A New Posterior Probability-Based Measure of Coherence

Stephan Hartmann (S.Hartmann@lmu.de)
Munich Center for Mathematical Philosophy and Department of Psychology, LMU Munich, 80539 Munich (Germany)

Borut Trpin (borut.trpin@lrz.uni-muenchen.de)
Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany)

Abstract

According to a common view in epistemology, a set of propositions is justified if it is coherent. Similarly, a new proposition should be accepted if it is coherent with the accepted body of beliefs. But what is coherence? And what, in turn, justifies the above claims? To answer these questions, various Bayesian measures of epistemic coherence have been proposed. Most of these measures are based on the prior probability distribution over the corresponding propositional variables. We criticize this "static" conceptualization of coherence and propose instead that the coherence of an information set is related to how well the information set responds when each of the propositions it contains is confirmed by an independent and partially reliable information source. The elaboration of this idea will show that the proposed "dynamic" perspective has several advantages and solves some open problems of coherentist epistemology. It also has implications for our understanding of reasoning and argumentation in science and beyond.

Keywords: Reasoning and Argumentation; Coherence; Formal Epistemology

Introduction

In scientific and in ordinary reasoning and argumentation, we often use coherence considerations (e.g. Thagard, 2000, 2007). These considerations guide us when it comes to forming hypotheses and possibly incorporating them into our belief system. In doing so, we strive for a coherent belief system and want new information to fit well into it. This pervasive practice raises several general questions, such as: What is epistemic[1] coherence anyway? How can we explicate the concept of coherence? And: What is good about coherence? Why should we strive for coherent information at all? Although these questions have already been addressed in the philosophical literature (see e.g. BonJour, 1985), many aspects are still unclear, mainly due to the fact that the concept of coherence is difficult to make precise. The use of formal methods promises progress here, as has already been shown in several studies (see Olsson, 2022 for a recent overview).

Formally, coherence is a property of an information set $\mathbf{S} = \{F_1, \ldots, F_n\}$, over which a (subjective) probability distribution P is defined. To illustrate this, we consider the propositions B: "My pet Tweety is a bird", G: "My pet Tweety cannot fly" and P: "My pet Tweety is a penguin" and let $\mathbf{S} = \{B, G\}$ and $\mathbf{S}' = \{B, G, P\}$. Fig. 1. shows possible probability distributions that an agent could choose in each case.

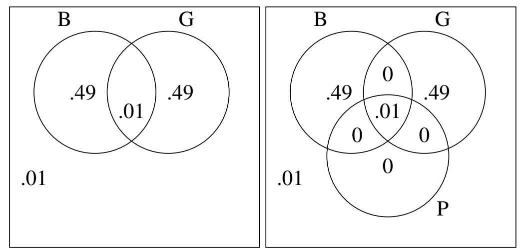

Figure 1: Venn diagrams of the Tweety example.

Here we would like to judge that $\mathbf{S}'$ is more coherent than $\mathbf{S}$. To do this, we need to define a suitable coherence measure as a function of the corresponding probability distribution P. This measure then induces an ordering of information sets according to their coherence. Moreover, it seems desirable that a measure allows us to decide whether an information set is coherent (or incoherent) in an absolute sense. For example, we would like to say that the information set $\mathbf{S}'$ above is (absolutely) coherent, while the information set $\mathbf{S}$ is (absolutely) incoherent. Again, we will see that some of the measures discussed in the literature allow such a judgment, while others do not. A natural way to formally arrive at such a judgment is to show that there is a threshold value τ and that an information set is (absolutely) coherent (or incoherent) if and only if the coherence value assigned to it by a given measure is greater (or less) than τ. However, it turns out that for some measures (such as the Olsson-Glass measure, which we will come back to) there is no such value for τ.

While at first glance it seems plausible that all information sets can be ordered according to their coherence, Bovens and Hartmann (2003, Ch. 1) have expressed doubts about this assertion by pointing to examples such as the following. Let $\mathbf{S} := \{F_1, F_2, F_3\}$ with F_1: "The culprit was a woman", F_2: "The culprit had a Danish accent", and F_3: "The culprit drove a Ford", and $\mathbf{S}' := \{F'_1, F'_2, F'_3\}$ with F'$_1$: "The culprit was wearing Coco Chanel shoes", F'$_2$: "The culprit had a French accent", and F'$_3$: "The culprit drove a Ford". We ask:

[1] We avoid explicitly referring to *epistemic* coherence in what follows because this is the only kind of coherence we assess in this paper. That is, our paper only engages with the type of coherence that refers to how information cognitively fits together. There are also other important related conceptions such as logical coherence, which revolves around logical consistency, and probabilistic coherence, which generalizes logical consistency. BonJour (1985, pp. 93-101) provides an insightful discussion of how these and several other aspects of coherence are related.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Which of the two sets is more coherent? On the one hand, the information set **S** seems to consist of independent sentences, which does not make it particularly coherent. But is it therefore less coherent than $\mathbf{S}'$? It should be noted that in $\mathbf{S}'$ two of the propositions seem to be positively correlated (namely F'$_1$ and F'$_2$) and two propositions (F'$_2$ and F'$_3$) seem to be negatively correlated. This suggests that there may be no fact of the matter as to which of the two sets is more coherent.

With this in mind, Bovens and Hartmann (2003) have proposed how to generate a partial coherence ordering of information sets instead. Specifically, this is achieved by assuming that each piece of information in **S** is confirmed by an independent and partially reliable information source. It is assumed that all information sources have the same reliability r. One can then calculate the posterior probability of **S** as a function of r and define a suitable coherence measure based on this posterior probability. Finally, it is postulated that an information set **S** is more coherent than an information set $\mathbf{S}'$ if, for all values of the reliability parameter r, the coherence measure of **S** is greater than the coherence measure of $\mathbf{S}'$. If the curves of the respective coherence measures cross at a point as a function of r, then **S** and $\mathbf{S}'$ cannot be ordered by their coherence.

The change from a "static", prior-based view to a "dynamic", posterior-based view of coherence thus allows us to take into account the plausible insight (suggested by examples such as the one mentioned above) that not all information sets can be ordered according to their coherence. This view is dynamic in that it takes into account the probabilistic response of the information set when confronted with confirming evidence. The central idea here, then, is that one must examine an information set in order to learn about its coherence. (This approach is reminiscent of how we learn about causal relationships. Here, too, the system in question must be investigated; in this case by means of interventions.)

Although this approach may sound promising, the specific measure proposed by Bovens and Hartmann (2003) has been criticized as it does not cope with some simple test cases (see Meijs, 2005). For example, if we allow a very minor probability that somebody has an ostrich for a pet (i.e., a ground-dwelling pet bird that is not a penguin) and accordingly slightly change the probability distribution in the Tweety example from Fig. 1, so that $P(\mathrm{B},\mathrm{G},\neg\mathrm{P}) = .000001$ instead of 0 and $P(\neg\mathrm{B},\neg\mathrm{G},\neg\mathrm{P}) = .009999$ instead of .01, then the measure by Bovens and Hartmann cannot determine whether learning that Tweety is a penguin increases the coherence or not (Meijs, 2005, pp. 58-59). In addition, there is another problem that also affects the other (prior probability-based) measures. All these measures either always allow us to judge an information set as absolutely coherent (or absolutely incoherent), or they never allow us to make such a judgment. The Bovens and Hartmann measure belongs to the second category. However, in analogy to the impossibility of always ordering information sets according to their coherence made plausible above, it is also possible that one cannot always decide whether an information set is absolutely coherent or absolutely incoherent. It may well be that some information sets have this property while others do not. Accordingly, the new coherence measure we propose in this paper will provide only a partial ordering of information sets according to their coherence. In particular, it will allow us to identify cases in which there is arguably no fact of the matter as to whether an information set is (absolutely) coherent or incoherent.

Prior-Based Measures of Coherence

We will now discuss three prior-based probabilistic coherence measures. These measures formalize one or more core intuitions associated with the notion of coherence. One of them is the intuition that the propositions in a coherent information set are interdependent and informationally relevant to each other. To formalize this, we consider an information set $S := \{\mathrm{F}_1, \ldots, \mathrm{F}_n\}$ and introduce the following definition.

Definition 1. *A probability distribution P is defined over a set of propositional variables* $V := \{F_1, \ldots, F_n\}$ *with the values* F_i *and* $\neg\mathrm{F}_i$ *for all* $i = 1, \ldots n$.[2]

(i) *V is* independent (relative to *P*) *iff* $P\big(\bigwedge_{i\in I}\mathrm{F}_i\big) = \prod_{i\in I} P(\mathrm{F}_i)$ *for all non-empty subsets* $I \subseteq \{1,\ldots,n\}$.

(ii) *V is* positively correlated (relative to *P*) *iff* $P\big(\bigwedge_{i\in I}\mathrm{F}_i\big) \geq \prod_{i\in I} P(\mathrm{F}_i)$ *for all non-empty subsets* $I \subseteq \{1,\ldots,n\}$ *and at least one of the "*$\geq$*" is a "*$>$*".*

(iii) *V is* negatively correlated (relative to *P*) *iff* $P\big(\bigwedge_{i\in I}\mathrm{F}_i\big) \leq \prod_{i\in I} P(\mathrm{F}_i)$ *for all non-empty subsets* $I \subseteq \{1,\ldots,n\}$ *and at least one of the "*$\leq$*" is a "*$<$*".*

The Shogenji measure (Shogenji, 1999), defined as

$$coh_S(\mathbf{S}) := \frac{P(\mathrm{F}_1,\ldots,\mathrm{F}_n)}{P(\mathrm{F}_1 \cdots P(\mathrm{F}_n)},$$

is based on this intuition: If **S** is independent, then $coh_S(\mathbf{S}) = 1$; if **S** is positively correlated, then $coh_S(\mathbf{S}) > 1$, and if **S** is negatively correlated, then $coh_S(\mathbf{S}) < 1$. Hence, the Shogenji measure allows us to define a notion of absolute coherence–**S** is absolutely coherent if $coh_S(\mathbf{S}) > 1$, and it is absolutely incoherent if $coh_S(\mathbf{S}) < 1$. As all information sets are assigned a coherence value, the Shogenji measure provides a complete ordering of information sets according to their coherence.

According to another intuition, coherence has something to do with the relative overlap of the propositions in probability space. The greater the relative overlap, the greater the coherence. The idea behind this proposal is that there is a lot of agreement between coherent propositions. The simplest measure that captures this intuition is the Olsson-Glass measure (Olsson, 2002; Glass, 2002). It is defined as follows:

$$coh_{OG}(\mathbf{S}) := \frac{P(\mathrm{F}_1,\ldots,\mathrm{F}_n)}{P(\mathrm{F}_1 \vee \cdots \vee \mathrm{F}_n)}$$

[2] We follow the convention of denoting propositional variables in italics and their values in roman script.

It should be noted that although this measure also allows a complete ordering of information sets according to their coherence, it does not have a natural threshold that would allow the notion of absolute coherence to be defined.

The third and final measure we will discuss here combines the intuition of relative overlap and deviation from independence. It is derived from the Olsson-Glass measure by normalizing it in an appropriate way. To illustrate the procedure, we start with the following definition:

Definition 2. *A probability distribution P is defined over a set of propositional variables $V := \{F_1, \ldots, F_n\}$. The associated probability distribution $\tilde{P}$ satisfies the following conditions: (i) $\tilde{P}$ is defined over the same set V; (ii) V is independent relative to $\tilde{P}$; (iii) $\tilde{P}(\mathrm{F}_i) = P(\mathrm{F}_i)$ for all $i = 1, \ldots, n$.*

Following Hartmann and Trpin (2023), we then define:

$$
\begin{aligned}
coh_{OG^+}(\mathbf{S}) \quad &:= \quad \frac{coh_{OG}^{(P)}(\mathbf{S})}{coh_{OG}^{(\tilde{P})}(\mathbf{S})} \\
&= \quad coh_S(\mathbf{S}) \cdot \frac{1 - P(\neg \mathrm{F}_1) \cdots P(\neg \mathrm{F}_\mathrm{n})}{1 - P(\neg \mathrm{F}_1, \ldots, \neg \mathrm{F}_\mathrm{n})}
\end{aligned}
$$

Note that here the Olsson-Glass measure evaluated under the probability distribution P is divided by the Olsson-Glass measure evaluated under the associated probability distribution $\tilde{P}$. It turns out that the resulting measure has many advantages: Not only does it combine the two main intuitions associated with the notion of coherence (i.e., dependence and relative overlap). It also makes it possible to define a threshold value τ above which a set of information is considered absolutely coherent (this threshold value is 1). Moreover, it is (in contrast to several other proposals in the literature; see e.g. Koscholke, Schippers, & Stegmann, 2019) easy to calculate. Finally, Hartmann and Trpin (2023) showed that coh_{OG^+} performs well in terms of truth-tracking. That is, if there is no further evidence confirming the items in the information set, then it makes sense to base the acceptance of an information set on its coherence, since a more coherent set – as measured by the OG^+-measure – has a higher chance of being true.

Before closing this section, it is interesting to note that one can define the Shogenji measure in terms of the associated probability distribution introduced in Definition 2:

$$coh_S(\mathbf{S}) := \frac{P(\mathbf{S})}{\tilde{P}(\mathbf{S})}$$

Posterior-Based Measures of Coherence

As we have seen, all the measures mentioned in the last section are based on the prior probability distribution. The decision as to whether an information set is (absolutely) coherent and how coherent it is follows solely from this. In contrast, we now turn to a class of coherence measures based on the posterior probability distribution. It turns out that these measures have numerous advantages. Originally proposed by Bovens and Hartmann (2003), one investigates how much the entire information set is confirmed when each element of the information set is confirmed by an independent and partially reliable information source. Specifically, an information set represented by a Bayesian network is considered, where the prior probability distribution takes into account not only the relevant *fact variables* F_i, but also the corresponding *report variables* R_i (for $i = 1, \ldots, n$). The following important independence assumption is made:

$$
\begin{aligned}
&R_i \perp\!\!\!\perp R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n, F_1, \ldots, F_{i-1}, F_{i+1}, \ldots, F_n | F_i \\
&\forall i \in \{1, \ldots, n\} \qquad (1)
\end{aligned}
$$

This means that each report variable is independent of all other report variables and all other fact variables if the value of the corresponding fact variable is given. This assumption models the idea that the reports are independent. See Fig. 2 for an illustration.

Assuming that the likelihoods of all reports are the same, one sets $P(\mathrm{R_i}|\mathrm{F_i}) = p$ and $P(\mathrm{R_i}|\neg\mathrm{F_i}) = q$ for all $i = 1, \ldots, n$. From this one can define the likelihood ratio $x := q/p$ and the reliability $r = 1 - x$. (For a discussion see Bovens & Hartmann, 2003, Ch. 2. All other strictly monotonically decreasing functions $\rho(x)$ with $\rho(0) = 0$ and $\rho(1) = 1$ are also possible.)[3] The resulting posterior probability $P^*(\mathbf{S}) = P(\mathrm{F_1}, \ldots, \mathrm{F_n}|\mathrm{R_1}, \ldots, \mathrm{R_n})$ can then be defined by the so-called *weight vector* $\langle a_0, a_1, \ldots, a_n \rangle$, where $a_0 := P(\mathrm{F_1}, \ldots, \mathrm{F_n})$, $a_1 := P(\neg\mathrm{F_1}, \mathrm{F_2}, \ldots, \mathrm{F_n}) + P(\mathrm{F_1}, \neg\mathrm{F_2}, \ldots, \mathrm{F_n}) + \ldots$ etc. Note that each a_i represents the probability that exactly i of n propositions are false. Note also that a_0 is the prior probability of $\mathbf{S}$ and that all a_i add up to 1: $\sum_{i=0}^{n} a_i = 1$. After receiving positive reports, the information set $\mathbf{S}$ gets a probability boost $boost(\mathbf{S}) := P^*(\mathbf{S})/P(\mathbf{S})$. Next, the authors note that an information set with completely overlapping propositions receives a maximum boost. In this case, the weight vector is simply given by[4] $\langle a_0, 0, 0, \ldots, 0, \overline{a_0} \rangle$ and $boost_{max}(\mathbf{S}) := P^*_{max}(\mathbf{S})/P(\mathbf{S})$. Finally, the coherence of $\mathbf{S}$ is defined as $coh_x^{(BH)}(\mathbf{S}) := boost(\mathbf{S})/boost_{max}(\mathbf{S}) = P^*(\mathbf{S})/P^*_{max}(\mathbf{S})$. Accordingly, one obtains:

$$coh_x^{(BH)}(\mathbf{S}) = \frac{a_0 + \overline{a_0} x^n}{\sum_{i=0}^{n} a_i x^i}$$

Furthermore, Bovens and Hartmann argue that an information set $\mathbf{S}$ is more coherent than an information set $\mathbf{S}'$ if for all $x \in (0, 1)$, $coh_x^{BH}(\mathbf{S}) > coh_x^{BH}(\mathbf{S}')$. This simple proposal led to a number of interesting results. For example, it could be shown that the "correct" judgment can be obtained for numerous test cases from the literature. However, as mentioned, Meijs (2005) also pointed out several counterexamples which

[3] Although we consider the likelihood ratio x and the reliability $r = 1 - x$ for the whole range of values between 0 and 1, we assume that x and the corresponding r are the same for all reports because we are assuming that all reports are on equal standing. It is an interesting task for future research to consider how different modelling assumptions related to x would affect our results. See Osimani and Landes (2023) for a related discussion in a different context.

[4] Here and in the following we use the short form $\overline{a_0} := 1 - a_0$.

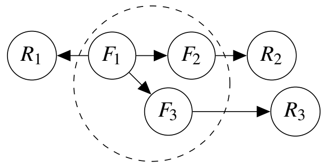

Figure 2: An example of a Bayesian network representing the probabilistic (in-)dependencies between the fact variables F_i and the corresponding report variables R_i (for $i = 1, 2$ and 3).

can occur due to minor changes in the probability distribution or, more generally, when mutual support relations of the set are not captured. Moreover, since the measure is in the tradition of the overlap measure due to the specific normalization used, there is no threshold that could be used to decide whether the information set is absolutely coherent or not.

These shortcomings, in conjunction with the discussion in the previous section, lead to our new proposal. Specifically, instead of using $P^*_{max}(\mathbf{S})$, we propose to normalize the posterior probability with $\tilde{P}^*(\mathbf{S})$, i.e. with the posterior probability that would be obtained if all propositions were independent, but without changing the marginal probabilities of the individual propositions. Accordingly, we propose

$$coh_x(\mathbf{S}) := \frac{P^*(\mathbf{S})}{\tilde{P}^*(\mathbf{S})} \tag{2}$$

as a new posterior probability-based measure of coherence.

Let us now explore the proposed measure in more detail. To do so, let us first define $c_i^{(n)}$ as the sum of the probabilities of i true propositions (for $i = 1, \ldots, n$). Note that $c_n^{(n)} = a_0$. For example, $c_1^{(3)} = P(\mathrm{F}_1) + P(\mathrm{F}_2) + P(\mathrm{F}_3)$ and $c_2^{(3)} = P(\mathrm{F}_1, \mathrm{F}_2) + P(\mathrm{F}_1, \mathrm{F}_3) + P(\mathrm{F}_2, \mathrm{F}_3)$. For convenience, let us furthermore set $c_0^{(n)} = 1$. Then the following proposition holds (all proofs are in the appendix):

Proposition 1. *Let $\mathbf{S} = \{\mathrm{F}_1, \ldots, \mathrm{F}_n\}$ be an information set and F_i the corresponding fact variables and R_i the corresponding report variables. A probability distribution P is defined over these variables, which fulfils the independence condition from Eq. (1). Then*

$$P^*(\mathbf{S}) = \frac{a_0}{\sum_{i=0}^{n} c_i^{(n)} x^i \,\overline{x}^{n-i}}$$

It is interesting to note that $P^*(\mathbf{S})$ decreases if one of the $c_i^{(n)}$ increases and everything else is kept fixed. For $n = 3$, we have

$$P^*(\mathbf{S}) = \frac{a_0}{x^3 + c_1^{(3)} x^2 \overline{x} + c_2^{(3)} x \overline{x}^2 + a_0 \overline{x}^3}$$

If one keeps $a_0, c_1^{(3)}$ and x fixed and increases $c_2^{(3)}$, then $P^*(\mathbf{S})$ decreases. This is surprising as the dependence between the variables increases and so one would expect that therefore the confirmation of one proposition "swaps over" to the other propositions so that the whole set gets more confirmed. The

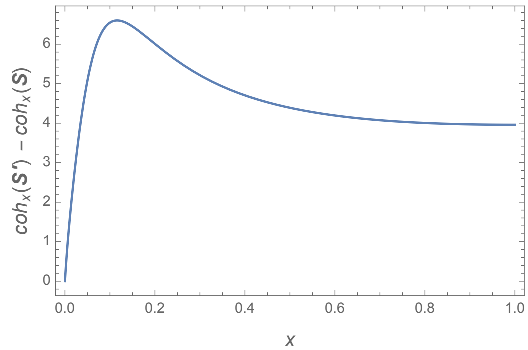

Figure 3: The difference function for the Tweety example.

reason is that a_0 and $c_1^{(3)}$ are kept fixed, which implies the introduction of negative dependence elsewhere in the information set as a result of increasing $c_2^{(3)}$.

An analogous result as the one stated in Proposition 1 holds for $\tilde{P}(\mathbf{S})$; just replace a_0 by $\tilde{a}_0$ and $c_i^{(n)}$ by $\tilde{c}_i^{(n)}$. Note that these expressions are all functions of the marginals $P(\mathrm{F_i})$ only and that $\tilde{P}(\mathbf{S})$ can also be written in factored form. However, for the present purpose, the representation in terms of the $\tilde{c}_i^{(n)}$ is most convenient. Finally, we obtain

$$coh_x(\mathbf{S}) = \frac{a_0}{\tilde{a}_0} \cdot \frac{\sum_{i=0}^{n} \tilde{c}_i^{(n)} x^i \,\overline{x}^{n-i}}{\sum_{i=0}^{n} c_i^{(n)} x^i \,\overline{x}^{n-i}}.$$

An interesting consequence of this representation is that the coherence of an information set decreases if one of the $c_i^{(n)}$ increases while keeping everything else fixed. This observation supports the claim made in Hartmann and Trpin (2023) that the principle of Dependence is not sacrosanct in the debate about coherence.

We now want to investigate what kind of coherence ordering the new posterior-based measure generates. As in the case of the Bovens-Hartmann measure, we say that an information set $\mathbf{S}$ is more coherent than an information set $\mathbf{S}'$ if for all $x \in (0, 1)$, $coh_x(\mathbf{S}) > coh_x(\mathbf{S}')$. It is therefore not surprising that the new measure, like the Bovens and Hartmann measure, allows scenarios in which there is no fact of the matter as to which of the two information sets is more coherent. The new measure generates only a partial ordering, which agrees with our intuitions (as discussed above).

As an illustration, let us consider the Tweety example. The plot in Fig. 3 shows the *difference function* $f(x) := coh_x(\mathbf{S}') - coh_x(\mathbf{S})$ with $\mathbf{S} := \{B, G\}$ and $\mathbf{S}' := \{B, G, P\}$ with the probability distributions from Fig. 1. We find that the difference function is always positive and conclude, in accordance with our intuitions, that $\mathbf{S}'$ is more coherent than $\mathbf{S}$.

It turns out that the new posterior-based measure also copes well with other test cases and avoids Meijs' counterexamples against the Bovens and Hartmann measure. For example, a very small change in the probability distribution of the Tweety example (as described earlier) has a negligible effect on coherence, just like we would expect.

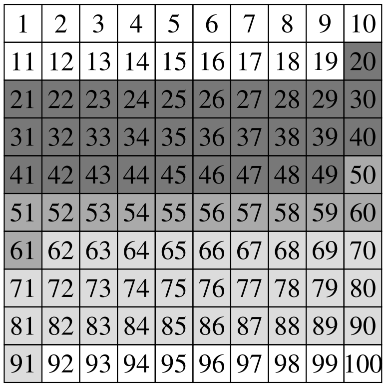

Figure 4: An illustration of sit_3 from the Tokyo corpse example. Witness 1 reports the corpse to be somewhere in squares 20-61, while Witness 2 reports it to be somewhere in 50-91.

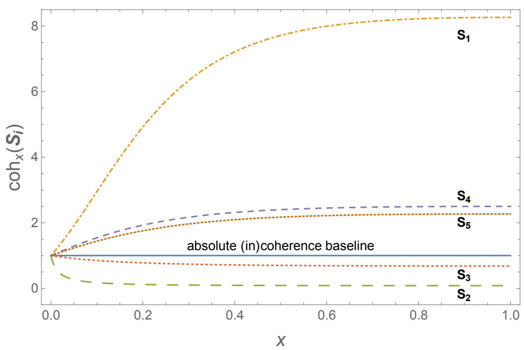

Figure 5: Coherence for $\mathbf{S}_1$ (orange, dash-dotted), $\mathbf{S}_2$ (green, dashed), $\mathbf{S}_3$ (red, dotted), $\mathbf{S}_4$ (violet, dashed), and $\mathbf{S}_5$ (dark orange, dotted) as a function of x.

To get a better feel of the new measure, it is helpful to take a look at an example from Bovens and Hartmann (2003):

A corpse in Tokyo. Suppose that we are trying to locate a corpse from a murder somewhere in Tokyo. We draw a grid of 100 squares over the map of the city and consider it equally probable that the corpse lies somewhere within each square. We interview two partially and equally reliable witnesses.

Consider the following situations $\text{sit}_i, i \in \{1,\ldots,5\}$, in which these two witnesses report that the corpse is to be found somewhere in the following numbered squares:

	sit_1	sit_2	sit_3	sit_4	sit_5
$\text{R}_{i.1}$	50-60	22-55	20-61	41-60	39-61
$\text{R}_{i.2}$	51-61	55-90	50-91	51-70	50-72

Bovens and Hartmann (2003) expect that $\mathbf{S}_1 = \{\text{R}_{1.1}, \text{R}_{1.2}\}$ should be more coherent than $\mathbf{S}_2 = \{\text{R}_{2.1}, \text{R}_{2.2}\}$ or $\mathbf{S}_3 = \{\text{R}_{3.1}, \text{R}_{3.2}\}$. $\mathbf{S}_4 = \{\text{R}_{4.1}, \text{R}_{4.2}\}$ and $\mathbf{S}_5 = \{\text{R}_{5.1}, \text{R}_{5.2}\}$ should have similar degrees of coherence. Our new measure masters the case with ease (see Figure 5): it judges $\mathbf{S}_1$ as highly coherent, $\mathbf{S}_2$ as highly incoherent and $\mathbf{S}_3$ as mildly incoherent. $\mathbf{S}_4$ and $\mathbf{S}_5$ are judged as absolutely coherent to similar degrees.

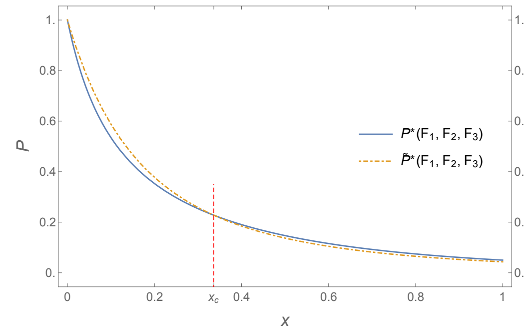

Figure 6: An example of the posterior probability functions P^* and $\tilde{P}^*$ for a randomly generated set $\mathbf{S} = \{\text{F}_1, \text{F}_2, \text{F}_3\}$. Note the criss-crossing at $x \approx .35$.

Discussion

All coherence measures proposed so far assume that it can either always be decided whether an information set is absolutely coherent (or absolutely incoherent), or never. For example, the Shogenji measure always allows us to make this decision, and the Olsson-Glass measure and the Bovens and Hartmann measure never allow us to make this decision. However, the examples given in the Introduction now suggest that some information sets should be judged as (absolutely) coherent or incoherent and others not. In the Tweety example, for instance, the set $\mathbf{S}$ seems to be incoherent and the set $\mathbf{S}'$ coherent. However, in the other example ("The culprit was wearing Coco Chanel shoes" etc.), it appeared that the set $\mathbf{S}'$ was neither coherent nor incoherent in the absolute sense. Some evidence pointed in this direction, and other evidence in the other direction.

This discussion suggests that an appropriate measure of coherence should allow us to judge *some* information sets as (absolutely) coherent or incoherent, while in other cases it should lead to the judgment that it is not a fact of the matter whether the particular set is (absolutely) coherent or incoherent. It turns out that the new measure allows exactly this. To illustrate this, let us consider the posterior probabilities $P^*(\mathbf{S})$ and $\tilde{P}^*(\mathbf{S})$ for a concrete example. See Fig. 6. We can see that the two curves cross at a certain value x_c of x. What should we conclude from this about the coherence of $\mathbf{S}$? If one focuses on values of $x > x_c$, one would come to the conclusion that $\mathbf{S}$ is absolutely coherent, since in this range $P^*(\mathbf{S}) > \tilde{P}^*(\mathbf{S})$. However, the situation is reversed if one considers values $x < x_c$. Here $P^*(\mathbf{S}) < \tilde{P}^*(\mathbf{S})$ and one would conclude that $\mathbf{S}$ is absolutely incoherent. We therefore get a contradiction. Note, however, that the judgment about whether an information set is (absolutely) coherent or not should not depend on the "dummy" variable x. It should only be determined by the "internal" properties of $\mathbf{S}$. We therefore conclude that in such cases there is no case of the matter whether the information set $\mathbf{S}$ is coherent or not. The following proposition provides a sufficient condition for the existence of a crisscrossing point.

Proposition 2. *Let **S** be an information set with a probability distribution P defined over it. P is characterized by the values $c_i^{(n)}$ for $i = 0,\dots,n$ as specified in Proposition 1. Then there is no fact of the matter as to whether **S** is coherent or incoherent if (i) $1 < a_0/\tilde{a}_0 < c_{n-2}^{(n)}/\tilde{c}_{n-2}^{(n)}$ or (ii) $1 > a_0/\tilde{a}_0 > c_{n-2}^{(n)}/\tilde{c}_{n-2}^{(n)}$.*

This is an interesting result which shows that even information sets with a considerable amount of dependencies may not be coherent in an absolute sense.

Conclusions

In science and in ordinary reasoning and argumentation, we often use coherence considerations to help us navigate an uncertain world. In doing so, we judge some information sets to be absolutely coherent or incoherent, and we judge some information sets to be more coherent than other sets. This paper has been concerned with clarifying the theoretical foundations for this practice. More specifically, we have argued that it is not always a fact of the matter as to whether a given information set is absolutely coherent or incoherent, and as to whether one information set is more coherent than another set. These claims have been motivated by examples, and we have shown that our theoretical account, in particular the new posterior-based coherence measure , can ground these claims.

In our further research, it will be important to better understand the conditions under which a set of information is absolutely coherent or incoherent. To this end, we also want to confront our theoretical findings with further case studies. In addition, we plan to conduct psychological experiments to compare our theoretical models with the way participants use coherence considerations in their reasoning.

Appendix

Proof of Proposition 1

Bovens and Hartmann (2003, Ch. 1) showed already that

$$P^*(\mathbf{S}) = \frac{a_0}{\sum_{i=0}^n a_i x^i}.$$

Let us now consider the denominator of the expression:

$$\begin{aligned}
\sum_{i=0}^n a_i x^i &= a_0 + \left(1 - a_0 - \sum_{i=1}^{n-1} a_i\right) x^n + \sum_{i=1}^{n-1} a_i x^i \\
&= a_0 + \overline{a_0}\, x^n + \sum_{i=1}^{n-1} a_i\,(x^i - x^n) \\
&= a_0 + \overline{a_0}\, x^n + x\overline{x} \sum_{i=1}^{n-1} a_i \sum_{j=i-1}^{n-2} x^j \\
&= a_0 + \overline{a_0}\, x^n + x\overline{x} \sum_{j=0}^{n-2} \left(\sum_{i=1}^{j+1} a_i\right) x^j \\
&= a_0 + \overline{a_0}\, x^n + x\overline{x} \sum_{i=0}^{n-2} b_i x^i, \qquad (3)
\end{aligned}$$

with $b_i := \sum_{j=1}^{i+1} a_j$. Multiplying relevant terms on the right-hand side of Eq. (3) with appropriate powers of $1 = x + \overline{x}$), we obtain

$$\sum_{i=0}^n a_i x^i = a_0\,(x+\overline{x})^n + \overline{a_0}\, x^n + x\overline{x} \sum_{i=0}^{n-2} b_i x^i\,(x+\overline{x})^{n-2-i}.$$

Expanding the binomials and some regrouping yields

$$\sum_{i=0}^n a_i x^i = a_0 \overline{x}^n + x^n + a_0 \sum_{k=1}^{n-1} \binom{n}{k} x^k \overline{x}^{n-k} + \psi_n,$$

with

$$\begin{aligned}
\psi_n &= \sum_{i=0}^{n-2} \sum_{j=0}^{n-2-i} b_i \binom{n-2-i}{j} x^{i+j+1}\, \overline{x}^{n-i-j-1} \\
&= \sum_{i=0}^{n-2} \sum_{k=i+1}^{n-1} b_i \binom{n-2-i}{k-1-i} x^k \overline{x}^{n-k} \\
&= \sum_{k=1}^{n-1} \sum_{i=0}^{k-1} b_i \binom{n-2-i}{n-1-k} x^k \overline{x}^{n-k} \\
&= \sum_{k=1}^{n-1} \left(\sum_{i=0}^{k-1} \sum_{j=1}^{i+1} a_j \binom{n-2-i}{n-1-k}\right) \cdot x^k \overline{x}^{n-k} \\
&= \sum_{k=1}^{n-1} \left(\sum_{j=1}^{k} a_j \sum_{i=j-1}^{k-1} \binom{n-2-i}{n-1-k}\right) \cdot x^k \overline{x}^{n-k} \\
&= \sum_{k=1}^{n-1} \left(\sum_{j=1}^{k} a_j \sum_{m=0}^{k-j} \binom{n-1-j-m}{n-1-k}\right) \cdot x^k \overline{x}^{n-k}.
\end{aligned}$$

Hence,

$$\sum_{i=0}^n a_i x^i = a_0 \overline{x}^n + x^n + \sum_{k=1}^{n-1} c_k x^k \overline{x}^{n-k},$$

with

$$c_k = a_0 \binom{n}{k} + \sum_{j=1}^{k} a_j \sum_{m=0}^{k-j} \binom{n-1-j-m}{n-1-k},$$

for $k = 1,\dots,n-1$. If we set $c_0 = a_0$ and $c_n = 1$, then we find

$$\sum_{i=0}^n a_i x^i = \sum_{k=0}^n c_k x^k \overline{x}^{n-k}.$$

A combinatorial argument then shows that the expressions c_k (for $k = 0,\dots,n-1$) are the sums of the joint probabilities of exactly k (true) propositions. □

Proof of Proposition 2

We have to find the values of x for which $P^*(\mathbf{S}) = \tilde{P}^*(\mathbf{S})$. Using Proposition 1, we obtain the following equation for x:

$$\phi_k^{(n)}(x) := \delta_0 x^{n-1} + \delta_0 s\, x^{n-2}\,\overline{x} + \sum_{k=1}^{n-2} \delta_k x^{k-1}\, \overline{x}^{n-k} = 0,$$

where we have set $\delta_0 := a_0 - \tilde{a}_0$, $\delta_k := a_0\,\tilde{c}_k^{(n)} - \tilde{a}_0\, c_k^{(n)}$ (for $k = 1,\dots,n-2$) and $s := \sum_{i=1}^n P(\mathrm{F_i})$. Note that $\phi_k^{(n)}(0) = c_{n-2}^{(n)}$ and $\phi_k^{(n)}(1) = \delta_0$. Hence, a criss-crossing occurs if either (i) $\delta_0 > 0$ and $c_{n-2}^{(n)} < 0$ or (ii) $\delta_0 < 0$ and $c_{n-2}^{(n)} > 0$. Note that these are only sufficient conditions. The conditions specified in the proposition follow immediately. □

Acknowledgments

This work was supported by the Arts and Humanities Research Council and The Deutsche Forschungsgemeinschaft [grant numbers HA 3000/20-1, HA 3000/21-1]. Thanks also to anonymous referees for their helpful comments.

References

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

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

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* (pp. 177–182). Berlin: Springer.

Hartmann, S., & Trpin, B. (2023). Why coherence matters. Forthcoming in: *Journal of Philosophy*.

Koscholke, J., Schippers, M., & Stegmann, A. (2019). New hope for relative overlap measures of coherence. *Mind*, *128*(512), 1261–1284.

Meijs, W. (2005). *Probabilistic Measures of Coherence*. Dissertation, Erasmus University, Rotterdam.

Olsson, E. J. (2002). What is the problem of coherence and truth? *Journal of Philosophy*, *99*(5), 246–272.

Olsson, E. J. (2022). *Coherentism*. Cambridge: Cambridge University Press.

Osimani, B., & Landes, J. (2023). Varieties of error and varieties of evidence in scientific inference. *The British Journal for the Philosophy of Science*, *74*(1), 117–170.

Shogenji, T. (1999). Is coherence truth conducive? *Analysis*, *59*(4), 338–345.

Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge, MA: MIT Press.

Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*, *74(1)*, 28–47.