To appear in *Synthese* (2018?)

# A Unified Account of the Conjunction Fallacy by Coherence

**Martin L. Jönsson and Tomoji Shogenji**

**Abstract:** We propose a coherence account of the conjunction fallacy applicable to both of its two paradigms (the M-A paradigm and the A-B paradigm). We compare our account with a recent proposal by Tentori, Crupi and Russo (2013) that attempts to generalize earlier confirmation accounts. Their model works better than its predecessors in some respects, but it exhibits only a shallow form of generality and is unsatisfactory in other ways as well: it is strained, complex, and untestable as it stands. Our coherence account inherits the strength of the confirmation account, but in addition to being applicable to both paradigms, it is natural, simple, and readily testable. It thus constitutes the next natural step for Bayesian theorizing about the conjunction fallacy.

## 1. The Confirmation Account of the Conjunction Fallacy

The conjunction fallacy challenges the naïve assumption of human rationality. The following Linda Scenario is most frequently cited in the discussion.

| | |
|---|---|
| $e$ | 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. |
| $h_1$ | Linda is a bank teller. |
| $h_1 \wedge h_2$ | Linda is a bank teller and is active in the feminist movement. |

As is very well documented, when given the description ($e$) of Linda, participants overwhelmingly judge the conjunctive statement ($h_1 \wedge h_2$) to be more probable than the isolated conjunct ($h_1$) even though this violates the conjunction law of probability theory.

In the face of apparent irrationality—and with an eye to finding some explanation—there have been attempts by psychologists and philosophers to spell out general conditions under which the conjunction fallacy occurs. There are two features in the Linda Scenario that are salient from the Bayesian perspective.[1] First, $e$ makes the probability of $h_1$ lower, and second, $e$

[1] Many influential accounts of the conjunction fallacy in the literature have been Bayesian. The main exception is the representativeness account by Tversky and Kahneman (1983). We agree with its critics (e.g. Gigerenzer 1996)

makes the probability of $h_2$ higher. It is not unreasonable to think that these features are responsible for the occurrence of the fallacy. Some Bayesians proposed to capture these features formally in terms of conditional probabilities, i.e. $\Pr(h_1|e)$ is low while $\Pr(h_2|e)$ is high. This difference by itself does not explain the fallacy since the participants are asked to compare $\Pr(h_1|e)$ and $\Pr(h_1 \wedge h_2|e)$. However, the suggestion is made that perhaps many participants compute $\Pr(h_1 \wedge h_2|e)$ by averaging $\Pr(h_1|e)$ and $\Pr(h_2|e)$. The faulty computation would explain the conjunction fallacy since $\Pr(h_1|e) < \Pr(h_2|e)$ makes their average higher than $\Pr(h_1|e)$. The averaging account proved empirically inadequate, though. We will not go over the empirical findings against it from the literature,[2] but its inadequacy is easy to understand with some reflection. Replace $h_2$ in the Linda Scenario by $h_3$: Linda owns a black pair of shoes. $\Pr(h_3|e)$ is quite high, presumably even higher than $\Pr(h_2|e)$, so that the average between $\Pr(h_1|e)$ and $\Pr(h_3|e)$ should be higher than $\Pr(h_1|e)$, but there is no intuitive pull toward the conjunction fallacy.

A more promising approach emerged that compares the degrees of confirmation in the incremental sense (Sides et al 2002). What matters, according to the confirmation account, is not whether the probabilities are high or low in the end, but in which direction the probabilities move due to *e*. It is clear in the Linda Scenario that the description (*e*) disconfirms the bank-teller hypothesis ($h_1$) in the incremental sense of $\Pr(h_1) > \Pr(h_1|e)$, while the description (*e*) confirms the feminist hypothesis ($h_2$) in the incremental sense of $\Pr(h_2) < \Pr(h_2|e)$. The account nicely explains why we should not expect the conjunction fallacy in the Black Shoes variant of the Linda Scenario. Though $\Pr(h_3|e)$ is high, it is only because $\Pr(h_3)$ is high to begin with, and not because *e* confirms $h_3$. The description (*e*) of Linda is irrelevant to her owning a black pair of shoes ($h_3$).

The confirmation account has a further advantage. While the averaging account must assert that a faulty computation is common, the confirmation account need not ascribe a faulty computation to the majority of participants. Let $\mathrm{Con}(h, e)$ be the degree of confirmation in the sense of increase in the probability from $\Pr(h)$ to $\Pr(h|e)$. When *e* disconfirms $h_1$ but confirms $h_2$, $\mathrm{Con}(h_1, e)$ is low and $\mathrm{Con}(h_2, e)$ is high. In such cases, it might be expected that the degree of confirmation $\mathrm{Con}(h_1 \wedge h_2, e)$ for the conjunction is somewhere between $\mathrm{Con}(h_1, e)$ and $\mathrm{Con}(h_2, e)$, so that $\mathrm{Con}(h_1 \wedge h_2, e)$ is higher than $\mathrm{Con}(h_1, e)$.[3] This is not a faulty computation, and allows the following simple formulation of the confirmation account:

> **The Confirmation Account:** The conjunction fallacy occurs commonly when and only when $\mathrm{Con}(h_1 \wedge h_2, e)$ is appreciably greater than $\mathrm{Con}(h_1, e)$; the greater the difference is, the more frequently the fallacy occurs.[4]

Of course, any account of a fallacy must ascribe some error to those who commit it. In the case

---

that the informal and fuzzy characterization of representativeness seriously limits its explanatory power. In this paper we only examine the Bayesian accounts of the conjunction fallacy.

[2] See, e.g. Tentori, Crupi and Russo (2013).

[3] The precise condition for $\mathrm{Con}(h_1, e) < \mathrm{Con}(h_1 \wedge h_2, e)$ depends on the way we measure confirmation. We will address this problem in the next section.

[4] We insert the qualification "appreciably" here because the conjunction fallacy is not expected if $\mathrm{Con}(h_1 \wedge h_2, e)$ is only negligibly greater than $\mathrm{Con}(h_1, e)$, so that the difference does not draw the participants' attention. The same point applies, mutatis mutandis, to the qualification "appreciably" in other accounts of the conjunction fallacy. We assume, in other words, that there are thresholds of appreciable difference in probability, confirmation, coherence, etc. to be determined empirically, though we do not investigate the issue in this paper.

of the confirmation account, it is an improper focus on the changes in the probabilities, where the actual question is about the probabilities. However, ascribing this error is not contrived or arbitrary. It is consistent with the broader phenomenon of the base rate neglect, where people asked about probabilities focus instead on the changes in the probability. The neglect of the prior probabilities (the base rates) is common and has been extensively documented.[5] So, in addition to being empirically superior to the averaging account, the confirmation account locates the conjunction fallacy in the broader map of fallacies in our probabilistic judgement.

## 2. Problems of the Confirmation Account

Despite the obvious attractions, there are two major problems with the confirmation account. One of them is the technical issue of measure sensitivity. There are many formal measures of confirmation proposed in the literature,[6] and most of them are not ordinally equivalent to each other.[7] Take two simple measures, the difference measure $D(h, e)$ and the ratio measure $R(h, e)$:

$$D(h,e) =_{\text{def}} \Pr(h \mid e) - \Pr(h)$$
$$R(h,e) =_{\text{def}} \frac{\Pr(h \mid e)}{\Pr(h)}$$

The following example reveals that they are not ordinally equivalent. Suppose $\Pr(h_1) = 0.01$ and $\Pr(h_1|e) = 0.1$ while $\Pr(h_2) = 0.4$ and $\Pr(h_2|e) = 0.8$. We use the two measures to evaluate the relative strengths of confirmation between $\text{Con}(h_1, e)$ and $\text{Con}(h_2, e)$. First, $e$ confirms $h_1$ much less than it confirms $h_2$ if we use the difference measure:

$$D(h_1, e) = \Pr(h_1 \mid e) - \Pr(h_1) = 0.1 - 0.01 = 0.09$$
$$D(h_2, e) = \Pr(h_2 \mid e) - \Pr(h_2) = 0.8 - 0.4 = 0.4$$

However, $e$ confirms $h_1$ much more than it confirms $h_2$ if we use the ratio measure:

$$R(h_1, e) = \frac{\Pr(h_1 \mid e)}{\Pr(h_1)} = \frac{0.1}{0.01} = 10$$
$$R(h_2, e) = \frac{\Pr(h_2 \mid e)}{\Pr(h_2)} = \frac{0.8}{0.4} = 2$$

This means that the confirmation account can make different predictions on the occurrence of the conjunction fallacy in some cases, depending on which measure we use.

Faced with the problem of measure sensitivity, Crupi, Fitelson and Tentori (2008) showed formally that the relation $\text{Con}(h_1, e) < \text{Con}(h_1 \wedge h_2, e)$ that induces the conjunction

[5] See Bar-Hillel (1980) and Koehler (1996) for influential discussions of the base rate fallacy.
[6] In a recent paper Roche and Shogenji (2014) list thirteen Bayesian measures of confirmation.
[7] Two measures $C(h, e)$ and $C^*(h, e)$ are ordinally equivalent to each other if and only if for any two ordered pairs $\langle h_1, e_1\rangle$ and $\langle h_2, e_2\rangle$, $C(h_1, e_1) < / = / > C(h_2, e_2)$ if and only if $C^*(h_1, e_1) < / = / > C^*(h_2, e_2)$.

fallacy is not affected by the choice of a measure, provided the following conditions hold:[8]

(i) $e$ is negatively (if at all) correlated with $h_1$
(ii) $e$ is positively correlated with $h_2$, *even conditionally on* $h_1$

It is a neat proof because these are the two salient features of the Linda Scenario. Note, however, that they cannot use their point in defense of the confirmation account as it is formulated earlier, viz. the conjunction fallacy occurs commonly when $\mathrm{Con}(h_1 \wedge h_2, e)$ is appreciably greater than $\mathrm{Con}(h_1, e)$. Think of the many cases where the two conditions (i) and (ii) do *not* hold but $\mathrm{Con}(h_1 \wedge h_2, e)$ is still appreciably greater than $\mathrm{Con}(h_1, e)$. This is possible because the two conditions are sufficient but not necessary for the relation $\mathrm{Con}(h_1, e) < \mathrm{Con}(h_1 \wedge h_2, e)$. The confirmation account predicts the occurrence of the conjunction fallacy in such cases, but then the problem of measure sensitivity arises again because $\mathrm{Con}(h_1 \wedge h_2, e)$ may be appreciably greater than $\mathrm{Con}(h_1, e)$ on one measure but not on another when the two conditions (i) and (ii) do not hold.

Crupi, Fitelson and Tentori are not concerned because the condition that induces the conjunction fallacy in their account is not different degrees of confirmation but the conditions (i) and (ii) themselves. They are using the degrees of confirmation to provide an *explanation* as to why (i) and (ii) induce the conjunction fallacy, viz. the conditions (i) and (ii) make people focus on changes in the probabilities instead of the probabilities. When the conditions (i) and (ii) do not hold, an appreciable difference in the degrees of confirmation does not cause the conjunction fallacy. The role of confirmation in their account is therefore limited. We call it "the confirmation analysis" of the conjunction fallacy to distinguish it from the confirmation account that uses different degrees of confirmation to predict the occurrence of the conjunction fallacy. If we want to defend the confirmation account, we still need to choose and defend a particular measure of confirmation.[9]

The second problem with the confirmation account is the scope of its application. In their classic article, Tversky and Kahneman (1983) distinguished two types of the conjunction fallacy, the M-A paradigm and the A-B paradigm.[10] The Linda Scenario belongs to the M-A paradigm, where $e$ lowers the probability of $h_1$ and raises the probabilities of $h_2$. The A-B paradigm does not share these features. Take the Health Survey Scenario, which is also known to induce the conjunction fallacy:

$e$ — $x$ is a randomly selected adult male.

$h_1$ — $x$ has had one or more heart-attacks.

$h_1 \wedge h_2$ — $x$ has had one or more heart-attacks and is older than 55.

[8] They prove that $\mathrm{Con}(h_1, e) < \mathrm{Con}(h_1 \wedge h_2, e)$ from the conditions (i) and (ii) on six popular measures of confirmation, where (i) and (ii) are understood formally as $\Pr(h_1|e) \leq \Pr(h_1)$ and $\Pr(h_2|e \wedge h_1) > \Pr(h_2|h_1)$, respectively.

[9] In his account of the conjunction fallacy Shogenji (2012) argues for a particular measure $J(h, e)$. He takes $J(h, e)$ to be the measure of "justification", but it behaves much like a measure of confirmation.

[10] "M", "B" and "A" in their paper correspond to "$e$", "$h_1$" and "$h_2$" in this paper, respectively. The M-A paradigm is so called because of the strong correlation between M and A ($e$ and $h_2$); the A-B paradigm is so called because of the strong correlation between A and B ($h_1$ and $h_2$).

Unlike the Linda scenario, *e* does not affect the probabilities of $h_1$ and $h_2$ in any obvious and significant way. So, explanation by confirmation—whether it is the confirmation account or the confirmation analysis—does not predict the occurrence of the conjunction fallacy.

A salient feature of the Health Survey Scenario, and more generally the conjunction fallacy of the A-B paradigm, is a positive correlation between the two conjuncts, $h_1$ and $h_2$. One possible stance in defense of the confirmation account is to concede that it is only applicable to the M-A paradigm and seek a different account for the A-B paradigm. The different degrees of confirmation—or the two conditions (i) and (ii) in the case of the confirmation analysis—are then a sufficient but not necessary condition of the conjunction fallacy. However, a unified account is preferable, other things being equal. The next two sections examine two approaches to a unified account of the conjunction fallacy.

## 3. The Coherence Account of the Conjunction Fallacy

Since we already have the confirmation account for the M-A paradigm, extending the account to make it applicable to the A-B paradigm is one obvious way forward. We will examine a proposal of that kind shortly, but there is another approach to consider. Some have already suggested with regard to the Linda Scenario that we can explain the fallacy by the notion of coherence.[11] The idea is to capture the two salient features of the scenario by coherence instead of confirmation.

The notions of confirmation and coherence are similar when applied to a pair of statements. This is because confirmation is qualitatively symmetrical in the sense that $\mathrm{Con}(x, y) > 0$ just in case $\mathrm{Con}(y, x) > 0$.[12] In other words, if *y* raises the probability of *x*, then *x* raises the probability of *y* and vice versa.[13] Disconfirmation is also qualitatively symmetrical in the sense that $\mathrm{Con}(x, y) < 0$ just in case $\mathrm{Con}(y, x) < 0$. We can use these symmetries to translate confirmation and disconfirmation into coherence and incoherence.

Take the point that *e* lowers the probability of $h_1$ in the Linda Scenario. By symmetry this means that *e* and $h_1$ disconfirm each other, so that they are incoherent with each other.[14]

---

[11] Siebel (2002) proposes a coherence account of the Linda Scenario. It should be noted that Siebel's account is not Bayesian since he adopts the constraint-satisfaction model of coherence (Thagard 1989; 1992, Ch. 4). Shogenji (2012) also mentions a coherence account of the Linda Scenario and his analysis of coherence is Bayesian, but he does not pursue the idea in part because it is not much different from the confirmation account with regard to the M-A paradigm, which is his focus. Schippers (2016) proposes an account of the Linda case in terms of "contrastive coherence". The concept of contrastive coherence is of limited use in analyzing the conjunction fallacy since it only allows for qualitative evaluation—whether *x* coheres better with *y* than with *z*. We need quantitative evaluation for predicting relative frequencies of the conjunction fallacy. The concept of contrastive coherence is also limited to pairwise coherence, while a unified account of the conjunction fallacy we develop in this section applies a measure of coherence to tripletons.

[12] Assuming that $C(x, y) = 0$ for the neutral point at which there is neither confirmation nor disconfirmation. Replace 0 with *k* if the neutral point is set at $k \neq 0$.

[13] This is easy to prove by Bayes' Theorem.

[14] It is understood here that incoherence is a probabilistic generalization of inconsistency, just as disconfirmation is a probabilistic generalization of refutation. Statements that are incoherent can still be true at the same time, though incoherence makes it less likely, just as a statement that is disconfirmed by evidence can still be true though disconfirmation makes it less likely. Some readers familiar with the literature may worry here about "the impossibility results" (Bovens and Hartmann 2003; Olsson 2005) that there is no probabilistic measure of coherence that is truth conducive, i.e. there is no probabilistic measure of coherence such that the more coherent the set is, the more probable the conjunction of its members is, ceteris paribus. However, this problem arises when the reliability of the evidence producer is a factor, which is not the case in the conjunction fallacy. It is easy to show (e.g. Shogenji

Meanwhile, $e$ makes the probability of $h_2$ higher. By symmetry, this means that $e$ and $h_2$ confirm each other, so that they are coherent with each other. If we let Coh($x$, $y$) be the degree of coherence between $x$ and $y$, then Coh($h_1$, $e$) is low and Coh($h_2$, $e$) is high. In such cases, it might reasonably be expected that the degree of coherence Coh($h_1 \wedge h_2$, $e$) is somewhere between Coh($h_1$, $e$) and Coh($h_2$, $e$), so that Coh($h_1 \wedge h_2$, $e$) is higher than Coh($h_1$, $e$).[15] If so, symmetry allows us to reinterpret the confirmation account in terms of coherence: what is responsible for the occurrence of the conjunction fallacy is the different degrees of coherence between Coh($h_1$, $e$) and Coh($h_1 \wedge h_2$, $e$). Though the confirmation account that compares Con($h_1$, $e$) and Con($h_1 \wedge h_2$, $e$) works fine for the purpose of predictions, that is only because confirmation is tied to coherence, according to this view. We call it "the coherence analysis" of the confirmation account.

How do we evaluate the coherence analysis? Is it confirmation or coherence that is responsible for the conjunction fallacy? Though Con($h$, $e$) and Coh($h$, $e$) are similar, there is still some difference. For example, coherence is symmetrical not just in the qualitative sense but also in the quantitative sense of Coh($x$, $y$) = Coh($y$, $x$). This is not generally true of confirmation. It is possible that Con($x$, $y$) ≠ Con($y$, $x$) on most measures of confirmation, so that the degree of confirmation in one direction may be greater than the degree of confirmation in the other direction. In short, confirmation still has a direction while coherence is devoid of direction. This seems to be a fairly minor difference, perhaps a minor disadvantage on the part of the coherence analysis because there seems to be a direction in the relation between the given description ($e$) on one hand and the hypothesis (either $h_1$ or $h_1 \wedge h_2$) on the other.

However, a different picture emerges when we look beyond the M-A paradigm in search of a unified account of the conjunction fallacy. Recall, first, that the salient feature of the A-B paradigm is a positive correlation between the two conjuncts, $h_1$ and $h_2$. There is no obvious direction in the relation between the two hypotheses. Their positive correlation is more naturally characterized by coherence. Further, and more importantly, coherence is not restricted to the binary relation between two statements, as confirmation is. We can intelligibly ask what is the degree of coherence Coh($e$, $h_1$, $h_2$) among the three members $e$, $h_1$ and $h_2$, and compare it with the degree of coherence Coh($e$, $h_1$). This allows us to state one characteristic shared by the M-A paradigm and the A-B paradigm: The addition of $h_2$ increases the degree of coherence from Coh($e$, $h_1$) to Coh($e$, $h_1$, $h_2$).

In the M-A paradigm Coh($e$, $h_1$, $h_2$) is greater than Coh($e$, $h_1$) because $h_2$ is coherent with $e$; in the A-B paradigm Coh($e$, $h_1$, $h_2$) is greater than Coh($e$, $h_1$) because $h_2$ is coherent with $h_1$. Whichever way it is, the conjunction fallacy occurs when and only when the addition of $h_2$ increases the degree of coherence from Coh($e$, $h_1$) to Coh($e$, $h_1$, $h_2$). The following is therefore the coherence account of the conjunction fallacy applicable to both the M-A paradigm and the A-B paradigm:

> **The Coherence Account:** The conjunction fallacy occurs commonly when and only when Coh($e$, $h_1$, $h_2$) is appreciably greater than Coh($e$, $h_1$); the greater the difference is, the more frequently the fallacy occurs.

---

1999) that coherence is truth conducive, ceteris paribus, with regard to the conjunction of the members when the reliability of the evidence producer is not a factor.

[15] The precise condition for Coh($h_1$, $e$) < Coh($h_1 \wedge h_2$, $e$) depends on the way we measure coherence. We will discuss measures of coherence in Section 5 and demonstrate the precise behavior of our preferred measure in Section 6.

Though coherence and confirmation are not much different with regard to the M-A paradigm, the notion of coherence allows a simple unified account of the conjunction fallacy when we expand the scope to include the A-B paradigm.

## 4. An Extension of the Confirmation Account

We can appreciate the strength of the coherence account when we compare it with the other approach mentioned earlier—to extend the confirmation account of the M-A paradigm to make it applicable to the A-B paradigm. In a recent article Tentori, Crupi and Russo (2013, hereafter TCR) attempt such an extension. Their approach is to capture the connection among $e$, $h_1$ and $h_2$ by two relations of confirmation. One is $\mathrm{Con}(h_2, e|h_1)$, which is the degree of confirmation that $e$ provides for $h_2$ against the background $h_1$. This is measured by the change in the probability from $\Pr(h_2|h_1)$ to $\Pr(h_2|e \wedge h_1)$. The other is $\mathrm{Con}(h_2, h_1|e)$, which is the degree of confirmation that $h_1$ provides for $h_2$ against the background $e$. This is measured by the change in the probability from $\Pr(h_2|e)$ to $\Pr(h_2|h_1 \wedge e)$. The strain is already palpable. The coherence account captures the ternary relation among $e$, $h_1$ and $h_2$ directly by $\mathrm{Coh}(e, h_1, h_2)$, but TCR are forced to capture it by two relations of confirmation, $\mathrm{Con}(h_2, e|h_1)$ and $\mathrm{Con}(h_2, h_1|e)$, with the third relatum in the background.

TCR apply the two relations of confirmation to the two paradigms, respectively. In the M-A paradigm the first relation $\mathrm{Con}(h_2, e|h_1)$ is high in comparison with $\mathrm{Con}(h_1, e)$, while in the A-B paradigm the second relation $\mathrm{Con}(h_2, h_1|e)$ is high in comparison with $\mathrm{Con}(h_1, e)$. Considering these roles, TCR make the following proposal:

> **The Confirmation Model:** The frequency of the conjunction fallacy is a decreasing function of $\mathrm{Con}(h_1, e)$ and an increasing function of $\mathrm{Con}(h_2, e|h_1)$ and $\mathrm{Con}(h_2, e|h_1)$.

We call it the confirmation "model" because it only identifies three determinants of the conjunction fallacy without stating how they are combined to make predictions. This is in contrast with the coherence account which has only two determinants and their comparison generates predictions.

The added complexity of the confirmation model gives rise to some concerns. First, it makes an implementation in our psychology more difficult: the participants must examine each of TCR's three determinants in a given scenario, especially the second and third determinants that are degrees of confirmation one statement provides for another statement against the background of a third statement. Of course, the participants need not evaluate each determinant consciously and there may be some psychological mechanism that computes each determinant and put them together, but a simpler account is preferable, other things being equal.[16]

Note that the three determinants are often in conflict. In the Linda Scenario, for example, the second determinant $\mathrm{Con}(h_2, e|h_1)$ is high because the description ($e$) of Linda raises the probability of the feminist hypothesis ($h_2$) against the background of the bank-teller hypothesis ($h_1$). The second determinant then increases the frequency of the fallacy. However, the third determinant $\mathrm{Con}(h_2, h_1|e)$ is low because the truth of the bank-teller hypothesis ($h_1$) lowers the probability of the feminist hypothesis ($h_2$) against the background of the description ($e$) of

[16] As we will discuss in the next section, simpler probabilistic models also have an advantage in their predictive accuracy.

Linda.[17] So, the third determinant decreases the frequency of the conjunction fallacy. In other words, the second and the third determinants of the model pull us in the opposite directions.

It may be pointed out that the first determinant Con($h_1$, $e$) is low in the Linda Scenario because the description ($e$) of Linda lowers the probability of the bank-teller hypothesis ($h_1$). So, the first determinant increases the frequency of the fallacy. The scale may then be tilted toward the occurrence of the conjunction fallacy overall. However, it is not true that the conjunction fallacy occurs commonly if two of the three determinants point in the direction of the conjunction fallacy. Take the Carol-Poetry Scenario (Shafir et al 1990) below, which is known to *not* induce the conjunction fallacy.

| | |
|---|---|
| $e$ | Carol is 34 years old, very ambitious, fluent in French, German and Spanish, and interested in current political events. |
| $h_1$ | Carol reads poetry for a hobby. |
| $h_1 \wedge h_2$ | Carol reads poetry for a hobby and works as a foreign correspondent. |

Two of the three determinants, Con($h_2$, $e|h_1$) and Con($h_2$, $h_1|e$), point in the direction of the conjunction fallacy. As in the M-A paradigm, the description ($e$) of Carol confirms the foreign correspondent hypothesis ($h_2$) against the background of the poetry hypothesis ($h_1$). Further, as in the A-B paradigm, the poetry hypothesis ($h_1$) confirms the foreign correspondent hypothesis ($h_2$) at least mildly against the background of the description ($e$) of Carol. So, two of the three determinants in the Carol-Poetry Scenario point in the direction of the conjunction fallacy, but the scenario does not induce the conjunction fallacy.

TCR could make the correct prediction in the Carol-Poetry Scenario by assigning different weights to the three determinants, but they do not specify any weights. They are also silent on the way the three determinants are combined, for example, whether they are added and subtracted, or they are multiplied and divided so that one determinant with a near zero value can have a decisive effect on the outcome. Without some formula for combining the determinants, their proposal is untestable.

## 5. Measuring Coherence

We proposed the coherence account of the conjunction fallacy and pointed out its advantage over the confirmation model, but there is one issue we need to address in order to complete the coherence account. Earlier in Section 3 we mentioned the problem of measure sensitivity with regard to the confirmation account. There are many formal measures of confirmation proposed in the literature, and the confirmation account can make different predictions on the occurrence of the conjunction fallacy, depending on which measure we use. The same is true with regard to the

[17] One reader has questioned this judgment for the reason that being a feminist may explain Linda's choice (surprising given $e$) to work as a bank teller. There may be some cultural and generational differences about the idea of being a feminist. If someone agrees with this reader (we don't), then consider some other case of the M-A paradigm. Our general point is that $h_1$ in the M-A paradigm is at odds with $e$ while $e$ and $h_2$ are positively correlated, so that $h_1$ is usually at odds with $h_2$. As a result, $h_1$ in the M-A paradigm usually lowers the probability of $h_2$ against the background of $e$.

coherence account. There are many formal measures of coherence proposed in the literature, and the coherence account can make different predictions on the occurrence of the conjunction fallacy, depending on which measure we use. In this section we propose and defend a particular measure of coherence, so that the coherence account can make unambiguous predictions. We will examine whether the predictions match the empirical data in the next section.

To clarify our task, our intent is to propose an *explication* of coherence in Carnap's sense (Carnap 1950, Ch. 1), in which an inexact prescientific concept (*explicandum*) is replaced by an exact scientific concept (*explicatum*). Carnap suggested that the explicatum should be (1) similar to the explicandum, (2) exact, (3) fruitful, and (4) as simple as (1), (2) and (3) permit.[18] With regard to conditions (1) and (3), we pay special attention to the conjunction fallacy, i.e. the formal measure should capture features of the notion of coherence that are in play in the conjunction fallacy, and make correct predictions on the occurrence of the conjunction fallacy. This means that the measure should be appropriately sensitive to changes in the degree of coherence due to the addition of $h_2$ from Coh($e$, $h_1$) to Coh($e$, $h_1$, $h_2$). With regard to condition (4), many measures of coherence proposed in the literature are not simple. For example, Douven and Meijs (2007) propose to compute the degree of coherence by averaging the degrees of confirmation among its members.[19] This means that the degree of coherence $DM(e, h_1, h_2)$ among $e$, $h_1$ and $h_2$ is the (possibly weighted) average of twelve degrees of confirmation: Con($e$, $h_1$), Con($h_1$, $e$), Con($e$, $h_2$), Con($h_2$, $e$), Con($h_1$, $h_2$), Con($h_2$, $h_1$), Con($e$, $h_1 \wedge h_2$), Con($h_1 \wedge h_2$, $e$), Con($h_1$, $e \wedge h_2$), Con($e \wedge h_2$, $h_1$), Con($h_2$, $e \wedge h_1$) and Con($e \wedge h_1$, $h_2$). This is particularly problematic since many of those complex measures are actually models with adjustable parameters. For example, $DM(e, h_1, h_2)$ produces different values, depending on the measure of confirmation to use as the basis, and possibly on different weights to assign between the first six components and the last six components. Of course, we can choose a specific confirmation measure and specify the weights that best accommodate the data on the conjunction fallacy. However, it is well known in statistical learning theory that a theory generated by a complex probabilistic model (a model with many adjustable parameters) from the data tends to make less accurate predictions, other things being equal, because a complex model is more prone to chase random variation (noise) in the data.[20] So, we start with the simplest measures of coherence available in the literature.

There are two measures of coherence that are notable for their simplicity. One is $S(x_1, \ldots, x_n)$ proposed by Shogenji (1999) and the other is $GO(x_1, \ldots, x_n)$ proposed by Glass (2002) and Olsson (2002).

---

[18] There is a strong emphasis on condition (1) in the literature often at the expense of conditions (3) and (4). See, for example, Koscholke (2016) on how different Bayesian measures of coherence perform in a variety of "test cases". We are skeptical that a single measure of coherence can accommodate all our intuitive judgments about coherence in different contexts. There may be such a measure, but what we seek here is a measure that is not just similar to the explicandum but illuminating, especially in analyzing the conjunction fallacy. If some people find our measure of coherence to be insufficiently similar to the prescientific concept of coherence, we have no objection to changing the term. Lewis (1946) called a set of mutually supporting beliefs "congruent", and Chisholm (1966) called a set of mutually confirming propositions "concurrent".

[19] Other complex measures of coherence with many components to weigh and tally include those proposed by Fitelson (2003), Meijs (2006), Roche (2013) and Schupbach (2011).

[20] This is commonly known as the problem of overfitting. See Burnham and Anderson (2002) on how the complexity of a probabilistic model affects the accuracy of predictions. Forster and Sober (1994) and Sober (2015, Ch. 2) provide an accessible account of the problem.

$$S(x_1,...,x_n) =_{\text{def}} \frac{\Pr(x_1 \wedge ... \wedge x_n)}{\Pr(x_1)...\Pr(x_n)}$$

$$GO(x_1,...,x_n) =_{\text{def}} \frac{\Pr(x_1 \wedge ... \wedge x_n)}{\Pr(x_1 \vee ... \vee x_n)}$$

The idea behind measure *S* is that a set of statements is coherent to the extent that there is, on average, a positive probabilistic dependence among its members. If the members are probabilistically independent, then $\Pr(x_1 \wedge \ldots \wedge x_n) = \Pr(x_1) \ldots \Pr(x_n)$ so that $S(x_1, \ldots, x_n) = 1$. This is the neutral point at which the set is neither coherent nor incoherent. On the other hand, if the members are on average positively probabilistically dependent, then $\Pr(x_1 \wedge \ldots \wedge x_n) > \Pr(x_1) \ldots \Pr(x_n)$ so that $S(x_1, \ldots, x_n) > 1$. Measure *GO* is even more straightforward. It is simply the ratio of the joint probability $\Pr(x_1 \wedge \ldots \wedge x_n)$ to the probability of the disjunction $\Pr(x_1 \vee \ldots \vee x_n)$. In the extreme case where the members are all logically equivalent, $\Pr(x_1 \wedge \ldots \wedge x_n) = \Pr(x_1 \vee \ldots \vee x_n)$, so that $GO(x_1, \ldots, x_n) = 1$. This is the highest value of coherence assigned by the measure.

Of these two simple measures, *GO* turns out to be unsuitable for our purpose. Note that $\Pr(x_1 \wedge \ldots \wedge x_n) \geq \Pr(x_1 \wedge \ldots \wedge x_n \wedge x_{n+1})$ while $\Pr(x_1 \vee \ldots \vee x_n) \leq \Pr(x_1 \vee \ldots \vee x_n \vee x_{n+1})$ with no exception. As a result, $GO(x_1, \ldots, x_n) \geq GO(x_1, \ldots, x_n, x_{n+1})$ with no exception. This means that we can never make a set of statements more coherent by adding another statement to it. So, if we adopt *GO* as our measure of coherence, the coherence account never predicts the occurrence of the conjunction fallacy because $GO(e, h_1, h_2)$ can never be greater than $GO(e, h_1)$.[21]

Measure *S* fares better in this regard since $S(e, h_1, h_2)$ can be greater than $S(e, h_1)$, as follows:

$$\begin{aligned} S(e,h_1,h_2) &= \frac{\Pr(e \wedge h_1 \wedge h_2)}{\Pr(e)\Pr(h_1)\Pr(h_2)} \\ &= \frac{\Pr(e \wedge h_1)\Pr(e \wedge h_1 \wedge h_2)}{\Pr(e)\Pr(h_1)\Pr(e \wedge h_1)\Pr(h_2)} \\ &= S(e,h_1)S(e \wedge h_1,h_2) \end{aligned}$$

If $S(e \wedge h_1, h_2) > 1$ and thus the added statement $h_2$ is pairwise coherent with the conjunction $e \wedge h_1$, then $S(e, h_1, h_2) > S(e, h_1)$ so that the addition of $h_2$ increases the degree of coherence.

However, measure *S* also has a questionable feature. By generalizing the reasoning above, we can compare $S(x_1, \ldots, x_n)$ and $S(x_1, \ldots, x_n, x_{n+1})$ as follows:

[21] See Koscholke and Schipper (2016) for this line of objection to "relative overlap measures" of coherence. We can think of even simpler (non-relative) overlap measures, such as $X(x_1, \ldots, x_n) = \Pr(x_1 \wedge \ldots \wedge x_n)$ and $Y(x_1, \ldots, x_n) = 1/\Pr(x_1 \wedge \ldots \wedge x_n)$, but they are not appropriately sensitive to changes in the degree of coherence due to the addition of a member. For example, $X(e, h_1, h_2)$ is never greater than $X(e, h_1)$, while $Y(e, h_1, h_2)$ is never smaller than $Y(e, h_1)$.

$$S(x_1,...,x_n,x_{n+1}) = \frac{\Pr(x_1 \wedge ... \wedge x_n \wedge x_{n+1})}{\Pr(x_1)...\Pr(x_n)\Pr(x_{n+1})}$$
$$= \frac{\Pr(x_1 \wedge ... \wedge x_n)\Pr(x_1 \wedge ... \wedge x_n \wedge x_{n+1})}{\Pr(x_1)...\Pr(x_n)\Pr(x_1 \wedge ... \wedge x_n)\Pr(x_{n+1})}$$
$$= S(x_1,...,x_n)S(x_1 \wedge ... \wedge x_n, x_{n+1})$$

The equation reveals that $S(x_1, \ldots, x_n) < / = / > S(x_1, \ldots, x_n, x_{n+1})$ if and only if $S(x_1 \wedge \ldots \wedge x_n, x_{n+1}) < / = / > 1$, except when $S(x_1, \ldots, x_n) = 0$, in which case $S(x_1, \ldots, x_n, x_{n+1}) = 0$ regardless of $S(x_1 \wedge \ldots \wedge x_n, x_{n+1})$. This means that as long as the original set is consistent and thus $S(x_1, \ldots, x_n) > 0$, whether the added statement $x_{n+1}$ increases coherence, decreases coherence, or keeps it at the same level, solely depends on whether the added member is pairwise coherent, incoherent, or neutral (neither coherent nor incoherent) with the conjunction of the extant members. This has some odd consequences.

Suppose the two members of the doubleton $\{x_1, x_2\}$ are strongly coherent with each other, but we keep adding to it a new statement that is neutral with the extant members, so that pairwise $S(x_1 \wedge x_2, x_3) = 1$, $S(x_1 \wedge x_2 \wedge x_3, x_4) = 1$, …, $S(x_1 \wedge \ldots \wedge x_{n-1}, x_n) = 1$. It follows from these by the equation above that $S(x_1, x_2) = S(x_1, x_2, x_3) = S(x_1, x_2, x_3, x_4) = \ldots = S(x_1, \ldots, x_n)$. This means that the resulting large set $\{x_1, \ldots, x_n\}$ whose members are mostly neutral with each other except for one pair is as strongly coherent as the original doubleton $\{x_1, x_2\}$ is, which is counterintuitive. If the successive additions are to keep the level of coherence at the same high level, the added statements should be coherent with the extant members. The same point also applies to the degree of incoherence. Suppose $x_1$ and $x_2$ are strongly incoherent with each other. The equation above also entails that if we keep adding a new statement that is neutral with the extant members, the resulting large set $\{x_1, \ldots, x_n\}$ is as strongly incoherent as the original doubleton $\{x_1, x_2\}$ is.

Schupbach (2011) calls this implication of Shogenji's measure "the problem of irrelevant addition" and proposes to avoid it by modifying Shogenji's measure.[22] He retains Shogenji's measure as the basis, but defines the degree of coherence of the set $\{x_1, \ldots, x_n\}$ by the weighted average of the degrees of coherence of its subsets, including $\{x_1, \ldots, x_n\}$ itself and excluding singletons and the empty set. For example, The Schupbach degree of coherence $Sc(x_1, x_2, x_3)$ of $\{x_1, x_2, x_3\}$ is the weighted average of $S(x_1, x_2, x_3)$, $S(x_1, x_2)$, $S(x_1, x_3)$ and $S(x_2, x_3)$. He also normalizes Shogenji's measure by logarithm to make the neutral point zero instead of one, but that does not affect the substance of the measure. The important point is that Schupbach's measure avoids the problem of irrelevant addition. Suppose, for example, that the first two members of $\{x_1, x_2, x_3\}$ are coherent while $x_3$ is an irrelevant addition. If we use the straight average, then $Sc(x_1, x_2, x_3)$ is smaller than $Sc(x_1, x_2)$, as follows.

[22] Schupbach (2011) also cites "the depth problem" as reason for modifying Shogenji's measure, viz. $S(x_1, x_2, x_3)$ ignores the degrees of coherence among members of the subsets. For example, $S(x_1, x_2, x_3)$ is solely dependent on $\Pr(x_1)$, $\Pr(x_2)$, $\Pr(x_3)$ and $\Pr(x_1 \wedge x_2 \wedge x_3)$, and is insensitive to $S(x_1, x_2)$, $S(x_2, x_3)$ and $S(x_3, x_1)$. (See also Fitelson (2003) for this line of objection to Shogenji's measure.) We set aside the depth problem here since there is no indication in cases of the conjunction fallacy that the degrees of coherence among members of the subsets are *additional* factors that influence the occurrence of the conjunction fallacy.

$$
\begin{aligned}
Sc(x_1, x_2, x_3) &= \frac{1}{4}[\log S(x_1, x_2, x_3) + \log S(x_1, x_2) + \log S(x_1, x_3) + \log S(x_2, x_3)] \\
&= \frac{1}{4}[\log S(x_1, x_2) + \log S(x_1, x_2) + \log 1 + \log 1] \\
&= \frac{1}{2}\log S(x_1, x_2) \\
&< Sc(x_1, x_2)
\end{aligned}
$$

It is easy to see that the inequality holds even if we use some other weights unless the weights for log $S(x_1, x_3)$ and log $S(x_2, x_3)$ are zero.

We take the problem of irrelevant addition seriously because change in the degree of coherence due to addition is central to the coherence account of the conjunction fallacy, but we also like the simplicity of Shogenji's measure where there is no need to combine different components by way of their weighted average. It is preferable if we can solve the problem in a simpler way, and our proposal is to adjust Shogenji's degree of coherence by the size $n$ of the set as follows.[23]

$$S^*(x_1,...,x_n) =_{\text{def}} S(x_1,...,x_n)^{\frac{1}{n-1}}$$

In the case of a doubleton there is no difference between $S^*(x_1, x_2)$ and $S(x_1, x_2)$ since the exponent attached to $S(x_1, \ldots, x_n)$ becomes one when $n = 2$. However, $S^*(x_1, x_2, x_3)$ is the square root of $S(x_1, x_2, x_3)$, $S^*(x_1, x_2, x_3, x_4)$ is the cubic root of $S(x_1, x_2, x_3, x_4)$, and so on, as $n$ increases. This has the effect of diluting coherence when an irrelevant member is added to the set. For example, if $x_1$ and $x_2$ are coherent while $x_3$ is an irrelevant addition, then $S^*(x_1, x_2) = S(x_1, x_2, x_3)$ because $S^*(x_1, x_2)$ is $S(x_1, x_2)$ itself and $S(x_1, x_2) = S(x_1, x_2, x_3)$ when $x_3$ is an irrelevant addition. Meanwhile, $S^*(x_1, x_2, x_3)$ is the square root of $S(x_1, x_2, x_3)$, which is smaller than $S(x_1, x_2, x_3)$ since $S(x_1, x_2, x_3) > 1$. It follows that $S^*(x_1, x_2) > S^*(x_1, x_2, x_3)$. An irrelevant addition therefore dilutes coherence by pulling the degree of coherence toward the neutral point. The same point applies when $x_1$ and $x_2$ are incoherent and $x_3$ is an irrelevant addition. As before, $S^*(x_1, x_2) = S(x_1, x_2, x_3)$, while $S^*(x_1, x_2, x_3)$ is the square root of $S(x_1, x_2, x_3)$, which is greater than $S(x_1, x_2, x_3)$ since $S(x_1, x_2, x_3) < 1$. An irrelevant addition dilutes incoherence by pulling the degree of coherence toward the neutral point.

Since adding an irrelevant member dilutes coherence toward the neutral point, maintaining the same high level of coherence requires that the added member be also coherent with the conjunction of the extant members, but how coherent does it have to be? Measure $S^*$ gives a clear answer to this question:

$$S^*(x_1,...,x_n,x_{n+1}) = S^*(x_1,...,x_n)$$

[23] If we want to make the neutral point zero instead of one, we can normalize $S^*(x_1, \ldots, x_n)$ by logarithm to obtain $\log S(x_1,...,x_n)^{\frac{1}{n-1}} = \frac{1}{n-1}\log S(x_1,...,x_n)$.

$$\text{iff} \quad S(x_1,...,x_n,x_{n+1})^{\frac{1}{n}} = S(x_1,...,x_n)^{\frac{1}{n-1}}$$

$$\text{iff} \quad S(x_1,...,x_n,x_{n+1}) = S(x_1,...,x_n)^{\frac{n}{n-1}}$$

$$\text{iff} \quad S(x_1,...,x_n)S(x_1 \wedge ... \wedge x_n, x_{n+1}) = S(x_1,...,x_n)^{\frac{n}{n-1}}$$

$$\text{iff} \quad S(x_1 \wedge ... \wedge x_n, x_{n+1}) = S(x_1,...,x_n)^{\frac{n}{n-1}-1}$$

$$\text{iff} \quad S(x_1 \wedge ... \wedge x_n, x_{n+1}) = S(x_1,...,x_n)^{\frac{1}{n-1}}$$

$$\text{iff} \quad S^*(x_1 \wedge ... \wedge x_n, x_{n+1}) = S^*(x_1,...,x_n)$$

This means that the degree of coherence remains at the same level if and only if the added member is as coherent with the conjunction of the extant members as the extant members are among themselves. To raise the degree of confirmation the added member must be more coherent with the conjunction of the extant members than the extant members are among themselves. We believe this feature of measure $S^*$ captures our intuitive sense of coherence that is in play in the conjunction fallacy.

## 6. Applications

In this section we examine the predictions of the coherence account with the use of measure $S^*$. We will write $S^*(e, h_1)$ and $S^*(e, h_1, h_2)$ as $\text{Coh}(e, h_1)$ and $\text{Coh}(e, h_1, h_2)$, respectively, since $S^*$ is now our official measure of coherence. To recall, the coherence account states that the conjunction fallacy occurs commonly when and only when $\text{Coh}(e, h_1, h_2)$ is appreciably greater than $\text{Coh}(e, h_1)$; the greater the difference is, the more frequent the occurrence is. Note also that $\text{Coh}(e, h_1) < \text{Coh}(e, h_1, h_2)$ if and only if $\text{Coh}(e, h_1) < \text{Coh}(e \wedge h_1, h_2)$ from the proof above. This means that we can compare $\text{Coh}(e, h_1)$ and $\text{Coh}(e \wedge h_1, h_2)$ to predict the occurrence of the conjunction fallacy. It is also helpful to note that $\text{Coh}(e \wedge h_1, h_2) = \text{Coh}(h_1, h_2|e)\text{Coh}(h_2, e)$, where $\text{Coh}(h_1, h_2|e)$ is the degree of pairwise coherence between $h_1$ and $h_2$ against the background $e$, as follows:

$$\begin{aligned}
\text{Coh}(e \wedge h_1, h_2) &= \frac{\Pr(e \wedge h_1 \wedge h_2)}{\Pr(e \wedge h_1)\Pr(h_2)} \\
&= \frac{\Pr(h_1 \wedge h_2 \mid e)\Pr(e)}{\Pr(h_1 \mid e)\Pr(e)\Pr(h_2)} \\
&= \frac{\Pr(h_1 \wedge h_2 \mid e)\Pr(e)\Pr(h_2 \mid e)}{\Pr(h_1 \mid e)\Pr(e)\Pr(h_2)\Pr(h_2 \mid e)} \\
&= \frac{\Pr(h_1 \wedge h_2 \mid e)\Pr(h_2 \wedge e)}{\Pr(h_1 \mid e)\Pr(h_2 \mid e)\Pr(e)\Pr(h_2)} \\
&= \text{Coh}(h_1, h_2 \mid e)\text{Coh}(h_2, e)
\end{aligned}$$

To start with the Linda Scenario of the M-A paradigm, the description ($e$) of Linda reduces the probability of the bank-teller hypothesis ($h_1$) considerably. Coh($e$, $h_1$) is therefore much smaller than the neutral point, one. Meanwhile, it is not immediately clear whether $e \wedge h_1$ and $h_2$ are pairwise coherent—one of the conjuncts ($e$) is strongly coherent with $h_2$ but the other ($h_1$) is at least mildly incoherent with $h_2$. It turns out, however, that this is unimportant for the purpose of making the prediction. What counts for the prediction is not whether $e \wedge h_1$ and $h_2$ are coherent, but whether they are more coherent than the two conjuncts $e$ and $h_1$ are. The answer to this question is clear: Coh($e \wedge h_1$, $h_2$) = Coh($h_1$, $h_2|e$)Coh($h_2$, $e$) is much greater than Coh($e$, $h_1$) since the first multiplicand Coh($h_1$, $h_2|e$) is no smaller than Coh($e$, $h_1$), and the second multiplicand Coh($h_2$, $e$) is much greater than the neutral point, one. More generally, in the M-A paradigm Coh($e \wedge h_1$, $h_2$) = Coh($h_1$, $h_2|e$)Coh($h_2$, $e$) may or may not be greater than one, but it is appreciably greater than Coh($e$, $h_1$). It follows that Coh($e$, $h_1$, $h_2$) is appreciably greater than Coh($e$, $h_1$), and the coherence account correctly predicts the common occurrence of the conjunction fallacy in the M-A paradigm.

In the Health Survey Scenario, and more generally in the A-B paradigm, the salient feature is the positive correlation between the two hypotheses $h_1$ and $h_2$. For example, the hypothesis that the selected person has had one or more heart-attacks ($h_1$) is coherent with the hypothesis that this person is older than 55 ($h_2$). The high degree of coherence remains against the background that the person is a randomly selected adult male ($e$). So, Coh($h_1$, $h_2|e$) is significantly higher than one. Meanwhile, Coh($e$, $h_1$) and Coh($e$, $h_2$) are not much different. As a result, Coh($e \wedge h_1$, $h_2$) = Coh($h_1$, $h_2|e$)Coh($h_2$, $e$) is appreciably greater than Coh($e$, $h_1$). It follows that Coh($e$, $h_1$, $h_2$) is appreciably greater than Coh($e$, $h_1$), and the coherence account correctly predicts the common occurrence of the conjunction fallacy in the A-B paradigm.

Finally in the Carol-Poetry Scenario, which does not induce the conjunction fallacy, the description ($e$) of Carol is coherent with both the poetry hypothesis ($h_1$) and the foreign correspondent hypothesis ($h_2$). This makes both Coh($e$, $h_1$) and Coh($e$, $h_2$) significantly greater than one, and not much different from each other. Meanwhile, $h_1$ and $h_2$ are coherent with each other, but not so against the background $e$, i.e. adding $h_1$ to the description $e$ does not change the probability of $h_2$ much, which is already high given $e$ alone. This makes Coh($h_1$, $h_2|e$) close to the neutral value, one. As a result, Coh($e \wedge h_1$, $h_2$) = Coh($h_1$, $h_2|e$)Coh($h_2$, $e$) is not much different from Coh($e$, $h_1$). It follows that Coh($e$, $h_1$, $h_2$) is *not* appreciably greater than Coh($e$, $h_1$). The coherence account therefore predicts, in line with the data, that Carol-Poetry Scenario does not induce the conjunction fallacy.

## 7. Conclusion

There have been many empirical studies of the conjunction fallacy by psychologists and analyses of its import by philosophers, but many of them focused on the M-A paradigm. We believe it is time to expand the scope and attempt a unified account. In this paper we proposed a unified Bayesian account of the conjunction fallacy by coherence, and examined its empirical adequacy in both the M-A paradigm and the A-B paradigm. Our account is natural, simple and readily testable in comparison with the only extant attempt at unification in the literature, TCR's confirmation model, which is strained, complex and untestable as it stands. In the course of working out the details of the coherence account, we proposed a new Bayesian measure of

coherence which improves upon Shogenji's measure. We are under no illusion that the account we proposed in this paper is the final word on the conjunction fallacy. Further evidence may indicate the two paradigms should be treated separately; some version of the confirmation account may be found that is more attractive. But we believe the coherence account and the results obtained in this paper point in promising new directions of research. The coherence account needs to be tested further on various scenarios for possible refinement.[24] It is also important to investigate whether the coherence account or some extension of it is applicable to similar phenomena, especially the base rate neglect.

**Acknowledgment** We would like to thank anonymous referees of this journal for detailed comments and helpful suggestions.

## Bibliography

Bar-Hillel, Maya. "The base-rate fallacy in probability judgments." *Acta Psychologica* 44.3 (1980): 211-233.

Bovens, Luc, and Hartmann, Stephan. *Bayesian Epistemology* (Oxford: Oxford University Press 2003).

Burnham, Kenneth P., and David R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (New York: Springer-Verlag).

Carnap, Rudolph. *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950).

Chisholm, Roderick. *Theory of Knowledge* (Englewood Cliffs NJ: Prentice Hall, 1966).

Crupi, Vincenzo, Branden Fitelson, and Katya Tentori. "Probability, confirmation, and the conjunction fallacy." *Thinking & Reasoning* 14.2 (2008): 182-199.

Douven, Igor, and Wouter Meijs. "Measuring coherence." *Synthese* 156.3 (2007): 405-425.

Fitelson, Branden. "A probabilistic theory of coherence." *Analysis* 63 (2003): 194-199.

Forster, Malcolm, and Elliott Sober. "How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions." *British Journal for the Philosophy of Science*, 45.1 (1994), 1-35.

Glass, David H. "Coherence, explanation, and Bayesian networks." M. O'Neal, R. F. E. Sutcliffe, C. Ryan, M. Eaton, and N. J. L. Griffith (eds.) *Artificial intelligence and cognitive science* (New York: Springer-Verlag 2002): 177-182.

Gigerenzer, Gerd. "On narrow norms and vague heuristics: A reply to Kahneman and Tversky." *Psychological Review* 103 (1996): 592-596.

Jönsson, Martin L., and Elias Assarsson. "A Problem for confirmation theoretic accounts of the conjunction fallacy." *Philosophical Studies* 173.2 (2016): 437-449.

[24] See, for instance, the Inverse Conjunction Fallacy due to Jönsson and Hampton (2006). See Jönsson and Assarsson (2015) for some problems that this fallacy gives rise to for confirmation theoretic accounts of the conjunction fallacy.

Jönsson, Martin L., and James A. Hampton. “The inverse conjunction fallacy.” *Journal of Memory and Language*, 33.5 (2006): 317-334.

Koehler, Jonathan J. "The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges." *Behavioral and Brain Sciences* 19.01 (1996): 1-17.

Koscholke, Jakob. “Evaluating test cases for probabilistic measures of coherence.” *Erkenntnis* 81.1 (2016): 155-181.

Koscholke, Jakob, and Michael Schippers. “Against relative overlap measures of coherence.” *Synthese* 193 (2016): 2805-2814.

Lewis, Clarence I. *An Analysis of Knowledge and Valuation* (La Salle IL: Open Court, 1946).

Olsson, Erik J. "What is the problem of coherence and truth?" *The Journal of Philosophy* 99.5 (2002): 246-272.

Olsson, Erik J. *Against Coherence: Truth, Probability, and Justification* (Oxford: Oxford University Press 2005).

Roche, William. "Coherence and probability: A probabilistic account of coherence" in Michał Araszkiewicz and Jaromír Šavelka (eds.) *Coherence: Insights from Philosophy, Jurisprudence and Artificial Intelligence* (New York: Springer-Verlag 2013), 59-91.

Roche, William, and Tomoji Shogenji. "Dwindling confirmation." *Philosophy of Science* 81.1 (2014): 114-137.

Schippers, Michael. “Competing accounts of contrastive coherence.” *Synthese* 193.10 (2016): 3383-3395.

Schupbach, Jonah N. "New hope for Shogenji's coherence measure." *The British Journal for the Philosophy of Science* 62.1 (2011): 125-142.

Shafir, Eldar, Edward B. Smith, and Daniel Osherson. "Typicality and reasoning fallacies." *Memory & Cognition* 18.3 (1990): 229-239.

Shogenji, Tomoji. "Is coherence truth conducive?" *Analysis* 59.264 (1999): 338-345.

Shogenji, Tomoji. "The degree of epistemic justification and the conjunction fallacy." *Synthese* 184.1 (2012): 29-48.

Sides, Ashley, Daniel Osherson, Nicolao Bonini, and Riccardo Viale. "On the reality of the conjunction fallacy." *Memory & Cognition* 30.2 (2002): 191-198.

Siebel, Mark. "There’s something about Linda: Probability, coherence and rationality." *First Salzburg Workshop on Paradigms of Cognition, Salzburg* (2002). http://www.uni-oldenburg.de/fileadmin/user_upload/philosophie/download/Mitarbeiter/Siebel/Siebel_Linda.pdf

Sober, Elliott. *Ockham's Razors: A User's Manual* (Cambridge UK: Cambridge University Press 2015).

Tentori, Katya, Vincenzo Crupi, and Selena Russo. "On the determinants of the conjunction fallacy: Confirmation versus probability." *Journal of Experimental Psychology: General* 142.1 (2013): 235-255.

Thagard, Paul. “Explanatory coherence.” *Behavioral and Brain Sciences* 12 (1989): 435–502.

Thagard, Paul. *Conceptual Revolutions* (Princeton: Princeton University Press 1992).

Tversky, Amos, and Daniel Kahneman. "Extensional versus intuitive reasoning: The conjunction fallacy in probability.