Journal for General Philosophy of Science
https://doi.org/10.1007/s10838-020-09508-4

ARTICLE

On the Assessed Strength of Agents' Bias

Jürgen Landes[1] · **Barbara Osimani**[1,2]

Abstract
Recent work in social epistemology has shown that, in certain situations, less communication leads to better outcomes for epistemic groups. In this paper, we show that, ceteris paribus, a Bayesian agent may believe less strongly that a single agent is biased than that an entire group of independent agents is biased. We explain this initially surprising result and show that it is in fact a consequence one may conceive on the basis of commonsense reasoning.

Keywords Social epistemology · Formal epistemology · Reliability · Bias · Bayesian networks · Conjunction fallacy

1 Introduction

Rational agents sometimes believe a conjunction more strongly than they believe every single literal in this conjunction. We show that this peculiar fact applies to Bayesian agents—in particular circumstances—and explain why.

In order to do so, we tackle the problem of how to assess a group of agents (e.g., scientists) providing testimony vis-à-vis a single agent (e.g., one scientist) providing testimony. Unlike previous works (e.g., Zollman 2013; Angere and Olsson 2017; Holman and Bruner 2015), which compared different communication structures of the same group of agents (N vs. N comparison), we here study how a group of agents compares to a single agent (N vs. 1 comparison).

Testimony consists of reports the agents provide based on their findings. The fallible agents considered here are either good inquirers; call them reliable; or not-so-good inquirers; call them biased. Intuitively, we are less likely to believe that a *group of N independent* agents each reporting a finding are all biased than we are to believe that one single agent providing these same N reports is biased, ceteris paribus. In other words: upon receiving the news, we assign a greater probability that at least one of the N independent agents is unbiased than we ascribe to the single agent being unbiased. We here show that this

✉ Jürgen Landes
juergen_landes@yahoo.de

[1] Munich Center for Mathematical Philosophy, LMU, Munich, Germany

[2] Università Politecnica delle Marche in Ancona, Ancona, Italy

Published online: 30 May 2020

intuitive probability judgement does not universally hold true (Theorems 1 and 2)[1] and explain why this is the case.

But why is it that we judge it more likely that, ceteris paribus, one single agent is biased than that a group of independent agents are all biased? Prior to obtaining evidence, the prior probability of a single agent being reliable is equal to some value, ρ say. The prior probability of the agent being unreliable (biased) is then $1 - \rho =: \bar{\rho}$. The ceteris paribus clause then entails that the probability of any one of N agents is biased with probability $\bar{\rho}$. The independence judgement then requires that the probability for all N agents being biased is $\bar{\rho}^N$. Clearly, $\bar{\rho} > \bar{\rho}^N$. The difference between $\bar{\rho}$ and $\bar{\rho}^N$ increases with growing N. As evidence accumulates, we have all reasons to believe that the posterior probabilities will continue to satisfy this inequality.

The probability functions considered here are those of a Bayesian agent receiving testimony from other agents (scientists). Since Bayesians agents are not prone to conjunction fallacies (holding that the probability of a conjunction is greater than the probability of a subset of conjuncts, see Tversky and Kahneman 1983) one may think that the lesson drawn from studying conjunction fallacies applies here.[2] However, we shall see that this lesson does not apply here and the intuitive answer is incorrect (Sect. 3.2).

The rest of this paper is organised as follows: next, we provide background and motivation for the area of research this paper contributes to (Sect. 2.1). Based on this exposition we introduce the formal model for our investigation (Sect. 2.2). Within the model we can formalise the Bayesian probability judgement we want to investigate (Sect. 3.1). We go on to derive (Sect. 3.2) and explain (Sects. 3.3 and 3.4) our main results and offer some conclusions regarding our immediate result and some wider implications (Sect. 4).

2 The Model

2.1 Background and Motivation

We consider a group of agents providing testimony for or against a hypothesis. We shall here not assume that we can fully rely on the reports provided by the agents, but instead we shall assess agents' reliability.

The Scandinavian School of Evidentiary Value conceived of unreliable agents as providing evidence which teaches us nothing about the hypothesis of interest, see further (Bovens and Hartmann 2003, 57) and Edman (1973), Hansson (1983), Schum (1988). In Bovens and Hartmann (2003), this notion of unreliability has been formalised in a Bayesian network model for determining the confirmation a body of evidence provided by a group of agents bestows on the hypothesis of interest. Their model has found applications in the philosophy of science concerning the epistemological Variety of Evidence Thesis (Bovens and Hartmann 2002; Claveau 2013; Claveau and Grenier 2019; Stegenga and Menon 2017; Landes 2020b, a), which states that varied evidence for a hypothesis confirms it more strongly than less varied evidence, ceteris paribus. Furthermore, it has been employed in Hahn et al. (2016) for modelling social debates of findings in climate science,

[1] More precisely, we show how this posterior probability judgement is inconsistent with particular prior probability judgements and Bayesian updating.

[2] It requires some serious effort to provide a Bayesian model of subjects committing conjunction fallacies (von Sydow 2011).

the philosophy of economics Casini and Landes (2020) and (the philosophy of) medicine (Abdin et al. 2019; Landes et al. 2018; De Pretis et al. 2019; 2020).

Crucial to this body of work is the *i*rrelevance of *u*nreliable *s*ources (Claveau 2013 calls this the IUS condition). But do we perceive of unreliable sources as providing no relevant information towards hypothesis confirmation? Collins et al. (2015; 2018) found that human subjects tend to favour the construal of unreliable sources put forward in Olsson (2011) over the approach of Bovens and Hartmann (2003), see also Merdes et al. (2021). In this approach, unreliable sources are construed as sources which always lie, i.e., the testimony of an unreliable agent is the exact opposite of what she thinks.

We are here interested in epistemic contexts in which fallible agents may be unreliable due to (possibly sub-conscious) biases.[3] We use sponsorship bias as our motivation which make agents' reports to be more likely to be in line with their sponsor's interest. Such a maximally strong bias is exhibited by agents who will always report findings in line with their sponsor's interest. Such agents are completely irrelevant for hypothesis confirmation since they provide no relevant information.

Agents which are biased to a non-maximal degree report findings with different probabilities than fully reliable agents; i.e., unbiased agents. We are here interested in biased agents that have a greater probability of reporting findings which *support* the hypothesis than unbiased agents and this probability is strictly less than one. That is, at times such agents do report findings which are not in their sponsor's interest. Reports from such agents *do provide* some information concerning the hypothesis. Reports supporting the hypothesis are (much) less confirmatory than reports from unbiased agents, whereas reports from biased agents conflicting with the hypothesis and thus with the sponsor's interest carry extra *dis*-confirmatory oomph.[4]

2.2 The Formal Model

We adopt the Bovens and Hartmann model by only changing their formalisation of unreliable agents. To the best of our knowledge, neither the Bovens and Hartmann model nor any of its derivatives have previously been employed to compare the posterior probabilities of unreliable sources. To keep this manuscript self-contained we now briefly describe the Bovens and Hartmann model (Sects. 2.2.1 and 2.2.2) and our adaptation (Sect. 2.2.3).

2.2.1 Variables

We employ a number of binary propositional variables: A variable *HYP* where the intended meaning for *Hyp* is that "the hypothesis is true" and for $\overline{Hyp}$ is that "the hypothesis is false". Next, we incorporate into our model that hypotheses may not always be directly tested, rather it is some of their observable consequences which are testable (Bovens and Hartmann 2003, 89). We employ consequence variables CON_n where Con_n ($\overline{Con_n}$) stands for the proposition that the n-th testable consequence of the hypothesis of interest holds (is false). Reported findings are modelled by means of a report variable *REP*. Reports pertain by definition to one consequence of the hypothesis only. *Rep* indicates that the consequence

[3] Reducing or even erasing biases one holds may be much harder than one thinks, see Kenyon (2014). This observation highlights the need to take into account such biases when assessing testimony.

[4] The term "oomph" is now a term of art mentioned by Stegenga and Menon (2017).

is reported to hold while $\overline{Rep}$ means that the consequence fails to hold is reported. Finally, every report is modulated by a single reliability variable *REL*, where *Rel* means that the reporting agent is assessed to be reliable and $\overline{Rel} = Bias$ stands for a biased agent. Report variables representing different reports originating from the same agent thus share their modulating reliability variable. Every agent is thus represented by a single variable formalising the agent's possible types: reliable or biased.

A Bayesian prior probability function, P, defined over the algebra generated by these variables, is selected. The choice of this probability function P is constrained by conditional independencies capturing the relation of variables, which are graphically represented in a Bayesian network.

2.2.2 Topology of Bayesian Networks

The topology of Bovens and Hartmann networks is generated by the following modelling choices regarding probabilistic independences and dependences.

These conditional independencies—denoted by $\perp$—are

$$\begin{gathered} HYP \perp REL_n \quad \text{for all } n \\ CON_i \perp REL_n \mid HYP \quad \text{for all } i, n \\ REP_i \perp HYP \mid REL_{n_i}, CON_{m_i} \quad \text{for all } i \\ \{CON_{m_i}, REL_{n_i} REP_i\} \perp \bigcup_{k \neq m_i} \bigcup_{j \neq n_i} \{CON_k, REL_j, REP_k\} \mid HYP \\ REP_i \perp X \mid REL_{n_i}, CON_{m_i} \quad \text{for all } i \text{ and all } X \notin \{REL_{n_i}, CON_{m_i}\}, \end{gathered}$$

where n_i is the reliability variable pertaining to REP_i and m_i the pertinent consequence variable.

The probability of whether a testable consequence is true or false is directly influenced by whether the hypothesis of interest is true or false. Similarly, the probabilities of reports that a testable consequence is reported depends on whether the relevant testable consequence of the hypothesis is true and on the reliability of the reporting agent. This motivates the edges and their orientations in such Bayesian networks; example topologies can be found in Figure 2.

2.2.3 Prior and Conditional Probabilities

The initial assessment of the hypothesis is expressed as the probability $0 < P(Hyp) < 1$. By initial we mean prior to receiving testimony. The initial assessment of an agent's reliability is captured by $0 < P(Rel) =: \rho = 1 - P(Bias) < 1$.

Consequences of the hypothesis are construed as being probabilistically entailed by the hypothesis, that is *Con* is more likely under *Hyp* than under its negation, $\overline{Hyp}$. Mathematically speaking:[5]

[5] Those interested in confirming the hypothesis of interest directly without the intermediate consequence variables may simply put $0 = P(Con|\overline{Hyp}) < P(Con|Hyp) = 1$ and thus "erase" the consequence variables from considerations. None of our results hinge on the inclusion/exclusion of consequence variables.

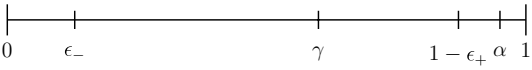

Fig. 1 Example parameter configuration for providing a positive report

$$0 < P(Con|\overline{Hyp}) < P(Con|Hyp) < 1 \quad \text{for all consequence variables } CON.$$

So far, we have been following Bovens and Hartmann (2003) from which we shall now deviate. The difference in models is explained by the different construals of unreliable (biased) agents (see Sect. 2.1) which give rise to a different formalisation.

We here consider fallible reliable agents, i.e., agents who sometimes fail to report the truth. $0 < \epsilon_+ < 1$ is a reliable agent's probability of reporting a false negative (reporting that the consequence is false while it is in fact true) and $0 < \epsilon_- < 1$ is a reliable agent's probability of reporting a false positive (reporting that the consequence is true while it is in fact false):[6,7]

$$P(\text{False Negative, reliable agent}) = P(\overline{Rep}|Con, Rel) = \epsilon_+$$
$$P(\text{False Positive, reliable agent}) = P(Rep|\overline{Con}, Rel) = \epsilon_-.$$

Intuitively, the more often an agent's testimony matches the true state of the world (truth value of *CON*) the greater an agent's competence. So, the smaller ϵ_+, ϵ_- the better the evidence an agent's testimony provides.

Agents biased in the above discussed sense are more likely to report findings supporting the hypothesis than reliable agents. That is, the probability that an agent assessed to be biased provides a report that a consequence has been observed is greater than the probability that an agent assessed to be reliable provides such a report.

In case the pertinent consequence is true, this means that

$$1-\epsilon_+ = P\underbrace{(Rep|Con, Rel)}_{\text{True Positive, reliable agent}} < P\underbrace{(Rep|Con, Bias)}_{\text{True Positive, unreliable agent}} =: \alpha.$$

In case the pertinent consequence is false, this means that

$$\epsilon_- = P\underbrace{(Rep|\overline{Con}, Rel)}_{\text{False Positive, reliable agent}} < P\underbrace{(Rep|\overline{Con}, Bias)}_{\text{False Positive, unreliable agent}} =: \gamma.$$

We are here interested only in fallible agents and thus agents assessed to be biased commit errors of both types. Hence, neither α nor γ can be equal to one. A possible configuration of parameters is shown in Figure 1, an overview is given in Table 1.[8]

There are two types of agents in our model, reliable ones characterised by ϵ_+, ϵ_- and biased agents represented by α, γ and one is unsure about each agent's type ($P(Rel)$). It poses no conceptual difficulty to model a situation in which agents may have multiple

[6] See Osimani and Landes (2020) for more motivation and background on our way of modelling unreliable agents.

[7] To streamline the exposition we suppress indices indicating the particular agent.

[8] $0 < \gamma < \epsilon_- < \alpha < 1 - \epsilon_+ < 1$ represents a biased agent more likely to report findings dis-confirming the consequence. $\alpha = P(Rep|Con, \overline{Rel}) = P(Rep|\overline{Con}, \overline{Rel}) = \gamma$ defines unreliable agents in the (Bovens and Hartmann 2003) sense.

Table 1 Overview of employed variables, their intended interpretation and (conditional) probabilities. To increase readability, we use ¬ to denote negation in this table

Variable	Intended interpretation	(Conditional) probabilities
HYP	Hypothesis of interest	$0 < P(Hyp) < 1$
CON	Testable consequence	$0 < P(Con\|\neg Hyp) < P(Con\|Hyp) < 1$
REL	Reliability of instrument	$0 < P(Rel) = \rho < 1$
REP	Report	See below
Rel	Reliable instrument	$0 < P(Rep\|Con, Rel) = 1 - \epsilon_+ < 1$ $0 < P(Rep\|\neg Con, Rel) = \epsilon_- < 1$
$\overline{Rel}$	Unreliable instrument	$1 > P(Rep\|Con, \neg Rel) = \alpha > 1 - \epsilon_+$ $1 > P(Rep\|\neg Con, \neg Rel) = \gamma > \epsilon_-$

types of bias and one is unsure about the type of bias a particular agent possesses. Technically, this is achieved by using variables *REL* of greater arity, adopting a prior over these greater arity variables and formalising different types and/or strengths of bias (Olsson 2005, Sect. 4.3).

Reports (even those from the same agent) are here taken to be independent from each other given the true state of the world and given the type(s) of the agent(s) the reports are obtained from. More precisely, the probability of a report stating that a consequence of the hypothesis holds (or fails) only depends on the reporting agent and the truth value of the consequence. This models a situation in which different reports are, for example, generated by independent random tosses of the same coin or by identically sampling from the same population. A report variable hence has only two parents (a reliability variable and a consequence variable) and no children.

All these assumptions are substantial assumptions and none of them will always hold in every situation. We do not want to make the case that our assumptions are appropriate in a wide range of situations. All we rely on is that there are *some* situations in which our assumptions are reasonable.

3 Analysis

3.1 Formalising the Probability Judgement

We can now return to asking the question raised in the introduction: "Ceteris paribus, do we always believe more strongly that a single agent is biased than we believe that an entire group of independent agents is biased?" As we argued in Section 1, the intuitive answer is affirmative. Before we can proceed to thoroughly answer this question we need to do two things.

First, we need to specify the evidence reports, the network structure of the reports and how the reports pertain to (the testable consequences of) the hypothesis of interest. In short, we have to specify the topology of Bayesian networks for our application. Bovens and Hartmann consider three scenarios, each scenario consists of two distinct set-ups (i.e.,

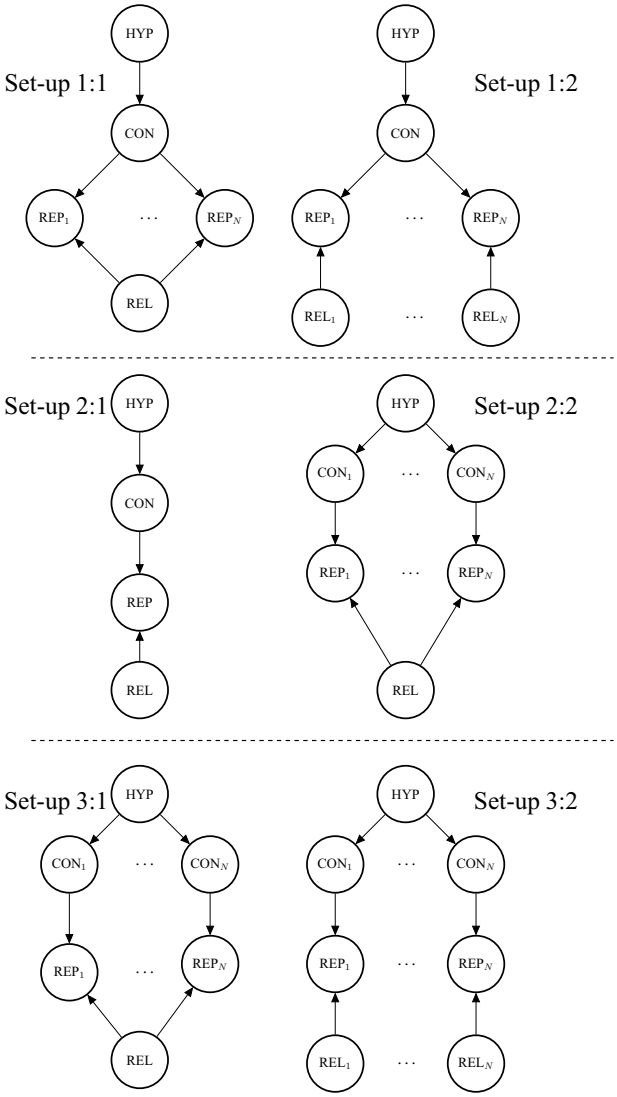

Fig. 2 The three scenarios described in Bovens and Hartmann (2003). Set-up 2:2 is the same as Set-up 3:1

network topologies). We here only discuss Scenario 1 and Scenario 3.[9] In the first set-up, one single agent provides all reports; in the second set-up, N agents each provide one report. See Figure 2—taken from Osimani and Landes (2020) for a graphical illustration—in which the first set-up is always pictured on the left and the second set-up on the right; dashed lines demarcate the three different scenarios. In the situations depicted on the left,

[9] In Scenario 2, a single agent provides all reports in both set-ups. Rather than presenting a problem for social epistemology it teaches us that more "confirmatory evidence" does not always lead to more confirmation (see Carnap 1962, 382), which is outside our current scope of interest.

a single agent provides all the reports. Since in this situation the reports are obtained from a single agent, we use one single variable to model the (un-)reliability of this source. In the situations depicted on the right, every report is obtained from a different agent. Consequently, we use a different reliability variable for every agent to capture the (un-)reliability of all the different agents.

Second, we obviously need to make sure that conditional probabilities in both compared set-ups are, ceteris paribus, the same. So, we impose the condition that the probabilities defined in Section 2.2.3 are the same for all agents. Furthermore, we assume that for all n the n-th report in both set-ups shows the same result. Finally, we require that all consequence variables are assigned the same conditional probabilities. Mathematically, this just means that we are now not abusing notation any more when dropping a great number of indices.

The probability function for the first set-up is denoted by P_1, the function for the second set-up by P_2. The bodies of evidence are respectively denoted by $\mathcal{E}_1$ and $\mathcal{E}_2$. Finally, we can formalise our probability judgement: "Ceteris paribus, we believe more strongly that a single agent is biased than we believe that an entire group of independent agents is biased" by

$$P_1(Bias|\mathcal{E}_1) > P_2\left(\bigwedge_{n=1}^{N} Bias_n|\mathcal{E}_2\right). \tag{1}$$

3.2 Results

We now state our main result:

Theorem 1 *In Scenario 1 and Scenario 3 for all* $0 < P(Hyp), P(Con|Hyp) < 1$, *if the following three conditions all hold*

$$\begin{aligned}
\frac{P(False\ Negative,\ reliable\ agent)}{P(False\ Negative,\ biased\ agent)} &= \frac{P(\overline{Rep}|ConRel)}{P(\overline{Rep}|ConBias)} = \frac{\epsilon_+}{1-\alpha} \geq 4(2^{N-1}-1) \\
\frac{P(True\ Negative,\ reliable\ agent)}{P(True\ Negative,\ biased\ agent)} &= \frac{P(\overline{Rep}|\overline{Con}Rel)}{P(\overline{Rep}|\overline{Con}Bias)} = \frac{1-\epsilon_-}{1-\gamma} \geq 4(2^{N-1}-1) \\
P(Rel) = \rho &\leq \frac{1}{1+\sqrt[N-1]{2}},
\end{aligned} \tag{2}$$

then it holds that

$$P_2\left(\bigwedge_{n=1}^{N} Bias_n|\bigwedge_{n=1}^{N}\overline{Rep_n}\right) > P_1\left(Bias|\bigwedge_{n=1}^{N}\overline{Rep_n}\right).$$

Proof All proofs can be found in the Appendix.

The answer to our question is thus *no*. For all probability assignments satisfying (2), we believe more strongly that the entire group of agents is biased than we believe that the single agent is biased, if all reports state that the pertinent consequence of the

hypothesis has not been observed. For example, for $2 \leq N \leq 5$ all probability assignments with $\epsilon_- \leq 10\%, \epsilon_+ \geq 10\%, \alpha \geq 99.9\%, \gamma \geq 99.1\%, \rho \leq 33\%$ satisfy (2).

Note that this result is untroubled by conjunction fallacies since we compare *two different* probability functions. Instead, holding that $\mathbf{P_2}(\bigwedge_{\mathbf{n=1}}^{\mathbf{N}} \mathbf{Bias_n} | \bigwedge_{n=1}^{N} \overline{Rep_n}) > \mathbf{P_2}(\mathbf{Bias_1} | \bigwedge_{n=1}^{N} \overline{Rep_n})$ would be committing a conjunction fallacy.

Since there is a canonical morphism induced by switching the truth-values of binary propositional variables, one may wonder whether there is a similar such phenomenon for reliability instead of bias. Indeed, there is

Theorem 2 *In Scenario 1 and Scenario 3, if*

$$\begin{aligned} \frac{P(True\ Positive,\ reliable\ agent)}{P(True\ Positive,\ biased\ agent)} &= \frac{P(Rep|ConRel)}{P(Rep|ConBias)} = \frac{1-\epsilon_+}{\alpha} \leq \frac{1}{4(2^{N-1}-1)} \\ \frac{P(False\ Positive,\ reliable\ agent)}{P(ccFalse\ Positive,\ biased\ agent)} &= \frac{P(Rep|\overline{Con}Rel)}{P(Rep|\overline{Con}Bias)} = \frac{\epsilon_-}{\gamma} \leq \frac{1}{4(2^{N-1}-1)} \\ P(Bias) = \bar{\rho} &\leq \frac{1}{1+\sqrt[N-1]{2}}, \end{aligned} \tag{3}$$

then for all $0 < P(Hyp), P(Con|Hyp) < 1$ *it holds that*

$$P_2\left(\bigwedge_{n=1}^{N} Rel_n | \bigwedge_{n=1}^{N} Rep_n\right) > P_1\left(Rel | \bigwedge_{n=1}^{N} Rep_n\right).$$

Having derived these results in our model we are next interpreting them in the setting we described. The obtained results also apply to other settings our model adequately represents. We discuss different types of biases which our model may adequately represent in Section 4.

3.3 A More Intuitive Picture

Having obtained the formal results we now know for which cases the probability of a conjunction behaves in an unexpected way. Based on this knowledge we paint a more intuitive picture of our results.

Consider a situation in which a person you believe to be unreliable tells you something you did not expect to hear. For example, the chief scientist of a pharmaceutical company publicly states that a drug they currently sell and recently researched is less effective than previously believed. Based on this information you believe more strongly that the agent is in fact reliable. Next suppose that there are a number (N, say) of chief scientists and each scientist tells you about the drug they have been exclusively selling and recently researching that their drugs are less effective than previously believed. What do you now think about the group of scientists? Your belief in their individual reliabilities has increased. This means that your belief in their individual *un*reliabilities has decreased. Supposing that there is no connection between the different companies, scientists and drugs your belief in all of them being unreliable decreases proportionally to the number of reports.

Now suppose instead that there is a single chief scientist working for a pharmaceutical company who tells you that a number of drugs (N) sold by her company which have all

recently been researched are less effective than previously thought. Let us make the picture more concrete by assuming that there is no connection between the different drugs (different research labs studying them, targeted at different diseases). To ease the comparison between this and the above set-up, we assume that the content of report i is the same in this and the above set-up. Furthermore, we suppose that all reports are all equally (un-)likely.

What do you now believe about the reliability of the single scientist? Clearly, your belief in her reliability is increasing. The increase in belief in her reliability is the stronger the more you initially believed the agent to be unreliable. Furthermore, the less likely you initially believed to hear such testimony from a biased agent, the stronger the reversal of the standing of the scientists in your eyes. Note that the reports from a single person have a cumulative effect on the assessed reliability. The situation resembles the accumulation of compound interest, the increase in the assessed reliability (interest) sky-rockets.

But since an increase in the assessed reliability means a decrease in the assessed *un*reliability, the latter plunges very quickly indeed. It is then conceivable that, ceteris paribus, in certain cases it is the case that you believe less strongly that the single agent is biased than you believe that all scientists in the group are biased.

We next discuss the parameter values for which this unexpected behaviour of the probability of a conjunction obtains.

3.4 Explanation of Results

Since these two results are natural duals of each other, we shall only discuss Theorem 1. Why is it that one believes more strongly that the entire group is biased than that the single agent is biased? We can explain this by looking at the parameter values for which this happens.[10]

We develop a deeper understanding of the first two conditions in (2) by re-writing them as

$$\frac{\epsilon_+}{1-\alpha} = \frac{P(\overline{Rep}|Con\,Rel)}{P(\overline{Rep}|Co\,n\overline{Rel})} \geq 2^{N+1} - 4 \leq \frac{P(\overline{Rep}|\overline{Con}\,Rel)}{P(\overline{Rep}|\overline{Con}\,\overline{Rel})} = \frac{1-\epsilon_-}{1-\gamma}. \tag{4}$$

This means that biased agents are strongly biased, $\alpha \gg 1-\epsilon_+$ and $\gamma \gg \epsilon_-$.

Holding the truth value of the *CON* variable fixed, we see that quotients on the left and on the right describe ratios of the likelihood of the reported findings. The literature on Bayesian statistics refers to these ratios as *Bayes factors*; which are—in this literature—considered to be *the* measure of the strength of evidence. Translated to our setting, this means that the received reports are strong evidence against the hypothesis that agents are biased, for large N. For $N = 2$, the Bayes factors are only required to be greater or equal than four; a Bayes factor equal to three is conventionally interpreted as relatively weak evidence for a hypothesis.

The third condition, $P(Rel) := \rho \leq \frac{1}{1+\sqrt[N-1]{2}} < 0.5$, says that, a priori, agents are assessed to more likely be biased than reliable. See Figures 3, 4 and 5 for illustrations of the parameter spaces in which this inequality and (4) hold.

[10] One might wonder why the condition that $P(Con|Hyp) > P(Con|\overline{Hyp})$ is not used in the proof. It seems like this condition is hence not required. Note however that we made implicit use of this condition when we decided upon our formalisation of a biased agent. Hence, in situations in which $P(Con|Hyp) < P(Con|\overline{Hyp})$ holds, our theorems continue to hold. In such situations, reporting that a consequence was observed *dis*confirms the hypothesis of interest.

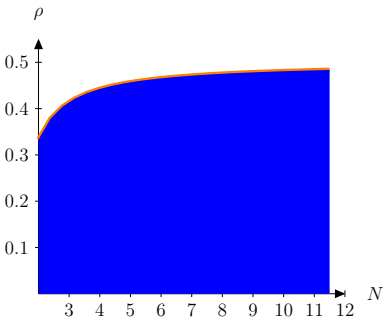

Fig. 3 The orange curve is a plot of $\rho = \frac{1}{1+\sqrt[N-1]{2}}$ in the $N-\rho$-plane. With increasing N the curve converges to $\rho = 0.5$. $\rho < \frac{1}{1+\sqrt[N-1]{2}}$ holds in the blue area, the size of the blue area increases with increasing N. To increase readability N is displayed as a continuous variable although it is discrete in the current setting. Our counter-intuitive results obtain in the blue area, if $\epsilon_+, \alpha, \epsilon_-, \gamma$ take suitable values, too. (Color figure online)

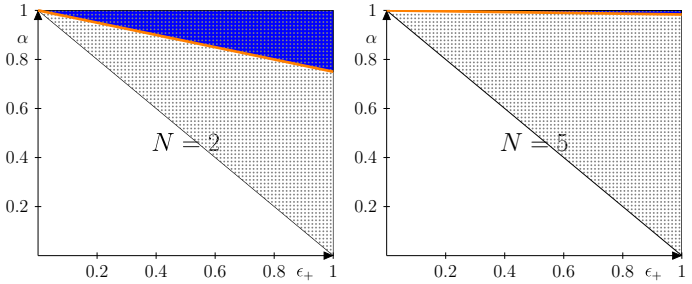

Fig. 4 The orange curve is a plot of $\alpha = 1 - \frac{\epsilon_+}{4(2^{N-1}-1)}$ in the $\epsilon_+ - \alpha$-plane. α is strictly greater than this value in the blue area where our assumption of $\alpha > 1 - \epsilon_+$ (dotted area) also holds. The number of agents N is equal to 2 in the left and and equal to 5 in the right plot. With increasing N the size of the blue area decreases quickly. Our counter-intuitive results obtain in the blue area, if ϵ_-, γ, ρ take suitable values, too. (Color figure online)

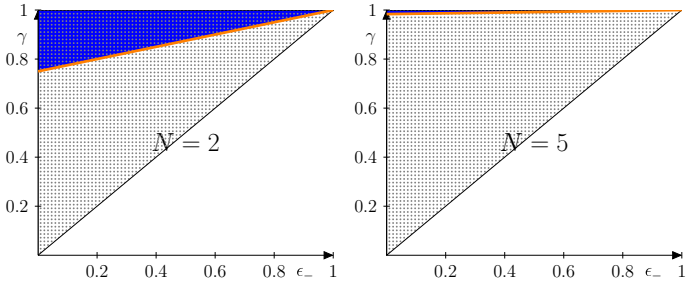

Fig. 5 The orange curve is a plot of $\gamma = \frac{\epsilon_-}{4(2^{N-1}-1)} + 1 - \frac{1}{4(2^{N-1}-1)}$ in the $\epsilon_- - \gamma$-plane. γ is greater than this value in the blue area where our assumption of $\gamma > \epsilon_-$ (dotted area) also holds. The number of agents N is equal to 2 in the left and and equal to 5 in the right plot. With increasing N the size of the blue area decreases quickly. Our counter-intuitive results obtain in the blue area, if ϵ_+, α, ρ take suitable values, too. (Color figure online)

So, upon receiving multiple reports dis-confirming the hypothesis from a single agent, the assessed reliability of this agent sky-rockets. In turn, the assessed bias of this agent falls through the floor. The stronger the assessed bias, the closer α, γ are to one (the closer $1-\alpha, 1-\gamma$ are to zero), the larger the "Bayes factors" in (4), the less likely one thinks that one agent consistently reports contrary to her bias. Hence, the stronger this effect.

Furthermore, the smaller the prior probability of agents being reliable, i.e., the smaller ρ (the greater $\bar{\rho}$), the more relevant the above considerations become. Hence, the stronger the effect.

Instead, if these findings are reported by a group of agents where every agent only makes one single report, then the assessed bias of every single agent decreases only somewhat. The assessed bias of the entire group hence also falls—but only moderately so.

For large enough Bayes factors, the drop in the assessed bias of the single agent outpaces the decrease of the assessed bias of the entire group of agents.

3.5 Further Observations

We also want to point out that the condition of *binary* report variables is unnecessarily restrictive. All results immediately generalise to report variables of finite arity, as long as the values of the received report variables satisfy the conditions in (4).

Observe that Theorems 1 and 2 apply to all $\alpha, \gamma, \epsilon_+, \epsilon_- \in (0,1)$ which satisfy (3). In particular, there is no constraint which couples α and γ, nor is there a constraint which couples ϵ_+ and ϵ_-. Hence, these theorems hold also for all $\alpha, \gamma, \epsilon_+, \epsilon_- \in (0,1)$ which satisfy (3), if $\alpha = \gamma, 1-\epsilon_+ = \epsilon_-$ or ($\alpha = \gamma$ and $1-\epsilon_+ = \epsilon_-$) hold. In case $\alpha = \gamma$ a biased agent is an unreliable agent in the Bovens and Hartmann-sense, in case $1-\epsilon_+ = \epsilon_-$ an agent assessed to be reliable in our setting is an unreliable agent in the Bovens and Hartmann-sense.

Furthermore, Theorems 1 and 2 also apply to incompetent agents with $1-\epsilon_+ < \epsilon_-$ and/or $\alpha < \gamma$. Faced with reports from such incompetent agents, one better believe the opposite of the reported findings. One hence perceives such agents as liars in the sense of Olsson (2011).

Finally, we observe that Theorems 1 and 2 do not distinguish between Scenario 1 and Scenario 3: that is, the constraints on the probability assessments are the same for Scenario 1 and Scenario 3. This observation should calm all remaining worries that somehow the *consequences* of the hypothesis of interest do the heavy lifting here; they do not. This contrasts the results in Bovens and Hartmann (2003) and Osimani and Landes (2020) for hypothesis confirmation, which distinguish between Scenario 1 and Scenario 3.

Finally, we remark that we obtain the counter-intuitive result for all group sizes, $N \geq 2$.

4 Conclusions

Recent work in social epistemology on the topology of group communications has brought the unexpected finding that sometimes epistemic groups fare better when agents (can) only communicate with few of their peers, see Zollman (2013) for an overview and Angere and Olsson (2017) for a recent point in case. Although, these results may depend on the

epistemic group being comprised of honest truth-seeking agents as argued by Holman and Bruner (2015) and on particular parameters Rosenstock et al. (2017), this part of the message is loud and clear: *Sometimes less is more in social epistemology*. This paper replicates this message for N versus 1 comparisons.

We draw two further immediate conclusions: intuitions in multi-agent settings must be appreciated with due care and formal modelling can help us discover interesting belief dynamics between epistemic notions (such as reliability, bias, group size, strength of evidence) that we are very unlikely to have discovered by any other means.

Let us for the moment switch point of view and take the perspective of the group of agents providing testimony. From here, it appears less than ideal that the entire group is perceived more strongly to be biased than a single agent. Group members may feel that the posterior probability assignment $P_1(Bias|\mathcal{E}_1) < P_2(\bigwedge_{n=1}^{N} Bias_n|\mathcal{E}_2)$ constitutes an epistemic injustice (Fricker 2007) caused by overly negative prior assessments ($\alpha, \gamma, \bar{\rho}$ large).[11] One wonders, given that all the agents have done is to report *contrary* to the perceived bias, is there nothing the group can do to overcome this unfortunate state of affairs? The short answer is: no. There is nothing to be done. Once the prior probabilities are set, Bayesian updating kicks in and finishes the job.

This means that the only road to salvage the standing of the group of agents is a more favourable assessment prior to reporting. This can be achieved by either a more favourable assessment of the strength of bias (smaller α, γ) or by a more favourable assessment of being reliable (greater ρ). This then demonstrates the importance of appearances and the value of a good public relations section as well as the importance of the choice of the prior probability function in Bayesian epistemology.

We also want to point out that the employed Bayesian network models are rather versatile having found applications in judgement aggregation, varied evidence reasoning and social epistemology. Future applications await exploration. Further future work may also address inequality (1) with different notions of (un-)reliability in mind, variables of greater arity, and/or bodies of evidence containing conflicting reports. Another interesting avenue are more complicated topologies of the Bayesian network with fewer independencies (more edges), see Claveau and Grenier (2019) and Landes (2020a).

We also remark that while sponsorship bias provided the motivation for our model of a biased agent (in terms of $1 - \epsilon_+ < \alpha$ and $\epsilon_- < \gamma$), our analysis applies to all other biases (or other cognitive states) which make false positives more likely and false negatives less likely. Furthermore, in case false negatives are less likely and false positives are more likely ($1 - \epsilon_+ > \alpha$ and $\epsilon_- > \gamma$), our analysis continues to apply after employing the canonical morphism permuting α and $1 - \epsilon_+$ as well as γ and ϵ_-. Since the list of biases is rather large (Bero and Grundy 2016; Hahn and Harris 2014) the analysis presented here may prove relevant for a variety of strands of research.

Finally, our analysis was motivated by considering agents which were either biased or reliable; agents hence had one of two possible types. The formal analysis presented here is, of course, blind to the motivation of the model. Our analysis is hence relevant to all other scenarios in which there is uncertainty about agents' types. Other instances of dichotomous types are right-wing versus left-wing, hawks versus doves (foreign policy), predator versus scavenger, authoritarianist versus anarchist and theist versus atheist.

[11] Many thanks to an anonymous reviewer for pointing out this connection to the literature on epistemic injustice. Spelling out this connection in detail would, in our view, take us too far away from the points we want to make here.

Acknowledgements Open Access funding provided by Projekt DEAL. Barbara Osimani is the PI of the European Research Council-founded project PhilPharm and gratefully acknowledges being fully funded by the project. Jürgen Landes gratefully acknowledges funding from the European Research Council ('PhilPharm' grant 639276) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—405961989 and 432308570. We also want to thank Lorenzo Casini and Stephan Hartmann for many helpful discussions and comments. Many thanks also to anonymous reviewers and the editor of this journal who helped us to improve the paper.

Appendix

We first prove a technical lemma. We need to introduce a little more notation. For variables, e.g., REP, we use Rep^1 to denote Rep and Rep^0 to denote $\overline{Rep}$. For all n we use $r_n \in \{0,1\}$ to denote the value of the n-th evidence report. Recall that r_n does not depend on the set-up, due to our above conventions. Note that we do not require that the r_n are equal.

Lemma 1 *In the first scenario we have*

$$\begin{aligned}
&\operatorname{sign}\left(P_2\left(\bigwedge_{n=1}^{N} Bias_n|\mathcal{E}_2\right) - P_1(Bias|\mathcal{E}_1)\right)\\
&\quad= \operatorname{sign}\left(\sum_{HYP} P(HYP)\sum_{CON} P(CON|HYP)\cdot\left(\bar{\rho}^{N-1}\rho\prod_{n=1}^{N} P(Rep^{r_n}|CON\,Rel)\right.\right.\\
&\qquad\left.\left.- \sum_{\substack{i_1,\dots,i_N\in\{0,1\}^N\\ i_1+i_2+\cdots+i_N\geq 1}}\prod_{n=1}^{N} P(Rel^{i_n})P(Rep^{r_n}|CON\,Rel^{i_n})\right)\right).
\end{aligned}$$

In the third scenario we have

$$\begin{aligned}
&\operatorname{sign}\left(P_2\left(\bigwedge_{n=1}^{N} Bias_n|\mathcal{E}_2\right) - P_1(Bias|\mathcal{E}_1)\right)\\
&\quad= \operatorname{sign}\left(\sum_{HYP} P(HYP)\sum_{c=1}^{N}\sum_{CON_c}\prod_{j=1}^{N} P(CON_j|HYP)\cdot\right.\\
&\left.\left[\bar{\rho}^{N-1}\rho P(Rep^{r_j}|CON_j\,Rel) - \sum_{\substack{i_1,\dots,i_N\in\{0,1\}\\ i_1+\cdots+i_N\geq 1}} P(Rep^{r_j}|CON_j\,Rel_j^{i_j})P(Rel_j^{i_j})\right]\right).
\end{aligned}$$

Proof For Scenario 1 we find

$$
\begin{aligned}
P_2\left(\bigwedge_{n=1}^{N} Bias_n|\mathcal{E}_2\right) - P_1(Bias|\mathcal{E}_1) &= \frac{P_2\left(\bigwedge_{n=1}^{N} Bias_n\,\mathcal{E}_2\right)}{P_2(\mathcal{E}_2)} - \frac{P_1(Bias\,\mathcal{E}_1)}{P_1(\mathcal{E}_1)} \\
&= \frac{\sum_{HYP}\sum_{CON} P_2\left(\bigwedge_{n=1}^{N} Bias_n\,\mathcal{E}_2\,HYP\,CON\right)}{\sum_{HYP}\sum_{CON}\sum_{n=1}^{N}\sum_{REL_n} P_2(\mathcal{E}_2\,HYP\,CON\,REL_1\ldots REL_N)} \\
&\quad - \frac{\sum_{HYP}\sum_{CON} P_1(Bias\,\mathcal{E}_1\,HYP\,CON)}{\sum_{HYP}\sum_{CON}\sum_{REL} P_1(\mathcal{E}_1\,HYP\,CON\,REL)} \\
&= \frac{P(Bias)^N\cdot\sum_{HYP}\sum_{CON}}{\sum_{HYP}\sum_{CON}\sum_{n=1}^{N}\sum_{REL_n}} \\
&\quad \frac{\prod_{n=1}^{N} P(HYP)P(CON|HYP)P(Rep^{r_n}|CON\,Bias)}{P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}|CON\,REL_n)P(REL_n)} \\
&\quad - \frac{P(Bias)\cdot\sum_{HYP}\sum_{CON}}{\sum_{HYP}\sum_{CON}\sum_{REL}} \\
&\quad \cdot \frac{P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}|CON\,Bias)}{P(REL)P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}|CON\,REL)}.
\end{aligned}
$$

The sign of this expression is equal to the sign of

$$
\begin{aligned}
&\frac{P(Bias)^N\cdot\sum_{HYP}\sum_{CON}}{P(Bias)\cdot\sum_{HYP}\sum_{CON}} \\
&\quad \cdot \frac{P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}|CON\,Bias)}{P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}|CON\,Bias)} \\
&\quad - \frac{\sum_{HYP}\sum_{CON}\sum_{n=1}^{N}\sum_{REL_n}}{\sum_{HYP}\sum_{CON}\sum_{REL}} \\
&\quad \cdot \frac{P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}|CON\,REL_n)P(REL_n)}{P(REL)P(HYP)P(CON|HYP)\prod_{n=1}^{N} P(Rep^{r_n}CON\,REL)}.
\end{aligned}
$$

Since the first term is equal to $\bar{\rho}^{N-1}$, the sign of this expression is equal to

$$\sum_{HYP}\sum_{CON} P(HYP)P(CON|HYP)\cdot$$
$$\left(\bar{\rho}^{N-1}\sum_{REL} P(REL)\prod_{n=1}^{N} P(Rep^{r_n}|CON\ REL)\right.$$
$$\left.-\sum_{n=1}^{N}\sum_{REL_n}\prod_{n=1}^{N} P(REL_n)P(Rep^{r_n}|CON\ REL_n)\right)$$
$$=\sum_{HYP}\sum_{CON} P(HYP)P(CON|HYP)\cdot\left(\bar{\rho}^{N-1}\rho\prod_{n=1}^{N} P(Rep^{r_n}|CON\ Rel)\right.$$
$$\left.-\sum_{\substack{i_1,\ldots,i_N\in\{0,1\}^N\\ i_1+i_2+\cdots+i_N\geq 1}}\prod_{n=1}^{N} P(Rel^{i_n})P(Rep^{r_n}|CON\ Rel^{i_n})\right).$$

For Scenario 3 we have

$$P_2\left(\bigwedge_{n=1}^{N} Bias_n|\mathcal{E}_2\right)-P_1(Bias|\mathcal{E}_1)=\frac{P_2\left(\bigwedge_{n=1}^{N} Bias_n\,\mathcal{E}_2\right)}{P_2(\mathcal{E}_2)}-\frac{P_1(Bias\,\mathcal{E}_1)}{P_1(\mathcal{E}_1)}$$
$$=\frac{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}}{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}\sum_{n=1}^{N}\sum_{REL_n}}$$
$$\frac{P_2\left(\bigwedge_{n=1}^{N} Bias_n\,\mathcal{E}_2\,HYP\,CON_1\ \ldots\ CON_N\right)}{P_2(\mathcal{E}_2\,HYP\,CON_1\ \ldots\ CON_N\,REL_1\ldots REL_N)}$$
$$-\frac{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c} P_1(Bias\,\mathcal{E}_1\,HYP\,CON_1\ \ldots\ CON_N)}{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}\sum_{REL} P_1(\mathcal{E}_1\,HYP\,CON_1\ \ldots\ CON_N\,REL)}$$
$$=\frac{P(Bias)^N\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}}{\sum_{HYP}\sum_{c=1}^{N}\sum_{n=1}^{N}\sum_{CON_c}\sum_{REL_n}}$$
$$\cdot\frac{P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,Bias)}{P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,REL_n)P(REL_j)}$$
$$-\frac{P(Bias)\cdot\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}}{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}\sum_{REL}}$$
$$\cdot\frac{P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,Bias)}{P(REL)P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,REL)}.$$

Since we are only interested in the sign of this equation we consider

$$
\begin{aligned}
&\frac{P(Bias)^N \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} P(HYP)}{P(Bias) \cdot \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} P(HYP)} \\
&\quad \frac{\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, Bias)}{\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, Bias)} \\
&\quad - \frac{\sum_{HYP} \sum_{c=1}^{N} \sum_{n=1}^{N} \sum_{CON_c} \sum_{REL_n}}{\sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} \sum_{REL}} \\
&\quad \cdot \frac{P(HYP) \prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, REL_n)P(REL_j)}{P(REL)P(HYP) \prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, REL_j)} \\
&= \bar{\rho}^{N-1} - \frac{\sum_{HYP} \sum_{c=1}^{N} \sum_{n=1}^{N} \sum_{CON_c} \sum_{REL_n}}{\sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} \sum_{REL}} \\
&\quad \cdot \frac{P(HYP) \prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, REL_n)P(REL_j)}{P(REL)P(HYP) \prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, REL_j)}.
\end{aligned}
$$

This is equal to the sign of

$$
\begin{aligned}
&- \bar{\rho}^{N-1} \rho \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c}^{N} P(HYP) \prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, Rel) \\
&\quad - \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} \sum_{\substack{i_1, \ldots, i_N \in \{0,1\} \\ i_1 + \cdots + i_N \geq 1}} P(HYP) \\
&\quad \cdot \prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\, Rel_j^{i_j})P(Rel_j^{i_j}) \\
&= \sum_{HYP} P(HYP) \sum_{c=1}^{N} \sum_{CON_c} \prod_{j=1}^{N} P(CON_j|HYP) \cdot \\
&\quad [\bar{\rho}^{N-1} \rho P(Rep^{r_j}|CON_j\, Rel) - \sum_{\substack{i_1, \ldots, i_N \in \{0,1\} \\ i_1 + \cdots + i_N \geq 1}} P(Rep^{r_j}|CON_j\, Rel_j^{i_j})P(Rel_j^{i_j})].
\end{aligned} \tag{5}
$$

□

Theorem 1 *In Scenario 1 and Scenario 3 for all* $0 < P(Hyp), P(Con|Hyp) < 1$, *if*

$$
\begin{aligned}
P(\overline{Rep}|ConRel) &= \epsilon_+ \geq 4(2^{N-1} - 1)(1 - \alpha) = 4(2^{N-1} - 1)P(\overline{Rep}|ConBias) \\
P(\overline{Rep}|\overline{Con}Rel) &= 1 - \epsilon_- \geq 4(2^{N-1} - 1)(1 - \gamma) = 4(2^{N-1} - 1)P(\overline{Rep}|\overline{Con}Bias) \\
\rho &\leq \frac{1}{1 + \sqrt[N-1]{2}},
\end{aligned}
$$

then it holds that

$$P_2\left(\bigwedge_{n=1}^{N} Bias_n \middle| \bigwedge_{n=1}^{N} \overline{Rep_n}\right) > P_1\left(Bias \middle| \bigwedge_{n=1}^{N} \overline{Rep_n}\right).$$

Proof First, observe that $\rho(1 + \sqrt[N-1]{2}) \leq 1$ is equivalent to $\rho \sqrt[N-1]{2} \leq (1 - \rho)$ what is in turn equivalent to $2\rho^{N-1} \leq \bar{\rho}^{N-1}$. We shall use to obtain the first strict inequality below (this implies $\rho < \bar{\rho}$).

To complete the proof for Scenario 1 it suffices to note that

$$\begin{aligned}
&\sum_{CON} P(CON|HYP) \sum_{\substack{i_1,\dots,i_N \in \{0,1\}^N \\ i_1+i_2+\cdots+i_N \geq 1}} \prod_{n=1}^{N} P(Rel^{i_n}) P(Rep_n^0 | CON\, Rel^{i_n}) \\
&= \sum_{CON} P(CON|HYP) \Big(\rho^N \prod_{n=1}^{N} P(Rep^0|CON\, Rel) \\
&\quad + \sum_{\substack{i_1,\dots,i_N \in \{0,1\}^N \\ i_1+i_2+\cdots+i_N \geq 1 \\ i_1 \cdot i_2 \cdot \cdots \cdot i_N = 0}} \prod_{n=1}^{N} P(Rel^{i_n}) P(Rep^0 | CON\, Rel^{i_n}) \Big) \\
&< \rho^N [P(Con|HYP)\epsilon_+^N + P(\overline{Con}|HYP)(1-\epsilon_-)^N] \\
&\quad + \sum_{CON} P(CON|HYP) \bar{\rho}^{N-1} \rho \sum_{\substack{i_1,\dots,i_N \in \{0,1\}^N \\ i_1+i_2+\cdots+i_N \geq 1 \\ i_1 \cdot i_2 \cdot \cdots \cdot i_N = 0}} \prod_{n=1}^{N} P(Rep^0 | CON\, Rel^{i_n}) \\
&\leq \rho^N [P(Con|HYP)\epsilon_+^N + P(\overline{Con}|HYP)(1-\epsilon_-)^N] \\
&\quad + \bar{\rho}^{N-1} \rho (2^N - 2) [P(Con|HYP)(1-\alpha)\epsilon_+^{N-1} \\
&\qquad + P(\overline{Con}|HYP)(1-\gamma)(1-\epsilon_-)^{N-1}] \\
&\leq \frac{1}{2} \bar{\rho}^{N-1} \rho [P(Con|HYP)\epsilon_+^N + P(\overline{Con}|HYP)(1-\epsilon_-)^N] \\
&\quad + \bar{\rho}^{N-1} \rho 2 (2^{N-1} - 1) [P(Con|HYP)(1-\alpha)\epsilon_+^{N-1} \\
&\quad + P(\overline{Con}|HYP)(1-\gamma)(1-\epsilon_-)^{N-1}] \\
&\leq \frac{1}{2} \bar{\rho}^{N-1} \rho [P(Con|HYP)\epsilon_+^N + P(\overline{Con}|HYP)(1-\epsilon_-)^N] \\
&\quad + \frac{1}{2} \bar{\rho}^{N-1} \rho [P(Con|HYP)\epsilon_+^N + P(\overline{Con}|HYP)(1-\epsilon_-)^N] \\
&= \bar{\rho}^{N-1} \rho [P(Con|HYP)\epsilon_+^N + P(\overline{Con}|HYP)(1-\epsilon_-)^N] \\
&= \sum_{CON} P(CON|HYP) \cdot \bar{\rho}^{N-1} \rho \prod_{n=1}^{N} P(\overline{Rep}|CON\, Rel).
\end{aligned}$$

The proof for Scenario 3 is analogous: rather than summing over the truth values of CON one sums over all possible combinations of truth values of $CON_1, \dots, CON_N$.

Theorem 2 *In Scenario 1 and Scenario 3, if*

$$4(2^{N-1}-1)P(Rep|ConRel) = 4(2^{N-1}-1)(1-\epsilon_+) \leq \alpha = P(Rep|ConBias)$$
$$4(2^{N-1}-1)P(Rep|\overline{Con}Rel) = 4(2^{N-1}-1)\epsilon_- \leq \gamma = P(Rep|\overline{Con}Bias)$$
$$\bar{\rho} \leq \frac{1}{1+\sqrt[N-1]{2}},$$

then for all $0 < P(Hyp), P(Con|Hyp) < 1$ *it holds that*

$$P_2\left(\bigwedge_{n=1}^{N} Rel_n \middle| \bigwedge_{n=1}^{N} Rep_n\right) > P_1\left(Rel \middle| \bigwedge_{n=1}^{N} Rep_n\right).$$

Proof The proof is obtained from the above by a suitable dualisation: switch *Rel* and *Bias*—this includes ρ and $\bar{\rho}$, as well as considering reports which confirm the consequences rather than dis-confirm them.

For Scenario 1 we find

$$P_2\left(\bigwedge_{n=1}^{N} Rel_n|\mathcal{E}_2\right) - P_1(Rel|\mathcal{E}_1) = \frac{P_2\left(\bigwedge_{n=1}^{N} Rel_n\, \mathcal{E}_2\right)}{P_2(\mathcal{E}_2)} - \frac{P_1(Rel\,\mathcal{E}_1)}{P_1(\mathcal{E}_1)}$$

$$\begin{aligned}
&= \frac{\sum_{HYP}\sum_{CON} P_2\left(\bigwedge_{n=1}^{N} Rel_n\, \mathcal{E}_2\, HYP\, CON\right)}{\sum_{HYP}\sum_{CON}\sum_{n=1}^{N}\sum_{REL_n} P_2(\mathcal{E}_2\, HYP\, CON\, REL_1 \dots REL_N)} \\
&\quad - \frac{\sum_{HYP}\sum_{CON} P_1(Rel\,\mathcal{E}_1\, HYP\, CON)}{\sum_{HYP}\sum_{CON}\sum_{REL} P_1(\mathcal{E}_1\, HYP\, CON\, REL)} \\
&= \frac{P(Rel)^N \cdot \sum_{HYP}\sum_{CON} P(HYP)P(CON|HYP)}{\sum_{HYP}\sum_{CON}\sum_{n=1}^{N}\sum_{REL_n} P(HYP)P(CON|HYP)} \\
&\quad \cdot \frac{\prod_{n=1}^{N} P(Rep^{r_n}|CON\, Rel)}{\prod_{n=1}^{N} P(Rep^{r_n}|CON\, REL_n)P(REL_n)} \\
&\quad - \frac{P(Rel) \cdot \sum_{HYP}\sum_{CON} P(HYP)P(CON|HYP)}{\sum_{HYP}\sum_{CON}\sum_{REL} P(REL)P(HYP)P(CON|HYP)} \\
&\quad \cdot \frac{\prod_{n=1}^{N} P(Rep^{r_n}|CON\, Rel)}{\prod_{n=1}^{N} P(Rep^{r_n}|CON\, REL)}.
\end{aligned}$$

The sign of this expression is equal to the sign of

$$\frac{P(Rel)^N \cdot \sum_{HYP} \sum_{CON}}{P(Rel) \cdot \sum_{HYP} \sum_{CON}}$$

$$\frac{P(HYP)P(CON|HYP) \prod_{n=1}^{N} P(Rep^{r_n}|CON\, Rel)}{P(HYP)P(CON|HYP) \prod_{n=1}^{N} P(Rep^{r_n}|CON\, Rel)} - \frac{\sum_{HYP} \sum_{CON} \sum_{n=1}^{N} \sum_{REL_n}}{\sum_{HYP} \sum_{CON} \sum_{REL}}$$

$$\cdot \frac{P(HYP)P(CON|HYP) \prod_{n=1}^{N} P(Rep^{r_n}|CON\, REL_n)P(REL_n)}{P(REL)P(HYP)P(CON|HYP) \prod_{n=1}^{N} P(Rep^{r_n} CON\, REL)}.$$

Since the first term is equal to ρ^{N-1}, the sign of this expression is equal to

$$\begin{aligned}
&\sum_{HYP} \sum_{CON} P(HYP)P(CON|HYP) \cdot \\
&\quad \left(\rho^{N-1} \sum_{REL} P(REL) \prod_{n=1}^{N} P(Rep^{r_n}|CON\, REL) \right. \\
&\quad \left. - \sum_{n=1}^{N} \sum_{REL_n} \prod_{n=1}^{N} P(REL_n)P(Rep^{r_n}|CON\, REL_n) \right) \\
&= \sum_{HYP} \sum_{CON} P(HYP)P(CON|HYP) \cdot \left(\rho^{N-1} \bar{\rho} \prod_{n=1}^{N} P(Rep^{r_n}|CON\, Bias) \right. \\
&\quad \left. - \sum_{\substack{i_1, \ldots, i_N \in \{0,1\}^N \\ i_1 \cdot i_2 \cdot \cdots \cdot i_N = 0}} \prod_{n=1}^{N} P(Rel^{i_n})P(Rep^{r_n}|CON\, Rel^{i_n}) \right).
\end{aligned}$$

For Scenario 3 we have

$$P_2\left(\bigwedge_{n=1}^{N} Rel_n|\mathcal{E}_2\right) - P_1(Rel|\mathcal{E}_1) = \frac{P_2\left(\bigwedge_{n=1}^{N} Rel_n\,\mathcal{E}_2\right)}{P_2(\mathcal{E}_2)} - \frac{P_1(Rel\,\mathcal{E}_1)}{P_1(\mathcal{E}_1)}$$

$$\begin{aligned}
&= \frac{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}}{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}\sum_{n=1}^{N}\sum_{REL_n}} \\
&\quad \frac{P_2\left(\bigwedge_{n=1}^{N} Rel_n\,\mathcal{E}_2\,HYP\,CON_1\ \ldots\ CON_N\right)}{P_2(\mathcal{E}_2\,HYP\,CON_1\ \ldots\ CON_N\,REL_1\ldots REL_N)} \\
&\quad - \frac{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c} P_1(Rel\,\mathcal{E}_1\,HYP\,CON_1\ \ldots\ CON_N)}{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}\sum_{REL} P_1(\mathcal{E}_1\,HYP\,CON_1\ \ldots\ CON_N\,REL)} \\
&= \frac{P(Rel)^N\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}}{\sum_{HYP}\sum_{c=1}^{N}\sum_{n=1}^{N}\sum_{CON_c}\sum_{REL_n}} \\
&\quad \frac{P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,Rel)}{P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,REL_n)P(REL_j)} \\
&\quad - \frac{P(Rel)\cdot\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}}{\sum_{HYP}\sum_{c=1}^{N}\sum_{CON_c}\sum_{REL}} \\
&\quad \cdot \frac{P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,Rel)}{P(REL)P(HYP)\prod_{j=1}^{N} P(CON_j|HYP)P(Rep^{r_j}|CON_j\,REL)}.
\end{aligned}$$

Since we are only interested in the sign of this equation we consider

$$\frac{P(Rel)^N \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c}}{P(Rel) \cdot \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c}}$$

$$\begin{aligned}
&\frac{P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, Rel)}{P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, Rel)} \\
&- \frac{\sum_{HYP} \sum_{c=1}^{N} \sum_{n=1}^{N} \sum_{CON_c} \sum_{REL_n}}{\sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} \sum_{REL}} \\
&\frac{P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, REL_n) P(REL_j)}{P(REL) P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, REL_j)} \\
= \rho^{N-1} &- \frac{\sum_{HYP} \sum_{c=1}^{N} \sum_{n=1}^{N} \sum_{CON_c} \sum_{REL_n}}{\sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} \sum_{REL}} \\
&\frac{P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, REL_n) P(REL_j)}{P(REL) P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, REL_j)}.
\end{aligned}$$

This is equal to the sign of

$$\begin{aligned}
&- \rho^{N-1} \bar{\rho} \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c}^{N} P(HYP) \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, Bias) \\
&\quad - \sum_{HYP} \sum_{c=1}^{N} \sum_{CON_c} \sum_{\substack{i_1, \ldots, i_N \in \{0,1\} \\ i_1 + \cdots + i_N \geq 1}} P(HYP) \\
&\quad \cdot \prod_{j=1}^{N} P(CON_j|HYP) P(Rep^{r_j}|CON_j\, Rel_j^{i_j}) P(Rel_j^{i_j}) \\
&= \sum_{HYP} P(HYP) \sum_{c=1}^{N} \sum_{CON_c} \prod_{j=1}^{N} P(CON_j|HYP) \cdot \\
&\left[\rho^{N-1} \bar{\rho} P(Rep^{r_j}|CON_j\, Rel) - \sum_{\substack{i_1, \ldots, i_N \in \{0,1\} \\ i_1 \cdot \cdots \cdot i_N = 0}} P(Rep^{r_j}|CON_j\, Rel_j^{i_j}) P(Rel_j^{i_j}) \right].
\end{aligned}$$

To complete the proof for Scenario 1 it suffices to note that

$$
\begin{aligned}
&\sum_{CON} P(CON|HYP) \sum_{\substack{i_1,\dots,i_N \in \{0,1\}^N \\ i_1 \cdot i_2 \cdot \dots \cdot i_N = 0}} \prod_{n=1}^{N} P(Rel^{i_n}) P(Rep|CON\, Rel^{i_n}) \\
&= \sum_{CON} P(CON|HYP) \Big(\bar{\rho}^N \prod_{n=1}^{N} P(Rep|CON\, Bias) \\
&\quad + \sum_{\substack{i_1,\dots,i_N \in \{0,1\}^N \\ i_1 + i_2 + \dots + i_N \geq 1 \\ i_1 \cdot i_2 \cdot \dots \cdot i_N = 0}} \prod_{n=1}^{N} P(Rel^{i_n}) P(Rep|CON\, Rel^{i_n}) \Big) \\
&< \bar{\rho}^N [P(Con|HYP)\alpha^N + P(\overline{Con}|HYP)\gamma^N] \\
&\quad + \sum_{CON} P(CON|HYP) \rho^{N-1} \bar{\rho} \sum_{\substack{i_1,\dots,i_N \in \{0,1\}^N \\ i_1 + i_2 + \dots + i_N \geq 1 \\ i_1 \cdot i_2 \cdot \dots \cdot i_N = 0}} \prod_{n=1}^{N} P(Rep|CON\, Rel^{i_n}) \\
&\leq \bar{\rho}^N [P(Con|HYP)\alpha^N + P(\overline{Con}|HYP)\gamma^N] \\
&\quad + \rho^{N-1} \bar{\rho} (2^N - 2) [P(Con|HYP)(1 - \epsilon_+)\alpha^{N-1} \\
&\qquad + P(\overline{Con}|HYP)\epsilon_- \gamma^{N-1}] \\
&\leq \frac{1}{2} \rho^{N-1} \bar{\rho} [P(Con|HYP)\alpha^N + P(\overline{Con}|HYP)\gamma^N] \\
&\quad + \rho^{N-1} \bar{\rho} 2 (2^{N-1} - 1) [P(Con|HYP)(1 - \epsilon_+)\alpha^{N-1} \\
&\quad + P(\overline{Con}|HYP)\epsilon_- \gamma^{N-1}] \\
&\leq \frac{1}{2} \rho^{N-1} \bar{\rho} [P(Con|HYP)\alpha^N + P(\overline{Con}|HYP)\gamma^N] \\
&\quad + \frac{1}{2} \rho^{N-1} \bar{\rho} [P(Con|HYP)\alpha^N + P(\overline{Con}|HYP)\gamma^N] \\
&= \rho^{N-1} \bar{\rho} [P(Con|HYP)\alpha^N + P(\overline{Con}|HYP)\gamma^N] \\
&= \sum_{CON} P(CON|HYP) \cdot \rho^{N-1} \bar{\rho} \prod_{n=1}^{N} P(Rep|CON\, Bias).
\end{aligned}
$$

The proof for Scenario 3 is again analogous and skipped in the interest of time.

References

Abdin, Y., Auker-Howlett, D. J., Landes, J., et al. (2019). Reviewing the mechanistic evidence assessors e-synthesis and EBM+: A case study of amoxicillin and drug reaction with eosinophilia and systemic symptoms (DRESS). *Current Pharmaceutical Design*, *25*(16), 1866–1880. https://doi.org/10.2174/1381612825666190628160603.

Angere, S., & Olsson, E. J. (2017). Publish late, publish rarely!: Network density and group performance in scientific communication. In C. Boyer-Kassem, T. Mayo-Wilson, & M. Weisberg (Eds.),

Scientific collaboration and collective knowledge: New essays. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190680534.003.0002.
Bero, L. A., & Grundy, Q. (2016). Why having a (nonfinancial) interest is not a conflict of interest. *PLOS Biology*, *14*(12), 1–8. https://doi.org/10.1371/journal.pbio.2001221.
Bovens, L., & Hartmann, S. (2002). Bayesian networks and the problem of unreliable instruments. *Philosophy of Science*, *69*(1), 29–72. https://doi.org/10.1086/338940.
Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.
Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago: University of Chicago Press.
Casini, L., & Landes, J. (2020). Confirmation by robustness analysis. A Bayesian account. *Economics & Philosophy* (unpublished manuscript).
Claveau, F. (2013). The independence condition in the variety-of-evidence thesis. *Philosophy of Science*, *80*(1), 94–118. https://doi.org/10.1086/668877.
Claveau, F., & Grenier, O. (2019). The variety-of-evidence thesis: A bayesian exploration of its surprising failures. *Synthese*, *196*, 3001–3028. https://doi.org/10.1007/s11229-017-1607-5.
Collins, P. J., Hahn, U., von Gerber, Y., & Olsson, E. J. (2015) The bi-directional relationship between source characteristics and message content. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, C. D. Matlock, T.and Jennings, & P. P. Maglio (Eds.), *Proceedings of CogSci* (pp. 423–428).
Collins, P. J., Hahn, U., von Gerber, Y., & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content. *Frontiers in Psychology*, *9*, 1–16. https://doi.org/10.3389/fpsyg.2018.00018.
De Pretis, F., Landes, J., & Osimani, B. (2019). E-synthesis: A Bayesian framework for causal assessment in pharmacosurveillance. *Frontiers in Pharmacology*, *10*, https://doi.org/10.3389/fphar.2019.01317.
De Pretis, F., Landes, J., Osimani, B., & Peden, W. J. (2020). Drug Safety and personalized medicine: A possible interaction through E-Synthesis. In M. Bertolaso, & S. Canali (Eds.), *Personalized medicine. A multidisciplinary approach to complexity*. Springer (submitted).
Edman, M. (1973). Adding independent pieces of evidence. In S. Halldén (Ed.), *Modality, Morality and other problems of sense and nonsense* (pp. 180–188). Lund: Gleerup.
Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press.
Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased. In *Psychology of learning and motivation: Motivated reasoning and rationality* (Vol. 61, pp. 41–102). https://doi.org/10.1016/B978-0-12-800283-4.00002-2.
Hahn, U., Harris, A. J. L., & Corner, A. (2016). Public reception of climate science: Coherence, reliability, and independence. *Topics in Cognitive Science*, *8*(1), 180–195. https://doi.org/10.1111/tops.12173.
Hansson, B. (1983). Epistemology and evidence. In P. Gärdenförs, B. Hansson, N.-E. Sahlin, & S. Halldén (Eds.), *Evidentiary value: Philosophical, judicial, and psychological aspects of a theory: essays dedicated to Sören Halldén on his sixtieth birthday*. Lund: Gleerup.
Holman, B., & Bruner, J. P. (2015). The problem of intransigently biased agents. *Philosophy of Science*, *82*(5), 956–968. https://doi.org/10.1086/683344.
Kenyon, T. (2014). False polarization: Debiasing as applied social epistemology. *Synthese*, *191*(11), 2529–2547. https://doi.org/10.1007/s11229-014-0438-x.
Landes, J. (2020a). The variety of evidence thesis and its independence of degrees of independence. *Synthese* (under review).
Landes, J. (2020b). Variety of evidence. *Erkenntnis*, *85*, 183–223. https://doi.org/10.1007/s10670-018-0024-6.
Landes, J., Osimani, B., & Poellinger, R. (2018). Epistemology of causal inference in pharmacology. *European Journal for Philosophy of Science*, *8*, 3–49. https://doi.org/10.1007/s13194-017-0169-1.
Merdes, C., von Sydow, M., & Hahn, U. (2021). Formal models of source reliability. *Synthese*. https://doi.org/10.1007/s11229-020-02595-2.
Olsson, E. J. (2005). *Against coherence: Truth, probability, and justification*. Oxford: Oxford University Press.
Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme*, *8*(2), 127–143. https://doi.org/10.3366/epi.2011.0012.
Osimani, B., & Landes, J. (2020). Varieties of error and varieties of evidence. *British Journal for the Philosophy of Science* (under review).
Rosenstock, S., Bruner, J., & O'Connor, C. (2017). In epistemic Networks, is less really more? *Philosophy of Science*, *84*(2), 234–252. https://doi.org/10.1086/690717.

Schum, D. A. (1988). Probability and the processes of discovery, proof and choice. In P. Tillers & E. D. Green (Eds.), *Probability and Inference in the law of evidence—The uses and limits of Bayesianism* (pp. 213–270). London: Kluwer.

Stegenga, J., & Menon, T. (2017). Robustness and independent evidence. *Philosophy of Science*, *84*(3), 414–435. https://doi.org/10.1086/692141.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. https://doi.org/10.1037/0033-295x.90.4.293.

von Sydow, M. (2011). The bayesian logic of frequency-based conjunction fallacies. *Journal of Mathematical Psychology*, *55*(2), 119–139. https://doi.org/10.1016/j.jmp.2010.12.001.

Zollman, K. J. S. (2013). Network epistemology: Communication in epistemic communities. *Philosophy Compass*, *8*(1), 15–27. https://doi.org/10.1111/j.1747-9991.2012.00534.x.