Madsen, J.K. and Hahn, Ulrike and Pilditch, T.D. (2020) The impact of partial source dependence on belief and reliability revision. Journal of Experimental Psychology: Learning, Memory, and Cognition, ISSN 0278-7393. (In Press)

The Impact of Partial Source Dependence on Belief and Reliability Revision

Jens Koed Madsen (jens.madsen@ouce.ox.ac.uk)*

University of Oxford, School of Geography and the Environment

South Parks Road, OX1 3QY, Oxford, United Kingdom

Ulrike Hahn (u.hahn@bbk.ac.uk)

Department of Psychological Sciences, Birkbeck, University of London,

Malet Street, London, WC1E 7HX, U.K.

Toby D. Pilditch (t.pilditch@ucl.ac.uk)

Department of Experimental Psychology, University College London,

26 Bedford Way, London, WC1H 0AP, U.K.

*Corresponding Author: Correspondence should be addressed to Jens Koed Madsen, School of Geography and the Environment, South Parks Road, OX1 3QY. Electronic mail can be sent to jens.madsen@ouce.ox.ac.uk

Declarations of interest: None.

In this paper, we explore how people revise their belief in a hypothesis and the reliability of sources in circumstances where those sources are either independent or are partially dependent because of their shared, common, background. Specifically, we examine people's revision of perceived source reliability by comparison with a formal model of reliability revision proposed by Bovens and Hartmann (2003). This model predicts a u-shaped trajectory for revision in certain circumstances: If a source provides a positive report for an unlikely hypothesis, perceived source reliability should decrease; as additional positive reports emerge, however, estimates of reliability should increase. Participants' updates in our experiment show this u-shaped pattern. Furthermore, participants' responses also respect a second feature of the model, namely that perceived reliability should once again decrease when it becomes known that the sources are partially dependent. Participants revise appropriately both when a specific shared reliability is observed (e.g. sources went to the same, low quality school) and when integrating the possibility of shared reliability. These findings shed light on how people gauge source reliability and integrate reports when multiple sources weigh in on an issue as seen in public debates.

Keywords: Bayesian reasoning; Reliability revision; Sequential testimonies

Introduction

In everyday life, we not only continuously receive evidence from others on everything from the weather, through politics, to science; we also frequently receive evidence from multiple sources *on the same issue*. These sources may be independent of one another, producing data and conclusions in isolation from our other sources. However, in many, if not most, circumstances, they will exhibit some degree of dependence: they may share a common background, share a common information source, or even discuss their evidence prior to providing individual reports.

A failure to appreciate the dependence of information can lead to potentially disastrous conclusions. For example, an intelligence agency may receive multiple reports concerning weapons of mass destruction in a foreign country, and increase their belief that those weapons of mass destruction exist. Multiple congruent reports may sway the agency to believe what was initially an improbable hypothesis. If it subsequently becomes known that all reports came from sources with a common, flawed approach, the corroboration these reports seemed to provide is compromised. That is, an appropriate appreciation for the dependence among, or independence of sources is critical to reasoning and decision-making.

The paper examines belief revision processes concerning both a claim at issue and the reliability of the reporting sources under conditions of either independence or partial dependence. In exploring this, we use a formal, Bayesian account of dependence and reliability (Bovens & Hartmann, 2003, chapter 3).

The impact of source reliability on belief revision[1]

The reliability of one's sources is crucial for everyday reasoning and decision-making. Because so much of human knowledge comes from the testimony of others, the impact of source reliability has received considerable empirical attention. Source reliability has been shown to influence the reception of persuasive messages (Petty & Cacioppo, 1984; Tormala & Clarkson, 2007), the development of children's perception of the world (Harris & Corriveau, 2011), legal reasoning (Lagnado et al., 2013), adherence with persuasion strategies (Cialdini, 2007), and how people are seen in social situations (Fiske, Cuddy & Glick, 2007; Cuddy, Glick & Beninger, 2011). There has also been increased interest in lay people's understanding of the impact of source dependence (see e.g., Yousif, Aboody & Keil, 2019).

Cognitive and social psychology have approached source reliability in a number of ways. While reliance on the reliability of others has typically been viewed as a shallow persuasive cue (Petty & Cacioppo, 1984), cognitive and developmental psychologists have tended to stress how sensitivity to source characteristics is rationally justified (see e.g., Hahn, Harris & Corner, 2009; Collins et al. 2018). At the same time, research has differed in the extent to which it separates the reliability of a source into distinct aspects (such as accuracy, trustworthiness, or bias, see e.g., Schum, 1994; Pornpitakpan, 2004), or simply rolls these into one overall measure of source reliability that reflects their combined net effect (e.g., Hahn et al., 2009).

---

[1] Some research uses the term 'source credibility' for 'reliability'. For the sake of parsimony, here we only use 'reliability'.

Harris, Hahn, Madsen & Hsu (2015) tested a probabilistic model that amalgamates two components of reliability: perceived trustworthiness and perceived expertise. Expertise refers to the *capacity* to provide accurate information about the topic. This is domain-dependent. For example, a carpenter can provide relevant and accurate information about types of wood, but may not be able to provide guidance on neurosurgery. Trust, on the other hand, refers to the *intention* to provide true and accurate information to the best of one's ability. For example, if the carpenter has a motive to sell surplus wood, she may falsely claim a particular type of wood is useful even in situations where it is not. Expertise and trustworthiness are orthogonal such that a person can be highly expert, but very untrustworthy and vice versa. However, these two factors ultimately combine to determine how likely it is that a source's testimony genuinely reflects the truth or falsity of a claim. This makes probabilistic models a natural way for thinking about sources and their reliability. As a result, Bayesian, probabilistic models of source reliability can now be found in the epistemology and philosophy of science literature (e.g., Olsson, 2005; Olsson, 2011; Bovens & Hartmann, 2003), in cognitive psychology (e.g., Harris et al., 2015; Hahn, Harris & Corner, 2016; Pilditch, Hahn & Lagnado, 2018) and in developmental psychology (e.g. Shafto, Goodman & Frank, 2012).

Bayesian approaches to reasoning represent (subjective) degrees of belief with probabilities, and use Bayes' rule for belief revision (Oaksford & Chater, 2007; Howson & Urbach, 1996). The Bayesian approach was suggested as an alternative to logicist approaches to reasoning (Oaksford & Chater, 1991) and has successfully been applied to evidential reasoning in a wide range of contexts (e.g. Lagnado, Gerstenberg & Zultan, 2015; Kemp & Tenenbaum, 2009; Oaksford & Chater, 1994) as well as to argumentation (Hahn & Oaksford, 2006; 2007; 2014). This work suggests Bayesian reasoning can capture much of information integration in everyday reasoning.

In this paper, we examine reasoning about source reliability and the veracity of a source's claim with reference to a Bayesian source credibility model first proposed by Bovens and Hartmann (2003). A full formal description of the model is provided in the Supplementary Material, so we limit our introduction here to its most salient features. The model represents an agent's beliefs about claims and sources, and is used to revise those beliefs in light of new evidence. The overall 'reliability' of a source (which may itself reflect a range of factors that are not modelled in detail) is represented by a probability, *P(Rel)*, that is, a number between 0 and 1. This number represents the agent's subjective degree of belief that the source has the capacity and willingness to provide accurate information about the hypothesis at issue. In Boven's and Hartmann's basic (2003) model, a source faithfully reports the truth when reliable, whereas if the source is unreliable, their testimony is unrelated to the truth or falsity of the claim at issue, and as good as flipping a coin.[2] Hence a probability of 1 for reliability (i.e., *P(Rel)* = 1) means the agent is 100% certain that the source reliably reports the truth. A probability of *P(Rel)* = .7 means the agent is 70% certain, and so on. This perceived reliability determines how much weight the source's testimonial report is given, that is, it determines the evidential value assigned to that report. On receiving a source's report, agents then use Bayes' rule to revise their belief in the claim at issue.

---

[2] It is also possible to represent bias on the part of the source (see Bovens & Hartmann, 2003; and, in the context of fitting experimental data, Jarvstad & Hahn, 2011), but we make no use of this in the modelling here.

At the same time, however, agents revise their beliefs *in the reliability of the reporting source*, once again using Bayes' rule. In other words, agents use testimonial reports to jointly revise beliefs both about the underlying claim and the source itself. Where the testimonial report has an unexpected content (i.e., prior to the current degree of belief about the claim at issue), agents modify their belief in the claim in line with that report, but they *also* somewhat reduce their degree of belief that the source is reliable. By contrast, when a report is in line with present beliefs about the claim, belief in the reliability of the source is increased.

In short, the Bovens and Hartmann (2003) model implements a strategy for dealing with the testimony of sources whose reliability is not fully known, as is frequent in many real-world contexts. It is a common feature of real-world testimony that the recipient does not know exactly how reliable the source is (and in some contexts, such as exchanges via social media, the recipient may, in fact, know nothing about the source at all). The Bovens and Hartmann model is an attempt to provide a rational solution to this common place difficulty, and it is one of a family of formal models that implement the intuitive reliability updating strategy just outlined (for other implementations see e.g., Olsson, 2011). This strategy, which has been labelled "expectation-based updating" by Collins et al. (2018), can be found in lay reasoners as evidenced both by experimental manipulation (e.g., Collins et al., 2018; Collins & Hahn, subm.) and model-fitting (e.g., Harris et al., 2015; Shafto et al., 2012). It has also been examined in simulations both with individual agents (Hahn, Merdes & von Sydow, 2018) and in societies of artificial agents (Madsen & Pilditch, 2018; Madsen, Bailey & Pilditch, 2018; Hahn, von Sydow & Merdes, 2019).

As detailed in the introduction, we often receive information about a claim from more than one party, whether these are multiple witnesses in a trial, reporters from different media organizations, scientists from different labs, or simply different friends and acquaintances speaking to the same issue (see e.g. Philipps, Hahn & Pilditch, 2018; Madsen, Hahn & Pilditch, 2018). This makes it essential to probe how lay reasoners deal with multiple sources, and how they accommodate a key feature of multiple sources in the real world, namely that these sources may not be wholly independent.

Shared reliability and reliability revision

Not only have past studies of belief revision tended to focus on contexts in which there is a single source of evidence, they have also tended to focus on belief in the claim at issue itself. This focus makes sense, as it will be that claim which motivates or influences decisions such as voting, economic behavior, etc. and, from that perspective, the reliability of a testimonial source seems simply like an auxiliary factor in revising beliefs about that claim.

However, perceived reliability should *moderate* the impact of testimonial evidence. This follows from normative Bayesian models that spell out how rational agents *should* revise their beliefs (e.g., Hahn et al., 2009), if they want those beliefs to be as accurate as possible (see Pettigrew, 2016), and it can be seen descriptively in many studies of evidential reasoning, argumentation and persuasion (e.g., Petty & Cacioppo, 1984; Corner & Hahn, 2009; Lagnado et al., 2015). This means the perception of source reliability *itself* is important in the belief revision process. Therefore, if the perceived reliability of the source changes, Bayesian normative models suggest the subsequent impact of that source should change also. For example, the Boy Who Cried Wolf made repeated bad forecasts (willingly), causing villagers to decrease their estimates of his reliability, with disastrous consequences.

Here, the villagers had access to actual outcomes (no wolves were actually observed), but in many contexts, cognitive agents do not have access to a forecasting history from their sources. In such circumstances, the expectation-based reliability updating strategy described above seems reasonable, and the Bovens and Hartmann (2003) model provides a normative, Bayesian implementation. This model can capture the belief dynamics involved with one or more evidence reports from a single source. But it can also be used to capture reports from multiple sources. In this case, the belief dynamics vary depending on whether or not those sources are independent.

Within the Bayesian Framework, graphical models (so-called Bayesian Belief Networks, see Pearl, 1988) are widely used to represent (in)dependence relations between variables, and these graphical models can be used to simplify Bayesian computations. Variables are represented as nodes, and arrows between nodes represent dependencies. Fig. 1 shows a simple graphical representation of multiple independent witnesses in the Bovens and Hartmann model.

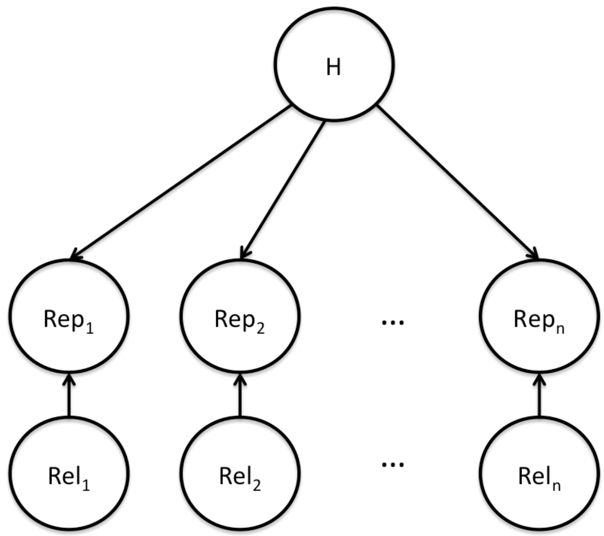

Fig. 1: Independent condition. Each source has an independent reliability (Rel) and provides a report (Rep) about a hypothesis (H), conditionally independent of other sources.

There is an underlying claim at issue (represented by the variable H, for hypothesis, which can take on the values true or false). The variables Rep_1 to Rep_N represent testimonial reports, which assert the hypothesis to be true or false. The variables Rel_1 to Rel_N represent the reliability of sources 1 to N respectively. The arrows leading to the report variables (from both hypothesis and reliability nodes) indicate that the source's report is determined by two factors: the true state of the world and their reliability. Crucially, there are no links between reporters or their reliabilities in Fig. 1, indicating that the sources provide their reports entirely independently. For example, climate scientists may conduct independent studies of the same phenomenon and produce reports of their findings without any knowledge of the conclusions of other teams. This would constitute fully independent sources, as they do not rely on the same apparatus, do not share results before making their reports known, and do not communicate between teams.

However, sources may also be dependent in multiple ways (see, in a climate context, for example Hahn, Corner & Harris, 2016). One such source of dependence is a shared common background. The graphical model of Fig. 2 represents this type of partial dependence. Here, there are still no direct links between report variables, indicating that sources provide these without conferring. However, there is an indirect dependence through the shared background variable (SR) which provides a common influence on all sources' reliability. In other words, the respective reliabilities, Rel_1 to Rel_N, may still vary individually, but are influenced by a common cause. An example of this might be a joint educational background (e.g., economists coming from the same good or bad school) that shapes the sources' interpretation of the data.

This dependency provides a constraint on the informativeness of each source and changes the belief dynamics.

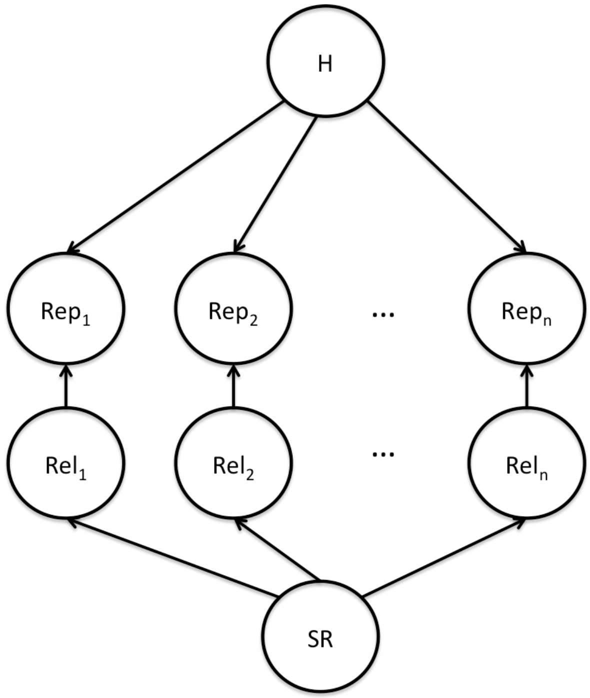

Fig. 2: Shared Reliability condition. All source reliabilities (Rel) now share a common ancestor "shared reliability" (SR; e.g. shared background).

In particular, the fact that the reliability of these sources is now conditional on a common background can weaken the *normative impact* of their reports. In other words, it can weaken the extent to which those reports *should* change beliefs about the claim. For example, if several doctors provide diagnoses for a patient, it makes an operational difference to the impact of their reports if they were found to all have attended the same low standard medical course. In comparison to a fully independent case, recipients should treat reports from these doctors as partially compromised. This shared background not only influences the reliability of each source, but in turn influences the degree to which reports from those sources impact the hypothesis / claim.

The graphical models of Fig. 1 and 2 are not just visual aids. Each corresponds to a set of equations (see Bovens and Hartmann, 2003, reprinted here in the Supplementary Material), that allows for calculation of not only how beliefs in the claim at issue change as one or more testimonial reports came in, but also how these reports lead to *changes in the perceived reliabilities.* It is these changes in perceived reliability that we examine in the present paper, together with participants degree of belief in the underlying claim.

More specifically, we test if lay people's intuitions are qualitatively in line with the dynamics of the model. We describe the experimental hypotheses in more detail in the following.

Present Research

In the paper, we explore four hypotheses.

H1: Do participants revise their belief in the reliability of the source in line with Bayesian predictions? For improbable hypotheses (in their example, Bovens and Hartmann use a probability of the hypothesis, P(H) = 0.3), a single positive report should *decrease the reliability of the source*. However, given further concurring reports from independent sources, the perceived source reliability should increase. This happens as the recipient receives multiple corroborative reports for the same unlikely hypothesis. In this study, sources report that the hypothesis is true (i.e., they provide a 'positive' report for the hypothesis).

H2: In line with Bayesian predictions, sources that provide positive statements for highly likely hypotheses (e.g. P(H) = .9) should increase in reliability, albeit less so than the decreases in H1.

H3: Do participants use source dependency to adjust reliability estimations? If sources are wholly independent, participants should, normatively, update in line with the predictions tested in the first hypothesis. If, however, sources are partially dependent (via their shared background), this pattern should change. Specifically, if the recipient learns that three sources, initially believed to be independent, draw their reports from the same source (e.g. a shared source or shared affiliation), the recipient should reduce her belief in the hypothesis and the credibility of sources to fit perception of the shared reliability. For example, if three economists independently state the economy is about to crash, a recipient may increase her belief in this hypothesis considerably given the independence. However, if the recipient subsequently learns the economists are employed at the same company, she may decrease the belief in the hypothesis, as the sources are dependent. We explore this hypothesis in two stages; first, when providing participants with a specific, observed instance of a shared background (e.g. high- or low-quality schooling of sources; Experiment 1), and second the more complex case where participants must integrate the *possibility* of a shared background (which could be high- or low-quality: Experiment 2).

H4: Do participants adjust reliability estimates of sources retrospectively, or do additional reports only result in updates to the most recent (reporting) sources? That is, after the first report, participants provide their reliability estimate for this first reporting source. We explore whether seeing subsequent positive reports from other sources for the same hypothesis leads to a revision of the reliability of the original source despite the fact that this source does not contribute additional reports. This should happen, as the recipient learns the source, initially discredited for providing an unlikely report, might be a reliable source given multiple corroborative reports from other (independent) sources. If participants revise their beliefs about the reliability of the source retroactively, we should see no differences between estimates of the source reliability for each of the sources (given new reports), as previous sources are also revised in light of new reports. If, however, participants do not revise beliefs retrospectively, the reliability of individual sources should differ, as participants learn additional information.

Experiment 1: Method[3]

To test the above hypotheses, we employ the following methodology: To test H1 & H2, the prior probability of the hypothesis is manipulated as high/low. This allows exploration of whether perceived reliability initially decreases and then subsequently increases given additional positive reports for highly unlikely statements (H1) and if providing positive reports for highly likely statements does not exhibit this effect (H2). To test H3, additional corroborating sources are incrementally introduced, followed by an SR manipulation. SR was presented as either a shared high reliability influence (a school with an excellent reputation) or a shared low reliability influence (a school with a poor reputation). This allowed us to explore the simpler case of sensitivity to the *observed quality* of the SR. To test H4, P(H) as well as $P(Rel_N)$ estimates are elicited after each report for both the hypothesis and for every source (meaning $P(Rel_1)$, the perceived reliability of $source_1$ is elicited three times, once after each positive report). Note that the initial, prior reliability of the three sources was elicited with a single judgment pertaining to that type of source.

[3] The study was approved by the Department of Psychological Sciences Research Ethics Committee of Birkbeck, University of London (reference 161754/5).

Participants: 100 participants (58 female, M_{age} = 30.09, SD = 10.44) were recruited from the online recruitment source Prolific Academic. All had to be aged 18+ and be native English speakers from either the UK or US. Median completion time was 5.25 min (SD = 3.95) and participants were paid £1.00 (resulting in an effective hourly wage of £11.42/hour for participation).

Experiment 1: Materials and procedure

Materials: In order to test the above hypotheses, two scenarios were used. In scenario 1 (low probability condition), participants were asked to consider the likelihood of a market crash within a 6-month period. Specifically, they saw the following:

"Imagine you are watching a news programme about the economy. Specifically, the programme considers whether or not the UK stock market will crash (i.e. fall by more than 30%) within the next 6 months. Historically, the likelihood of a crash occurring within a 6-month window is 5%.

In your opinion, how likely is the UK stock market to crash within the next 6 months?"

Scenario 2 (high probability condition) considers the likelihood that the salmon population will grow within a 5-year period. Specifically, participants saw the following:

"Imagine you are watching a nature programme about fish. Specifically, the programme considers whether or not the salmon population of Norway will grow (i.e. increase by more than 10%) over the next 5 years. Historically, the likelihood of an increase in the salmon population in Norway within a 5-year window is 85%.

In your opinion, how likely is the salmon population of Norway will grow over the next 5 years?"

In addition to the scenarios, participants were presented with statements from experts in the field (economist and biologists)[4]. This allowed for reliability measures of the sources. For the biological scenario, they saw the following:

"Reliability can be defined as having access to relevant information about a topic, and a willingness to say what you believe to be the true state of the world.

How reliable are biologists in predicting the growth of species?"

To generate reports about the hypothesis, participants were told experts had been interviewed. Specifically, they saw the following:

"Now, imagine that a biologist, Linda, is being interviewed about the salmon. Linda states the following: "I am completely certain the salmon population of Norway will grow over the next 5 years."

---

[4] 50% of the named experts were female.

Given Linda's report, how likely is it the salmon population of Norway will grow over the next 5 years?"

Finally, to generate SR conditions (high and low), the participants were told the experts had attended the same school. Specifically, they saw the following:

"It turns out, all the interviewed biologists studied at the same school and subscribe to the same biological models. Their school has a very good reputation for excellent teaching and accurate approaches to biology [High-quality SR condition]// Their school has a very bad reputation for sloppy teaching and out-dated approaches to biology [Low-quality SR condition].

Given the fact that they all studied at the same school and follow the same biological models, how likely is it the salmon population of Norway will grow over the next 5 years?"

In all, the materials included elicitation of prior beliefs (P(H) and the prior reliability of the type of source(s), $P(Rel_{Profession})$, and posterior beliefs regarding the hypothesis and the reliability of the source after each report, i.e. $P(H|Rep_1,\ldots,Rep_N)$ and $P(Rel_{1\text{-}N}|Rep1, \ldots Rep_N)$, where N refers to the total number of sources who have provided reports at the point of measurement. Posterior beliefs are elicited after each report (in the current design, after reports 1, 2, and 3) and after the subsequent SR manipulation (high and low-quality).

Procedure: Participants first provided prior estimates for their beliefs in the hypothesis on a scale from 0-100 (0: I am completely certain the stock market will NOT crash within the next 6 months; 100: I am completely certain the stock market will crash within the next 6 months) and their belief in the reliability of the type of source (economist or biologist) from 0-100 (0: biologists are completely unreliable; 100: biologists are completely reliable).

Having provided their priors, participants saw sequential reports from experts (in total, participants read 3 reports, all of whom were positive). After each report, participants provided their degree of belief in the hypothesis as well as their degree of belief in the reliability of each source that had reported so far (thus, the reliability of $source_1$ was elicited three times, but the reliability of $source_3$ was only elicited once, after the third report).

Finally, participants were asked to "…consider two possible continuations to the scenario, providing your assessments for each". They then read both SR conditions and were asked to provide their degree of belief in the hypothesis and in the reliability of each expert given the dependency between the experts.

The study was a mixed design. Within-subjects, participants saw both scenarios (high and low likelihood). Between-subjects, and in order to manipulate shared reliability, half of participants were told the sources came from a good school (high-quality SR condition) and the other half were told the sources had graduated from a school with a poor reputation (low-quality SR condition). Thus, for each scenario (low and high probability), participants were given 3 reports for each scenario (all confirming the hypothesis), followed by the SR condition statement. Which scenario was presented first (high or low probability) was counterbalanced, and whether a scenario was followed by a high or low-quality SR condition statement was manipulated between-subjects, independently for each scenario.

Experiment 1: Results

All inferential statistics reported below were Bayesian[5], and were conducted using the JASP statistical software (JASP Team, 2019). The resulting Bayes factors (BF_{10}) detail the likelihood ratio of the data given the experimental hypothesis, over the data given the null. In other words, BFs indicate how much more likely the data are assuming that the experimental hypothesis is true, than it would be under the null hypothesis of no difference. BFs of 1-3 may be considered anecdotal support; 3-10 as "substantial" support; 10-30 as "strong" support; 30-100 as "very strong"; and >100 as "decisive" (Jeffreys, 1961; but for further explanation on the use of Bayesian statistics, see Kruschke & Liddell, 2018; Wagenmakers, 2007).

The probability manipulations were successful in generating high and low estimates for the two scenarios: The market crash scenario was rated as unlikely (M = .32, SD = .23) and the salmon growth scenario was rated as likely (M = .81, SD = .15) . The scenario with the unlikely prior was particularly fortunate, as Bovens and Hartmann use P(H) = .3 as the example in their book, making the current scenario comparable to their example. For reliability, Bovens and Hartmann use P(Rel) = .5. Participants rated both sources in our scenarios higher ($P(Rel_{Economist})$: M = .60, SD = .22; $P(Rel_{Biologist})$: M = .73, SD = .16). Importantly, though, both sources were rated positively, which allowed us to

[5] All analyses used the default JASP (uninformed) prior.

test whether positive reports of unlikely hypotheses influenced reliability estimates negatively.

To test whether participants revise their belief in the reliability of the source in line with Bayesian predictions (hypothesis 1), we explore if participants adjusted reliability estimates in the initial source given sequential testimonies. Bayesian predictions dictated that positive reports of an unlikely hypothesis should initially decrease estimates of reliability.

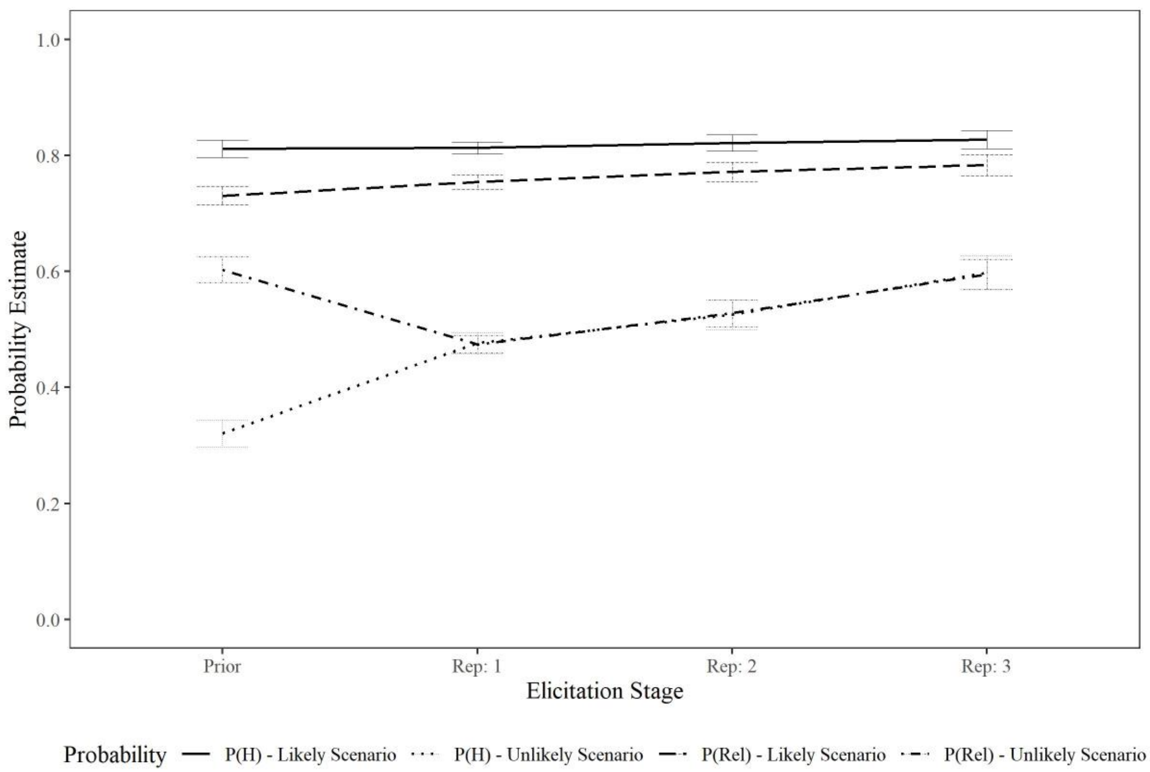

Figure. 3. Displayed are the mean participant estimate of the probability that the claim, H, is true, before receipt of the reports and then after receipt of each of the three, sequentially received, testimonial reports, along with the mean estimate of the reliability of source 1, Rel_1, prior to any reports, $P(Rel_1)$, and then after each additional report (i.e., $P(Rel_1|Rep_1)$, $P(Rel_1|Rep_1,Rep_2)$ and $P(Rel_1|Rep_1,Rep_2,Rep_3)$) both for likely, (P(H) = .81,

and unlikely, P(H) = .32), scenarios. Error bars reflect +/- 1 Standard Error. The y-axis shows estimated probabilities expressed in decimal form (0-1). Data are plotted separately for the high- and low likelihood condition

In line with the predictions of the Bayesian model, using a repeated measures ANOVA ($P(Rel_1) - P(Rel_1|Rep_1)$)[6], the perceived reliability of source 1 is revised negatively (i.e. decreases compared with baseline reliability) given a positive report of an unlikely hypothesis (*N* = 100), BF_{10} = 10907.7 (in the current design, the source predicts the stock market will crash within a 6-month period). However (and also in line with Bayesian model predictions), as participants learn other experts provide corroborating reports ($P(Rel_1|Rep_1) - P(Rel_1|Rep_1, Rep_2,Rep_3)$), they revise their belief in the initial source and revise reliability in a positive direction (*N* = 100), $BF_{10} = 1.36 * 10^{11}$. This specifically tested a scenario with a low prior probability (here, P(H) = .32).

In addition, participants *increase* their belief in the likelihood of the hypothesis whilst they simultaneously *decrease* belief in the source reliability (P(H) to $P(H|Rep_1)$; *N* = 100), $BF_{10} = 1.94 * 10^6$. While the paper focuses on testing the former (change in reliability), it is worth noting the participants do follow previous findings (e.g. Harris et al., 2015) that show P(H|Rep) increases if the source is viewed as reliable.

To test whether sources that provide positive statements for highly likely hypotheses increase their reliability (hypothesis 2), we conducted a repeated measures ANOVA ($P(Rel_1)$ to $P(Rel_1|Rep_1, Rep_2, Rep_3)$) to see if the reports change the estimation

[6] Where $P(Rel_1) = P(Rel_{Profession})$, that is, the prior reliability of source 1 is the generic prior for that type of source (economist or biologist).

of the reliability of the biologist. While there is a substantial increase in reliability as more reports are given (N = 100), BF_{10} = 3.21, we note this difference is small (can be seen visually in Fig. 3 where P(Rel) remains fairly flat).

The introduction of shared reliability

To test whether source independence impacts reliability estimations (hypothesis 3), we compared posterior degrees of belief in the hypothesis and the reliability of the sources before and after a shared reliability (SR) of either observed high- or low-quality is introduced. As described in the above, we manipulate the reliability of the SR and compare $P(H|Rep_1, Rep_2, Rep_3)$ with $P(H| Rep_1, Rep_2, Rep_3, SR)$ and $P(Rel_1|Rep_1, Rep_2, Rep_3)$ with $P(Rel_1| Rep_1, Rep_2, Rep_3, SR)$ for high and low quality SR conditions.

A repeated measures ANOVA was conducted on belief in the hypothesis (P(H)) for the introduction of the shared reliability information (i.e. $P(H|Rep_1, Rep_2, Rep_3)$ to $P(H| Rep_1, Rep_2, Rep_3, SR)$), with the inclusion of the shared reliability condition (high/low-quality) as a between-subjects condition.

In line with predictions for the unlikely scenario, there was a substantial decrease in the belief that the claim at issue was true (economic crash), given the introduction of shared reliability (main effect of introduction), $BF_{Inclusion}$[7] = $5.05 * 10^8$, along with a main effect of shared reliability condition (low-quality < high-quality), $BF_{Inclusion}$ = 3193.73, demonstrating a successful manipulation check. Importantly, the substantial interaction of shared reliability condition, and its introduction, $BF_{Inclusion}$ = 209.81, revealed belief in the

[7] $BF_{Inclusion}$ shows the change in odds from the sum of the prior probabilities of models including the effect, to the sum of the posterior probabilities of models including the effect.

hypothesis decreased more substantially when the shared reliability was low-quality. Consequently, the model with all the above terms included was the best fit, BF_M[8] = 209.81, and substantial overall, $BF_{10} = 6.16 * 10^{10}$. We observe the same effects for revision of reliability estimates. There is not only a general decrease in reliability estimates given the introduction of a shared reliability, $BF_{Inclusion} = 3.64 * 10^9$, but also a main effect of shared reliability condition (low-quality < high-quality), $BF_{Inclusion}$ = 756483.62, with more substantial decreases in estimated reliability when the introduced shared reliability is low-quality, $BF_{Inclusion}$ = 9987.20. Once again, the model with all the above terms included was the best fit, BF_M = 9987.2, and substantial overall, $BF_{10} = 2.57 * 10^{12}$.

The above analyses were then repeated for the likely scenario, where, in line with predictions, the belief in the hypothesis (salmon growth) was found to substantially decrease when a shared reliability is introduced, $BF_{Inclusion}$ = 210.95. When shared reliability was low-quality, belief in the hypothesis was substantially lower than when shared reliability was high-quality, $BF_{Inclusion}$ = 1031.85, demonstrating a successful manipulation check. Lastly, the substantial interaction of shared reliability introduction, and its quality, $BF_{Inclusion}$ = 347.93, indicated that the decrease in belief in the hypothesis was local to the low-quality shared reliability condition only. The model with the above terms included was the best fit, BF_M = 347.93, and substantial overall, BF_{10} = 5394.15.

We again observe the same trends for the updating of source reliability given the introduction of a shared reliability. More precisely, the introduction of shared reliability is found to decrease estimations of source reliability, $BF_{Inclusion} = 1.24 * 10^{10}$, and whether

[8] BF_M shows the change from prior to posterior odds, given the model.

a shared reliability was high or low-quality led to higher or lower reliability estimates (respectively), $BF_{Inclusion} = 4.64 * 10^{10}$, once more passing the manipulation check. Critically, once more reductions in reliability, given the introduction of a shared reliability among sources, is found to be localized to when the introduced shared reliability is of low-quality (right-hand facet, Fig. 4), $BF_{Inclusion} = 2.71 * 10^{7}$. Finally, the model with the above terms included was the best fit, $BF_{M} = 2.71 * 10^{7}$, and substantial overall, $BF_{10} = 1.65 * 10^{14}$.

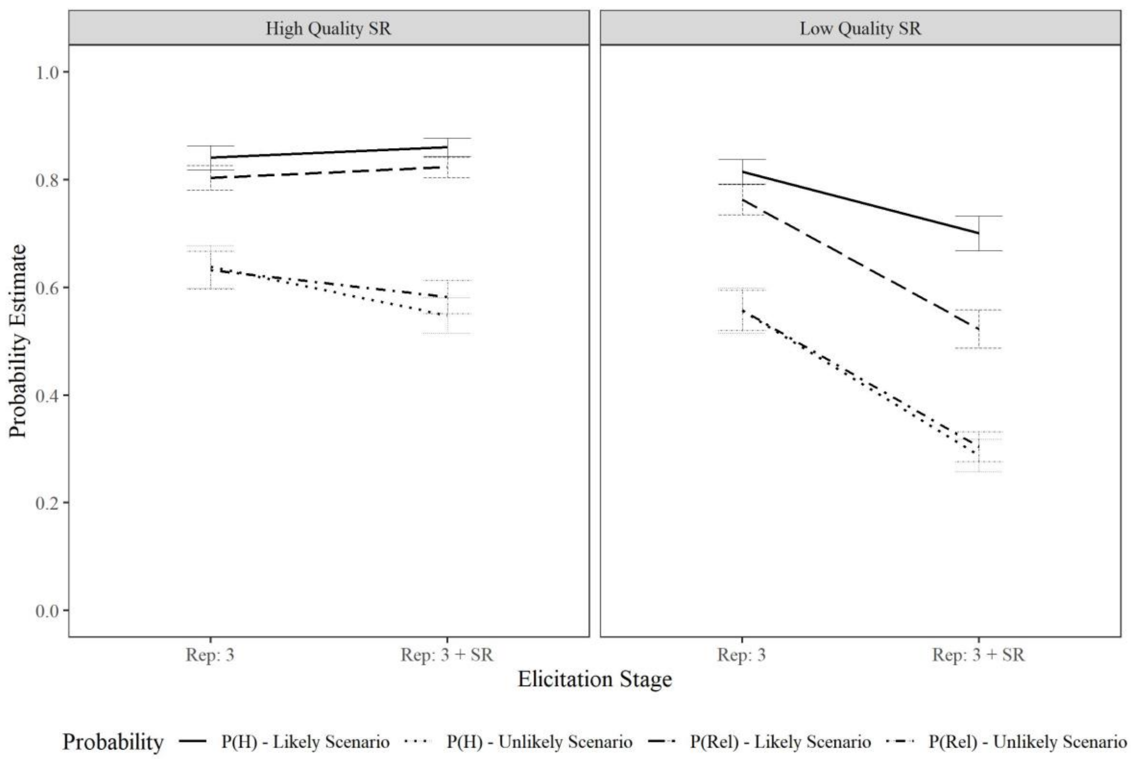

Figure. 4: Mean participant estimate of the probability that the claim after the third testimonial report, $P(H|Rep_1,Rep_2,Rep_3)$ along with the mean estimate of the reliability of source 1 after the third testimonial report $P(Rel_1|Rep_1,Rep_2,Rep_3)$. Mean participant estimate of the probability that the claim after learning sources are dependent, P(H| Rep_1,

Rep_2, $Rep_{3,}$SR) along with the mean estimate of the reliability of source 1 after learning sources are dependent P(Rel_1| Rep_1, Rep_2, $Rep_{3,}$SR), split by SR condition (facet) and scenario (linetype; P(H) = .81 and P(H) = .32). Error bars reflect +/- 1 Standard Error. The y-axis shows estimated probabilities expressed in decimal form (0-1). Data are plotted separately for the high- and low likelihood condition

To test whether participants adjust reliability estimates of sources retrospectively, or if additional reports only reflect on the most recent sources (hypothesis 4), we conducted a repeated measures ANOVA across the source reliability estimates of all sources. Specifically, we compared participant reliability estimates of $source_1$ after all reports have been provided. This revealed a null difference[9] in estimated reliabilities between sources in the unlikely scenario (N = 100), BF_{10} = 0.167, and likely scenario (N = 100), BF_{10} = 0.066. This suggests people update their belief in $source_1$ retroactively given new reports from additional sources, even if $source_1$ does not contribute with additional reports, in line with Bayesian model predictions.

Experiment 1: Discussion

Broadly, the results from Experiment 1 support the main hypotheses that participants are appropriately sensitive to the likelihood of a hypothesis, the impact of single confirmatory reports, and subsequent corroboration (both in terms of the updated likelihood of the hypothesis, and the updated reliabilities of reporting sources). Importantly, when specific observed instances of shared reliability are introduced (e.g.,

[9] The use of Bayesian statistics allowed us to infer the strength for the null hypothesis, whereby BFs less than 1/3rd may be considered substantial support for the null (Dienes, 2014).

high- or low-quality schooling), participants update their models of the world appropriately. This suggests participants can handle reasoning from specific observations of shared reliability.

Notably, in Exp. 1 participants are asked to revise their beliefs in light of a particular value for the shared reliability variable. This leaves open the question of whether lay reasoners would be sensitive to the structural dependency *per se*. Experiment 2 builds upon this by introducing to participants only the structure (i.e. possibility) of shared reliability, such that participants must evaluate whether this shared dependence influences their model of the world (relative to the independent equivalent) in and of itself. Put another way, does learning that the experts went to the same school lead to an appropriate revision of reliabilities and the hypothesis, *despite not knowing the quality of that schooling*?

Experiment 2: Method

Experiment 2 tests the introduction of shared reliability structure, without specific observations of its quality. As such, Experiment 2 follows the method of Experiment 1, except for the manipulation of shared reliability versus independence (see below).

Participants: As in Experiment 1, 100 participants (64 female. M_{age} 35.19, SD = 12.77), subject to the same selection criteria, were recruited from Prolific Academic. Median completion time was 5.87 min (SD = 2.41) and participants were paid £1.50 (resulting in an effective hourly wage of £15.33/hour for participation).

Experiment 2: Materials and procedure

The materials and procedure were identical to Experiment 1 for all aspects aside from the shared reliability/independence manipulation. In Experiment 1, the description of shared reliability contained a valuation of the excellence of the school (excellent or bad reputation). Here, having read the same initial reports as in Experiment 1, participants saw the following:

"It turns out, all the interviewed economists [biologists] studied at the **same [different]** school and subscribe to the same economic theories. You do not yet know whether this school has **either** a bad reputation for poor teaching and out-dated approaches to the economy, **or**, a good reputation for excellent teaching and up-to-date approaches to the economy.

Given the fact that they all studied at the **same** [**different]** school and follow the same economic theories, how likely is the UK stock market to crash within the next 6 months?"

Following this, posterior beliefs, ($P(H|Rep_1)$ and $P(H|Rep_1,\ldots,Rep_N)$),were elicited in the same way as in Experiment 1. Participants either saw only the shared dependency condition or the independent condition.

Experiment 2: Results

The probability manipulations were successful in generating high and low estimates for the two scenarios: The market crash scenario was rated as unlikely (M = .28, SD = .24) and the salmon growth scenario was rated as likely (M = .79, SD = .13), replicating Experiment 1. Participants rated both sources in our scenarios higher ($P(Rel_{Economist})$: M = .56, SD = .20; $P(Rel_{Biologist})$: M = .73, SD = .15). Importantly, both sources were once again rated positively, which allowed us to test whether positive reports of unlikely hypotheses influenced reliability estimates negatively.

As in Experiment 1, to test whether participants revise their belief in the reliability of the source in line with Bayesian predictions (hypothesis 1), we explore if participants adjusted reliability estimates in the initial source given sequential testimonies. Bayesian predictions dictated that positive reports of an unlikely hypothesis should initially decrease estimates of reliability.

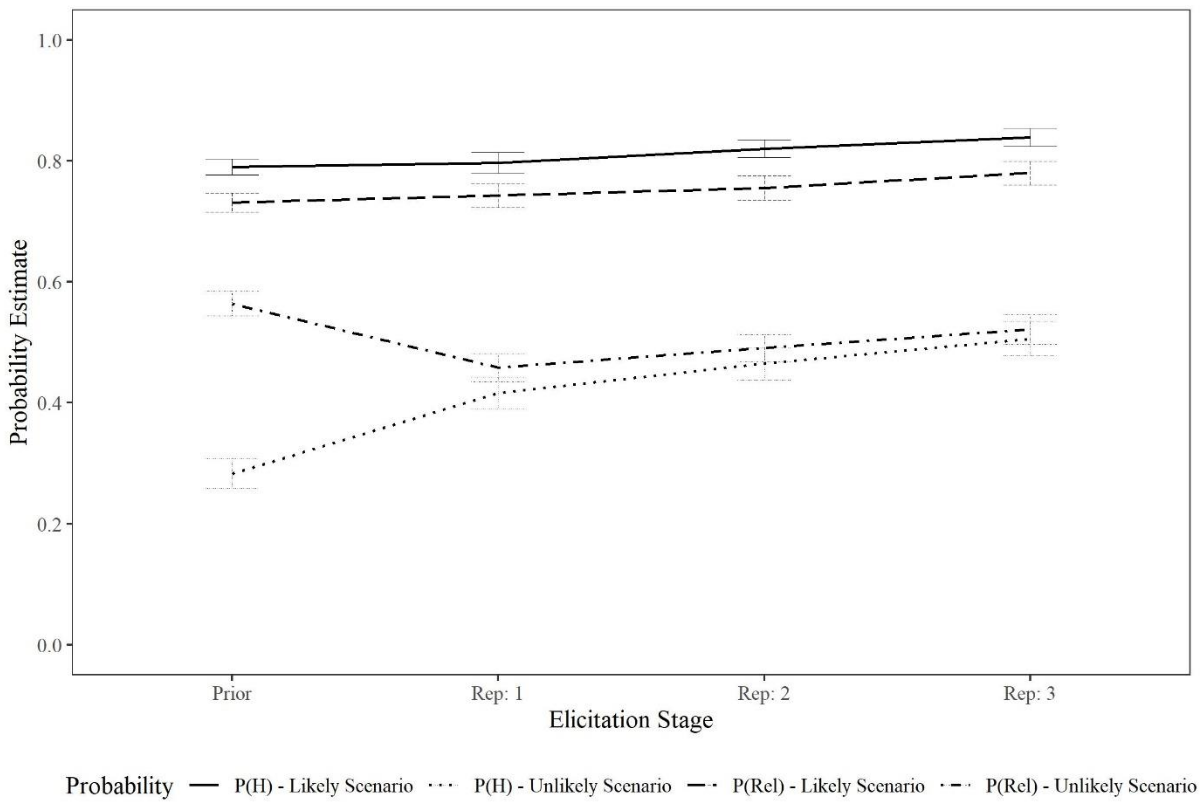

Figure. 3. Displayed are the mean participant estimate of the probability that the claim, H, is true, before receipt of the reports and then after receipt of each of the three, sequentially received, testimonial reports, along with the mean estimate of the reliability of source 1, Rel_1, prior to any reports, $P(Rel_1)$, and then after each additional report (i.e., $P(Rel_1|Rep_1)$, $P(Rel_1|Rep_1,Rep_2)$ and $P(Rel_1|Rep_1,Rep_2,Rep_3)$) both for likely, (P(H) = .79,

and unlikely, $P(H) = .28$), scenarios. Error bars reflect +/- 1 Standard Error. The y-axis shows estimated probabilities expressed in decimal form (0-1). Data are plotted separately for the high- and low likelihood condition

In line with predictions, using a repeated measures ANOVA ($P(Rel_1)$ – $P(Rel_1|Rep_1)$)[10], the perceived reliability of source 1 is revised negatively (i.e. decreases compared with baseline reliability) given a positive report of an unlikely hypothesis (N = 100), $BF_{10} = 352.67$ (in the current design, the source predicts the stock market will crash within a 6-month period). However (and also in line with Bayesian model predictions), as participants learn other experts provide corroborating reports ($P(Rel_1|Rep_1)$ – $P(Rel_1|Rep_1, Rep_2, Rep_3)$), they revise their belief in the initial source and revise reliability in a positive direction (N = 100), $BF_{10} = 73321.36$, again replicating Experiment 1.

In addition, we again find that participants *increase* their belief in the likelihood of the hypothesis whilst they simultaneously *decrease* belief in the source reliability ($P(H)$ to $P(H|Rep_1)$; N = 100), $BF_{10} = 2.43 * 10^6$.

To test whether sources that provide positive statements for highly likely hypotheses increase their reliability (hypothesis 2), we conducted a repeated measures ANOVA ($P(Rel_1)$ to $P(Rel_1|Rep_1, Rep_2, Rep_3)$) to see if the reports change the estimation of the reliability of the biologist. We find a substantial increase in reliability as more reports are given (N = 100), $BF_{10} = 1269.77$, replicating Experiment 1.

[10] Where $P(Rel_1) = P(Rel_{Profession})$, that is, the prior reliability of source 1 is the generic prior for that type of source (economist or biologist).

The introduction of shared reliability

To test whether source dependence impacts reliability estimations (hypothesis 3), we compared posterior degrees of belief in the hypothesis and the reliability of the sources before and after a shared reliability (SR) is introduced. As described above, we manipulate the reliability of the SR and compare $P(H|Rep_1, Rep_2, Rep_3)$ with $P(H| Rep_1, Rep_2, Rep_3, SR)$ and $P(Rel_1|Rep_1, Rep_2, Rep_3)$ with $P(Rel_1| Rep_1, Rep_2, Rep_3, SR)$ when that SR is an indication of *shared* or *independent* (different) schooling, with the former predicted by Bayesian models to reduce belief in the reported hypothesis and estimated reliability of sources.

A repeated measures ANOVA was conducted on belief in the hypothesis (P(H)) for the introduction of the shared reliability information (i.e. $P(H|Rep_1,Rep_2,Rep_3)$ to $P(H| Rep_1, Rep_2, Rep_3,SR)$), with the inclusion of the shared reliability condition (shared/independent) as a between-subjects condition.

In the unlikely scenario, we do not observe significant main effects for the introduction of additional reliability information, or whether that reliability was shared or different. However, focusing on the introduction of *shared* reliability information alone (N = 51), there is a significant reduction in belief in the hypothesis, BF_{10} = 23.19. This reduction is not observed in the condition where the additional reliability information indicates sources went to different schools (N = 49), BF_{10} = 0.22.

There is a general decrease in reliability estimates given the introduction of additional reliability information, $BF_{Inclusion}$ = 31.01, but no main effect of shared reliability condition or a significant interaction. The model with all terms included was the best fit, BF_M = 3.28, and strong overall, BF_{10} = 41.22. Following the same split as above, we note that it is the introduction of shared reliability information alone that drives the reduction in reliability estimates (N = 51), BF_{10} = 105.17, whilst the introduction of different school reliability information leads to no such reduction (N = 49), BF_{10} = 0.32.

Taken together, we find evidence in support of hypothesis 3, that the introduction of the possibility of shared reliability (i.e., source dependence), leads to a decrease in estimates of reliability and belief in the hypothesis.

The above analyses were then repeated for the likely scenario, where, in line with predictions, the belief in the hypothesis (salmon growth) was found to substantially decrease when additional reliability information is introduced, $BF_{Inclusion}$ = 657.1. When this information indicated a shared reliability, belief in the hypothesis was substantially lower than when this information indicated different schooling, $BF_{Inclusion}$ = 80.5. The model with all terms included was the best fit, BF_M = 120.39, and strong overall, BF_{10} = 31290.43. As with the unlikely scenario, we confirm that the reduction in belief in the hypothesis is driven by the introduction of *shared* reliability information (N = 51), BF_{10} = 1621.75, whilst there is no evidence for a reduction in belief in the hypothesis as a consequence of introducing *different* reliability information (N = 49), BF_{10} = 0.3.

Lastly, the additional reliability information was found to decrease reliability estimates, $BF_{Inclusion}$ = 250.70, but significantly more so when that information indicated a shared reliability, $BF_{Inclusion}$ = 4.04. The model with the above terms included was the best fit, BF_M = 5.68, and substantial overall, BF_{10} = 545.55. Again, as in the unlikely scenario, we confirm that the reduction in reliability estimates is driven by the introduction of *shared* reliability information (N = 51), BF_{10} = 91.13, whilst there is no evidence for a reduction reliability estimates as a consequence of introducing *different* reliability information (N = 49), BF_{10} = 0.67.

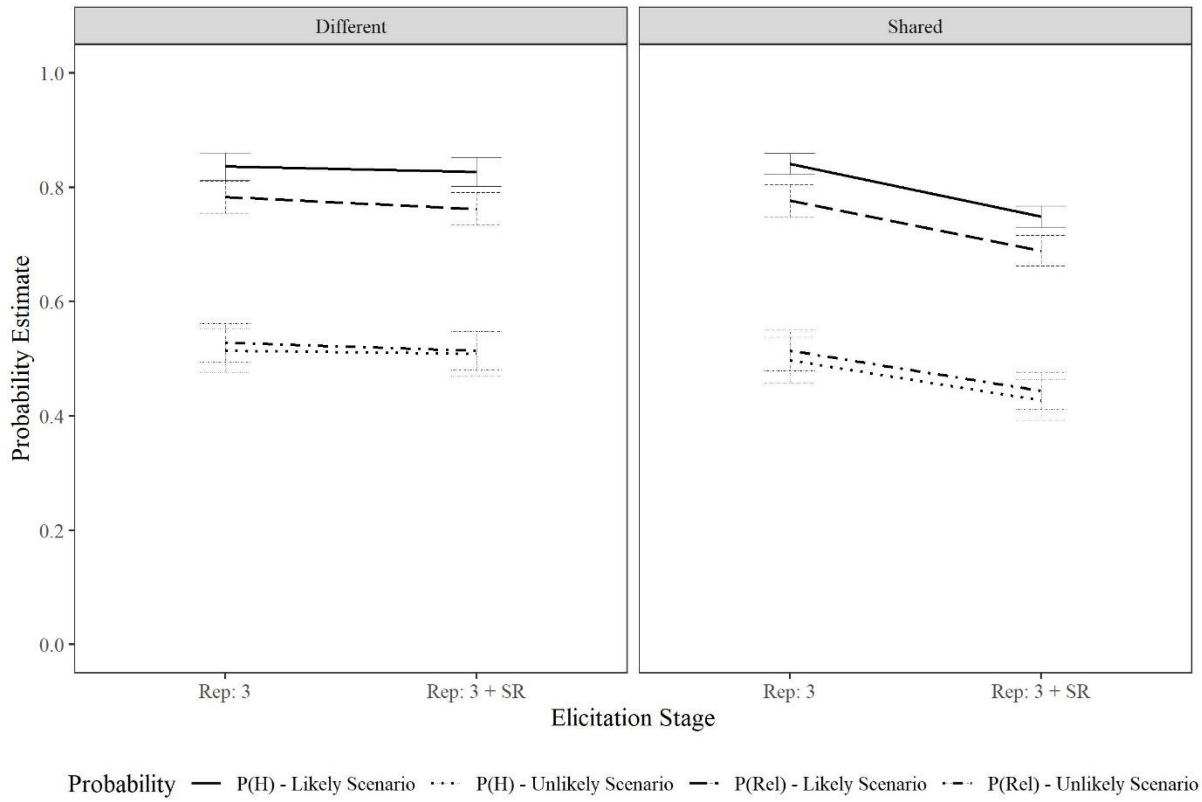

Figure. 4: Mean participant estimate of the probability that the claim after the third testimonial report, $P(H|Rep_1,Rep_2,Rep_3)$ along with the mean estimate of the reliability of source 1 after the third testimonial report $P(Rel_1|Rep_1,Rep_2,Rep_3)$. Mean participant estimate of the probability that the claim after learning sources are dependent, $P(H| Rep_1, Rep_2, Rep_3,SR)$ along with the mean estimate of the reliability of source 1 after learning sources are dependent $P(Rel_1| Rep_1, Rep_2, Rep_3,SR)$, split by SR condition (facet) and scenario (linetype; P(H) = .79 and P(H) = .28). Error bars reflect +/- 1 Standard Error. The y-axis shows estimated probabilities expressed in percentages. Data are plotted separately for the high- and low likelihood condition

Finally, as in Experiment 1, to test whether participants adjust reliability estimates of sources retrospectively, or if additional reports only reflect on the most recent sources (hypothesis 4), we conducted a repeated measures ANOVA across the source reliability estimates of all sources. Specifically, we compared participant reliability estimates of $source_1$ after all reports have been provided. This revealed a null difference in the estimated reliability of sources in the unlikely scenario (N = 100), BF_{10} = 0.067, but a minor – yet significant – difference in the likely scenario (N = 100), BF_{10} = 12.4. Although imperfect, this is again suggestive of people updating their belief in $source_1$ retroactively given new reports from additional sources, even if $source_1$ does not contribute with additional reports, in line with Bayesian model predictions.

Discussion and concluding remarks

The paper explores how sequential testimonies and partial dependence modulates reliability estimates of sources. Across two experiments we explored four hypotheses:

First, we tested whether participants revised their posterior degree of belief in the reliability of sources in line with Bayesian predictions. The data supports this prediction, as P(Rel) initially decreased given a positive report of an unlikely hypothesis, but subsequently increased as more positive reports were observed.

Second, we tested whether sources that provide positive statements for highly likely hypotheses increase their reliability, but to a lesser extent. The data provides indicative support for this, as reliability of sources remained almost constant in the scenario with a likely prior probability.

Third, we tested whether the introduction of a partial source dependence (i.e., shared background) led to an adjustment in reliability estimates in line with Bayesian predictions. The data provides support for this hypothesis. In Experiment 1, we showed that participants could reason appropriately from specific observations of a shared background: When participants learned the experts attended the same school, they adjusted their posterior degree of belief negatively, both for the hypothesis and the reliability of each source. In line with expectations, this effect was stronger if the experts' school was of low-quality as compared to high-quality. In Experiment 2, we extended this finding to show that participants can appropriately reason from the introduction of a shared reliability form of dependence – irrespective of an observed value (i.e., revising estimates appropriately after learning experts attended the *same* school, without yet knowing the quality of that school). Put another way, we find participants are appropriately sensitive to the introduction of a shared reliability dependency *relation*.

Finally, we tested if participants revised their posterior degree of belief in the reliability of sources retrospectively. The data supports this hypothesis, as $source_1$ initially decreased when reporting an unlikely hypothesis. Yet, as $sources_{2\text{-}3}$ provided similar reports, reliability of $source_1$ was adjusted in line with the n^{th} source (enjoying the same reliability as $source_2$ after 2 reports, *mutatis mutandis* with $source_{2\text{-}3}$ after 3 reports). Overall, the data provides preliminary empirical evidence that lay reasoners are, at least qualitatively, sensitive to key features of the dynamics of the Bayesian model predictions (Bovens and Hartmann, 2003)

Interestingly, along with the revision of reliability, participants also revised beliefs in the hypothesis along expected lines, increasing their degrees of belief when a reliable source provides a positive report for the hypothesis – this is in line with previous Bayesian experiments on the impact of source credibility (see e.g. Hahn et al., 2009; Harris et al., 2015; Madsen, 2016). Experiment 1 and 2 both support the above hypotheses and interpretations.

Future work

In our experimental conditions, participants are told that sources have a shared background (shared reliability). However, sources may be dependent in other ways. For example, they may communicate directly and reach consensus before providing their reports (e.g. a jury does so) or dependency may be one-way (source n can see the reports of source n-1 before making her statement whilst source n-1 cannot see the reports of subsequent sources, see e.g. Pilditch, Hahn, Fenton & Lagnado, under review; Pilditch, Hahn & Lagnado, 2018; 2019). Future work should explore how different types of dependencies influence the beliefs of the recipient with regards to claim and the reliability of the sources.

The current work provides experimental corroboration of key qualitative features of reliability and dependency found in the normative, Bayesian model. Future work could use the method employed here to test reliability updating given a much wider range of social and information structures, a wider range of hypotheses, different signal strength, and differences in shared reliability. In addition, sources in the present study were domain-general (biologists and economists). Future work might interrogate the degree domain-specific expertise functions (e.g. biologists who specialize in Norwegian salmon compared with more generic expertise markers). Finally, in the current design, participants were told of the shared reliability after having seen (supposedly) independent reports. Future studies might manipulate when and how participants experience the shared reliability between sources. The findings shed light on how people gauge source reliability and integrate reports when multiple sources weigh in on an issue as seen in public debates.

Author Note

The research reported in this paper was supported by the Alexander von Humboldt Foundation's Anneliese Maier Research Award to UH. We have registered and made the data available on the OSF (project id: DOI 10.17605/OSF.IO/X64TH; project title: "The Impact of Partial Source Dependence on Belief and Reliability Revision".)

References

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.

Cialdini, R. B. (2007). *Influence: The Psychology of Persuasion*, Collins Business.

Collins, P.J. & Hahn, U. (in press). We might be wrong, but we think that hedging doesn't protect your reputation. *JEP:LMC*.

Collins, P., Hahn, U., Gerber, Y. & Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content, *Frontiers in Psychology*, 9-18.

Corner, A., & Hahn, U. (2009). Evaluating science arguments: evidence, uncertainty, and argument strength. *Journal of Experimental Psychology: Applied*, *15*(3), 199.

Cuddy, A. J. C., Glick, P. & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations, *Research in Organizational Behavior* 31, 73-98.

Dienes, Z. (2014). Using Bayes to get the most out of nonsignificant results. *Frontiers in Psychology*, *5*, Article 781. doi:10.3389/fpsyg.2014.00781

Duggins, P. (2016). A Psychologically-Motivated Model of Opinion Chance with Applications to American Politics, *arXiv:1406.7770.*

Fiske, Susan T., Cuddy, A. J. C. & Click, P. (2007). Universal dimensions of social cognition: warmth and competence, *Trends in Cognitive Sciences*, 11 (2), 77-83.

Hahn, U., Harris, A. J. L., & Corner, A. (2009). Argument content and argument source: An exploration, *Informal Logic*, *29*, 337-367

Hahn, U., Harris, A. J., & Corner, A. (2016). Public reception of climate science: Coherence, reliability, and independence. *Topics in cognitive science*, *8*(1), 180-195.

Hahn, U. Merdes, C. & Sydow, M. (2018). How good is your evidence and how would you know it? *Topics in Cognitive Science*, 10 (4), 660-678.

Hahn, U., & Oaksford, M. (2006). A normative theory of argument strength, *Informal Logic*, 26, 1-24.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies, *Psychological Review*, *114*, 704-732.

Hahn, U., Sydow, M. & Merdes, C. (2019). How communication can make voters choose less well, *Topics in Cognitive Science*, 11, 194-206.

Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2015). The Appeal to Expert Opinion: Quantitative support for a Bayesian Network Approach. *Cognitive Science*, 40, 1496-1533.

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants, *Philosophical Transactions of the Royal Society B*, *366*, 1179-1187.

Howson, C., & Urbach, P. (1996). *Scientific Reasoning: The Bayesian Approach (2nd Edition).* Chicago, IL: Open Court.

Jarvstad, A. & Hahn, U. (2011). Source reliability and the conjunction fallacy, *Cognitive Science*, 1-30.

JASP Team (2019). JASP (Version 0.11.1)[Computer software].

Jeffreys, H. (1961). *Theory of probability*, Oxford: Oxford University Press.

Kemp, C. & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning, *Psychological Review*, 116 (1), 20-58.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective, *Psychonomic Bulletin & Review, 25* (1), 178-206.

Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning, *Argument & Computation,* 4 (1), 46-63.

Lagnado, D., Gerstenberg, T. & Zultan, R. (2015). Causal responsibility and counterfactuals, *Cognitive Science,* 37 (6), 1036-1073.

Madsen, J. K. (2016). Trump supported it?! A Bayesian source credibility model applied to appeals to specific American presidential candidates' opinions, Papafragou, A., Grodner, D., Mirman, D., & Trueswell, J.C. (Eds.) *Proceedings of the 38th Annual Conference of the Cognitive Science Society,* Austin, TX: Cognitive Science Society, 165-170.

Madsen, J. K. (2020) Open Science Framework: The Impact of Partial Source Dependence on Belief and Reliability Revision (project id: DOI 10.17605/OSF.IO/X64TH)

Madsen, J. K., Bailey, R. M. & Pilditch, T. (2018). Large networks of rational agents form persistent echo chambers, *Scientific Reports* 8:12391, 1-8

Madsen, J. K., Hahn, U. & Pilditch, T. (2018). Partial source dependence and reliability revision: the impact of shared backgrounds, In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.) *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, 720-726

Madsen, J. K. & Pilditch, T. (2018). A method for evaluating cognitively informed micro-targeted campaign strategies: An agent-based model proof of principle, *PLoS One,* 13 (4), e0193909, 1-14.

Oaksford, M. & Chater, N. (1991). Against logicist cognitive science, *Mind and Language*, 6, pp. 1-38.

Oaksford, M. & Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection, *Psychological Review*, 101 (4), 608-631.

Oaksford, M. & Chater, N. (2007). *Bayesian Rationality: The probabilistic approach to human reasoning*. Oxford, UK: Oxford University Press.

Olsson, E. J. (2005). Against coherence: Truth, probability, and justification. Oxford University Press.

Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme*, *8*(2), 127-143.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*, Morgan Kaufmann.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Petty, R. E. & Cacioppo, J. T. (1984). Source Factors and the Elaboration Likelihood Model of Persuasion, *Advances in Consumer Research*, 11, 668-672.

Philipps, K., Hahn, U. & Pilditch, T.D. (2018). Evaluating testimony from multiple witnesses: single cue satisficing or integration? *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.

Pilditch, T. D., Hahn, U., Fenton, N., & Lagnado, D. (*under review*). Dependencies in evidential reports: The case for informational advantages.

Pilditch, T.D., Hahn, U. & Lagnado, D. (2018). Integrating dependent evidence: naïve reasoning in the face of complexity**.** *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.

Pilditch, T.D., Hahn, U., & Lagnado, D. (2019). Shared Evidence: It all depends... In Goel, A., Seifert, C., & Freska, C. (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2571-2577). Austin, TX: Cognitive Science Society.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34, 243-281.

Schum, D. A. (1994). *The evidential foundations of probabilistic reasoning*, Northwestern University Press.

Shafto, P., Goodman, N.D., & Frank, M.C. (2012). Learning from others: The consequences of psychological reasoning for human learning, *Perspectives on Psychological Science, 7,* 341-351.

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology*, *38*(2), 317-346.

Tormala, Z. L. & Jackson, J. J. (2007). Assimilation and Contrast in Persuasion: The Effects of Source Credibility in Multiple Message Situations, *Personality and Social Psychology Bulletin*, 33 (4), 559-571.

Vallinder, A. & Olsson, E. J. (2014). Trust and the value of overconfidence: A Bayesian perspective on social network communication, *Synthese*, 191, 1991-2007

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p-values, *Psychonomic Bulletin & Review*, *14* (5), 779-804.

Yousif, S. R., Aboody, R. & Keil, F. C. (2019). The Illusion of Consensus: A Failure to Distinguish Between True and False Consensus, *Psychological Science*, 30 (8), 1195-1204.