Synthese (2008) 161:167–182
DOI 10.1007/s11229-006-9155-4

ORIGINAL ARTICLE

Equivalent testimonies as a touchstone of coherence measures

Mark Siebel · Werner Wolff

Received: 30 August 2005 / Accepted: 18 December 2006 / Published online: 10 February 2007

Abstract Over the past years, a number of probabilistic measures of coherence have been proposed. As shown in the paper, however, many of them do not conform to the intuitition that equivalent testimonies are highly coherent, regardless of their prior probability.

Keywords Probabilistic measures of coherence · Equivalent testimonies

1 A test case

A bankrobbery has occurred. Detective Jones asks the witnesses Susan and Tom about the robbers' height and colour of hair. Susan says: 'Both of them were at least six feet tall. One had blonde and the other one red hair.' Tom answers: 'None of them was smaller than six feet. Both had dark brown hair.' Back in the police department, Jones tells his colleagues that, although Susan's and Tom's testimonies on the robbers' colour of hair do not fit together, they cohere perfectly with respect to their height.

Jones's judgement is quite reasonable. After all, what Susan said about the robbers' height is logically equivalent to what Tom claimed: if the proposition put forward by Susan is true, the proposition put forward by Tom must be true as well, and conversely. Jones is therefore right in reporting that, in this respect, their testimonies are coherent to an outstanding degree.

M. Siebel (✉)
Philosophisches Seminar, Universität Hamburg,
Von-Melle-Park 6, 20146 Hamburg, Germany
e-mail: mark.siebel@uni-hamburg.de

W. Wolff
Institut für Logik und Wissenschaftstheorie, Universität Leipzig
Beethovenstr. 15, 04107 Leipzig, Germany
e-mail: wwolff@uni-leipzig.de

The same holds for equivalent results of measurements. Consider a physicist who needs to know the air pressure in her laboratory. She makes two measurements, one with a barometer measuring in bars and one with a barometer displaying hectoPascals. The readings are '0.985 bars' and '985 hectoPascals'. Since bars are converted into hectoPascals by multiplying them by 1,000, these results are equivalent and thus cohere perfectly.

We take it that intuitions are very strong here, so that examples of this kind may be used as a touchstone for theories of coherence.[1] The theories we want to expose to this test are the *probabilistic* measures of coherence which have been proposed over the last years (plus some variants of one of them).[2] These measures consist in functions taking as input certain probabilities relating to the propositions in question to calculate from them a number which is supposed to represent their degree of coherence. We will show which probabilistic measures conform to the intuition that equivalent testimonies (results of measurements ...) are highly coherent and which do not, thereby adding, as we hope, a substantial piece to the ongoing discussion on the pros and cons of these measures.[3] (For the sake of brevity, we shall often refrain from providing a motivation for the proposed functions.)

To avoid misunderstandings, let us mention two points before we start. First, there is an act/content ambiguity in words like 'testimony' (or 'report') which may easily be overlooked. When Susan says that the robbers were at least six feet tall, then both the particular speech act she performs—an event in space and time—and the content of this act—the proposition she puts forward—may be called Susan's testimony. Our claim that equivalent testimonies are coherent is bound to the latter sense. That is, although we take the *contents* of testimonial acts to fit together if they are equivalent, we do not state that the same holds for the corresponding *acts*. Assume, for example, that Susan displays a psychotic aversion to Tom which is so strong that everything he says provokes vehement protest from her. Assume furthermore that she knows what Tom has reported about the robbers' height. Against this background, her testifying that they were at least six feet tall might not dovetail with Tom's testifying that none of them was smaller than six feet. But this does not alter the fact that the *propositions* they claim to be true are coherent.

Second, that equivalent testimonies fit together well does not mean that one is always justified in believing witnesses who put forward equivalent claims. The good fit may come about by an arrangement of the witnesses which was carried out in order to shield the real culprit. Or suppose one of the subjects was not at the scene of the crime actually but just parrots what the other witness has said. Roughly, if the reports fail to be independent, their coherence is of little consequence.[4] But we are not concerned here with the question of what to *conclude* from coherent testimonies. The assumption that equivalent witness reports are highly coherent does not entail

[1] Cf. Fitelson (2003, p. 194); Bovens and Hartmann (2003b, p. 32). An equally strong test is provided by two *contradictory* witness reports, which are intuitively highly *in*coherent. However, this case does not enable one to sort out further measures from the ones to be examined because none of them passes the equivalency test without also scoring well with respect to contradictions.

[2] For a different type of mathematical approach to coherence, cf. Thagard (1992, Ch. 4); Thagard and Verbeurgt (1998); Siebel (2005b).

[3] For further cons, see Bovens and Hartmann (2003b); Douven and Meijs (2007); Fitelson (2003); Moretti and Akiba (2007); Siebel (2005a).

[4] More on the independence of witnesses in Bovens and Olsson (2002, p. 143f.); Bovens and Hartmann (2003b, 15f.). Cf. also Lewis (1946, p. 352).

that one should, come what may, put one's trust in them. There is, at most, a default rule stating that one is warranted in believing them unless there is reason for taking them to be dependent.

Something analogous holds for situations where the equivalent testimonies do not fit into the wider picture of what is known already about the crime. Then it might be unwise to trust the witnesses because their claims would lower *global* coherence. Nonetheless, the propositions put forward are still *locally* coherent: although they are not in harmony with the given background, they dovetail with each other.[5]

Closely connected with these topics is the question of whether coherence is *truth-conducive*, i.e. whether a more coherent system has (at least in fortunate circumstances and in a ceteris paribus sense) a higher probability of being true.[6] This is surely an important question, but it is hard to see how it could be answered without first spelling out in more detail what is meant by 'coherence' in this context. After all, a standard objection to coherence theories is that they offer only vague characterisations of their central notion instead of assigning a definitive meaning to the term 'coherence'. The probabilistic accounts discussed in this paper try to meet this challenge by giving a formal explication of the concept of coherence; and we are interested in whether they succeed. Like Carnap (1950, pp. 5–7), we assume that an explication should respect fundamental intuitions associated with the concept in question; and it seems to us that what we said about equivalent testimonies is among these intuitions when it comes to coherence.

2 Shogenji's measure

According to Shogenji (1999, p. 240), the coherence of a set of propositions $\{A_1, \ldots, A_n\}$ is to be determined as follows:

$$C_{Sh}(A_1, \ldots, A_n) = \frac{P(A_1 \,\&\cdots\&\, A_n)}{P(A_1) \times \cdots \times P(A_n)}$$

$P(A_1) \times \cdots \times P(A_n)$ is the probability the conjunction $A_1 \,\&\cdots\&\, A_n$ would have if its constituents were statistically independent. Shogenji's formula thus tells us to which extent the actual probability that the propositions are true together deviates from the probability they would have if they were statistically independent.[7] If the C_{Sh}-value is 1, the set is taken to be neither coherent nor incoherent because the statements are independent. If it is greater than 1, the set is coherent; if it is smaller than 1, the statements do not fit together.

As to equivalent testimonies, Shogenji's measure has a startling consequence. If A is logically equivalent to B, then $P(A \,\&\, B) = P(B)$. Hence:

$$C_{Sh}(A,\ B) = \frac{P(A \,\&\, B)}{P(A) \times P(B)} = \frac{1}{P(A)}$$

5 On local versus global coherence, cf. Bartelborth (1996, Ch. IV.C).

6 For arguments against the truth-conducivity of coherence, see Bovens and Hartmann (2003b, Ch. 1.4); Klein and Warfield (1994); Olsson (2005, Ch. 7).

7 $P(A \,\&\, B)/[P(A) \times P(B)] = P(A|B)/P(A) = P(B|A)/P(B)$. According to one of the accounts to be introduced in Sect. 3, the latter formulas represent B's degree of support for A and A's degree of support for B. Hence, in two-member cases, Shogenji's proposal may be interpreted as identifying coherence with the support the propositions provide for each other.

Given that $P(\mathrm{A})$ is smaller than 1, our witness reports are of course *coherent* because $C_{\mathrm{Sh}}(\mathrm{A, B})$ will be above the threshold 1. However, how *much* they cohere would depend on how likely the propositions put forward are: the more probable it is that they are true, the less they dovetail. Applied to our test case, if the probability of the bankrobbers' being at least six feet tall is 0.1, then Susan's and Tom's testimonies show a coherence of 10. But if it was already likely to degree 0.9 that the bankrobbers are that tall, then their reports would be close to being neither coherent nor incoherent because $C_{\mathrm{Sh}} \approx 1$. In other words, their testimonies would be only slightly (or minimally) coherent.

Expressions such as 'close', 'slight' and 'minimal' are always a bit tricky because they are vague. To make the difficulty with Shogenji's formula more precise, let us say that a pair of propositions is only slightly coherent on a measure if the difference between the pair's coherence and the measure's neutral point does not go beyond a positive number ε which may be as small as you like. In the case of Shogenji's measure, whose neutral point is 1, this means that $C_{\mathrm{Sh}}(\mathrm{A,B}) \leq 1 + \varepsilon$. Since for our set $C_{\mathrm{Sh}}(\mathrm{A,B}) = 1/P(\mathrm{A})$, it follows that this set is only minimally coherent if and only if $P(\mathrm{A}) \geq 1/(1 + \varepsilon)$.

The problem then is that it does not matter how close you want the coherence to come to 1 in order to call the testimonies only slightly coherent. For any small difference ε you choose, there is a probability for A, namely $1/(1 + \varepsilon)$, leading to the result that $C_{\mathrm{Sh}}(\mathrm{A, B})$ is close enough to 1. For example, take ε to be 0.01. Then Susan's and Tom's reports fit together only to a minimum degree if the probability that the robbers' are at least six feet tall is greater than or equal to approximately 0.99. Shogenji's formula thus does not respect the intuition that equivalent testimonies always cohere perfectly.

A possible objection should be considered. Bovens and Olsson comment on a different example:[8]

> One could argue that *coherence* is an ambiguous notion. One can think of coherence as a measure of *agreement* or as a measure of *striking agreement*. Coherence is sensitive to the specificity [and therefore to the probability] of the information on the latter but not on the former notion. (Bovens and Olsson, 2000, p. 688f., fn. 1)

The idea here is that there is a concept of coherence according to which propositions cohere more the less probable they are, ceteris paribus. If this was true, then our test case would not be as clear-cut as we considered it to be. In *this* sense of 'coherent', equivalent witness reports would not necessarily be highly coherent because their coherence would depend on how likely it is that the claims put forward are true. Shogenji's formula could then be viewed as measuring coherence in this sense.

However, we do not see why 'coherence' (as it is used by detectives, scientists ...) should be ambiguous in the way described by Bovens and Olsson. Just imagine Detective Jones saying to his colleagues: 'If we have had no prior evidence for the robbers' being at least six feet tall, I would have said that Susan's and Tom's testimonies fit together very well. But it was already quite probable that the robbers are that tall because another credible witness assured us of this before we asked Susan and Tom. I must therefore admit that the latters' reports cohere only slightly'. If there was Bovens and Olsson's sense of 'cohere', this should be a reasonable remark. But, obviously, it will cause general disapproval among Jones's colleagues.

[8] Cf. also Olsson (2002, p. 250f.); and Olsson & Shogenji (2004, p. 28).

Perhaps it makes sense to say that the agreement between Susan's and Tom's reports is not really striking if there is already strong evidence for the robbers' being at least six feet tall. For, given that Susan and Tom are reliable, it is to be expected that both of them provide true information about the robbers' height. And if it is already quite probable that the robbers are at least six feet tall, it is thus to be expected that Susan's and Tom's descriptions agree in this respect. But this does not mean that what they say is less *coherent*. It only means that it is less *striking* that their reports cohere. Briefly, equivalent testimonies *are* highly coherent, no matter whether such a strong coherence was to be *expected* or not.

3 Douven and Meijs's measure

Douven and Meijs (2007, Sects. 2f.) do not only advocate a specific account but also present a general recipe for generating probabilistic measures of coherence. Their starting point is simple and appealing: a system's degree of coherence depends on the degree of confirmation (alias support) its elements provide for each other. This idea is implemented as follows.

Choose a probabilistic measure of support S. Calculate the extent to which each proposition and each conjunction of propositions in the set is supported by each remaining proposition and each conjunction of them. Then take the straight average of the results. For example, if we are confronted with a pair {A, B}, its coherence is

$$C(\mathrm{A},\mathrm{B}) = [S(\mathrm{A},\mathrm{B}) + S(\mathrm{B},\mathrm{A})]/2,$$

that is, the average of the degree to which B confirms A and the degree to which A confirms B. For a trio we get a more complicated formula:

$$\begin{aligned} C(\mathrm{A},\mathrm{B},\mathrm{C}) = [&S(\mathrm{A},\mathrm{B}) + S(\mathrm{A},\mathrm{C}) + S(\mathrm{B},\mathrm{A}) + S(\mathrm{B},\mathrm{C}) + S(\mathrm{C},\mathrm{A}) + S(\mathrm{C},\mathrm{B}) \\ &+ S(\mathrm{A},\mathrm{B}\ \&\ \mathrm{C}) + S(\mathrm{B},\mathrm{A}\ \&\ \mathrm{C}) + S(\mathrm{C},\mathrm{A}\ \&\ \mathrm{B}) + S(\mathrm{A}\ \&\ \mathrm{B},\mathrm{C}) \\ &+ S(\mathrm{A}\ \&\ \mathrm{C},\mathrm{B}) + S(\mathrm{B}\ \&\ \mathrm{C},\mathrm{A})]/12 \end{aligned}$$

More formally, let A' and A'' be non-empty and disjoint subsets of $\{\mathrm{A}_1,\ldots,\mathrm{A}_n\}$, and let CA' stand for the conjunction of the elements in A' while CA'' is the conjunction of the members of A'' (in the case of singletons, we allow the single element to be the corresponding 'conjunction'). Furthermore, take an ordering $\langle \mathrm{PCA}_1,\ldots,\mathrm{PCA}_m\rangle$ including all of the ordered pairs $\langle \mathrm{CA}',\mathrm{CA}''\rangle$, where m is the number of elements in this ordering. Then the degree of coherence of $\{\mathrm{A}_1,\ldots,\mathrm{A}_n\}$ relative to a support measure S is given by:

$$C_{\mathrm{S}}(\mathrm{A}_1,\ldots,\mathrm{A}_n) = \left[\sum_{i=1}^{m} S(\mathrm{PCA}_i)\right] \Big/ m$$

Intuitively, coherence is thus identified with average confirmation.

To convert this general schema into a measure of coherence, a specific measure of confirmation is required. Friends of probabilistic (or Bayesian) epistemology agree on how to answer the *qualitative* question under which conditions B provides any confirmation for A at all. Evidence supports a hypothesis, so they claim, if it raises its probability. This is laid down in the so-called relevance criterion:[9]

> B confirms A iff $P(\mathrm{A}|\mathrm{B}) > P(\mathrm{A})$; B disconfirms A iff $P(\mathrm{A}|\mathrm{B}) < P(\mathrm{A})$; and B is confirmationally irrelevant to A iff $P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A})$.

There is, however, substantial disagreement on the *quantitative* question of how to measure the degree to which evidence supports a hypothesis. Many people rely on the difference between $P(\mathrm{A}|\mathrm{B})$ and $P(\mathrm{A})$, others prefer the ratio of these probabilities or a logarithm thereof, and there is a significant number of further proposals on the agenda.[10]

Note that it really matters from which support measure one starts because different choices may entail significantly different outcomes. For example, consider the difference measure S_{d}—the favourite of Gillies (1986), Jeffrey (1992, p. 72) and Rosenkrantz (1994)—, the ratio measure S_{r}—which is advocated by Horwich (1998) and Schlesinger (1995)—and the log-likelihood account—for which Good (1984) and Fitelson (2001, Ch. 3.2) plead. According to these proposals, B's support for A is to be calculated as follows:

$$S_{\mathrm{d}}(\mathrm{A},\mathrm{B}) = P(\mathrm{A}|\mathrm{B}) - P(\mathrm{A})$$
$$S_{\mathrm{r}}(\mathrm{A},\mathrm{B}) = \frac{P(\mathrm{A}|\mathrm{B})}{P(\mathrm{A})}$$
$$S_{\mathrm{ll}}(\mathrm{A},\mathrm{B}) = \log \frac{P(\mathrm{B}|\mathrm{A})}{P(\mathrm{B}|\neg\mathrm{A})}$$

Now let there be 100 suspects for a murder, where it is sure that one of them committed it. The assumptions at issue are:

(A) The murderer is male.
(B) The murderer has blonde hair.

In case 1, 63 of the subjects are both male and blonde, 7 are males and not blonde, 7 are females with blonde hair, and the remaining 23 suspects are neither male nor blonde. In case 2, there are 8 male and blonde persons, 12 males who are not blonde, 12 females with blonde hair and 68 suspects who are neither male nor blonde. That is, in situation 1, $P(\mathrm{A}|\mathrm{B}) = P(\mathrm{B}|\mathrm{A}) = 0.9$, $P(\mathrm{A}) = P(\mathrm{B}) = 0.7$, $P(\mathrm{A}|\neg\mathrm{B}) = P(\mathrm{B}|\neg\mathrm{A}) \approx 0.23$. And in situation 2, $P(\mathrm{A}|\mathrm{B}) = P(\mathrm{B}|\mathrm{A}) = 0.4$, $P(\mathrm{A}) = P(\mathrm{B}) = 0.2$, $P(\mathrm{A}|\neg\mathrm{B}) = P(\mathrm{B}|\neg\mathrm{A}) = 0.15$.

Thus, in both cases, the propositions confirm each other on the relevance criterion because $P(\mathrm{A}|\mathrm{B}) > P(\mathrm{A})$ and $P(\mathrm{B}|\mathrm{A}) > P(\mathrm{B})$. When it comes to the *degree* of support, however, the three measures differ considerably. The difference measure gives identical degrees of confirmation:

$$S^1_{\mathrm{d}}(\mathrm{A},\mathrm{B}) = S^1_{\mathrm{d}}(\mathrm{B},\mathrm{A}) = 0.9 - 0.7 = 0.2$$
$$S^2_{\mathrm{d}}(\mathrm{A},\mathrm{B}) = S^2_{\mathrm{d}}(\mathrm{B},\mathrm{A}) = 0.4 - 0.2 = 0.2$$

9 Strictly speaking, the relevance criterion specifies *incremental* confirmation, in contrast to *absolute* confirmation. B is taken to support A in the absolute sense iff $P(\mathrm{A}|\mathrm{B}) > k$, where the threshold k should be 0.5 at least so that $P(\mathrm{A}|\mathrm{B}) > P(\neg\mathrm{A}|\mathrm{B})$.

10 For overviews and the pros and cons of different measures, see Eells and Fitelson (2002); Fitelson (1999, 2001); Kyburg (1983, Sect. IV).

The ratio account states the support is higher in situation 2:

$$S_{\mathrm{r}}^{1}(\mathrm{A},\mathrm{B}) = S_{\mathrm{r}}^{1}(\mathrm{B},\mathrm{A}) = 0.9/0.7 \approx 1.3$$
$$S_{\mathrm{r}}^{2}(\mathrm{A},\mathrm{B}) = S_{\mathrm{r}}^{2}(\mathrm{B},\mathrm{A}) = 0.4/0.2 = 2$$

And the log-likelihood measure tells us that it is greater in case 1:

$$S_{\mathrm{ll}}^{1}(\mathrm{A},\mathrm{B}) = S_{\mathrm{ll}}^{1}(\mathrm{B},\mathrm{A}) \approx \log(0.9/0.23) \approx \log 3.9$$
$$S_{\mathrm{ll}}^{2}(\mathrm{A},\mathrm{B}) = S_{\mathrm{ll}}^{2}(\mathrm{B},\mathrm{A}) \approx \log(0.4/0.15) \approx \log 2.7$$

Hence, by choosing the difference measure, one is committed to the claim that the assumptions in question own the same coherence in these situations:

$$C_{\mathrm{d}}^{1}(\mathrm{A},\mathrm{B}) = [S_{\mathrm{d}}^{1}(\mathrm{A},\mathrm{B}) + S_{\mathrm{d}}^{1}(\mathrm{B},\mathrm{A})]/2 = [S_{\mathrm{d}}^{2}(\mathrm{A},\mathrm{B}) + S_{\mathrm{d}}^{2}(\mathrm{B},\mathrm{A})]/2 = C_{\mathrm{d}}^{2}(\mathrm{A},\mathrm{B})$$

In contrast, the ratio measure rules that the coherence is higher in the second setting, whereas using the log-likelihood measure leads to the opposite result. The given example thus shows that, depending on the involved support measures, the coherence theories which are generated after Douven and Meijs's recipe may provide substantially different outcomes.

Back to Douven and Meijs's specific theory of coherence. They choose the difference measure, thereby taking the coherence of a two-member set to be given by:

$$C_{\mathrm{d}}(\mathrm{A},\mathrm{B}) = [P(\mathrm{A}|\mathrm{B}) - P(\mathrm{A}) + P(\mathrm{B}|\mathrm{A}) - P(\mathrm{B})]/2$$

The values of this function range from −1 to 1. Unlike Shogenji, Douven and Meijs do not explicitly distinguish a threshold above which coherence begins. But their account suggests that they have 0 in mind. For it appears that they consider pairs of propositions which speak neither for nor against each other to be neither coherent nor incoherent. Now, in terms of the relevance criterion, A and B are confirmationally irrelevant to each other just in case $P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A})$, which entails that $P(\mathrm{B}|\mathrm{A}) = P(\mathrm{B})$. But then $P(\mathrm{A}|\mathrm{B}) - P(\mathrm{A}) = P(\mathrm{B}|\mathrm{A}) - P(\mathrm{B}) = 0$; and thus $C_{\mathrm{d}}(\mathrm{A}, \mathrm{B})$ is also 0. It is therefore reasonable to assume that Douven and Meijs interpret a value of 0 as meaning that the pair is neither coherent nor incoherent. Analogously, if there is mutual support between the propositions—viz., $P(\mathrm{A}|\mathrm{B}) > P(\mathrm{A})$ so that also $P(\mathrm{B}|\mathrm{A}) > P(\mathrm{B})$—then $C_{\mathrm{d}}(\mathrm{A},\mathrm{B}) > 1$, suggesting that a value greater than 0 stands for coherence.

To be fair, it should be pointed out that Douven and Meijs (2007, Sect. 5.1) restrict their measure to sets of pairwise logically independent propositions. Although they wish to get rid of a different problem thereby, this constraint also concerns our test case because equivalent statements are logically dependent. However, such a scope restriction in view of counter-examples smacks of an 'easy way out'. We therefore take the liberty of subjecting their proposal in its unqualified form to our 'experimentum crucis'.

Given that A and B are logically equivalent, $P(\mathrm{A}|\mathrm{B}) = P(\mathrm{B}|\mathrm{A}) = 1$ and $P(\mathrm{A}) = P(\mathrm{B})$. Therefore:

$$C_{\mathrm{d}}(\mathrm{A},\mathrm{B}) = [1 - P(\mathrm{A}) + 1 - P(\mathrm{A})]/2 = 1 - P(\mathrm{A})$$

In agreement with Shogenji's function, the account of Douven and Meijs thus rules that equivalent testimonies are the less coherent the more likely the propositions are. What Susan says about the height of the robbers would be far from fitting perfectly what Tom claims when it is already highly probable that the robbers are at least six feet tall.

To be more precise again, let there be a small positive number ε such that {A, B} is only slightly coherent if its coherence differs from Douven and Meijs's neutral point 0 at most by ε, i.e. $C_d(A,B) \leq \varepsilon$. Since $C_d(A, B)$ is in our example equal to $1 - P(A)$, this entails that our set is only slightly coherent if $P(A) \geq 1 - \varepsilon$. But this means that, however small ε is, it is always possible to assign the propositions put forward by Susan and Tom a probability which makes them coherent to a very low degree. For instance, if you want ε to be 0.01, then Susan's and Tom's testimonies are only minimally coherent just in case the probability that the robbers' are at least six feet tall is greater than or equal to 0.99. Like Shogenji's formula, Douven and Meijs's proposal does not pass our test because it permits equivalent witness reports to be close to being neither coherent nor incoherent.

4 Further measures after Douven and Meijs's recipe

C_d is only one of the coherence functions which may be generated on the basis of Douven and Meijs's pattern. Let us see what happens if we insert further probabilistic measures of confirmation into their schema. Besides the already mentioned ratio measure S_r, there is a normalised variant of it, the log-ratio measure S_{lr}—which Milne (1996) takes to be 'the one true measure of confirmation'—, Finch's (1960) proposal S_{Fi}, Popper's (1954) S_{Po}, Levi's (1962) S_{Le}, Rescher's (1958) S_{Re} and Carnap's (1950, Sect. 67) S_{Ca}:

$$S_r(A,B) = \frac{P(A|B)}{P(A)}$$
$$S_{lr}(A,B) = \log\frac{P(A|B)}{P(A)}$$
$$S_{Fi}(A,B) = \frac{P(B|A) - P(B)}{P(B)}$$
$$S_{Po}(A,B) = \frac{P(B|A) - P(B)}{P(B|A) + P(B)} \times [1 + P(A) \times P(A|B)]$$
$$S_{Le}(A,B) = P(A|B) \times P(\neg A) - P(\neg A|B) \times P(A)$$
$$S_{Re}(A,B) = \frac{P(A|B) - P(A)}{1 - P(A)} \times P(B)$$
$$S_{Ca}(A,B) = P(A\&B) - P(A) \times P(B)$$

If we put the ratio measure to use, we get the same erroneous result for our test case as on Shogenji's account of coherence:[11]

$$C_r(A,B) = \left(\frac{P(A|B)}{P(A)} + \frac{P(B|A)}{P(B)}\right) \Big/ 2 = \left(\frac{1}{P(A)} + \frac{1}{P(A)}\right) \Big/ 2 = \frac{1}{P(A)}$$

Inserting the log-ratio measure leads to:

$$C_{lr}(A,B) = \left(\log\frac{P(A|B)}{P(A)} + \log\frac{P(B|A)}{P(B)}\right) \Big/ 2 = \log\frac{1}{P(A)}$$

Thus, again, the higher the probability of the robbers' being at least six feet tall is, the lower the coherence of Susan's and Tom's equivalent testimonies would be. The same

[11] This should come as no surprise. As said in footnote 7, for pairs of propositions Shogenji's formula may be rewritten as $P(A|B)/P(A)$, which is identical with $P(B|A)/P(B)$. Therefore, $C_r(A,B) = C_{Sh}(A,B)$.

holds for the coherence measures which rest on Finch's, Popper's and Levi's accounts of confirmation (in the latter cases the results equal the one for Douven and Meijs's formula):

$$\begin{aligned}
C_{\mathrm{Fi}}(\mathrm{A},\mathrm{B}) &= \left(\frac{P(\mathrm{B}|\mathrm{A})-P(\mathrm{B})}{P(\mathrm{B})}+\frac{P(\mathrm{A}|\mathrm{B})-P(\mathrm{A})}{P(\mathrm{A})}\right)\Big/2\\
&= \left(\frac{1-P(\mathrm{A})}{P(\mathrm{A})}+\frac{1-P(\mathrm{A})}{P(\mathrm{A})}\right)\Big/2=\frac{1}{P(\mathrm{A})}-1\\
C_{\mathrm{Po}}(\mathrm{A},\mathrm{B}) &= \left(\frac{P(\mathrm{B}|\mathrm{A})-P(\mathrm{B})}{P(\mathrm{B}|\mathrm{A})+P(\mathrm{B})}\times[1+P(\mathrm{A})\times P(\mathrm{A}|\mathrm{B})]\right.\\
&\quad\left.+\frac{P(\mathrm{A}|\mathrm{B})-P(\mathrm{A})}{P(\mathrm{A}|\mathrm{B})+P(\mathrm{A})}\times[1+P(\mathrm{B})\times P(\mathrm{B}|\mathrm{A})]\right)\Big/2\\
&= \left(\frac{1-P(\mathrm{A})}{1+P(\mathrm{A})}\times[1+P(\mathrm{A})]+\frac{1-P(\mathrm{A})}{1+P(\mathrm{A})}\times[1+P(\mathrm{A})]\right)\Big/2=1-P(\mathrm{A})\\
C_{\mathrm{Le}}(\mathrm{A},\mathrm{B}) &= [P(\mathrm{A}|\mathrm{B})\times P(\neg\mathrm{A})-P(\neg\mathrm{A}|\mathrm{B})\times P(\mathrm{A})\\
&\quad+P(\mathrm{B}|\mathrm{A})\times P(\neg\mathrm{B})-P(\neg\mathrm{B}|\mathrm{A})\times P(\mathrm{B})]/2\\
&= [P(\neg\mathrm{A})+P(\neg\mathrm{A})]/2=1-P(\mathrm{A})
\end{aligned}$$

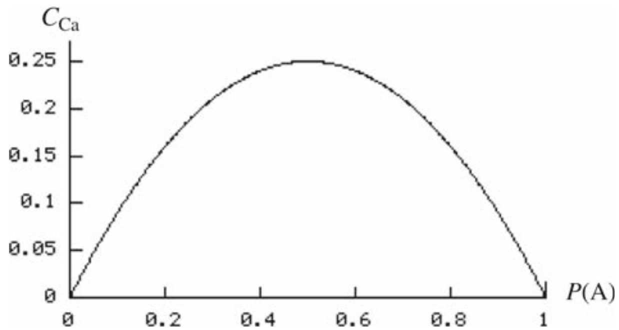
By employing Rescher's proposal, we are led to the opposite claim that the coherence of equivalent witness reports goes *up* with their probability because the former boils down to the latter:

$$\begin{aligned}
C_{\mathrm{Re}}(\mathrm{A},\mathrm{B}) &= \left(\frac{P(\mathrm{A}|\mathrm{B})-P(\mathrm{A})}{1-P(\mathrm{A})}\times P(\mathrm{B})+\frac{P(\mathrm{B}|\mathrm{A})-P(\mathrm{B})}{1-P(\mathrm{B})}\times P(\mathrm{A})\right)\Big/2\\
&= \left(\frac{1-P(\mathrm{A})}{1-P(\mathrm{A})}\times P(\mathrm{A})+\frac{1-P(\mathrm{B})}{1-P(\mathrm{B})}\times P(\mathrm{A})\right)\Big/2=P(\mathrm{A})
\end{aligned}$$

A coherence measure à la Douven and Meijs making use of Carnap's support measure leads to an even stranger result:

$$\begin{aligned}
C_{\mathrm{Ca}}(\mathrm{A},\mathrm{B}) &= [P(\mathrm{A}\&\mathrm{B})-P(\mathrm{A})\times P(\mathrm{B})+P(\mathrm{B}\&\mathrm{A})-P(\mathrm{B})\times P(\mathrm{A})]/2\\
&= P(\mathrm{A})-P(\mathrm{A})\times P(\mathrm{A})=P(\mathrm{A})\times(1-P(\mathrm{A}))
\end{aligned}$$

The graph of this function looks as follows:



Hence, in the light of this measure, the coherence of equivalent witness reports depends on their probability in a rather odd way. In the range from 0 to 0.5, the more likely it is that the bankrobbers are at least six feet tall, the more coherent

Susan's and Tom's reports would be; and for $P(\mathrm{A}) > 0.5$, the degree of coherence goes down in a symmetrical way. This is a behaviour for which we see no justification.

Why do all of these coherence measures fail? Let us go one step back and have a closer look at the measures of confirmation involved. In his landmark *Studies in Bayesian Confirmation Theory*, Fitelson directs our attention to *conclusive* evidence, viz., cases where the evidence B deductively implies the hypothesis A. Fitelson (2001, p. 42) holds that "the strength of the support [B] provides for [A] in this case should not depend on how probable [A] is (a priori). [...] After all, evidential support is supposed to be a measure of how strong the evidential *relationship* between [B] and [A] is, and deductive entailment is the strongest that such a relationship can possibly get." If B implies A, then, so Fitelson claims, a measure of confirmation is to tell us that B speaks for A to the maximal extent. It must not permit the support to vary with the prior probability of A (or the one of B).

We are not sure whether this is tenable. What does Fitelson mean by 'evidential relationship'? Actually, probabilistic measures of support rest on the relevance criterion, which states that B confirms A if B raises A's probability. It thus appears that these measures are meant to give the degree to which $P(\mathrm{A}|\mathrm{B})$ *goes beyond* $P(\mathrm{A})$. But consider an analogous case. Two plants are fertilised with a newly developed manure. After some weeks, both of them have a height of ten inches, but one of them was only two inches tall before the fertilisation whereas the other one measured six inches. It is natural to say that the first plant has grown *more rapidly* under the influence of the manure than the second because it started from a smaller height. In other words, its posterior height surpasses its prior height *to a greater extent*.

Now, doesn't the same hold for the extent to which $P(\mathrm{A}|\mathrm{B})$ goes beyond $P(\mathrm{A})$? If A_1 starts from a smaller prior probability than A_2, but ends with the same posterior, it also seems natural to claim that A_1's probability was raised *more* than A_2's. It is thus far from obvious that a hypothesis' prior plays no role for the degree to which its probability is increased by conclusive evidence. To be sure, there is nothing to be said against the idea that the support should be maximal in the case of conclusive evidence if this is taken to mean that the elbowroom left on account of the hypothesis' prior must be entirely exhausted. But this merely entails that the posterior probability of the hypothesis goes up to the maximum of 1. It does not entail that its prior is irrelevant. The previous considerations rather suggest that the increase in probability is the higher the less likely the hypothesis was.

Hence, Fitelson might be right in claiming that "deductive entailment is the strongest that [the evidential] relationship [between B and A] can possibly get". But this may simply be understood as meaning that $P(\mathrm{A}|\mathrm{B})$ is 1 if B implies A. And from this it does not follow that "the strength of the support [B] provides for [A] in this case should not depend on how probable [A] is (a priori)". For if, in conformity with the relevance criterion, strength of support is equated with increase in probability, then there is reason for assuming that A's prior has an influence indeed.

However, whether we agree with Fitelson or not, the accounts in question do not meet his desideratum. They do not assign maximal confirmation in Fitelson's sense if A is a consequence of B, but let the degree of support be sensitive to the probability of A and/or the one of B:

$$S_{\mathrm{r}}(\mathrm{A},\mathrm{B}) = \frac{1}{P(\mathrm{A})}$$

$$S_{\mathrm{lr}}(\mathrm{A},\mathrm{B}) = \log\frac{1}{P(\mathrm{A})}$$

$$S_{\mathrm{Fi}}(\mathrm{A},\mathrm{B}) = \frac{\frac{P(\mathrm{B})\times P(\mathrm{A}|\mathrm{B})}{P(\mathrm{A})} - P(\mathrm{B})}{P(\mathrm{B})} = \frac{\frac{P(\mathrm{B})}{P(\mathrm{A})} - P(\mathrm{B})}{P(\mathrm{B})} = \frac{1}{P(\mathrm{A})} - 1$$

$$S_{\mathrm{Po}}(\mathrm{A},\mathrm{B}) = \frac{\frac{P(B)}{P(A)} - P(\mathrm{B})}{\frac{P(B)}{P(A)} + P(\mathrm{B})} \times [1 + P(\mathrm{A})] = \frac{1 - P(\mathrm{A})}{1 + P(\mathrm{A})} \times [1 + P(\mathrm{A})] = 1 - P(\mathrm{A})$$

$$S_{\mathrm{Le}}(\mathrm{A},\mathrm{B}) = P(\neg\mathrm{A}) - 0 = 1 - P(\mathrm{A})$$

$$S_{\mathrm{Re}}(\mathrm{A},\mathrm{B}) = \frac{1 - P(\mathrm{A})}{1 - P(\mathrm{A})} \times P(\mathrm{B}) = P(\mathrm{B})$$

$$S_{\mathrm{Ca}}(\mathrm{A},\mathrm{B}) = P(\mathrm{B}) - P(\mathrm{A}) \times P(\mathrm{B}) = P(\mathrm{B}) \times [1 - P(\mathrm{A})]$$

It is tempting, now, to put down the failure of the coherence measures in question to this behaviour of the support measures they are based on. A and B are equivalent just in case they imply each other. In order for the given measures of coherence to grant maximal coherence to pairs of equivalent propositions, it therefore appears necessary that the corresponding accounts of confirmation let an assumption maximally support its logical consequences. Since the above-mentioned accounts do not conform to this principle, we seem to have found the reason for why the theories of coherence built on them do not pass our test.

However, this diagnosis is too simplistic. The fact that a measure of confirmation refrains from giving maximal support for cases of conclusive evidence does not automatically make the corresponding coherence measure deficient. Consider Christensen's (1999) and Nozick's (1981, p. 252) proposals for capturing support in terms of probability:

$$S_{\mathrm{Ch}}(\mathrm{A},\mathrm{B}) = P(\mathrm{A}|\mathrm{B}) - P(\mathrm{A}|\neg\mathrm{B})$$
$$S_{\mathrm{No}}(\mathrm{A},\mathrm{B}) = P(\mathrm{B}|\mathrm{A}) - P(\mathrm{B}|\neg\mathrm{A})$$

These measures are also prior-sensitive. Given that B implies A, $P(\mathrm{B}|\neg\mathrm{A})$ is 0; and so:

$$S_{\mathrm{Ch}}(\mathrm{A},\mathrm{B}) = 1 - P(\mathrm{A}|\neg\mathrm{B}) = 1 - \frac{P(A) \times P(\neg\mathrm{B}|\mathrm{A})}{P(\neg\mathrm{B})}$$

$$S_{\mathrm{No}}(\mathrm{A},\mathrm{B}) = \frac{P(\mathrm{B}) \times P(\mathrm{A}|\mathrm{B})}{P(\mathrm{A})} - 0 = \frac{P(\mathrm{B})}{P(\mathrm{A})}$$

Hence, the fact that B entails A does not guarantee that its confirmatory power is at a maximum. For example, if $P(\mathrm{A}) = 0.4, P(\mathrm{B}) = 0.3$ and $P(\neg\mathrm{B}|\mathrm{A}) = 0.25$, then $S_{\mathrm{Ch}}(\mathrm{A},\mathrm{B}) \approx 0.14$ and $S_{\mathrm{No}}(\mathrm{A},\mathrm{B}) = 0.75$. But if $P(\mathrm{A}) = 0.2, P(\mathrm{B}) = 0.1$ and $P(\neg\mathrm{B}|\mathrm{A}) = 0.5$, then $S_{\mathrm{Ch}}(\mathrm{A},\mathrm{B}) \approx 0.11$ and $S_{\mathrm{No}}(\mathrm{A},\mathrm{B}) = 0.5$. These measures provide different values, depending on the probabilities the premise B and the conclusion A possess. Like the previously discussed accounts, they do not accord with Fitelson's intuition that the extent of support is maximal, and thus constant, in the case of conclusive evidence.

However, the measures of *coherence* resulting from Christensen's and Nozick's proposals are immune to the problem of equivalent testimonies. If A and B are equivalent, then:

$$C_{\mathrm{Ch}}(\mathrm{A},\mathrm{B}) = [P(\mathrm{A}|\mathrm{B}) - P(\mathrm{A}|\neg\mathrm{B}) + P(\mathrm{B}|\mathrm{A}) - P(\mathrm{B}|\neg\mathrm{A})]/2$$
$$= [1 - 0 + 1 - 0]/2 = 1$$

The same holds if we use Nozick's formula because this just means interchanging the first and the last two addends. Since 1 is the highest possible value for both measures, they pass our test: these formulas rule that equivalent witness reports are coherent to an outstanding degree.

The reason is that the underlying accounts of confirmation, although they do not automatically let B confirm A maximally if B implies A, at least provide this result in the case of *mutual* entailment, i.e., if A implies B, too. One could argue that this does not make sense. Why should it matter for the strength to which an assumption speaks for one of its logical consequences whether the latter also entails the former or not? However, the crucial point here is that, for a Douven/Meijs-style measure of coherence to assign maximal fit to equivalent propositions, the measure of support on which it rests need not give the maximum value for *all* cases of conclusive evidence. It suffices that it honours propositions with maximally confirming each other if *both* of them are conclusive evidence for the other. Equivalent testimonies cause trouble for the accounts generated with the help of $S_{\mathrm{d}}, S_{\mathrm{r}}, S_{\mathrm{lr}}, S_{\mathrm{Fi}}, S_{\mathrm{Po}}, S_{\mathrm{Le}}, S_{\mathrm{Re}}$ and S_{Ca} because these measures do not even concede maximum support in the case of *mutual* implication.

What about Good's (1984) and Fitelson's (2001, Ch. 3.2) favourite, the log-likelihood measure of confirmation? Since log (a/b) = log a − log b, the coherence of equivalent testimonies is to be calculated as follows:

$$C_{\mathrm{ll}}(\mathrm{A},\mathrm{B}) = \left(\log\frac{P(\mathrm{B}|\mathrm{A})}{P(\mathrm{B}|\neg\mathrm{A})} + \log\frac{P(\mathrm{A}|\mathrm{B})}{P(\mathrm{A}|\neg\mathrm{B})}\right)\Big/2$$
$$= [\log P(\mathrm{B}|\mathrm{A}) - \log P(\mathrm{B}|\neg\mathrm{A}) + \log P(\mathrm{A}|\mathrm{B}) - \log P(\mathrm{A}|\neg\mathrm{B})]/2$$
$$= 0 - \log 0$$

Although −log a goes against ∞ if a approaches 0, logarithm functions are only defined for numbers above 0. So, strictly speaking, the log-likelihood account results in a coherence measure which remains silent on equivalent witness reports. This could be one of the reasons for why Fitelson's own theory of coherence does not make use of the account of confirmation he has advocated but an ordinally equivalent variant of it. This leads us to the next section.

5 Fitelson's measure

In his article 'A Probabilistic Theory of Coherence' (2003), Fitelson offered a measure which he revised a bit in the online-paper 'Two Technical Corrections to My Coherence Measure' (2004). Although developed independently of Douven and Meijs's recipe, it is a cake which can be baked after it.

Unlike Douven and Meijs, Fitelson does not employ the difference measure of confirmation, but adopts a slight modification of Kemeny and Oppenheim's (1952) proposal. The values of Fitelson's support measure range from −1 to 1, where a number greater than 0 represents confirmation of A by B and a number smaller than 0 disconfirmation. If B implies A (and is not logically false), then Fitelson takes B to support A to the maximum degree: $S_{\mathrm{F}}(\mathrm{A},\mathrm{B}) = 1$. If B implies ¬A, the support is as small as it can be: $S_F(A, B) = -1$.[12] And if neither A nor its negation follows from B, the extent to which B speaks for A arises from Kemeny and Oppenheim's function:

$$S_F(A, B) = \frac{P(B|A) - P(B|\neg A)}{P(B|A) + P(B|\neg A)}$$

The coherence of a set is then defined in accordance with Douven and Mejis's pattern. For a pair {A, B} it is simply $[S_F(A, B) + S_F(B, A)]/2$. Like the values of the support function S_F, the numbers provided by this coherence measure lie in the range from −1 to 1. A value above 0 means that the set is coherent, a value below 0 that it is incoherent.

The result for our test case is obvious. Since what Susan says implies what Tom says, and conversely, their reports confirm each other to the highest possible degree 1. Therefore, they also cohere maximally: $C_F(A, B) = (1 + 1)/2 = 1$. That is, equivalent testimonies fit together perfectly on Fitelson's measure, regardless of how likely the propositions put forward are.

6 Olsson's measure

The same holds for the formula which Olsson (2002, p. 250) tentatively suggested. It is quite similar to Shogenji's:

$$C_{Ol}(A_1, \ldots, A_n) = \frac{P(A_1 \,\&\, \cdots \,\&\, A_n)}{P(A_1 \vee \cdots \vee A_n)}$$

The idea here is: the closer the probability that *each* proposition is true comes to the probability that *at least one* of them is true, the more they cohere. Since the probability of a conjunction cannot exceed the probability of the corresponding disjunction, the values of this function range from 0 to 1. (There is no universal threshold here where incoherence passes into coherence, but this is irrelevant to the following argumentation.)

It is easy to see that Olsson's measure scores a hit if we subject it to our test. If A and B entail each other, then A & B is equivalent to $A \vee B$, so that $P(A \,\&\, B) = P(A \vee B)$. Hence, $C_{Ol}(A, B)$ is in such a case 1. Olsson's account rules that the reports of two witnesses cohere to the highest degree when they are logically equivalent.

7 Conclusion

Our touchstone allows us to throw a significant number of probabilistic coherence measures overboard. Neither Shogenji's proposal nor Douven and Meijs's nor the ones which result from supplementing Douven and Meijs's general schema by the ratio, the log-ratio, the log-likelihood, Finch's, Popper's, Levi's, Rescher's or Carnap's measure of support pass this test. The winners are the coherence theories of Fitelson, Olsson and the two accounts which are generated after Douven and Meijs's recipe from Christensen's and Nozick's measures of confirmation, respectively. They are in the semi-final because they let equivalent testimonies be maximally coherent, no matter how likely it is that they are true.

[12] These additions make sure that you get maximal values for all of these deductive cases. The Kemeny/Oppenheim measure is undefined if $P(B)$ is 0 or $P(A)$ is 0 or 1.

Further considerations may pick out two of these measures as the participants in the final; and perhaps it is possible to distinguish one of them as the final winner. On the other hand, it might turn out that none of our semi-finalists, but only a measure still to be developed, has the makings of adequately capturing coherence. Or even that no purely probabilistic account will ever be able to get a grip on this notion.[13] We would be quite happy, however, with having the reader convinced that the above-mentioned measures are definitively out of play.

Appendix: Bovens and Hartmann's quasi-ordering

We were concerned with probabilistic *measures* of coherence. For reasons which do not matter here, Bovens and Hartmann think that it is merely possible to construct a so-called *quasi-ordering*. An ordering does not enable absolute but only relative judgements, which means that it does not give *degrees* of coherence but tells us only whether one set is *more* (*equally* or *less*) coherent than another set. Moreover, a quasi-ordering is not complete insofar as the relative coherence of some sets is not defined by it.

In a nutshell, Bovens and Hartmann argue that the following formula may be used to compare systems of statements with respect to coherence:[14]

$$\mathrm{F_r}(\mathrm{A}_1, \ldots, \mathrm{A}_n) = \frac{\mathrm{a}_0 + (1 - \mathrm{a}_0) \times (1 - \mathrm{r})^n}{\sum_{i=0}^{n} [\mathrm{a}_i \times (1 - \mathrm{r})^i]},$$

where a_i is the probability that i of the n statements are false (e.g., a_0 is $P(\mathrm{A}\ \&\ \mathrm{B})$ for a two-member set) and r is the reliability of the statements' sources. The sources are supposed to be independent in a certain way; and they are partially reliable, so that, according to the definition of $r, 0 < r < 1$.

$\mathrm{F_r}$ is functionally dependent on the credibility of the sources. Since, intuitively, the coherence of a set of propositions is not sensitive to how reliable their sources are,[15] $\mathrm{F_r}$ must not be viewed as a measure of coherence. Bovens and Hartmann's claim is rather that supplementing this formula with a simple assumption makes it possible to determine the relative coherence of two sets of propositions:

$$S_1 \text{ is at least as coherent as } S_2 \text{ iff for all values of } r, \mathrm{F_r}(S_1) \geq \mathrm{F_r}(S_2)$$

Analogously, if the $\mathrm{F_r}$-values for S_1 are, for all degrees of partial reliability, greater (smaller) than the corresponding values for S_2, then S_1 is more (less) coherent than S_2.

What happens in the case of logically equivalent testimonies? If one of them is false, the other one must be false as well. Hence a_1, the probability that one, and only one, of them is false, equals 0, entailing that $1 - \mathrm{a}_0$ is identical with a_2. Therefore:

$$\mathrm{F_r}(\mathrm{A}, \mathrm{B}) = \frac{\mathrm{a}_0 + \mathrm{a}_2 \times (1 - \mathrm{r})^2}{\mathrm{a}_0 + 0 \times (1 - \mathrm{r}) + \mathrm{a}_2 \times (1 - \mathrm{r})^2} = 1$$

13 For some general arguments against probabilistic measures of coherence, cf. Siebel (2005a).

14 Cf. Bovens and Hartmann (2003a, Sects. 4f.); Bovens and Hartmann (2003b, Chs. 1.4, 2.2).

15 One could wonder why Bovens rejects a sense of coherence in which it is sensitive to the reliability of the information's sources (Bovens and Hartmann, 2003b, p. 34) while allowing a sense according to which the probability of the information has an influence (Bovens and Olsson, 2000, p. 688f., fn. 1).

That is, regardless of how credible the witnesses are, the F_r-value for a pair of equivalent claim will always be at least as high as the value for any other set of statements whose sources are reliable to the same extent. The reason for this is that the nominator of the given fraction, the sum of a_0 and $a_n \times (1-r)^n$, cannot be greater than the denominator, the sum of a_0 and $a_n \times (1-r)^n$ *plus* the further addends $a_i \times (1-r)^i$. For any system $\{A_1, \ldots, A_n\}$ and any value of the reliability parameter r, $F_r(A_1, \ldots, A_n)$ cannot exceed 1. Bovens and Hartmann's quasi-ordering thus passes our test by entailing that a pair of equivalent witness reports is at least as coherent as every other set of propositions.[16]

References

Bartelborth, T. (1996). *Begründungsstrategien. Ein Weg durch die analytische Erkenntnistheorie*. Berlin: Akademie-Verlag.

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

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

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

Bovens, L., & Olsson, E. J. (2002). Believing more, risking less: On coherence, truth and non-trivial extensions. *Erkenntnis, 57*, 137–150.

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

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

Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese* (Forthcoming), DOI: 10.1007/s11229-006-9131-z.

Eells, E., & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies, 107*, 129–142.

Finch, H. A. (1960). Confirming power of observations metricized for decisions among hypotheses. *Philosophy of Science, 27*, 293–307.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science, 66* (Proceedings), S362–S378.

Fitelson, B. (2001). *Studies in Bayesian confirmation theory*. Ph.D. thesis, University of Wisconsin at Madison. Online: http://fitelson.org/thesis.pdf.

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

Fitelson, B. (2004). Two technical corrections to my coherence measure. Online: http://fitelson.org/coherence2.pdf.

Gillies, D. (1986). In defense of the Popper-Miller argument. *Philosophy of Science, 53*, 110–113.

Good, I. J. (1984). The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation, 19*, 294–299.

Horwich, P. (1998). Wittgensteinian Bayesianism. In M. Curd & J. A. Cover (Eds.), *Philosophy of science: The central issues* (pp. 607–624). New York and London: Norton.

Jeffrey, R. (1992). *Probability and the art of judgement*. Cambridge: Cambridge University Press.

Kemeny, J., & Oppenheim, P. (1952). Degrees of factual support. *Philosophy of Science, 19*, 307–324.

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

Kyburg, H. E., Jr. (1983). Recent work in inductive logic. In T. Machan & K. Lucey (Eds.), *Recent work in philosophy* (pp. 87–150). Totowa/NJ: Rowman & Allanheld.

Levi, I. (1962). Corroboration and rules of acceptance. *British Journal for the Philosophy of Science, 13*, 307–313.

Lewis, C. I. (1946). *An analysis of knowledge and valuation*. Chicago: Open Court.

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

[16] A closer look reveals that Bovens and Hartmann's account actually rests on our intuition. For they take a maximally coherent information set to be a set of equivalent elements, and "assess the coherence of an information set by measuring the proportion of the confidence boost that we actually receive, relative to the confidence boost that we would have received *had we received this very same information in the form of maximally coherent information*" (2003b, p. 32f.).

Moretti, L., & Akiba, K. (2007). Probabilistic measures of coherence and the problem of belief individuation. *Synthese*, 154, 73–95.
Nozick, R. (1981). *Philosophical explanations*. Cambridge/MA: Harvard University Press.
Olsson, E. J. (2002). What is the problem of coherence and truth? *The Journal of Philosophy, 94*, 246–272.
Olsson, E. J. (2005). *Against coherence*. Oxford: Oxford University Press.
Olsson, E. J., & Shogenji, T. (2004). Can we trust our memories? C. I. Lewis's coherence argument. *Synthese, 142*, 21–41.
Popper, K. R. (1954). Degree of confirmation. *British Journal for the Philosophy of Science, 5*, 143–149.
Rescher, N. (1958). Theory of evidence, *Philosophy of Science, 25*, 83–94.
Rosenkrantz, R. (1994). Bayesian confirmation: Paradise regained. *British Journal for the Philosophy of Science, 45*, 467–476.
Schlesinger, G. (1995). Measuring degrees of confirmation. *Analysis, 55*, 208–212.
Shogenji, T. (1999). Is coherence truth conducive? *Analysis, 59*, 338–345.
Siebel, M. (2005a). Against probabilistic measures of coherence. *Erkenntnis*, *63*, 335–360.
Siebel, M. (2005b). Thagard's measure of coherence: Corrected and compared with probabilistic accounts. Manuscript.
Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.
Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science, 22*, 1–24.