How to Confirm the Disconfirmed

On conjunction fallacies and robust confirmation

David Atkinson, Jeanne Peijnenburg and Theo Kuipers

Abstract

Can some evidence confirm a *conjunction* of two hypotheses more than it confirms *either* of the hypotheses separately? We show that it can, moreover under conditions that are the same for nine different measures of confirmation. Further we demonstrate that it is even possible for the conjunction of two *disconfirmed* hypotheses to be *confirmed* by the same evidence.

Key words: Probability, confirmation, conjunction fallacy.

1 Introduction

Jack Author has just made an important discovery. From his calculations it follows that recent evidence e supports the conjunction of two popular hypotheses, h_1 and h_2. With great gusto he sets himself to the writing of a research proposal in which he explains his idea and asks for time and money to work out all its far-reaching consequences. Author's proposal is sent to Jill Reviewer, who however writes a devastating report. Reviewer first recalls what is common knowledge within the scientific community, namely that e strongly *dis*confirms not only h_1, but h_2 as well. Then Reviewer intimates that Author is clearly not familiar with the relevant literature; for if he were he would have realized that any calculation that results in confirming the conjunction of two disconfirmed hypotheses must contain a mistake. At any rate he should never have launched this preposterous idea, which will make him the laughing stock of his peers.

Is Reviewer right? Did Author indeed make a blunder by assuming that e might confirm a conjunction of hypotheses, $h_1 \wedge h_2$, given that the same e

disconfirms h_1 and h_2 separately? In this paper we will argue that Author may not have been mistaken. If c is some generic confirmation function, then it can happen that $c(h_1 \wedge h_2, e)$ is confirmed — which here means greater than zero — while $c(h_1, e)$ and $c(h_2, e)$ are both disconfirmed — here less than zero — while zero corresponds to neutrality. Jack Author might have written a defensible proposal after all.

Our paper can be seen as a reinforcement of recent claims made by Crupi, Fitelson and Tentori (2007). In this very stimulating paper, Crupi, Fitelson and Tentori (CFT) introduce the two inequalities

$$c(h_1, e) \leq 0 \quad \& \quad c(h_2, e|h_1) > 0\,.^1 \tag{1}$$

Next they convincingly argue that (1) is a sufficient condition for

$$c(h_1 \wedge h_2, e) > c(h_1, e)\,. \tag{2}$$

In other words, if e disconfirms h_1 but confirms h_2 when h_1 is true, then e gives more confirmation to $h_1 \wedge h_2$ than to h_1 alone — although it confirms h_2 alone even more: $c(h_2, e) > c(h_1 \wedge h_2, e)$.

The fact that e can give more confirmation to $h_1 \wedge h_2$ than to one of its conjuncts is of course intriguing, since the corresponding claim for conditional probabilities is false. Under no condition whatsoever can it be true that

$$P(h_1 \wedge h_2|e) > P(h_1|e)\,. \tag{3}$$

Indeed, (3) instantiates the notorious conjunction fallacy. Following a suggestion by Sides et al. (2002), CFT surmise that the conjunction fallacy might arise from a confusion between (2) and (3). We find this idea plausible and promising, but we believe that its scope can be significantly expanded. For if we replace CTF's condition (1) by another sufficient condition, then we are able to obtain a much stronger result. In this paper we will argue that if we take as condition

$$P(h_1 \wedge \neg h_2|e) = 0 \quad \& \quad P(\neg h_1 \wedge h_2|e) = 0\,, \tag{4}$$

then it can be shown that

$$c(h_1 \wedge h_2, e) \geq c(h_1, e) \quad \& \quad c(h_1 \wedge h_2, e) \geq c(h_2, e)\,. \tag{5}$$

[1] The notation '$c(h_2, e|h_1)$' is used by CFT to mean '$c(h_2, e)$ on the assumption that h_1 is true'. As we will further explain in footnote 2, $c(h_2, e|h_1)$ is not the same as $c(h_2, e \wedge h_1)$.

There are two senses in which (5) is more general, and cuts deeper than CFT's result (2). Firstly, (5) implies that there is a situation in which the conjunction $h_1 \wedge h_2$ is more highly confirmed than *either* of the conjuncts – whereas (2) implies that *only one* of the conjuncts receives less confirmation than the conjunction. Secondly, (5) is also applicable when both h_1 and h_2 are confirmed – whereas (2) requires that one of the hypotheses must be disconfirmed. Hence CFT's explanatory net can in fact be cast much wider, since there appear to be more ways in which the occurrence of a conjunction fallacy might actually be guided by a sound assessment of a confirmation relation.

The condition (4) states that, if e is the case, then either both h_1 and h_2 are true, or neither is: $P(h_1 \leftrightarrow h_2|e) = 1$. This is sufficient for (5), as we will prove in Sect. 2 and in Appendix A. However, it is by no means necessary, as will become clear from Sect. 4 and Appendix B, cf. inequality (12). There we describe a sufficient condition under which $c(h_1 \wedge h_2, e)$ is positive, but both $c(h_1, e)$ and $c(h_2, e)$ are negative:

$$c(h_1 \wedge h_2, e) > 0 \quad \& \quad c(h_1, e) < 0 \quad \& \quad c(h_2, e) < 0\,. \tag{6}$$

Note that (6) describes the case in which Jack Author would after all be vindicated. As we will explain later, the sufficient condition for (6) is more detailed than (4). In fact, as we will see from (12), it consists of a relaxed version of (4) – in the sense that small, nonzero values of the two conditional probabilities are tolerated – plus some extra constraints. Since (6) implies (5) but not the other way around, the sufficient condition for (6) also suffices for (5). In this sense a relaxed version of (4), together with some contraints that we will spell out later, would still be enough to deduce (5).

2 Robust Confirmation

CFT justly emphasize that their analysis is *robust*: it holds for various specifications of the generic function $c(h, e)$, that is for various measures of confirmation (cf. Fitelson 1999). These authors list six prevailing confirmation measures and they prove that their conclusions follow for any of them. We will demonstrate that our argument also goes through robustly in this sense. Indeed, we will add two more measures of confirmation to those listed by CFT, making eight measures in all. The eight measures in question are the following, where as usual $P(h|e)$ denotes a conditional, and $P(h)$ an unconditional probability:

$$\begin{aligned}
C(h,e) &= P(h \wedge e) - P(h)P(e) \\
D(h,e) &= P(h|e) - P(h) \\
S(h,e) &= P(h|e) - P(h|\neg e) \\
Z(h,e) &= \frac{P(h|e) - P(h)}{P(\neg h)} \quad \text{if} \quad P(h|e) \geq P(h) \\
&= \frac{P(h|e) - P(h)}{P(h)} \quad \text{if} \quad P(h|e) < P(h) \\
R(h,e) &= \log \left[\frac{P(h|e)}{P(h)} \right] \\
L(h,e) &= \log \left[\frac{P(e|h)}{P(e|\neg h)} \right] \\
N(h,e) &= P(e|h) - P(e|\neg h) \\
F(h,e) &= \frac{P(h|e) - P(h|\neg e)}{P(h|e) + P(h|\neg e)} \, .
\end{aligned}$$

The first six of these measures have been discussed by CFT. Measure C is attributed to Carnap (1950), D to Carnap (1950) and Eells (1982), S to Christensen (1999) and Joyce (1999), Z to Crupi, Tentori and Gonzalez (2007), R to Keynes (1921) and Milne (1996), and L to Good (1950) and Fitelson (2001).

The two extra measures that we have added are N and F. Measure N has been taken from Tentori et al. (2007), who attribute it to Nozick (1981). Note that N is similar to S, the only difference being in the positioning of h and e. F is inspired by a measure introduced by Fitelson (2003). The latter is in fact a measure of *coherence* rather than of confirmation. However, it is well known that there are close conceptual connections between coherence and confirmation, and confirmation measures are sometimes used to indicate coherence (cf. Douven and Meijs 2006). For example, Carnap's confirmation measure D, to which Carnap himself gives special attention in (Carnap 1962), is the favourite measure of coherence of Douven and Meijs (2007). Further, the exponent of the confirmation measure R of Keynes and Milne, $\exp R(h,e)$, is equal to a coherence measure of Shogenji (1999).

CFT succeed in showing that their inference from (1) to (2) remains valid under any of their six measures of confirmation. Similarly, we can prove that our inference from (4) to (5) remains valid as one passes from one of the eight specifications of $c(h, e)$ to another. Take for example the case in which $c(h, e)$ is specified as the Carnap-Eells measure $D(h, e)$. We prove in Appendix A that, if our condition (4) is fulfilled, then

$$D(h_1 \wedge h_2, e) - D(h_1, e) = P(h_1 \wedge \neg h_2 \wedge \neg e) \geq 0\,, \tag{7}$$

and similarly with h_1 and h_2 interchanged. Clearly, if (7) holds, then $D(h_1 \wedge h_2, e) \geq D(h_1, e)$, which is the first half of our conclusion (5), with D substituted for c. An analogous argument applies to h_2, and this will give us the second half of (5).[2]

Similarly, but with more effort, it can be shown that all the remaining specifications of c in terms of C, S, N, R, L, Z and F will do the trick: they lead to quotients of products of probabilities that are necessarily nonnegative. In this manner we will have robustly deduced the two inequalities of (5) from (4). Full details are given in Appendix A, where in passing we also explain under which additional conditions (5) is deducible with $>$ in place of $\geq$.[3]

As said above, the condition (4) is sufficient but by no means necessary for the conclusion (5). For (5) still follows robustly if the probabilities in (4) are small but not zero, and if some extra constraints are in force. In Sect. 4 and in Appendix B we will formulate upper bounds for $P(h_1 \wedge \neg h_2|e)$ and $P(\neg h_1 \wedge h_2|e)$ as well as extra constraints such that (5) still holds robustly, i.e. under any of the eight confirmation measures in our list. But first, in Sect. 3, we will illustrate the validity of our inference from (4) to (5) with some examples. The purpose of this exercise is to make the inference intuitively reasonable and to explain its connection to a (particular type of) conjunction fallacy.

---

[2]In terms of D, the second inequality in CFT's condition (1) reads

$$D(h_2, e|h_1) \equiv P(h_2|e \wedge h_1) - P(h_2|h_1) \geq 0\,.$$

This function is not the same as $D(h_2, e \wedge h_1) = P(h_2|e \wedge h_1) - P(h_2)$.

[3]In Appendix C we present yet another specification of c — the ninth — under which our inference from (4) to (5) goes through. In that appendix we discuss a coherence measure of Bovens and Hartmann (2003a and 2003b), which can be treated as a comparative measure of confirmation. Since the approach of Bovens and Hartmann is radically different from the above eight cases, we do not consider the Bovens-Hartmann measure here, but devote a separate appendix to it. See also Meijs and Douven (2005) for a critique of the Bovens-Hartmann measure, and the reply of Bovens and Hartmann (2005).

3 Conjunction Fallacies

Suppose a die is cast in secrecy and the number that came up is recorded by the gamemaster. Consider two gamblers who entertain different hypotheses about what the number is. Hypothesis h_1 is that the number is 1, 2 or 3, whereas h_2 is that it is 2, 3, or 4. We conclude that

$$P(h_1) = P(h_2) = \tfrac{3}{6} = \tfrac{1}{2}.$$

Suppose next that the gamemaster provides the clue that the number is prime. This can be treated as incoming evidence $e = \{2, 3, 5\}$, with $h_1 = \{1, 2, 3\}$ and $h_2 = \{2, 3, 4\}$. Since e contains three primes, whereas h_1 and h_2 contain only two primes apiece, we find for the conditional probabilities

$$P(h_1|e) = P(h_2|e) = \tfrac{2}{3}.$$

Let us now work out this example in terms of the Carnap-Eells measure D, keeping in mind that similar results apply to any of the other eight measures. Since by definition $D(h, e) = P(h|e) - P(h)$, it is the case that

$$D(h_1, e) = P(h_1|e) - P(h_1) = \tfrac{2}{3} - \tfrac{1}{2} = \tfrac{1}{6},$$

and similarly $D(h_2, e) = \frac{1}{6}$. However $h_1 \wedge h_2 = \{2, 3\}$, so $P(h_1 \wedge h_2|e) = \frac{2}{3}$ too, but $P(h_1 \wedge h_2) = \frac{2}{6} = \frac{1}{3}$. Therefore the degree of confirmation of $h_1 \wedge h_2$ is

$$D(h_1 \wedge h_2, e) = P(h_1 \wedge h_2|e) - P(h_1 \wedge h_2) = \tfrac{2}{3} - \tfrac{1}{3} = \tfrac{1}{3},$$

which is twice as large as the degrees of confirmation of h_1 or h_2 alone. So

$$D(h_1 \wedge h_2, e) > D(h_1, e) \quad \& \quad D(h_1 \wedge h_2, e) > D(h_2, e),$$

and this is our conclusion (5), with D substituted for c, and with $\geq$ replaced by $>$. Note that $\neg h_1 = \{4, 5, 6\}$, so $\neg h_1 \wedge h_2 = \{4\}$, and since 4 is not a prime number, $P(\neg h_1 \wedge h_2|e) = 0$. Similarly $P(h_1 \wedge \neg h_2|e) = 0$, so the two equalities of our condition (4) are in fact respected.

It is of course intuitively clear why $e = \{2, 3, 5\}$ should confirm $h_1 \wedge h_2 = \{2, 3\}$ more than $h_1 = \{1, 2, 3\}$ or $h_2 = \{2, 3, 4\}$, for $h_1 \wedge h_2$ contains only primes, whereas h_1 and h_2 are each 'diluted' by a nonprime. However, by the same token it also seems clear how this example can trigger the commission of a conjunction fallacy. Imagine that subjects are given the following information: "A gamemaster casts a die. He records the number, but does not tell anybody what it is. The only thing he makes known is that the number is prime (e)." Suppose that now the question is posed: "What do you think is more probable? That the number recorded by the gamemaster is 1 or 2 or 3 (h_1)? That it is 2 or 3 or 4 (h_2)? Or that it is 2 or 3 ($h_1 \wedge h_2$)?" It is quite likely that many people would choose the last option, thereby committing a conjunction fallacy (although they would have been right had the question been about confirmation rather than probability).

Here is another example. Imagine that you have a little nephew of whom you are very fond. One day you receive an email from his mother, telling you that the child is suffering a severe bout of measles (e). You immediately decide to visit the boy, bringing him some of the jigsaw puzzles that you know he likes so much. What do you think is more probable to find upon your arrival? That the child has a fever (h_1), that he has red spots all over his body (h_2), or that he has a fever and red spots ($h_1 \wedge h_2$)? Again we expect that many people will opt for the third possibility. Although this answer is fallacious, the nephew story does instantiate our inference from (4) to (5). For under the assumption that measles always comes with fever *and* red spots, so that (4) is satisfied, our story makes it true that

$$c(h_1 \wedge h_2, e) \geq c(h_1, e) \quad \& \quad c(h_1 \wedge h_2, e) \geq c(h_2, e)\,.$$ [4]

It should be noted that the conjunction fallacies committed in the two examples above differ from those discussed by CFT. For CFT focus exclusively on conjunction fallacies of the familiar Linda-type: when hearing about a person named Linda, who is 31 years old, single, bright and outspoken, concerned about discrimination and social justice, involved in anti-nuclear demonstrations (e), most people find 'Linda is a bank teller and is active in the feminist movement' ($h_1 \wedge h_2$) more probable than 'Linda is a bank teller' (h_1). Linda-type conjunction arguments have the following characteristics in common:

[4] The assumption that there are no measles without both fever and red spots does *not* imply

$$c(h_1 \wedge h_2, e) = c(h_1, e) = c(h_2, e),$$

although it does entail

$$P(h_1 \wedge h_2|e) = P(h_1|e) = P(h_2|e)\,,$$

which in fact are all equal to 1. However, it is less probable that a patient has fever and spots than that he has just one of these afflictions, if one does *not* conditionalize on his having measles:

$$P(h_1 \wedge h_2) < P(h_1) \quad \& \quad P(h_1 \wedge h_2) < P(h_2)\,,$$

and therefore $c(h_1 \wedge h_2, e)$ is greater than either $c(h_1, e)$ or $c(h_2, e)$.

(i) e is negatively (if at all) correlated with h_1;

(ii) e is positively correlated with h_2, *even conditionally on* h_1;

(iii) h_1 and h_2 are mildly (if at all) negatively correlated

(Crupi, Fitelson and Tentori 2007, forthcoming, Sect. 3 – emphasis by the authors). It is solely in relation to fallacies of this type that CFT submit their idea that people may actually rely on assessments of confirmation when judging probabilities. Indeed, CFT's condition (1), that is robustly sufficient for their conclusion (2), is an "appropriate confirmation-theoretic rendition of (i) and (ii)" (ibid.).

In Sect. 1 we have already intimated that there are two differences between CFT's conclusion (2) and our result (5). The first was that (5) can handle the case in which the conjunction $h_1 \wedge h_2$ might be more confirmed than *either* of its conjuncts, the second that (5) also applies when *both* h_1 and h_2 are confirmed, i.e. $c(h_1, e)$ and $c(h_2, e)$ are positive. We are now in a position to understand that these differences stem from the dissimilarity between Linda-type fallacies on the one hand and 'our' conjunction fallacies on the other. Linda-type fallacies treat h_1 and h_2 asymmetrically in the sense that one must be disconfirmed, while the other is confirmed. In our case, by contrast, all options are possible: both may be confirmed or both disconfirmed, or indeed one may be confirmed and the other disconfirmed. Moreover, in Linda examples the confirmation degree of the conjunction lies between those of the two conjuncts. In our case, on the other hand, the confirmation degree of the conjunction is greater than that of either of the conjuncts.

Given this dissimilarity, it is perhaps appropriate to say that Linda-like problems are best analyzed by CFT's inference from (1) to (2), whereas 'our' conjunction fallacies are best analyzed by our inference from (4) to (5). In both cases the fallacious arguments are explained by pointing to a confusion between probability judgements and confirmation assessments. And both cases satisfy the minimal constraint that the confirmation degree of the conjunction is greater than that of at least one of the conjuncts.

It is to be expected that there are many different kinds of conjunction fallacy, all of which will satisfy this minimal contraint. We suspect that, for each kind of fallacy, it will be possible to reconstruct and spell out the corresponding legitimate inference that ought to take the place of the fallacious probabilistic claim.

4 Why Jack Author Can Be Correct

In this section we will formulate a sufficient condition for (6). That is, we will explain how Jack Author might be correct when he claims that a conjunction of two disconfirmed hypotheses can itself be confirmed. In the process, it will become clear that the rather strict condition (4) is not necessary for obtaining (5). Indeed, (5) is consistent with much looser forms of (4), viz. ones in which the conditional probabilities $P(h_1 \wedge \neg h_2|e)$ and $P(\neg h_1 \wedge h_2|e)$ are positive and can be numerically different from one another. Granted, it is only under certain restrictions that (5) follows from weaker versions of (4). But we will show that a precise specification can be given of these restrictions, together with bounds for $P(h_1 \wedge \neg h_2|e)$ and $P(\neg h_1 \wedge h_2|e)$.

Since we are dealing with two hypotheses and one piece of evidence, the following eight triple probabilities exhaust all the possibilities that are open to us:

$$\begin{array}{ll} P(h_1 \wedge h_2 \wedge e) & P(\neg h_1 \wedge h_2 \wedge e) \\ P(h_1 \wedge \neg h_2 \wedge e) & P(\neg h_1 \wedge \neg h_2 \wedge e) \\ P(h_1 \wedge h_2 \wedge \neg e) & P(\neg h_1 \wedge h_2 \wedge \neg e) \\ P(h_1 \wedge \neg h_2 \wedge \neg e) & P(\neg h_1 \wedge \neg h_2 \wedge \neg e) \end{array}$$

Note that all the (un)conditional probabilities and all the (un)conditional degrees of confirmation are functions of the above eight probabilities. Hence looking for (in)equalities between confirmation degrees and probability functions reduces itself to (in)equalities between the values of these eight triple probabilities. Because these triples are probabilities, they cannot be negative; and because there are not more triples than these eight, their sum is unity.

Since $P(\neg h_1 \wedge \neg h_2 \wedge \neg e)$ describes the area outside the Venn diagram (see Appendix A), it is not very interesting, and we will set its value to 1 minus the sum of the remaining seven triples. These seven triples are independent of one another, subject only to the requirement that their sum be not greater than 1. Two of them, $P(h_1 \wedge \neg h_2 \wedge e)$ and $P(\neg h_1 \wedge h_2 \wedge e)$, are both zero under our condition (4). However, here we will relax (4) and allow them to be positive. We shall restrict our attention to the case in which these two triples are equal:

$$z \;=\; P(h_1 \wedge \neg h_2 \wedge e) = P(\neg h_1 \wedge h_2 \wedge e)\,, \tag{8}$$

where z need not be zero. There are now five other independent triples left, and to further reduce the search to manageable proportions we will give them all the same numerical value:

$$\begin{aligned} x &= P(h_1 \wedge h_2 \wedge e) = P(\neg h_1 \wedge \neg h_2 \wedge e) \\ &= P(h_1 \wedge h_2 \wedge \neg e) = P(\neg h_1 \wedge h_2 \wedge \neg e) = P(h_1 \wedge \neg h_2 \wedge \neg e)\,. \end{aligned} \tag{9}$$

We stress that the artifices (8)-(9) are purely for convenience: they limit the search from 7 to 2 dimensions. Many more possibilities remain open: no matter, our ambition is only to find a sufficient (not a necessary) condition for $h_1 \wedge h_2$ to be confirmed, while both h_1 and h_2 are disconfirmed.

The key question that we now have to ask ourselves is:

Can we find values of z and x such that (6) follows?

If we can, then we will have discovered a sufficient condition for $c(h_1 \wedge h_2, e)$ to be positive (confirmation), while both $c(h_1, e)$ and $c(h_2, e)$ are negative (disconfirmation). Hence we would have shown that there is at least one way in which Jack Author could have written a defensible reaearch proposal. Moreover, we would also have demonstrated that our general conclusion (5) does not require the rather rigorous condition (4), but is also compatible with a modified form of that condition. For if (6) were to follow from a relaxed version of (4), then so would (5), since the latter follows from (6).

In Appendix B we prove that the answer to our key question above is yes. In particular we show that, if $0 \leq z \leq \frac{1}{12}$, and

$$\tfrac{1-2z}{6} < x < \tfrac{1}{5} - \tfrac{2z}{5}\,,$$

then (6) holds. Actually there is more, for z may be a little bigger, as large as $\frac{1}{8}$, but then the upper limit on x decreases. If $\frac{1}{12} \leq z \leq \frac{1}{8}$, the restrictions on x are

$$\tfrac{1-2z}{6} < x < \tfrac{1}{4} - z\,.$$

A proof of what we might call the 'Jack Author Theorem' can be found in Appendix B.

In Fig. 1 we show the area in which Jack Author's proposal would made sense. The straight diagonal line in Fig. 1 gives the lower bound of x for values of z between 0 and $\frac{1}{8}$, while the bent line gives the upper bound for the x-values. For every x and z between these lines the inequalities (6) are observed. The region in which the Jack Author effect occurs might perhaps look small, but in fact it is quite large. If $z = 0$, which corresponds to the strict condition (4), x can be between $\frac{1}{6}$ and $\frac{1}{5}$, as can be seen from Fig. 1. Since the 5 triples given in (9) are all equal to one another, their sum can be between $\frac{5}{6}$ and 1, which is a large part of the whole probability space. True, at the maximum allowed value of z, namely $\frac{1}{8}$, the allowed values for x shrink to the point $x = \frac{1}{8}$; but at $z = \frac{1}{12}$, for instance, which corresponds to the kink in the bent line of Fig. 1, one finds $\frac{5}{36} < x < \frac{1}{6}$.

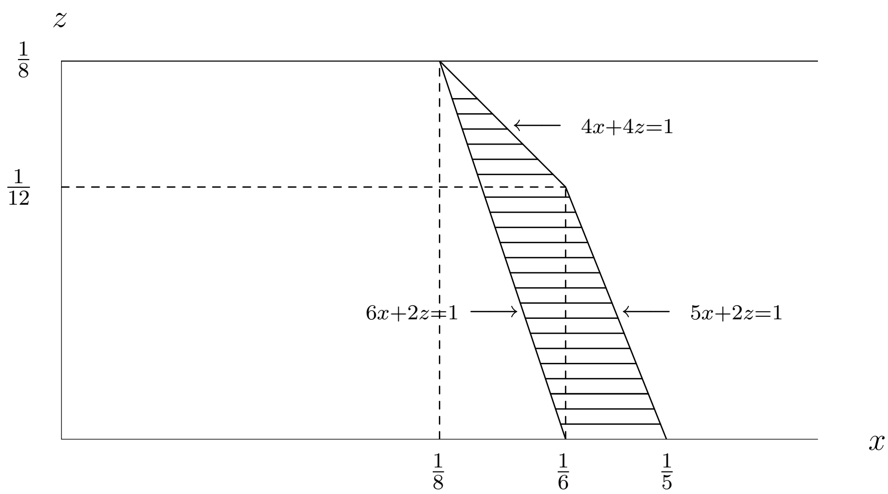
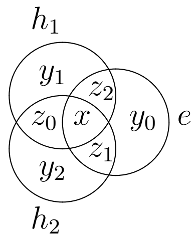


Figure 1: Allowed region of x-z plane for Jack Author effect

Affiliation:
Faculty of Philosophy, University of Groningen
Oude Boteringestraat 52, 9712 GL Groningen
The Netherlands

Email:
d.atkinson@rug.nl
jeanne.peijnenburg@rug.nl
t.a.f.kuipers@rug.nl

Appendix A: Proof of confirmation



Here is a compact notation for the probabilities of the elementary intersections of the sets h_1, h_2, e, as displayed pictorially in the Venn diagram:

$$
\begin{array}{rcl}
 & x = P(h_1 \wedge h_2 \wedge e) & \\
y_0 = P(\neg h_1 \wedge \neg h_2 \wedge e) & y_1 = P((h_1 \wedge \neg h_2 \wedge \neg e) & y_2 = P(\neg h_1 \wedge h_2 \wedge \neg e) \\
z_0 = P(h_1 \wedge h_2 \wedge \neg e) & z_1 = P(\neg h_1 \wedge h_2 \wedge e) & z_2 = P(h_1 \wedge \neg h_2 \wedge e)
\end{array}
$$

An important point is that these seven triple probabilities are nonnegative, i.e. each one must be positive or zero. Further their sum cannot be larger than unity, for

$$x + y_0 + y_1 + y_2 + z_0 + z_1 + z_2 = 1 - P(\neg h_1 \wedge \neg h_2 \wedge \neg e) \leq 1 \,. \qquad (10)$$

Under this global condition (10) the seven triple probabilities are independent variables, spanning a hypervolume in seven dimensions.

The purpose of the calculations in this appendix is to show that, under the restrictions $z_1 = 0 = z_2$, it is the case that $c(h_1 \wedge h_2, e) \geq c(h_1, e)$ and $c(h_1 \wedge h_2, e) \geq c(h_2, e)$, for all of the eight realizations of the generic measure c, that is for C, D, S, Z, R, L, N, F. To do this, it will be proved, for each of the measures, that $c(h_1 \wedge h_2, e) - c(h_1, e)$ can be reduced to an expression involving the five remaining independent triple probabilities that is manifestly nonnegative. Terms that can be recognized as being nonnegative are of course products of triple probabilities with a plus sign, or terms involving $1 - \Sigma$, where Σ is a triple probability or the sum of two or more of them (cf. inequality (10)). Once that has been done the job is finished, for then $c(h_1 \wedge h_2, e) \geq c(h_1, e)$ has been demonstrated; and $c(h_1 \wedge h_2, e) \geq c(h_2, e)$ follows immediately by interchanging the suffices 1 and 2 throughout the proof.

Let us begin with the Carnap measure C. We read off from the Venn diagram that

$$
\begin{aligned}
C(h_1 \wedge h_2, e) - C(h_1, e) &= x - (x + z_0)(x + y_0) - x + (x + y_1 + z_0)(x + y_0) \\
&= (x + y_0) y_1 \,.
\end{aligned}
$$

Since $(x + y_0)y_1$ is manifestly nonnegative, we conclude that $C(h_1 \wedge h_2, e) \geq C(h_1, e)$. This concludes the demonstration for this simple measure.

The Carnap-Eells measure D and Christensen's S are not much harder:

$$D(h_1 \wedge h_2, e) - D(h_1, e) = \frac{C(h_1 \wedge h_2, e) - C(h_1, e)}{P(e)} = \frac{(x + y_0)y_1}{P(e)} \geq 0 \,,$$

Since $(x + y_0) = P(e)$, given that $z_1 = 0 = z_2$, and $y_1 = P(h_1 \wedge \neg h_2 \wedge \neg e)$, we have thereby proved inequality (7). Similarly,

$$S(h_1 \wedge h_2, e) - S(h_1, e) = \frac{C(h_1 \wedge h_2|e) - C(h_1|e)}{P(e)P(\neg e)} = \frac{(x + y_0)y_1}{P(e)P(\neg e)} \geq 0 \,.$$

For the Z measure of Crupi, Tentori and Gonzalez, we have to distinguish the cases in which $P(h_1 \wedge h_2|e)$ is larger from those in which it is smaller than $P(h_1|e)$. In the former case, $Z(h_1 \wedge h_2, e) \geq 0$: either $Z(h_1, e) < 0$, so $Z(h_1 \wedge h_2, e) > Z(h_1, e)$ trivially, or $Z(h_1, e) \geq 0$, and then we find

$$Z(h_1 \wedge h_2, e) - Z(h_1, e) = \frac{N_{Z_+}}{P(e)P(\neg(h_1 \wedge h_2))P(\neg h_1)} \,,$$

where N_{Z_+} has the form

$$[x-(x+z_0)(x+y_0)][1-x-y_1-z_0]-[x-(x+y_1+z_0)(x+y_0)][1-x-z_0] = y_0 y_1 \,,$$

which is nonnegative. The alternative is $Z(h_1 \wedge h_2, e) < 0$, and then

$$Z(h_1 \wedge h_2, e) = \frac{D(h_1 \wedge h_2, e)}{P(h_1 \wedge h_2)} \,,$$

so $Z(h_1 \wedge h_2, e) < 0$ implies $D(h_1 \wedge h_2, e) < 0$. Since $z_1 = z_2 = 0$, we know that $D(h_1, e) \leq D(h_1 \wedge h_2, e)$, so $D(h_1, e) < 0$. Thus $Z(h_1, e) < 0$ and we can calculate the difference

$$Z(h_1 \wedge h_2, e) - Z(h_1, e) = \frac{N_{Z_-}}{P(e)P(h_1 \wedge h_2)P(h_1)} \,,$$

where N_{Z_-} is given by

$$[x - (x + z_0)(x + y_0)][x + y_1 + z_0] - [x - (x + y_1 + z_0)(x + y_0)][x + z_0] = xy_1 \,,$$

which is also nonnegative. This concludes the proof for the measure Z.

As to the two measures involving logarithms,

$$R(h_1 \wedge h_2|e) - R(h_1|e) = \log\left[\frac{x + y_1 + z_0}{x + z_0}\right] = \log\left[1 + \frac{y_1}{x + z_0}\right] \geq 0 \,.$$

L is a little more complicated, and we find

$$\begin{aligned} L(h_1 \wedge h_2|e) - L(h_1|e) &= \log\left[\frac{(1-x-z_0)(x+y_1+z_0)}{(1-x-y_1-z_0)(x+z_0)}\right] \\ &= \log\left[\left(1+\frac{y_1}{1-x-y_1-z_0}\right)\left(1+\frac{y_1}{x+z_0}\right)\right] \geq 0\,. \end{aligned}$$

Nozick's measure yields

$$N(h_1 \wedge h_2|e) - N(h_1|e) = \frac{N_N}{P(h_1)P(\neg h_1)P(h_1 \wedge h_2)P(\neg(h_1 \wedge h_2))}\,,$$

where

$$\begin{aligned} N_N &= [x-(x+y_0)(x+z_0))](x+y_1+z_0)(1-x-y_1-z_0) \\ &\quad -[x-(x+y_0)(x+y_1+z_0)](x+z_0)(1-x-z_0) \\ &= xy_1(1-x-z_0)(1-x-y_1-z_0) + y_0y_1(x+z_0)(x+y_1+z_0) \geq 0\,. \end{aligned}$$

Finally, Fitelson's form has a complicated denominator:

$$F(h_1 \wedge h_2|e) - F(h_1|e) = \frac{N_F}{[P(h_1 \wedge h_2 \wedge e)P(\neg e) + P(h_1 \wedge h_2 \wedge \neg e)P(e)][P(h_1 \wedge e)P(\neg e) + P(h_1 \wedge \neg e)P(e)]}\,,$$

but the numerator simplifies greatly, giving

$$\begin{aligned} N_F &= [x-(x+z_0)(x+y_0)][x+(y_1+z_0-x)(x+y_0)] \\ &\quad -[x-(x+y_1+z_0)(x+y_0)][x+(z_0-x)(x+y_0)] \\ &= 2xy_1(x+y_0)(1-x-y_0) \geq 0\,. \end{aligned}$$

This concludes the proof that Eq.(4) is a robust sufficient condition for the validity of the inequalities (5). However, by scrutinizing the forms that we have obtained for each of the expressions for $c(h_1 \wedge h_2, e) - c(h_1, e)$, we observe that, if we *add* to Eq.(4) the requirement that none of the remaining five triple probabilities, x, y_0, y_1, y_2, z_0 is zero, which is always a valid option, then the inequalities in (5) can be replaced by strict inequalities, i.e.

$$c(h_1 \wedge h_2, e) > c(h_1, e) \quad \& \quad c(h_1 \wedge h_2, e) > c(h_2, e)\,.$$

It is of course interesting that the equality option can be excluded so easily, and moreover robustly, i.e. in a manner that works for all the measures considered.

Appendix B: Disconfirmed hypotheses

In this appendix we will describe sufficient conditions for the validity of the three strict inequalities $c(h_1, e) < 0$, $c(h_2, e) < 0$ and $c(h_1 \wedge h_2, e) > 0$. To do this we will set $x = y_0 = y_1 = y_2 = z_0$ and $z \equiv z_1 = z_2$. Thus five of the marked areas in the Venn diagram are equal to x, while the remaining two are equal to z. It will be required that x is nonzero, while z may be zero or nonzero, so in this appendix the condition (4) is being relaxed: it will turn out that z should be small, but need not be zero. The reason for reducing the seven-dimensional problem to a two-dimensional one is purely one of convenience. With more trouble, for instance, one could allow z_1 and z_2 to be different. However the purpose here is only to show that the inequalities (6) are possible, not to explore every part of the 7-dimensional hypervolume for which they hold.

Consider first the Carnap confirmation function $C(h, e)$. We read off from the Venn diagram that

$$\begin{aligned} C(h_1, e) &= x + z_2 - (x + y_1 + z_0 + z_2)(x + y_0 + z_1 + z_2) \\ &= x + z - (3x + z)(2x + 2z) = (x + z)(1 - 6x - 2z) \\ C(h_1 \wedge h_2, e) &= x - (x + z_0)(x + y_0 + z_1 + z_2) \\ &= x - 2x(2x + 2z) = x(1 - 4x - 4z)\,. \end{aligned}$$

If $1-6x-2z < 0$ and $1-4x-4z > 0$, then $C(h_1, e) < 0$ and $C(h_1 \wedge h_2, e) > 0$. The first inequality yields $6x > 1 - 2z$, while the second gives $4x < 1 - 4z$, which are consistent with each other if $(1 - 4z)/4 > (1 - 2z)/6$, and that is only possible if $z < \frac{1}{8}$. When this holds,

$$\tfrac{1}{6} - \tfrac{z}{3} < x < \tfrac{1}{4} - z\,. \tag{11}$$

In addition, there is the constraint $x + y_0 + y_1 + y_2 + z_0 + z_1 + z_2 \leq 1$, which means that $5x + 2z \leq 1$. Note that $z = 0$ is possible, for then the inequalities simply reduce to $\frac{1}{6} < x < \frac{1}{5}$. We see that $C(h_1, e) < 0$ and $C(h_1 \wedge h_2, e) > 0$ are simultaneously possible, and because of the symmetry between h_1 and h_2, also $C(h_2, e) < 0$ (indeed, with the symmetries that we have imposed, $C(h_2, e) = C(h_1, e)$). As we have seen, this can occur under the strict condition (4), but also when this condition is relaxed.

We will now show that these inequalities guarantee $c(h_1, e) < 0$ and $c(h_1 \wedge h_2, e) > 0$ also when c is realized by the other measures of confirmation.

This is obvious for D, S, N and F because

$$D(h,e) = \frac{C(h,e)}{P(e)} \qquad S(h,e) = \frac{C(h,e)}{P(e)P(\neg e)} \qquad N(h,e) = \frac{C(h,e)}{P(h)P(\neg h)}$$

$$F(h,e) \;=\; \frac{C(h,e)}{P(h \wedge e)P(\neg e) + P(h \wedge \neg e)P(\neg e)}\,.$$

Whenever C is positive (or negative), D, S, N and F are likewise positive (or negative), so the same sufficient conditions are applicable. The same applies to the measure Z, but we have to distinguish between two cases:

$$\begin{aligned} Z(h,e) &= \frac{C(h,e)}{P(e)P(\neg h)} \quad \text{if} \quad P(h|e) \geq P(h) \\ &= \frac{C(h,e)}{P(e)P(h)} \quad \text{if} \quad P(h|e) < P(h)\,. \end{aligned}$$

The first form must be used for $h \equiv h_1 \wedge h_2$, when C and Z are both positive. The second form is needed for $h \equiv h_1$, for then C and Z are negative. We have used the fact that C, D, S, N and F are all equivalent up to normalization, which indeed inspired Crupi, Tentori and Gonzalez (2007) to produce their Z-measure, that has the property that it is equal to 1 if e implies h, and -1 if e implies $\neg h$.

For the logarithmic measures R and L we find

$$\begin{aligned} R(h_1,e) &= \log \frac{P(h_1 \wedge e)}{P(e)P(h_1)} \\ &= \log \frac{x+z_2}{(x+y_0+z_1+z_2)(x+y_1+z_0+z_2)} \\ &= \log \frac{x+z}{2(x+z)(3x+z)} = -\log(6x+2z)\,. \\ R(h_1 \wedge h_2,e) &= \log \frac{P(h_1 \wedge h_2 \wedge e)}{P(e)P(h_1 \wedge h_2)} \\ &= \log \frac{x}{(x+y_0+z_1+z_2)(x+z_0)} \\ &= \log \frac{x}{2(x+z)(2x)} = -\log(4x+4z)\,. \end{aligned}$$

We require $6x + 2z > 1$ and $4x + 4z < 1$, and this yields Eq.(11) again. Finally,

$$\begin{aligned}
L(h_1, e) &= \log \frac{P(h_1 \wedge e)[1 - P(h_1)]}{[P(e) - P(h_1 \wedge e)]P(h_1)} \\
&= \log \frac{(x + z_2)(1 - x - y_1 - z_0 - z_2)}{(y_0 + z_1)(x + y_1 + z_0 + z_2)} \\
&= \log \frac{(x + z)(1 - 3x - z)}{(x + z)(3x + z)} = \log \frac{1 - 3x - z}{3x + z} . \\
L(h_1 \wedge h_2, e) &= \log \frac{P(h_1 \wedge h_2 \wedge e)[1 - P(h_1 \wedge h_2)]}{[P(e) - P(h_1 \wedge h_2 \wedge e)]P(h_1 \wedge h_2)} \\
&= \log \frac{x(1 - x - z_0)}{(y_0 + z_1 + z_2)(x + z_0)} \\
&= \log \frac{x(1 - 2x)}{(x + 2z)(2x)} = \log \frac{1 - 2x}{2(x + 2z)} .
\end{aligned}$$

Now we require $1 - 3x - z < 3x + z$ and $1 - 2x > 2x + 4z$, and this leads to Eq.(11) once more.

In conclusion, we have proved the following robust result, which we call the Jack Author Theorem:

A sufficient condition for $c(h_1, e) < 0$, $c(h_2, e) < 0$ *and* $c(h_1 \wedge h_2, e) > 0$ *is*

$$\tfrac{1-2z}{6} < x < \min \left[\tfrac{1-2z}{5} \, , \, \tfrac{1-4z}{4} \right] \qquad \& \qquad 0 \leq z \leq \tfrac{1}{8} \, , \tag{12}$$

where c stands for C, D, S, Z, R, L, N, F, and where

$$\begin{aligned}
x &= P(h_1 \wedge h_2 \wedge e) = P(\neg h_1 \wedge \neg h_2 \wedge e) \\
&= P((h_1 \wedge \neg h_2 \wedge \neg e) = P(\neg h_1 \wedge h_2 \wedge \neg e) = P(h_1 \wedge h_2 \wedge \neg e) \\
z &= P(\neg h_1 \wedge h_2 \wedge e) = P(h_1 \wedge \neg h_2 \wedge e) \, .
\end{aligned}$$

The upper bound, $z \leq \frac{1}{8}$, is in fact already implied by the first part of (12).

Appendix C: Bovens and Hartmann coherence

In Sect. 2 it was noted that there is a close conceptual link between confirmation and coherence: measures of confirmation can be put to work as measures of coherence and vice versa. The eight measures of confirmation that we have considered so far were all quantitative, i.e. the confirmation that e gives to h is expressed as a number between -1 and $+1$.

Bovens and Hartmann (2003a & 2003b) have introduced a coherence measure that is comparative rather than quantitative. Given two pairs of propositions, for example $\{h, e\}$ and $\{h', e'\}$, Bovens and Hartmann describe a condition such that, when it is fulfilled, it tells us which of the two pairs is the more coherent. An ordering of pairs is thus introduced, but it is what Bovens and Hartmann call a quasi-ordering. For it can happen that the condition is *not* fulfilled, and then the relative coherence of two different pairs is simply not defined.

In more detail, the quasi-ordering can be explained as follows. Imagine a witness who reports on a hypothesis and some evidence about which he has heard. Let h be his report concerning the hypothesis, and e be his formulation of the evidence. Define

$$\begin{aligned} a_0 &= P(h \wedge e) \\ a_1 &= P(\neg h \wedge e) + P(h \wedge \neg e) \\ a_2 &= P(\neg h \wedge \neg e) = 1 - a_0 - a_1 \,. \end{aligned}$$

Consider the function

$$B(h, e; x) = \frac{a_0 + (1 - a_0)x^2}{a_0 + a_1 x + a_2 x^2}\,, \qquad (13)$$

where x is a number in the interval $[\,0\,,1]$ that represents the *unreliability* of the witness, with 1 corresponding to total unreliability and 0 to total reliability. According to Bovens and Hartmann, the pair $\{h', e'\}$ is not less coherent than the pair $\{h, e\}$ if

$$B(h', e'; x) \geq B(h, e; x)\,, \qquad \forall x \in [\,0\,,1]\,. \qquad (14)$$

The salient point is that this inequality must hold for all x, i.e. for all possible degrees of unreliability of the witness, x being the same for both pairs $\{h', e'\}$ and $\{h, e\}$. If (14) holds with $\geq$ replaced by $\leq$, then $\{h', e'\}$ is said to be not more coherent than the pair $\{h, e\}$. However, it can happen that (14) holds neither with $\geq$ nor with $\leq$, and then the relative coherence of the two pairs is undefined: they are not ordered in respect of their Bovens-Hartmann coherence.

Bovens and Hartmann illustrate cases in which the inequality (14) holds by means of a graph that shows the left- and the right-hand sides as curves that do not intersect one another. Here we introduce however a simple algebraic alternative. Since $a_2 = 1 - a_0 - a_1$, it follows from Eq.(13) that

$$x(1-x)\frac{B(h,e;x)}{1-B(h,e;x)} = \frac{a_0}{a_1}(1-x^2) + \frac{x^2}{a_1}\,.$$

Further, since the left-hand side of this equation is a monotonic increasing function of $B(h, e; x) \in [\,0\,, 1]$, for fixed x, it follows that (14) is equivalent to

$$\frac{a_0'}{a_1'}(1-x^2) + \frac{x^2}{a_1'} \geq \frac{a_0}{a_1}(1-x^2) + \frac{x^2}{a_1}\,, \qquad \forall x \in [\,0\,,1]\,. \tag{15}$$

With the notation $X = x^2$, this can be rewritten

$$\frac{a_0'}{a_1'}(1-X) + \frac{X}{a_1'} \geq \frac{a_0}{a_1}(1-X) + \frac{X}{a_1}\,,\,. \tag{16}$$

The above inequality holds at $X = 0$ if $\frac{a_0'}{a_1'} \geq \frac{a_0}{a_1}$, and it holds at $X = 1$ if $\frac{1}{a_1'} \geq \frac{1}{a_1}$. This is a necessary and sufficient condition that the two straight lines in (16), to the left and right of $\geq$, do not cross, and therefore that (15) is true. These requirements are equivalent to

$$a_1' \leq a_1 \qquad \& \qquad \frac{a_1'}{a_0'} \leq \frac{a_1}{a_0}\,. \tag{17}$$

We now set $h = h_1$, $h' = h_1 \wedge h_2$, $e' = e$; and we use Bovens-Hartmann coherence as a (relative) measure of confirmation. Conditions (17) read

$$\begin{aligned} P\Big(\neg(h_1 \wedge h_2) \wedge e\Big) + P(h_1 \wedge h_2 \wedge \neg e) \;&\leq\; P(\neg h_1 \wedge e) + P(h_1 \wedge \neg e) \\ \frac{P\Big(\neg(h_1 \wedge h_2) \wedge e\Big) + P(h_1 \wedge h_2 \wedge \neg e)}{P(h_1 \wedge h_2 \wedge e)} \;&\leq\; \frac{P(\neg h_1 \wedge e) + P(h_1 \wedge \neg e)}{P(h_1 \wedge e)}\,. \end{aligned}$$

On referring to the Venn diagram of Appendix A, we transcribe these conditions as follows:

$$\begin{aligned} y_0 + z_0 + z_1 + z_2 &\leq y_0 + y_1 + z_0 + z_1 \\ \frac{y_0 + z_0 + z_1 + z_2}{x} &\leq \frac{y_0 + y_1 + z_0 + z_1}{x + z_2}\,, \end{aligned}$$

which reduce respectively to

$$z_2 \leq y_1 \qquad \& \qquad z_2(x + y_0 + z_0 + z_1 + z_2) \leq x y_1\,. \tag{18}$$

Evidently both of these equalities are satisfied automatically if $z_2 = 0$. Under this condition we conclude that $h_1 \wedge h_2$ is not less highly Bovens-Hartmann confirmed by e than is h_1 alone. A similar argument, with h_1 and h_2 interchanged, shows that, if $z_1 = 0$, also $h_1 \wedge h_2$ is not less highly confirmed by e than is h_2 alone.

Since $z_1 = 0 = z_2$ is equivalent to the conditions (4) of Sect. 2, we have shown thereby that the robustness of these conditions extends also to the measure of Bovens and Hartmann. Moreover, if neither x nor y_1 nor y_2 is zero, $\leq$ may be replaced by $<$, and under these conditions $h_1 \wedge h_2$ is strictly more confirmed by e than are h_1 or h_2.

The work of this appendix may be seen as an extension of Appendix A to Bovens-Hartmann confirmation. No analogous extension in the spirit of Appendix B is possible, for it does not make sense to say that $h_1 \wedge h_2$ is confirmed, nor that h_1 or h_2 are disconfirmed in the sense of Bovens and Hartmann.

References

Bovens, L. and S. Hartmann (2003a). Solving the riddle of coherence. *Mind,* 112, pp. 601-633.

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

Bovens, L. and S. Hartmann (2005). Coherence and the Role of Specificity: A Response to Meijs and Douven. *Mind,* 114, pp. 365-369.

Carnap, R. (1950). *Logical foundations of probability.* Chicago: University of Chicago Press.

Carnap, R. (1962). Preface to the second edition of *Logical foundations of probability.* Chicago: University of Chicago Press.

Christensen, D. (1999). Measuring confirmation. *Journal of Philosophy* 96, pp. 437-461.

Crupi, V., Fitelson, B., and Tentori, K. (2007). Probability, confirmation, and the conjunction fallacy. *Thinking and Reasoning,* to appear.

Crupi, V., Tentori, K., and Gonzalez, M. (2007). On Bayesian measures of evidential support: theoretical and empirical issues. *Philosophy of Science,* 74 pp. 229-252.

Douven, I. and W. Meijs (2007). Bootstrap confirmation made quantitative. *Synthese,* 149 pp. 97-132.

Douven, I. and W. Meijs (2007). Measuring coherence. *Synthese,* 156 pp. 405-425.

Eells, E. (1982). *Rational decision and causality.* Cambridge, UK: Cambridge University Press.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science,* 66, pp. S363-S378.
Fitelson, B. (2001). A Bayesian account of independent evidence with applications. *Philosophy of Science,* 68, pp. 123-140.

Fitelson, B. (2003). A probabilistic theory of coherence. *Analysis,* 63, pp. 194-199.

Good, I.J. (1950). *Probability and the weighing of evidence.* London: Griffin.
Joyce, J. (1999). *The foundations of causal decision theory.* Cambridge, UK: Cambridge University Press.

Keynes, J. (1921). *A treatise on probability.* London: Macmillan.

Meijs, W. and Douven, I. (2005). Bovens and Hartmann on coherence. *Mind,* 114, pp. 355-363.

Milne, P. (1996). Log$[p(h/eb)/p(h/b)]$ is the one true measure of confirmation. *Philosophy of Science* 63, pp. 21-26.

Nozick, R. (1981). *Philosophical explanations.* Oxford, UK: Clarendon.

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

Sides, A., Osherson, D., Bonini, N., and Viale, R. (2002). On the reality of the conjunction fallacy. *Memory and Cognition,* 30, pp. 191-198.

Tentori, K., Crupi, V., Bonini, N., and Osherson, D. (2007). Comparison of confirmation measures. *Cognition,* 103, pp. 107-119.