enlarged class of cases, in which systematic biases of instruments are dependent to some degree, his version of the VET fails in certain cases, see (Claveau, 2013, Section 5). Seemingly puzzled and/or unsatisfied by his findings, (Claveau, 2013, p. 113) ends his discussion with: "*The fate of the variety-of-evidence thesis is not yet settled*".

In Claveau and Grenier (2018), the model is extended to consequence variables which are, given the hypothesis, dependent to a degree; in other words: (1) is not imposed. The conclusions drawn are ambivalent: "*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.*"

Approaches to the VET are summarised in Table 1. Lacking in all these approaches is a widely-applicable explication of the notion of varied evidence. Neither Horwich nor his successors attempt an explication. Earman only considers the variety of a body of evidence which follows deductively from the hypothesis and the background knowledge. The only other ingredient to his explication of variety of evidence is the agent's prior. To us, this seems too meagre a recipe for a well-rounded explication. Bovens and Hartmann do not offer an explication. Claveau's explications only apply to bodies of evidence consisting of two items of evidence.

This introduction and the above summary summarise the forthcoming *blinded for review* which gives more details and perspectives.

[2] This is consistent with the Bovens and Hartmann approach since their analysis "*does not apply to unreliable instruments that do not randomize*" (Bovens and Hartmann, 2003, p. 95).

Approach	Variety in terms of
Correlation	Correlation of items of evidence
Elimination	Disconfirmation of competing hypotheses
Sources of Evidence & Consequences of the Hypothesis	Multiple consequences of the hypothesis and/or multiple instruments

Table 1: Different approaches to the Variety of Evidence Thesis.

2.2 The next Chapter

Bayesians who hold the VET dear and are thus deeply troubled by the negative results obtained by Bovens and Hartmann and Claveau face the challenge of squaring their bond to the VET and these negative results which spell an end to the unconditional love of the VET.

The plan for battle is as follows: First, we devise models of scientific inference in which it makes sense to think about varied evidence. We put forward such models in Section 3 drawing on the Bovens and Hartmann approach. We then explicate a notion of varied evidence within our models (Section 4). This puts us in a position to offer an explication of the VET within these models of scientific inference (Section 5). We argue in Section 5.3 that our model and explication are better suited for the philosophy of science than those of Bovens and Hartmann and Claveau. Furthermore, we argue that their models do not track how scientists think about the reliability of their instruments, conclusions drawn within their models are hence of reduced interest in (the philosophy of) science. In Section 6, we show that the VET holds in many of our models.

Interestingly, our models not only pronounce on confirmatory evidence but also on disconfirmatory and discordant evidence, see Theorems 3 and 4. Disconfirmation is rarely discussed in the Bayesian literature while featuring prominently in Popperian writings. Our models give the – we think – intuitively right result for disconfirmatory and discordant evidence, see Section 6.2.2 and Section 6.2.3.

We claim to establish that our understanding of the VET can be given a formal justification within a variety of models of scientific inference (Section 7.2.2) and gesture at a re-examination of intuitions which drive the VET (Section 7.1). Although, it needs to be admitted that another plausible intuitive interpretation of the VET fails to hold in our models. For applications in science, we admit defeat: the VET does fail to hold in some of our models pertinent to science. Conditions of VET failure, explanations of VET failure and a graphical exploration of the parameter space of VET failure are presented in Section 7.2.

The theoretical discussion is accompanied by an episode in the history of science in which variety of evidence clinched the argument: the thirteen ways of determining Avogardo's constant from the hypothesis that matter consists of molecules (atomism), see Perrin (1924). (Poincaré, 1963, p. 91) famously conceded his opposition to atomism when presented with a highly diverse body of evidence: *"What makes it all the more convincing are the multiple correspondences between results obtained by totally different processes."* Poincaré's reasoning can be captured by our models as we see in Section 6.2.1.

3 Models of Scientific Inference

3.1 Variables

We base our models on two widely-held beliefs. i) Scientific inference is amenable to Bayesian modelling and ii) concerned with the status of scientific hypotheses.

Our models of scientific inference use a binary propositional hypothesis variable, H, where h stands for the proposition that the hypothesis of interest is true (e.g., 'matter consists of molecules'). It is irrelevant for our discussion whether the hypothesis in question is deterministic or statistical, e.g., '85% of patients treated with drug D recover more quickly than patients not receiving any treatment'. The reader interested in the descriptive adequacy of a hypothesis may replace 'true' by 'descriptively adequate' and 'false' by 'not descriptively adequate' throughout.

For a body of evidence, $\mathcal{E}$, we shall be interested in the Bayesian probability of H being true, $P_{\mathcal{E}}(h)$. We freely admit that not all scientific inferences can be captured by our models, for example we only consider direct evidence for or against a hypothesis of interest; indirect evidence such as the No Alternative Argument of Dawid et al. (2015) is not considered here.

Next, we incorporate into our models that scientific hypotheses are typically not directly testable. It is rather some of their observable consequences which are testable (Bovens and Hartmann, 2003, p. 89). We shall use two sets of binary propositional consequence variables, the C_n and the D_k. The values c_n, respectively, d_k, stand for the proposition that the testable consequence C_n, respectively, D_k, of H holds.[3] Evidence pertaining to the first set of variables will vary for ev-

[3]Variety of Evidence reasoning in climate science has recently been analysed to proceed via different consequences of the hypothesis that temperatures are rising, see Vezér (2017). Consequence considered were patterns in temperature profiles in ice, rock and soil as well as the lengths of mountain glaciers and sizes of tree rings.

idential situations we compare, while the evidence pertaining to the D_k variables will not vary. $\mathcal{C}$ denotes the set of all consequence variables.

In our running example for the hypothesis H that 'matter consists of molecules', the testable consequences might be: i) 'the sky is blue' (Perrin, 1924, §83, pp. 197), ii) 'at room temperature, the viscosity of hydrogen is $0.88 \cdot 10^{-5} Nms^{-2}$' (Perrin, 1924, §46, pp. 107), iii) 'the pressure in in an emulsion is exponentially decreasing with decreasing depth' (Perrin, 1924, §54, pp. 129) and iv) 'black bodies emit electromagnetic radiation with the black body spectrum' (Perrin, 1924, §88-90, pp. 211).

Finally, measurements of testable consequences are modelled by evidential propositional variables. We use variables E_l to represent evidence pertaining to the C_n. Furthermore, the evidence variables F_l pertain to the D_k. e_l, respectively, f_l, stand for the proposition that a (series of) measurement(s) of the quantity of interest has resulted in some value. Let e^+ stand for the proposition that this value is consistent with C (e.g., 'the average measured quantity of helium produced by one gram of radium per day in our lab was 0.02mm^3'). As usual, closes misses count as hits. What counts as a close in a particular application depends. Evidential variables have finite arity which is obtained by suitably discretising the set of possible measurements.

For a variable V we also let $v^1 := v$ and for the negation we put $v^0 := \bar{v}$.

3.2 Topology of Bayesian Networks

We use Bayesian networks to represent and reason with probability functions defined over the space spanned by these variables. The topology of the directed acyclic graphs of the Bayesian networks is adopted from Bovens and Hartmann (2002).

The hypothesis variable H is the unique root node. The children of H are the consequence variables in $\mathcal{C}$. They have, in turn, a number of evidential variables as children. Every child of a consequence variable has exactly one parent. There are no further edges in the Bayesian network. See Figures 1 and 2 for example networks. The assumed independences can be read off the Bayesian networks using the usual d-separation criterion, see for example Pearl (2009). Consequence variables without children represent lines of scientific inquiry which are currently unexplored. A generalisation of the network topology is presented in Section 5.2.

This topology is generated by the following modelling choices regarding probabilistic independences and dependences. We first consider the independences, i.e., the missing edges. Firstly, H screens off every consequence variable from every other consequence variable:

$$C \perp C' | H \text{ for all different } C, C' \in \mathcal{C} \quad . \tag{1}$$

For example, the probability that the colour of the sky is blue, $P(c)$, depends on whether atomism (H) holds but does not also depend on the viscosity of hydrogen at room temperature (C').

Furthermore, a testable consequence variable screens C off its evidence from the hypothesis H

$$E \perp H | C \text{ for all } C \in \mathcal{C} \text{ and all children } E \text{ of } C. \tag{2}$$

The probability that the colour of the sky is observed to be blue depends on whether the sky is blue; but this probability does not also depend on the truth of atomism, $P(e^+|ch) = P(e^+|c)$.

Finally, a consequence variable C screens off its evidence variables

$$E \perp E' | C \text{ for all } C \in \mathcal{C} \text{ and all different children } E, E' \text{ of } C \quad . \tag{3}$$

This says that conditionalising on a consequence variable renders their children variables independent.[4]

We now turn to the edges present in our models. The probability of whether a testable consequence holds or not is directly influenced by whether the hypothesis of interest holds or not. Similarly, the probabilities of measurements coming out a certain way directly depends on whether the relevant testable consequence of the hypothesis holds. This motivates the edges and their orientations within our networks.

3.3 Probabilities for Bayesian Networks

To avoid unnecessary technical complications, we assume that all conditional probabilities in our Bayesian networks are non-zero.

That C is a consequence of H means that C is probabilistically entailed by H and hence that C is more likely under H than under its negation. We hence require (as do Bovens and Hartmann (2003)) that $P(c|h) > P(c|\bar{h})$ or equivalently

$$P(\bar{c}|h) < P(\bar{c}|\bar{h}) \quad . \tag{4}$$

[4] Evidence variables which are confirmationally independent regarding the hypothesis variable and their role in confirming the hypothesis with respect to different confirmation measures are investigated in Fitelson (2001). Note that conditional independence with respect to consequence variables (our approach) is not immediately transferable to conditional independence with respect to the hypothesis variable (Fitelson).

$P(\bar{c}|h)$ is a probability of false negatives, while $P(c|\bar{h})$ is a probability of false positives.

4 Varied Evidence

4.1 Varieties of Evidence

We are interested in how varied a body of evidence is with respect to the hypothesis of interest. We can think of two natural ways in which bodies of evidence formalised within the present models may vary thusly. Firstly and more importantly, the items in a body of evidence may pertain in different ways to different consequences of the hypothesis. This sort of variety is thus fully determined by the topology of the Bayesian network and is pursued here. The idea to understand variety in terms of the topology of the Bayesian network is not new, it is already in Bovens and Hartmann (2002, 2003).

Secondly, the items in a body of evidence may be informative to different degrees about consequences of the hypothesis. This second sort of variety depends on actual conditional probabilities, which is left to further study. Clearly, two bodies of evidence formalised within the present models may vary in both these senses at the same time and in further senses.

Our explication of the notion of varied evidence is but one way to formalise *some* intuitions regarding varied evidence. We side with (Howson and Urbach, 2006, p. 125) in doubting that the grand project of giving one explication which captures all senses of varied evidence can ever successfully be completed since our intuitions seem too vague.

4.2 Explicating Varied Evidence

A variety of objects is a set of objects in which few (very few or even no) objects look the same. Clearly, variety is the sort of thing which comes in degrees. A degree of variety then measures how different objects are in some fixed set of objects. A degree of variety thus measures the degree of dis-order or chaos in some set. The most widely used measure of (dis-)order in (the philosophy of) the sciences is Shannon Entropy, see Shannon (1948).[5]

[5] A variety of measures of entropy have been put forward. The well of measures of entropy shows no sign of drying up any time soon, see Crupi (2019); Csiszár (2008) for overviews of entropies and their axiomatic characterisations.

With this understanding of what 'variety of evidence' may mean and with models of scientific inference in place we now proceed to explicate the degree of variety of a body of evidence. While we think that this sense of variety is highly pertinent for the issues at hand, we do not want to suggest that our understanding (and hence our explanation) captures all relevant senses; the reliability independence sense of Bovens and Hartmann (2003); Claveau (2013) is one of the sense outside the scope of this paper, see Section 5.3 for more details.

We use $|\mathcal{E}|$ to denote the total number of items of evidence in $\mathcal{E}$ and $|C|$ for the number of children of consequence variable C. The fraction $\frac{|C|}{|\mathcal{E}|}$ then gives the proportion of items of evidence which are children of C. The variety V of a body of evidence $\mathcal{E}$ can then be defined by[6]

$$V(\mathcal{E}) := -\sum_{C \in \mathcal{C}} \frac{|C|}{|\mathcal{E}|} \cdot \log\Big(\frac{|C|}{|\mathcal{E}|}\Big) \ . \tag{5}$$

This is in close analogy with Shannon Entropy for probability functions w defined on some finite not-empty set of possible worlds Ω by

$$H(w) := -\sum_{\omega \in \Omega} w(\omega) \cdot \log(w(\omega)) \ .$$

Our explication of variety formalises the degree of uniformity with which the items of evidence pertain to a variety of testable consequences of the hypothesis.

Those who prefer science to be as objective as possible might appreciate that our topological explication is based on facts of the matter[7]: which item pertains to which consequence. The choice of a prior – consistent with the topology – does *not* affect the degree of variety. This can be seen as an answer to (Wayne, 1995, pp. 115-116) who claimed that Bayesians still had to give a satisfactory account of their judgement of diversity.

4.3 Shannon's Axioms for Explicating Varied Evidence

Shannon gave an axiomatisation of his measure in terms of three axioms: H1, H2 and H3.

Shannon's axiom $H1$ says that the measure of entropy should be continuous in the probabilities $w(\omega)$. In our discrete application (a finite number of items

[6] Adopting the standard convention that $0 \cdot \log(0) := 0$.

[7] See Colombo (2017) for some of the latest on objectivity in science.

of evidence are children of a finite number of consequences of the hypothesis), this amounts to the requirement that the variety of a large body of evidence varies only little, if one item of evidence is adopted by a new parent. This strikes us reasonable, small changes in the topology of the Bayesian network entail only small changes in the variety of evidence. Our measure V tracks this thought.

The degree of variety $V(\mathcal{E})$ increases with increasingly larger bodies of evidence in which every consequence variable has exactly one child. This is (Shannon, 1948, Section 6)'s axiom $H2$. This seems right, the variety of a body of evidence which has maximal variety (no two evidence variables look the same [have the same parent]) increases with the size of the body of evidence.

Shannon's third and final axiom H3 formalises "conditional entropy". Conditionals have long interested philosophers and a number of competing analyses have been put forward. It is hence not surprising that different measures of entropy do not satisfy Shannon's axiom H3.

In general, different measures of entropy disagree about which of two probability functions has greater entropy. For an illustration of different intuitively plausible measures of entropy disagreeing on which probability function has greater entropy see, e.g., (Landes and Williamson, 2013, Figure 1). For example, our variety measure (employing Shannon Entropy) pronounces on whether body of evidence $\mathcal{E}$ with $|C_1| = 5, |C_2| = 3, |C_3| = 2, |C_4| = 2$ or body of evidence $\mathcal{E}'$ with $|C_1'| = 4, |C_2'| = 4, |C_3'| = 3, |C_4| = 1$ has greater variety.[8] An argument as to whether $\mathcal{E}$ or $\mathcal{E}'$ has greater variety could be made either way: the evidence pertaining to C_1 and C_2 is more varied for $\mathcal{E}'$ while the evidence pertaining to C_3 and C_4 is more varied for $\mathcal{E}$. One could also argue that both bodies of evidence have equal variety.

4.4 Two Conditions: Novel Evidence and Adoption

We next identify two conditions every entropy measure employed in an explication of varied evidence ought to satisfy. If one of these conditions is satisfied, then it is intuitive which of two bodies of evidence has greater evidential variety. We shall hence only consider ordered pairs of bodies of evidence which satisfy one of these conditions. There may be further conditions under which such comparisons are intuitive.

Condition A: The variety of a body of evidence $\mathcal{E}$ increases when a

[8]For the record, $\mathcal{E}$ has a slightly greater variety score than $\mathcal{E}'$: $V(\mathcal{E}) = 1.3086 > 1.2861 = V(\mathcal{E}')$.

novel item of evidence is discovered which is the only-child of its parent consequence variable, see Figure 1 for a network representation. For this larger body of evidence $\mathcal{E}'$, we have that $V(\mathcal{E}) < V(\mathcal{E}')$ (Proposition 4).

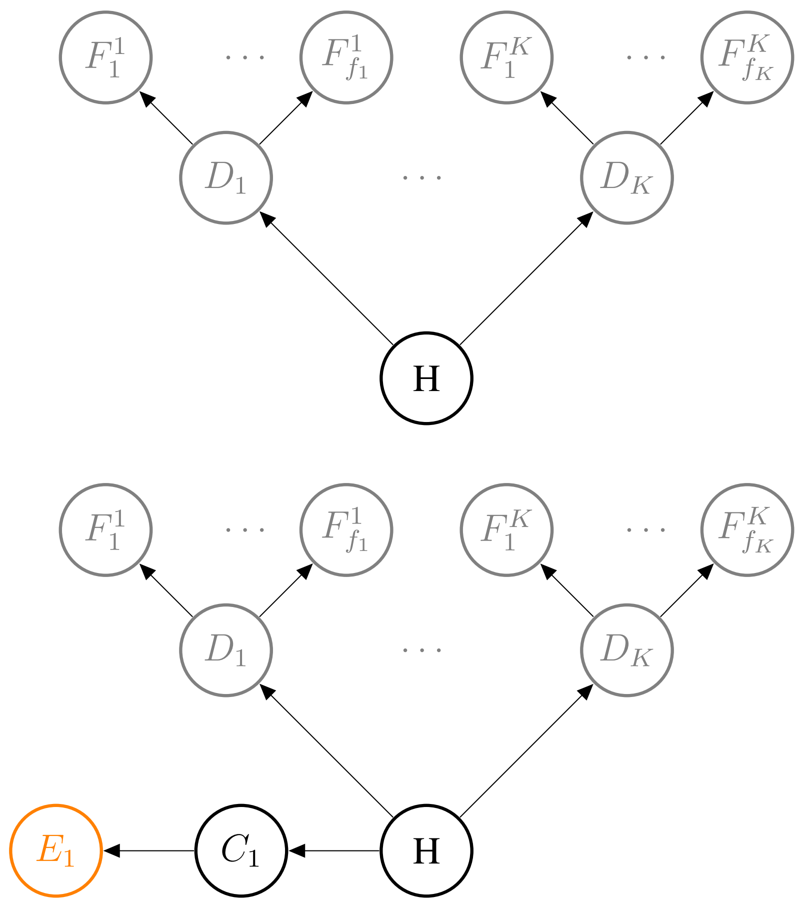

Figure 1: General Bayesian network model of Condition A in which a novel item of evidence has been added. It pertains to a consequence for which there was previously no evidence available. The body of evidence pertaining to the D variables is the same in both networks and is shaded grey.

Condition B: The variety of a body of evidence increases $\mathcal{E}$ when an item of evidence switches its parent variable from C to C' such that, prior to the switch, $|C| \geq |C'|+2$. Figuratively speaking: C' adopts a child of C such that C' ends up not having more children than C, see Figure 2 for graphical illustration. This is intuitively right, the more evenly spread out the items of evidence are, the greater the variety of evidence, ceteris paribus.

Denoting the body of evidence after the adoption by $\mathcal{E}'$, we now show that $V(\mathcal{E}) < V(\mathcal{E}')$. Shannon Entropy strictly increases when the probability of some $\omega \in \Omega$ is reduced by $\alpha \in (0,1)$ and added to $w(\nu)$ some $\nu \in \Omega$, if $w(\omega) - \alpha > w(\nu)$. This follows immediately from the strict concavity of H. We can conclude that $V(\mathcal{E}) < V(\mathcal{E}')$.

Our "continuous" measure V (compare to H1) satisfies H2 and agrees with these two intuitive comparisons of evidential variety.

4.5 The Non-Issue of Normalisation

Shannon Entropy is defined for probability functions while our measure of variety $V(\mathcal{E})$ is defined on formal representations of bodies of evidence. To strengthen the analogy, we normalised by dividing by $|\mathcal{E}|$. Had we not normalised, we could have defined

$$V_1(\mathcal{E}) := -\sum_{C \in \mathcal{C}} |C| \cdot \log(|C|) \quad . \tag{6}$$

Since we are not interested in the absolute variety of a body of evidence but rather in which body of evidence has the comparatively greater variety, the following proposition shows that normalising is unproblematic for bodies of evidence of equal size:

Proposition 1. *For bodies of evidence $\mathcal{E}, \mathcal{E}'$ with $|\mathcal{E}| = |\mathcal{E}'|$ it holds that*

$$V(\mathcal{E}) \geq V(\mathcal{E}'), \quad \textit{if and only if} \quad V_1(\mathcal{E}) \geq V_1(\mathcal{E}') \quad .$$

Non-trivial proofs of main arguments, such as this one, can be found in Appendix A.

Whatever the technical details, the key idea is this: A measure of entropy can be used to explicate the notion of varied evidence.

5 The Variety of Evidence Thesis

With an explication of variety of evidence in place, we can now proceed to offer an explication of the VET.

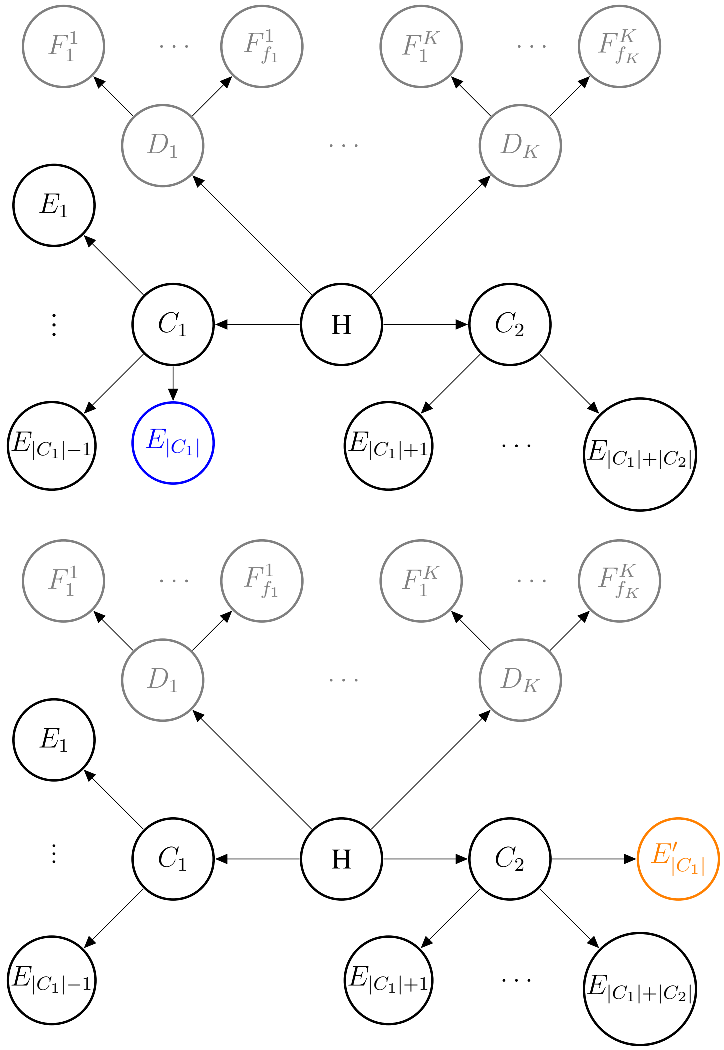

Figure 2: General Bayesian network model of Condition B where $E_{|C_1|}$ has switched parents and become a child of the consequence variable C_2. The switching variable is highlighted in blue (prior to the switch) and orange (after the switch).

5.1 Our Variety of Evidence Thesis

We follow (Claveau, 2013, p. 95) and state the thesis of interest in ordinary language

Variety of Evidence Thesis. *Ceteris paribus, the strength of confirmation of a hypothesis by an evidential set of independent and confirmatory items of evidence increases with the diversity of the evidential elements in that set.*

which we shall explicate in our models of scientific inference. We added the independence clause and the condition concerning confirmatory evidence[9] which we will discuss in Section 6.2.3 and Section 7. Here, we shall only consider two Scenarios. In Scenario A, two bodies of evidence $\mathcal{E}, \mathcal{E}'$ satisfy Condition A; in Scenario B they satisfy Condition B.

5.1.1 Explicating the Ingredients

Ceteris paribus means *all other* things being equal, respective conditional probabilities included. In both scenarios, (conditional) probabilities of the same variables remain unchanged. In Scenario B, this means that the conditional probabilities of $E_{|C_1|}$ given C_1 correspond to the conditional probabilities of $E'_{|C_1|}$ given C_2 and the conditional probabilities of C_1 given H have to equal those of C_2 given H.[10]

We explicate the term "independent items of evidence" as follows: evidence variables which share the same parent are conditionally independent given their parent. That is, we require that (3) holds.

How then shall we explicate the idea that evidence is *evidence for* the hypothesis? In Scenario A, the weakest condition one might imagine seems to be to say that e_1 is more likely given c_1 than given the complement $\bar{c}_1$. This can be expressed in terms of a "Bayes factor"

$$Bf_1 := \frac{P(e_1|c_1)}{P(e_1|\bar{c}_1)} > 1 \ , \tag{7}$$

equivalently as $P(e_1|c_1) > P(e_1|\bar{c}_1)$. Traditionally, a Bayes factor is defined as the conditional probability of observing the evidence given the hypothesis over

[9]We suspect a slip in Claveau's in his statement of the VET. Claveau's proofs in the appendix of his paper only deal with evidence which is in agreement with the predicted value. Furthermore, in (Claveau, 2011, p. 241) he writes "*I take evidence to be always evidence* for *a specific proposition*" [emphasis original]. He might have simply forgotten to state this convention in Claveau (2013). Bovens and Hartman and Earman also only consider confirmatory evidence.

[10]The case for treating two observable consequence on equal epistemological footing in (the reconstruction of) scientific inference has recently been defended in Parkkinen (2016).

the conditional probability of the evidence given the negation of the hypothesis. Throughout, we (ab-)use the term “Bayes factor” for a quotient of probabilities conditional on a consequence of the hypothesis rather than the hypothesis itself.

In Scenario B, the body of evidence is split into two parts: i) the evidence variable which has a change in parents and ii) all the other evidence variables. It appears that the weakest understanding of *evidence for* the hypothesis is that this evidence is more likely when the; respective; consequence variable is true than when the consequence variable is false. Furthermore, ignoring the evidence variable being adopted, the evidence for C_1 is stronger than the evidence for C_2. Finally, the entire body of evidence needs to be evidence for the hypothesis. Again, we can formalise this in terms of Bayes factors (in our sense):

$$Bf_{C_1} := \frac{P(e_{C_1}|c_1)}{P(e_{C_1}|\bar{c}_1)} = \frac{P(e'_{C_1}|c_2)}{P(e'_{C_1}|\bar{c}_2)} > 1 \tag{8}$$

$$Bf_{C_1-1} := \prod_{n=1}^{|C_1|-1} \frac{P(e_n|c_1)}{P(e_n|\bar{c}_1)} > \prod_{g=1}^{|C_2|} \frac{P(e_{|C_1|+g}|c_2)}{P(e_{|C_1|+g}|\bar{c}_2)} =: Bf_{C_2} \tag{9}$$

$$Bf_{\mathcal{E}} := Bf_{C_1} \cdot Bf_{C_1-1} \cdot Bf_{C_2} = \prod_{n=1}^{|C_1|} \frac{P(e_n|c_1)}{P(e_n|\bar{c}_1)} \prod_{g=1}^{|C_2|} \frac{P(e_{|C_1|+g}|c_2)}{P(e_{|C_1|+g}|\bar{c}_2)} > 1 \ . \tag{10}$$

These evidential conditions (8)–(10) are jointly satisfied, if, for example, every evidence variable pertaining to C_1 or C_2 has the same Bayes factor which is strictly greater than one.

5.1.2 Our Explication

We are ready to state our

> **Explication of the Variety of Evidence Thesis**
> In Scenario A and in Scenario B, it holds that $P_{\mathcal{E}}(h) < P_{\mathcal{E}'}(h)$.

Note that our explication of the VET in Table 2 does not require that the evidence variables $E_{|C_1|}$ and $E'_{|C_1|}$ have the same arity. Worth pointing out is that; with the exception of variable E_1 in Scenario A and variables $E_{|C_1|}, E'_{|C_1|}$ in Scenario B; there is no requirement that any particular item of evidence is confirmatory, bodies of evidence may be discordant. Also observe that the only ceteris paribus condition we impose for any particular evidence variable is (8).

Network Topology	Condition A
Confirmatory Evidence	$Bf_1 > 1$

Network Topology	Condition B
Confirmatory Evidence	$Bf_{C_1} > 1$ $Bf_{C_1-1} > Bf_{C_2}$ $Bf_{\mathcal{E}} > 1$
Ceteris Paribus	$\frac{P(e_{C_1}\mid c_1)}{P(e_{C_1}\mid \bar{c}_1)} = \frac{P(e'_{C_1}\mid c_2)}{P(e'_{C_1}\mid \bar{c}_2)}$ $P(c_1\mid h) = P(c_2\mid h)$ $P(c_1\mid \bar{h}) = P(c_2\mid \bar{h})$

Table 2: Scenario A (top) and Scenario B (bottom) under which more varied evidence ought to be more confirmatory, according to our explication of the VET.

The reader might be surprised that we include Scenario A in our explication: is it not vacuous? She might think that Scenario A is simply a case of *more* confirmatory evidence contributing more to confirmation. While it is in general not the case that more confirmatory evidence leads to more confirmation, see (Carnap, 1962, p. 382) and (Bovens and Hartmann, 2003, Section 4.3), the addition of one further confirmatory evidence variable does increase confirmation in our evidential (network) structure (Section 6.1). We reply that we include Scenario A in the explication of the VET since it captures one of our core intuitions of variety. Intuitively, variety increases when an object is added to a set which does not contain any object of this sort. This is the thought which makes Shannon's axiom H2 so appealing, see Section 4.3. It is easy to show, that Shannon Entropy does capture this intuition, see Proposition 4.

5.2 Alternative Explications

We are interested in which of two bodies of evidence is more confirmatory. Hence, we are interested in which of the posterior probabilities $P_{\mathcal{E}}(h), P_{\mathcal{E}'}(h)$ is greater. Our choice of confirmation measure is thus the difference of two posterior probabilities. Those interested in other confirmation measures in this context are referred to (Bovens and Hartmann, 2002, Appendix M) for discussion. Trivially, every choice of a confirmation measure which is ordinally equivalent to the difference measure yields same the comparisons of posterior probabilities. Thus, it delivers the same verdict about the VET.

One could also allow edges between the D_k or allow the children of the D_k to have multiple parents among the D_k. However, this renders the defined measure $V(\mathcal{E})$ meaningless since it ought to matter how exactly the evidence variables pertain to their parents. While the measure $V(\cdot)$ is meaningless, one can still talk about the variety V of the restrictions of $\mathcal{E}, \mathcal{E}'$ to the consequence variables where they disagree $(C_1, \ldots, C_N)$. If one thusly defines *comparative degrees of variety*, then all results derived here hold, too.

Alternatively, one could define a measure of variety V' under which the variety of the evidence pertaining to the C_n is additive, i.e., $V'(\mathcal{E}) = V'(\mathcal{E}_D) + V'(\mathcal{E}_C)$ where the index denotes the restriction to these variables. Again, the obtained results hold in this alternative approach, too.

In the same vein, one can allow for consequence variables D_k which have arbitrary infinite arity and our results still hold – still assuming non-zero conditional probabilities.

Another generalisation is to employ a hypothesis variable with countable (finite or infinite) arity and consequence variables which all have the same countable (finite or infinite) arity. We have so far only be interested in a *binary* hypothesis variable and *binary* consequence variables. We require that the consequence variables have equal arity to ensure that our explication of variety makes sense for these more general models. Given – we think – natural ceteris paribus conditions, the established technical results also hold for these more general models as we show in Appendix B.

Further properties of our explication of the VET are discussed in more detail in Section 7.

5.3 Comparison to Bovens and Hartmann and Claveau

In the approaches of Bovens and Hartmann and Claveau, bodies of evidence vary by the agent's judgements of the dependence of the sources which provided the evidence. Imposing ceteris paribus conditions, they study the confirmation provided by different bodies of evidence under different (in-)dependence assumptions. The key idea is that the more independent the sources the greater their notion of variety of evidence. With thusly understood variety, they discover cases in which less varied bodies of evidence confirm the hypothesis more strongly than more varied bodies of evidence.

(Bovens and Hartmann, 2003, Section 4.3) also describes cases in which more evidence (in accordance with consequences of the hypothesis) decreases confirmation of the hypothesis. In (Bovens and Hartmann, 2003, Section 4.4), they discuss this as a failure of the VET. They hence seem to agree with our assessment made in Section 5.1 that a novel item of evidence pertaining to a previously untested consequence increases the variety of the evidential set. In our models, our explication of the VET holds unequivocally in such a case, we show this in Section 6.1.

Since the main contributions of these works is to show that there are cases in which purportedly varied evidence confirms less strongly than less varied evidence, failing to offer an explication of variety is, in our view, not the main issue with these arguments. It seems that some explication of their notion of variety should be possible. We feel that the main problems with their arguments lie elsewhere, their use and construal of the notion of reliability of scientific instruments.

We avoid the notion of reliability by considering variety of evidence with respect to the hypothesis. Conceptually, we *exclude* reliability from our analysis of the VET since the notion of reliability is – neither directly nor indirectly – invoked in the VET. Our analysis is thus closer in spirit to the epistemological thesis under investigation. Reliability is an exogenous part of the model and subsumed here in the all-important Bayesian prior probability function.

(Bovens and Hartmann, 2002, p. 33) model scientific instruments as either being fully reliable or as being a complete randomiser: either a scientific instrument distinguishes correctly whether the pertinent testable consequence C holds [fully reliable] or it reports the value predicted by the pertinent testable consequence C with fixed probability a, regardless of whether C holds or not [complete randomiser]. In the latter case, all measurements from this instrument provide no information whatsoever.

We agree with (Claveau, 2013, p. 96) "*The unreliable sources in their model are not like unreliable sources in actual science (i.e., their unreliable sources are randomly biased while systematic bias is far more likely to be the issue)*". There is another problem with modelling scientific instruments as randomisers. Two conflicting measurements from the same instrument force us to conclude that the instrument is a randomiser. This seems much too strong.

While Claveau's model of (un-)reliability is an improvement, it suffers from a similar problem. It is inconsistent with the agent's beliefs that the readings of an instrument on two runs of the same experiment differ, see (Claveau, 2013, p. 105) and (Claveau and Grenier, 2018, Table 2). This cannot be right. When repeating a scientific experiment, it is normally very hard to guarantee that the (exact) initial and boundary conditions hold on multiple occasions. One should hence expect that the measurements disagree (to some degree) without suspecting a problem with measuring instruments. It hence seems, to us, that Claveau's formalisation of (un-)reliability of scientific instruments does not adequately capture our understanding of (un-)reliable scientific instruments.

Our model is more general in that we consider: a) consequence variables which may have three or more children as suggested by (Bovens and Hartmann, 2002, p. 65, (II)); see also Hartmann and Bovens (2001), b) non-binary evidence variables, c) non-binary hypothesis and non-binary consequence variables, see Appendix B for details, d) a relationship between evidence variables and consequence variables which is always non-deterministic and informative at the same time and e) bodies of evidence which may be discordant.

5.4 Comparison to the Correlational and the Eliminative Approach

Our approach extends results in Earman (1992) since we consider bodies evidence which may not be entailed by the hypothesis, some items of evidence may even be inconsistent with the hypothesis.

There do not seem to be deep connections to the eliminative approach, in general. For two competing (mutually exclusive and jointly exhaustive) hypothesis $(H, \bar{H})$ we readily find for their posterior probabilities $P_{\mathcal{E}}(h) + P_{\mathcal{E}}(\bar{h}) = 1$. Hence, evidence which confirms H more strongly is better at disconfirming the single competing hypothesis.[11]

With the generalisation to higher arity variables, our approach may be seen as an eliminative approach, see Appendix B. We leave it to the reader to judge whether this does justice to the spirit of the eliminative approach.

6 The Status of the Variety of Evidence Thesis

We now discuss the status of the VET in our models.

6.1 Novel Evidence – Scenario A

We consider a fixed body of evidence $\mathcal{E}$ and add a new item of evidence which pertains to a consequence C for which there is no other evidence available. The

[11] A rare meeting of Bayes and Popper, see also Section 6.2.2 for disconfirmatory evidence.

variety of this new body of evidence $\mathcal{E}'$ has increased (Proposition 4). The posterior probability of the hypothesis increases, if the added new item of evidence is in agreement with its predicted value:

Theorem 1. *In case of Condition A*

$$\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \mathrm{sign}(P(e_1|\bar{c}_1) - P(e_1|c_1)) = \mathrm{sign}(1 - Bf_1) \ .$$

Put differently

Corollary 1 (The VET – Scenario A)**.** *Condition A and* $P(e_1|c_1)/P(e_1|\bar{c}_1) = Bf_1 > 1$ *imply*

$$P_{\mathcal{E}}(h) < P_{\mathcal{E}'}(h) \ .$$

In other words, the more independent items of evidence which confirm the consequence they pertain to, the greater the posterior probability of the hypothesis.

6.2 Adoption – Scenario B

As explained in Section 4.2, the variety of a body of evidence increases, when an item of evidence changes parents where the new parent has – after the switch – not more children than the previous parent; Condition B. The Bayesian networks for the corresponding bodies of evidence $\mathcal{E}, \mathcal{E}'$ are depicted in Figure 2, with $|C_1| \geq |C_2| + 2$. For the remainder of this section, $\mathcal{E}, \mathcal{E}'$ will denote an ordered pair of such bodies of evidence.

We obtain a surprisingly simple formula for the sign of $P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)$:

Theorem 2. *Condition B and the Ceteris Paribus Conditions jointly entail that*

$$\begin{aligned}&\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h))\\ =&\,\mathrm{sign}\Big((1 - Bf_{C_1})(Bf_{C_1-1} - Bf_{C_2})\Big(Bf_{\mathcal{E}} - \frac{P(\bar{c}|h)\cdot P(\bar{c}|\bar{h})}{P(c|h)\cdot P(c|\bar{h})}\Big)\Big) \ . \end{aligned} \tag{11}$$

We note, all terms using the extra evidence $\vec{f}$ and the prior probability of h have disappeared. The fate of the VET, in our models, hinges on Bayes factors and how likely the consequences are given that the hypothesis is true/false.

Next, we consider different evidential situations. First, we see what happens, if the body evidence is confirmatory on the whole. Then, we consider bodies of evidence which are disconfirmatory on the whole. Finally, we consider more discordant evidence.

6.2.1 Confirmatory Evidence

Turning to confirmatory evidence, we find by applying Theorem 2 that for evidence which is more probable under a consequence C than under its negations that:

Corollary 2 (The VET – Scenario B)**.** *Condition B, the Ceteris Paribus Conditions, $Bf_{C_1} > 1$ and $Bf_{C_1-1} > Bf_{C_2}$ jointly entail that*

$$\text{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \text{sign}\Big(\frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})} - Bf_{\mathcal{E}}\Big) \ .$$

In other words, the following are equivalent:

- the VET holds,
- the entire body of evidence is sufficiently confirmatory, $\frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})} < Bf_{\mathcal{E}}$.

So, when is it the case that the VET holds? Using that $Bf_{\mathcal{E}} > 1$, the VET holds if $P(\bar{c}|h) \cdot P(\bar{c}|\bar{h}) = (1 - P(c|h))(1 - P(c|\bar{h})) \leq P(c|h) \cdot P(c|\bar{h})$. We thus find that the VET holds, if[12]

$$P(\bar{c}|h) \leq P(c|\bar{h}) \ . \tag{12}$$

We now discuss the plausibility of our main results for confirmatory evidence, Corollary 1 and Corollary 2, in the Perrin case. Discussion of the status of the VET, in particular failures of the VET, can be found in Section 7, after we have taken a look at disconfirmatory and discordant bodies of evidence.

Let us assume that we have conducted a number of independent experiments to measure the viscosity of hydrogen at room temperature (C_1) and have not investigated any other consequence of H ('matter consists of molecules'). All measurements were (around) $0.88 \cdot 10^{-5} Nms^{-2}$. Our degree of belief in the hypothesis being true will take some value d. Further experiments regarding the viscosity of hydrogen at room temperature which agree with the predicted value will do little to change our belief in H being true, even if these experiments are carried out at different places, at different times, by different people with different equipments.

But why then was Poincaré prepared to accept atomism when he started with a low prior belief in it[13]? Corollary 2 provides an answer. All consequences

[12] there are two equivalent formulations, i) $P(\bar{c}|\bar{h}) \leq P(c|h)$ and ii) $P(c|h) + P(c|\bar{h}) \geq 1$.

[13] Let us simply assume for a moment that Poincaré was a Bayesian.

of atomism are (very close to being) deductive consequences of the hypothesis that atomism is true, $P(c|h) \approx 1$. Thus, $\frac{P(\bar{c}|h)\cdot P(\bar{c}|\bar{h})}{P(c|h)\cdot P(c|\bar{h})}$ is less or equal than one. Furthermore, the evidence obtained is very unlikely under the assumptions that these consequences fail to hold, i.e., the Bayes factors are large. $P_{\mathcal{E}}(h) \ll P_{\mathcal{E}'}(h)$ follows for even a small degree of prior belief in the hypothesis being true.

Our models indeed track this line of thought that there is an upper bound to confirmation by only investigating a single consequence – even when all experiments turn out as predicted:

Corollary 3. *If $E_1, \ldots, E_{|C_1|}$ are the children of C_1, then for all possible measurements $E_1 = e_1, \ldots, E_{|C_1|} = e_{|C_1|}$*

$$P(h|e_1 \ldots e_{|C_1|}\vec{f}) < P(h|c_1\vec{f}) = \frac{1}{1 + \frac{P(\bar{h})\cdot P(c_1|\bar{h})\cdot P(\vec{f}|\bar{h})}{P(h)\cdot P(c_1|h)\cdot P(\vec{f}|h)}} < 1 \ .$$

The upper bound for the confirmation is smaller the further C_1 is from being a perfect indicator from H, i.e., the greater $P(c_1|\bar{h})$ and the smaller $P(c_1|h)$. That is, the greater the error probabilities $P(c|\bar{h})$ and $P(\bar{c}|h)$.

Instead of conducting yet another experiment on the viscosity of hydrogen let us contemplate conducting an experiment on the radiation emitted by a black body (C_2). If we detect the predicted black body radiation, then we are confident – to some degree – that black bodies do emit black body radiation. We hence become more confident that the consequences of H are borne out across a variety of phenomena. Our degree of belief in the hypothesis H being true increases non-negligibly beyond our prior degree of belief d.

This thinking is captured by Corollary 2 for the situation when there is no further background evidence, i.e., $|C_2| = 0$ and $\vec{f} = \emptyset$. Corollary 1 and Corollary 2 demonstrate that our Bayesian models capture this reasoning in a wider range of circumstances, e.g., when there is further background knowledge ($\vec{f} \neq \emptyset$) and/or a few experiments regarding the black body radiation had measured results in agreement with the predicted spectrum ($|C_2| > 0$).

In fact, we can calculate when it pays more to stop investigating the viscosity of hydrogen and move to investigating black body radiation. We shall assume that $P(c_1|h) = 1 = P(c_2|h)$, i.e., (12) holds. To ease notation we put:

$$\chi_{1s} := \prod_{n=1}^{|C_1|-1} P(e_n|c_1^s) \quad \text{and} \quad \chi_{2s} := \prod_{g=1}^{|C_2|} P(e_{|C_1|+g}|c_2^s) \ . \tag{13}$$

With this bit of notation we obtain

Proposition 2. *If Condition B, the last two ceteris paribus condition and if* $P(c_1|h) = 1 = P(c_2|h)$ *hold, then*

$$\operatorname{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \operatorname{sign}\Big(\frac{\sum_{j,l=0}^{1} P(c_1^j|\bar{h})P(c_2^l|\bar{h})P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}}{P(e'_{|C_1|}|c_2)} - \frac{\sum_{j,l=0}^{1} P(c_1^j|\bar{h})P(c_2^l|\bar{h})P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}}{P(e_{|C_1|}|c_1)}\Big) \ .$$

While this formula might appear somewhat opaque at first glance, we can already note that all the background $\vec{f}$ and the prior probability $P(h)$ have disappeared. In order to make progress, we need to have that the conditional probabilities of $E_{L+1}|C_1$ and $E'_{L+1}|C_2$ are – at least for certain truth values – comparable. Making such an assumption we find

Proposition 3. *If Condition B, the last two ceteris paribus condition,* $P(c_1|h) = 1 = P(c_2|h)$ *and if* $P(e_{|C_1|}|c_1) = P(e'_{|C_1|}|c_2) > P(e'_{|C_1|}|\bar{c}_2)$ *hold, then*

$$\operatorname{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \operatorname{sign}\Big(\frac{\alpha_{0,0} \cdot \beta_{0,0} + \alpha_{0,1} \cdot \beta_{0,1}}{|\alpha_{1,0} \cdot \beta_{1,0}|} - \prod_{n=1}^{|C_1|-1} \frac{P(e_n|c_1)}{P(e_n|\bar{c}_1)}\Big) \ ,$$

where the α *and* β *are parameters independent of* $|C_1|$ *which are defined in* (19).

In a given epistemic situation (α and β parameters, $P(e_{|C_1|}|\bar{c}_1)$, $P(e'_{|C_1|}|\bar{c}_2)$ and $|C_2| \geq 0$ fixed to arbitrary values) there comes a point at which evidence in favor of black body radiation (C_2) is more confirmatory than evidence in favor of the predicted viscosity of hydrogen (C_1), the second summand eventually outgrows the first. For example,

Corollary 4. *Under the assumptions of Proposition 3, if there also exists some* $\delta > 0$ *such that* $\frac{P(e_n|c_1)}{P(e_n|\bar{c}_1)} > 1 + \delta$ *is constant for all* n*, then for bodies of evidence* $\mathcal{E}$ *with*

$$|C_1| \geq 1 + \log(1+\delta) - \log\Big(\frac{\alpha_{1,0} \cdot \beta_{1,0}}{\alpha_{0,0} \cdot \beta_{0,0} + \alpha_{0,1} \cdot \beta_{0,1}}\Big)$$

it holds that

$$P_{\mathcal{E}}(h) < P_{\mathcal{E}'}(h) \ .$$

The main point is this, there comes a point in life when investigating the exact same consequence yet again cannot provide significant further confirmation for the hypothesis of interest. Investigating a different consequence becomes more confirmatory. A weaker version of this result is given in (Howson and Urbach, 2006, p. 95). Our argument is more general in that items of evidence are not necessarily logical consequences of the testable consequences, some items of evidence may even be strongly disconfirmatory in our result.

6.2.2 Disconfirmatory Evidence

Having discussed evidence confirmatory bodies of evidence, we now turn to evidence which does not confirm to expectations. That is, we consider evidence which is more likely when consequences of the hypothesis are false.

Theorem 3. *Condition B, the Ceteris Paribus Conditions, $Bf_{C_1} < 1$ and $Bf_{C_1-1} < Bf_{C_2}$ jointly entail that*

$$\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = -\,\mathrm{sign}\Big(Bf_{\mathcal{E}} - \frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})}\Big) \ .$$

Theorem 3 says that a more varied body of evidence which speaks *against* testable consequences entails a higher probability of H being true than a less varied body of evidence, if $P(c|h) \cdot P(c|\bar{h})$ is large. In other words, the probability of H being true is lower, if evidence against the testable consequences is clustered around few or even just a single testable consequence.

Those interested in showing that a scientific hypothesis is (likely) false are thus interested in discovering disconfirmatory independent items of evidence which are concentrated on as few as possible different testable consequences, ceteris paribus.

Let us consider whether this intuitively right. Assume for the moment that the C_n are deductive consequences of the hypothesis. So, if only a single C_n is known to be false, then we know that H is false and assign zero posterior probability to H being true. In the running Perrin case, if the experiments conclusively rule out that black bodies emit black body radiation, then we can be sure that matter does not consist of molecules (holding the auxiliary background assumptions fixed as usual). A body of evidence which strongly discredits belief in black body radiation and in that the viscosity of hydrogen is $0.88 \cdot 10^5 Nms^{-2}$ at room temperature, makes for a (very) low, but non-zero, degree of belief in atomism.

We can now use a continuity argument to pass from categorical ruling out of a consequence C to evidence which strongly indicates (but not entails) the falsity of C: a body of evidence strongly telling against one testable consequence C is preferable for disconfirming the hypothesis to amassing a body of evidence which discredits the belief in the truth of multiple testable consequences, ceteris paribus. This thought is formalised in Theorem 3 – if $P(c|h) \cdot P(c|\bar{h})$ is large.

On the other hand, for small $P(c|h) \cdot P(c|\bar{h})$, i.e., when $P(\bar{c}|h) \geq P(c|\bar{h})$, then concentrating disconfirmatory evidence on a single consequence does not do so much for hypothesis dis-confirmation. This is because one suspects that a plausible reason for the disconfirmatory evidence is the fact that $\bar{c}$ is the likely cause of observing such evidence. Whereas a diverse body of disconfirmatory evidence is better at decreasing the belief in the hypothesis since one has good evidence for multiple consequences being false.

6.2.3 Discordant Evidence

There are a wide range of cases in which some of the evidence confirms a hypothesis while other evidence speaks against it. We here only consider a case in which one item of evidence, $E_{|C_1|}$, goes against the other items of evidence. We easily find

Theorem 4. *Condition B, the Ceteris Paribus Conditions and one of the following two conditions*

- $Bf_{C_1} > 1$ *and* $Bf_{C_1-1} < Bf_{C_2}$
- $Bf_{C_1} < 1$ *and* $Bf_{C_1-1} > Bf_{C_2}$

jointly entail that

$$\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \mathrm{sign}\Big(Bf_{\mathcal{E}} - \frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})}\Big) \ .$$

Let us consider an overall confirmatory body of evidence $\mathcal{E}$, $Bf_{\mathcal{E}} > 1$, which is discordant with $Bf_{C_1} < 1$ and $Bf_{C_1-1} > Bf_{C_2}$. Let us also suppose that $Bf_{\mathcal{E}} > \frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})}$. Theorem 4 expresses the thought that making the item of evidence which is discordant with other items of evidence more prominent by switching it to a new parent with fewer children makes for a comparatively lower posterior probability of the hypothesis. Let us see whether this is right.

The theorem says that evidence in disagreement with one predicted value according to C_1 leads to a lower posterior probability of h, if the other evidence, which is in agreement with the predicted value, pertains to another testable consequence (C_2), other things being equal. This is in line with our reasoning for the disconfirmation case. Disconfirmation is strongest, if disconfirmatory evidence is unopposed.

We conclude our discussion of bodies of evidence (Sections 6.2.1–6.2.3) by stating the hope that the reader agrees with us that our models also pronounce rightly on disconfirmation and discordant evidence for the considered cases.

7 Discussion

We now discuss the independent evidence condition (3) and the condition that $P(c|\bar{h}) \geq P(\bar{c}|h)$ which is (12). We consider two respects: i) their applicability in science and ii) their adequacy for explicating the VET.

7.1 Independent Evidence

Addressing both respects we point out that idealisations made, such as the probabilistic independence conditions [(1)–(3)], may only hold approximately in applications/may only be approximately adequate for explication of the VET. Since our main results (Corollary 1 and Corollary 2 for $P(\bar{c}|h) \cdot P(\bar{c}|\bar{h}) < Bf_{\mathcal{E}} \cdot P(c|h) \cdot P(c|\bar{h})$) are *strict* inequalities, these inequalities continue to hold in a neighbourhood of the idealisations, too. These neighbourhoods are the larger, the stricter the inequalities. In particular, in cases in which the evidence which is in good agreement with the predicted value, very large Bayes factors, we expect the VET to be robust under non-drastic changes of idealisations made.

As for the applicability of the assumption of independent evidence we point to (Fitelson, 2001, Footnote 20) who defends an idealisation of *independent evidence* thusly: "*I do not mean to suggest that confirmational independence can be used to undergird all of our intuitions about the value of diverse evidence. But, I do think that there are* many important scientific cases *that fit this mold*" (our emphasis). We agree with these sentiments. However, when evidence is strongly dependent, for instance on sources, then neither our models nor our explication apply.

Fitelson does not include consequences of the hypothesis into his model. The probabilistic independence he is thus concerned with is that of $E \perp E'|H$ since the evidence *directly* dis-confirms the hypothesis. The corresponding DAG is $E \leftarrow H \rightarrow E'$. Our model includes consequences and the evidence (dis-)confirms the hypothesis *through* testable consequences. The corresponding DAG is $E \leftarrow$

$C \rightarrow E'$. With this in mind, it seems permissible to transfer Fitelson's arguments for $E \leftarrow H \rightarrow E'$ in his model into an argument for $E \leftarrow C \rightarrow E'$ in our model.

Concerning the adequacy of the independence assumption for the explication of the VET, we feel that the intuitive appeal of the VET is mainly driven by judgements of independence. Again, we refer to (Fitelson, 2001, p. 131) who is thinking along similar lines: "*I suspect that the notion of* independent *evidence can undergrid, at least partially, (some of) our intuitions about the significance of* diverse *evidence. At least one recent philosophers seems to share this suspicion. Sober (1989) shows [...]*" (emphasis original).

As we discussed in Section 6.2.3, varied discordant evidence supports the hypothesis less than less varied evidence, ceteris paribus and under certain conditions. Out of this discussion a proponent of the VET might re-examine her intuitions and discover that they were mainly driven by an implicit assumption of *independent* evidence which *supports* the hypothesis. The intuition regarding dependent and/or discordant evidence may be felt less clearly or less strongly.

Strongly dependent evidence is another kettle of fish altogether. While we have not considered strongly dependent evidence here, we did cast some doubt on the applicability of the models by Bovens and Hartmann and Claveau for the philosophy of science, in which their formulation of the VET fails to hold for strongly dependent evidence. Clearly, there are many more ways (to formalise) in which evidence could be dependent, e.g., Claveau and Grenier (2018) – none of which we considered here.

7.2 Error Rates and Bayes Factors

But what about condition (12) that $P(c|\bar{h}) \geq P(\bar{c}|h)$ – or equivalently that $P(c|h) + P(\bar{c}|h) \geq 1$ and $P(c|h) \geq P(\bar{c}|\bar{h})$?[14] The condition that $P(c|h)$ is large is plausible in applications in science. After all, C is a *consequence* of H. In our running example, the consequences are (almost) deductive consequence of the hypothesis. So, $P(c|h)$ is only marginally – if at all – less than one. Hence, any addition of a non-negligible amount of probability will make this sum greater or equal than one, i.e., (12) holds. The same is true for the consequences considered in Landes et al. (2017); Vezér (2017). The VET holds in all these cases.

Of course, in some applications $P(c|\bar{h}) < P(\bar{c}|h)$ holds. See Maxim et al.

[14] In statistical lingo, $P(c|h)$ describes a rate of true positives (sensitivity) while $P(\bar{c}|\bar{h})$ describes a rate of true negatives (specificity). $P(c|\bar{h}) \geq P(\bar{c}|h)$ may also be read as saying that an error of Type II is more or equally likely than a Type I error.

(2014) for more discussion. (Maxim et al., 2014, Table 1) suggests that in medicine there are many tests which are more specific than sensitive. In cases with $P(c|\bar{h}) < P(\bar{c}|h)$ the VET fails, if and only if

$$\frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})} \geq Bf_{\mathcal{E}} \ . \tag{14}$$

In other words, the *VET fails*, if and only if the evidence is relatively weak ($Bf_{\mathcal{E}}$ small) and the consequences are relatively unlikely ($P(c|h)$ and $P(c|\bar{h})$ are small). Why does the VET fail to hold? What is going on?

Let us first consider $P(c|\bar{h}) \approx 0$ where the effect occurs par excellence. Under the standing convention that all other conditional probabilities are not very close to zero, this entails that $P(c\bar{h}) \approx 0$ and hence establishing that a single consequence of the hypothesis is true, establishes the hypothesis. Ceteris paribus, weakly confirmatory bodies of evidence which do better at establishing a single consequence are more confirmatory than bodies of evidence which spread the evidence over a number of consequences.

Ceteris paribus, the more confirmatory the entire body of evidence, the stronger the varied body of evidence establishes a consequence. So, this difference in confirmation is stronger, the weaker the confirmatory value of the entire body of evidence.

Furthermore, the greater $P(c|\bar{h})$ becomes, the smaller this effect becomes. At $P(c|\bar{h}) = 0.5$ we have $P(c|h) + P(c|\bar{h}) > 1$ (since $P(c|h) > P(c|\bar{h})$ formalises the notion of a consequence, see (4)). At this point, the VET holds universally. This is because the consequences are a priori relatively likely. Obtaining evidence for a single consequence does relatively little for confirming the hypothesis. Ceteris paribus, confirmation of the hypothesis increases with increased diversity of the body of evidence.

Before going into details, let us observe that the key equation (11) and its derivation do not depend on the number of evidence variables pertaining to consequence variables. The entire contribution of the evidence to (11) is via Bayes factors. This means that our results hold in evidential situations in which C_2 has more children than C_1, too. The relevant property is not the number of evidence variables but rather the strength of the evidence, i.e., $Bf_{C_2} < Bf_{C_1-1}$ which is (9). This suggests a possible explication of the VET in terms of the strength of evidence rather than in terms of the topology a Bayesian network. For the adoption scenario, this explication may be obtained from ours by simply dropping our requirement that $|C_1| \geq |C_2| + 2$. Details are left to future work.[15]

7.2.1 Applications in Science

Quantifying the percentage of applications in science in which (12) holds provides an enormous challenge. More feasible is an investigation of the parameter space in which the VET fails. Fixing $Bf_{C_2} < Bf_{C_1-1}$ and $Bf_{|C_1|} > 1$, the parameter space to investigated can be parametrised by three variables: $Bf_{\mathcal{E}}, P(c|h), P(c|\bar{h})$. Solving (14) for $P(c|\bar{h})$ with parameter $Bf_{\mathcal{E}}$ we find that the following equality fully describes the border of VET failure in the $P(c|h) - P(c|\bar{h})$-plane:

$$P(c|\bar{h}) = \frac{P(\bar{c}|h)}{P(\bar{c}|h) + P(c|h)Bf_{\mathcal{E}}} = \frac{1 - P(c|h)}{1 - P(c|h) \cdot (1 - Bf_{\mathcal{E}})} \ . \tag{15}$$

To investigate the parameter space, let us consider Bayes factors which traditionally formalise evidence in favour of a hypothesis. A Bayes factor of one is neither evidence for nor against the hypothesis, a Bayes factor of three and greater is taken as positive evidence for the hypothesis, a Bayes factor of 20 and greater is taken to represent strong evidence for the hypothesis and a Bayes factor of 150 and greater is very strong evidence. The adoption case involves at least two evidence variables. The larger the Bayes factor, the smaller the area of VET failure. Hence, VET failure is most prominent for cases with only two evidence variables. We shall focus on the two variable case.

Assuming that both evidence variables have the same Bayes factor, we obtain Bayes factors for the entire body of evidence of $1, 9, 400, 22500$ for the four cases above. Figure 3 clearly shows the drastic shrinking of the area of VET failure as the evidence becomes ever stronger. For the minimum for strong evidence, $Bf_{\mathcal{E}} = 400$, the area of VET failure is rather small; for two items of very strong evidence the area of VET failure is barely noticeable to the naked eye; if it all.

To conclude the discussion of the status of the VET in applications in science, we would like to echo (Claveau, 2013, p. 113) "*One could thus 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.*". Here, weak evidential sources are those with a small enough Bayes factors such that (12) fails.[16]

[15] It is also conjectured that the two conditions $P(c_1|h) = P(c_2|h)$ and $P(c_1|\bar{h}) = P(c_2|\bar{h})$ can be weakened into the single condition $P(c_1|h)/P(c_1|\bar{h}) = P(c_2|h)/P(c_2|\bar{h})$ and all observations for Scenario B continue to hold.

[16] This also dovetails nicely with, as yet, our unpublished results on VET failure in the original topology of Bovens & Hartmann; *blinded for review*.

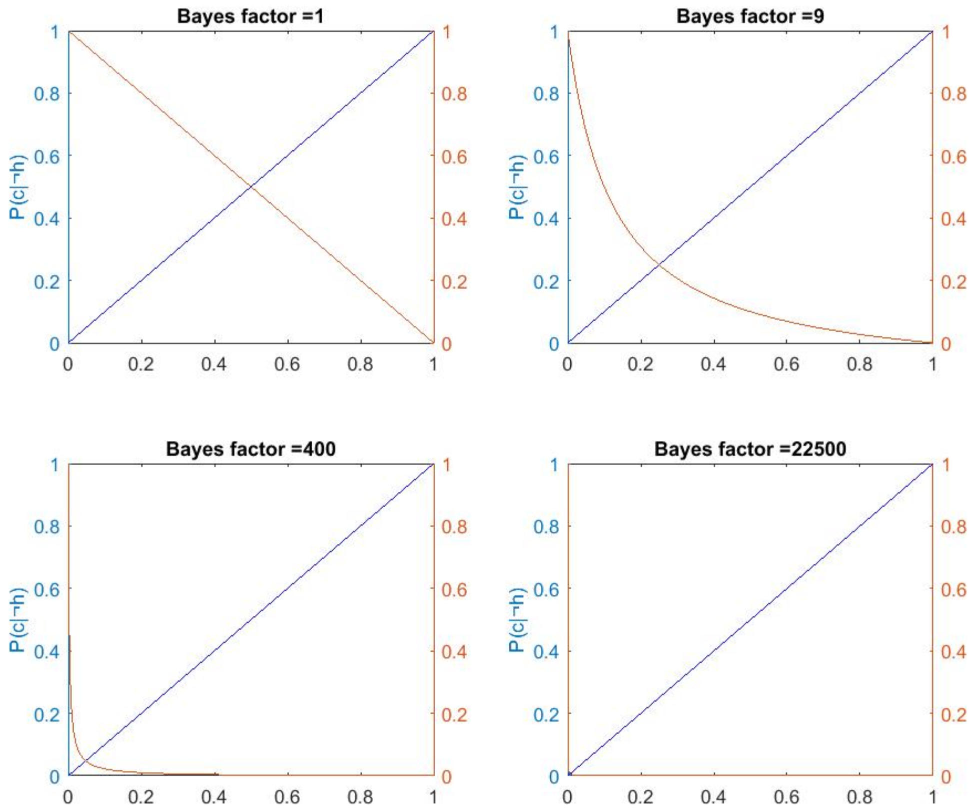

Figure 3: VET failure in the $P(c|h) - P(c|\bar{h})$-plane for different Bayes factors in the area under the blue bisecting line $P(c|h) = P(\bar{c}|h)$ which is also under brown curve (15). These lines intersect at the point $\langle P(c|h), P(c|\bar{h})\rangle = \langle (1 + \sqrt{Bf_{\mathcal{E}}})^{-1}, (1 + \sqrt{Bf_{\mathcal{E}}})^{-1}\rangle$; this can be verified by inserting this point into (15).

7.2.2 Adequacy for the Explication

The adequacy of condition (12) for the explication is also not obvious. While *we* feel that it may be reasonable to assume that (12) holds, others may disagree.

Firstly, *we* think that *evidence for* the hypothesis should be understood as observations that are predicted by the consequences of the hypothesis. Hence, evidence for the hypothesis is modelled by a reasonably large Bayes factor, a rather plausible minimum is $Bf_{\mathcal{E}} \geq 9$ – to our eyes.

We think there is something special about scientific hypotheses and their consequences. Elementary logic tells us that if 'H entails C', then it does not follow that '$\bar{H}$ entails $\bar{C}$'. So, C being a consequence of H does not entail that $\bar{C}$ is a consequence of $\bar{H}$. For example, if matter would not consist of atoms, then it would not be a consequence that the sky is not blue. Likewise in (Landes et al., 2017, Figure 4), if a drug does not cause a side-effect, then it is not a consequence that there is no probabilistic dependence between drug use and observed side-effect; the observed side-effect may well be due to confounding. Classical scientific hypotheses seem to tend to be specific and their testable consequences seem to tend to be (almost) deductive consequences.[17] Hence, $P(c|h)$ is so large that $P(c|h) + P(c|\bar{h}) \geq 1$ becomes true.

To *us*, (12) captures our intuitions regarding the VET. Hence, the analysis on offer here vindicates our intuitive understanding of the VET.

Other supporters of the VET may disagree. They might point out that there is a symmetry between a consequence and its negation. Hence, if $P(c|h)+P(c|\bar{h}) \geq 1$ were to hold, it would also need to hold for the negation of the consequence, $P(\bar{c}|h) + P(\bar{c}|\bar{h}) \geq 1$, due to a symmetry argument. This however imposes the unwarranted condition that $P(c|h) = P(\bar{c}|\bar{h})$. They may hence think that $P(c|h)+P(c|\bar{h}) \geq 1$ is by no means a formalisation of their intuitive understanding of the VET.

Furthermore, *others* may also think that a smaller Bayes factor, $Bf_{\mathcal{E}} = 1.5$ say, is compatible with their understanding of *evidence for* the hypothesis.

They may hence think that (12) does not do justice to their intuitions regarding the VET. Assuming that our derivation of (12) is technically correct, they are hence forced into at least one of two options, none of which appeals to us. Either they re-consider their intuitions regarding the VET and reject some of them and/or they find fault with our explication of the VET.

Without an arbitration between (possibly) conflicting intuitions, we leave it to the readers to decide for themselves whether the here offered analysis supports (some of) their intuitions concerning the VET.

8 Conclusions

Despite enjoying widespread intuitive support the VET has received bad press of late Bovens and Hartmann (2002, 2003); Claveau (2013); Claveau and Grenier (2018); Fitelson (1996). A model-independent or explication-independent vindication of the VET is – given these negative results – a pipe-dream. Hence, at best

[17] It must be admitted that during writing intuitions were also driven by legal reasoning; the hypothesis that a suspect committed the crime deductively entails that the suspect had a motive, an opportunity and the means.

one can hope for a wide class of pertinent models in which explications of the VET hold.

We here put forward a novel explication of the notion of varied evidence within a class of models of scientific inference by appealing to measures of entropy. Our models pronounce rightly on disconfirmation and discordant evidence, we argued. This lends pertinence to the class of discussed models.

In this paper, we chose to model the reliability of the evidence exogenously, i.e., we did not employ variables to model the (un-)reliability of the sources of evidence. Assessments of reliability are subsumed in the all important prior probability distribution. This choice was motivated by the desire to devise a model of scientific inference which only employs the most pertinent ingredients of our intuitive understanding of the VET. The reliability of the sources of evidence is not part of this understanding. VET failure in the models of Bovens and Hartmann (2003); Claveau (2013); Claveau and Grenier (2018); and *blinded for review*; is explained by the endogenous modelling of the notion of reliability. It is hence note-worthy that our analysis which does not explicitly model the reliability of sources of evidence also unearthes cases of VET failure.

Our explication of the VET holds in our models for our intuitive understanding of the VET. Despite VET failure in our models for certain parameter values, we have argued that the intuitive case for the Variety of Evidence Thesis emerges strengthened.

VET failure in our model is due to the interaction between the non-deterministic relationship between consequences and evidence on one side and the error probabilities $P(c|\bar{h}), P(\bar{c}|h)$ on the other side, which jointly drive hypothesis confirmation. The only terms in the key equation (11) are Bayes factors and error probabilities. Our main results are hence directly interpretable in mainstream Bayesian statistics.

Connecting the philosophical literature on the VET and its implications to research in other areas of research holds great promise. Connections to Bayesian statistics emerged here, connections to the organisation of science and research funding (e.g., Perović et al. (2016)) as well as the current debate on the replication crisis are immediate. Progress may also be made by exploring areas which are less obviously connected. For example, the ecology literature is much interested in developing measures of diversity, e.g., Tuomisto (2010).

Much remains to be learned about the Variety of Evidence Thesis.

Acknowledgements Blinded for review.

References

Borm, G. F., Lemmers, O., Fransen, J., and Donders, R. (2009). The evidence provided by a single trial is less reliable than its statistical analysis suggests. *Journal of Clinical Epidemiology*, 62(7):711–715.

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.

Carnap, R. (1962). *Logical Foundations of Probability*. University of Chicago Press, 2 edition.

Claveau, F. (2011). Evidential variety as a source of credibility for causal inference: beyond sharp designs and structural models. *Journal of Economic Methodology*, 18(3):233–253.

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

Claveau, F. and Grenier, O. (2018). The Variety-of-Evidence Thesis: A Bayesian Exploration of its Surprising Failures. *Synthese*. forthcoming.

Colombo, M and Gervais, R. and Sprenger, J. (2019) Introduction: objectivity in science, 194(12):4641–4642. *Synthese*

Crupi, V. and Nelson, J. and Meder, B. and Cevolani, G. and Tentori, K. (2019). Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search. forthcoming.

Csiszár, I. (2008). Axiomatic Characterizations of Information Measures. *Entropy*, 10(3):261–273.

Dawid, R., Hartmann, S., and Sprenger, J. (2015). The No Alternatives Argument. *British Journal for the Philosophy of Science*, 66(1):213–234.

Earman, J. (1992). *Bayes or Bust?* MIT Press.

Fitelson, B. (1996). Wayne, Horwich, and Evidential Diversity. *Philosophy of Science*, 63(4):652–660.

Fitelson, B. (2001). A bayesian account of independent evidence with applications. *Philosophy of Science*, 68(3):S123–S140.

Franklin, A. and Howson, C. (1984). Why do scientists prefer to vary their experiments? *Studies in History and Philosophy of Science Part A*, 15(1):51–62.

Glymour, C. (1980). *Theory and Evidence*. Princeton University Press.

Hartmann, S. and Bovens, L. (2001). The Variety-of-Evidence Thesis and the Reliability of Instruments: A Bayesian-Network Approach. PhilSci-Archive, last modified: 07 Oct 2010 15:10.

Hempel, C. (1966). *Philosophy of Natural Science*. Prentice Hall.

Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press.

Horwich, P. (1998). Wittgensteinian Bayesianism. In Curd, M. and Cover, J. A., editors, *Philosophy of Science: The Central Issues*, pages 607–624. W. W. Norton & Company.

Howson, C. and Urbach, P. (2006). *Scientific Reasoning*. Open Court, 3 edition.

Hüffmeier, J., Mazei, J., and Schultze, T. (2016). Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology*, 66:81–92.

Keynes, J. M. (1921). *A Treatise on Probability*. MacMillan.

Kuorikoski, J. and Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, 83(2):227–247.

Landes, J., Osimani, B., and Poellinger, R. (2017). Epistemology of Causal Inference in Pharmacology. *European Journal for Philosophy of Science*. 47 pages.

Landes, J. and Williamson, J. (2013). Objective Bayesianism and the maximum entropy principle. *Entropy*, 15(9):3528–3591.

Lloyd, E. A. (2015). Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science Part A*, 49:58–68.

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

Meehl, P. E. (1990). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It. *Psychological Inquiry*, 1(2):108–141.

Myrvold, W. C. (1996). Bayesianism and diverse evidence: A reply to Andrew Wayne. *Philosophy of Science*, 63(4):661.

Parkkinen, V.-P. (2016). Robustness and evidence of mechanisms in early experimental atherosclerosis research. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 60:44–55.

Pearl, J. (2009). *Causality Models, Reasoning and Inference*. Cambridge University Press, 2 edition. Zbl 1188.68291.

Perović, S., Radovanović, S., Sikimić, V., and Berber, A. (2016). Optimal research team composition: data envelopment analysis of Fermilab experiments. *Scientometrics*, 108(1):83–111.

Perrin, J. (1924). *Les Atomes*. Félix Alcan.

Poincaré, H. (1963). *Mathematics and Science: Last Essays*. Dover.

Schupbach, J. N. (2015). *Robustness, Diversity of Evidence, and Probabilistic Independence*, pages 305–316. Springer.

Schupbach, J. N. (2017). Robustness analysis as explanatory reasoning. *British Journal for the Philosophy of Science*. forthcoming.

Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423.

Steel, D. (1996). Bayesianism and the value of diverse evidence. *Philosophy of Science*, 63(4):666–674.

Stegenga, J. and Menon, T. (2017). Robustness and Independent Evidence. *Philosophy of Science*, 84(3):414–435.

Tuomisto, H. (2010). A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecologia*, 164(4):853–860.

Vezér, M. A. (2017). Variety-of-evidence reasoning about the distant past. *European Journal for Philosophy of Science*, 7(2):257–265.

Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science*, 62(1):111–121.

A Proofs of Main Results

We now give the longer proofs. The propositions to be proved are re-stated for ease of reference.

Proposition 1. *For bodies of evidence* $\mathcal{E}, \mathcal{E}'$ *with* $|\mathcal{E}| = |\mathcal{E}'|$ *it holds that*

$$V(\mathcal{E}) \geq V(\mathcal{E}'), \quad \textit{if and only if} \quad V_1(\mathcal{E}) \geq V_1(\mathcal{E}') \ .$$

Proof. Using that $\sum_{C \in \mathcal{C}} |C| = |\mathcal{E}|$, we find

$$\begin{aligned}
V(\mathcal{E}) &= -\sum_{C \in \mathcal{C}} \frac{|C|}{|\mathcal{E}|} \cdot \log\Big(\frac{|C|}{|\mathcal{E}|}\Big) \\
&= -\frac{1}{|\mathcal{E}|} \cdot \Big(\sum_{C \in \mathcal{C}} |C| \cdot \log\Big(\frac{|C|}{|\mathcal{E}|}\Big)\Big) \\
&= -\frac{1}{|\mathcal{E}|} \cdot \Big(\sum_{C \in \mathcal{C}} |C| \cdot (\log(|C|) - \log(|\mathcal{E}|))\Big) \\
&= -\frac{1}{|\mathcal{E}|} \cdot \Big(-V_1(\mathcal{E}) + \sum_{C \in \mathcal{C}} |C| \cdot (-\log(|\mathcal{E}|)\Big) \\
&= \frac{V_1(\mathcal{E})}{|\mathcal{E}|} + \log(|\mathcal{E}|) \ .
\end{aligned}$$

Hence, $V_1(\mathcal{E})$ is a positive-slope affine-linear transformation of $V(\mathcal{E})$. Positive-slope affine-linear transformation preserve ordinal comparisons. $\square$

Proposition 4. *For all finite and not-empty sets* Ω *and all probability functions* w *on* Ω, $\vec{x} := \langle w(\omega) : \omega \in \Omega, w(\omega) \in [0,1] \& \sum_\omega w(\omega) = 1 \rangle$, *it holds that* $H(\vec{x}) < H(\frac{|\Omega|}{|\Omega|+1}\vec{x}, \frac{1}{|\Omega|+1})$.

Proof. Using that $\sum_{i=1}^{|\Omega|} x_i = 1$ and that $H(\vec{x}) \leq -\log(\frac{1}{|\Omega|}) = \log(|\Omega|)$ we find

$$
\begin{aligned}
&H(\frac{|\Omega|}{|\Omega|+1}\vec{x}, \frac{1}{|\Omega|+1}) - H(\vec{x}) \\
&= -\frac{1}{|\Omega|+1}\Big(\log(\frac{1}{|\Omega|+1}) + \sum_{i=1}^{|\Omega|} |\Omega| x_i \cdot \log\Big(\frac{|\Omega|}{|\Omega|+1} x_i\Big)\Big) - H(\vec{x}) \\
&= -\log(\frac{1}{|\Omega|+1}) - \frac{|\Omega|}{|\Omega|+1} \cdot \Big(\sum_{i=1}^{|\Omega|} x_i \cdot [\log(x_i) + \log(|\Omega|)]\Big) - H(\vec{x}) \\
&= \log(|\Omega|+1) - \frac{|\Omega|}{|\Omega|+1}(-H(\vec{x}) + \log(|\Omega|)) - H(\vec{x}) \\
&= \log(|\Omega|+1) - \frac{|\Omega|}{|\Omega|+1} \cdot \log(|\Omega|) - \frac{1}{|\Omega|+1} H(\vec{x}) \\
&\geq \log(|\Omega|+1) - \frac{|\Omega|}{|\Omega|+1} \log(|\Omega|) - \frac{1}{|\Omega|+1} \log(|\Omega|) \\
&= \log(|\Omega|+1) - \log(|\Omega|) \\
&> 0 \ .
\end{aligned}
$$

□

Theorem 1. *In case of Condition A*

$$\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \mathrm{sign}(P(e_1|\bar{c}_1) - P(e_1|c_1)) = \mathrm{sign}(1 - Bf_1) \ .$$

Proof. To simply notation we let $\vec{f}$ denote the conjunction of all items evidence pertaining to the D_k and obtain:

$$
\begin{aligned}
P_{\mathcal{E}}(h) = \frac{P(h\vec{f})}{P(\vec{f})} &= \frac{P(\vec{f}|h) \cdot P(h)}{\sum_{s=0}^{1} P(\vec{f}|h^s) \cdot P(h^s)} \\
&= \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h})}{P(\vec{f}|h) \cdot P(h)}} \ .
\end{aligned}
$$

$$P_{\mathcal{E}'}(h) = \frac{P(he_1\vec{f})}{P(e_1\vec{f})} = \frac{P(\vec{f}|h) \cdot P(h) \cdot (\sum_{i=0}^{1} P(c_1^i|h) P(e_1|c_1^i))}{\sum_{s=0}^{1} P(\vec{f}|h^s) \cdot P(h^s) \cdot (\sum_{i=0}^{1} P(c_1^i|h^s) P(e_1|c_1^i))}$$

$$= \frac{1}{1 + \frac{P(\vec{f}|\bar{h})\cdot P(\bar{h})\cdot(\sum_{i=0}^{1} P(c_1^i|\bar{h})P(e_1|c_1^i))}{P(\vec{f}|h)\cdot P(h)\cdot(\sum_{i=0}^{1} P(c_1^i|h)P(e_1|c_1^i))}} \quad .$$

Hence,

$$\begin{aligned}
&\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) \\
&= \mathrm{sign}\Big(\frac{1}{1 + \frac{P(\vec{f}|\bar{h})\cdot P(\bar{h})}{P(\vec{f}|h)\cdot P(h)}} - \frac{1}{1 + \frac{P(\vec{f}|\bar{h})\cdot P(\bar{h})\cdot(\sum_{i=0}^{1} P(c_1^i|\bar{h})P(e_1|c_1^i))}{P(\vec{f}|h)\cdot P(h)\cdot(\sum_{i=0}^{1} P(c_1^i|h)P(e_1|c_1^i))}}\Big) \\
&= \mathrm{sign}\Big(\frac{\sum_{i=0}^{1} P(c_1^i|\bar{h})P(e_1|c_1^i)}{\sum_{i=0}^{1} P(c_1^i|h)P(e_1|c_1^i)} - 1\Big) \\
&= \mathrm{sign}\Big(\sum_{i=0}^{1} P(c_1^i|\bar{h})P(e_1|c_1^i) - \sum_{i=0}^{1} P(c_1^i|h)P(e_1|c_1^i)\Big) \\
&= \mathrm{sign}\Big(P(e_1|c_1)\cdot[P(c_1|\bar{h}) - P(c_1|h)] + P(e_1|\bar{c}_1)\cdot[P(\bar{c}_1|\bar{h}) - P(\bar{c}_1|h)]\Big) \\
&= \mathrm{sign}\Big(P(e_1|c_1)\cdot[P(c_1|\bar{h}) - P(c_1|h)] - P(e_1|\bar{c}_1)\cdot[P(c_1|\bar{h}) - P(c_1|h)]\Big) \\
&= \mathrm{sign}\Big([P(e_1|c_1) - P(e_1|\bar{c}_1)]\cdot[P(c_1|\bar{h}) - P(c_1|h)]\Big) \quad .
\end{aligned}$$

By (4) we have $P(c_1|h)-P(c_1|\bar{h}) < 0$ and hence the proof is completed by noting that

$$\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = \mathrm{sign}\Big(P(e_1|\bar{c}_1) - P(e_1|c_1)\Big) \quad .$$

□

Theorem 2. *Condition B and the Ceteris Paribus Conditions jointly entail that*

$$\begin{aligned}
&\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) \\
&= \mathrm{sign}\Big((1 - Bf_{C_1})(Bf_{C_1-1} - Bf_{C_2})\Big(Bf_{\mathcal{E}} - \frac{P(\bar{c}|h)\cdot P(\bar{c}|\bar{h})}{P(c|h)\cdot P(c|\bar{h})}\Big)\Big) \quad .
\end{aligned}$$

Proof. We employ the Ceteris Paribus conditions and simply write $P(c|h)$ and $P(c|\bar{h})$; dropping the subscript of c. To simplify notation we let for $j, l \in \{0, 1\}$

$$\alpha_{jl} := \prod_{n=1}^{|C_1|-1} P(e_n|c_1^j) \cdot \prod_{g=1}^{|C_2|} P(e_{|C_1|+g}|c_2^l) = \chi_{1j}\cdot\chi_{2l} \quad .$$

We begin by calculating the posterior probabilities in turn

$$
\begin{aligned}
P_{\mathcal{E}}(h) &= \frac{P(he_1 \dots e_{|C_1|+|C_2|}\vec{f})}{P(e_1 \dots e_{|C_1|+|C_2|}\vec{f})} \\
=& \frac{P(\vec{f}|h) \cdot P(h) \cdot \sum_{j,l=0}^{1} P(c_1^j c_2^l e_1 \dots e_{|C_1|+|C_2|}|h)}{\sum_{a=0}^{1} P(\vec{f}|h^a) \cdot P(h^a) \cdot \sum_{j,l=0}^{1} P(c_1^j c_2^l e_1 \dots e_{|C_1|+|C_2|}|h^a)} \\
=& \frac{P(\vec{f}|h) \cdot P(h) \cdot \left(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e_{|C_1|}|c_1^j)\chi_{1j} \cdot \chi_{2l}\right)}{\sum_{a=0}^{1} P(\vec{f}|h^a) \cdot P(h^a) \cdot \left(\sum_{j,l=0}^{1} P(c^j|h^a)P(c^l|h^a)P(e_{|C_1|}|c_1^j)\chi_{1j} \cdot \chi_{2l}\right)} \\
=& \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot \left(\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}\right)}{P(\vec{f}|h) \cdot P(h) \cdot \left(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}\right)}} \quad ,
\end{aligned}
\tag{16}
$$

recalling that the χ were defined in (13). Similarly, we find for $P_{\mathcal{E}'}(h)$ that

$$
\begin{aligned}
P_{\mathcal{E}'}(h) &= \frac{P'(he_1 \dots e_{|C_1|-1}e'_{|C_1|}e_{|C_1|+1} \dots e_{|C_1|+|C_2|}\vec{f})}{P(e_1 \dots e_{|C_1|-1}e'_{|C_1|}e_{|C_1|+1} \dots e_{|C_1|+|C_2|}\vec{f})} \\
=& \frac{P(\vec{f}|h) \cdot P(h) \cdot \sum_{j,l=0}^{1} P'(c_1^j c_2^l e_1 \dots e_{|C_1|-1}e'_{|C_1|}e_{|C_1|+1} \dots e_{|C_1|+|C_2|}|h)}{\sum_{a=0}^{1} P(\vec{f}|h^a)P(h^a) \sum_{j,l=0}^{1} P'(c_1^j c_2^l e_1 \dots e_{|C_1|-1}e'_{|C_1|}e_{|C_1|+1} \dots e_{|C_1|+|C_2|}|h^a)} \\
=& \frac{P(\vec{f}|h) \cdot P(h) \cdot \left(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}\right)}{\sum_{a=0}^{1} P(\vec{f}|h^a) \cdot P(h^a) \cdot \left(\sum_{j,l=0}^{1} P(c^j|h^a)P(c^l|h^a)P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}\right)} \\
=& \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot \left(\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}\right)}{P(\vec{f}|h) \cdot P(h) \cdot \left(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}\right)}} \quad .
\end{aligned}
\tag{17}
$$

Note that the only difference between these two posteriors is that the first posterior contains the term $P(e'_{|C_1|}|c_1^j)$ while the second posterior contains the term $P(e_{|C_1|}|c_2^l)$.

The leading factors play no role, the sign of $P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)$ is thus equal to the sign of

$$\frac{\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}}{\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}} - \frac{\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}}{\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}} \quad .$$

The sign of this expression is equal to the sign of

$$\begin{aligned}
&\Big(\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}\Big) \\
&\quad \cdot \Big(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}\Big) \\
&-\Big(\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}\Big) \\
&\quad \cdot \Big(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l}\Big) \quad .
\end{aligned}$$

To simplify notation we let for $j, l \in \{0, 1\}$

$$\alpha_{jl} := \prod_{n=1}^{|C_1|-1} P(e_n|c_1^j) \cdot \prod_{g=1}^{|C_2|} P(e_{|C_1|+g}|c_2^l) = \chi_{1j} \cdot \chi_{2l} \quad .$$

We obtain the more manageable

$$\begin{aligned}
&\Big(\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e'_{|C_1|}|c_2^l) \cdot \alpha_{jl}\Big) \cdot \Big(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e_{|C_1|}|c_1^j) \cdot \alpha_{jl}\Big) \\
-&\Big(\sum_{j,l=0}^{1} P(c^j|\bar{h})P(c^l|\bar{h})P(e_{|C_1|}|c_1^j) \cdot \alpha_{jl}\Big) \cdot \Big(\sum_{j,l=0}^{1} P(c^j|h)P(c^l|h)P(e'_{|C_1|}|c_2^l) \cdot \alpha_{jl}\Big).
\end{aligned}$$

Spelling this out we obtain

$$\Big(P(\bar{c}|\bar{h})^2 P(e'_{|C_1|}|\bar{c}_2)\alpha_{00} + P(c|\bar{h})^2 P(e'_{|C_1|}|c_2)\alpha_{11}$$

$$
\begin{aligned}
&\quad + P(\bar{c}|\bar{h})P(c|\bar{h})[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10} + P(e'_{|C_1|}|c_2)\alpha_{01}]\Big) \\
&\cdot\Big(P(\bar{c}|h)^2 P(e_{|C_1|}|\bar{c}_1)\alpha_{00} + P(c|h)^2 P(e_{|C_1|}|c_1)\alpha_{11} \\
&\quad + P(\bar{c}|h)P(c|h)[P(e_{|C_1|}|c_1)\alpha_{10} + P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\Big) \\
- \quad &\Big(P(\bar{c}|\bar{h})^2 P(e_{|C_1|}|\bar{c}_1)\alpha_{00} + P(c|\bar{h})^2 P(e_{|C_1|}|c_1)\alpha_{11} \\
&\quad + P(\bar{c}|\bar{h})P(c|\bar{h})[P(e_{|C_1|}|c_1)\alpha_{10} + P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\Big) \\
&\cdot\Big(P(\bar{c}|h)^2 P(e'_{|C_1|}|\bar{c}_2)\alpha_{00} + P(c|h)^2 P(e'_{|C_1|}|c_2)\alpha_{11} \\
&\quad + P(\bar{c}|h)P(c|h)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10} + P(e'_{|C_1|}|c_2)\alpha_{01}]\Big) \ .
\end{aligned}
$$

Fortunately, all those terms which do not contain α_{00} nor α_{11} (these are precisely those terms with $P(\bar{c}|\bar{h})P(c|\bar{h})P(\bar{c}|\bar{h})P(c|\bar{h})$) cancel out. Furthermore, all terms which contain α_{00}^2 and all terms containing α_{11}^2 cancel out.

For the terms containing α_{00} and α_{11} we find

$$
\begin{aligned}
\alpha_{00}\alpha_{11}\Big(&P(\bar{c}|\bar{h})^2 P(c|h)^2 [P(e'_{|C_1|}|\bar{c}_2)P(e_{|C_1|}|c_1) - P(e_{|C_1|}|\bar{c}_1)P(e'_{|C_1|}|c_2)] \\
&+ P(c|\bar{h})^2 P(\bar{c}|h)^2 [P(e'_{|C_1|}|c_2)P(e_{|C_1|}|\bar{c}_1) - P(e_{|C_1|}|c_1)P(e'_{|C_1|}|\bar{c}_2)]\Big) \ .
\end{aligned}
$$

Under the standing assumption that

$$
Bf_{C_1} = \frac{P(e_{|C_1|}|c_1)}{P(e_{|C_1|}|\bar{c}_1)} = \frac{P(e'_{|C_1|}|c_2)}{P(e'_{|C_1|}|\bar{c}_2)}
$$

we note that also these terms cancel out.

What remains is the much more manageable

$$
\begin{aligned}
&\Big(P(\bar{c}|\bar{h})^2 P(e'_{|C_1|}|\bar{c}_2)\alpha_{00} + P(c|\bar{h})^2 P(e'_{|C_1|}|c_2)\alpha_{11}\Big) \\
&\cdot \Big(P(\bar{c}|h)P(c|h)[P(e_{|C_1|}|c_1)\alpha_{10} + P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\Big) \\
&- \Big(P(\bar{c}|\bar{h})^2 P(e_{|C_1|}|\bar{c}_1)\alpha_{00} + P(c|\bar{h})^2 P(e_{|C_1|}|c_1)\alpha_{11}\Big) \\
&\cdot \Big(P(\bar{c}|h)P(c|h)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10} + P(e'_{|C_1|}|c_2)\alpha_{01}]\Big) \\
+&\Big(P(\bar{c}|h)^2 P(e_{|C_1|}|\bar{c}_1)\alpha_{00} + P(c|h)^2 P(e_{|C_1|}|c_1)\alpha_{11}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&\cdot\Big(P(\bar{c}|\bar{h})P(c|\bar{h})[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10}+P(e'_{|C_1|}|c_2)\alpha_{01}]\Big)\\
&-\Big(P(\bar{c}|\bar{h})^2P(e'_{|C_1|}|\bar{c}_2)\alpha_{00}+P(c|\bar{h})^2P(e'_{|C_1|}|c_2)\alpha_{11}\Big)\\
&\cdot\Big(P(\bar{c}|\bar{h})P(c|\bar{h})[P(e_{|C_1|}|c_1)\alpha_{10}+P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\Big)\\
=&P(\bar{c}|h)P(c|h)\Big([P(\bar{c}|\bar{h})^2\alpha_{00}+P(c|\bar{h})^2Bf_{C_1}\alpha_{11}]P(e_{|C_1|}|\bar{c}_2)[P(e_{|C_1|}|c_1)\alpha_{10}+P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\\
&-[P(\bar{c}|\bar{h})^2\alpha_{00}+P(c|\bar{h})^2Bf_{C_1}\alpha_{11}]P(e_{|C_1|}|\bar{c}_1)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10}+P(e'_{|C_1|}|c_2)\alpha_{01}]\Big)\\
+&P(\bar{c}|\bar{h})P(c|\bar{h})\Big([P(\bar{c}|h)^2\alpha_{00}+P(c|h)^2Bf_{C_1}\alpha_{11}]P(e'_{|C_1|}|\bar{c}_1)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10}+P(e'_{|C_1|}|c_2)\alpha_{01}]\\
&-[P(\bar{c}|\bar{h})^2\alpha_{00}+P(c|\bar{h})^2Bf_{C_1}\alpha_{11}]P(e'_{|C_1|}|\bar{c}_2)[P(e_{|C_1|}|c_1)\alpha_{10}+P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\Big)\\
=&P(\bar{c}|h)P(c|h)[P(\bar{c}|\bar{h})^2\alpha_{00}+P(c|\bar{h})^2Bf_{C_1}\alpha_{11}]\\
&\Big(P(e'_{|C_1|}|\bar{c}_2)[P(e_{|C_1|}|c_1)\alpha_{10}+P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]-P(e_{|C_1|}|\bar{c}_1)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10}+P(e'_{|C_1|}|c_2)\alpha_{01}]\Big)\\
+&P(\bar{c}|\bar{h})P(c|\bar{h})[P(\bar{c}|h)^2\alpha_{00}+P(c|h)^2Bf_{C_1}\alpha_{11}]\\
&\Big(P(e_{|C_1|}|\bar{c}_1)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10}+P(e'_{|C_1|}|c_2)\alpha_{01}]-P(e'_{|C_1|}|\bar{c}_2)[P(e_{|C_1|}|c_1)\alpha_{10}+P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]\Big)\\
=&\Big(P(\bar{c}|h)P(c|h)[P(\bar{c}|\bar{h})^2\alpha_{00}+P(c|\bar{h})^2Bf_{C_1}\alpha_{11}]-P(\bar{c}|\bar{h})P(c|\bar{h})[P(\bar{c}|h)^2\alpha_{00}+P(c|h)^2Bf_{C_1}\alpha_{11}]\Big)\\
&\cdot\Big(P(e'_{|C_1|}|\bar{c}_2)[P(e_{|C_1|}|c_1)\alpha_{10}+P(e_{|C_1|}|\bar{c}_1)\alpha_{01}]-P(e_{|C_1|}|\bar{c}_1)[P(e'_{|C_1|}|\bar{c}_2)\alpha_{10}+P(e'_{|C_1|}|c_2)\alpha_{01}]\Big).
\end{aligned}
$$

We now investigate both factors in turn. We begin with the second and find that it is equal to

$$
\begin{aligned}
&P(e'_{|C_1|}|\bar{c}_2)P(e_{|C_1|}|\bar{c}_1)([Bf_{C_1}\alpha_{10}+\alpha_{01}]-[\alpha_{10}+Bf_{C_1}\alpha_{01}])\\
=&P(e'_{|C_1|}|\bar{c}_2)P(e_{|C_1|}|\bar{c}_1)(Bf_{C_1}-1)(\alpha_{10}-\alpha_{01}) \quad .
\end{aligned}
$$

We note that

$$
\begin{aligned}
\alpha_{10}-\alpha_{01}=&\prod_{n=1}^{|C_1|-1}P(e_n|c_1)\cdot\prod_{g=1}^{|C_2|}P(e_{|C_1|+g}|\bar{c}_2)-\prod_{n=1}^{|C_1|-1}P(e_n|\bar{c}_1)\cdot\prod_{g=1}^{|C_2|}P(e_{|C_1|+g}|c_2)\\
=&\Big(\prod_{n=1}^{|C_1|-1}P(e_n|\bar{c}_1)\prod_{g=1}^{|C_2|}P(e_{|C_1|+g}|\bar{c}_2)\Big)\cdot\Big(\prod_{n=1}^{|C_1|-1}\frac{P(e_n|c_1)}{P(e_n|\bar{c}_1)}-\prod_{g=1}^{|C_2|}\frac{P(e_{|C_1|+g}|c_2)}{P(e_{|C_1|+g}|\bar{c}_2)}\Big)\\
=&\Big(\prod_{n=1}^{|C_1|-1}P(e_n|\bar{c}_1)\prod_{g=1}^{|C_2|}P(e_{|C_1|+g}|\bar{c}_2)\Big)\cdot\Big(Bf_{C_1-1}-Bf_{C_2}\Big) \quad .
\end{aligned}
$$

After some algebra we find for the negative of the first factor that

$$\begin{aligned}
&[P(\bar{c}|h)^2\alpha_{00} + P(c|h)^2 Bf_{C_1}\alpha_{11}] \cdot P(\bar{c}|\bar{h})P(c|\bar{h}) \\
&- [P(\bar{c}|\bar{h})^2\alpha_{00} + P(c|\bar{h})^2 Bf_{C_1}\alpha_{11}] \cdot P(\bar{c}|h)P(c|h) \\
=&\alpha_{00} \cdot [P(\bar{c}|h)^2 \cdot P(\bar{c}|\bar{h})P(c|\bar{h}) - P(\bar{c}|\bar{h})^2 \cdot P(\bar{c}|h)P(c|h)] \\
&+ Bf_{C_1}\alpha_{11} \cdot [P(c|h)^2 \cdot P(\bar{c}|\bar{h})P(c|\bar{h}) - P(c|\bar{h})^2 \cdot P(\bar{c}|h)P(c|h)] \\
=&\alpha_{00} \cdot [P(\bar{c}|h)P(\bar{c}|\bar{h})] \cdot [P(\bar{c}|h)P(c|\bar{h}) - P(\bar{c}|\bar{h})P(c|h)] \\
&+ Bf_{C_1}\alpha_{11} \cdot [P(c|h)P(c|\bar{h})] \cdot [P(\bar{c}|\bar{h})P(c|h) - P(\bar{c}|h)P(c|\bar{h})] \\
=&[P(\bar{c}|\bar{h})P(c|h) - P(\bar{c}|h)P(c|\bar{h})] \cdot \\
&[Bf_{C_1} \cdot \alpha_{11} \cdot P(c|h) \cdot P(c|\bar{h}) - \alpha_{00} \cdot P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})] \ . \qquad (18)
\end{aligned}$$

Since $P(c|h) > P(c|\bar{h})$, we have $P(\bar{c}|\bar{h}) = 1 - P(c|\bar{h}) > 1 - P(c|h) = P(\bar{c}|h)$. Thus, $P(\bar{c}|\bar{h})P(c|h) - P(\bar{c}|h)P(c|\bar{h}) > 0$.

Using that $Bf_{C_1}\alpha_{11}/\alpha_{00} = Bf_{\mathcal{E}}$, i.e., (10) holds, we find that

$$\begin{aligned}
&\operatorname{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) \\
=&\operatorname{sign}\Big((Bf_{C_1} - 1)(Bf_{C_1-1} - Bf_{C_2}) \\
&\cdot [-Bf_{C_1} \cdot \alpha_{11} \cdot P(c|h) \cdot P(c|\bar{h}) + \alpha_{00} \cdot P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})]\Big) \\
=&\operatorname{sign}\Big((1 - Bf_{C_1})(Bf_{C_1-1} - Bf_{C_2})(Bf_{\mathcal{E}} - \frac{P(\bar{c}|h) \cdot P(\bar{c}|\bar{h})}{P(c|h) \cdot P(c|\bar{h})}\Big) \ .
\end{aligned}$$

□

Corollary 3. *If $E_1, \ldots, E_{|C_1|}$ are the children of C_1, then for all possible measurements $E_1 = e_1, \ldots, E_{|C_1|} = e_{|C_1|}$*

$$P(h|e_1 \ldots e_{|C_1|}\vec{f}) < P(h|c_1\vec{f}) \ .$$

Proof. The proof is a relatively simple exercise in Bayesian network calculations

$$P(h|e_1 \ldots e_{|C_1|}\vec{f}) = \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot (\sum_{j=0}^{1} P(c^j|\bar{h}) \prod_{n=1}^{|C_1|} P(e_n|c^j))}{P(\vec{f}|h) \cdot P(h) \cdot (\sum_{j=0}^{1} P(c^j|h) \prod_{n=1}^{|C_1|} P(e_n|c^j))}} \ .$$

$$P(h|c_1\vec{f}) = \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot (P(c|\bar{h}))}{P(\vec{f}|h) \cdot P(h) \cdot P(c|h)}} \ .$$

Hence,

$$
\begin{aligned}
&\operatorname{sign}(P(h|e_1\dots e_{|C_1|}\vec{f})-P(h|c_1\vec{f}))\\
&=\operatorname{sign}\Big(\frac{P(c|\bar{h})}{P(c|h)}-\frac{\sum_{j=0}^{1}P(c^j|\bar{h})\prod_{n=1}^{|C_1|}P(e_n|c^j)}{\sum_{j=0}^{1}P(c^j|h)\prod_{n=1}^{|C_1|}P(e_n|c^j)}\Big)\\
&=\operatorname{sign}\Big(P(c|\bar{h})\sum_{j=0}^{1}P(c^j|h)\prod_{n=1}^{|C_1|}P(e_n|c^j)-P(c|h)\cdot\sum_{j=0}^{1}P(c^j|\bar{h})\prod_{n=1}^{|C_1|}P(e_n|c^j)\Big)\\
&=\operatorname{sign}\Big(P(c|\bar{h})[P(c|h)\prod_{n=1}^{|C_1|}P(e_n|c)+P(\bar{c}|h)\prod_{n=1}^{|C_1|}P(e_n|\bar{c})]\\
&\quad-P(c|h)\cdot[P(c|\bar{h})\prod_{n=1}^{|C_1|}P(e_n|c)+P(\bar{c}|\bar{h})\prod_{n=1}^{|C_1|}P(e_n|\bar{c})]\Big)\\
&=\operatorname{sign}\Big(P(c|\bar{h})P(\bar{c}|h)\prod_{n=1}^{|C_1|}P(e_n|\bar{c})-P(c|h)\cdot P(\bar{c}|\bar{h})\prod_{n=1}^{|C_1|}P(e_n|\bar{c})\Big)\\
&=\operatorname{sign}\Big(P(c|\bar{h})P(\bar{c}|h)-P(c|h)\cdot P(\bar{c}|\bar{h})\Big)\ .
\end{aligned}
$$

By (4) we have that $P(c|\bar{h})<P(c|h)$ and that $P(\bar{c}|h)<P(\bar{c}|\bar{h})$. Hence, the bracket is negative and it follows that

$$P(h|e_1\dots e_{|C_1|}\vec{f})<P(h|c_1\vec{f})\ .$$

□

Proposition 2. *If Condition B, the last two ceteris paribus condition and if* $P(c_1|h)=1=P(c_2|h)$ *hold, then*

$$
\begin{aligned}
\operatorname{sign}(P_{\mathcal{E}}(h)-P_{\mathcal{E}'}(h))=\operatorname{sign}\Big(&\frac{\sum_{j,l=0}^{1}P(c_1^j|\bar{h})P(c_2^l|\bar{h})P(e'_{|C_1|}|c_2^l)\cdot\chi_{1j}\cdot\chi_{2l}}{P(e'_{|C_1|}|c_2)}\\
&-\frac{\sum_{j,l=0}^{1}P(c_1^j|\bar{h})P(c_2^l|\bar{h})P(e_{|C_1|}|c_1^j)\cdot\chi_{1j}\cdot\chi_{2l}}{P(e_{|C_1|}|c_1)}\Big)\ .
\end{aligned}
$$

Proof. The proof is a simple exercise in Bayesian network calculations. First, we use (16) and (17) to obtain

$$P_{\mathcal{E}}(h)=\frac{1}{1+\frac{P(\vec{f}|\bar{h})\cdot P(\bar{h})\cdot\left(\sum_{j,l=0}^{1}P(c_1^j|\bar{h})P(c_2^l|\bar{h})P(e_{|C_1|}|c_1^j)\cdot\chi_{1j}\cdot\chi_{2l}\right)}{P(\vec{f}|h)\cdot P(h)\cdot\left(\sum_{j,l=0}^{1}P(c_1^j|h)P(c_2^l|h)P(e_{|C_1|}|c_1^j)\cdot\chi_{1j}\cdot\chi_{2l}\right)}}\ .$$

$$P_{\mathcal{E}'}(h) = \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot \left(\sum_{j,l=0}^{1} P(c_1^j|\bar{h}) P(c_2^l|\bar{h}) P(e'_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l} \right)}{P(\vec{f}|h) \cdot P(h) \cdot \left(\sum_{j,l=0}^{1} P(c_1^j|h) P(c_2^l|h) P(e'_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l} \right)}} \ .$$

Using that $P(c_1|h) = 1 = P(c_2|h)$ and hence $P(\bar{c}_1|h) = 0 = P(\bar{c}_2|h)$ the expressions simplify to

$$P_{\mathcal{E}}(h) = \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot \left(\sum_{j,l=0}^{1} P(c_1^j|\bar{h}) P(c_2^l|\bar{h}) P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l} \right)}{P(\vec{f}|h) \cdot P(h) \cdot \left(P(e_{|C_1|}|c_1) \cdot \chi_{11} \cdot \chi_{21} \right)}}$$

and

$$P_{\mathcal{E}'}(h) = \frac{1}{1 + \frac{P(\vec{f}|\bar{h}) \cdot P(\bar{h}) \cdot \left(\sum_{j,l=0}^{1} P(c_1^j|\bar{h}) P(c_2^l|\bar{h}) P(e'_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l} \right)}{P(\vec{f}|h) \cdot P(h) \cdot \left(P(e'_{|C_1|}|c_1) \cdot \chi_{11} \cdot \chi_{21} \right)}} \ .$$

The claimed result follows by re-substituting the definitions of χ_{1j} and χ_{2l}. □

Proposition 3. *If Condition B, the last two ceteris paribus condition, $P(c_1|h) = 1 = P(c_2|h)$ and if $P(e_{|C_1|}|c_1) = P(e'_{|C_1|}|c_2) > P(e'_{|C_1|}|\bar{c}_2)$ hold, then*

$$\mathrm{sign}(P_{\mathcal{E}}(h) - P_{\mathcal{E}'}(h)) = -\,\mathrm{sign}\Big(\frac{\alpha_{0,0} \cdot \beta_{0,0} + \alpha_{0,1} \cdot \beta_{0,1}}{\alpha_{1,0} \cdot \beta_{1,0}} + \prod_{n=1}^{|C_1|-1} \frac{P(e_n|c_1)}{P(e_n|\bar{c}_1)} \Big) \ ,$$

where the α and β are parameters independent of $|C_1|$ which are defined in (19).

Proof. Under the assumption that $P(e_{|C_1|}|c_1) = P(e'_{|C_1|}|c_2)$ the denominators are equal. Furthermore, the terms with $j = l = 1$ cancel out. We hence find

$$\begin{aligned} \mathrm{sign}\Big(& P(h|e_1 \dots e_L e_{|C_1|} e_{|C_1|+1} \dots e_{|C_1|+|C_2|} \vec{f}) \\ & - P(h|e_1 \dots e_{|C_1|-1} e'_{|C_1|} e_{|C_1|+1} \dots e_{|C_1|+|C_2|} \vec{f}) \Big) \\ = \mathrm{sign}\Big(& \sum_{\substack{j,l=0 \\ (j,l) \neq (1,1)}}^{1} P(c_1^j|\bar{h}) P(c_2^l|\bar{h}) P(e'_{|C_1|}|c_2^l) \cdot \chi_{1j} \cdot \chi_{2l} \end{aligned}$$

$$- \sum_{\substack{j,l=0 \\ (j,l)\neq(1,1)}}^{1} P(c_1^j|\bar{h})P(c_2^l|\bar{h})P(e_{|C_1|}|c_1^j) \cdot \chi_{1j} \cdot \chi_{2l}\Big) \ .$$

With (subscripts represent the values for j and l)

$$\begin{aligned}
\alpha_{0,0} &:= P(\bar{c}_1|\bar{h})P(\bar{c}_2|\bar{h}) \cdot \prod_{g=1}^{|C_2|} P(e_{|C_1|+g}|\bar{c}_2) \\
\beta_{0,0} &:= P(e'_{|C_1|}|\bar{c}_2) - P(e_{|C_1|}|\bar{c}_1) \\
\alpha_{1,0} &:= P(c_1|\bar{h})P(\bar{c}_2|\bar{h}) \cdot \prod_{g=1}^{|C_2|} P(e_{|C_1|+g}|\bar{c}_2) \\
\beta_{1,0} &:= P(e'_{|C_1|}|\bar{c}_2) - P(e_{|C_1|}|c_1) \\
\alpha_{0,1} &:= P(\bar{c}_1|\bar{h})P(\bar{c}_2|\bar{h}) \cdot \prod_{g=1}^{|C_2|} P(e_{|C_1|+g}|c_2) \\
\beta_{0,1} &:= P(e'_{|C_1|}|c_2) - P(e_{|C_1|}|\bar{c}_1)
\end{aligned} \tag{19}$$

this becomes equal to

$$\begin{aligned}
&\operatorname{sign}\Big(\alpha_{0,0} \cdot \beta_{0,0} \cdot \prod_{n=1}^{|C_1|-1} P(e_n|\bar{c}_1) + \alpha_{0,1} \cdot \beta_{0,1} \cdot \prod_{n=1}^{|C_1|-1} P(e_n|\bar{c}_1) \\
&\quad + \alpha_{1,0} \cdot \beta_{1,0} \cdot \prod_{n=1}^{|C_1|-1} P(e_n|c_1)\Big) \\
&= \operatorname{sign}\Big([\alpha_{0,0} \cdot \beta_{0,0} + \alpha_{0,1} \cdot \beta_{0,1}] \cdot \prod_{n=1}^{|C_1|-1} P(e_n|\bar{c}_1) + \alpha_{1,0} \cdot \beta_{1,0} \cdot \prod_{n=1}^{|C_1|-1} P(e_n|c_1)\Big) \ .
\end{aligned}$$

Note that the α and the β parameters do not change value with varying $|C_1|$ and are hence treated as constants.

Using that $\alpha_{1,0} \cdot \beta_{1,0} < 0$ and $\prod_{n=1}^{|C_1|-1} P(e_n|\bar{c}_1) > 0$ we find that

$$\begin{aligned}
&\operatorname{sign}(P(h|e_1 \dots e_{|C_1|-1}e_{|C_1|}e_{|C_1|+1} \dots e_{|C_1|+|C_2|}\vec{f}) \\
&\quad - P(h|e_1 \dots e_{|C_1|-1}e'_{|C_1|}e_{|C_1|+1} \dots e_{|C_1|+|C_2|}\vec{f}) \\
&= -\operatorname{sign}\Big(\frac{\alpha_{0,0} \cdot \beta_{0,0} + \alpha_{0,1} \cdot \beta_{0,1}}{\alpha_{1,0} \cdot \beta_{1,0}} + \frac{\prod_{n=1}^{|C_1|-1} P(e_n|c_1)}{\prod_{n=1}^{|C_1|-1} P(e_n|\bar{c}_1)}\Big) \ .
\end{aligned}$$

□

B High Arity Variables

We now address the claim that the so-far established technical results, also hold for models which employ higher arity hypothesis and/or consequence variables.

Denote by $h^2, h^3, \ldots$ the values of H different from h and by $c^2, c^3, \ldots$ the values of consequence variable C different from c. h^0 is the (possibly infinite) disjunction of the h^i with $i \geq 2$. c^0 is the (possibly infinite) disjunction of the c^k with $k \geq 2$. To simplify notation we let $P(c_j^0|h^i) := 1 - P(c_j|h^i)$.

The ceteris paribus conditions are that for every consequence variables C_j and every evidence variable E pertaining to the consequence variable C_j it holds that

$$P(c_j^0|h^i) = P(c_j^0|h^2) = P(\bar{c}_j|\bar{h}) \tag{20}$$

$$P(e|c_j^k) = P(e|c_j^2) = P(e|\bar{c}_j) \ . \tag{21}$$

This formalises the thought that all values of the hypothesis variable H different from h are equal. Furthermore, the conditional probability of the evidence does not depend on the particular value c^k. We can hence define our Bayes factors in the usual way unequivocally as $P(e|c)/P(e|\bar{c})$.

Proposition 5. *If* (20) *and* (21) *hold, then Corollary 1 and Corollary 2 hold for higher arities, too.*

Proof. Using the ceteris paribus condition for the second and third equality, we find

$$\begin{aligned}
\sum_{i\geq 2}\sum_{k\geq 2} P(h^i)\cdot P(c_j^k|h^i)\cdot P(e|c_j^k) &= \sum_{i\geq 2} P(h^i)\cdot \sum_{k\geq 2} P(c_j^k|h^i)\cdot P(e|c_j^k) \\
&= P(e|c_j^2)\cdot \sum_{i\geq 2} P(h^i)\cdot \sum_{k\geq 2} P(c_j^k|h^i) \\
&= P(e|\bar{c}_j)\cdot \sum_{i\geq 2} P(h^i)\cdot P(c_j^0|h^i) \\
&= P(e|\bar{c}_j)\cdot P(\bar{c}_j|\bar{h})\cdot P(\bar{h}) \ .
\end{aligned}$$

To complete the proof it suffices make the following observations:
1) Whenever the term '$P(e|\bar{c}_j)\cdot P(\bar{c}_j|\bar{h})\cdot P(\bar{h})$' appears in a proof for the basic model it is replaced by the term '$\sum_{i=2}\sum_{k=2} P(h^i)\cdot P(c_j^k|h^i)\cdot P(e|c_j^k)$'. Since both terms are equal, no new difficulties arise.
2) The terms of the form '$P(\bar{c}_j|\bar{h})$' for $i \geq 2$ are replaced by terms of the form $\sum_{k\geq 2} P(c_j^k|h^i)$'. The latter is equal to $P(c_j^0|h^i)$. By the first ceteris paribus assumption for greater arities, these terms are equal. □