Erkenn
DOI 10.1007/s10670-016-9820-z

ORIGINAL ARTICLE

Carnap's Relevance Measure as a Probabilistic Measure of Coherence

Jakob Koscholke[1]

Received: 11 August 2015 / Accepted: 21 May 2016

Abstract Tomoji Shogenji is generally assumed to be the first author to have presented a probabilistic measure of coherence. Interestingly, Rudolf Carnap in his *Logical Foundations of Probability* discussed a function that is based on the very same idea, namely his well-known relevance measure. This function is largely neglected in the coherence literature because it has been proposed as a measure of evidential support and still is widely conceived as such. The aim of this paper is therefore to investigate Carnap's measure regarding its plausibility as a candidate for a probabilistic measure of coherence by comparing it to Shogenji's. It turns out that both measures (i) satisfy and violate the same adequacy constraints, (ii) despite not being ordinally equivalent exhibit a strong correlation with each other in a Monte Carlo simulation and (iii) perform similarly in a series of test cases for probabilistic coherence measures.

1 Introduction

In his response to Klein and Warfield's (1994) rejection of the truth-conduciveness of coherence, Shogenji (1999) presented a mathematical function which is supposed to measure the degree of coherence of two propositions x_1 and x_2 under some joint probability distribution P, namely $C_{sho}(x_1, x_2) = P(x_1 \wedge x_2)/P(x_1) \times P(x_2)$. He also provided a straightforward generalization of this measure for any finite, non-empty, non-singleton set X of propositions under some probability function P:

$$C_{sho}(X) = P\left(\bigwedge_{x_i \in X} x_i\right) \bigg/ \prod_{x_i \in X} P(x_i)$$

✉ Jakob Koscholke
jakob.koscholke@uni-oldenburg.de

[1] Philosophy Department, University of Oldenburg, 26111 Oldenburg, Germany

Published online: 02 June 2016

For a set of probabilistically independent propositions Shogenji's measure equals 1 while for negatively/positively dependent propositions the measure takes values below/above this threshold value.[1] Accordingly, a value of 1 is treated as neutral coherence value while values below/above 1 are interpreted as degrees of incoherence/coherence. The measure is therefore commonly understood as measuring coherence in terms of the ratio-wise deviation from probabilistic independence (cf. Shogenji 1999, 339–340; for criticism of this view or Shogenji's measure in general cf. e.g. Akiba 2000; Fitelson 2003; Wheeler 2009).

Now, it is quite interesting to see that Carnap (1950, §67) in his seminal *Logical Foundations of Probability* suggested a very similar function for measuring the degree to which a piece of evidence e (incrementally, not absolutely) confirms/disconfirms a hypothesis h under some probability distribution P, his relevance measure $\mathfrak{r}(h, e) = P(h \wedge e) - P(h) \times P(e)$. Carnap's measure is obviously a special case of the following function which similar to Shogenji's generalization is defined for any non-empty, non-singleton set X rather than only pairs of propositions (h, e) under some probability function P:

$$C_{car}(X) = P\left(\bigwedge_{x_i \in X} x_i\right) - \prod_{x_i \in X} P(x_i)$$

Apparently, this function only differs from Shogenji's in using the difference instead of the ratio. But just like Shogenji's measure it also quantifies degree to which the propositions deviate from probabilistic independence. For this measure independence is associated with a value of 0 and values below/above this threshold indicate negative/positive dependence. Following the general interpretation of the Shogenji measure, it seems quite natural to interpret Carnap's measure as a candidate explication for coherence, measured in terms of the difference-wise deviation from probabilistic independence. This interpretation is not only very natural but it also provides new hope for Carnap's relevance measure since it is well documented in the literature that the measure fares rather poorly as a measure of confirmation (cf. e.g. Eells and Fitelson 2002; Crupi et al. 2007).[2]

There is also an improved, subset-sensitive version of Shogenji's measure by Schupbach (2011).[3] Based on Fitelson's (2003) observation that for a set X with n propositions Shogenji's measure only takes into account n-wise dependence neglecting the dependencies in the subsets, Schupbach suggested to apply a log-

[1] By definition, a set of propositions X is negatively dependent/independent/positively dependent if and only if $P(\bigwedge_{x_i \in X} x_i) < / = / > \prod_{x_i \in X} x_i$ (cf. e.g. Kolmogorov 1956, §I.5).

[2] It is also worth noticing that Carnap's interpretation of the measure slightly differs from what is usually assumed to be a confirmation measure. According to Carnap, the measure is supposed to quantify the *mutual* relevance two propositions have for each other as opposed to the *one-way* relevance one proposition, a piece of evidence e, might have for another proposition, a hypothesis h (cf. Carnap 1950, §67). This idea of mutual relevance also suggests to investigate Carnap's measure as a probabilistic measure of coherence.

[3] Other authors have presented probabilistic coherence measures as well, e.g. Douven and Meijs (2007), Fitelson (2003), Glass (2002), Meijs (2005), Olsson (2002), Roche (2013) and Schippers (2014). However, these measures will not be discussed here since they are not based on the idea of coherence as deviation from probabilistic independence. Nevertheless, a comparative study between the omitted and the included measures should be taken into consideration in future research.

normalization of Shogenji's measure to each X'_{ij} denoting X's ith subset with $j \geq 2$ propositions and to calculate a weighted average over the resulting values, formally:

$$C'_{sho}(X) = \frac{\sum_{j=2}^{n} \sum_{i=1}^{\binom{n}{j}} \log\left(C_{sho}(X'_{ij})\right) \times \binom{n}{j}^{-1}}{n-1}$$

The same generalization can be applied to Carnap's measure, resulting in a more fine-grained version of the measure. But since unlike Shogenji's measure Carnap's measure is point symmetric around its independence threshold 0, it does not need to be log-normalized:

$$C'_{car}(X) = \frac{\sum_{j=2}^{n} \sum_{i=1}^{\binom{n}{j}} C_{car}(X'_{ij}) \times \binom{n}{j}^{-1}}{n-1}$$

Now, the aim of this paper is rather straightforward. It is to examine whether C_{car} or C'_{car} are at least as plausible as candidates for probabilistic measures of coherence as C_{sho} and C'_{sho} respectively—especially if we keep in mind that C_{sho} and C'_{sho} are widely accepted as probabilistic measures of coherence. In order to do this we will compare the measures with respect to a set of coherence desiderata in Sect. 2 and with respect to a collection of coherence test cases from the literature in Sect. 3. Our findings are summarized in Sect. 4.

2 Coherence Desiderata

In order to evaluate the adequacy of probabilistic coherence measures, several desiderata have been suggested in the literature. The idea underlying such desiderata is to formulate structural properties a coherence measure should have in order to be adequate. Recently, Schippers (2014) presented a survey of the desiderata discussed so far. We will inspect the introduced measures with respect to these desiderata and one further, more recent desideratum which is not included in Schippers' investigation. Proofs for the corresponding observations are given in the Appendix.

The first desideratum is due to Fitelson (2003). According to him, the degree of coherence of a set of propositions should be a function of the joint probabilistic dependencies of all subsets of this set:

(D_1) For any probability distribution P over a set X and any coherence measure C with neutrality threshold θ_C:

1. $C(X) > \theta_C$ if all $X' \subseteq X$ are positively dependent
2. $C(X) = \theta_C$ if all $X' \subseteq X$ are independent
3. $C(X) < \theta_C$ if all $X' \subseteq X$ are negatively dependent

As indicated above, the threshold values are $\theta_{C_{sho}} = 0$ and $\theta_{C'_{sho}} = \theta_{C_{car}} = \theta_{C_{car}} = 1$. It can easily be shown that the following holds:

Observation 1 C_{sho}, C_{car}, C'_{sho} and C'_{car} satisfy D_1.

Since the measures are defined to measure the deviation from probabilistic independence, it is no wonder that they satisfy Fitelson's desideratum. Nevertheless, it is worth noticing that the idea underlying this desideratum is that the degree of coherence of a set of propositions does not only depend on the probabilistic dependence of the whole set but also depends on the probabilistic dependencies of all subsets. While C_{sho} and C_{car} only consider the total set, C'_{sho} are and C'_{car} also take into account its subsets.

The next desideratum has been proposed by Bovens and Olsson (2000) and generalized by Schippers (2014). Their intuition is that increasing conditional probabilities among the proposition in some set should increase its coherence, more formally:

(D_2) Let X_1 and X_2 denote a set X under distribution P_1 and P_2 respectively and let X', X'' be non-empty, disjoint subsets of X. Then for any X_1, X_2 and any coherence measure C: If $P_1(\bigwedge_{x_i \in X'} x_i | \bigwedge_{x_j \in X''} x_j) > P_2(\bigwedge_{x_i \in X'} x_i | \bigwedge_{x_j \in X''} x_j)$ for all X', X'', then $C(X_1) > C(X_2)$.

Schippers (2014) elegantly showed that D_1 and D_2 are jointly inconsistent. Relying on this result we can state the following:

Observation 2 C_{sho}, C_{car}, C'_{sho} and C'_{car} do not satisfy D_2.

Fitelson (2003) and Siebel and Wolff (2008) have suggested another desideratum. For them, logical equivalence is the prime example of maximal coherence. Since Fitelson's desideratum is restricted to satisfiable propositions, we refer to Siebel and Wolff's more general version for any kinds of propositions:

(D_3) For any probability distribution P over a set X of logically equivalent propositions and any coherence measure C: $C(X) = \max(C)$.

As can be seen by counterexample, the following holds:

Observation 3 C_{sho}, C_{car}, C'_{sho} and C'_{car} do not satisfy D_3.

Still, it is worth noticing that the measures do not behave identically. For a set X of logically equivalent propositions $x_1, \ldots, x_n$ it holds that $P(x_1) = \ldots = P(x_n)$, which entails that $C_{sho}(X) = 1/P(x_i)^{n-1}$ and $C_{car}(X) = P(x_i) - P(x_i)^n$ for any $x_i \in X$. Hence, in cases of low but non-zero/high $P(x_i)$ Shogenji's measure assigns high/low or neutral degrees of coherence. Carnap's measure, however, strongly depends on the number of propositions n. For instance, for $n = 2$ Carnap's measure has a maximum of 0.25 only if $P(x_i) = 0.5$ and assigns low degrees of coherence in cases of low and high $P(x_i)$. For $n > 2$ the maximum value of 0.25 can be exceeded and the measure assigns high degrees of coherence only if $P(x_i)$ is high but not 1. Fitelson (2003) already noted that C_{sho}'s behaviour in the aforementioned cases is highly counter-intuitive. Now, it seems that C_{car} does not behave any better. Since these effects carry over to the alternative generalizations as well, the measures should be avoided for sets of equivalent propositions.

The next desideratum by Fitelson (2003) can be viewed as the counterpart to D_3. Just as logical equivalence can be considered the prime example of maximal coherence, unsatisfiability of all subsets can be considered the prime example of minimal coherence:

(D_4) For any probability distribution P over a set X where each $X' \subseteq X$ is unsatisfiable and any coherence measure C: $C(X) = \min(C)$.

Regarding this desideratum we can state the following:

Observation 4 C_{sho}, C_{car}, C'_{sho} and C'_{car} do not satisfy D_4.

As with the aforementioned desideratum there are some subtle differences between the measures. If each $X' \subseteq X$ is unsatisfiable including the singleton subsets, then $P(x_i) = 0$ for all $x_i \in X$, which entails that $C_{sho}(X)$ and hence $C'_{sho}(X)$ are undefined, whereas $C_{car}(X) = C'_{car}(X) = 0$. However, if all singleton subsets of X are satisfiable and only all non-singleton subsets are unsatisfiable it holds that $C_{sho}(X) = 0$ and $C'_{sho}(X)$ is undefined whereas $C_{car}(X) < 0$ and $C'_{car}(X) < 0$. Hence, Shogenji's measure is minimal indicating the strongest degree of incoherence while Carnap's measure assigns values ranging from neutral coherence to degrees of incoherence. It seems that the measures are only of limited use for the assessment of unsatisfiable sets of propositions.

Recently, Shogenji (2013) himself formulated a desideratum. The idea is that the degree of coherence of some set of propositions should increase *ceteris paribus* with increasing joint probability of the propositions and decrease *ceteris paribus* with an increasing product of the marginal probabilities of the propositions:

(D_5) For any probability distribution P over a set $X = \{x_1, x_2\}$ and any coherence measure C:

1. $C(X)$ is an increasing function of $P(x_1 \wedge x_2)$
2. $C(X)$ is a decreasing function of $P(x_1) \times P(x_2)$
3. $C(X)$ is neutral if $P(x_1 \wedge x_2) = P(x_1) \times P(x_2)$

According to Shogenji, the following holds:

Observation 5 C_{sho}, C_{car}, C'_{sho} and C'_{car} satisfy D_5.

It is quite interesting to see that Shogenji (2013) himself has noticed that aside from his own measure Carnap's measure also satisfies his desideratum.

The results from above can be summarized in the following Table 1 where 1 indicates satisfaction and 0 violation of the corresponding desideratum:

Table 1 Summary of the results

	C_{sho}	C_{car}	C'_{sho}	C'_{car}
D_1	1	1	1	1
D_2	0	0	0	0
D_3	0	0	0	0
D_4	0	0	0	0
D_5	1	1	1	1

Table 2 Cross-correlation matrix for the measures

	C_{sho}	C_{car}	C'_{sho}	C'_{car}
C_{sho}	1	0.972	0.992	0.954
C_{car}	0.972	1	0.966	0.980
C'_{sho}	0.992	0.966	1	0.968
C'_{car}	0.954	0.980	0.968	1

Apparently, all measures satisfy and violate the same five desiderata that have been put forward in the coherence literature. These similarities might raise the suspicion that the measures are just ordinally equivalent versions of each other.[4] But this can easily be refuted:

Observation 6 No pair in $\{C_{sho}, C_{car}, C'_{sho}, C'_{car}\}$ is ordinally equivalent.

This observation guarantees that the measures can disagree in judging a set of propositions more or less coherent than another set. Hence, despite the aforementioned similarities the measures are in fact different and have distinct characteristics regarding their coherence assessments.

Nevertheless, the measures are highly correlated with each other, as can be shown using Monte Carlo simulation methods.[5] For this simulation 1 million joint probability assignments over 3 propositions were generated using the Mersenne-Twister algorithm (cf. Matsumoto and Nishimura 1998). For each assignment, the corresponding values of the measures have been recorded in order to calculate a cross-correlation matrix containing Spearman's (1904) rank correlation coefficient (Pearson's (1895) product moment correlation coefficient cannot be applied here since it assumes a linear relationship between the variables) ranging from -1 indicating maximum anti-correlation to 1 indicating maximum correlation (Table 2).

As the Table 2 indicates, there is a strong correlation of 0.972 between C_{sho} and C_{car} and 0.968 between C'_{sho} and C'_{car}. This means that increasing/decreasing C_{sho} or C'_{sho} values are strongly associated with increasing/decreasing C_{car} or C'_{car} values.

The evaluation of the coherence desiderata and the simulation results suggest that Carnap's measure and its alternative generalization are indeed very closely related to Shogenji's and its alternative generalization. And since Shogenji's measure and Schupbach's alternative version are commonly interpreted as probabilistic measures of coherence, there seems to be no good reason why this interpretation should not be applicable to Carnap's measure and its alternative generalization as well. With this in mind, we will examine the measures in a series of test cases for probabilistic coherence measures in the next section.

[4] By definition, two measures C_1 and C_2 are ordinally equivalent if and only if for all sets X_1, X_2 it holds that $C_1(X_1) </=/> C_1(X_2)$ if and only if $C_2(X_1) </=/> C_2(X_2)$.

[5] A similar approach has been employed by Tentori et al. (2007) for Bayesian measures of confirmation. For a comprehensive introduction to Monte Carlo methods see e.g. Hammersley and Handscomb (1964).

3 Coherence Test Cases

Test cases have proven to be both convenient and efficient for the evaluation of probabilistic coherence measures. They typically consist of a scenario providing probabilistic information about one or more sets of propositions and a normative coherence assessment for these sets. If a measure disagrees with this coherence assessment and the assessment has strong intuitive support, then the measure is to be considered inadequate.

In a recent study, Koscholke (2015) submitted several measures to 11 test cases from the literature. They include Akiba's (2000) die case T_1, BonJour's (1985) raven case T_2, Bovens and Hartmann's (2003) Tweety case T_3, their Tokyo case T_4, their culprit case T_5, Glass' (2005) dodecahedron case T_6, Meijs' (2005) Samurai sword case T_7 and his (2005) rabbit case T_8, Meijs and Douven's (2005) plane lottery case T_9, Schupbach's (2011) robber case T_{10} and Siebel's 2004 pickpocketing case T_{11}. We will complement this collection by Schippers and Siebel's (2015) inconsistent testimonies case T_{12} which is not included in Koscholke's survey. The results are summarized in the Table 3. Here, 1 indicates a positive test case result while 0 indicates a negative result.

Apparently, there are only four test cases in which the measures differ, namely T_7, T_9, T_{10} and T_{12}. Let us briefly take a look at them.

Meijs' (2005) samurai sword test case runs as follows. Imagine that a murder occurred in a large city and we are interested in finding the murderer. Imagine the following two situations:

Situation 1: There are ten million independent and equally likely suspects. 1059 suspects are Japanese, 1059 suspects own a Samurai sword, 9 suspects are Japanese and own a Samurai sword.

Situation 2: There are 100 independent and equally likely suspects. 10 suspects are Japanese, 10 suspects own a Samurai sword, 9 suspects are Japanese and own a Samurai sword.

Now, consider the following two propositions:

x_1 : The murderer is Japanese.
x_2 : The murderer owns a Samurai sword.

Meijs argues that X_1 denoting $\{x_1, x_2\}$ in situation 1 is less coherent than X_2, the same set in situation 2. The measures give the following verdicts (Table 4).

Table 3 Summary of the test case results

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
C_{sho}	0	1	1	1	1	0	0	0	1	0	0	0
C_{car}	0	1	1	1	1	0	1	0	1	0	0	0
C'_{sho}	0	1	1	1	1	0	0	0	0	1	0	0
C'_{car}	0	1	1	1	1	0	1	0	1	1	0	1

Table 4 Results for Meijs' samurai sword case

	X_1	X_2	Result
C_{sho}	80.3	9	0
C_{car}	0.000000889	0.08	1
C'_{sho}	1.9	0.954	0
C'_{car}	0.000000889	0.08	1

Here, Shogenji's measure and its alternative version disagree with Meijs' coherence assessment by assigning a higher degree of coherence to X_1 than to X_2. By contrast, Carnap's measure and its alternative generalization do the trick.

In Meijs and Douven's (2005) plane lottery test case we are asked to imagine a person participating in a lottery consisting of a random flight in a windowless plane. Her chances for flying to the following locations are as follows: 4/100 for the North Pole, 49/100 for the South Pole and 47/100 for New Zealand. The probability of seeing a penguin given she is on the South Pole is 10/49, in New Zealand 1/47 and on the North Pole 0. Suppose that she leaves the plane not knowing where she has landed, facing an unrecognizable animal. Now, consider that she is confronted with the following three propositions:

x_1 : The animal you see is a penguin.
x_2 : You are on the North Pole.
x_3 : You are on the South Pole.

According to Meijs and Douven, the set $X_1 = \{x_1, x_2\}$ is less coherent than $X_2 = \{x_1, x_3\}$ since there are no penguins on the Northpole. The results are as follows (Table 5).

Apparently, only the alternative generalization of Shogenji's measure fails in this test case due to an undefined function value for X_1.[6]

The next test case is Schupbach's (2011) robber case. Imagine eight equally likely suspects for a robbery and consider the following independent and equally reliable witness reports:

x_1 : The robbery was committed by suspect 1, 2 or 3.
x_2 : The robbery was committed by suspect 1, 2 or 4.
x_3 : The robbery was committed by suspect 1, 3 or 4.
x_4 : The robbery was committed by suspect 1, 4 or 5.
x_5 : The robbery was committed by suspect 1, 6 or 7.

According to Schupbach, $X_1 = \{x_1, x_2, x_3\}$ is more coherent than $X_2 = \{x_1, x_4, x_5\}$ since the agreement about who is the robber is much stronger in X_1 than in X_2. Let us take a look at the measures' verdicts (Table 6).

[6] Here and in the following tables 'NaN' is short for 'not a number' indicating undefined function values due to division by 0 or log(0). Following Siebel and Wolff (2008), undefined function values are interpreted as suspended coherence assessments.

Table 5 Results for Meijs and Douven's plane lottery case

	X_1	X_2	Result
C_{sho}	0	1.86	1
C_{car}	−0.004	0.046	1
C'_{sho}	NaN	0.268	0
C'_{car}	−0.004	0.046	1

Table 6 Results for Schupbach's robber case

	X_1	X_2	Result
C_{sho}	2.37	2.37	0
C_{car}	0.0723	0.0723	0
C'_{sho}	0.312	0.162	1
C'_{car}	0.0908	0.0283	1

This case is a problem for C_{sho} and C_{car} since both measures treat X_1 and X_2 as equally coherent. Their alternative generalizations, however, master the test case. This is no surprise since Schupbach formulated this test case as an argument for his subset-sensitive generalization of Shogenji's measure.

Finally, with their inconsistent testimonies case Schippers and Siebel (2015) presented a modification of Schupbach's test case. Suppose there are 6 suspects and each of them is equally likely to be the culprit. Now, consider the following independent and equally reliable witness reports:

x_1 : The criminal was either suspect 1 or 2.
x_2 : The criminal was either suspect 2 or 3.
x_3 : The criminal was either suspect 1 or 3.
x_4 : The criminal was either suspect 3 or 4.
x_5 : The criminal was either suspect 5 or 6.

Obviously, $X_1 = \{x_1, x_2, x_3\}$ and $X_2 = \{x_1, x_4, x_5\}$ are inconsistent. But the propositions in X_1 are at least pairwise consistent whereas in X_2 they are not. Accordingly, Schippers and Siebel argue that X_1 should be more coherent than X_2. Here are the results (Table 7).

Apparently, only the alternative generalization of Carnap's measure masters this case. The other measures fail because they are either undefined or treat both sets as equally coherent.

Table 7 Results for Siebel and Schippers' inconsistent testimonies case

	X_1	X_2	Result
C_{sho}	0	0	0
C_{car}	−0.037	−0.037	0
C'_{sho}	NaN	NaN	0
C'_{car}	0.009	−0.074	1

The evaluation of the test cases shows that Carnap's measure and its alternative generalization perform quite well compared to Shogenji's and its alternative version. Every test case mastered by Shogenji's measure is also mastered by Carnap's measure and every test case mastered by the alternative generalization of Shogenji's measure is also mastered by the alternative generalization of Carnap's measure. Thus, Carnap's measure and especially its alternative version represent quite promising candidates for probabilistic measures of coherence.

4 Conclusion

The results from Sect. 2 point out strong similarities between Shogenji's coherence measure and Schupbach's alternative generalization on the one hand and Carnap's relevance measure and its alternative generalization on the other. Based on these similarities it seems rational to claim that Carnap's measure and its alternative version are at least as plausible as candidates for a probabilistic coherence measure as Shogenji's measure and its refined version. Additionally, the results from Sect. 3 show that in a number of test cases Carnap's measure and its refined version provide coherence assessments several philosophers have argued to be intuitively correct for the respective cases. Hence, Carnap's measure and especially its fine-grained version can be considered promising candidates for a probabilistic measure of coherence as Shogenji's measure and Schupbach's alternative generalization. This does, of course, not mean that Carnap actually proposed a probabilistic measures of coherence. Rather, it means that the function Carnap proposed as a candidate for a measure of confirmation can more fruitfully be interpreted as a candidate probabilistic measure of coherence. It should therefore be included in future research on probabilistic coherence measures.

Acknowledgments I would like to thank (in alphabetical order) Michael Schippers, Jonah Schupbach, Mark Siebel and Elia Zardini for helpful comments or discussion. This work was funded by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program SPP 1516 New Frameworks of Rationality (grant SI 1731/1-1 to Mark Siebel).

Appendix

Proof (*Observation* 1) For C_{sho} cf. Schippers (2014). For C_{car} inspect the definition of independence. If all $X' \subseteq X$ are positively dependent/independent/negatively dependent then (trivially) X is also positively dependent/independent/negatively dependent. It is obvious that the following equivalent transformation holds:

$$P\left(\bigwedge_{x_i \in X} x_i\right) \gtreqless \prod_{x_i \in X} P(x_i) \;\Leftrightarrow\; P\left(\bigwedge_{x_i \in X} x_i\right) - \prod_{x_i \in X} P(x_i) \gtreqless 0$$

The expression on the right hand side is nothing but Carnap's measure C_{car} and 0 is its threshold value. It therefore satisfies D_1. Since C'_{sho} and C'_{car} take into account the degrees of coherence of all subsets of a target set X they trivially satisfy D_1, too.

Proof (*Observation* 2) Cf. Schippers (2014) for a proof that there can be no coherence measure C which satisfies both D_1 and D_2. Since by observation 1 C_{sho}, C_{car}, C'_{sho} and C'_{car} satisfy D_1, they cannot satisfy D_2.

Proof (*Observation* 3) Consider the following probability distribution over the set $X = \{x_1, x_2\}$ of logically equivalent propositions as counterexample: $P(x_1) = P(x_2) = P(x_1 \wedge x_2) = 1$ and hence $C_{sho} = 1$ and $C'_{sho} = C_{car}(X) = C'_{car}(X) = 0$ which is not the maximum value of any of the measures. Hence, the measures do not satisfy D_3.

Proof (*Observation* 4) Consider the following probability distribution over the set $X = \{x_1, x_2\}$ of inconsistent propositions as counterexample: $P(x_1 \wedge x_2) = P(x_1) = P(x_2) = 0$. For $C_{sho}(X)$ and $C'_{sho}(X)$ the function values are not defined while $C_{car}(X) = C'_{car}(X) = 0$. Hence, none of the measures yields its minimum value and therefore none satisfies D_4.

Proof (*Observation* 5) For C_{sho} and C_{car} cf. Shogenji (2013). Since D_5 is formulated for 2 propositions, C'_{sho} and C'_{car} trivially satisfy D_5, too.

Proof (*Observation* 6) Meijs' (2005) test case is a counterexample against the ordinal equivalence of the pairs (C_{sho},C_{car}), (C_{sho},C'_{car}), (C_{sho},C_{car}) and (C'_{sho},C'_{car}). Schupbach's (2011) test case is a counterexample against the ordinal equivalence of the pairs (C_{sho}, C'_{sho}) and (C_{car}, C'_{car}).

References

Akiba, K. (2000). Shogenji's probabilistic measure of coherence is incoherent. *Analysis*, *60*, 356–359.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.

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

Bovens, L., & Olsson, E. J. (2000). Coherentism, reliability and Bayesian networks. *Mind*, *109*(436), 685–719.

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

Crupi, V., Tentori, K., & Gonzales, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, *74*, 229–252.

Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese*, *156*, 405–425.

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

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

Glass, D. H. (2002). Coherence, explanation, and Bayesian networks. In M. O'Neill, R. F. E. Sutcliffe, C. Ryan, M. Eaton, & N. J. L. Griffith (Eds.), *Artificial intelligence and cognitive science. 13th Irish conference, AICS 2002, Limerick, Ireland, September 2002* (pp. 177–182). Berlin: Springer.

Glass, D. H. (2005). Problems with priors in probabilistic measures of coherence. *Erkenntnis*, *63*, 375–385.

Hammersley, J. M., & Handscomb, D. C. (1964). *Monte Carlo methods*. London: Methuen & Co Ltd.

Klein, P., & Warfield, T. A. (1994). What price coherence? *Analysis*, *54*(3), 129–132.

Kolmogorov, A. (1956). *Foundations of the theory of probability*. New York: AMS Chelsea Publishing.

Koscholke, J. (2015). Evaluating test cases for probabilistic measures of coherence. *Forthcoming in Erkenntnis*, pp. 1–29. doi:10.1007/s10670-015-9734-1.

Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, *8*(1), 3–30.

Meijs, W. (2005). *Probabilistic measures of coherence*. PhD thesis, Erasmus University, Rotterdam.
Meijs, W., & Douven, I. (2005). Bovens and Hartmann on coherence. *Mind*, *114*, 355–363.
Olsson, E. J. (2002). What is the problem of coherence and truth? *The Journal of Philosophy*, *94*, 246–272.
Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions. Royal Society of London. Series A. Mathematical and Physical Sciences*, *186*, 343–414.
Roche, W. (2013). Coherence and probability: A probabilistic account of coherence. In M. Araszkiewicz & J. Savelka (Eds.), *Coherence: Insights from philosophy, jurisprudence and artificial intelligence* (pp. 59–91). Dordrecht: Springer.
Schippers, M. (2014). Probabilistic measures of coherence: From adequacy constraints towards pluralism. *Synthese*, *191*(16), 3821–3845.
Schippers, M., & Siebel, M. (2015). Inconsistency as a touchstone for coherence measures. *Theoria*, *30*, 11–41.
Schupbach, J. N. (2011). New hope for Shogenji's coherence measure. *British Journal for the Philosophy of Science*, *62*(1), 125–142.
Shogenji, T. (1999). Is coherence truth conducive? *Analysis*, *59*, 338–345.
Shogenji, T. (2013). Reductio, coherence, and the myth of epistemic circularity. In F. Zenker (Ed.), *Bayesian argumentation* (pp. 165–184). Berlin: Springer.
Siebel, M. (2004). On Fitelson's measure of coherence. *Analysis*, *64*, 189–190.
Siebel, M., & Wolff, W. (2008). Equivalent testimonies as a touchstone of coherence measures. *Synthese*, *161*, 167–182.
Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101.
Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007). Comparison of confirmation measures. *Cognition*, *103*, 107–119.
Wheeler, G. (2009). Focused correlation and confirmation. *British Journal for the Philosophy of Science*, *60*(1), 79–100.