Why Coherence Matters*

Stephan Hartmann† and Borut Trpin‡

November 23, 2023

Forthcoming in *The Journal of Philosophy*

Abstract

Explicating the concept of coherence and establishing a measure for assessing the coherence of an information set are two of the most important tasks of coherentist epistemology. To this end, several principles have been proposed to guide the specification of a measure of coherence. We depart from this prevailing path by challenging two well-established and prima facie plausible principles: Agreement and Dependence. Instead, we propose a new probabilistic measure of coherence that combines basic intuitions of both principles, but without strictly satisfying either of them. It is then shown that the new measure outperforms alternative measures in terms of its truth-tracking properties. We consider this feature to be central and argue that coherence matters because it is likely to be our best available guide to truth, at least when more direct evidence is unavailable.

KEYWORDS: coherence, epistemic principles, truth-tracking, Bayesian coherentism, formal epistemology

If an information set is highly coherent in the sense that its elements fit well together, then it is intuitively more plausible than a set where this is not the case.[1] To illustrate this, consider the following scenario: Study 1 reports that glaciers are melting, Study 2 reports that global sea levels are rising, and Study 3 reports that sea surface temperature is rising. In another scenario, Study 1 also reports glaciers melting, while Study 2 reports global sea level falling, and Study 3 reports sea surface temperature remaining more or less constant. Obviously, the three studies mentioned in the first scenario are coherent, while in the second scenario, according to everything else we know (that is, taking into account our scientific background knowledge), they are in some tension with each other. Now, the higher degree of coherence of the three studies in the first scenario is not a guarantee of their truthfulness, and yet their higher coherence makes the studies in the first scenario seem more plausible overall than the studies in the second scenario.

While such coherence considerations are regularly used in everyday reasoning (see, for example, Harris and Hahn, 2009 and Hahn et al., 2016), it is debatable whether they also have normative epistemological significance. To resolve this issue, we need to be clear about what coherence refers to, as its definitions often remain vague and the use of the concept is not uniform (Olsson, 2022). Furthermore, the question arises whether rational degrees of belief should

---

*Both authors contributed equally and are listed in alphabetical order. We would like to thank Ulrike Hahn and the other members of our research group—Leon Assaad, Rafael Fuchs, Ammar Jalalimanesh, Anita Keshmirian, Leon Schöppl, and Corina Strößner—for their comments, which helped us shape this paper. Thanks also to Christopher von Bülow for his help in preparing the final version. We also gratefully acknowledge helpful feedback from audiences at the Epistemic Justification conference at LMU Munich, the SILFS Triennial Conference 2023 in Urbino, and the Working Group in History and Philosophy of Logic, Mathematics and Science, University of California at Berkeley, where we presented earlier versions of this paper. Finally, we would like to thank the German Research Foundation (DFG, projects 448424181 and 455912038) and the Humboldt Foundation (postdoctoral fellowship; BT) for their support.

†Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://www.stephanhartmann.org – s.hartmann@lmu.de.

‡Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – https://boruttrpin.weebly.com – Borut.Trpin@lrz.uni-muenchen.de.

[1] We follow Bovens and Hartmann (p. 10 2003) in taking coherence to be a property of information *sets* and not, for example, of propositions. Formally, if we obtain the information items $R_1, \ldots, R_n$ from n independent and partially reliable sources, then $\mathbf{S} = \{R_1, \ldots, R_n\}$ is an information set over which a (subjective) probability distribution is defined. Throughout the paper we assume that the prior probability of information items is in the open interval $(0, 1)$.

be oriented to the degree of coherence of the corresponding information sets. To answer such questions we need a way to quantify coherence. The literature in formal epistemology provides a number of probabilistic measures of coherence that are supposed to do just this (see Shogenji, 1999; Glass, 2002, Olsson, 2002; Fitelson, 2003; Bovens and Hartmann, 2003; Douven and Meijs, 2007; Schupbach, 2011; Koscholke et al., 2019). It turns out, however, that the various proposed measures may differ fundamentally in their assessment. We therefore also need to identify which measure is most appropriate for determining the normative role of coherence considerations.

Three types of arguments are used for this purpose. First, the proposed measures are confronted with test cases for which we have a clear intuition (for a survey, see Koscholke (2016)). Unfortunately, these test cases usually involve only information pairs and triples, as it is difficult to develop reliable intuitions for larger information sets. Second, empirical studies are conducted to determine which coherence measure best represents our coherence intuitions (Harris and Hahn, 2009; Koscholke and Jekel, 2017). In addition to the controversial is-to-ought inference (Elqayam and Evans, 2011), the available empirical results also cannot be used as a normative guide because they are too diverse. Third, one can point to plausible normative principles that a proposed coherence measure satisfies. Interestingly, it turned out that the two most important normative principles, AGREEMENT and DEPENDENCE, are mutually exclusive (Schippers, 2014). We will examine these two principles in more detail below and argue that while they have some plausibility for smaller information sets, they are too strict for larger information sets. Finally, we will argue that the best way to evaluate a proposed coherence measure is to show that it serves a desirable function, namely that of helping us figure out which information sets contain true information.

In the next section, we introduce the principles of AGREEMENT and DEPENDENCE, and use counterexamples to show that both principles are generally implausible as well as largely inapplicable. If nothing else, this is bad news for the relative overlap measures currently discussed in the literature, as we will show in Section 2. In Section 3, we introduce a new relative overlap measure and show that it satisfies DEPENDENCE for small information sets, which we also consider desirable from a normative point of view. In the following Section 4, we will then argue that we should generally evaluate coherence measures according to how well they fulfill a particular function. To do this, we examine the truth-tracking properties of coherence measures and show that our new measure is a particularly good indicator of truth. We conclude with some thoughts on the wider implications of our results (Section 5).

1 Agreement and Dependence

It is clear that not all information sets are equally coherent. For instance, there is an obvious difference between the degrees of coherence of the information sets $\mathbf{S} = \{$"This man speaks Italian", "This man owns a copy of Dante's *La Divina Commedia*"$\}$ and $\mathbf{S}' = \{$"This man speaks Italian", "This man owns a copy of the US Declaration of Independence"$\}$. The information items in $\mathbf{S}$ fit more strongly together than those in $\mathbf{S}'$. Any adequate measure of coherence should reflect this. However, once we consider larger information sets, the intuitions are not as clear. For instance, compare the information sets $\mathbf{S}'' = \{$"This man speaks Italian", "This man owns a copy of *La Divina Commedia*", "This man is a linguist", "This man is from South Korea"$\}$ and $\mathbf{S}''' = \{$"This man speaks Italian", "This man owns a copy of the US Declaration of Independence", "This man is a linguist", "This man is from Italy"$\}$. It is not clear which of the two is more coherent. Some authors even claim that it cannot always be determined whether one information set is more coherent than another (see, for example, Bovens and Hartmann, 2003).

Instead of referring to intuitions about specific test cases, an alternative approach to determining which measure of coherence is the most adequate relies on normative principles that have a certain intuitive appeal. The idea is that any adequate measure of coherence should satisfy these principles. For instance, consider the so-called Principle of Agreement (hereafter simply AGREEMENT). The principle goes back to Bovens and Olsson (2000) and has been revived by, for example, Schippers (2014) and Koscholke et al. (2019). It roughly states that increasing conditional probabilities of information items given the other information items from the considered information set should increase the coherence because there is then more mutual support. Here is a precise definition (following Koscholke et al., 2019):

Definition 1. (AGREEMENT) *Let us assume that the following inequality holds for all non-empty disjoint subsets* $\mathbf{S}'$ *and* $\mathbf{S}''$ *of an information set* $\mathbf{S}$ *for two probability distributions* P_1 *and* P_2*:*

$$P_1\left(\bigwedge_{s_j\in\mathbf{S}'} s_j \,\middle|\, \bigwedge_{s_m\in\mathbf{S}''} s_m\right) > P_2\left(\bigwedge_{s_j\in\mathbf{S}'} s_j \,\middle|\, \bigwedge_{s_m\in\mathbf{S}''} s_m\right)$$

Given a coherence measure Coh, we say that it satisfies AGREEMENT *if it also holds that* $Coh_{P_1}(\mathbf{S}) > Coh_{P_2}(\mathbf{S})$*, where the subscripts* P_1 *and* P_2 *refer to the two probability distributions.*

Another principle that has an intuitive appeal is the Principle of Dependence (hereafter DEPENDENCE). Simply put, DEPENDENCE states that the coherence of an information set is above (below) a certain threshold if the information it contains is positively (negatively) correlated. To make the principle more precise, some definitions are in order:

Definition 2. *A probability distribution* P *is defined over a set of propositional variables* $V{:=}\{H_1,\ldots,H_n\}$ *with the values* H_i *and* $\neg\mathrm{H}_i$ *for all* $i = 1,\ldots,n$*.*

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

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

Thus, the principle can be formulated as follows (again following Koscholke et al., 2019):

Definition 3. (DEPENDENCE) *Given a coherence measure Coh, we say that it satisfies* DEPENDENCE *if there is a threshold* τ *such that for any information set* $\mathbf{S}$*:*

- $Coh(\mathbf{S}) > \tau$ *if* $\mathbf{S}$ *is positively correlated,*
- $Coh(\mathbf{S}) = \tau$ *if* $\mathbf{S}$ *is independent,*
- $Coh(\mathbf{S}) < \tau$ *if* $\mathbf{S}$ *is negatively correlated.*

Before proceeding, we note that DEPENDENCE refers to a threshold τ which can be used to define absolute (or categorical) coherence and absolute (or categorical) incoherence: an information set $\mathbf{S}$ is absolutely coherent if $Coh(\mathbf{S}) > \tau$ and absolutely incoherent if $Coh(\mathbf{S}) < \tau$. (For the Shogenji measure, $\tau = 1$, and for the Fitelson measure, $\tau = 0$.) These are useful terms, whose further investigation we leave open here. Instead, we examine how the two principles are related.

In a seminal article, Koscholke et al. (2019) state (in their Theorem 5) that AGREEMENT and DEPENDENCE are mutually exclusive: any measure that satisfies one principle, they claim, cannot satisfy the other. Their proof refers to Schippers (2014), who shows that another principle, which he calls INDEPENDENCE and which is weaker than DEPENDENCE, is inconsistent with AGREEMENT for $n = 2$. It should be noted, however, that this result allows for the possibility that AGREEMENT and DEPENDENCE are compatible for $n \geq 3$, which is the more interesting case anyway. While this conjecture should be investigated further, for now we accept the incompatibility of the two principles.[2] This would be bad news, since both principles are intuitively appealing to begin with. Koscholke, Schippers, and Stegmann accordingly advocate a pluralistic position according to which we should, for each principle, find the best measures that satisfy it. We do not think this is a good proposal because we do not think that AGREEMENT is an acceptable condition for coherence measures. To this end, we consider the following counterexample.

[2] At this point, it is helpful to note that increased mutual conditional probabilities are usually accompanied by increased correlation. However, AGREEMENT and DEPENDENCE are defined as *general principles*, so they must hold in all cases (or at least for all cases at a certain cardinality of the information sets under consideration). Since there are pairs of probability functions over a given information set where increased conditional probability can change the correlation from positive to independent or even to negative, it follows that the two principles are inconsistent (at least for $n = 2$). This is because the two principles suggest inconsistent verdicts in these cases.

PARTY PARTICIPATION: There is a large party tonight. Now consider two situations:

1. Mr A is very likely to attend the party. Ms B also plans to attend it. However, she tends to avoid A. So, if A is not going to be at the party, B is very likely to join it. If A goes, however, B may have second thoughts.

   Suppose that the probability distribution is defined by the following three values: $P(\mathrm{A}) = P(\mathrm{B} \mid \neg\mathrm{A}) = .9$ and $P(\mathrm{B} \mid \mathrm{A}) = .6$, where A and B represent Mr A and Ms B attending the party, respectively.

2. It is unlikely that Mr A attends the party. If he does not attend it, then Ms B is also unlikely to attend it. However, she enjoys A's company, so if A is going, B might change her mind.

   Let us assume that the probability distribution is defined by the following three values: $P(\mathrm{A}) = P(\mathrm{B} \mid \neg\mathrm{A}) = .1$ and $P(\mathrm{B} \mid \mathrm{A}) = .5$.

In both cases, two independent witnesses each give the reports

R_1: Mr A attended the party.

R_2: Ms B attended the party.

In which situation do the two reports fit together better, 1 or 2?

The question is not in which case the two reports are more likely, but rather in which situation the two reports fit together better or, equivalently, in which situation there is less tension between them. Taking this into account, it is rather clear that the two reports R_1 and R_2 fit together better in situation 2. In the first version, both Mr A and Ms B are likely to attend the party, but they are negatively correlated: if Mr A goes, Ms B is less likely to go too. In situation 2, on the other hand, Mr A's attendance encourages Ms B's, and as DEPENDENCE requires, the reports are therefore more coherent.

However, any measure that respects AGREEMENT will give an opposite response because all the mutual conditional probabilities are greater in situation 1 than in situation 2: $P_1(\mathrm{A} \mid \mathrm{B}) = .86 > P_2(\mathrm{A} \mid \mathrm{B}) = .36$ and $P_1(\mathrm{B} \mid \mathrm{A}) = .6 > P_2(\mathrm{B} \mid \mathrm{A}) = .5$. So, according to these measures, situation 1 is the more coherent one. We believe that this is clearly wrong, because in situation 1 there is more tension between A and B than in situation 2, which suggests that AGREEMENT is a problematic principle and that the same holds for any measure that satisfies it.

It also seems that AGREEMENT (or consideration of mutual conditional probabilities) does not really have much to do with the concept of coherence. Take BonJour (1985, p. 93) as a classical reference for what it means for an information set to be coherent:

> It is reasonably clear that this "hanging together" depends on the various sorts of inferential, evidential, and explanatory relations which obtain among the various members of a system of beliefs, and especially on the more holistic and systematic of these.

However, AGREEMENT is not related to any relevant aspects of inferential, evidential or explanatory relations. The principle only considers conditional probabilities of various subsets, so it does not capture relevant inferential relations between information items because it does not measure whether some information item may be inferred from another. It also does not capture evidential relations, because we cannot tell whether some information items provide evidence for each other just by looking at the relations among conditional probabilities. Finally, it is also clear that AGREEMENT does not capture the explanatory relations in an information set, as these do not depend on conditional probabilities of various subsets. We therefore suggest that AGREEMENT should not be seen as a desideratum for measures of coherence and that any measure which satisfies it should be rejected.

Since DEPENDENCE corresponds to evidential relations, as it captures how information items are mutually confirmatory, and since the principles AGREEMENT and DEPENDENCE are known to be mutually exclusive, we still have the possibility of motivating coherence measures that satisfy DEPENDENCE. However, it turns out that DEPENDENCE is too strict for larger information sets and should thus not be a general desideratum for an acceptable coherence measure. This is essentially because there are information sets in which all propositions are positively correlated but have almost no overlap. Without a sufficient amount of overlap, however, there can be no "hanging together" in the first place, and thus no coherence.

We can show the issue with the following example. Suppose that there is a town where it rains frequently. Let R represent that it is raining in the town, let B represent that a person in the town is reading a book, and let C represent that a person in the town is wearing a raincoat. Suppose that the probability of R (rain) is .5. The probability of B (reading a book) given R is .2, meaning that when it rains, 20 % of people are expected to read books. The probability of B given not-R (no rain) is .1 because fewer people read books on a non-rainy day. The probability of C (wearing a raincoat) given R is .9 and only .3 given not-R. Assuming that the propositional variable R probabilistically screens off the propositional variable B from the propositional variable C (or, in causal language, R is the common cause of B and C), we have enough information to calculate the joint probability distribution over all three variables.

Note that the set $\mathbf{S} = \{\mathrm{R}, \mathrm{B}, \mathrm{C}\}$ is positively dependent, although there is very little relative overlap of the three information items because most people of the town do not read books regardless of the weather or what they are wearing. Hence, it is not clear whether the set $\mathbf{S}$ should be considered as absolutely coherent. The set $\mathbf{S}' = \{\mathrm{R}, \neg\mathrm{B}, \mathrm{C}\}$ (rain, not reading a book, but wearing a raincoat) has, after all, a much higher degree of relative overlap.

More generally, both DEPENDENCE and AGREEMENT may be criticized because they consider subsets of an information set (see Olsson's (2022, pp. 51-57) similar argument against subset measures of coherence). The conditions specified in Definition 1 (for the latter) and in Definition 2 (for the former principle), after all, refer to subsets of a given information set. Consequently, these principles only provide guidance for cases where all non-empty subsets of an information set are correlated in the same way (DEPENDENCE), or for cases where the conditional probabilities of information conjuncts from specifically defined subsets are all greater under one probability function than under another (AGREEMENT). For the vast majority of sets, these conditions do not hold, as they are very demanding, and the principles do thus not apply.

For instance, if we use a few lines of code to generate 10,000 random probability distributions for variously-sized information sets we find the following: For sets with three information items, DEPENDENCE only applies in 2400 cases. For sets with four information items, the number of cases where it is applicable drops to 227, and further to nine for sets with five, and all the way down to zero for sets with six information items. This is because the larger a random information set is, the more likely it is that it is neither independent nor positively/negatively correlated as defined by the conditions of Definition 2. Notably, any information pair is either independent or (positively/negatively) correlated. This provides a case in point of using DEPENDENCE as a normative principle of coherence for information pairs.

On the other hand, AGREEMENT involves a comparison of two probability distributions, so we assume that it provides normative guidance in even fewer cases than DEPENDENCE. In fact, AGREEMENT does not even cover all information pairs. For instance, consider an information set $\mathbf{S} = \{A, B\}$ and two probability distributions P_1 and P_2 such that[3] $P_1(\mathrm{A}, \mathrm{B}) = P_2(\mathrm{A}, \mathrm{B})$, but $P_1(\mathrm{A}, \neg\mathrm{B}) = P_2(\neg\mathrm{A}, \mathrm{B})$ and $P_1(\neg\mathrm{A}, \mathrm{B}) = P_2(\mathrm{A}, \neg\mathrm{B})$. It is easy to see that in this case $P_1(\mathrm{A} \mid \mathrm{B}) \gtreqless P_2(\mathrm{A} \mid \mathrm{B})$, but $P_1(\mathrm{B} \mid \mathrm{A}) \lesseqgtr P_2(\mathrm{B} \mid \mathrm{A})$, so AGREEMENT does not apply.

Moreover, in cases of information pairs where AGREEMENT applies, it sometimes provides the intuitively wrong assessment (recall the PARTY PARTICIPATION example). DEPENDENCE, on the other hand, always provides reasonable guidance for information pairs but increasingly fails to apply when larger information sets are considered. In summary, the two principles are unlikely to be of much use in practice.

2 Relative Overlap Measures

Let us now consider the class of relative overlap measures. The idea underlying these measures is that an inconsistent information set, whose joint probability is therefore zero, is maximally incoherent, while an information set consisting wholly of equivalent propositions is maximally coherent. The coherence of an information set then corresponds to the degree of relative overlap of the propositions in the probability space (see Figure 1). These considerations motivate the following simple measure of relative overlap, $\mathrm{Coh}_{\mathrm{OG}}(\mathbf{S})$, of an information set $\mathbf{S} = \{\mathrm{H}_1, \ldots, \mathrm{H}_n\}$,

[3] When appropriate, we use the convention of representing the conjunction $\mathrm{A} \wedge \mathrm{B} \wedge \ldots$ as $\mathrm{A}, \mathrm{B}, \ldots$.

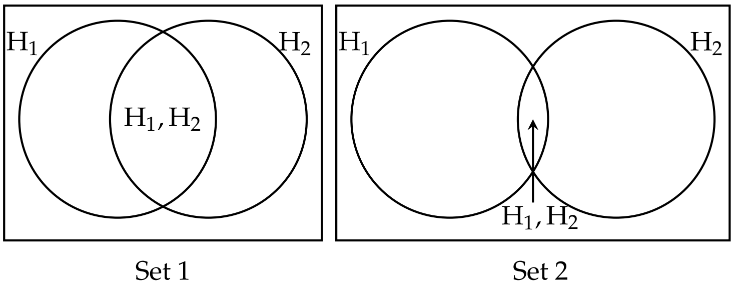

Figure 1: The idea behind relative overlap measures: set 1 is more coherent than set 2 because the two information items—H_1 and H_2—overlap more in set 1 than in set 2.

independently proposed by Olsson (2002) and Glass (2002). It is given by

$$\mathrm{Coh}_{\mathrm{OG}}(\mathbf{S}) := \frac{P(\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_n)}{P(\mathrm{H}_1 \vee \mathrm{H}_2 \vee \cdots \vee \mathrm{H}_n)} \tag{1}$$

This measure has some initial plausibility. It has also been shown by Koscholke et al. (2019) that it satisfies AGREEMENT for $n = 2$. Furthermore, when restricted to information pairs, such as when considering the two-place relationship between explanans and explanandum, it may be fruitfully used to rank explanations and to guide inference to the best explanation (Glass, 2012). Numerical studies suggest that it also satisfies AGREEMENT for $n > 2$, but a strict proof is still lacking. Besides the already mentioned criticism of AGREEMENT, this measure faces other serious objections.

Perhaps the most serious problem takes its cue from the observation that the joint probability of all propositions (that is, the numerator in eq. (1)) can never increase when another proposition is added. Similarly, the probability of a disjunction (that is, the denominator in eq. (1)) cannot decrease when another proposition is added. Consequently, when measured by $\mathrm{Coh}_{\mathrm{OG}}$, the coherence of an information set cannot be increased by adding new information. The following counterexample shows that this is counterintuitive (from Bovens and Hartmann, 2003). Consider the information pair $\mathbf{S} = \{\mathrm{B}, \mathrm{G}\}$ with B: "Tweety is a bird" and G: "Tweety is a ground-dweller". Obviously, $\mathbf{S}$ is not very coherent, since almost all birds can fly. An overlap measure like $\mathrm{Coh}_{\mathrm{OG}}$ captures this correctly, since there is not much overlap between ground-dwelling animals and birds. However, adding the proposition P: "This animal is a penguin" should increase the coherence, since all penguins are ground-dwelling birds, so the three propositions fit together very well. That is, the coherence of $\mathbf{S}' = \{\mathrm{B}, \mathrm{G}, \mathrm{P}\}$ should be greater than that of $\mathbf{S}$. Unfortunately, $\mathrm{Coh}_{\mathrm{OG}}$ cannot provide this result, because the joint probability of B, G, and P (in the numerator) cannot be larger than the joint probability of B and G alone, for purely mathematical reasons, while the probability of the disjunction $\mathrm{B} \vee \mathrm{G} \vee \mathrm{P}$ (in the denominator) cannot be smaller than the probability of $\mathrm{B} \vee \mathrm{G}$, also for purely mathematical reasons. Thus, according to $\mathrm{Coh}_{\mathrm{OG}}$, adding P does not increase coherence, although "Tweety is a penguin" reduces the tension between "Tweety is a bird" and "Tweety is a ground-dweller". This and similar counterexamples were considered so damaging to the Olsson–Glass measure that it was largely abandoned (note, however, that Olsson (2022) defends $\mathrm{Coh}_{\mathrm{OG}}$ as a measure of agreement, not coherence).

The aforementioned problem can be avoided if, instead of using only the relative overlap of the entire information set $\mathbf{S}$, that is, $\mathrm{Coh}_{\mathrm{OG}}(\mathbf{S})$, one also considers the relative overlap of all non-empty non-singleton subsets of $\mathbf{S}$ and then takes the average of all these values (see Meijs, 2005, 2006). This yields the following measure $\mathrm{Coh}_{\mathrm{OG}'}$:

$$\mathrm{Coh}_{\mathrm{OG}'}(\mathbf{S}) := \frac{1}{m} \sum_{i=1}^{m} \mathrm{Coh}_{\mathrm{OG}}(\mathbf{S}'_i) \tag{2}$$

where $\mathbf{S}'_i$ represents any of the m non-empty non-singleton subsets of the information set $\mathbf{S}$ and $m := (2^n - n) - 1$. This measure resolves the Tweety counterexample we mentioned earlier because, contrary to $\mathrm{Coh}_{\mathrm{OG}}$, the coherence of an information set may now increase, decrease, or remain unchanged when we add new information. Alas, the measure $\mathrm{Coh}_{\mathrm{OG}'}$ suffers from a formal limitation. It cannot judge a given information set $\mathbf{S}$ as more coherent than its most coherent two-element subset (see Koscholke and Schippers, 2016). This is a persistent issue: even if the relative overlaps of subsets are averaged in some other way, the issue still remains.

Koscholke et al. (2019) therefore propose a new measure to overcome these problems. Their measure takes the idea of relative overlap and combines it with the principle of average mutual support (Douven and Meijs, 2007); that is, to judge how coherent an information set is, we should take a look at how much all information items in subsets of the information set under consideration support each other. The idea behind the measure is that we first consider all pairs $\mathbf{S}'$ and $\mathbf{S}''$ of non-empty disjoint subsets of $\mathbf{S}$. Then we take the conjunction of the information items in these subsets and use the simple measure $\mathrm{Coh}_{\mathrm{OG}}$ to measure how much relative overlap there is between them. In other words, to assess how coherent an information set $\mathbf{S}$ is, we need to consider the average relative overlap of non-empty disjoint subsets of conjunctions in $\mathbf{S}$. Formally:

$$\mathrm{Coh}_{\mathrm{OG}^*}(\mathbf{S}) \ := \ \frac{1}{k}\sum_{i=1}^{k}\mathrm{Coh}_{\mathrm{OG}}\left(\bigwedge_{s_j\in\mathbf{S}'} s_j, \bigwedge_{s_m\in\mathbf{S}''} s_m\right)_i \tag{3}$$

where $k := \left[(3^n - 2^{n+1}) + 1\right]/2$ and $\mathbf{S}'$ and $\mathbf{S}''$ are subsets as described above.

This measure is relatively complicated, but that may be just the price we have to pay to get a satisfactory solution. At first glance, it seems to do just that: it avoids the weaknesses of the two relative overlap measures mentioned earlier, $\mathrm{Coh}_{\mathrm{OG}}$ and $\mathrm{Coh}_{\mathrm{OG}'}$. It also provides the intuitively expected judgment in some standard test cases from the literature and satisfies the principles one might expect from coherence measures. In particular, it satisfies AGREEMENT in general (Koscholke et al., 2019, p. 1272).

Before continuing, it needs to be made clear that all three measures of relative overlap, $\mathrm{Coh}_{\mathrm{OG}}$, $\mathrm{Coh}_{\mathrm{OG}'}$, and $\mathrm{Coh}_{\mathrm{OG}^*}$, are equivalent when information pairs are considered. Moreover, it can easily be proven that they satisfy AGREEMENT for information pairs (Koscholke et al., 2019, p. 1272). However, although AGREEMENT is praised as an important principle, it is an open question whether $\mathrm{Coh}_{\mathrm{OG}}$ and $\mathrm{Coh}_{\mathrm{OG}'}$ satisfy the principle for sets with three or more information items (Schippers, 2014, p. 3830). As discussed earlier, we take AGREEMENT to be implausible, as the PARTY PARTICIPATION example and other identified issues from the previous section show. Yet, the fact that it is unsettled whether it is satisfied by any additional relative overlap measures (besides $\mathrm{Coh}_{\mathrm{OG}^*}$) further decreases the principle's importance.

Unfortunately, the measure $\mathrm{Coh}_{\mathrm{OG}^*}$ also falls short. The measure is defined as the average coherence of conjunctions from all respective subsets. Hence, for n information items, the measure averages over $\left[(3^n - 2^{n+1}) + 1\right]/2$ computations of relative overlaps (each computed using $\mathrm{Coh}_{\mathrm{OG}}$) in specifically defined subsets (Koscholke et al., 2019). This means that the computational load exponentially increases when the set under consideration increases and the measure is therefore computationally intractable if we were to use it for large information sets. The authors praise the measure for the fact that the amount of computation required is only half of what is required to compute coherence using the recipe of Douven and Meijs (2007). This is not much consolation, however, because the number of calculations required increases exponentially with the size of the information set. For example, to compute the coherence of an information set with ten information items, we need to perform 28,501 computations, and for twenty information items we already need to perform almost 2 billion computations (1,742,343,625, to be exact). The measure is therefore limited to information sets with low cardinality.

3 A New Relative Overlap Measure

All problems related to the so-far-mentioned measures of relative overlap seem to stem from the fact that they do not respect DEPENDENCE—they may judge an information pair (as in the PARTY PARTICIPATION example) or a triple of positively correlated information items as less coherent than another information set with negatively correlated information. Note also that although DEPENDENCE is plausible for smaller information sets (AGREEMENT is already implausible for information pairs), for larger n it is not clear that strict versions of AGREEMENT or DEPENDENCE make sense from a normative point of view. Requesting that either principle hold may be too strict a requirement.

For example, consider two information sets, each containing one million information items. The first information set is only slightly positively correlated, while the second is strongly positively correlated, but there is some information that is slightly negatively correlated with the rest of the information set. According to DEPENDENCE, the coherence of the first information set is higher. But is this plausible? Should not rather the strong positive correlation of the bulk of the second information set compensate for the (low) negative correlation of a few information items with the rest, so that in this case the second information set can be classified as more coherent than the first information set?

Note also that a measure of coherence may very well fail to satisfy either AGREEMENT or DEPENDENCE. It may also only satisfy either principle in a limited sense, for example, only for information sets of lower cardinality. What is clear, though, is that any admissible measure should satisfy DEPENDENCE at least for information pairs. If two information items disconfirm each other, then the information pair simply cannot be more coherent than another pair of mutually confirmatory information items. Our PARTY PARTICIPATION case from above shows that measures from the literature which satisfy AGREEMENT for information pairs fail in this regard. However, this does not mean that all measures which take into account the extent of relative overlap in one way or another suffer from this limitation.

To construct one such measure that avoids this problem, let us first consider the standard independence deviation (that is, correlational) measure of coherence, which was provided by Shogenji (1999):

$$\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S}) := \frac{P(\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_n)}{P(\mathrm{H}_1)P(\mathrm{H}_2)\cdots P(\mathrm{H}_n)} \tag{4}$$

Although this measure suffers from its own problems[4], it is easy to see that it satisfies DEPENDENCE. If the information set under consideration is positively (negatively) correlated, then the numerator is greater (less) than the denominator and the measure will judge an information set to have coherence above (below) the threshold value of 1. If it is independent, then the numerator and the denominator are equal and the amount of coherence is exactly at the threshold value of 1.

We also observe that $\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S})$ is defined as the ratio between the joint probability of the information items in the information set $\mathbf{S}$ and the probability of the same items if they were probabilistically independent and had the same marginal probabilities. We can generalize this construction principle to introduce independence deviation intuitions into a relative overlap measure. Before that, however, we need to introduce a new concept, which will prove useful in the further course.

Definition 4. *A probability distribution P is defined over a set of propositional variables $V := \{H_1, \ldots, H_n\}$. 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*) *$\tilde{P}(\mathrm{H}_i) = P(\mathrm{H}_i)$ for all $i = 1, \ldots, n$.*

The Shogenji measure of coherence of an information set $\mathbf{S}$ can then be written as

$$\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S}) = \frac{P(\mathbf{S})}{\tilde{P}(\mathbf{S})} \tag{5}$$

Note that one of the simplest prima facie measures of coherence is obtained by merely considering the joint probability of the information set $\mathbf{S}$, that is, $\mathrm{coh}_P^{(0)}(\mathbf{S}) := P(\mathbf{S})$ (Olsson, 2021).[5] This measure is not particularly convincing, but we can consider $\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S})$ as its improvement, which results from normalizing $\mathrm{coh}_P^{(0)}(\mathbf{S})$ by $\mathrm{coh}_{\tilde{P}}^{(0)}(\mathbf{S}) = \tilde{P}(\mathbf{S})$. Since we use the associated probability distribution $\tilde{P}$ in this expression, we can say that $\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S})$ is the coherence measure associated with $\mathrm{coh}_P^{(0)}(\mathbf{S})$. This example leads to the following definition, which provides a general construction recipe for including independence deviation intuitions in a coherence measure:

Definition 5. *Let* $\mathbf{S}$ *be an information set and P be a probability distribution defined over the corresponding set of propositional variables. Furthermore, let coh_P be a prima facie measure of coherence* (*relative*

[4]For some standard counterexamples see Fitelson, 2003; Bovens and Hartmann, 2003.

[5]Coherence measures are always relative to a probability distribution. To avoid confusion, this is made clear by adding a subscript where appropriate.

to P) and let $\tilde{P}$ be the associated probability measure. Then

$$\mathrm{Coh}_P(\mathbf{S}) := \frac{coh_P(\mathbf{S})}{coh_{\tilde{P}}(\mathbf{S})} \tag{6}$$

is the associated measure of coherence if $coh_{\tilde{P}}(\mathbf{S}) > 0$.

This approach is indeed a simple way to include the deviation from independence in a coherence measure, and it is easy to see that $\mathrm{Coh}_P(\mathbf{S}) = 1$ when $\mathbf{S}$ is independent. That is, the recipe always satisfies the principle of independence (Schippers, 2014): if an information set is independent, its degree of coherence is always at a certain baseline value τ (here: $\tau = 1$). Note also that the qualification $\mathrm{coh}_{\tilde{P}}(\mathbf{S}) > 0$ rules out certain measures of coherence, such as Fitelson's (2003) measure Coh_F, since its baseline value is 0. This problem can be simply solved by using $\mathrm{Coh}'_\mathrm{F} = \mathrm{Coh}_\mathrm{F} + 1$ instead, whose baseline value is 1.

Besides this limitation, the recipe remains open about what should be admitted as a prima facie measure. However, suppose that we use Coh_Sh as our prima facie measure of choice. Then the denominator $\mathrm{coh}_{\tilde{P}}(\mathbf{S})$ will always be 1 because 1 is the independence baseline for Coh_Sh and the obtained measure will simply reduce to Coh_Sh. The same holds for any measure that has 1 as its independence baseline (including, for example, Coh'_F). It is easy to see that the only genuine candidate measures will therefore be those that are not already compatible with the intuition of independence deviation mentioned above.

In addition, there are other requirements that we place on an admissible prima facie measure. One of these requirements is SYMMETRY: the coherence of an information set does not depend on the order in which the information items are presented. This requirement is automatically satisfied in our discussion, since we have assumed from the beginning that the argument of a coherence measure is an information set, and for sets it is always true that, for example, $\{\mathrm{A}, \mathrm{B}\} = \{\mathrm{B}, \mathrm{A}\}$. It should be noted, however, that the symmetry requirement rules out most confirmation measures as candidates for prima facie coherence measures (although symmetrized sums of them are still an option).

Let us now construct the associated measure of coherence from the Olsson–Glass measure Coh_OG. This new measure of coherence provides a promising compromise between the two main intuitions behind the notion of coherence—probabilistic relevance (that is, DEPENDENCE) and relative overlap (that is, AGREEMENT)—without strictly satisfying either of them. We obtain:

$$\begin{aligned}
\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}) \quad &:= \quad \frac{\mathrm{Coh}_{\mathrm{OG}_P}(\mathbf{S})}{\mathrm{Coh}_{\mathrm{OG}_{\tilde{P}}}(\mathbf{S})} && (7)\\
&= \quad \frac{P(\mathrm{H}_1, \ldots, \mathrm{H}_n)}{P(\mathrm{H}_1 \vee \cdots \vee \mathrm{H}_n)} \bigg/ \frac{\tilde{P}(\mathrm{H}_1, \ldots, \mathrm{H}_n)}{\tilde{P}(\mathrm{H}_1 \vee \cdots \vee \mathrm{H}_n)}\\
&= \quad \frac{P(\mathrm{H}_1, \ldots, \mathrm{H}_n)}{\tilde{P}(\mathrm{H}_1, \ldots, \mathrm{H}_n)} \cdot \frac{\tilde{P}(\mathrm{H}_1 \vee \cdots \vee \mathrm{H}_n)}{P(\mathrm{H}_1 \vee \cdots \vee \mathrm{H}_n)}\\
&= \quad \mathrm{Coh}_\mathrm{Sh}(\mathbf{S}) \cdot \frac{\tilde{P}(\mathrm{H}_1 \vee \cdots \vee \mathrm{H}_n)}{P(\mathrm{H}_1 \vee \cdots \vee \mathrm{H}_n)}\\
&= \quad \mathrm{Coh}_\mathrm{Sh}(\mathbf{S}) \cdot \frac{1 - P(\neg\mathrm{H}_1) \cdots P(\neg\mathrm{H}_n)}{1 - P(\neg\mathrm{H}_1, \ldots, \neg\mathrm{H}_n)} && (8)
\end{aligned}$$

This measure corresponds to the ratio between the actual relative overlap (that is, for P) and the relative overlap that would exist if the propositions in question were independent and had the same marginal probabilities (that is, for $\tilde{P}$). $\mathrm{Coh}_{\mathrm{OG}^+}$ may therefore also be used as a measure of absolute coherence: If $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}) > 1$, then we are above the independence baseline and the information set $\mathbf{S}$ is absolutely coherent. And if $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}) < 1$, then $\mathbf{S}$ is absolutely incoherent. It is also interesting to note that, as eq. (8) shows, $\mathrm{Coh}_{\mathrm{OG}^+}$ is closely related to the Shogenji measure (and only a little more difficult to compute).[6]

[6]The second factor in eq. (8) is greater than 1 if $P(\neg\mathrm{H}_1, \ldots, \neg\mathrm{H}_n) > P(\neg\mathrm{H}_1) \cdots P(\neg\mathrm{H}_n)$. For $n = 2$, this is exactly the case when $P(\mathrm{H}_1, \mathrm{H}_2) > P(\mathrm{H}_1)\, P(\mathrm{H}_2)$. That is, for $n = 2$, Coh_Sh and $\mathrm{Coh}_{\mathrm{OG}^+}$ agree on whether a given information set is coherent or not. Interestingly, the generalization of the above equivalence does not hold for larger information sets, so Coh_Sh and $\mathrm{Coh}_{\mathrm{OG}^+}$ may come to different judgments.

$P(\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3, \mathrm{H}_4)$	$= .0696$	$P(\neg\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3, \mathrm{H}_4)$	$= .1057$
$P(\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0564$	$P(\neg\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0619$
$P(\mathrm{H}_1, \mathrm{H}_2, \neg\mathrm{H}_3, \mathrm{H}_4)$	$= .0701$	$P(\neg\mathrm{H}_1, \mathrm{H}_2, \neg\mathrm{H}_3, \mathrm{H}_4)$	$= .0572$
$P(\mathrm{H}_1, \mathrm{H}_2, \neg\mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0018$	$P(\neg\mathrm{H}_1, \mathrm{H}_2, \neg\mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0627$
$P(\mathrm{H}_1, \neg\mathrm{H}_2, \mathrm{H}_3, \mathrm{H}_4)$	$= .0830$	$P(\neg\mathrm{H}_1, \neg\mathrm{H}_2, \mathrm{H}_3, \mathrm{H}_4)$	$= .1017$
$P(\mathrm{H}_1, \neg\mathrm{H}_2, \mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0303$	$P(\neg\mathrm{H}_1, \neg\mathrm{H}_2, \mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0901$
$P(\mathrm{H}_1, \neg\mathrm{H}_2, \neg\mathrm{H}_3, \mathrm{H}_4)$	$= .0266$	$P(\neg\mathrm{H}_1, \neg\mathrm{H}_2, \neg\mathrm{H}_3, \mathrm{H}_4)$	$= .0849$
$P(\mathrm{H}_1, \neg\mathrm{H}_2, \neg\mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0591$	$P(\neg\mathrm{H}_1, \neg\mathrm{H}_2, \neg\mathrm{H}_3, \neg\mathrm{H}_4)$	$= .0389$

Table 1: An example of a probability distribution P of an information quadruple for which $\mathrm{Coh}_{\mathrm{OG}^+}$ does not satisfy DEPENDENCE.

Interestingly, and contrary to other measures of relative overlap considered so far, we can show that $\mathrm{Coh}_{\mathrm{OG}^+}$ satisfies DEPENDENCE for information pairs and triples. That is, our new measure satisfies DEPENDENCE for information sets where any adequate measure of coherence arguably should satisfy this principle, whereas the other overlap measures from the literature all fail to do so. The following proposition shows that our proposed measure satisfies this desideratum (all proofs are in the Appendix):

Proposition 1. *An agent considers information items* H_1*,* H_2*, and* H_3 *with a prior probability distribution P defined over the corresponding propositional variables. Let* $\mathbf{S}_2 := \{\mathrm{H}_1, \mathrm{H}_2\}$ *and* $\mathbf{S}_3 := \{\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3\}$*. Then the following hold for* $i = 2, 3$*: (i)* $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_i) > 1$ *if* $\mathbf{S}_i$ *is positively correlated; (ii)* $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_i) = 1$ *if* $\mathbf{S}_i$ *is independent; (iii)* $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_i) < 1$ *if* $\mathbf{S}_i$ *is negatively correlated.*

This is a significant result because it shows that a measure that takes the intuition of relative overlap seriously can account for probabilistic relevance considerations, even though the corresponding baseline measure of coherence, $\mathrm{Coh}_{\mathrm{OG}}$, does not satisfy DEPENDENCE (Schippers, 2014, p. 3840). Remarkably, $\mathrm{Coh}_{\mathrm{OG}^+}$ satisfies DEPENDENCE even for information triples.

Note, however, that DEPENDENCE does not generally hold for $\mathrm{Coh}_{\mathrm{OG}^+}$ when larger information sets are considered, as the example of an information quadruple $\mathbf{S} = \{\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3, \mathrm{H}_4\}$ in Table 1 shows. It is easy to verify that this information set is positively correlated. However, $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}) \approx .996 < 1$, which is below the threshold for absolute coherence according to this measure. Hence, DEPENDENCE does not hold in general for $\mathrm{Coh}_{\mathrm{OG}^+}$. This is a welcome result because DEPENDENCE is a demanding principle and it is doubtful whether it is reasonable to require it for information sets of any size. And yet, DEPENDENCE is eminently plausible for smaller information sets and $\mathrm{Coh}_{\mathrm{OG}^+}$ satisfies it for $n = 2$ and 3. We therefore conclude that $\mathrm{Coh}_{\mathrm{OG}^+}$ provides a good compromise between relative overlap and dependence (or relevance) considerations.

However, the objection could be raised that it is not even necessary to consider relative overlap: it seems that we would be fine with a simple measure that considers only correlative aspects of coherence, since such measures also satisfy DEPENDENCE. As we will now show, this is not the case: the amount of relative overlap between the information items should also be taken into account when determining the coherence of an information set.

If we want to measure only the deviation from the independence baseline, then we can use a measure such as Shogenji's $\mathrm{Coh}_{\mathrm{Sh}}$ (Shogenji, 1999). This, however, cannot be correct. As Fitelson (2003, p. 196) points out, an information set can be j-wise independent (correlated), but not i-wise independent (correlated) for any $i \neq j$.[7] The Shogenji measure only corresponds to the n-wise (in)dependence for information sets of n information items, that is, the only (in)dependence that matters for this measure is the overall joint coherence.

Fortunately, however, these problems can be avoided with our new measure, which combines relative overlap and dependence considerations. Suppose that we have an information triple $\mathbf{S} = \{\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3\}$ and that H_1, H_2, and H_3 are jointly independent, but pairwise positively correlated. This means that H_1 and H_2, H_2 and H_3, and H_1 and H_3 are each in mutual confirmatory relations, even though H_1, H_2, and H_3 are jointly independent. The amount of 2-wise correlation, however, suggests that the information set should be considered to be absolutely

[7] An information set is i-wise independent (correlated) when all and only the subsets of size i are independent (correlated).

coherent. This is exactly the result that our new measure, contrary to the Shogenji measure, gives. Hence, it seems that we should not just take the overall deviation from the independence baseline into consideration, but also how much agreement there is (see also Olsson, 2022, pp. 44-45, for a helpful discussion). This point is made precise in the following corollary to Proposition 1.

Corollary 1. *An agent considers an information triple* $\mathbf{S}_3 = \{\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3\}$ *with a prior probability distribution P defined over the corresponding propositional variables. The information items are pairwise positively correlated, but jointly independent. Then the following hold: (i)* $\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S}_3) = 1$*, that is,* $\mathbf{S}_3$ *is assessed to be neither absolutely coherent nor absolutely incoherent on the Shogenji measure. (ii)* $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_3) > 1$*, that is,* $\mathbf{S}_3$ *is assessed to be absolutely coherent on the proposed new overlap measure.*

In contrast to the Shogenji measure, which measures the amount of independence deviation alone, $\mathrm{Coh}_{\mathrm{OG}^+}$ takes into account not only the overall independence or correlation of an information set, but also the amount of relative overlap. The $\mathrm{Coh}_{\mathrm{OG}^+}$ measure is moreover computationally tractable, unlike subset-based measures such as $\mathrm{Coh}_{\mathrm{OG}'}$ and $\mathrm{Coh}_{\mathrm{OG}^*}$ (which are based on the idea of relative overlap) as well as Fitelson's (2003) and Schupbach's (2011) measures (which are based on the idea of independence deviation), and the measures discussed by Douven and Meijs (2007) (which are based on the idea of mutual support). In summary, the proposed new measure $\mathrm{Coh}_{\mathrm{OG}^+}$ represents a fruitful compromise between relative overlap and independence deviation while remaining computationally manageable.

4 The Truth-Tracking Argument

We have seen that our new measure satisfies the desirable principle DEPENDENCE for information pairs and triples. This gives it an advantage over pure relative overlap measures. However, it is not clear what principles, if any, should apply to larger information sets. Furthermore, our intuitions about the coherence of larger information sets are not clear at all (which is suggested by the lack of test cases with more than four information items; see, for example, the standard test cases of Koscholke (2016). This is problematic because coherentism should say something about the coherence of larger information sets in order to investigate the holistic justification of our beliefs—a central concern of coherentism.

Therefore, coherence measures should be evaluated for larger information sets to see how well they perform a particular function. In our case, this is to indicate the truth of the information set in question. An ideal coherence measure should correlate the degree of coherence of the information set with its truth content. But how do we know that more coherent information sets are more likely to contain true information? Here, different coherence measures are likely to cut off differently, and are correspondingly more or less suitable for identifying whether information is true or not.

This is not to say that it is not important to consider the performance of the measures in standard toy cases from the literature, or that the intuitions behind the measures do not matter. These two aspects help us to develop appropriate coherence measures. How well they are able to measure the truth content of an information set, however, may be quite another matter. If we are to claim that a certain degree of coherence (as determined by a measure) justifies our beliefs, we also need to know which measures are the most sensitive and reliable indicators of truth. Our goal is therefore twofold: First, we want to show that, under some plausible assumptions which we describe below, coherence measures, in general, may indicate whether an information set contains true information. Second, we want to show that the new measure $\mathrm{Coh}_{\mathrm{OG}^+}$ is at least as good in this respect as other pure measures of relative overlap that we have criticized above for satisfying AGREEMENT when information pairs are considered. There we argued that AGREEMENT is not a plausible principle, but if the measures that (at least partially)[8] satisfy this principle (that is, $\mathrm{Coh}_{\mathrm{OG}}$, $\mathrm{Coh}_{\mathrm{OG}'}$, and $\mathrm{Coh}_{\mathrm{OG}^*}$) were much more reliable indicators of truth, then this would be a strong reason to reconsider this principle.[9]

[8]Recall that it is yet unsettled whether $\mathrm{Coh}_{\mathrm{OG}}$ and $\mathrm{Coh}_{\mathrm{OG}'}$ satisfy AGREEMENT for information sets with three or more items.

[9]This is also why we do not here further explore other measures that do not satisfy AGREEMENT at all, such as $\mathrm{Coh}_{\mathrm{Sh}}$, $\mathrm{Coh}_{\mathrm{F}}$, and the measures proposed by Douven and Meijs (2007).

Douven (2021) has recently shown that we can use computer simulations to study how well different probabilistic confirmation measures discriminate between true and false hypotheses. In answering the puzzle of the truth-tracking abilities of coherence, we follow Douven's approach and adapt it for our needs. We also note the previous simulation-based research on coherence and truth-tracking, in particular by Angere (2007, 2008) and Glass (2012).Our interest here, however, is not whether higher coherence of an information set means higher probability on average. Instead, we focus on how well the coherence of an information set is able to distinguish true and false information sets. Accordingly, the method of Douven (2021) is more suitable here.

All relative overlap measures of coherence provide a numerical value that expresses how coherent a given information set **S** supposedly is (relative to a probability distribution P). Moreover, the information set **S** can be described as true if all information items in **S** are true, and as false otherwise. We can then use statistical techniques to find a particular threshold for coherence that best reflects the truth value of the information set. But even once we have found these optimal thresholds, we can still assume that some measures are better able to distinguish between true and false information sets. Thus, it can be assumed that some coherence measures are better indicators of truth than others.

The flow of our simulations can be roughly described as follows (see Appendix A.3 for details): We generate n possible worlds over which we define a random probability distribution (clearly, the probabilities of all possible worlds sum up to 1).[10] We also generate an information set where each information item obtains in randomly selected possible worlds. This suffices to calculate the coherence of the information set by various probabilistic measures of coherence. However, it is well known that internal coherence may be completely detached from truth. Fairy tales, for instance, are internally highly coherent, but lack a veridical connection to reality. In addition to each randomly generated information item, we therefore also randomly generate a corresponding true piece of evidence which confirms it. The truth of evidence in our simulations simply means that all pieces of evidence contain the true (actual) possible world.

We consequently randomly generate two sets: the information set $\mathbf{S} = \{\mathrm{H}_1, \ldots, \mathrm{H}_n\}$, which may or may not contain only true information (depending on whether the randomly chosen actual world is included in the subset of all information items), and the set of corresponding true evidence $\mathbf{E} = \{\mathrm{E}_1, \ldots, \mathrm{E}_n\}$, all of which include the actual world. Finally, to establish a reasonable connection to the truth, we require that each (true) item of evidence is positively correlated with some information, that is, that $P(\mathrm{H}_i \mid \mathrm{E}_i) > P(\mathrm{H}_i \mid \neg\mathrm{E}_i)$ for all $i = 1, \ldots, n$.

This does not mean that any information item is automatically true (they may all be false), but only that *if* it is true, it is true for a sensible reason. For example, suppose that Mary lives on a farm. The information item "Mary has a little lamb" is then positively correlated with a true piece of evidence, "Mary lives on a farm," even if it turns out that Mary has another animal (or none). Thus, this condition provides a sensible connection to the truth while keeping the randomization process intact.

Once we have generated the information set **S** and a probability distribution P in the manner described, we can calculate how coherent **S** is according to various coherence measures. We can also easily determine whether the information in **S** is true by checking whether all information items in **S** are true. Figure 2 illustrates how this works in a randomly generated case of three information items.

After 100 simulations, we can then examine which coherence measure most reliably predicts (discriminates) the truth value of the simulated information sets. Given a binary dependent variable (here with the values "true" and "false") and a continuous independent variable (that is, the degree of coherence), this is accomplished by logistic regression by default. This statistical technique then allows us to estimate which measure of coherence provides the best model for discriminating between the truth and falsehood of the information set, using the so-called area under curve (AUC) value. The AUC value is usually understood to indicate the probability with which an independent continuous measure distinguishes dependent binary categories (here: true/false). In simple terms, an AUC value of 1 means that the amount of coherence is a perfect predictor of the truth of information. This could happen, for instance, if all and only the information sets with true information were determined to be maximally coherent and

[10] Similar to Douven (2021), we also note that the number of possible worlds we use in our simulations can matter, as it affects how fine-grained the sample space Ω can be. To keep the simulations nontrivial, we only consider cases with $2n$ possible worlds or more, where n is the number of information items in a simulated information set.

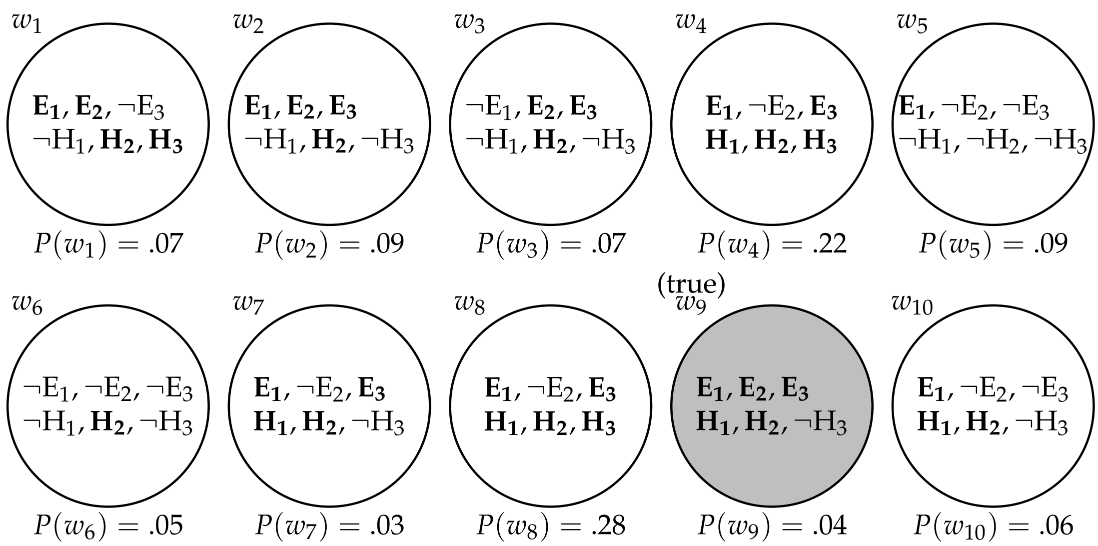

Figure 2: An example of a randomly generated set $\mathbf{S} = \{\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3\}$ with the prior probability $P(\mathbf{S}) = P(w_4) + P(w_8) = .5$ when the probability space contains 10 possible worlds. The world w_9 is true. Each information item is confirmed by true evidence: $P(\mathrm{H}_1 \mid \mathrm{E}_1) = .72 > P(\mathrm{H}_1 \mid \neg\mathrm{E}_1) = 0$; $P(\mathrm{H}_2 \mid \mathrm{E}_2) = 1 > P(\mathrm{H}_2 \mid \neg\mathrm{E}_2) = .88$; $P(\mathrm{H}_3 \mid \mathrm{E}_3) = .68 > P(\mathrm{H}_3 \mid \neg\mathrm{E}_3) = .25$. In this case, all considered measures of coherence determine the set $\mathbf{S}$ as quite coherent, that is, $\mathrm{Coh}_{\mathrm{OG}}(\mathbf{S}) = .54$, $\mathrm{Coh}_{\mathrm{OG}'}(\mathbf{S}) = .65$, $\mathrm{Coh}_{\mathrm{OG}^*}(\mathbf{S}) = .67$, and $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}) = 1.66$. Note that the former three range between 0 and 1, and the last between 0 and ∞, with values over 1 denoting absolutely coherent sets.

all and only information sets with false information were determined to be minimally coherent. An AUC value of .5, on the other hand, indicates that a measure is about as reliable as a random guess in predicting whether a given information set contains only true or also false information.[11]

Finally, it has also been pointed out that the prior joint probability of an information set plays an important role in evaluating its coherence (see, for example, Bovens and Hartmann, 2003, 2005, 2006). Therefore, we keep the prior joint probability of our simulated information sets constant across all simulations—the example in Figure 2, for instance, has a prior joint probability of .5. Given an information set $\mathbf{S}$, we consider prior probabilities $P(\mathbf{S})$ ranging from .1 to .9 in steps of .2. In terms of cardinality, we consider information sets ranging in size from two to seven information items. We avoid larger information sets for three reasons: First, our results are relatively stable as we increase the size of the simulated information sets. Second, the computations for the measures $\mathrm{Coh}_{\mathrm{OG}'}$ and $\mathrm{Coh}_{\mathrm{OG}^*}$, which account for the average coherence sets in differently defined subsets, become increasingly expensive. Already to compute $\mathrm{Coh}_{\mathrm{OG}^*}$ for an information set with seven information items, we need to perform 966 computations in each of the millions of simulations; for each additional information item, we need about three times as many computations (3025 for a set with eight information items and 9330 for a set with ten information items). If we want to maintain randomization, it also becomes increasingly difficult to generate suitable probability distributions that satisfy our conditions that each information is confirmed by true evidence and that the prior probability is fixed. To achieve the desired reliability of our results, we repeat each of these simulations 100 times.

Figure 3 shows the average AUC values over 100 simulations for different numbers of possible worlds in all variations (cardinality of information sets from 2 to 7 and prior probabilities between .1 and .9 in steps of .2). It is immediately apparent from the figure that our new measure $\mathrm{Coh}_{\mathrm{OG}^+}$ performs better in terms of its truth-finding capabilities than the other relative overlap measures. Its performance in this respect is similar to that of $\mathrm{Coh}_{\mathrm{OG}^*}$, but it should be stressed that $\mathrm{Coh}_{\mathrm{OG}^+}$ is computationally much less demanding as it does not consider how

[11] The technical details behind the procedure are beyond the scope of this paper, but an interested reader can find a brief summary in Douven, 2021, p. 404. For more details, see Fawcett, 2006.

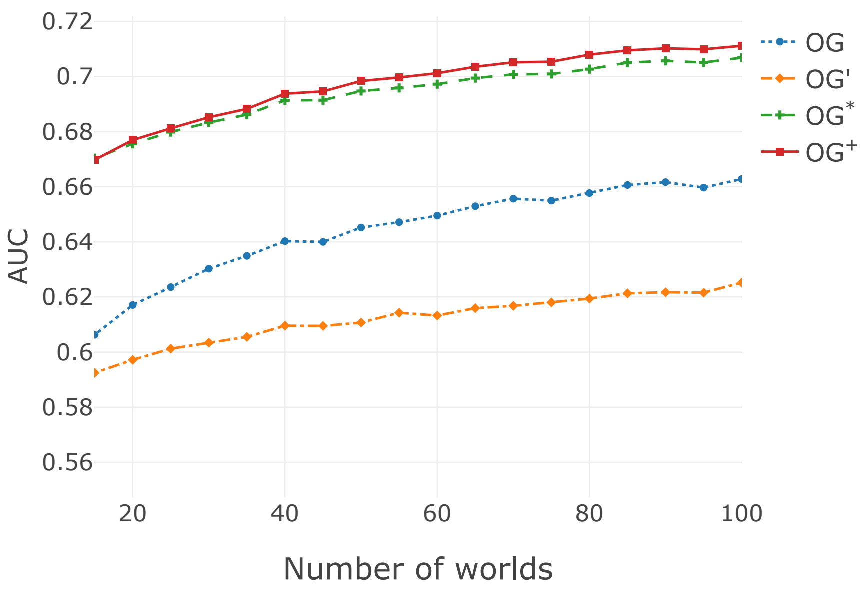

Figure 3: Average AUC values for different measures of coherence, averaged over all simulated variations.

much each subset of an information set coheres. This means that in addition to having the desirable properties we expect of a coherence measure, we can also use $\mathrm{Coh}_{\mathrm{OG}^+}$ in distinguishing truth from falsehood quite reliably under the assumption that the assessed information set is confirmed by true evidence.

It should be noted that these results are aggregated over all variations in terms of the size and the prior probability of the simulated sets. However, we get a more detailed insight when we focus on the simulated information sets of specific sizes. Specifically, when we consider information sets consisting of two, three, or four information items, the differences between the results of the respective measures are less apparent and are also not statistically significant (despite $\mathrm{Coh}_{\mathrm{OG}^+}$ being the best in truth-tracking on average). This is expected because there is less variability in smaller sets, so the measures also provide more similar assessments. On the other hand, if we consider larger information sets, the differences among the measures are amplified and two measures come out as significantly better indicators of truth than others: our new measure $\mathrm{Coh}_{\mathrm{OG}^+}$ and the measure $\mathrm{Coh}_{\mathrm{OG}^*}$ proposed by Koscholke et al. (2019) (see Figure 4). Although this may seem like a somewhat troubling result (the performance of our new measure is not significantly better than that of $\mathrm{Coh}_{\mathrm{OG}^*}$), it actually is not. Quite the opposite: it shows that our new measure is among the best indicators of truth despite its computational simplicity and without considering any specific subsets even when we consider larger sets. Furthermore, it is interesting to note that all coherence measures discussed here are better than chance at predicting the truth of information in the considered information sets.

Levi (1967, p. 58) famously argued that truth and relief from agnosticism are the most important desiderata of scientific reasoning and that other intuitively appealing desiderata may be reducible to them. However, the impossibility results of Bovens and Hartmann (2003) and Olsson (2005) show that coherence and truth are not related in a straightforward way. And yet, our investigation suggests an answer to the question of why we value coherent information: coherence matters because it is likely to be our best available guide to truth, at least when more direct evidence is unavailable and when the information items are confirmed by veridical evidence. In other words, coherence considerations provide an important heuristic of uncertain reasoning.[12]

[12] See also Angere (2008) for a similar claim.

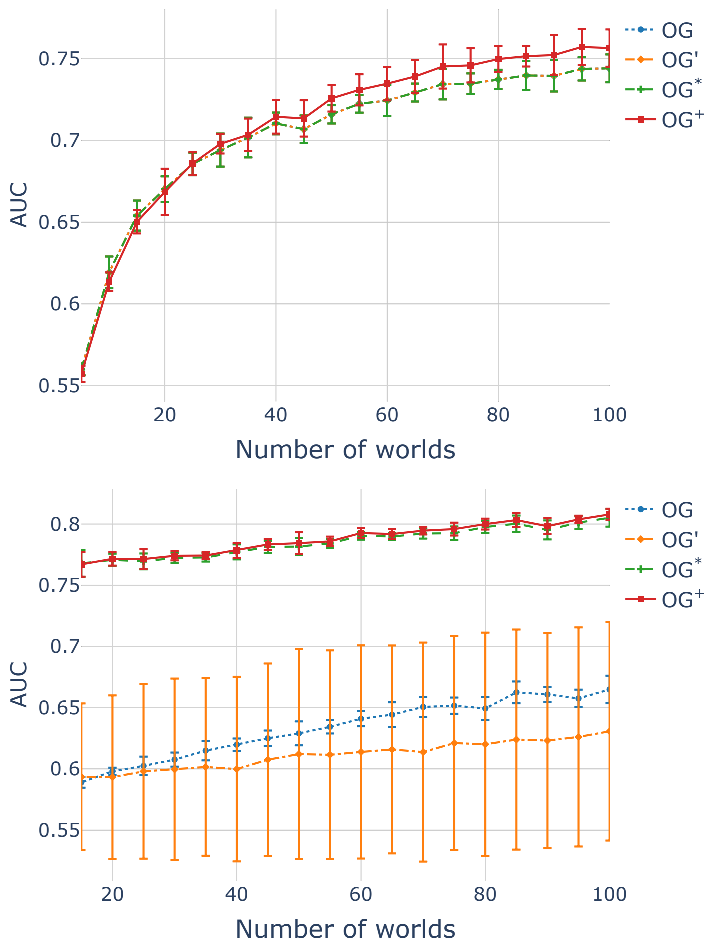

Figure 4: Average AUC values for information pairs (top) and for information sets with seven information items (bottom). Error bars represent the standard deviations. $\mathrm{Coh}_{\mathrm{OG}'}$, unlike other measures, is very sensitive to the prior probability of larger information sets (bottom), which explains the wide error bars.

5 Conclusions

The coherence measure $\mathrm{Coh}_{\mathrm{OG}^*}$ proposed by Koscholke, Schippers, and Stegmann seemed to give new hope to the possibility of measuring the coherence of an information set by the relative overlap of propositions in probability space. However, as we have shown, this measure fails for a number of reasons. The most important reason is that it completely ignores how strongly the information set in question is statistically correlated. Nevertheless, there is hope for relative overlap measures if, in addition to relative overlap, we consider correlations in the information set. The proposed new measure $\mathrm{Coh}_{\mathrm{OG}^+}$ does just that. It is also superior to pure correlation measures, such as the Shogenji measure $\mathrm{Coh}_{\mathrm{Sh}}$. Our results suggest that all measures that focus on only one main intuition about the coherence of an information set (that is, either deviation from independence or relative overlap) perform less well than measures that, like the proposed one, seek a productive compromise between the two main intuitions.

As we have seen in our exploratory computer simulations, the new measure $\mathrm{Coh}_{\mathrm{OG}^+}$ is also better at identifying true information than pure overlap measures. This is another argument that both relative overlap and deviation from the independence baseline should be considered when we want to determine how coherent a given set of information is. Examining the truth-tracking properties of various measures of coherence also provides us with yet another argument for the importance of coherence considerations in justifying beliefs. Although pointing to the degree of coherence of an information set is hardly sufficient to convince the hardcore epistemic skeptic, our results may help to convince a mild skeptic who concedes the possibility that there is at least some connection between coherence and the truth of our beliefs.

Finally, it is plausible that scientific research resembles a weakly skeptical scenario because we can reasonably assume that scientific insights provide us with at least somewhat reliable representations of slices of the world. This may also be why coherence considerations seem to play such an important role in scientific reasoning (see, for example, Thagard, 2007). Further investigation of the role of coherence considerations in science (for example, using case studies) and of their normative underpinnings remains the task of future studies. Measures such as $\mathrm{Coh}_{\mathrm{OG}^+}$ could prove useful in these investigations.

A Proofs and Algorithms

A.1 Proof of Proposition 1

We assume that $\mathbf{S}_2$ and $\mathbf{S}_3$ are positively correlated. The proofs for negatively correlated and for independent information sets can be obtained by the same steps as below, but with "$<$" (for negatively correlated information sets) or "$=$" (for independent sets) instead of "$>$".

Let us begin with $\mathbf{S}_2$. Then $P(\mathrm{H}_1, \mathrm{H}_2) > P(\mathrm{H}_1)\, P(\mathrm{H}_2)$ implies that $\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S}_2) > 1$. It therefore suffices to show that the second factor in eq. (8) is greater than 1. This follows because that factor is greater than 1 iff $P(\neg\mathrm{H}_1, \neg\mathrm{H}_2) > P(\neg\mathrm{H}_1)\, P(\neg\mathrm{H}_2)$, which is equivalent to $P(\mathrm{H}_1, \mathrm{H}_2) > P(\mathrm{H}_1)\, P(\mathrm{H}_2)$. Hence, $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_2) > 1$.

For $\mathbf{S}_3$, we define $\alpha_1 := P(\mathrm{H}_1) + P(\mathrm{H}_2) + P(\mathrm{H}_3)$, $\alpha_2 := P(\mathrm{H}_1, \mathrm{H}_2) + P(\mathrm{H}_2, \mathrm{H}_3) + P(\mathrm{H}_1, \mathrm{H}_3)$, $\alpha_3 := P(\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3)$, $\beta_1 := P(\mathrm{H}_1) + P(\mathrm{H}_2) + P(\mathrm{H}_3) = \alpha_1$, $\beta_2 := P(\mathrm{H}_1)P(\mathrm{H}_2) + P(\mathrm{H}_2)P(\mathrm{H}_3) + P(\mathrm{H}_1)P(\mathrm{H}_3)$, and $\beta_3 := P(\mathrm{H}_1)\, P(\mathrm{H}_2)\, P(\mathrm{H}_3)$. Then $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_3) > 1$ iff

$$
\begin{aligned}
& \frac{P(\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3)}{P(\mathrm{H}_1 \vee \mathrm{H}_2 \vee \mathrm{H}_3)} > \frac{\tilde{P}(\mathrm{H}_1, \mathrm{H}_2, \mathrm{H}_3)}{\tilde{P}(\mathrm{H}_1 \vee \mathrm{H}_2 \vee \mathrm{H}_3)} \\
\Leftrightarrow \quad & \frac{\alpha_3}{\alpha_1 - \alpha_2 + \alpha_3} > \frac{\beta_3}{\beta_1 - \beta_2 + \beta_3} = \frac{\beta_3}{\alpha_1 - \beta_2 + \beta_3} \\
\Leftrightarrow \quad & \alpha_3\, (\alpha_1 - \beta_2) > \beta_3\, (\alpha_1 - \alpha_2) \\
\Leftrightarrow \quad & (\alpha_2 - \beta_2)\, \beta_3 + (\alpha_1 - \beta_2)(\alpha_3 - \beta_3) > 0
\end{aligned}
\tag{9}
$$

Since $\mathbf{S}_3$ is positively correlated, it holds that $\alpha_2 > \beta_2$ and $\alpha_3 > \beta_3$. To complete the proof, we only need to show that $\alpha_1 > \beta_2$. This holds because each term in α_1 is greater than the corresponding term in β_2. □

A.2 Proof of Corollary 1

We use the same shorthands as in A.1 and note that the conditions specified in Proposition 1 imply that $\alpha_2 > \beta_2$ and $\alpha_3 = \beta_3$. Then $\mathrm{Coh}_{\mathrm{Sh}}(\mathbf{S}_3) = 1$ and eq. (9) implies that $\mathrm{Coh}_{\mathrm{OG}^+}(\mathbf{S}_3) > 1$. □

A.3 The Possible-Worlds Approach to Simulations of Truth-Tracking

The following pseudo-algorithm describes the simulations:[13]

FOR EACH cardinality of the information set **S** between 2 and 7
  FOR EACH prior joint probability of **S** between .1 and .9 in .2-increments
    START simulation – REPEAT 100 times
      FOR EACH number of possible worlds from 5 to 100 in increments of 5
        REPEAT 100 times
          WHILE not: $P(\mathrm{H}_i \mid \mathrm{E}_i) > P(\mathrm{H}_i \mid \neg\mathrm{E}_i)$ for all i,
            1. RANDOMLY SELECT which possible worlds are contained by each evidence $\mathrm{E}_1, \ldots, \mathrm{E}_n$ and ensure $\bigcap\{\mathrm{E}_1, \ldots, \mathrm{E}_n\} \neq \emptyset$;
            2. RANDOMLY SELECT which possible worlds are contained by each information $\mathrm{H}_1, \ldots, \mathrm{H}_n$;
            3. GENERATE a random probability distribution P over all possible worlds;
          4. RANDOMLY SELECT the true world w_T from $\bigcap\{\mathrm{E}_1, \ldots, \mathrm{E}_n\}$;
          5. IF $w_\mathrm{T} \in \bigcap\{\mathrm{H}_1, \ldots, \mathrm{H}_n\}$ (from step 2), THEN information is true, OTHERWISE false;
          6. CALCULATE $\mathrm{Coh}_{\mathrm{OG}}$, $\mathrm{Coh}_{\mathrm{OG}'}$, $\mathrm{Coh}_{\mathrm{OG}^*}$, and $\mathrm{Coh}_{\mathrm{OG}^+}$ according to the probability distribution P (from step 3) and the probability of the possible worlds contained by information (from step 2);
          7. RETURN a vector with true/false (from step 5) and values of coherence according to various measures (from step 6);
        8. CONDUCT logistic regression on the data (from step 7) obtained in 100 repetitions to determine the AUC of each measure of coherence;
9. AGGREGATE the data (from step 8);
10. GENERATE AUC plots.

References

Angere, S. (2007). The defeasible nature of coherentist justification. *Synthese 157*(3), 321–335.

Angere, S. (2008). Coherence as a heuristic. *Mind 117*(465), 1–26.

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

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

Bovens, L. and S. Hartmann (2005). Why there cannot be a single probabilistic measure of coherence. *Erkenntnis 63*, 361–374.

Bovens, L. and S. Hartmann (2006). An impossibility result for coherence rankings. *Philosophical Studies 128*(1), 77–91.

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

Douven, I. (2021). Tracking confirmation. *Philosophy of Science 88*(3), 398–414.

Douven, I. and W. Meijs (2007). Measuring coherence. *Synthese 156*(3), 405–425.

Elqayam, S. and J. S. B. T. Evans (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences 34*(5), 233–248.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters 27*(8), 861–874.

Fitelson, B. (2003). A probabilistic theory of coherence. *Analysis 63*(3), 194–199.

[13] The actual code used to conduct the simulations (written in Julia; extended and adapted from the code originally developed by Douven, "Tracking Confirmation".) as well as additional plots and data are available here: `https://github.com/philosophy-simul/truth-tracking-coherence`.

Glass, D. H. (2002). Coherence, explanation, and Bayesian networks. In M. O'Neill, R. F. E. Sutcliffe, C. Ryan, M. Eaton, and N. J. L. Griffith (Eds.), *Artificial Intelligence and Cognitive Science, 13th Irish Conference, AICS 2002*, pp. 177–182. Berlin: Springer.

Glass, D. H. (2012). Inference to the best explanation: does it track truth? *Synthese 185*, 411–427.

Hahn, U., A. J. L. Harris, and A. Corner (2016). Public reception of climate science: Coherence, reliability, and independence. *Topics in Cognitive Science 8*(1), 180–195.

Harris, A. J. L. and U. Hahn (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition 35*(5), 1366–1373.

Koscholke, J. (2016). Evaluating test cases for probabilistic measures of coherence. *Erkenntnis 81*(1), 155–181.

Koscholke, J. and M. Jekel (2017). Probabilistic coherence measures: A psychological study of coherence assessment. *Synthese 194*(4), 1303–1322.

Koscholke, J. and M. Schippers (2016). Against relative overlap measures of coherence. *Synthese 193*(9), 2805–2814.

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

Levi, I. (1967). *Gambling with Truth: An Essay on Induction and the Aims of Science*. New York and London: Alfred A. Knopf and Routledge & Kegan Paul, Ltd.

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

Meijs, W. (2006). Coherence as generalized logical equivalence. *Erkenntnis 64*(2), 231–252.

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

Olsson, E. J. (2005). *Against Coherence: Truth, Probability, and Justification*. Oxford: Oxford University Press.

Olsson, E. J. (2021). Coherentist theories of epistemic justification. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

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

Schippers, M. (2014). Probabilistic measures of coherence: From adequacy constraints towards pluralism. *Synthese 191*(16), 3821–3845.

Schupbach, J. N. (2011). New hope for Shogenji's coherence measure. *The British Journal for the Philosophy of Science 62*(1), 125–142.

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

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