Structured probabilistic coherence and the usual counterexamples to probabilistic measures of coherence

Contents

1 Introduction & motivations

The notion of coherence is often used in many philosophical, especially epistemological, discussions (for instance, in discussions about the truth-conduciveness of coherence). When we talk about the coherence of a set of propositions or about the coherence of a story, we seem to refer to how well their individual pieces fit together. How are we to understand and apply this notion systematically, though? In particular, we will be interested in probabilistic explications of this notion, as Bayesian epistemology strives to be a general epistemological project and as such it should be able to accommodate coherence-oriented considerations.

There is also a more practical reason to develop a better understanding of the notion: a plausible measure of coherence could be used to better evaluate the quality of some stories or narrations. For example in the legal context we would like to be able to assess the quality of a testimony in the court of law.

REFS

Multiple probabilistic explications of coherence have been proposed (Douven & Meijs, 2007; Fitelson, 2003; Glass, 2002; Meijs & Douven, 2007; Olsson, 2001; Roche, 2013; Shogenji, 1999). However, clear general principles to choose between them are hard to come by. One paper where some such principles have been formulated is (Schippers, 2014), where a list of seemingly plausible adequacy conditions for a coherence measure is proposed and shown to be inconsistent to argue for pluralism about the notion of coherence. However, some of those requirements are quite non-trivial.[1]

The general point here is not that the approach taken in (Schippers, 2014) is flawed, but rather that the task of formulating general principles for coherence is a challenge, and that no clear list of such uncontroversial desiderata is on the horizon.

One approach to obtaining some clarity on which abstract conditions are plausible is looking at various thought experiments in which our intuitions about what the coherence scores should be (at least comparatively) are more robust than direct assessment of general requirements. In fact, looking at examples is what the main stream of literature on probabilistic coherence focused on, and each probabilistic measure of coherence faces a selection of seemingly intuitive counterexamples.

We decided to work with this methodology. We first gathered key examples that occur in the literature, represented them in terms of Bayesian networks, and developed **R** scripts calculating all coherence scores for the Bayesian networks at play, pushing further the results obtained by Koscholke (2016).[2] Then we reflected on the results, noticing that one weakness of the measures is that they pay little attention to the underlying structure of a given narration in the calculation of its coherence.

Inspired by this observation, we formulate our own proposal, which diverges from the known purely probabilistic measures of coherence in three important respects: (i) It is not a function of a probability measure and a set of propositions alone, because it is also sensitive to the selection and direction of edges in a Bayesian network representing an agent's credal state. (ii) Unlike in the case of quite a few coherence measures, it is sensitive to the weakest links in the narration. (iii) It is not obtained by simply averaging confirmation levels between all possible combinations of elements.

We described this approach in a more detailed introduction to this measure [ANONYMIZED], which explains the method and some of the theoretical decisions that we have made, and show how it works using a Bayesian network developed for the well-known Sally Clark case (Fenton & Neil, 2018). The goal of the current paper is to discuss a range of philosophical counterexamples to the existing probabilistic measures of coherence and evaluate the performance of our approach using those as a benchmark, arguing that it performs much better than the existing ones.

Accordingly, in Section 2 we introduce all the coherence measures, including the key motivations for

[1] Let us illustrate this. The (Dependence) condition formulated there requires that the coherence score of a set of propositions is above (below) the neutral score if for all pairs of non-empty subsets the posterior of an element of a pair conditional on the other element is higher than the prior of the former. This makes some of the features of the coherence measure dependent on the priors, and whether it should be so is not obvious. On the other hand, (Agreement) is formulated in terms of conditional probabilities between such pairs. If on a given measure P all conditional probabilities (between pairs already mentioned) are higher than on P', the coherence of a set given P should be higher than given P'. The (Equivalence) requirement is that any finite set of logically equivalent propositions should be maximally coherent. This is suspicious, as the set $\{0 = 1, 2, = 5\}$ is a set of equivalent propositions (with sufficiently strong notion of logical equivalence in the background), but we would intuitively hesitate to say it's maximally coherent.

[2] The whole work has been made possible by all those who contributed to the development of **R** language, and Marco Scutari, the author of **bnlearn** package, who was kind enough to extend his package with additional features upon our requests (Scutari & Denis, 2015).

and a pseudo-code description of our measure. In Section 3 we describe the thought experiments meant as counterexamples to coherence measures, their corresponding desiderata and their status on various coherence measures, including ours. The order of the discussion of any given example is straightforward: we first explain what the situation we are to consider is, what the intuitive desiderata related to it are supposed to be, how the situation is represented by means of a Bayesian network(s), and what happens when we apply all coherence measures. We end with Section 4 in which we compare all of the results and draw some general conclusions.

2 Probabilistic coherence measures and structured coherence

Quite a few different measures of coherence have been proposed in the literature. Two early proposals are:

- Shogenji's **deviation from independence** (Shogenji, 1999), is defined as the ratio between the probability of the conjunction of all claims, and the probability that the conjunction would get if all its conjuncts were probabilistically independent (scaling from 0 to ∞ with neutral point 1):

$$\mathscr{C}_S(S) = \frac{P(\bigwedge S)}{\prod_{i=1}^{|S|}\{P(S_i)|i \in S\}} \qquad \text{(Shogenji)}$$

This measure was later generalized by Meijs & Douven (2007). According to this approach, (Shogenji) is applied not only to the whole set of propositions, but to each non-empty non-singleton subset of the set, and the final value is defined as the average of all sub-values thus obtained.

- **Relative overlap** (Glass, 2002; Olsson, 2001), is defined as the ratio between the intersection of all propositions and their union (scaling from -1 to 1 with no clear neutral point):

$$\mathscr{C}_O(S) = \frac{P(\bigwedge S)}{P(\bigvee S)} \qquad \text{(Olsson)}$$

It has also been generalized in a way analogous to the one used in the generalization of the Shogenji's measure (Meijs & Douven, 2007).

Both of these approaches are susceptible to various objections and counterexamples (Akiba, 2000; Bovens & Hartmann, 2004; Crupi, Tentori, & Gonzalez, 2007; Koscholke, 2016; Merricks, 1995; Schippers & Koscholke, 2019; Shogenji, 1999, 2001, 2006; Siebel, 2004, 2006). To overcome them, more recent works proposed **average mutual support** measures, starting with (Fitelson, 2003). The general recipe for such measures is as follows.

- Given that S is a set whose coherence is to be measured, let P indicate the set of all ordered pairs of non-empty, disjoint subsets of S.
- First, define a confirmation function (of a hypothesis H by evidence E): $\mathsf{conf}(H,E)$.
- For each pair $\langle X,Y\rangle \in P$, calculate $\mathsf{conf}(\bigwedge X, \bigwedge Y)$, where $\bigwedge X$ is the conjunction of all the elements of X (and $\bigwedge Y$ is to be understood analogously).
- Take the mean of all the results:

$$\mathscr{C}(S) = \mathsf{mean}\left(\left\{\mathsf{conf}(\bigwedge X_i, \bigwedge Y_i)|\langle X_i, Y_i\rangle \in P\right\}\right).$$

Different measures of coherence result from different choices of a confirmation measure. Here are the key candidates present in the literature:

- Fitelson (2003) uses the following confirmation function (the resulting coherence measure ranges from -1 to 1 with neutral point at 0):

$$F(H,E) = \begin{cases} 1 & E \models H, E \not\models \bot \\ -1 & E \models \neg H \\ \frac{P(E|H)-P(E|\neg H)}{P(E|H)+P(E|\neg H)} & \text{o/w} \end{cases}$$

$$\mathscr{C}_F(S) = \mathsf{mean}\left(\left\{F(\bigwedge X_i, \bigwedge Y_i)|\langle X_i, Y_i\rangle \in P\right\}\right) \qquad \text{(Fitelson)}$$

- Douven & Meijs (2007) use the difference confirmation measure (with coherence ranging from -1 to 1 with neutral point at 0):

$$D(H,E) = P(H|E) - P(H)$$
$$\mathscr{C}_{DM}(S) = \mathsf{mean}\left(\left\{D(\bigwedge X_i, \bigwedge Y_i) | \langle X_i, Y_i\rangle \in P\right\}\right) \tag{DM}$$

- Roche (2013) uses the absolute confirmation measure (the resulting coherence measure ranges from 0 to 1 with neutral point at 0.5):

$$A(H,E) = \begin{cases} 1 & E \models H, E \not\models \bot \\ 0 & E \models \neg H \\ P(H|E) & \text{o/w} \end{cases}$$
$$\mathscr{C}_{R}(S) = \mathsf{mean}\left(\left\{A(\bigwedge X_i, \bigwedge Y_i) | \langle X_i, Y_i\rangle \in P\right\}\right) \tag{Roche}$$

Mind your head: different measures use different scales and have different neutral points (values taken for any set of probabilistically independent propositions; not all measures have neutral points). This is worth keeping in mind when it comes to various desiderata that we will discuss.

As we already mentioned in the introduction, formulating abstract formal requirements for a coherence measure and investigating whether a given coherence measure satisfies them has not resulted in an agreement. For this reason, we follow another path, which has dominated the literature on the topic. We look at how the measures behave in test scenarios. Many putative scenarios were put forward as counterexamples. They usually have the form of a few propositions formulated in natural language, such that intuitive judgments of coherence involved and the formal coherence calculations seem to diverge (Akiba, 2000; Bovens & Hartmann, 2004; Koscholke, 2016; Meijs & Douven, 2007, 2007; Merricks, 1995; Schippers & Koscholke, 2019, 2019; Shogenji, 1999, 2001, 2006; Siebel, 2004, 2006). We will focus on these examples in what follows. To spoil the experience, let us already point out that the probabilistic measures we introduced above do not seem to handle these examples very well (read on for details).

Inspired by these failures, in [REFERENCE SUPRESSED FOR ANONYMITY] we proposed to take a different perspective. Putting the earliest measures aside (they were problematic for various reasons), we noticed that the problems with the average mutual support measures stem from the fact that the coherence score is an average confirmation score for all possible combinations of the parts of a narration. Therefore we proposed to take a more fine-grained account. First, we represented an agent's credal state by means of a Bayesian network, which comprises not only a probabilistic measure but additional structural information. Then we used this structural information in our definition of coherence, so that only those directions of support are considered which in fact are indicated by the structure of the agent's belief state.

While we refer the reader to a more extensive treatment in [REFERENCE SUPRESSED FOR ANONYMITY], we now briefly discuss the main idea behind it. A Bayesian network (BN) represents agent's probabilistic belief state with respect to the relevant nodes. Additionally, some of the nodes are distinguished as narration nodes—the agent holds definite beliefs about which states of these nodes occur, without assigning probabilities equal 1 to them.[3] If actual learning takes place, the evidence nodes and their states need to be specified as well.

Each parented node in the BN receives its expected confirmation score (ECS). It is calculated by looking at all combinations of its states and states of its parents not excluded by agents' binary beliefs.[4] For each of these combinations, the confirmation score between the parents' states and the child state is calculated (in the pseudo-code, we use confirmation measure Z, in further calculations we also use measures LR and L for comparison)[5]. Then, a weighted average of these scores is obtained. The weights are the normalized probabilities of the combinations of parents' states obtained in the BN (updated with the evidence, if it has been specified). If both the child and the parent nodes belong to the narration, there is only one possible combination so after normalization it gets weight 1. The final coherence score is either the mean of the ecs scores for all the child nodes, if all of them are positive, or it is a weighted average of their mean and their minimum, $(1 - |\mathsf{min}(\mathsf{ecs})|) \times \mathsf{mean}(\mathsf{ecs}) + |\mathsf{min}(\mathsf{ecs})| \times$

[3] Also, we assume that the agent does not assign probability 0 to the narration being true.

[4] Conceptually, it is possible to not restrict ourselves this way and look at all combinations of states of the nodes which are assigned non-null probabilities. But then, effectively, the binary content of the narration would have no impact on the coherence score, and we would obtain a coherence measure for the purely probabilistic part of agents' convictions. While this might be a worthy enterprise, we do not purse this idea in this paper. Calculations of a measure thus modified can be achieved by fairly straighforward modification of our code.

[5] DEFINITIONS

$\min(\mathsf{ecs})$, otherwise.[6]

[6] We have developed **R** code calculating this and other measures to handle calculations that will be discussed further on, the code with documentation is available at ANONYMIZED.

```
FUNCTION parents_child_possible_states(parents,child)
    IF child included in narration THEN
        consequentStates <- the unique state of child as reported in the narration
    ELSE
        consequentStates <- all possible states of child
    FOR EACH parent in parents
        IF parent included in narration THEN
            parentStates[parent] <- the unique state of parent as reported in the narration
        ELSE
            parentStates[parent] <- all possible states of parent
    parentsStates <- all combinations of parentStates
    variants <- list of all possible combinations of consequentStates and parentsStates
    RETURN variants


FUNCTION coherence_from_ecs(ecs)
    IF min(ecs) <= 0 THEN
        RETURN mean(ecs) * (min(ecs)+1) - min(ecs)min(ecs)
        #this is equivalent to (1- |min(ecs)|) * mean(ecs) + |min(ecs)| * min(ecs)
    ELSE
        RETURN mean(ecs)

FUNCTION structured_coherence(BN,evidenceNodes,evidenceStates)
    parentedNodes <- vector of non-root nodes in BN
    ecsList <- []
    FOR EACH parentedNode IN parentedNodes
        variants <- parents_child_possible_states(parents,parentedNode)
        variants_count <- length(variants)
        sumParentsNarr <- 0
        FOR EACH variant IN variants
            childPrior <- prior probability of the child state in variant
            childPosterior <- posterior probability of the child state in variant,
                        obtained by updating on the parents states
                        from this variant
            parentsEvidence[variant] <- joint probability of
                        the parents states in variant in BN updated with evidenceStates of evidenceNodes
            sumParentsEvidence <- sumParentsEvidence +  parentsEvidence[variant]
            z[variant] <- z_confirmation_measure(childPrior, childPosterior)
        ecs <- 0  #expected confirmation score
        FOR EACH variant IN variants
            IF parentsEvidence[variant] > 0 THEN
                weight <- parentsEvidence[variant]/sumParentsEvidence
            ELSE
                weight <- 1/variants_count
            zScaled <- z[variant] * weight
            ecs <- ecs + zScaled
        ecsList.add(ecs)
    RETURN coherence_from_ecs(ecsList)
```

Having introduced the coherence measures at play, let us now move to the key counterexamples discussed in the literature.

3 Challenges and their treatment

We will now go through a list of key counterexamples proposed in the literature, each time explaining the relevant desiderata. We represent those scenarios as Bayesian networks. Then we calculate the coherence scores for those scenarios using all the measures we have introduced. Finally, we test whether the desiderata are satisfied.

Here are the counterexamples put forward against various coherence measures in the literature. We ignored only a few where both we didn't share the authors' intuitions and the examples were not picked up in further discussion in theliterature.[7]

REFS

[7] One such an example, involves Sarah and her pregnancy (Tomoji Shogenji, 2006), but it focused more on truth-conduciveness of coherence, which is beyond the scope of our paper. We also do not discuss a few other examples involving fossils and voltage (T. Shogenji, 2001; Mark Siebel, 2006). In some respects, they were quite similar to the dice and depth problems that we do discuss, and some of their variants simply did not inspire our agreement. For instance, Siebel thinks that for voltage levels $\{V = 1, V = 2\}$ is more coherent than $\{V = 1, V = 50\}$, while we think that both sets are maximally incoherent (there

3.1 Penguins

The scenario. This is a challenge to the Olsson-Glass measure discussed in (Bovens & Hartmann, 2004, p. 50) and (Meijs & Douven, 2007). It consists of the propositions (instead of *letters* or *abbreviations*, we'll talk about *nodes*, as these will be used later on in Bayesian networks) displayed in Table 1.

node	content
B	Tweety is a bird.
G	Tweety is a grounded animal.
P	Tweety is a penguin.

Table 1: Propositions in the Penguins scenario

Desiderata. Meijs & Douven (2007) claim that the set {B,G}, which doesn't contain the information about Tweety being a penguin, should be less coherent than the one that does contain this information: {B,G,P}.

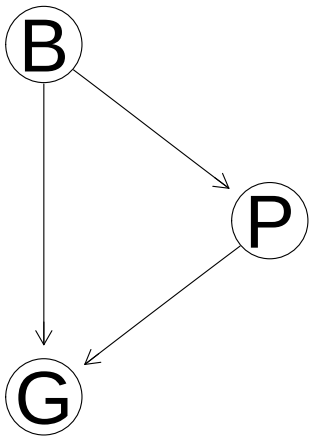
(BG$<$BGP) {B,G} should be less coherent than {B,G,P}.

Another intuition about this scenario (Schippers & Koscholke, 2019) is that when you consider a set which says that Tweety is both a bird and a penguin: {B,P}, adding proposition about not flying (G) shouldn't increase the coherence of the set as much as moving from {B,G} to {B,G, P}. It's a well-known fact that penguins don't fly and by adding G explicitly to the set, one wouldn't gain as much information. However, as G is not a logical consequence of P, it can be argued that {B,P} and {B,P,G} represent different information sets, and so some difference in their coherence is to be expected.

(BG $\ll$ BP$\leq$BGP) {B,P} should be notably above {B,G}, and less than {B,P,G}.

Formally, we'll require that the absolute difference between BG and BP be greater than .1 (the exact placement of the threshold doesn't make a huge difference, unless it's at an unintuitive value below .01) and that {B,G} $\leq$ {B,P,G}.

Bayesian network. We used the distribution used in the original formulation (Meijs & Douven, 2007) to build a BN corresponding to the narrations at play (Fig. 1).[8]



B	Pr
1	0.5
0	0.5

P	B	
	1	0
1	0.02	0
0	0.98	1

	B	P	
G			Pr
1	1	1	1.00
0	1	1	0.00
1	0	1	0.00
0	0	1	1.00
1	1	0	0.00
0	1	0	1.00
1	0	0	0.98
0	0	0	0.02

Figure 1: Bayesian network for the Penguins problem.

might be some claims in the vicinity that are not incoherent, say, focusing on results of separate measurements, but an example along these lines has not been properly formulated in the literature).

[8]Not without concerns. There are around 18 000 species of birds, and around 60 of them are flightless. We couldn't find information about counts, but it seems the probability of being a penguin if one is grounded is overestimated by philosophers. Also, there are many things that are not grounded but are not birds, mostly insects, and there's plenty of them. We did spend some time coming up with plausible ranges of probabilities to correct for such factors, and none of them actually makes a difference to the main point. So, for the sake of simplicity, we leave the original unrealistic distribution in our discussion.

Double check satisfaction for LR

Results. Now, let's calculate the coherence scores (Table 2) and see if the desiderata are satisfied (Table 3). The measures are: Olsson-Glass, generalized Olsson-Glass, Shogenji, generalized Shogenji, Douven-Meijs, Roche, Fitelson, Structured with Z, LR, and L used as a confirmation measure.

	OG	OGGen	Sh	ShGen	DM	R	Fi	SZ	SLR	SL
Penguins: BGP 111	0.01	0.015	4.00	2.01	0.255	0.51	0.453	0.505	Inf	0.669
Penguins: BG 11	0.01	0.010	0.04	0.04	-0.480	0.02	-0.960	-0.960	0.02	-0.960
Penguins: BP 11	0.02	0.020	2.00	2.00	0.255	0.51	0.669	0.010	2.02	0.338

Table 2: Coherence scores for the Penguins scenario (rounded). Note how LR might result in Inf if a conditional probability of 1 at an arrow used in the calculations is involved.

	OG	OGGen	Sh	ShGen	DM	R	Fi	SZ	SLR	SL
Penguins: BG<BGP	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Penguins: BG<< BP< BGP	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Table 3: Desiderata satisfaction for the Penguins scenario.

3.2 Dice

The scenario. This scenario was offered by Schippers & Koscholke (2019). You're either tossing a regular die, or a dodecahedron, X is the result (there is nothing particular about this choice of dice; *mutatis mutandis* this should hold for other possible pairs of dice as well). Consider the coherence of:

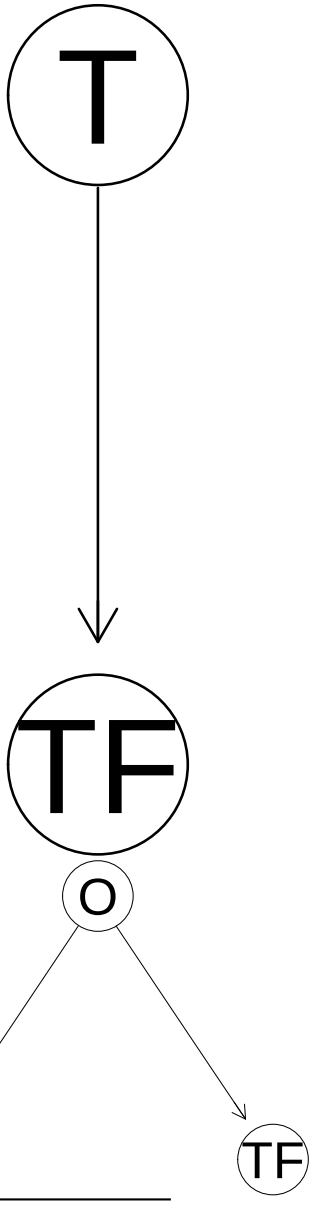
$$D = \{X = 2, (X = 2 \vee X = 4)\}.$$

Desiderata. In this scenario posterior conditional probabilities are fixed: getting 2 or 4 logically follows from getting 2 ($P(X = 2 \vee X = 4 | X = 2) = 1$), and you always have 50% chance to get 2 given that the outcome was 2 or 4 ($P(X = 2 | X = 2 \vee X = 4) = 0.5$). Therefore, according to Schippers & Koscholke (2019), the coherence of the set D shouldn't change no matter which die you use.

(D=const) the coherence of D should not change.

tabela z nodami

Bayesian networks. We'll follow the strategy similar to the one we already used. Since neither the example nor the narrations involve information about how probable it is that we're dealing with a regular die, as opposed to a dodecahedron, we avoid using a node representing this. Moreover, if at a given time the agent claims that the result is both two and (two or four), their cognitive situation at that time cannot be represented using uniform distribution for possible toss outcomes. Instead, we start with initial separate BNs for a regular die and a dodecahedron which do have uniform distributions for the O (outcome) node (Fig. 2), but when weighing the antecedent nodes which are not strictly speaking part of the narration, we use the probabilities updated in light of the narration content itself.



O	Pr
1	0.167
2	0.167
3	0.167
4	0.167
5	0.167
6	0.167

(a) Root CPT for the regular die.

O	Pr
1	0.083
2	0.083
3	0.083
4	0.083
5	0.083
6	0.083
7	0.083
8	0.083
9	0.083
10	0.083
11	0.083
12	0.083

(b) Root CPT for the dodecahedron.

T	O					
	1	2	3	4	5	6
1	0	1	0	0	0	0
0	1	0	1	1	1	1

TF	O					
	1	2	3	4	5	6
1	0	1	0	1	0	0
0	1	0	1	0	1	1

(c) Conditional probabilities for the regular die.

T	O											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0	1	0	0	0	0	0	0	0	0	0	0
0	1	0	1	1	1	1	1	1	1	1	1	1

TF	O											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0	1	0	1	0	0	0	0	0	0	0	0
0	1	0	1	0	1	1	1	1	1	1	1	1

(d) Conditional probabilities for the dodecahedron.

Figure 2: BNs for the dice problem.

Results. Calculation of coherence scores of the scenarios in the respective BNs yield the result in Table **??**, and the status of the desiderata is pictured in Table **??**.

make sure Dice precede robbers

3.3 Dunnit

The scenario. Another challenge, introduced by Merricks (1995) goes as follows: Mr. Dunnit is a suspect in the murder case. Detectives first obtained the body of evidence specified in Table 4.

node	content
I	Witnesses claim to have seen Dunnit do it (incriminating testimony).
M	Dunnit had a motive for the murder.
W	A credible witness claims to have seen Dunnit two hundred miles from the scene of the crime at the time of the murder.

Table 4: Initial evidence in the Dunnit scenario.

In light of this information they try to assess whether Dunnit is responsible for the crime (Table 5).

node	content
G	Dunnit is guilty.

Table 5: The guilt statement in the Dunnit scenario.

Now, suppose the detectives learn Dunnit has a twin brother (Table 6).

node	content
Tw	Dunnit has an identical twin which was seen by the credible witness two hundred miles from the scene of the crime during the murder.

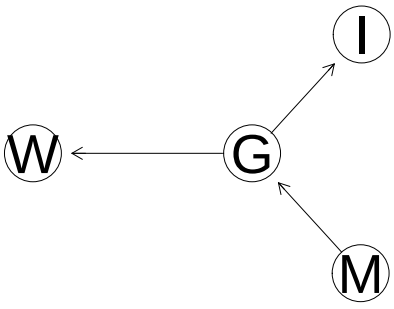
Table 6: New evidence in the Dunnit scenario.

What are our intuitions when we compare the coherence of $\{\mathsf{I,M,W,G}\}$ with the coherence of $\{\mathsf{I,M,W,G,Tw}\}$?

Desideratum. It seems that adding proposition about a twin should increase the coherence of the set.

(Dunnit<Twin) $\{\mathsf{I,M,W,G}\}$ should be less coherent than $\{\mathsf{I,M,W,G,Tw}\}$.

Bayesian networks. Here, we deal with two separate BNs. One, before the Twin node is even considered (Figure 3), and one with the Twin node (Figure 4). The CPTs for the no-twin version are in agreement with those in the ones in the Twin case. Since Merricks (1995) did not specify probabilities in the original example (and its further discussion focused on the general relationship between coherence and the addition of propositions), we came up with plausible values.



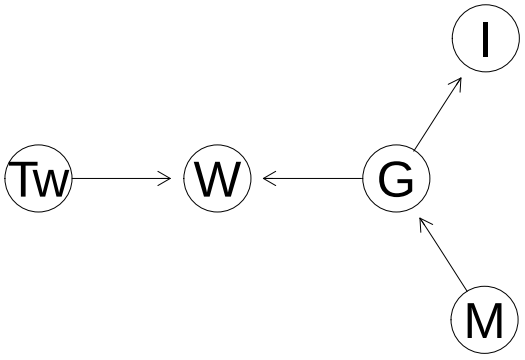
M	Pr
1	0.4
0	0.6

G	M	
	1	0
1	0.05	0.005
0	0.95	0.995

I	G	
	1	0
1	0.8	0.005
0	0.2	0.995

W	G	
	1	0
1	0.012	0.207
0	0.988	0.793

Figure 3: Twin-less BN for the Dunnit problem.



W	G	Tw	Pr
1	1	1	0.200
0	1	1	0.800
1	0	1	0.400
0	0	1	0.600
1	1	0	0.005
0	1	0	0.995
1	0	0	0.200
0	0	0	0.800

Figure 4: BN for the Dunnit problem. The key difference for the twin version lies in the construction of the CPT for W. The table gives conditional probabilities for W given various joint states of Tw and G.

Add table for TW

3.4 Japanese swords

The scenario. The next challenge comes from (Meijs & Douven, 2007, p. 414):

> We start by considering two situations in both of which it is assumed that a murder has been committed in a street in a big city with 10,000,000 inhabitants, 1,059 of them being Japanese, 1,059 of them owning Samurai swords, and 9 of them both being Japanese and owning Samurai swords. In situation I we assume that the murderer lives in the city and that everyone living in the city is equally likely to be the murderer. In situation II, on the other hand, we make the assumption that the victim was murdered by someone living in the street in which her body was found. In that street live 100 persons, 10 of them being Japanese, 10 owning a Samurai sword, and 9 both being Japanese and owning a Samurai sword. [...] [In situation III] we have 12 suspects who all live in the same house, and 10 of them are Japanese, 10 own a Samurai sword, and 9 are both Japanese and Samurai sword owners.

The nodes involved are as in Table 7.

node	content
J	The murderer is Japanese.
O	The murderer owns a Samurai sword.

Table 7: Nodes in the Japanese swords scenario.

Now, we look at three separate scenarios: (1) The murderer lives in the city, (2) The murderer lives in the street popular among Japanese owners of Samurai swords, and (3) The murderer lives in the house with many Japanese owners of Samurai swords.

Desiderata. In all of the above situations the number of Japanese owners of Samurai swords remains the same. However, situations 1 and 2 differ in the relative overlap of J and O. Because J and O are more correlated in situation 2, it seems more coherent than situation 1.

(JO2>JO1) {J,O,2} should be more coherent than {J,O,1}.

However, bigger overlap doesn't have to indicate higher coherence. In situation 3 J and O confirm each other to a lesser extent than in situation 2 (compare $P(J|O) - P(J)$ and $P(O|J) - P(O)$ in both cases), and for this reason Douven and Meijs claim that situation 2 is more coherent than situation 3.

(JO2>JO3) {J,O,2} should be more coherent than {J,O,3}.

Bayesian networks. There is a common DAG for the three scenarios, but the CPTs differ (Figure 5).

J	Pr
1	0
0	1

O	J	
	1	0
1	0.008	0
0	0.992	1

(a) Scenario 1.

J	Pr
1	0.1
0	0.9

O	J	
	1	0
1	0.9	0.011
0	0.1	0.989

(b) Scenario 2.

J	Pr
1	0.833
0	0.167

O	J	
	1	0
1	0.9	0.5
0	0.1	0.5

(c) Scenario 3.

Figure 5: A common DAG and three sets of CPTs for the Japanese Swords problem.

–> –> –>
–> –>
–> –> –>

3.5 Robbers

The scenario. A challenge put forward by Siebel (2004, p. 336) goes as follows:

> Let there be ten equiprobable suspects for a murder. All of them previously committed at least one crime, two a robbery, two pick-pocketing, and the remaining six both crimes. There is thus a substantial overlap: of the total of eight suspects who committed a robbery, six were also involved in pick-pocketing, and conversely.

The nodes involved are Table 8.

node	content
W	Real perpetrator status (states: OnlyP, OnlyR, Both).
P	The murderer is a pickpocket.
R	The murderer is a robber.

Table 8: Nodes in the Robbers scenario.

Desiderata. The first observation is that the set of propositions that corresponds to the situation in which a murderer committed both crimes should be regarded coherent. Most suspects committed both crimes, so this option is even the most probable one.

(PR>neutral) {P,R} should be regarded coherent.

According to Siebel (2004, p. 336) committing both crimes by the murderer should also be regarded more coherent than committing only one crime.

(PR>P¬R) {P,R} should be more coherent than {P,¬R} and {¬P,R}.

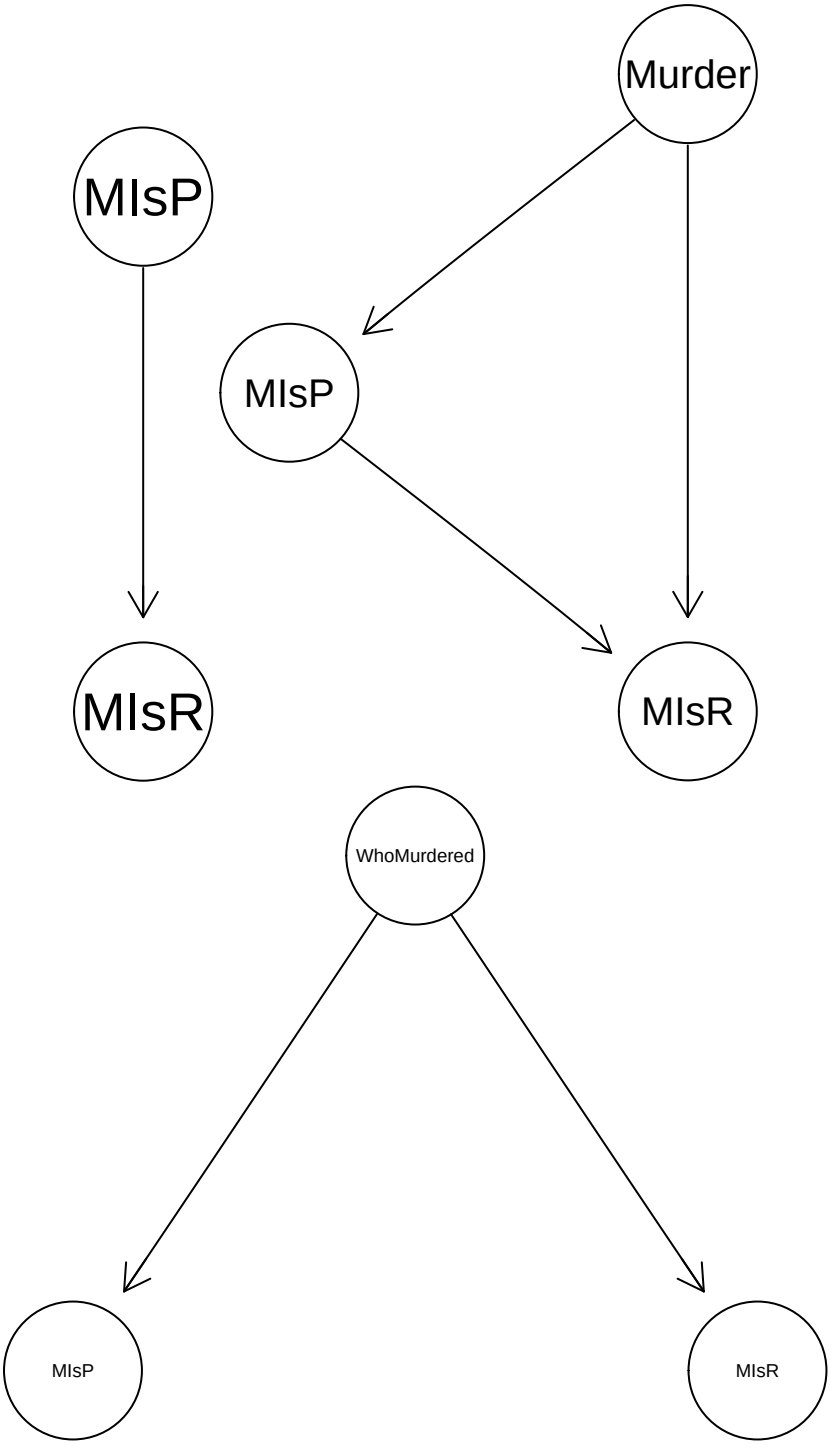
This requirement is slightly more controversial. Even though {P,R} is the most probable setup, P and R disconfirm each other ($Pr(P|R) < Pr(P)$ and $Pr(R|P) < Pr(R)$). Moreover, the intuition behind this desideratum seems to conflict with the intuition behind (JO2>JO3).

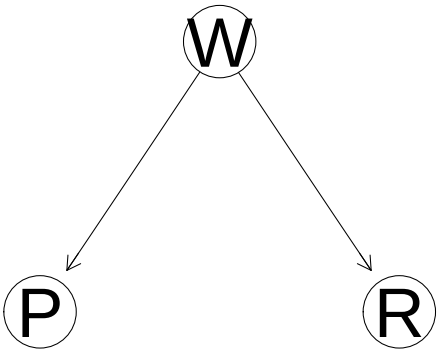
Bayesian networks.

The robbers counterexample illustrates the point we already discussed when we talked about the Dice example: a purely probabilistic assumption may or may not be explicitly included as part of the narration. Its explicit inclusion, achieved by the addition of a node in a BN can play a role in coherence calculations. Here, such an inclusion is in line with the example as it was proposed in the literature, where the underlying prior probabilities were explicitly listed as part of the story.

the information about the prior probabilities is supposed to be part of the narration or not. If we want to include this information in our coherence assessment, we can do this employing a single BN.

is this comment still needed?


Murder
MIsP
MIsP
MIsR
MIsR
WhoMurdered
MIsP
MIsR



	Pr
OnlyP	0.2
OnlyR	0.2
Both	0.6

MisP	WhoMurdered		
	OnlyP	OnlyR	Both
1	1	0	1
0	0	1	0

MisR	WhoMurdered		
	OnlyP	OnlyR	Both
1	0	1	1
0	1	0	0

Figure 6: BN for the Robbers problem.

Results. Coherence calculations yield the results in Table **??**, and the performance of the coherence measures with respect to the desiderata is illustrated in Table **??**.

–>

–> –> –>

–> –>

–>

3.6 The Beatles

Consider moving to top of the list.

The scenario. The challenge has been offered by Shogenji (1999, p. 339) to criticize defining coherence in terms of pairwise coherence — it shows there are jointly incoherent pairwise coherent sets. The scenario consists of the claims displayed in Table 9.

node	content
D	Exactly one of the Beatles (John, Paul, George and Ringo) is dead.
J	John is alive.
P	Paul is alive.
G	George is alive.
R	Ringo is alive.

Table 9: Nodes in the Beatles scenario.

REF

Desiderata. The set consisting of all of these propositions is logically inconsistent (even though the propositions are pairwise consistent), so it seems quite intuitive that it should be incoherent.[9]

(below neutral) {D,J,P,G,R} should be incoherent.

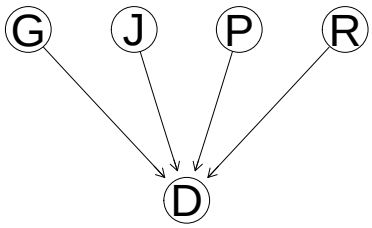
We can make this desideratum a bit stronger by requiring that the coherence score for {D,J,P,G,R} should be minimal.

(minimal) {D,J,P,G,R} should get the lowest possible coherence value.

For the sake of example, we assume the prior probability of each individual band member being dead to 0.5 (as in the above table), and the CPT for D is many-dimensional and so difficult to present

[9]One may argue that some coherence measures also measure the degree of incoherence, therefore logically inconsistent sets don't need to get the minimal score. We do not focus on such an understanding of coherence in this paper. If you think different inconsistent scenarios can differ in coherence—in line with ACCURACY, COHERENCE AND EVIDENCE—our measure can accommodate this move by revising the calculations of coherence based on the ecs scores (for example, the penalty for the weakest link can be lowered, or dropped).

concisely, but the method is straightforward: probability 1 is given to D in all combinations of the parents in which exactly one is true, and otherwise D gets conditional probability 0.



G	Pr
1	0.5
0	0.5

Figure 7: Bayesian network for the Beatles scenario.

–> –> –>
–> –>
–> –> –>

3.7 The Witnesses

The scenario. This counterexample comes from (Olsson, 2005, p. 391). Equally reliable witnesses try to identify a criminal. Consider the reports listed in Table 10 (we extended the original scenario by adding W5).

node	content
W1	Witness no. 1: "Steve did it"
W2	Witness no. 2: "Steve did it"
W3	Witness no. 3: "Steve, Martin or David did it"
W4	Witness no. 4: "Steve, John or James did it"
W5	Wittness no. 5: "Steve, John or Peter did it"
D	Who committed the deed (6 possible values)

Table 10: Testimonies in the Witnesses scenario.

Note that this time each proposition has the structure "Witness no. X claims that ..." instead of explicitly stating the witness' testimony.

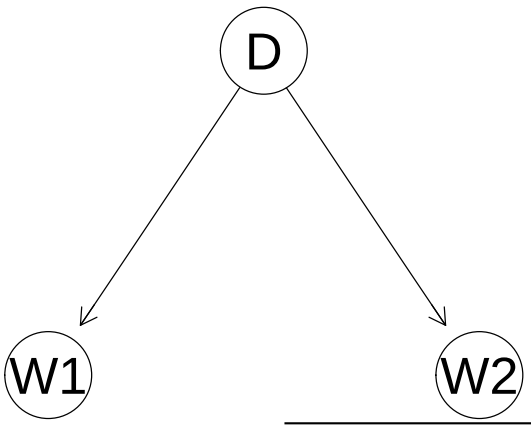
Desiderata. First, we can observe that W1 and W2 fully agree. Testimonies of W3 and W4 overlap only partially, therefore it seems that {W1,W2} is more coherent than {W3,W4}.

(W1W2>W3W4) {W1,W2} should be more coherent than {W3,W4}.

Similarly, there is a greater agreement between W4 and W5 than W3 and W4, so {W4,W5} seems more coherent than {W3,W4}.

(W4W5>W3W4) {W4,W5} should be more coherent than {W3,W4}.

Bayesian networks. The basic idea behind the CPTs we used is that for any particular witness we take the probability of them including the perpetrator in their list to be 0.8, and the probability of including an innocent to be .05. Of course, the example can be run with different conditional probability tables. Moreover, in this case, the fact that the witnesses provided their testimonies constitutes evidence (in contrast with the Robbers scenario, where there is no evidence as to who the perpetrator is), and so we update on it in our weights calculations. Let's first take a look at the BN for the first scenario (Figure 8).



D	Pr
Steve	0.167
Martin	0.167
David	0.167
John	0.167
James	0.167
Peter	0.167

W1	D					
	Steve	Martin	David	John	James	Peter
1	0.8	0.05	0.05	0.05	0.05	0.05
0	0.2	0.95	0.95	0.95	0.95	0.95

Figure 8: BN for the W1W2 narration in the Witness problem. CPT for W2 is identical to the one for W1.

In the remaining two BNs for the problem the CPT for D remains the same, and the CPTs for the witness nodes are analogous to the one for W1. The remaining BNs have the following DAGs (Fig. 9).

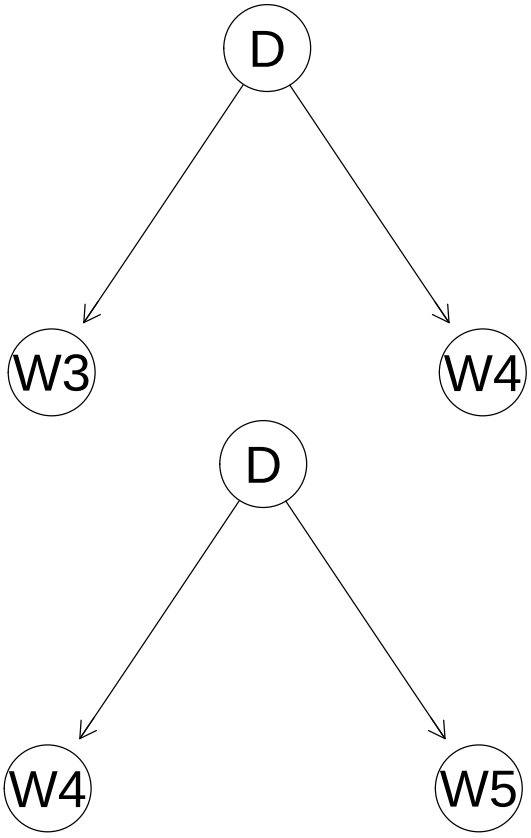


Figure 9: Two remaining DAGs for the Witness problem.

Results. Note that in this case we're dealing with the perspective of someone who starts with a uniform prior for D, and subsequently considers what would happen if she obtained a given set of testimonies as evidence. Clearly, then, the distribution for D would no longer be uniform, but rather result from updating on this evidence.

	Steve	Martin	David	John	James	Peter
Pr	0.981	0.004	0.004	0.004	0.004	0.004

Table 11: Propagated probabilities for D in the Witnesses scenario (rounded).

In the coherence calculations we first obtain how specific states of the narration child nodes in a narration would be confirmed by various states of D. In calculating these confirmation scores we do *not* use the updated BN (after all, confirmation by evidence if you start with BN according to which you already received it would not be positive). However, the weights assigned to confirmation scores in the ecs calculations should be the (normalized) probabilities obtained by updating on the evidence. If given the evidence included in the scenario Steve is the most likely perpetrator, in your coherence calculations you give the most weight to the confirmation score obtained if he indeed is the perpetrator.

–> –> –>

–> –>

–> –> –>

3.8 Depth

add tables explaining nodes

The scenario. There are eight equally likely suspects $1,\ldots,8$, and three equally reliable witnesses each trying to identify the person responsible for the crime. Compare two sets of claims about who is responsible. In contrast with Witnesses, the example ingores the fact that these are testimonies and by focusing on the material disjunctive content.

X1 and X2: Revise

$$X_1 = \{(1 \vee 2 \vee 3), (1 \vee 2 \vee 4), (1 \vee 3 \vee 4)\}$$
$$X_2 = \{(1 \vee 2 \vee 3), (1 \vee 4 \vee 5), (1 \vee 6 \vee 7)\}$$

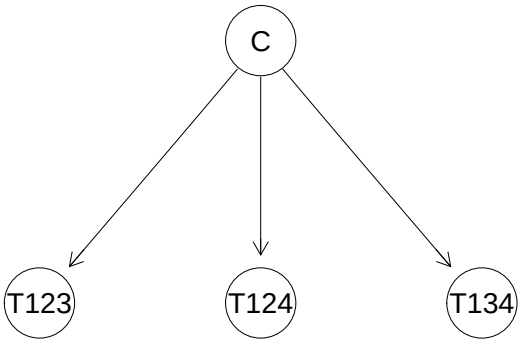
Desiderata. In X1 witnesses' testimonies have bigger overlap, between each pair of the witnesses 2 suspects are the same, and in X2 only 1 suspect is always the same. Following Schupbach (2008), one may have an intuition that the first situation is more coherent.

(X1>X2) X_1 should be more coherent than X_2.

In our framework, the BN with narration and evidence nodes and states is supposed to represent the conviction of a single agent. From this perspective, calculating the coherence of X_1 (or of X_2) in our way would presume these are the beliefs of a single agent. This being the case, if the agent accepts an uncontroversial application of the conjunction closure

convert to table

Bayesian networks. We start with representing the two scenarios with two fairly natural BNs (C stands for who Committed the crime, TXYZ stands for Testimony that $X \vee Y \vee Z$), see Figures 10 and 11.



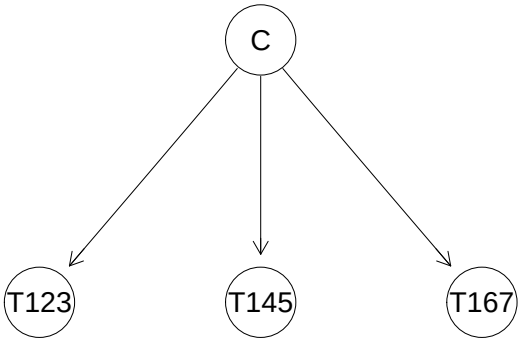
C	Pr
1	0.125
2	0.125
3	0.125
4	0.125
5	0.125
6	0.125
7	0.125
8	0.125

T123	C							
	1	2	3	4	5	6	7	8
1	1	1	1	0	0	0	0	0
0	0	0	0	1	1	1	1	1

T124	C							
	1	2	3	4	5	6	7	8
1	1	1	0	1	0	0	0	0
0	0	0	1	0	1	1	1	1

T134	C							
	1	2	3	4	5	6	7	8
1	1	0	1	1	0	0	0	0
0	0	1	0	0	1	1	1	1

Figure 10: BN for X1 in the Depth problem.



C	Pr
1	0.125
2	0.125
3	0.125
4	0.125
5	0.125
6	0.125
7	0.125
8	0.125

T123	C							
	1	2	3	4	5	6	7	8
1	1	1	1	0	0	0	0	0
0	0	0	0	1	1	1	1	1

T145	C							
	1	2	3	4	5	6	7	8
1	1	0	0	1	1	0	0	0
0	0	1	1	0	0	1	1	1

T167	C							
	1	2	3	4	5	6	7	8
1	1	0	0	0	0	1	1	0
0	0	1	1	1	1	0	0	1

Figure 11: BN for X2 in the Depth problem.

Results. One effect of dropping the "the witness testified that" and using the testimony contents themselves is that the CPTs for the narration nodes are deterministically connected with the root node. In result, the coherence calculations give the results in Table **??**, and the status of the desiderata is pictured in Table **??**.

–> –> –> –>

–> –>

need to revise this

Note that this time we listed two values for our measure. Structured 1 shows the values obtained if we do not update the weighting of the node not included in the narration, and Structured 2 is the result of such an updated weighing (analogous to the updating involved in the Witness problem). Now, what are we to make of this?

Structured 1 is negative. This isn't too surprising: after all, this is the coherence of the narration with the probabilistic assumption that the distribution for C is uniform, and this probabilistic assumption undermines the narration. Why, however, does Structured 2 equal 1, and why are the results identical for both narrations? This, upon reflection, isn't too suprising either. If the BN and the narration is supposed to represent a single agent's credal state, there is only one state of C in which the whole narration X_1 is true – trivially, it is the one in which suspect 1 is guilty, and it is the same unique state of C in which the whole narration X_2 is true. Since seen as narrations these sets have exactly the same truth conditions, there is no surprise in them being equally coherent.

What if the sentences in the set are not claims made by one agent and there is no single underlying

credal state? We aren't convinced that our tool is optimal for measuring the agreement of multiple witnesses. Instead, there already exists a working measure of such an agreement — Cohen's κ – which already gives the desired results.

To illustrate, let's think of a simplified situation (devoid of three-dimensional tables) with two witnesses $w1$ and $w2$, where the respective sets are $A = \{1 \vee 2 \vee 3, 1 \vee 2 \vee 4\}$ and $B = \{1 \vee 2 \vee 3, 1 \vee 4 \vee 5\}$ and in each set the first proposition comes from $w1$ and the second from $w2$. The information for these two sets is pictured in Tables 12 and 13.

	w2: suspect	w2: innocent
w1:suspect	2	1
w1: innocent	1	4

Table 12: Situation A in the Depth problem.

	w2: suspect	w2: innocent
w1: suspect	1	2
w1: innocent	2	3

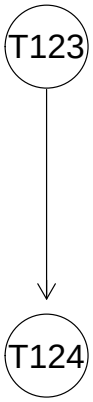
Table 13: Situation B in the Depth problem.

Standard calculations using the vcd package results in the unweighted values of Cohen's κ pictured in Table 14.

	A	B
value	0.467	-0.067

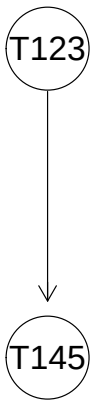
Table 14: Kappas for the Depth scenario.

Let's further illustrate our point about the requirement that the BN should represent a single agent's cognitive state. For instance, you can represent, the situation in A from the perspective of the first witness. This suggests we should focus only on the nodes involved in the narration, and on the fact that from the witness' perspective the suspects are not equally likely. The example doesn't provide us enough information to build a table for C. In fact, no information about the wintess attitude towards this node is given, but given they say what they say, it's unlikely they think the distribution is uniform. So let's take one of the witness' own statements as the root (which ones we choose doesn't change the outcome). Clearly (or, at least, hopefully, if we talk about witnesses), the agent thinks her own claim is very likely and evaluates the probability of the other statements in A or B from its perspective. This gives us two different BNs, and when we calculate the respective coherence scores we actually do get the desired result, which isn't too hard for the other measures either.



T124	T123	
	1	0
1	0.667	0.2
0	0.333	0.8

Figure 12: A witness perspective for the agreement problem, set A.



T124	T123	
	1	0
1	0.333	0.4
0	0.667	0.6

Figure 13: A witness perspective for the agreement problem, set B.

Now, the resulting calculations and outcomes are in Tables **??** and **??**, and the situation is no longer problematic for any of the measures.

–> –> –>

–> –> –>

4 Discussion

Ultimately, all the calculations are displayed in Table 15 and the desiderata yield Table 16, with corresponding success rates (Table 17).

	OG	OGen	Sh	ShGen	Fit	DM	R	S
Penguins: BGP 111	0.010	0.015	4.000	2.010	0.453	0.255	0.255	0.010
Penguins: BG 11	0.010	0.010	0.040	0.040	-0.960	-0.480	-0.480	-0.960
Penguins: BP 11	0.020	0.020	2.000	2.000	0.669	0.255	0.255	0.010
Dunnit: MGWI 1111	0.000	0.087	4.294	11.012	0.169	0.167	0.167	-0.932
Dunnit: MTGWI 11111	0.000	0.042	73.836	13.669	0.385	0.150	0.150	-0.100
Japanese Swords 1: JO 11	0.004	0.004	80.251	80.251	0.976	0.008	0.008	0.008
Japanese Swords 2: JO 11	0.818	0.818	9.000	9.000	0.976	0.800	0.800	0.889
Japanese Swords 3: JO 11	0.818	0.818	1.080	1.080	0.286	0.067	0.067	0.400
Robbers: MIsPMIsR 11	0.600	0.600	0.937	0.937	-0.143	-0.050	-0.050	0.600
Robbers: MIsPMIsR 10	0.250	0.250	1.250	1.250	0.571	0.125	0.125	-0.600
Robbers: MIsPMIsR 01	0.250	0.250	1.250	1.250	0.571	0.125	0.125	-0.600
Beatles: JPGRD 11111	0.000	0.202	0.000	1.423	-0.036	0.025	0.025	-1.000
Books: AR 11	0.014	0.014	1.493	1.493	0.212	0.027	0.027	0.055
Books: AR 10	0.009	0.009	0.945	0.945	-0.127	-0.025	-0.025	-0.055
Books: AR 01	0.100	0.100	0.995	0.995	-0.101	-0.003	-0.003	-0.005
Books: AR 00	0.892	0.892	1.001	1.001	0.016	0.001	0.001	0.005
Witness: W1W2 11	0.451	0.451	3.551	3.551	0.771	0.446	0.446	0.729
Witness: W3W4 11	0.187	0.187	0.740	0.740	-0.234	-0.110	-0.110	0.494
Witness: W4W5 11	0.365	0.365	1.260	1.260	0.218	0.110	0.110	0.602
DepthA: T123T124 11	0.664	0.664	1.014	1.014	0.280	0.012	0.012	0.027
DepthB: T123T145 11	0.331	0.331	0.996	0.996	-0.047	-0.003	-0.003	-0.004
Regular: TTF 11	0.500	0.500	3.000	3.000	0.833	0.500	0.500	1.000
Dodecahedron: TTF 11	0.500	0.500	6.000	6.000	0.917	0.625	0.625	1.000

Table 15: Coherence scores for all the examples.

	OG	OGen	Sh	ShGen	Fit	DM	R	S
Penguins: BG<BGP	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Penguins: BP≈ BGP	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
Dunnit: Dunnit<Twin	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE
Swords: JO2>JO1	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
Swords: JO2>JO3	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Robbers: PR>P¬R	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
Robbers: PR>neutral	NA	NA	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
Beatles: below neutral	NA	NA	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE
Beatles: minimal	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
Books: AR>A¬R	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Books: AR> ¬AR	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Books: ¬A¬R>A¬R	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Books: ¬A¬R> ¬AR	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Witness: $W_1W_2 > W_3W_4$	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Witness: $W_4W_5 > W_3W_4$	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Depth: $X_1 > X_2$	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Dodecahedron: Regular = Dodecahedron	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE

Table 16: Overall desiderata satisfaction in the examples discussed).

OG	OGen	Sh	ShGen	Fit	DM	R	S
0.733	0.733	0.706	0.647	0.706	0.647	0.706	1

Table 17: Success rates in the examples discussed.

References

Akiba, K. (2000). Shogenji's probabilistic measure of coherence is incoherent. *Analysis*, *60*(4), 356–359. https://doi.org/10.1093/analys/60.4.356

Bovens, L., & Hartmann, S. (2004). *Bayesian epistemology*. Oxford University Press.

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential Support: theoretical and empirical Issues. *Philosophy of Science*, *74*(2), 229–252. https://doi.org/10.1086/520779

Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese*, *156*(3), 405–425. https://doi.org/10.1007/s11229-006-9131-z

Fenton, N., & Neil, M. (2018). *Risk assessment and decision analysis with bayesian networks*. Chapman; Hall.

Fitelson, B. (2003). A Probabilistic Theory of Coherence. *Analysis*, *63*(3), 194–199.

Glass, D. H. (2002). Coherence, Explanation, and Bayesian Networks. In G. Goos, J. Hartmanis, J. van Leeuwen, M. O'Neill, R. F. E. Sutcliffe, C. Ryan, … N. J. L. Griffith (Eds.), *Artificial Intelligence and Cognitive Science* (Vol. 2464, pp. 177–182). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45750-X_23

Koscholke, J. (2016). Evaluating test cases for probabilistic measures of coherence. *Erkenntnis*, *81*(1), 155–181. https://doi.org/10.1007/s10670-015-9734-1

Meijs, W., & Douven, I. (2007). On the alleged impossibility of coherence. *Synthese*, *157*(3), 347–360. https://doi.org/10.1007/s11229-006-9060-x

Merricks, T. (1995). Warrant entails truth. *Philosophy and Phenomenological Research*, *55*, 841–855.

Olsson, E. J. (2001). Why Coherence Is Not Truth-Conducive. *Analysis*, *61*(3), 236–241.

Olsson, E. J. (2005). The Impossibility of Coherence. *Erkenntnis*, *63*(3), 387–412. https://doi.org/10.1007/s10670-005-4007-z

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*. *191*(16), 3821–3845. https://doi.org/10.1007/s11229-014-0501-7

Schippers, M., & Koscholke, J. (2019). A General Framework for Probabilistic Measures of Coherence. *Studia Logica*. https://doi.org/10.1007/s11225-019-09848-3

Schupbach, J. N. (2008). On the alleged impossibility of bayesian coherentism. *Philosophical Studies*, *141*(3), 323–331. https://doi.org/10.1007/s11098-007-9176-y
Scutari, M., & Denis, J.-B. (2015). *Bayesian networks in* ***r***. CRC Press.
Shogenji, T. (1999). Is coherence truth-conducive? *Analysis*, *59*(4), 338–345.
Shogenji, T. (2001). Reply to Akiba on the probabilistic measure of coherence. *Analysis*, *61*(2), 147–150. https://doi.org/10.1093/analys/61.2.147
Shogenji, T. (2006). Why does coherence appear truth-conducive? *Synthese*, *157*(3), 361–372. https://doi.org/10.1007/s11229-006-9062-8
Siebel, M. (2004). On Fitelson's measure of coherence. *Analysis*, *64*, 189–190.
Siebel, M. (2006). Against probabilistic measures of coherence. In *Coherence, truth and testimony* (pp. 43–68). Springer.