On the value of varied evidence for imprecise probabilities

Jürgen Landes [1] **Sébastien Destercke** [2]

[1]Munich Center for Mathematical Philosophy and Open Science Center, LMU Munich, Germany
[2]UMR CNRS 7253 Heudiasyc, Université de Technologie de Compiègne, France

Abstract

It has long been considered a truism that we can learn more from a variety of sources than from highly correlated sources. This truism is captured by the Variety of Evidence Thesis. To the surprise of many, this thesis turned out to fail in a number of Bayesian settings. In other words, replication can trump variation. Translating the thesis into IP we obtain two distinct, a priori plausible formulations in terms of 'increased confirmation' and 'uncertainty reduction', respectively. We investigate both formulations, which both fail for different parameters and different reasons, that cannot be predicted prior to formal analysis. The emergence of two distinct formulations distinguishing confirmation increase from uncertainty reduction, which are conflated in the Bayesian picture, highlights fundamental differences between IP and Bayesian reasoning.

Keywords. Bayesian network, variety of evidence, credal network, testimony, confirmation, value of information

1. Introduction

That statement that 'good information is informative' is rather obvious. The more interesting question investigated here is the question, which information is informative? A number of answers suggest themselves; we want our data to be big, timely, high-resolution (image, video, audio), unbiased and, of course, truth-conducive (veracity). Methodologists of science have long insisted that varied evidence provides good information [4, 10, 16, 21, 22, 39, 50, 76].

1.1. Some methodology and philosophy of science. Philosophers of science have cast this idea as follows: varied evidence allows us to draw firmer conclusions than less varied evidence, ceteris paribus [8, 22, 26, 35, 38]. More formally: More varied evidence for a hypothesis confirms it more strongly than less varied evidence, ceteris paribus. This formulation is known as the *Variety of Evidence Thesis* (VET), cf. Figure 1. This thesis was thought to be intuitive and one of the very few things philosophers agreed upon. Worries that the thesis could fail [23, 74] were widely ignored. It was not until after the turn of the century that the work of Bovens and Hartmann sank the VET, at least when formalised in a Bayesian way.

1.2. Bovens & Hartmann. If we first hear interesting news from Charles and then his identical twin brother Charley with whom he shares his information, reports the same news, then Charley's report does not impress us (much). However, if we first hear interesting news from Ann and then Bob, who is a stranger to Ann, reports the same news, then Bob's report impresses us (more). Intuitively, the fact that Ann and Bob are independent makes the news more credible than in the case of the twin brothers. This intuitive story was formalised by Bovens & Hartmann [5, 6].

Their crucial conceptual idea is to understand *evidential variety* as *variety of sources* providing the evidence. The more independent sources are the greater the variety. Sources that are (probabilistically) independent have maximal variety, while reports originating from the same source have minimal variety. The set of two sources {Charles, Charley} is less varied than the set of two sources {Ann, Bob}, because Charles and Charley are taken to be, in some sense, dependent sources whereas Ann and Bob are strangers to each other. Dependencies between the twins can arise due to their shared experiences, similar (physical or political) points of view and/or comparable perception deficiencies.

Bovens & Hartmann are Bayesians and thus took confirmation to mean Bayesian confirmation, i.e., confirmation is measured by how much posterior probabilities change with respect to the prior. The more the posterior changes, the more confirmatory the evidence – ceteris paribus. The VET thus implies that given n reports, each equally confirming a hypothesis of interest equally strong, the more varied the sources of these reports, the greater our posterior probability that the hypothesis is true. Surprisingly, there are counter-examples!

Let us return to the twin brothers, Ann and Bob. Let us now suppose that we take their testimony to possibly be unreliable and thus fallible. Clearly, if we can somehow learn that they are reporting the truth, then we would be (more) certain about the event in question.

Since Charles and Charley are twins, their reports are – in our minds – correlated ; for example, they both are near-sighted and observed the same event taking place at some distance. Thus, hearing them consistently reporting the same news teaches us something valuable about their reliabilities, e.g., near-sightedness is either to blame for two bad reports or they both did see what happened. In particular, if our belief in them reporting these news is low if they are unreliable, and large if they are reliable, then hearing their testimony significantly increases our belief in them being both reliablein this instance, e.g., both report what took place. In other words, reports from dependent sources can significantly increase our belief in their shared reliability. On the other hand, since Ann and Bob are strangers to each other, learning that they report the same news does not change much our belief in their reliabilities.

There is, of course, also something to be said in favour of independent sources. Intuitively, independence makes the information provided by sources less prone to common deficiencies. That is, we have identified two factors that influence our beliefs:

1. dependent sources are informative about each other,
2. independent sources are less likely to suffer from common deficiencies.

Bovens & Hartmann showed that – in some cases – the former can outweigh the latter. Hearing the same news from Charles and Charley makes you believe the news more than hearing the news from Ann and Bob – the VET fails. Section 2 gives the formal framework and derivations of this surprising result.

Note that this kind of analysis can have a direct impact on practical application: for instance, information sharing in vehicular network, or more generally in distributed environment [20] has to account for such dependencies and their effects.

1.3. An imprecise probability approach to the VET. While Bovens & Hartmann showed that a Bayesian interpretation of the VET can fail, we know nothing about the VET in the setting of imprecise probabilities. Before we can even begin to analyse the thesis by formal means, we must first formally state it. While we do not see any reason to modify the formalisation of the notion of evidential variety for imprecise probabilities, we have to rethink the notion of confirmation. We think that this notion is arguably best captured in the imprecise probability setting by a reduction in uncertainty; 1) the closer the posterior is to one, the greater the confirmation, 2) the closer the lower and the upper probabilities are to each other, the lower the remaining uncertainty. The question then naturally arises whether a thusly formalised VET holds or fails. The remainder of this paper is dedicated to addressing this question.

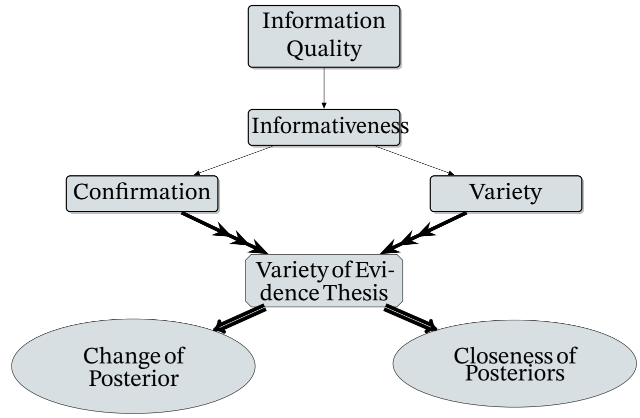


Figure 1. *Information quality increases with informativeness, which itself increases with increasing evidential variety and confirmation, increases are represented by a simple arrow. Confirmation and evidential variety are combined in the Variety of Evidence Thesis, the combination is represented by arrows with multiple heads. This thesis can be formalised in a Bayesian way in terms of a change of prior probability, formalisations are represented by double arrows.*

Structure of the paper. First, in Section 2, we provide the basic Bayesian story of the VET. Section 3 translates the basic Bayesian approach into IP and investigates its consequences. Section 4 extends more recent Bayesian analyses of the VET to IP. Section 5 offers some conclusions.

2. The Bayesian story of Bovens & Hartmann

2.1. Variables and prior probabilities. As indicated above, Bovens & Hartmann were interested in the probability that one should rationally assign to a hypothesis of interest.[1] To keep matters simple, they use a binary variable HYP, with truth values Hyp, $\neg Hyp$, and assign it a prior Bayesian probability, $P(Hyp) = h \in (0,1)$. This probability changes with new information, which comes in terms of reports from witnesses. Report variables, one variable for every report indexed by its number, formalise the content of testimony. The binary report variable being true, $REP_i = Rep_i$, means that the report says that the hypothesis is true while $REP_i = \neg Rep_i$ means that the report says that the hypothesis is false. The compound variable E represents all the available evidence, unless otherwise stated E are two positive reports.

Reports are not taken at face value. Instead, the reliability of the source that makes the report is also considered.

[1] None of the following depends on whether the hypothesis is causal or not.

It is formalised by a binary variable REL_i (the index i denotes the i-th source) with prior probability $P(Rel_i) = \rho_i$ and $P(\neg Rel_i) = \tilde{\rho}_i$. A tilde over a probability x, $\tilde{x}$, signals that $x + \tilde{x} = 1$. We use bars over and under small letters to denote probability bounds. The prior probabilities of the hypothesis and the reliabilities are independent of each other.

The probabilities of the report variables Rep_i instead depend on the reliability of the source of the report, Rel_i, and the hypothesis:[2]

$$P(Rep_i|Hyp, Rel_i) = 1 \qquad P(Rep_i|\neg Hyp, Rel_i) = 0$$
$$P(Rep_i|Hyp, \neg Rel_i) = a_i \quad P(Rep_i|\neg Hyp, \neg Rel_i) = a_i$$

Note that the parameters a_i are all also independent of the prior probabilities of the hypothesis, h, and the prior reliability of sources, ρ_i.

Assigning equal probability to a hypothesis being true, whether the hypothesis is true or not, formalises a particular notion of unreliability. In the words of Bovens & Hartmann [6, P. 57], such a source casts an (imaginary) die to determine whether they will provide a positive report (with probability a_i) or a negative report (with probability $1 - a_i$) without even looking at the state of the world. These sources are also known as *randomisers*, since they provide random reports independently of the state of the world. Such an unreliable source does not provide any information about the hypothesis of interest

$$P(Hyp) = P(Hyp|Rep_i, \neg Rel_i) = P(Hyp|\neg Rep_i, \neg Rel_i) \ .$$

This follows from the fact unreliable sources have a Bayes factor of one (for $a_i \in (0, 1)$)

$$\frac{P(Rep_i|\neg Hyp, \neg Rel_i)}{P(Rep_i|Hyp, \neg Rel_i)} = \frac{a_i}{a_i} = 1$$
$$= \frac{1 - a_i}{1 - a_i} = \frac{P(\neg Rep_i|\neg Hyp, \neg Rel_i)}{P(\neg Rep_i|Hyp, \neg Rel_i)} \ .$$

Clearly, not all unreliable sources are (best modelled as) randomisers; Section 4 investigates different notions of (un-)reliability and Section 5 offers some further considerations.

Throughout, probabilities (Bayesian and imprecise) are assumed by default to be in the open unit interval – unless stated otherwise.

For more background on (the models of) the reliability of a source see [15, 48, 51, 64, 79]. The case of causal relationships between evidence and the hypothesis of interest has been investigated in [81].

[2] The original model from Bovens and Hartmann contains one further (type of) variable between the hypothesis node and report nodes representing testable consequences of the hypothesis. These nodes are well known to be irrelevant for the questions we consider, the interested reader is referred to literature on the hierarchy of hypotheses [69, 71].

2.2. The Variety of Evidence Thesis. The *variety* of a body of evidence is now taken to be the variety of the sources, reports from independent sources have high variety while reports originating from the same (type of) source have low variety. Let us now consider n reports that all state that the hypothesis of interest holds. Holding the contents of all reports fixed, one may now expect that the more varied the sources, the greater the posterior probability of the hypothesis.

The situation is shown in Figure 2 for $n = 2$ in terms of a Bayesian network [53]. Bayesian networks are a convenient tool for representing the independences of variables in a graphical way. The relevant aspect for this paper is that the variable HYP and the variables REL are probabilistically independent, they become dependent upon conditionalisation on the reports. The only technical relevance of Bayesian networks for this paper is that the probability of an elementary event is equal to the conditional probability of the report given the other two variables multiplied by the prior probabilities of these variables.

The Variety of Evidence Thesis applied to the case of $n = 2$ then states that ceteris paribus

$$P_V(Hyp|Rep_1, Rep_2) > P_R(Hyp|Rep_1, Rep_2) \ ,$$

where P_V stands for the case of greater variety (two independent sources) and P_R stands for the replication case (one source).

The ceteris paribus conditions ensure that the difference of the probabilities $P_V(Hyp|Rep_1, Rep_2)$ and $P_R(Hyp|Rep_1, Rep_2)$ is only due to the difference in evidential variety:

1. All sources have the same a priori reliability: $\rho_i = \rho_{i+1} = \rho$.
2. All sources provide the same sort of unreliable report: $a_i = a_{i+1} = a$.

In other words, none of the indices matter, and we will hence drop them from now on.

In the next sections, we will often be interested in computing the posterior probability $P(Hyp|E)$; it is hence useful to derive complete forms of it in the two scenarios we consider.

Lemma 2.1. *The posterior probabilities are*

$$h_R := P_R(Hyp|E) = \frac{1}{1 + \frac{\tilde{h}(a^2\tilde{\rho})}{h(\rho + a^2\tilde{\rho})}} \tag{1}$$
$$h_V := P_V(Hyp|E) = \frac{1}{1 + \frac{\tilde{h}(a^2\tilde{\rho}^2)}{h(\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho})^2}} \ . \tag{2}$$

Proof. First note (also for later use) that

$$P(Hyp|E) = \frac{P(Hyp, E)}{P(E)} = \frac{1}{1 + \frac{P(\neg Hyp, E)}{P(Hyp, E)}} \ , \tag{3}$$

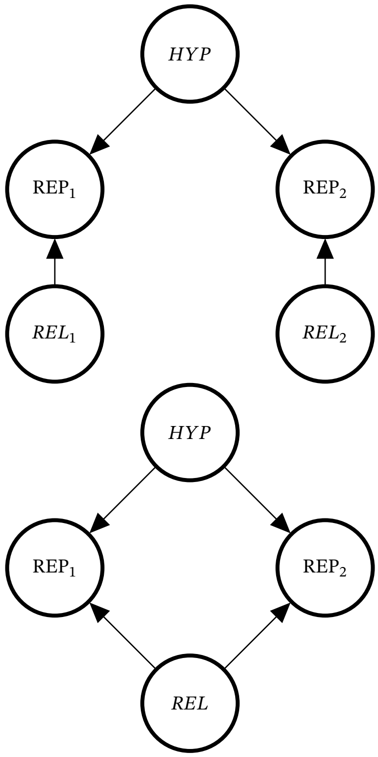


Figure 2. *Topologies of Bayesian networks for two reports concerning the hypothesis of interest that are independent (top) and dependent (bottom).*

where E is the default evidence of two positive reports Rep, Rep. Depending on our assumption, we then have to derive the terms of (3). We will only do so for $P_R(\neg Hyp, E) = \tilde{h}(a^2\tilde{\rho})$, and leave the rest to the reader, as the reasoning is similar. Recalling that $\tilde{h} = 1 - h$ we find,

$$\begin{aligned}
&P_R(\neg Hyp, Rep, Rep)\\
=&P_R(\neg Hyp, Rep, Rep, Rel) + P_R(\neg Hyp, Rep, Rep, \neg Rel)\\
=&P_R(Rep, Rep|\neg Hyp, \neg Rel)P_R(\neg Rel|\neg Hyp)P_R(\neg Hyp)\\
=&\tilde{h}P_R(\neg Rel)P_R(Rep|\neg Hyp, \neg Rel)P_R(Rep|\neg Hyp, \neg Rel)\\
=&\tilde{h}\tilde{\rho}a^2,
\end{aligned}$$

where the before last equality follows from the structure of the network in Figure 2. □

Surprisingly, the VET fails for $n = 2$ for certain parameters.

Theorem 2.1 (Bovens & Hartmann)**.** *If $h \in (0,1), a \in (0,1), \rho \in (0,1)$, then*

$$\begin{aligned}
&\operatorname{sign}(P_V(Hyp|Rep, Rep) - P_R(Hyp|Rep, Rep))\\
&= \operatorname{sign}(1 - 2\tilde{a}\tilde{\rho}) \ .
\end{aligned}$$

For small a and small ρ, the P_V posterior probability of Hyp is less than P_R posterior probability of Hyp.

Proof. If $h \in (0,1), a \in (0,1), \rho \in (0,1)$, then

$$\begin{aligned}
&\operatorname{sign}(P_V(Hyp|Rep, Rep) - P_R(Hyp|Rep, Rep))\\
&= \operatorname{sign}\Big(\frac{P_R(\neg Hyp, E)}{P_R(Hyp, E)} - \frac{P_V(\neg Hyp, E)}{P_V(Hyp, E)}\Big)\\
&= \operatorname{sign}(P_R(\neg Hyp, E)P_V(H, E) - P_R(H, E)P_V(\neg Hyp, E))\\
&= \operatorname{sign}(a^2\tilde{\rho}(\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2) - (\rho + a^2\tilde{\rho})a^2\tilde{\rho}^2)\\
&= \operatorname{sign}(\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2 - (\rho\tilde{\rho} + a^2\tilde{\rho}^2))\\
&= \operatorname{sign}(1 - \tilde{\rho} + 2a\tilde{\rho} - \tilde{\rho}) = \operatorname{sign}(1 - 2\tilde{a}\tilde{\rho}) \ ,
\end{aligned}$$

where the second equality follows from Equation (3). □

Theorem 2.1 contradicts the VET. Our next endeavour will be to see what happens to the VET when considering imprecise probabilistic models.

3. Imprecision and Variety of Evidence

We now turn to the Variety of Evidence Thesis in the framework of Imprecise Probabilities. We thus first formalise the thesis in two different ways § 3.1 and then investigate consequences of making some sets of parameters § 3.2 – § 3.4. Technically, this means that we will keep the same network structure of Figure 2 and will still apply Bayesian rules, we will however replace some of the marginal and conditional probabilities by sets of probabilities. We will thus employ the so-called *credal networks* [18] instead of Bayesian networks.

3.1. Imprecise probabilities and evidential variety. In the Bayesian framework, the intuition driving the variety of evidence thesis is that the varied evidence confirms more strongly and thus pushes the posterior probability closer to 1 and thereby reduces uncertainty. In other words, confirmation and uncertainty reduction go hand in hand. By contrast, these notions come apart in the IP framework, giving rise to different notions of the VET. More confirmation means that posterior probabilities are greater for varied evidence, whereas a reduction in uncertainty can be understood as upper and lower posterior being closer together. Before we can formalise this, we need some more notation.

We use $\mathbb{P}_R = [\underline{h}_R, \overline{h}_R]$ to denote the set of posterior probabilities which are assigned to Hyp in the replication case and $\mathbb{P}_V = [\underline{h}_V, \overline{h}_V]$ these probabilities in the variation case. We also define a partial order $\succeq_{Max}$ between intervals $\mathbb{Q} = [\underline{q}, \overline{q}]$ and $\mathbb{R} = [\underline{r}, \overline{r}]$ by:

$$\mathbb{Q} \succeq_{Max} \mathbb{R} :\Leftrightarrow (\underline{q} \geq \underline{r}) \wedge (\overline{q} \geq \overline{r})$$

to denote the pairwise dominance of bounds. We refer to such sets as *shifted*. If at least one of $\underline{q} \geq \underline{r}, \overline{q} \geq \overline{r}$ is strict, we write $\mathbb{Q} \succ_{Max} \mathbb{R}$.

We can thus frame the VET in the IP framework capturing an increase in confirmation afforded by more varied evidence by

$$\text{VET}_{\succeq} \quad : \quad \mathbb{P}_V \succeq_{Max} \mathbb{P}_R. \tag{4}$$

Alternatively, capturing the uncertainty reduction intuition one may formalise the VET by

$$\mathrm{VET}_{\subseteq} \quad : \quad \mathbb{P}_V \subseteq \mathbb{P}_R. \tag{5}$$

Nothing in the following hinges on whether we define $\mathrm{VET}_{\geq}$ or $\mathrm{VET}_{\subseteq}$ in terms of strict inequalities. We instead believe that the more interesting questions arise by comparing the two formulations of the VET put forward here.

A priori, it seems to us that both formulations are viable formalisations of the underlying intuitions. We take the formulation of two IP VETs already as a first valuable observation demonstrating how the Bayesian and the IP approach can come apart, when it comes to the value of information. Although a change of posterior probabilities is common in both approaches, a distance between posteriors makes sense only for imprecise probabilities.

3.2. Imprecise prior of the hypothesis. We will consider here and in the next section that our prior can become imprecise, and have values in the interval $[\underline{h}, \overline{h}]$ with $\underline{h} < \overline{h}$.

Such a treatment will, in particular, embed the situation of so-called near-ignorance prior [80, Section 4.6.9], modelled by the open set $(0, 1)$, that we will differentiate from complete ignorance, modelled by the closed interval $[0, 1]$. It is indeed well known that complete ignorance about priors leads to an inability to learn (and we will see that this is also the case here). Note also that near-ignorance does not guarantee the ability to learn [60], but we will not go into such considerations here (as near-ignorance allows us to learn).

$\mathbf{P(Hyp) = h \in [\underline{h}, \overline{h}]}$ *with* $\mathbf{0 < \underline{h} \leq \overline{h} < 1}$. Equation (3) gives directly that all posteriors of Hyp are in the open unit interval, $(0, 1)$. Furthermore, the posterior increases monotonically with the prior h. Hence, the minimum of the posterior obtains for a minimal prior, $\underline{h}$, and the maximum obtains for a maximal prior, $\overline{h}$.

Applying Theorem 2.1 and the fact that the posterior is monotonically increasing with the prior h, we easily find the following.

Proposition 3.1. *For all* $a \in (0, 1), \rho \in (0, 1)$ *and* $h \in [\underline{h}, \overline{h}]$ *with* $0 < \underline{h} \leq \overline{h} < 1$,

$$\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R, \text{ iff } 2\tilde{a}\tilde{\rho} > 1 \tag{6}$$
$$\underline{h}_V = \underline{h}_R \text{ and } \overline{h}_V = \overline{h}_R, \text{ iff } 2\tilde{a}\tilde{\rho} = 1 \tag{7}$$
$$\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R, \text{ iff } 2\tilde{a}\tilde{\rho} < 1 \ . \tag{8}$$

In particular, there is no case in which either $\mathbb{P}_V \subsetneqq \mathbb{P}_R$ or $\mathbb{P}_V \supsetneqq \mathbb{P}_R$. The closest we get to proper subset inclusion are the cases with $2\tilde{a}\tilde{\rho} = 1$ where the set of posterior probabilities for the variation and the replication case of Hyp are equal. In the other cases, we have $\mathbb{P}_V \succ_{Max} \mathbb{P}_R$ or $\mathbb{P}_V \prec_{Max} \mathbb{P}_R$ depending on the sign of $2\tilde{a}\tilde{\rho} - 1$, meaning that the VET also does not hold by making h imprecise since $\mathrm{VET}_{\geq}$ fails at times as badly as possible, sometimes replication confirms more strongly.

$\mathbf{P(Hyp) = h \in [\underline{h}, 1]}$. Allowing the prior probability of Hyp to reach 1 has important consequences. Firstly, the maximal posterior probability of Hyp will always be one, since a prior probability of one can, in the usual Bayesian framework, never be undone. The next proposition follows easily from

$$\underline{h}_V < \underline{h}_R, \text{ iff } 2\tilde{a}\tilde{\rho} > 1 \tag{9}$$
$$\underline{h}_V = \underline{h}_R, \text{ iff } 2\tilde{a}\tilde{\rho} = 1 \tag{10}$$
$$\underline{h}_V > \underline{h}_R, \text{ iff } 2\tilde{a}\tilde{\rho} < 1 \ . \tag{11}$$

Proposition 3.2. *For all* $a \in (0, 1), \rho \in (0, 1)$ *and* $h \in [\underline{h}, 1]$ *with* $0 < \underline{h} < 1$,

$$\mathbb{P}_V \prec_{Max} \mathbb{P}_R, \text{ iff } \mathbb{P}_V \supsetneqq \mathbb{P}_R, \text{ iff } 2\tilde{a}\tilde{\rho} > 1 \tag{12}$$
$$\mathbb{P}_V = \mathbb{P}_R, \text{ iff } 2\tilde{a}\tilde{\rho} = 1 \tag{13}$$
$$\mathbb{P}_V \succ_{Max} \mathbb{P}_R, \text{ iff } \mathbb{P}_V \subsetneqq \mathbb{P}_R, \text{ iff } 2\tilde{a}\tilde{\rho} < 1 \ . \tag{14}$$

Although the proof of this proposition is trivial, it shows that there are cases in which obtaining the information from two sources reduces uncertainty (in Hyp) more than obtaining the information from a single source ($2\tilde{a}\tilde{\rho} < 1$); however, sometimes obtaining the information from two sources *increases* uncertainty (in Hyp) more than obtaining the information from a single source ($2\tilde{a}\tilde{\rho} > 1$). In other words, the Bayesian VET holds, iff the $\mathrm{VET}_{\geq}$ holds, iff the $\mathrm{VET}_{\subseteq}$ holds.

See Figure 3 for a graphical summary.

3.3. Imprecise reliability parameter a. We still assume $0 < \underline{h} \leq \overline{h} < 1$ in this subsection, but we now turn to the case where the parameter $a = [\underline{a}, \overline{a}]$ is also imprecise. Recall that $P(Rep|Hyp, \neg Rel) = a$ is the probability that the source provides a positive report, if it is unreliable. Letting a be imprecise can be interpreted as the fact that if the source is unreliable, we have vacuous knowledge about the randomisation parameter a governing the casting of the imaginary die.

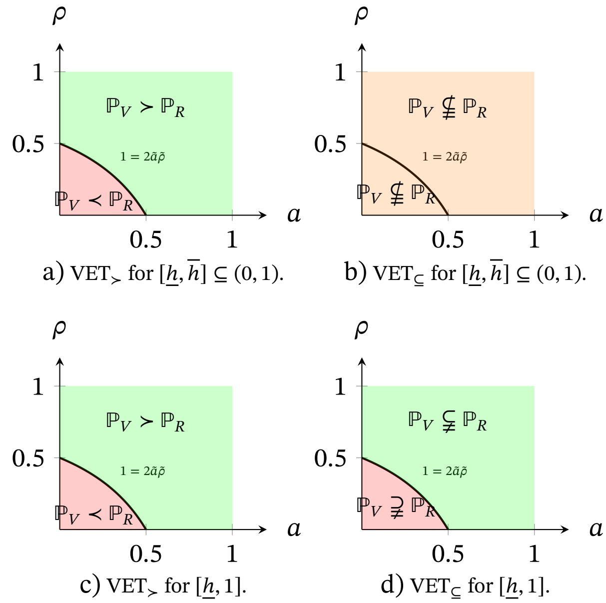
We first let a range over the entire unit interval:

Proposition 3.3. *Let* $0 = \underline{a}, \overline{a} = 1$. *For all* $\rho \in (0, 1)$ *and* $h \in [\underline{h}, \overline{h}]$ *with* $0 < \underline{h} \leq \overline{h} < 1$,

$$[\frac{1}{1 + \frac{\tilde{h}}{h}\tilde{\rho}^2}, 1] = \mathbb{P}_V \quad \begin{matrix} \succ_{Max} \\ \subsetneqq \end{matrix} \quad \mathbb{P}_R = [\frac{1}{1 + \frac{\tilde{h}}{h}\tilde{\rho}}, 1] \ .$$

Proof. First, note that for $a = 0$ an unreliable report will surely be negative. Hence, obtaining a positive report entails that the report is instead reliable and the posterior of Hyp is thus one.

In order to find minimal and maximal posterior probabilities, we compute derivatives of posterior probabilities



a) $\text{VET}_{\succ}$ for $[\underline{h},\overline{h}] \subseteq (0,1)$. b) $\text{VET}_{\subseteq}$ for $[\underline{h},\overline{h}] \subseteq (0,1)$.

c) $\text{VET}_{\succ}$ for $[\underline{h},1]$. d) $\text{VET}_{\subseteq}$ for $[\underline{h},1]$.

Figure 3. *IP VET satisfaction for imprecise h as a function of ρ, a. Green means that the VET holds, red that replication trumps variation, while orange means that neither is the case.*

of Hyp with respect to a (at the boundaries of the unit interval we take the right and left derivative since probabilities are not defined outside the unit interval):

$$\begin{aligned}\text{Replication}: \quad &\frac{\partial}{\partial a}\frac{a^2\tilde{\rho}}{\rho + a^2\tilde{\rho}}\\ =&\frac{2a\tilde{\rho}(\rho + a^2\tilde{\rho}) - 2a^3\tilde{\rho}^2}{(\rho + a^2\tilde{\rho})^2} = \frac{2a\tilde{\rho}\rho}{(\rho + a^2\tilde{\rho})^2} > 0\\ \text{Variation}: \quad &\frac{\partial}{\partial a}\frac{a^2\tilde{\rho}^2}{\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2}\\ =&\frac{2a\tilde{\rho}^2(\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2) - a^2\tilde{\rho}^2(2\rho\tilde{\rho} + 2a\tilde{\rho}^2)}{(\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2)^2}\\ =&\frac{2a\tilde{\rho}^2(\rho^2 + a\rho\tilde{\rho})}{(\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2)^2} > 0 \ .\end{aligned}$$

To see that these inequalities holds, simply observe that all numerator and denominator terms are positive probabilities. So, minimal posteriors of Hyp, $\underline{h}_R$ and $\underline{h}_V$, obtain for maximal $a, \overline{a}$, and maximal posteriors of Hyp, $\overline{h}_R$ and $\underline{h}_V$, obtain for minimal a, $\underline{a}$.

Secondly, if $a = 1$, then

$$\begin{aligned}P_R(Hyp|Rep, Rep) &= \frac{1}{1 + \frac{\tilde{h}}{h}\frac{a^2\tilde{\rho}}{\rho + a^2\tilde{\rho}}} = \frac{1}{1 + \frac{\tilde{h}}{h}\tilde{\rho}}\\ P_V(Hyp|Rep, Rep) &= \frac{1}{1 + \frac{\tilde{h}}{h}\frac{a^2\tilde{\rho}^2}{\rho^2 + 2a\rho\tilde{\rho} + a^2\tilde{\rho}^2}} = \frac{1}{1 + \frac{\tilde{h}}{h}\tilde{\rho}^2} \ .\end{aligned}$$

From $\rho, h \in (0,1)$ we thus obtain that $\underline{h}_R < \underline{h}_V$ and so

$$[\frac{1}{1 + \frac{\tilde{h}}{h}\tilde{\rho}^2}, 1] = \mathbb{P}_V \quad \underset{\subsetneqq}{\succ_{Max}} \quad \mathbb{P}_R = [\frac{1}{1 + \frac{\tilde{h}}{h}\tilde{\rho}}, 1] \ .$$

□

This means that, at least in this setting, $\text{VET}_{\succeq}$ and $\text{VET}_{\subseteq}$ both hold, yet provide a trivial upper bound. It is however a promising result, as IP presents at least one way to verify both VETs, in contrast with precise probabilities.

Let us now see what happens for generic intervals $[\underline{a}, \overline{a}] \subseteq (0,1)$, capturing near-ignorance information on a as a special yet interesting limit case.

$\mathbf{0 < \underline{a} \leq \overline{a} < 1}$. As the next proposition shows, bounding a away from extreme probabilities does lead to different behaviour of the posterior of Hyp.

Proposition 3.4. *Let $0 < \underline{a} < \overline{a} < 1$ and $h \in [\underline{h}, \overline{h}]$ with $0 < \underline{h} \leq \overline{h} < 1$.*
For all $\rho \in [0.5, 1)$, $a \in [\underline{a}, \overline{a}]$,

$$\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R \ .$$

If $\rho \in (0, 0.5)$, then

$$\begin{aligned}&\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R, \quad &&\text{if } \underline{a} < 1 - \frac{1}{2\tilde{\rho}} < \overline{a}\\ &\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R, \quad &&\text{if } \overline{a} < 1 - \frac{1}{2\tilde{\rho}}\\ &\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R, \quad &&\text{if } \underline{a} > 1 - \frac{1}{2\tilde{\rho}} \ .\end{aligned}$$

Proof. For all $\rho \in [0.5, 1)$, $h \in (0,1)$, $a \in (0,1)$ it holds by Proposition 3.1 and Equation (8) that $P_R(Hyp|E) < P_V(Hyp|E)$. So, $\underline{h}_R < \underline{h}_V$ and $\overline{h}_R < \overline{h}_V$. Indeed, if $\rho \geq 0.5$, the constraint $2\tilde{a}\tilde{\rho} \leq 1$ is always satisfied.

For all $\rho \in (0, 0.5)$ things change. The following considerations hold for all fixed $P(Hyp) = h \in [\underline{h}, \overline{h}]$. If a is small, $2\tilde{\rho}\tilde{a} > 1$, then $P_R(Hyp|E_R) > P_V(Hyp|E_V)$. If instead a is large, $2\tilde{\rho}\tilde{a} < 1$, then $P_R(Hyp|E_R) < P_V(Hyp|E_V)$. Having established that the posteriors of Hyp are declining with increasing a in proof of Proposition 3.3 (positive partial derivatives of denominators $\frac{\partial}{\partial a}$), we see that the lower bound of the posterior of Hyp obtains for large a where $P_R(Hyp|E_R) < P_V(Hyp|E_V)$ meaning that $\underline{h}_R < \underline{h}_V$. The upper bound of the posterior of Hyp obtains for small a where $P_R(Hyp|E_R) > P_V(Hyp|E_V)$ meaning that $\overline{h}_R > \overline{h}_V$. □

In words: For large ρ, $\mathbb{P}_R$ is shifted to the left of $\mathbb{P}_V$, $\mathbb{P}_V \succ_{Max} \mathbb{P}_R$ and the $\text{VET}_{\succeq}$ holds while the $\text{VET}_{\subseteq}$ fails.

For small ρ the relationship between $\mathbb{P}_R$ and $\mathbb{P}_V$ is more delicate. For small $\overline{a}$ (where the VET fails in the

Bovens & Hartmann model) $\mathbb{P}_R$ is shifted to the left of $\mathbb{P}_V$, $\mathbb{P}_V \succ_{Max} \mathbb{P}_R$ and the VET$_{\geq}$ holds while the VET$_{\subseteq}$ fails. For large $\underline{a}$ (where the VET holds in the Bovens & Hartmann model) $\mathbb{P}_R$ is shifted to the right of $\mathbb{P}_V$, $\mathbb{P}_V \prec_{Max} \mathbb{P}_R$ and the VET$_{\geq}$ fails while the VET$_{\subseteq}$ holds.

If $1 - \frac{1}{2\tilde{\rho}}$ is in between $\underline{a}$ and $\overline{a}$, we have inclusion between the two posterior intervals. This is the only case in which the VET$_{\subseteq}$ holds par excellence in the sense that $\mathbb{P}_V$ is in the interior of $\mathbb{P}_R$, $[\underline{h}_V, \overline{h}_V] \subset (\underline{h}_R, \overline{h}_R)$. Note that this situation covers the near-ignorance prior one, where $\underline{a}$ is arbitrarily close to 0, and $\overline{a}$ close to 1.

Finally, we have yet to find a case, in which the converse is true and the VET $_{\geq}$ fails par excellence.

3.4. Full imprecision. Previously, we made the conditional probability of a report imprecise allowing a to range over an interval of probabilities. We however kept this probability to be the same for Hyp and $\neg Hyp$, that is assuming that $P(Rep|Hyp, \neg Rel) = P(Rep|\neg Hyp, \neg Rel)$.

Yet, in IP it is common to assume that we have identical intervals for $P(Rep|Hyp, \neg Rel)$ and $P(Rep|\neg Hyp, \neg Rel)$, but to allow one to pick different precise probabilities within each of them. This would mean in an IP setting that our beliefs about receiving a positive report in case of unreliable sources are the same, but that the probabilities within such intervals could be different. This is, for instance, what happens when considering epistemic irrelevance [19].

We could therefore have gone one step further and allowed for more imprecision in the following sense:

$$P(Rep|Hyp, \neg Rel) = a_1 \in [\underline{a}, \overline{a}] \subseteq [0,1] \text{ with } \underline{a} < \overline{a}$$
$$P(Rep|\neg Hyp, \neg Rel) = a_0 \in [\underline{a}, \overline{a}] \subseteq [0,1] \text{ with } \underline{a} < \overline{a}$$

Splitting the parameter a into two independent parameters has however drastic consequences. Consider what we would then think about the hypothesis given a single positive report from an unreliable source:

$$P(Hyp|\neg Rel, Rep) = \frac{h\tilde{\rho}a_1}{h\tilde{\rho}a_1 + \tilde{h}\tilde{\rho}a_0} = \frac{ha_1}{ha_1 + \tilde{h}a_0}$$

For $a_1 = 1, a_0 = 0$ this posterior is 1 and for $a_1 = 0, a_0 > 0$ this posterior is 0. So, the posterior probability of interest can be everywhere in the entire unit interval – given just one single report from an unreliable source. This would be true even if we started with a precise prior h. However, if we are sure that the source is unreliable, it would make sense to require that our rational belief in Hyp remains unchanged, whatever the information provided by the source. Due to this counter-intuitive fact we are here not considering "full imprecision" of the parameter a.

The requirement that the posterior probability of the hypothesis given a report from an unreliable source remains constant is given by

$$P(Hyp) = h = P(Hyp|\neg Rel, Rep) = \frac{ha_1}{ha_1 + \tilde{h}a_0} \ ,$$

which is equivalent to $a_0 = a_1$. This however means that a_0 and a_1 are the same after all. So, even in case of non-extreme probabilities for a_0, a_1, it seems not desirable to allow them to come apart.

4. The VET of Osimani & Landes

The VET has also been studied using different notions of reliability in the Bayesian framework. Most pertinent is the approach by Osimani & Landes [56], which we now investigate in our IP formalisation. We are again only interested in posterior probabilities of the hypothesis of interest given two positive reports, which either share a reliability variable as parent or have different reliability parent variables, see Figure 2.

4.1. Their model. There are two main differences to the Bovens & Hartmann model concerning the reliability of a source in the Osimani & Landes model. The first substantial change is that even reliable sources are not 100% accurate and get things wrong from time to time, $1 > \epsilon_+, \epsilon_- > 0$:

$$P(Rep|Hyp, Rel) = 1 - \epsilon_+ \qquad P(Rep|\neg Hyp, Rel) = \epsilon_-$$
$$P(Rep|Hyp, \neg Rel) = \alpha \quad P(Rep|\neg Hyp, \neg Rel) = \gamma \ .$$

Furthermore, reliable sources can be trusted in the sense that a positive report increases our belief that the hypothesis is true. This is captured in terms of a Bayes factor, a ratio of likelihoods, being greater than one

$$\frac{P(Rep|Hyp, Rel)}{P(Rep|\neg Hyp, Rel)} = \frac{1 - \epsilon_+}{\epsilon_-} > 1 \ .$$

This condition is obviously equivalent to $\epsilon_+ + \epsilon_- < 1$.

Secondly, unreliability is now understood differently. Osimani & Landes construe unreliability as *biased towards reporting a particular way*. The motivation for this construal are (financial) incentives increasing the probability of reports that align with incentives. For example, a study funded by the manufacturer of the investigated drug is more likely to report that the drug has the intended consequence than a study without such funding [47, 52]. Note in particular that there is no requirement to model unreliable sources as randomisers. In the Osimani & Landes approach, we can thus learn even from unreliable sources. The main idea is that while unreliable sources grant some learning, reliable sources ought to be "epistemically superior" in some sense.

There are 4 main options to capture a directed bias:[3] 1) $\alpha > 1 - \epsilon_+$ and $\gamma > \epsilon_-$. Such an unreliable source

[3] Further options arise by allowing some parameters to be equal. We shall however concentrate on these four options.

makes a positive report more often than a reliable source – independently of the truth of the hypothesis. The contrast is such between a reliable source and a source saying "yes" more often.
2) $1-\epsilon_+ > \alpha$ and $\epsilon_- > \gamma$. Such an unreliable source makes a negative report more often than a reliable source – independently of the truth of the hypothesis. The contrast is such between a reliable source and a source saying "no" more often.
3) $1-\epsilon_+ > \alpha$ and $\gamma > \epsilon_-$. Such an unreliable source makes errors of the first and second kind more often than the reliable source. The contrast is such between a better and a worse source.
4) $\alpha > 1-\epsilon_+$ and $\epsilon_- > \gamma$. Such an unreliable source makes errors of the first and second kind less often than the reliable source. The contrast is such between a better and a worse source, however we've mis-labelled which source is reliable.

Swapping reliability and unreliability, we note that formally the first two and the second two formalisations are mathematically equivalent. For $\epsilon_+ = \epsilon_- = 0$ and $\alpha = \gamma$ we obtain the Bovens & Hartmann model as a limiting case of Osimani & Landes model.

The general formula determining whether $P_R(Hyp|Rep, Rep)$ or $P_V(Hyp|Rep, Rep)$ is greater, is known but somewhat involved [42, Proposition 9]:

$$
\begin{aligned}
&\operatorname{sign}(P_V(Hyp|E) - P_R(Hyp|E)) \\
&\quad = \operatorname{sign}\Big((\alpha\epsilon_- - \gamma(1-\epsilon_+)) \quad \cdot \\
&\qquad (\rho[\epsilon_-[(1-\epsilon_+)-\alpha] + (1-\epsilon_+)[\epsilon_- - \gamma]] \\
&\qquad + \tilde{\rho}[\gamma((1-\epsilon_+)-\alpha) + \alpha(\epsilon_- - \gamma)]\Big) \ .
\end{aligned}
\tag{15}
$$

Osimani & Landes studied sources, which were unreliable due to sponsorship bias, and took Option 1.

The key thing to remember and going forward is that for all four options there are parameters such that $P_V(Hyp|E) > P_R(Hyp|E)$, but there are also parameters such that $P_V(Hyp|E) < P_R(Hyp|E)$.

4.2. Imprecise prior of the hypothesis.

$\mathbf{P(Hyp)} = \mathbf{h} \in [\underline{\mathbf{h}}, \overline{\mathbf{h}}]$ *with* $\mathbf{0} < \underline{\mathbf{h}} \leq \overline{\mathbf{h}} < \mathbf{1}$. Exactly like in Section 3.2, Equation (3) gives directly that all posteriors of Hyp are in the open unit interval, $(0, 1)$. Furthermore, the minimum of the posterior obtains for a minimal prior, $\underline{h}$, and the maximum obtains for a maximal prior, $\overline{h}$.

Proposition 4.1. *Let $a \in (0,1)$, $\rho \in (0,1)$ and $h \in [\underline{h}, \overline{h}]$ with $0 < \underline{h} \leq \overline{h} < 1$.*
If $\alpha > 1-\epsilon_+$ and $\gamma > \epsilon_-$ (Option 1), then

$$
\begin{aligned}
\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R,\ \textit{iff}\ \frac{\alpha}{\gamma} > \frac{1-\epsilon_+}{\epsilon_-} \\
\underline{h}_V = \underline{h}_R \text{ and } \overline{h}_V = \overline{h}_R,\ \textit{iff}\ \frac{\alpha}{\gamma} = \frac{1-\epsilon_+}{\epsilon_-} \\
\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R,\ \textit{iff}\ \frac{\alpha}{\gamma} < \frac{1-\epsilon_+}{\epsilon_-} \ .
\end{aligned}
$$

If $1-\epsilon_+ > \alpha$ and $\epsilon_- > \gamma$ (Option 2), then

$$
\begin{aligned}
\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R,\ \textit{iff}\ \frac{\alpha}{\gamma} < \frac{1-\epsilon_+}{\epsilon_-} \ , \\
\underline{h}_V = \underline{h}_R \text{ and } \overline{h}_V = \overline{h}_R,\ \textit{iff}\ \frac{\alpha}{\gamma} = \frac{1-\epsilon_+}{\epsilon_-} \ , \\
\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R,\ \textit{iff}\ \frac{\alpha}{\gamma} > \frac{1-\epsilon_+}{\epsilon_-} \ .
\end{aligned}
$$

If $1-\epsilon_+ > \alpha$ and $\gamma > \epsilon_-$ (Option 3), then

$$
\begin{aligned}
&\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R,\ \textit{if}\ 2\alpha\gamma > \gamma(1-\epsilon_+) + \alpha\epsilon_- \ , \\
&\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R,\ \textit{if}\ 2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_- \\
&\qquad\qquad \text{and } \rho > -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+) - \alpha)}, \\
&\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R,\ \textit{if}\ 2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_- \\
&\qquad\qquad \text{and } \rho < -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+) - \alpha)} \ .
\end{aligned}
$$

If $\alpha > 1-\epsilon_+$ and $\epsilon_- > \gamma$ (Option 4), then

$$
\begin{aligned}
&\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R,\ \textit{if}\ 2\alpha\gamma > \gamma(1-\epsilon_+) + \alpha\epsilon_- \ , \\
&\underline{h}_V < \underline{h}_R \text{ and } \overline{h}_V < \overline{h}_R,\ \textit{if}\ 2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_- \\
&\qquad\qquad \text{and } \rho > -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+) - \alpha)} \ , \\
&\underline{h}_V > \underline{h}_R \text{ and } \overline{h}_V > \overline{h}_R,\ \textit{if}\ 2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_- \\
&\qquad\qquad \text{and } \rho < -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+) - \alpha)} \ .
\end{aligned}
$$

Aligning with Proposition 3.1 we find that – depending on parameters – either $\mathbb{P}_V \prec_{Max} \mathbb{P}_R$, $\mathbb{P}_V = \mathbb{P}_R$ or $\mathbb{P}_V \succ_{Max} \mathbb{P}_R$.

Proof. According to Option 1, the bracket in (15) containing $\rho, \tilde{\rho}$ is negative and we find [56, Theorem 1]

$$
\operatorname{sign}(P_R(Hyp|E) - P_V(Hyp|E)) = \operatorname{sign}\Big(\frac{\alpha}{\gamma} - \frac{1-\epsilon_+}{\epsilon_-}\Big) \ .
$$

According to Option 2 the bracket containing $\rho, \tilde{\rho}$ is instead positive and we obtain the exact opposite of Option 1:

$$
\operatorname{sign}(P_V(Hyp|E) - P_R(Hyp|E)) = \operatorname{sign}\Big(\frac{\alpha}{\gamma} - \frac{1-\epsilon_+}{\epsilon_-}\Big) \ .
$$

Unsurprisingly, the first two options can be obtained from another by swapping α with $1-\epsilon_+$ and γ with ϵ_-.

According to Option 3 and 4, $\alpha\epsilon_- - \gamma(1-\epsilon_+) \neq 0$. The sign of $P_R(Hyp|E) - P_V(Hyp|E)$ hence depends the bracket in Equation 15 containing $\rho, \tilde{\rho}$ and is also explicitly depending on $\rho = P(Rel)$. We find for Option 3 ($1-\epsilon_+ > \alpha$ and $\gamma > \epsilon_-$),

$$\begin{aligned}
&\operatorname{sign}(P_R(Hyp|E) - P_V(Hyp|E)) \\
&\quad = \operatorname{sign}(\rho[\epsilon_-[(1-\epsilon_+)-\alpha] + (1-\epsilon_+)[\epsilon_- - \gamma]] \\
&\qquad + \tilde{\rho}[\gamma((1-\epsilon_+)-\alpha) + \alpha(\epsilon_- - \gamma)]) \\
&\quad = \operatorname{sign}(\gamma(1-\epsilon_+) - 2\alpha\gamma + \alpha\epsilon_- \\
&\qquad + 2\rho[\epsilon_-((1-\epsilon_+)-\alpha) - \gamma((1-\epsilon_+)-\alpha)]) \\
&\quad = \operatorname{sign}(\gamma(1-\epsilon_+) - 2\alpha\gamma + \alpha\epsilon_- \\
&\qquad + 2\rho(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)) \\
&\quad =(-1)\operatorname{sign}\Big(\rho + \frac{\gamma(1-\epsilon_+) - 2\alpha\gamma + \alpha\epsilon_-}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)}\Big) \ .
\end{aligned}$$

The denominator is always negative. If $2\alpha\gamma > \gamma(1-\epsilon_+) + \alpha\epsilon_-$, then so is the numerator. This means that the term inside the large bracket is always positive. So, $P_V(Hyp|E) > P_R(Hyp|E)$ for all $P(Hyp) = h$.

Instead, if $2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_-$, then the numerator is positive. Hence, for $\rho > -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)}$ we have $P_V(Hyp|E) > P_R(Hyp|E)$ and for $\rho < -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)}$ we have $P_V(Hyp|E) < P_R(Hyp|E)$.

Option 4: Here $\operatorname{sign}(\frac{\alpha}{\gamma} - \frac{1-\epsilon_+}{\epsilon_-})$ is different than in Option 3. Hence,

$$\begin{aligned}
&\operatorname{sign}(P_R(Hyp|E) - P_V(Hyp|E)) \\
&\quad = \operatorname{sign}\Big(\rho + \frac{\gamma(1-\epsilon_+) - 2\alpha\gamma + \alpha\epsilon_-}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)}\Big) \ .
\end{aligned}$$

The denominator is always negative. If $2\alpha\gamma > \gamma(1-\epsilon_+) + \alpha\epsilon_-$, then so is the numerator. This means that the term inside the large bracket is always positive. So, $P_V(Hyp|E) < P_R(Hyp|E)$ for all $P(Hyp) = h$.

Instead, if $2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_-$, then the numerator is positive. Hence, for $\rho > -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)}$ we have $P_V(Hyp|E) < P_R(Hyp|E)$ and for $\rho < -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)}$ we have $P_V(Hyp|E) > P_R(Hyp|E)$. □

$\mathbf{P(Hyp) = h \in [\underline{h}, 1]}$. If we allow for $P(Hyp) = 1$, then the maximal posterior of Hyp is one, which is reached for a prior of Hyp of one. So, $\underline{h}_V > \underline{h}_R$ entails $\mathbb{P}_V \subsetneqq P_R$. Conversely, $\underline{h}_V < \underline{h}_R$ allows us to conclude $\mathbb{P}_V \supsetneqq P_R$. We now obtain as a corollary from Proposition 4.1 focussing only on Option 1 and 3:

Proposition 4.2. *Let $a \in (0,1), \rho \in (0,1)$ and $h \in [\underline{h}, 1]$ with $0 < \underline{H} < 1$.*
If $\alpha > 1-\epsilon_+$ and $\gamma > \epsilon_-$ (Option 1), then

$$\mathbb{P}_V \supsetneqq P_R, \ \textit{iff} \ \frac{\alpha}{\gamma} > \frac{1-\epsilon_+}{\epsilon_-}$$
$$\mathbb{P}_V = P_R, \ \textit{iff} \ \frac{\alpha}{\gamma} = \frac{1-\epsilon_+}{\epsilon_-}$$
$$\mathbb{P}_V \subsetneqq P_R, \ \textit{iff} \ \frac{\alpha}{\gamma} < \frac{1-\epsilon_+}{\epsilon_-} \ .$$

If $1-\epsilon_+ > \alpha$ and $\gamma > \epsilon_-$ (Option 3), then

$$\begin{aligned}
&\mathbb{P}_V \subsetneqq P_R \ , \textit{if } 2\alpha\gamma > \gamma(1-\epsilon_+) + \alpha\epsilon_- \ , \\
&\mathbb{P}_V \subsetneqq P_R \ , \textit{if } 2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_- \\
&\qquad \textit{and } \rho > -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)} \ , \\
&\mathbb{P}_V \supsetneqq P_R \ , \textit{if } 2\alpha\gamma < \gamma(1-\epsilon_+) + \alpha\epsilon_- \\
&\qquad \textit{and } \rho < -\frac{2\alpha\gamma - \alpha\epsilon_- - \gamma(1-\epsilon_+)}{2(\epsilon_- - \gamma)((1-\epsilon_+)-\alpha)} \ .
\end{aligned}$$

Analogous to Proposition 3.2 we here have subset inclusions between sets of posterior probabilities, but no shifts.

Since there is no randomisation parameter a in the Osimani & Landes model but rather 4 parameters capturing the quality of sources, it makes no sense here to consider an imprecise parameter a.

5. Conclusions

Brief summary. We formalised the Variety of Evidence Thesis in two ways within the framework of imprecise probabilities and showed that they sometimes hold and sometimes fail. Unlike previous work, we identified a case in which the VET is supported by formal analysis – even in both formalisations (Proposition 3.3). The emerging relationships between posterior probabilities and parameters are intricate and next to impossible to predict from the outset. The status of the IP Variety of Evidence Theses hangs on a knife's edge, changing parameter values by the slightest of margins can result in a reversal.

Personally, it seems to us that there are no *a priori* grounds for categorically preferring one of the two IP formalisations of the VET over the other. They are formulations of two different intuitions that both appear to be sensible. The question of whether the intuitive pull of the Bayesian formalisation of the VET capturing both intuitions simultaneously is greater; we leave for our readers to ponder.

Relevance. The results reported in this note are immediately relevant to i) the Theory of Imprecise Probability and its relation to the Bayesian approach [66], ii) the value of information [2, 30], iii) the importance of replication studies, in particular in light of initial results that are possibly unreliable [1, 3, 12, 24, 33, 55, 65, 68, 78, 82], iv) judgement aggregation and the value of information [36, 59], in particular for updating imprecise probabilities [7, 49, 72].

The strong constraints of the precise Bayesian framework often enforces that distinct notions take the same formal appearance, and increasing the expressiveness of the Bayesian framework through the use of IP and probability sets often allow to gain new insights by being able to formally distinguish those same notions. This is the case, for instance, for the notion of independence [17] or conditioning [29]. This is also the case of our study, as confirmation increase and uncertainty reduction are conflated in the Bayesian approach, but come apart in IP. *If* one commits to the Bayesian formulation of the VET in all its power, then one also seems to be committed to *both* formulations of the VET in IP and subsequently faces a considerable increase in VET failures. It has furthermore demonstrated that IP offers the possibility of a more fine-grained investigation of the undeniable intuitive appeal of the VET and thus offers prospects of insights that are not achievable via Bayesian analyses. We particularly want to raise the possibility of further formulations of the VET, leveraging other languages and theories such as the theory of evidence. All of this further challenges our pre-conceived notions of evidential variety and its consequences for confirmation. Clearly, there is much more to be learned.

Less obvious is perhaps the relevance to decision-making and (science) policy-making. We here showed that it is sometimes better to use the same method of investigation twice rather than diversifying modes of inquiry. Hence, it can be beneficial to concentrate the acquisition of information on one type of inquiry. For example, there is a long-standing debate in the methodology of medicine (and now also other sciences) about how to evaluate causal hypotheses [43, 67, 83]. Should one rely on *one kind of method* (randomised studies) to determine the benefits of a drug [37, 77] or should one draw on a *variety of methods* [27, 31, 46, 57]? This paper shows that even if one grants a ceteris paribus assumption between randomised studies and non-randomised studies (bench research and observational studies); proponents of the former view emphatically reject such assumptions; then methodological variety can be inferior. The existence of instances in which varied evidence confirms less strongly arguably decreases the intuitive appeal of the second position.

Furthermore, there is some relevance to the principles of engineering: given that one may use two sensors, should both of them be (of the) same (type) or should they be (of) different (types)?

Future work. may investigate translations of different Bayesian versions of Variety of Evidence Thesis [38, 40, 41] into the realm of imprecise probabilities; for example, the study of other forms of (un-)reliability, formalised by a reliability variable of arity three or greater [54, Section 4.3], cf. [13, 14, 32, 58]. One could also investigate whether it ever happens that $\mathbb{P}_R$ is to the left or right of $\mathbb{P}_V$ (non-overlapping); i.e., do there exist cases with $\overline{h}_R < \underline{h}_V$ or $\overline{h}_V < \underline{h}_R$, ?

Currently, the study of three or more items of evidence does not appear promising, as none of the previous works on multiple items of evidence have found qualitatively new results [34, 56].

More widely, one could study the value of information and formalisations of the Variety of Evidence Thesis in other terms or within other frameworks of uncertain reasoning such as *ranking functions* [75] or *imaging* [28, 44, 45], some connections to voting theory are traced in [25, § 3.2], see also [31, 63, 82]. An immediate first question is whether the current conclusions extend to valuation-based systems [73] and Dempster-Shafer modelling of reliability [62]. In particular, the work of Pichon *et al.* [61, 62] draws a distinction between the reliability of a source and its trustworthiness within IP. The idea of multiple distinct dimensions of evidence has also emerged elsewhere [43, § 3.3], [9, 11, 70]. Such distinctions may motivate further ways of reasoning in the Bovens & Hartmann and/or the Osimani & Landes model.

Additional Author Information

Acknowledgements. Jürgen Landes gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 528031869. Part of this research was done during a visit of Jürgen Landes at UTC, funded by the SAFE AI chair.

References

[1] Harald Atmanspacher and Sabine Maasen, eds. *Reproducibility: Principles, Problems, Practices, and Prospects*. Hoboken: Wiley, 2016.

[2] Bengt Autzen. “Significance testing, p-values and the principle of total evidence”. In: *European Journal for Philosophy of Science* 6 (2 2016), pp. 281–295. DOI: 10.1007/s13194-016-0138-0.

[3] Florian Johannes Boge. “Why trust a simulation? Models, parameters, and robustness in simulation-infected experiments”. In: *British Journal for the Philosophy of Science* 75.4 (2024). DOI: 10.1086/716542.

[4] George F. Borm, Oscar Lemmers, Jaap Fransen, and Rogier Donders. “The evidence provided by a single trial is less reliable than its statistical analysis suggests”. In: *Journal of Clinical Epidemiology* 62.7 (2009), pp. 711–715. DOI: 10.1016/j.jclinepi.2008.09.013.

[5] Luc Bovens and Stephan Hartmann. “Bayesian Networks and the Problem of Unreliable Instruments”. In: *Philosophy of Science* 69.1 (2002), pp. 29–72. DOI: 10.1086/338940.

[6] Luc Bovens and Stephan Hartmann. *Bayesian Epistemology*. Oxford: Oxford University Press, 2003.

[7] Michele Caprio and Teddy Seidenfeld. “Constriction for Sets of Probabilities”. In: *Proceedings of Machine Learning Research* 215 (2023), pp. 84–95. URL: https://proceedings.mlr.press/v215/caprio23b/caprio23b.pdf.

[8] Rudolf Carnap. *Logical Foundations of Probability*. 2nd ed. Chicago: University of Chicago Press, 1962.

[9] Nancy Cartwright. “Evidence-based policy: what’s to be done about relevance?” In: *Philosophical Studies* 143.1 (2009), pp. 127–136. DOI: 10.1007/s11098-008-9311-4.

[10] Nancy Cartwright. “Rigour versus the need for evidential diversity”. In: *Synthese* 199 (2021), pp. 13095–13119. DOI: 10.1007/s11229-021-03368-1.

[11] Nancy Cartwright and Jacob Stegenga. “A Theory of Evidence for Evidence-Based Policy”. In: *Evidence, Inference and Enquiry*. Ed. by Mimi Vasilaki Phil Dawid William Twinning. OUP, 2011. Chap. 11, pp. 291–322.

[12] Lorenzo Casini and Jürgen Landes. “Confirmation by Robustness Analysis. A Bayesian Account”. In: *Erkenntnis* 89 (2024), pp. 367–409. DOI: 10.1007/s10670-022-00537-7.

[13] François Claveau. “The Independence Condition in the Variety-of-Evidence Thesis”. In: *Philosophy of Science* 80.1 (2013), pp. 94–118. DOI: 10.1086/668877.

[14] François Claveau and Olivier Grenier. “The Variety-of-Evidence Thesis: A Bayesian Exploration of its Surprising Failures”. In: *Synthese* 196 (8 2019), pp. 3001–3028. DOI: 10.1007/s11229-017-1607-5.

[15] Peter J. Collins, Ulrike Hahn, Ylva von Gerber, and Erik J. Olsson. “The Bi-directional Relationship between Source Characteristics and Message Content”. In: *Frontiers in Psychology* 9 (2018), pp. 1–16. DOI: 10.3389/fpsyg.2018.00018.

[16] Nathan Couch. “The diversity principle and the evaluation of evidence”. In: *Psychonomic Bulletin & Review* 29 (2022), pp. 1270–1294. DOI: 10.3758/s13423-022-02065-5.

[17] Inés Couso, Serafín Moral, and Peter Walley. “A survey of concepts of independence for imprecise probabilities”. In: *Risk, Decision and Policy* 5.2 (2000), pp. 165–181. DOI: 10.1017/S1357530900000156.

[18] Fabio G Cozman. “Credal networks”. In: *Artificial Intelligence* 120.2 (2000), pp. 199–233. DOI: 10.1016/S0004-3702(00)00029-1.

[19] Jasper De Bock. “Credal networks under epistemic irrelevance”. In: *International Journal of Approximate Reasoning* 85 (2017), pp. 107–138. DOI: 10.1016/j.ijar.2014.07.002.

[20] Thierry Denoeux. “Distributed combination of belief functions”. In: *Information Fusion* 65 (2021), pp. 179–191. DOI: 10.1016/j.inffus.2020.09.001.

[21] Paul Downward and Andrew Mearman. “Retroduction as mixed-methods triangulation in economic research: reorienting economics into social science”. In: *Cambridge Journal of Economics* 31.1 (2007), pp. 77–99. DOI: 10.1093/cje/bel009.

[22] John Earman. *Bayes or Bust?* Cambridge: MIT Press, 1992. URL: http://pitt.edu/~jearman/Earman_1992BayesOrBust.pdf.

[23] Branden Fitelson. “Wayne, Horwich, and Evidential Diversity”. In: *Philosophy of Science* 63.4 (1996), pp. 652–660. DOI: 10.1086/289982.

[24] Samuel C. Fletcher. “How (not) to measure replication”. In: *European Journal for Philosophy of Science* 11.2 (2021). DOI: 10.1007/s13194-021-00377-2.

[25] Samuel C. Fletcher, Jürgen Landes, and Roland Poellinger. “Evidence Amalgamation in the Sciences: An Introduction”. In: *Synthese* 196 (8 2019), pp. 3163–3188. DOI: 10.1007/s11229-018-1840-6.

[26] Allan Franklin and Colin Howson. “Why do scientists prefer to vary their experiments?” In: *Studies in History and Philosophy of Science Part A* 15.1 (1984), pp. 51–62. DOI: 10.1016/0039-3681(84)90029-3.

[27] Jonathan Fuller, Benjamin Chin-Yee, and Ross E. G. Upshur. “The argument framework is a flexible approach to evidence in healthcare”. In: *Nature Medicine* (2024). DOI: 10.1038/s41591-024-02930-x.

[28] Peter Gärdenfors. “Imaging and Conditionalization”. In: *Journal of Philosophy* 79.12 (1982), p. 747. DOI: 10.2307/2026039.

[29] Ruobin Gong and Xiao-Li Meng. “Judicious Judgment Meets Unsettling Updating”. In: *Statistical Science* 36.2 (2021), pp. 169–190. DOI: 10.1214/19-STS765.

[30] I. J. Good. “On the Principle of Total Evidence”. In: *British Journal for the Philosophy of Science* 17.4 (1967), pp. 319–321. DOI: 10.1093/bjps/17.4.319.

[31] Sirena Gutierrez, M. Maria Glymour, and George Davey Smith. "Evidence triangulation in health research". In: *European Journal of Epidemiology* (2025). DOI: 10.1007/s10654-024-01194-6.

[32] Rolf Haenni and Stephan Hartmann. "Modeling partially reliable information sources: A general approach based on Dempster–Shafer theory". In: *Information Fusion* 7.4 (2006), pp. 361–379. DOI: 10.1016/j.inffus.2005.06.005.

[33] Margherita Harris. "The epistemic value of independent lies: false analogies and equivocations". In: *Synthese* 199 (5-6 2021), pp. 14577–14597. DOI: 10.1007/s11229-021-03434-8.

[34] Stephan Hartmann and Luc Bovens. *The Variety-of-Evidence Thesis and the Reliability of Instruments: A Bayesian-Network Approach*. PhilSci-Archive, last modified: 07 Oct 2010 15:10. 2001. URL: http://philsci-archive.pitt.edu/235/.

[35] Carl Hempel. *Philosophy of Natural Science*. Upper Saddle River: Prentice Hall, 1966.

[36] Leah Henderson and Alexander Gebharter. "The role of source reliability in belief polarisation". In: *Synthese* 199.3-4 (2021), pp. 10253–10276. DOI: 10.1007/s11229-021-03244-y.

[37] Bennett Holman. "Philosophers on drugs". In: *Synthese* 196 (11 2019), pp. 4363–4390. DOI: 10.1007/s11229-017-1642-2.

[38] Paul Horwich. *Probability and evidence*. Cambridge: Cambridge University Press, 1982.

[39] Joachim Hüffmeier, Jens Mazei, and Thomas Schultze. "Reconceptualizing replication as a sequence of different studies: A replication typology". In: *Journal of Experimental Social Psychology* 66 (2016), pp. 81–92. DOI: 10.1016/j.jesp.2015.09.009.

[40] Jürgen Landes. "Variety of evidence and the elimination of hypotheses". In: *European Journal for Philosophy of Science* 10 (2 2020). DOI: 10.1007/s13194-019-0272-6.

[41] Jürgen Landes. "Variety of Evidence". In: *Erkenntnis* 85 (1 2020), pp. 183–223. DOI: 10.1007/s10670-018-0024-6.

[42] Jürgen Landes. "The variety of evidence thesis and its independence of degrees of independence". In: *Synthese* 198 (11 2021), pp. 10611–10641. DOI: 10.1007/s11229-020-02738-5.

[43] Jürgen Landes, Barbara Osimani, and Roland Poellinger. "Epistemology of Causal Inference in Pharmacology". In: *European Journal for Philosophy of Science* 8 (1 2018), pp. 3–49. DOI: 10.1007/s13194-017-0169-1.

[44] Hannes Leitgeb. "Imaging all the people". In: *Episteme* 14 (4 2017), pp. 463–479. DOI: 10.1017/epi.2016.14.

[45] David Lewis. "Probabilities of Conditionals and Conditional Probabilities". In: *Philosophical Review* 85.3 (1976), pp. 297–315. DOI: 10.2307/2184045.

[46] José Luis Luján. "Evidence-based policies: Lessons from regulatory science". In: *Politics & Policy* (2023). DOI: 10.1111/polp.12543.

[47] Andreas Lundh, Joel Lexchin, Barbara Mintzes, Jeppe B Schroll, and Lisa Bero. "Industry sponsorship and research outcome (Review)". In: *Cochrane Database of Systematic Reviews* (2 2017). Art. No.: MR000033. DOI: 10.1002/14651858.MR000033.pub3.

[48] George Masterton and Erik J. Olsson. "Argumentation and belief updating in social networks: a Bayesian approach". In: *Trends in belief revision and argumentation dynamics*. Ed. by E Fermé, D. Gabbay, and G. Simari. College Publications, 2013.

[49] Conor Mayo-Wilson and Gregory Wheeler. "Scoring Imprecise Credences: A Mildly Immodest Proposal". In: *Philosophy and Phenomenological Research* 93.1 (2016), pp. 55–78. DOI: 10.1111/phpr.12256.

[50] Paul E. Meehl. "Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It". In: *Psychological Inquiry* 1.2 (1990), pp. 108–141. DOI: 10.1207/s15327965pli0102_1.

[51] Christoph Merdes, Momme von Sydow, and Ulrike Hahn. "Formal models of source reliability". In: *Synthese* 198 (23 2021), pp. 5773–5801. DOI: 10.1007/s11229-020-02595-2.

[52] National Academies of Sciences, Engineering, and Medicine. *Sponsor Influences on the Quality and Independence of Health Research*. Ed. by Dara Rosenberg, Elizabeth Boyle, Alexandra McKay, and Joe Alper. Washington: National Academies Press, 2023. DOI: 10.17226/27056.

[53] Richard E. Neapolitan. *Learning Bayesian Networks*. Upper Saddle River: Pearson, 2003.

[54] Erik J. Olsson. *Against Coherence: Truth, Probability, and Justification*. Oxford: Oxford University Press, 2005.

[55] Open Science Collaboration. "Estimating the reproducibility of psychological science". In: *Science* 349.6251 (2015). DOI: 10.1126/science.aac4716.

[56] Barbara Osimani and Jürgen Landes. "Varieties of Error and Varieties of Evidence in Scientific Inference". In: *British Journal for the Philosophy of Science* 74.1 (2023), pp. 117–170. DOI: 10.1086/714803.

[57] Veli-Pekka Parkkinen et al. *Evaluating evidence of mechanisms in medicine: Principles and procedures*. Springer, 2018. DOI: 10.1007/978-3-319-94610-8.

[58] Kong Ngai Pei and Chin Shing Arthur Chin. "Towards an empirically informed normative Bayesian scheme-based account of argument from expert opinion". In: *Thinking & Reasoning* 29.4 (2023), pp. 726–759. DOI: 10.1080/13546783.2022.2134926.

[59] Richard Pettigrew. "Aggregating agents with opinions about different propositions". In: *Synthese* 200.5 (2022). DOI: 10.1007/s11229-022-03850-4.

[60] Alberto Piatti, Marco Zaffalon, Fabio Trojani, and Marcus Hutter. "Learning about a Categorical Latent Variable under Prior Near-Ignorance". In: *Proceedings of ISIPTA*. Ed. by Gert de Cooman, Jirina Vejnarová, and Marco Zaffalon. 2007, pp. 347–356. URL: https://core.ac.uk/reader/55855370#page=369.

[61] Frédéric Pichon, Didier Dubois, and Thierry Denœux. "Quality of Information Sources in Information Fusion". In: *Information Quality in Information Fusion and Decision Making*. Cham: Springer, 2019. Chap. 2, pp. 31–49. DOI: 10.1007/978-3-030-03643-0_2.

[62] Frédéric Pichon, Didier Dubois, and Thierry Denoeux. "Relevance and truthfulness in information correction and fusion". In: *International Journal of Approximate Reasoning* 53.2 (2012), pp. 159–175. DOI: 10.1016/j.ijar.2011.02.006.

[63] Toby D. Pilditch, Ulrike Hahn, and David Lagnado. "The problem of dependency". In: *Synthese* 205.4 (2025). DOI: 10.1007/s11229-025-04969-w.

[64] Charlie Pilgrim, Adam Sanborn, Eugene Malthouse, and Thomas T. Hills. "Confirmation bias emerges from an approximation to Bayesian reasoning". In: *Cognition* 245 (2024), p. 105693. DOI: 10.1016/j.cognition.2023.105693.

[65] Florian Prinz, Thomas Schlange, and Khusru Asadullah. "Believe it or not: how much can we rely on published data on potential drug targets?" In: *Nature Reviews Drug Discovery* 9.10 (712 2011), pp. 328–329. DOI: 10.1038/nrd3439-c1.

[66] Mantas Radzvilas, William Peden, and Francesco De Pretis. "The Ambiguity Dilemma for Imprecise Bayesians". In: *British Journal for the Philosophy of Science* (2024). ahead of print. DOI: 10.1086/729618.

[67] Elena Rocca. "The judgements that evidence-based medicine adopts". In: *Journal of Evaluation in Clinical Practice* 24.5 (2018), pp. 1184–1190. DOI: 10.1111/jep.12994.

[68] Felipe Romero and Jan Sprenger. "Scientific Self-Correction: The Bayesian Way". In: *Synthese* 198 (2021), pp. 5803–5823. DOI: 10.1007/s11229-020-02697-x.

[69] Masahiro Ryo, Jonathan M. Jeschke, Matthias C. Rillig, and Tina Heger. "Machine learning with the hierarchy-of-hypotheses (HoH) approach discovers novel pattern in studies on biological invasions". In: *Research Synthesis Methods* 11.1 (2019), pp. 66–73. DOI: 10.1002/jrsm.1363.

[70] David Schum. "Classifying Forms and Combinations of Evidence: Necessary in a Science of Evidence". In: *Evidence, Inference and Enquiry*. Ed. by Mimi Vasilaki Phil Dawid William Twinning. OUP, 2011. Chap. 2, pp. 11–36.

[71] Gerhard Schurz. "The Hierarchy-of-Hypotheses Approach from a Philosophy of Science Perspective". In: *BioScience* 71 (4 2021), pp. 350–356. DOI: 10.1093/biosci/biaa097.

[72] Teddy Seidenfeld and Larry Wasserman. "Dilation for Sets of Probabilities". In: *The Annals of Statictics* 21.3 (1993), pp. 1139–1154. DOI: 10.1214/aos/1176349254.

[73] Prakash P Shenoy. "Conditional independence in valuation-based systems". In: *International Journal of Approximate Reasoning* 10.3 (1994), pp. 203–234. DOI: 10.1016/0888-613X(94)90001-9.

[74] Elliott Sober. "Independent Evidence about a Common Cause". In: *Philosophy of Science* 56.2 (1989), pp. 275–287. DOI: 10.1086/289487.

[75] Wolfgang Spohn. *The Laws of Belief. Ranking Theory and its Philosophical Applications*. Oxford: Oxford University Press, 2012.

[76] Jacob Stegenga. "Robustness, Discordance, and Relevance". In: *Philosophy of Science* 76.5 (2009), pp. 650–661. DOI: 10.1086/605819.

[77] Doohyun Sung and Bennett Holman. "Against Evidential Pluralism in Pharmaceutical Regulation". In: *Philosophy of Science* 90 (5 2023), pp. 1276–1285. DOI: 10.1017/psa.2023.40.

[78] Borut Trpin. "Against Methodological Gambling". In: *Erkenntnis* 88 (2021), pp. 907–927. DOI: 10.1007/s10670-021-00386-w.

[79] Aron Vallinder and Erik J. Olsson. “Trust and the value of overconfidence: a Bayesian perspective on social network communication”. In: *Synthese* 191.9 (2014), pp. 1991–2007. DOI: 10.1007/s11229-013-0375-0.

[80] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1991.

[81] Gregory Wheeler and Richard Scheines. “Coherence and Confirmation through Causation”. In: *Mind* 122.485 (2013), pp. 135–170. DOI: 10.1093/mind/fzt019.

[82] David H. Wolpert and David B. Kinney. “A Stochastic Model of Mathematics and Science”. In: *Foundations of Physics* 5.2 (2024). DOI: 10.1007/s10701-024-00755-9.

[83] John Worrall. “Evidence in Medicine and Evidence-Based Medicine”. In: *Philosophy Compass* 2.6 (2007), pp. 981–1022. DOI: 10.1111/j.1747-9991.2007.00106.x.