Walter the banker: the conjunction fallacy reconsidered

Stephan Hartmann · **Wouter Meijs**

Abstract   In a famous experiment by Tversky and Kahneman (Psychol Rev 90:293–315, 1983), featuring Linda the bank teller, the participants assign a higher probability to a conjunction of propositions than to one of the conjuncts, thereby seemingly committing a probabilistic fallacy. In this paper, we discuss a slightly different example featuring someone named Walter, who also happens to work at a bank, and argue that, in this example, it is rational to assign a higher probability to the conjunction of suitably chosen propositions than to one of the conjuncts. By pointing out the similarities between Tversky and Kahneman's experiment and our example, we argue that the participants in the experiment may assign probabilities to the propositions in question in such a way that it is also rational for them to give the conjunction a higher probability than one of the conjuncts.

Keywords   Conjunction fallacy · Linda problem · Psychology of reasoning · Bayesian epistemology

1 Introduction

In recent years, a famous experiment of Tversky and Kahneman (1983) has attracted the attention of formal epistemologists.[1] In the experiment participants are provided

---

[1]   See Crupi et al. (2008) and the contributions to this special issue. For a recent review of the cognitive science literature on the conjunction fallacy, see Hertwig et al. (2008). For a recent review of Bayesian epistemology, see Hájek and Hartmann (2009).

S. Hartmann (✉) · W. Meijs
Tilburg Center for Logic and Philosophy of Science, Tilburg University,
P.O. Box 90153, 5000 LE, Tilburg, The Netherlands
e-mail: S.Hartmann@uvt.nl
URL: http://www.stephanhartmann.org

W. Meijs
e-mail: woutertjuh@hotmail.com
URL: http://www.uvt.nl/tilps/people/rESEARCHaSSOCIATES/#meijs

with information about a woman named Linda: *Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.* The participants are then asked to rank the following statements in terms of their probability:

  i) Linda is a bank teller.
 ii) Linda is bank teller and Linda is a feminist.

In the experiment, Tversky and Kahneman found that the majority of the participants ranked the second statement as more probable than the first. As it is a law of probability that the probability of the conjunction of two propositions cannot be larger than the probability of any of the conjuncts, it seems inevitable to conclude that in some cases people behave in violation of the laws of probability.

However, the fact that a conjunction can never be more probable than one of its conjuncts only holds in a *ceteris paribus* sense: that is, we should assume that all relevant factors are equal in the case that we compare the probability of the conjunction to the probability of one of its conjuncts. That this is not a mute point is shown by the recent discussion in epistemology about whether coherence can be a truth conducive property of an information set. In their (1994), Klein and Warfield use the same theorem of the probability calculus to show that coherence cannot be truth conducive in the sense that a more coherent information set will also have a higher probability. For, as they argue, in some cases the coherence of an information set can increase by adding a proposition to the set. But as a conjunction cannot be more probable than any of its conjuncts, the new set cannot be more probable and, therefore, coherence cannot be truth conducive.

However, as Bovens and Olsson (2000) have pointed out, Klein and Warfield's argument no longer works if we do not consider the propositions on their own, but instead as reports by different sources. For if each proposition is reported by a source, then the background conditions in the case of the conjunction are different from those in case of each of its conjuncts: in the first case we have two reports, while in the second only one. And then the condition that all relevant factors must be equal no longer holds. Since then, this approach to analyzing the relation between coherence and probability has been further elaborated upon in various other publications.[2]

In this paper we will apply the same approach to a thought experiment that is similar to Tversky and Kahneman's Linda experiment. We will show that for this example at least, we can present a model similar to the one used in the coherence discussion according to which the participants are rational in preferring the conjunction over one of its conjuncts. The example we present will be slightly different from the Linda example. So the conclusion that the participants are rational in our example does not immediately tell us anything about the Linda example. Nonetheless, we will argue in the last two sections that the two examples are much more similar than one may have imagined.

---

[2] See, for instance, Bovens and Hartmann (2003) and Meijs (2006, 2009). Meijs and Douven (2007) and Olsson (2005) argue that coherence is not truth conducive, even if Klein and Warfield's dismissal was too quick.

2 Walter the banker

Consider the following situation:

> Suppose you are a philosopher who knows a little, but not much about the current status of formal epistemology. Suppose also that you are intimately connected with the university of Leuven, and that through this you know that there will be a conference on probabilistic paradoxes next week. Suppose that on one morning you meet a stranger in the commuter train to his work and you start talking. Now suppose that when the stranger finds out that you are a philosopher, he starts talking about an acquaintance of his, who happens to be a philosopher as well. In summary, you learn that this guy's name is Walter, that he is 31 years old, outspoken, bright, with a background in philosophy and physics ("would you believe that?!"—the stranger exclaims) and did some work on probabilistic modeling. The stranger continues to make some comments about the merits of formal epistemology in general and then proceeds to inform you that, by the way, Walter is a banker, working as a manager for one of the world's largest financial services providers. Now suppose you arrive at your stop and after saying goodbye, you leave the train. Call this scenario 1.
> Now consider scenario 2, which is like scenario 1, but before you say your goodbyes, the stranger mentions one more fact: that Walter will be speaking at a conference in Leuven next week.

Clearly, this example is quite similar to the Linda example. But there are also a number of important differences between the two examples. First, in the Walter example we have a clear source who reports the information and the credibility we assign to that source is crucial for determining the probability that what he reports is true.

Second, there is a difference in how we should treat the background information. In the Linda case, the information provided about Linda is presented as being beyond doubt – the participants should assign unit probability to these statements. In the current example, this is not the case: the information provided about Walter being bright, outspoken, etc., is just as much part of the story as the rest of the information.

Third, intuitively, it is no longer clear whether it is irrational to assign a higher probability to the information reported in the second scenario of the Walter example than to the information reported in the first. And, as the next section will show, the laws of probability may actually support this statement.

3 Walter formalized

It goes without saying that we will not propose to change one of the axioms of probability theory in this paper. But as we saw, the fact that the probability of a conjunct cannot be higher than the probability of any of its conjuncts only holds *ceteris paribus*. In the Walter example there is an explicit mention of the source that reports the information, the "stranger" that you meet in the train. And when we condition on the reports of the stranger, it no longer needs to be the case that the probability of the conjunct is equal to or larger than the probability of each of the conjuncts.

To formalize this, we first define the following propositions:

H_1: Walter is 31 years old, outspoken, bright, and has a background in physics and philosophy.
H_2: Walter works at a bank.
H_3: Walter attends a conference on formal epistemology in Leuven.

These propositions represent *facts*. Note that the first scenario is represented by the set $\mathcal{S}_I := \{H_1, H_2\}$ and the second scenario is represented by the set $\mathcal{S}_{II} := \{H_1, H_2, H_3\}$.

Next, we define propositions that represent *reports*: R_i is the proposition that the stranger reports that H_i is the case. For example, R_2 stands for "The stranger reports that Walter works at a bank".

We will then show that it can be the case that

$$p(H_1, H_2, H_3 \mid R_1, R_2, R_3) > p(H_1, H_2 | R_1, R_2) \tag{1}$$

for plausible probability distributions.[3] Hence, the probability of the conjunction of three propositions can be larger than the probability of two of the conjuncts, if one conditions on the corresponding reports.

The value of a posterior probability, such as $p(H_1, H_2 | R_1, R_2)$, depends critically on the credibility of the witness. If we know, for instance, for certain that the witness always and only report the truth, then the posterior probability will be unity. If, on the other hand, we know that the witness is fully untrustworthy, then the posterior probability will equal the prior probability $p(H_1, H_2)$: the reports do not tell us anything about the facts in question.

Moreover, if the reports of the witness are fully dependent on one another, then the second, third, etc, witness reports will not add anything to the value of the posterior probability. The more independent the witness reports are, the larger the posterior probability, *ceteris paribus*.

In the remainder of this section we will construct a model that allows us to study the case at hand in detail. To do so, we propose an account of witness credibility and show how different independent reports of one partially reliable witness can be combined. We will then show that for plausible probability distributions over the different propositions, we will be more justified in believing information set $\mathcal{S}_{II}$ than information set $\mathcal{S}_I$.

It proves useful to introduce the following binary propositional variables: For $i = 1, \ldots, n$, let us call H_i (in italics) a *fact variable* and R_i a *report variable*. H_i can take two values: H_i and $\neg H_i$ (both in roman script). Similarly, R_i can take two values: R_i and $\neg R_i$. R_i is the proposition that, after consultation with the proper source, there is a report to the effect that H_i is the case. $\neg R_i$ is the proposition that, after consultation with the proper source, there is no report to the effect that H_i is the case.

---

[3] Here $p(A, B)$ denotes the probability of the conjunction of the propositions A and B. $p(A)$ is the prior probability of A, and $p(A|B)$ is the posterior probability of A given B.

Witness credibility. There are various ways to model the credibility of a witness.[4] Following the 'Scandinavian School of Evidentiary Value' (see Hansson 1983), we assume that a witness is either reliable or unreliable. If the witness is reliable, he is a truth-teller. If the witness is unreliable, learning about his report does not teach us anything about the hypothesis. In this case, the witness randomizes.

To formalize this, we introduce a further binary propositional variable: The reliability variable Rel, which can take two values: Rel is the proposition that the witness is reliable (and, hence, a truth-teller), and ¬Rel is the proposition that the witness is not reliable (and, hence, a randomizer). We then set for all $i, j = 1, \ldots, n$:

$$p(\mathrm{R_i|H_j, Rel}) = \delta_{ij}, \quad p(\mathrm{R_i|H_j, \neg Rel}) = a_{ij} \qquad (2)$$

The parameter a_{ij} is called the randomization parameter.[5] Following Olsson (2005), we set for all $i, j = 1, \ldots, n$:

$$a_{ij} = p(\mathrm{H_j}) \qquad (3)$$

This choice makes sense in our case: Without any additional information, a witness will randomize on the basis of his prior beliefs (see also Bovens and Hartmann 2003, ch. 5).

Independence conditions. Next, we formalize what it means that different reports are probabilistically independent. To do so, we introduce further propositional variables.

We make the following assumptions: First, each fact variable is independent of the credibility of the witness. In formal notation and for all $i, j = 1, \ldots, n$:

$$H_i \perp\!\!\!\perp Rel \qquad (4)$$

In terms of probabilities, this requirement amounts to:

$$p(\mathrm{H_i|Rel}) = p(\mathrm{H_i}) \qquad (5)$$

Intuitively, this assumption makes sense: learning something about the credibility of a witness should not tell us anything about the probability of a proposition without a report by the witness.

Second, the witness report is independent of all other witness reports, given the (values of the) corresponding fact and the reliability variables. Put formally, we consider the case that the witness makes *n* reports and assume for all *i*,

$$R_i \perp\!\!\!\perp R_1, \ldots, R_{i-1}, R_{i+1}, \ldots R_n | H_i, Rel. \qquad (6)$$

---

[4] See Bovens and Hartmann (2003) for a discussion of several options.

[5] The "Kronecker delta" δ_{ij} equals 1 if $i = j$ and 0 otherwise.

In terms of probabilities, this condition amounts to

$$p(R_i|R_1, \ldots R_{i-1}, R_{i+1}, \ldots, R_n, H_i, Rel) = p(R_i|H_i, Rel). \tag{7}$$

For a similar condition, (see Bovens and Hartmann 2003: 65).

Note that the probabilistic versions of conditions (4) and (6) hold for all instantiations of the corresponding propositional variables (see Bovens and Hartmann 2003, ch. 3).

Our independence condition (6) may be somewhat difficult to read and the reader may wonder why we would not opt for a simpler proposal such as for all $i, j = 1, \ldots, n$,

$$R_i \perp\!\!\!\perp R_j \tag{8}$$

or, probabilistically,

$$p(R_i \,|\, R_j) = p(R_i). \tag{9}$$

However, this condition would be too strict since it would exclude the possibility that a report can increase the probability that the witness is a truth teller, which in turn may influence the probability of a next report. More specifically, it may increase or decrease it, depending on the prior probability of that proposition.[6]

There is also a more pragmatic defense of our independence condition (6). For the only reason to disagree with an independence condition is if it would be unreasonable to assume that the reports are independent in the sense specified by the condition. Therefore, we have an a priori reason to prefer a weaker condition over a stronger. The only reason not to drop the independence condition completely is that in that case we would never see an increase in probability at all.

The independence conditions (4) and (6) and our model for the witness credibility are captured by the Bayesian network represented in Fig. 1.

The posterior probability. In Appendix A we show that given the above conditions, we can derive an expression for the posterior probability $p^*(\mathcal{S}) := p(H_1, \ldots, H_n|R_1, \ldots, R_n)$ of an information set $\mathcal{S} := \{H_1, \ldots, H_n\}$ given a corresponding set of reports $\mathcal{R} := \{R_1, \ldots, R_n\}$. Although the derivation can be done without the theory of Bayesian networks, it turns out that the calculation simplifies considerably if we use this powerful tool. We obtain:

$$p^*(\mathcal{S}) = \frac{p(Rel) + p(H_0) \cdots p(H_n)\, p(\neg Rel)}{p(\mathcal{S})\, p(Rel) + p(H_0) \cdots p(H_n)\, p(\neg Rel)}\, p(\mathcal{S}) \tag{10}$$

With this expression we can determine the posterior probability of both scenarios involving Walter the Banker:

---

[6] This case is argued at length in Meijs (2009).

Fig. 1 The Bayesian network
representing our witness
scenario. The *dashed lines*
connecting the H_is represent any
arrows between the various fact
variables
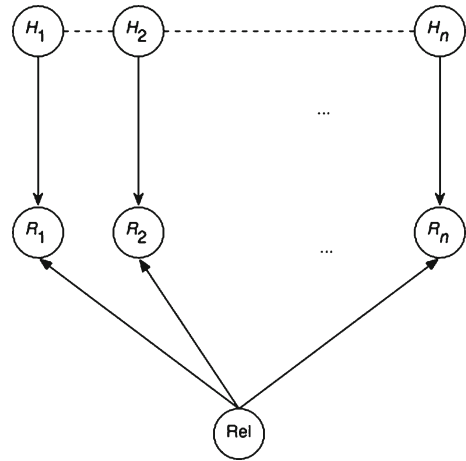
$$p^*(\mathcal{S}_I) = \frac{p(\text{Rel}) + p(H_1)\,p(H_2)\,p(\neg\text{Rel})}{p(H_1, H_2)\,p(\text{Rel}) + p(H_1)\,p(H_2)\,p(\neg\text{Rel})}\,p(H_1, H_2) \tag{11}$$

$$p^*(\mathcal{S}_{II}) = \frac{p(\text{Rel}) + p(H_1)\,p(H_2)\,p(H_3)\,p(\neg\text{Rel})}{p(H_1, H_2, H_3)\,p(\text{Rel}) + p(H_1)\,p(H_2)\,p(H_3)\,p(\neg\text{Rel})}\,p(H_1, H_2, H_3) \tag{12}$$

The difference function. Note that set $\mathcal{S}_{II}$ is more probable than set $\mathcal{S}_I$ if and only if the *difference function*

$$D(\mathcal{S}_I, \mathcal{S}_{II}) := p^*(\mathcal{S}_{II}) - p^*(\mathcal{S}_I) \tag{13}$$

is positive. The following theorem provides a necessary and sufficient condition for the positivity of the difference function.

Theorem $D(\mathcal{S}_I, \mathcal{S}_{II}) > 0$ *if and only if* $p(H_3|H_1, H_2) > p(H_3) \cdot (1 + \delta)$ *with the error measure*

$$\delta = \frac{p(H_1, H_2)\,p(\text{Rel}) + p(H_1)\,p(H_2)\,(1 - p(\text{Rel}))}{\left[1 - p(H_1, H_2)\,(1 - p(H_3))\right]\,p(\text{Rel}) + p(H_1)\,p(H_2)\,p(H_3)\,(1 - p(\text{Rel}))} \cdot (1 - p(H_3)).$$

Note that $\delta \ll 1$ if $p(H_1)$, $p(H_2)$ and $p(H_1, H_2)$ are small and if (approximately) $p(\text{Rel}) \geq .5$. See Appendix B for a detailed discussion and the proof of the theorem.

To understand the theorem and its relevance, note that the principle at work in our example is that H_1 and H_3 positively support each other, while H_1 and H_2 are negatively relevant for each other. That is, we have:

(a) $p(H_2 \mid H_1) < p(H_2)$
(b) $p(H_3 \mid H_1) > p(H_3)$

The main condition in the theorem (neglecting the error term),

(c) $p(H_3|H_1, H_2) > p(H_3)$,

states that the second set has a higher posterior probability than the first set if H_2 does not destroy the positive relevance between H_1 and H_3. We believe this is a plausible assumption in this case. That is, we believe that the relevance between the proposition that Walter is 31 years old, with a background in physics and philosophy etc. with the proposition that Walter will attend the epistemology conference is strong enough to not be completely destroyed by the report that he works at a bank. We show in Appendix B that our result can also be put in the following way: the second set has a higher posterior probability than the first set if adding H_3 to H_1 and H_2 increases the coherence, i.e. if $\mathcal{S}_{II}$ is more coherent than $\mathcal{S}_I$. The coherence increase has to be the larger, the smaller the witness credibility.

Summarizing this section, we conclude that given some plausible assumptions about the probabilities of the propositions in the Walter example, the second set does indeed have a higher posterior probability than the first. And anyone who does not believe that the second set is more probable than the first is irrational.

4 What Walter tells us about Linda

So far we have done a couple of things: We have introduced an example in which intuitively it makes sense to argue that we are more justified in believing a conjunction than one of its conjuncts. Furthermore, we have shown how taking the witness reports into account supports this conclusion. Finally, we have shown that given a number of assumptions about the witness and his reports, we can calculate that the posterior probability for the conjunction of the three propositions is indeed larger than the posterior probability of the two propositions. This led us to the conclusion that participants who do not assign the conjunction a higher probability are in fact irrational. The fallacy has completely disappeared.

In these final two sections, we would like to pursue the argument that just as in the Walter case, the participants in the Linda experiment, too, should assign a higher probability to the conjunction. We will do this by pointing out how the same model that we used above for Walter the Banker can also be applied to the Linda case.

Clearly, for this argument to succeed, we would need to show that both cases are similar in all relevant respects. We believe that the three assumptions we make about the probabilities (conditions (a), (b) and (c)), hold in both examples. Just replace the propositions H_1, H_2 and H_3 by the propositions H_1', H_2' and H_3' with

H_1': Linda is 31 years old, single, outspoken and very bright, etc.
H_2': Linda is bank teller.
H_3': Linda is a feminist.

Nonetheless, there are also important differences between the Walter and the Linda examples. We believe the two most important ones are:

(1)   In the Linda experiment, the background information about Linda is assumed to be certain while in the Walter example it is part of the analysis.
(2)   In the Linda experiment the statements about Linda are not presented as reports while in the Walter example they are.

Let us take up each of these in turn. First, the reason why we do not take the background information about Walter as given is that we do not fully trust the source. Clearly, in the Linda case the information is reported by a source as well: we have an "experimenter" who tells us certain things about Linda. However, in the Linda case the experimenter is supposed to be fully credible—the participants are supposed to take for granted anything he or she tells them is the case. So even if we were to agree to see the experimenter as a witness, we should assume this witness to be fully reliable and should assign any statements he or she presents as facts with unit probability. As assigning unit probability to the background information about Linda (i.e. H'_1) would make it impossible to exploit the confirmation given by the background information to the proposition that Linda is a feminist (i.e. H'_3), we could no longer adopt the same reasoning as we did in the case of Walter above.

Nonetheless, we believe we should not immediately dismiss the possibility that the source that reports the background information to us (to wit, the experimenter) is not fully trustworthy. No doubt many of you have tried this example with friends or students and, if so, we would expect that you, just as we did, recognized in your audience at least some distrust, some slight uneasiness while they were trying to determine if, and if so in which way, you were trying to trick them. Such uneasiness may even be larger when put into an official experiment as everybody in the experiment knows that the experimenter is not genuinely interested in what you believe about Linda.

However, if we cannot fully trust the experimenter to present true facts, we should not assign any of his or her statements unit probability. This would immediately lead us to assign less than unit probability to the background information provided about Linda.

It would also make it much easier to argue that the statements the participants are given are interpreted as reports by the experimenter as well. That is, it would allow us to argue against difference (2) above by pointing out that participants may very well view these statements as reports by the experimenter and feel as if they are asked to judge which of these reports are more probable.[7]

And so we conclude that if we accept that the participants do not view the background information as having unit probability and if we accept that they view the statements about Linda as reports by the experimenter, then the fallacy no longer arises for plausible probability distributions.

In the recent literature there are quite a number of arguments that attempt to show that the participants in the Linda example do not commit a fallacy. In Hertwig et al. (2008), an overview is given of different arguments that attempt to show that the use of the 'AND' operator in ordinary language is different from the 'AND' operator used in propositional logic. Such arguments would lead us to conclude that the 'and' in 'Linda is a bank teller *and* a feminist' would be different from the corresponding logical operator and so no fallacy arises.

---

[7] Participants have also been found to commit the conjunction fallacy if no background information is provided by the experimenter. To fit these cases into our account, we conjecture that the participants also consider their background beliefs about the situation at stake (= H''_1) and proceed as described above. If this conjecture is correct, however, our model has to be modified as the experimenter now only presents the propositions H''_2 and H''_3. See also our fifth objection in the next section.

Others have argued that although the fallacy does arise, it can be explained by showing that the participants in the example are using some sort of rule that in other cases would lead them to favor the proposition with the highest probability. So for instance, Shogenji (2009) argues that participants favor the proposition with the highest *epistemic justification*, which in many cases coincides with the proposition with the highest probability, though not always, as the Linda experiment shows. In a similar vein, Crupi et al. (2008) and Atkinson et al. (2009) argue that participants choose the proposition that has the highest degree of *confirmation*, which, again, often but not always points us to the proposition with the highest probability (cf. Schupbach 2009a; Tentori and Crupi 2009).

But note that our argument is different from either of these. However, we agree with the second approach that we should analyze the 'and' by using the logical 'AND' operator, and we agree with the first approach that the participants are not making a mistake in the Linda experiment. And so the participants would commit a fallacy if they were not to assign the conjunction the highest probability. And we believe that this is an argument that has not yet been defended by anybody.

Our argument is inspired by a proposal by Bovens and Hartmann (2003, ch. 3), who analyze the Linda example by taking the corresponding reports into account and showed that $p(H'_2, H'_3|R'_2, R'_3)$ can be greater than $p(H'_3|R'_3)$ for plausible probability distributions. However, the analysis of Bovens and Hartmann differs from ours in important respects: First, they assume that the background information about Linda is certain, i.e. they effectively work with a probability function that conditions on H'_1. Secondly, Bovens and Hartmann model the witness credibility differently. While we assume that an unreliable witness randomizes with the prior probability of the corresponding proposition, Bovens and Hartmann introduce a randomization parameter *a* (see Eq. 2) which is fixed and hence independent of the prior probability of the proposition in question. We think that our assumption is descriptively more realistic and normatively more appealing. What is more, our main results, captured in our theorem, crucially depend on our way of modeling the witness credibility.

5 Anticipated objections

Clearly, the argument in the previous section is still quite tentative. For one thing, we would need actual experimental data to support some of our statements about the beliefs of the participants. But let us at least discuss a few possible objections that we believe are especially noteworthy, as they may further explain our claim above.

First, one could insist that the experimenter is not part of the experiment and that she can therefore simply not be modelled as a witness with a certain credibility. Clearly, this is a critical assumption for our argument to succeed. Without it, we cannot argue that the background information has a probability different from unity or that the statements about Linda can be modelled as witness reports. We argued in favour of it by pointing out that there is good reason to distrust the experimenter. We could argue at length in favor of this argument, but in the end this is an empirical claim which should be tested experimentally. This is simply not a question that philosophy alone can answer.

Second, one may object that no matter how the participants view the experimenter, they can never be rational. For this type of experiment has also been conducted using bets with actual money and the participants who put their money on the conjunction instead of on one of the conjuncts will certainly end up losing money. But this would be a misinterpretation of our argument. For we certainly do not believe that the participants are assigning the *correct* probabilities to the propositions. Indeed, they are not. For we all know that the experimenter can indeed be trusted (or at least with respect to the propositions that he or she is reporting). Our point is rather that the participants in the experiment *may* believe otherwise. That is, there is good reason to believe that many of the participants will assign probabilities to the different propositions such that they are fully rational in assigning a higher probability to the conjunction.

Third, similar experiments have been conducted in which experimental data are used instead of a made-up story about Linda. In such experiments (which Hertwig et al. 2008: 741 call 'frequency representations'), the participants should know the probabilities involved and we really have an independent experimenter. However, in that case, too, the experimenter is still the person relating the story and the participants still seem to have no a priori reason to expect the experimenter to be fully reliable with regard to the information that he or she is reporting.

Fourth, one may wonder why, provided that our argument is correct, still a substantial number of participants assign a larger probability to the first set. This, however, is not difficult to answer: Our theorem provides a necessary and sufficient condition for $D(\mathcal{S}_I, \mathcal{S}_{II}) > 0$. If this condition is not satisfied, then it is rational to assign the first set a higher posterior probability than the second set.

Fifth, one may point out that our model assumption that the witness is either a truth-teller or a randomizer is highly idealized. After all, there seems to be a whole spectrum of possibilities between truth-telling and randomizing. And why don't we take into account that the witness is deliberately lying? Moreover, one could criticize our independence assumption either by arguing that it is too weak or mistaken in another way or even by arguing that it is too strong and does not make sense in the case where we have only a single witness. We respond that it is an empirical matter which mechanism for the witness credibility is at work and which independences hold. This clearly depends on the particular scenario under investigation. The goal of this paper was rather modest: We only wanted to show that for a specific (though plausible) case rational participants should assign a higher probability to the larger set of propositions. And this we did.

Finally, one may argue that condition (c) is implausible or ad hoc. It certainly is not clear a priori whether participants who consider $\mathcal{S}_{II}$ to be more probably than $\mathcal{S}_I$ do indeed comply with this condition. We agree. This is an empirical questions and corresponding experiments have to be conducted to test it. To do so, our model will presumably have to be elaborated and better adapted to the Linda experiment.

Fortunately for us, this paper was meant as a first step in that direction and we think that we can safely conclude that its goal has been accomplished.

Appendix A: the posterior probability

From the Bayesian network in Fig. 1, we can read off an expression for the joint probability over all variables (see Neapolitan 1990):

$$p(H_1, \ldots, H_n, R_1, \ldots, R_n, \text{Rel}) = p(H_1, \ldots, H_n) \, P(\text{Rel}) \, \Pi_{i=1}^n \, p(R_i|H_i, \text{Rel})$$

With this we can calculate

$$p(H_1, \ldots, H_n | R_1, \ldots, R_n) = \frac{p(H_1, \ldots, H_n, R_1, \ldots, R_n)}{p(R_1, \ldots, R_n)}.$$

Let's first calculate

$$
\begin{aligned}
p(H_1, \ldots, H_n, R_1, \ldots, R_n) &= \sum_{\text{Rel}} p(H_1, \ldots, H_n, R_1, \ldots, R_n, \text{Rel}) \\
&= \left[p(\text{Rel}) \cdot 1 + p(\neg \text{Rel}) \, \Pi_{i=1}^n \, p(H_i) \right] \cdot p(H_0, H_1, \ldots, H_n).
\end{aligned}
$$

Similarly, we obtain:

$$
\begin{aligned}
p(R_1, \ldots, R_n) &= \sum_{H_1, \ldots, H_n, \text{Rel}} p(H_1, \ldots, H_n, R_1, \ldots, R_n, \text{Rel}) \\
&= p(\text{Rel}) \sum_{H_1, \ldots, H_n} p(H_1, \ldots, H_n) \, \Pi_{i=1}^n \, p(R_i|H_i, \text{Rel}) \\
&\quad + p(\neg \text{Rel}) \sum_{H_1, \ldots, H_n} p(H_1, \ldots, H_n) \, \Pi_{i=1}^n \, p(R_i|H_i, \neg \text{Rel}) \\
&= p(\text{Rel}) \, p(H_1, \ldots, H_n) + p(\neg \text{Rel}) \, \Pi_{i=1}^n \, p(H_i)
\end{aligned}
$$

Putting it all together, we obtain Eq. 10.

Appendix B: proof of the theorem

To simplify the notation, we introduce the following abbreviations: $h_i := p(H_i)$ (for $i = 1, 2, 3$), $h_{12} := p(H_1, H_2)$, $h_{123} := p(H_1, H_2, H_3)$, $r := p(\text{Rel})$ and $\overline{x} := 1 - x$ for all variables x. We assume that $0 < h_1, h_2, h_3, h_{12}, h_{123}, r < 1$ and obtain for the difference function:

$$
\begin{aligned}
D(\mathcal{S}_I, \mathcal{S}_{II}) &= \frac{r + h_1 h_2 h_3 \overline{r}}{h_{123} \, r + h_1 h_2 h_3 \overline{r}} \, h_{123} - \frac{r + h_1 h_2 \overline{r}}{h_{12} \, r + h_1 h_2 \overline{r}} \, h_{12} \\
&= \frac{h_1 h_2 \overline{r} \left\{ \left[(1 - h_{12} \overline{h_3}) r + h_1 h_2 h_3 \overline{r} \right] h_{123} - \left[r + h_1 h_2 \overline{r} \right] h_{12} h_3 \right\}}{(h_{123} \, r + h_1 h_2 h_3 \overline{r}) (h_{12} \, r + h_1 h_2 \overline{r})} \\
&= \frac{h_1 h_2 h_{12} \overline{r} \left[(1 - h_{12} \overline{h_3}) r + h_1 h_2 h_3 \overline{r} \right]}{(h_{123} \, r + h_1 h_2 h_3 \overline{r}) (h_{12} \, r + h_1 h_2 \overline{r})} \cdot \{ h_{123} / h_{12} - h_3 (1 + \delta) \}
\end{aligned}
$$

with

$$\delta = \frac{h_{12}\,r + h_1\,h_2\,\overline{r}}{(1 - h_{12}\,\overline{h_3})\,r + h_1\,h_2\,h_3\,\overline{r}} \cdot \overline{h_3}.$$

Note that $\delta > 0$. We conclude that $D(\mathcal{S}_I, \mathcal{S}_{II}) > 0$ iff

$$h_{123}/h_{12} > h_3 \cdot (1 + \delta), \tag{14}$$

which proofs the first part of the theorem as $h_{123}/h_{12} = p(H_3|H_1, H_2)$.

It is interesting to note that δ is a negative function of h_3 (proof omitted), i.e. the larger h_3, the smaller δ (and hence the "better" condition (c)).

Equation 14 can be rewritten in the following way:

$$\frac{h_{123}}{h_1\,h_2\,h_3} > \frac{h_{12}}{h_1\,h_2} \cdot (1 + \delta)$$

This equation can be expressed in terms of Shogenji's measure of coherence (Shogenji 1999)[8]:

$$c_S^{(n)} := \frac{h_{1\ldots n}}{h_1 \cdots h_n}.$$

Hence, condition (14) is equivalent to

$$c_S^{(3)} > c_S^{(2)} \cdot (1 + \delta). \tag{15}$$

Let us now examine the error measure δ. To do so, we plot δ as a function of the probability of the background story (i.e. h_1 with $.01 < h_1 < .99$) for $r = .5$ and three different sets of parameters, which satisfy conditions (a) and (b). The parameter sets are:

Set 1: $h_2 = h_3 = 0.1$, $h_{12} = 0.001$, $h_{13} = 0.009$
Set 2: $h_2 = h_3 = 0.05$, $h_{12} = 0.0005$, $h_{13} = 0.0495$
Set 3: $h_2 = h_3 = 0.01$, $h_{12} = 0.0001$, $h_{13} = 0.0099$

As Fig. 2 shows, the error is small even for large values of the parameters.

Next, we examine the r-dependence of δ. To begin with, note that δ is strictly monotonically decreasing for $0 < r < 1$ with

$$\lim_{r \searrow 0} \delta = \frac{1}{h_3} - 1, \tag{16}$$

$$\lim_{r \nearrow 1} \delta = \frac{h_{12}\,(1 - h_3)}{1 - h_{12}\,(1 - h_3)} = h_{12}\,(1 - h_3)\,(1 + \mathcal{O}(h_{12})). \tag{17}$$

---

[8] Shogenji's measure has been criticized. See, for instance, Bovens and Hartmann (2003) and Fitelson (2003). For a recent defence of (a modified version of) Shogenji's measure, see Schupbach (2009b). It remains to be examined whether a condition similar to Eq. 15 obtains for other coherence measures discussed in the literature.
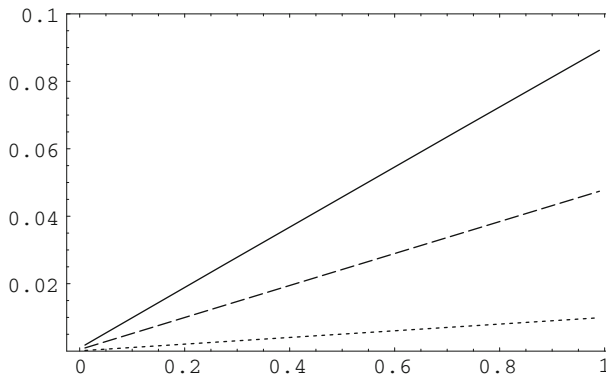
Fig. 2 The error measure δ for $r = 0.5$ as a function of h_1 for set 1 (solid line), set 2 (dashed line) and set 3 (dotted line)
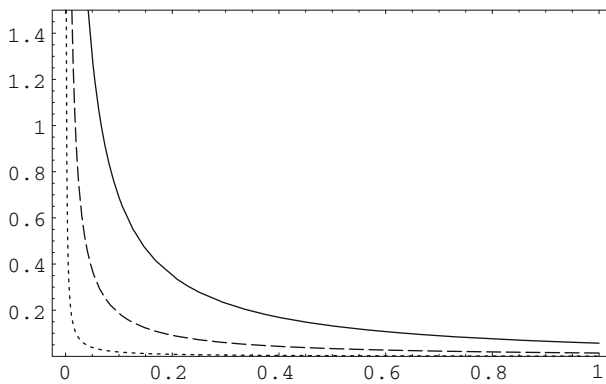

Fig. 3 The error measure δ as a function of r for set 4 (solid line), set 5 (dashed line) and set 6 (dotted line)

(Note that $\lim_{r \nearrow 1} D(\mathcal{S}_I, \mathcal{S}_{II}) = 0$ as $p^*(\mathcal{S}_I), p^*(\mathcal{S}_{II}) \to 1$ as r approaches 1 from below.) As h_{12} is assumed to be small (see condition (a)), condition (c) and the equivalent coherence condition

$$c_S^{(3)} > c_S^{(2)} \tag{18}$$

become increasingly good as r increases. Figure 3 illustrates this for the following three parameter sets:

Set 4: $h_1 = 0.8, h_2 = h_3 = 0.1, h_{12} = 0.06$
Set 5: $h_1 = 0.4, h_2 = h_3 = 0.05, h_{12} = 0.015$
Set 6: $h_1 = 0.2, h_2 = h_3 = 0.01, h_{12} = 0.001$

Figure 3 also show that condition (c) and the equivalent coherence condition (18) are not good guides for values of r smaller than approximately .5. Equation 16 shows that they are also not good guides for small values of h_3. In these cases it is rational to assign a higher probability to $\mathcal{S}_{II}$ than to $\mathcal{S}_I$ only if the (Shogenji) coherence of $\mathcal{S}_{II}$ is *considerably* greater than the (Shogenji) coherence of $\mathcal{S}_I$.

References

Atkinson, D., Peijnenburg, J., & Kuipers, T. (2009). How to confirm the disconfirmed. On conjunction fallacies and robust confirmation. *Philosophy of Science, 76*, 1–21.

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

Bovens, L., & Olsson, E. J. (2000). Coherentism, reliability and Bayesian networks. *Mind, 109*, 685–719.

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking and Reasoning, 14*(2), 182–199.

Fitelson, B. (2003). A probabilistic theory of coherence. *Analysis, 63*, 195–199.

Hájek, A., & Hartmann, S. (2009). Bayesian epistemology. In M. Steup (ed.), *Blackwell companion to epistemology*. Oxford: Blackwell.

Hansson, B. (1983). Epistemology and evidence. In P. Gärdenfors, B. Hansson, & N. Sahlin (eds.), *Evidentiary value: Philosophical, judicial and psychological aspects of a theory—Essays dedicated to Sören Halldén on his sixtieth birthday (Library of Theoria 15)* (pp. 75–97). Lund: Gleerup.

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of *and*. *Cognition, 108*, 740–753.

Klein, P., & Warfield, T. (1994). What price coherence? *Analysis, 54*, 129–132.

Meijs, W. (2006). A corrective to Bovens and Hartmann's measure of coherence. *Philosophical Studies, 133*, 151–180.

Meijs, W. (2009). Coherentism and justification. Unpublished manuscript.

Meijs, W., & Douven, I. (2007). On the alleged impossibility of coherence. *Synthese, 157*, 347–360.

Neapolitan, R. (1990). *Probabilistic reasoning in expert systems*. New York: Wiley.

Olsson, E. J. (2005). *Against coherence*. Oxford: Oxford University Press.

Schupbach, J. N. (2009a). Is the conjunction fallacy tied to probabilistic confirmation? *Synthese*, this special issue.

Schupbach, J. N. (2009b). New hope for Shogenji's coherence measure. Unpublished manuscript.

Shogenji, T. (1999). Is coherence truth-conducive? *Analysis, 59*, 338–345.

Shogenji, T. (2009). The degree of epistemic justification and the conjunction fallacy. *Synthese*, this special issue.

Tentori, K., & Crupi, V. (2009). How the conjunction fallacy is tied to probabilistic confirmation: A reply to Schupbach (2009). *Synthese*, this special issue.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.