Confirmation, Coherence and the Strength of Arguments

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

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

Abstract

Alongside science and law, argumentation is also of central importance in everyday life. But what characterizes a good argument? This question has occupied philosophers and psychologists for centuries. The theory of Bayesian argumentation is particularly suitable for clarifying it, because it allows us to take into account in a natural way the role of uncertainty, which is central to much argumentation. Moreover, it offers the possibility of measuring the strength of an argument in probabilistic terms. One way to do this, implicit in much work, is to identify the strength of an argument with the degree to which the premises of the argument confirm the conclusion. We criticize this prima facie plausible proposal and suggest instead that the strength of an argument has something to do with how much the premises and the conclusion of the argument cohere with each other. This leads to a new probabilistic measure whose properties we examine in more detail.

Keywords: Bayesian Argumentation; Argument Strength; Confirmation; Coherence

Introduction

Argumentation is central to science, law, and everyday life. Scientists use arguments, for example, to convince their colleagues that a proposed explanation of a phenomenon is correct and that the proposed research project is worthy of funding. Lawyers present arguments for their clients' innocence in court, and we all argue with our partner, for example, why this investment is better than that one. But how should we evaluate these arguments? What makes an argument a good argument? And why is this argument better than that one? Answering these questions is the goal of the research area of rational argumentation, which has occupied philosophers and psychologists since Aristotle.

While related research has long focused on the study of deductive arguments, the realization that most interesting arguments involve uncertain premises and conclusions has led to the initiation of research programs that incorporate non-classical logics and other approaches. Last but not least, in this context, the theory of Bayesian argumentation was developed, which explores various argument patterns that are not necessarily deductively valid and involve uncertain premises. The motivation for this is twofold. First, there are deductively valid arguments that are not particularly strong. To see this, consider the following example from Oaksford and Chater (2007, p. 127):

A_1: If I turn the key, the car starts.

A_2: The car didn't start.

C: I didn't turn the key.

This argument is an example of the deductively valid argument pattern modus tollens, but it does not seem particularly strong because it is easy to find an alternative explanation for why the car did not start even though the key was turned.

Second, deductively invalid arguments can be strong. For example, Ulrike Hahn and her collaborators have convincingly demonstrated that various so-called fallacious arguments can actually be decidedly strong arguments (e.g., Hahn & Oaksford, 2006, 2007; Hahn, 2020). Compare the following arguments:

A: This treatment has worked in all 5,000 trials so far.

C: This treatment is efficient,

and

A': This treatment has worked in the only trial so far.

C: This treatment is efficient.

Of course, the conclusion that the treatment is efficient does not follow with necessity in either case, but there is also an obvious difference in the strengths of the two arguments: 5,000 successful trials support the conclusion more than a single trial. These considerations suggest that the strength of an argument is not simply determined by the logical structure of the argument and depends on the context.

Within the framework of Bayesian argumentation, these examples can be analyzed as follows (see Eva & Hartmann, 2018): We consider an agent (= agent 1) who entertains a set of propositions $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_\mathrm{n}, \mathrm{C}$ (in roman script) with a probability distribution P defined over the corresponding algebra of propositional variables $A_1, A_2, \ldots, A_n, C$ (in italics). This probability distribution represents the corresponding degrees of belief of the agent. Let one of these propositions be the conclusion (C) of an argument; the others function in the premises. Then another agent (= agent 2) comes along and wants to convince agent 1 of the conclusion of the argument in question. To do this, agent 2 supports the premises of the argument, which causes agent 1 to change the probabilities of the premises "by hand". For example, agent 1 may shift the probability of all premises to 1. Consequently, agent 1

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

"learns" the premises of the argument and then updates the entire probability distribution P to continue to have coherent beliefs. In the case of a deductively valid argument, this results in a new probability of 1 being assigned to the argument's conclusion. If the probabilities of the premises do not all shift to 1, the new probability of the conclusion will always increase (but usually not to 1) if the argument pattern is deductively valid. If the argument pattern is not deductively valid, then the probability of the conclusion may be greater than, less than, or equal to the original one (see Adams, 1975, 1996).

It is now natural to identify the strength of an argument with the change in the probability of the conclusion. We will examine this proposal in the next section.

Confirmation-based Measures

To make the mentioned proposal more precise, it is useful to recall the basics of Bayesian Confirmation Theory. Here we consider an agent who entertains a hypothesis H to which they assign a prior probability $P(\mathrm{H})$. To test H, the agent explores whether a (deductive or inductive) consequence E of H (= the evidence) holds or not. Before finding out, the agent has a certain expectation $P(\mathrm{E})$ whether E will turn out or not. $P(\mathrm{E})$ can be computed from the *likelihoods* $P(\mathrm{E}|\mathrm{H})$ and $P(\mathrm{E}|\neg\mathrm{H})$ and the prior probability of the hypothesis $P(\mathrm{H})$ using the law of total probability. If the agent then observes that E is indeed the case, they update the probability distribution P to obtain a new ("posterior") probability distribution P' which follows from P using the principle of Conditionalization ("Bayes Theorem") and set $P'(\mathrm{H}) = P(\mathrm{H}|\mathrm{E})$. Using the definition of conditional probability, one then obtains

$$P'(\mathrm{H}) = \frac{P(\mathrm{E}|\mathrm{H})P(\mathrm{H})}{P(\mathrm{E})}. \tag{1}$$

Introducing the *likelihood ratio* $x := P(\mathrm{E}|\neg\mathrm{H})/P(\mathrm{E}|\mathrm{H})$, this can also be written as[1]

$$P'(\mathrm{H}) = \frac{P(\mathrm{H})}{P(\mathrm{H}) + x \cdot P(\neg\mathrm{H})}. \tag{2}$$

If $P'(\mathrm{H}) > P(\mathrm{H})$, then E confirms H, if $P'(\mathrm{H}) < P(\mathrm{H})$, then E disconfirms H, and if $P'(\mathrm{H}) = P(\mathrm{H})$, then E is irrelevant for H. Equivalently, if E obtains and $x < 1$, then E confirms H, if $x > 1$, then E disconfirms H, and if $x = 1$, then E is irrelevant for H.

To apply Bayesian Confirmation Theory to argumentation, we have to assume, as described above, that the agent has a prior probability distribution P over the premises and conclusion of the argument and then learns the premises e.g. via the testimony of another agent. Using Bayes Theorem to update their degrees of belief, the agent obtains a new (posterior) probability of the conclusion C. The argument is *good* if the new probability of the conclusion is greater than its prior probability, i.e. if the premises confirm the conclusion. All measures of confirmation follow this general principle (Fitelson, 1999, S362-3) and nothing of what follows will depend on a choice of a specific measure of confirmation. Still, it is perhaps the most straight-forward and natural to measure the degree of confirmation with the difference measure $d(\mathrm{H},\mathrm{E}) := P(\mathrm{H}|\mathrm{E}) - P(\mathrm{H})$. The measure has many proponents (e.g., Earman, 1992; Eells, 1982; Gillies, 1986; Jeffrey, 1992; Rosenkrantz, 1994). When we apply d to argumentation, H represents the conclusion and E represents the conjunction of all premises of the argument.

This proposal has an intuitive appeal and it is also implicit in the above-mentioned work of Ulrike Hahn and collaborators. It is also used, e.g., in the assessment of various argument patterns such as the no-alternatives argument (see Dawid, Hartmann, & Sprenger, 2015). In these examples, the number of premises is small and all premises are propositions. It is less clear how the proposal works when there is a conditional among the premises, since natural language conditionals may perhaps not be representable by a proposition (Edgington, 1995). Suggestions on how to deal with such cases can be found, for example, in Douven (2012), Eva, Hartmann, and Rad (2020) and Günther and Trpin (2022).

Here we want to point out three more problems for identifying the strength of an argument with the degree of confirmation of the conclusion by the premises.

1. **Irrelevant Premises:** Consider the following two arguments:

   A_1: It rained earlier today.
   C: The street is wet,

   and

   A_1: It rained earlier today.
   A_2: There are no people on Mars.
   C: The street is wet.

   The first argument is clearly strong because it provides a good reason (i.e., that it rained) in support of the conclusion (i.e., that the street is wet). It is stronger than the second argument which in addition to the first premise also invokes the irrelevant premise that there are no people on Mars. Whether or not the second premise is true has no implications for our assessment of the conclusion. The second argument is therefore weaker than the first argument; the additional premise distracts us from appreciating the force of the first premise. To put it slightly differently, the information set the agent entertains (i.e. the set comprising all premises and the conclusion of the argument) of the first argument is more coherent than the information set of the second argument. Here we have used the intuitive notion of coherence which has something to do with how well the propositions in the set "hang together" (BonJour, 1985, p. 93).

2. **Additional Information:** Consider the following arguments:

[1] In defining the likelihood ratio, we follow the convention used, for example, in Bovens and Hartmann (2003). Other authors define it as the reciprocal of the expression used here.

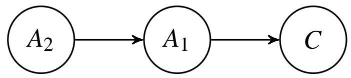

Figure 1: The Bayesian network of the additional information argument.

A_1: The air will rise, cool down and form clouds.
C: It will rain today,

and

A_2: Air pressure is lowering.
A_1: The air will rise, cool down and form clouds.
C: It will rain today.

Clearly, the second argument is stronger than the first because premise A_2 provides additional information which is relevant for A_1. However, if we assume the (scientifically quite plausible) causal chain in Figure 1, then the propositional variable A_1 screens off C from A_2 and thus $P(\mathrm{C}|\mathrm{A}_1,\mathrm{A}_2) = P(\mathrm{C}|\mathrm{A}_1)$. Accordingly, both arguments have the same strength if one uses the difference measure of confirmation.

3. **Many Premises:** The proposal under study only considers the relevance of the premises to the conclusion and is arguably not sensitive enough to the probabilistic relationships between the premises. This is because there are only two quantities that have to be specified to compute the argument strength, i.e. the probability of the conclusion given the premises and the prior probability of the conclusion. This does not matter much if the number of premises is small. For larger premise sets, however, the way the premises relate to each other will play a role in the assessment of the argument as a whole. This suggests that a more fine-grained measure needs to be found.

Let's take stock: What the three problems discussed above have in common is the problematic feature that confirmation-based measures only consider the relationship between two elements: the prior probability of the conclusion and the probability of the conclusion given the premises. However, we also care about the fit amongst the premises and about the fit of the premises and the conclusion. Taking these considerations into account is, or so we will argue, also epistemologically relevant. In other words, a more adequate and normatively appealing measure of argument strength should not be confirmation-based but coherence-based.

Coherence-based Measures

In a good argument, the premises not only support the conclusion, but also fit together well. Tensions or even contradictions between the premises certainly weaken the argument. This suggests the use of the notion of coherence for the explication of the notion of argument strength. In a first attempt, one might want to identify the strength of an argument with the coherence of the premises and the conclusion. However, this proposal does not work because the coherence of a set of propositions is *symmetric* in its arguments: Changing the order of the propositions in an information set does not change the coherence of the set. This is clearly not a desirable property of the strength of an argument: The argument from a given premise A to a conclusion C is typically stronger than the corresponding argument from C to A. For example, from the fact that I dumped a bucket of water on the street (A), I can conclude that the street is wet (C). However, the reverse conclusion from C to A is not possible without further ado.

To fix this problem, we propose to consider not only the extent to which the premises cohere with the conclusion C, but also the extent to which the conclusion coheres with the *negations* of the premises. More specifically, we consider the set of premises $\mathbf{A} = \{\mathrm{A}_1, \dots \mathrm{A}_\mathrm{n}\}$ and a conclusion $\mathrm{C} \notin \mathbf{A}$ and set $\mathbf{S} := \{\mathrm{A}_1, \dots, \mathrm{A}_n, \mathrm{C}\}$ and $\mathbf{S}^\dagger := \{\neg\mathrm{A}_1, \dots, \neg\mathrm{A}_n, \mathrm{C}\}$. (Note that we disregard arguments where the conclusion is identical to one of the premises. These petitio principii arguments "beg the question" and are excluded here.) Furthermore, Coh is a coherence measure that assigns a non-negative number to information sets, based on the probability distribution over the relevant propositional variables. The difference between $\mathrm{Coh}(\mathbf{S})$ and $\mathrm{Coh}(\mathbf{S}^\dagger)$ is then an improved measure of argument strength. To arrive at our final proposal, we normalize the resulting expression to ensure that the range of values of the measure is in the interval $[-1, 1]$. In summary, we propose the following general coherentist measure of argument strength:

Definition 1. *An agent considers the propositions* $\mathrm{A}_1, \dots, \mathrm{A}_n$ *(= the premises) and* C *(= the conclusion) with a prior probability distribution P defined over the corresponding propositional variables. The probabilistic measure of coherence* Coh: $\mathbf{S} \to \mathbb{R}^+$ *assigns a non-negative number to the information sets* $\mathbf{S}$ *and* $\mathbf{S}^\dagger$ *as defined above. Then*

$$\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \dots, \mathrm{A}_\mathrm{n}) := \frac{\mathrm{Coh}(\mathbf{S}) - \mathrm{Coh}(\mathbf{S}^\dagger)}{\mathrm{Coh}(\mathbf{S}) + \mathrm{Coh}(\mathbf{S}^\dagger)}$$

is a coherentist measure of argument strength.

We call an argument a *good argument* if $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \dots, \mathrm{A}_\mathrm{n}) > 0$. In this case, the premises and the conclusion cohere better than the negation of the premises and the conclusion. Analogously, an argument is a *bad argument* if $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \dots, \mathrm{A}_\mathrm{n}) < 0$. In this case, the premises and the conclusion cohere worse than the negation of the premises and the conclusion. Finally, if the premises and the conclusion are irrelevant for each other, then $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \dots, \mathrm{A}_\mathrm{n}) = 0$ and we do not have an argument at all.

To proceed, we need to choose a specific measure of coherence. Fortunately there is a rich literature in formal epistemology from which one can choose (see, e.g., Olsson, 2021, 2022 for surveys). The simplest measure is perhaps the Shogenji measure (Shogenji, 1999) which assigns a non-negative number to a set of propositions $\mathbf{S} := \{\mathrm{A}_1, \dots, \mathrm{A}_\mathrm{n}, \mathrm{C}\}$

and which is defined as follows:

$$\mathrm{Coh_{Sh}}(\mathbf{S}) := \frac{P(\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n, \mathrm{C})}{P(\mathrm{A}_1)P(\mathrm{A}_2)\cdots P(\mathrm{A}_n)P(\mathrm{C})} \tag{3}$$

For $n = 2$, the Shogenji measure simply measures how much the truth of one of the propositions increases the probability of the other proposition. This follows from the observation that $\mathrm{Coh_{Sh}}(\{\mathrm{A}_1, \mathrm{A}_2\}) = P(\mathrm{A}_1|\mathrm{A}_2)/P(\mathrm{A}_1) = P(\mathrm{A}_2|\mathrm{A}_1)/P(\mathrm{A}_2)$. However, the proposed generalization of this idea to larger information sets is problematic as the resulting expression is not sensitive to the dependencies in the various subsets of **S** (see Fitelson, 2003).

Perhaps a more important problem, however, is that the use of the Shogenji measure in our proposed measure of argument strength also suffers from the irrelevant premises problem mentioned above, as the following proposition states (all proofs are in the Appendix).

Proposition 1. *Let P be a probability distribution defined over the propositional variables $A_1, \ldots, A_{n+1}, C$ such that A_{n+1} is probabilistically independent of all other variables. Then $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \ldots, \mathrm{A}_{\mathrm{n+1}}) = \mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \ldots, \mathrm{A}_{\mathrm{n}})$ if one uses the Shogenji measure of coherence.*

Therefore, it is useful to look for other measures of coherence. Let's first consider the Olsson-Glass measure (Olsson, 2002; Glass, 2002), which measures the relative overlap of the propositions in probability space:

$$\mathrm{Coh_{OG}}(\mathbf{S}) := \frac{P(\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n, \mathrm{C})}{P(\mathrm{A}_1 \vee \mathrm{A}_2 \vee \cdots \vee \mathrm{A}_n \vee \mathrm{C})}. \tag{4}$$

Despite its intuitive appeal, this measure is fraught with serious problems. For example, according to this measure the coherence of an information set always decreases if one adds a proposition to it. This does not make sense, since adding a proposition to an information set often results in a more coherent set. And yet, the idea that coherence has something to do with relative overlap in probability space has enough plausibility to motivate the search for an improved overlap measure. Instead of other pure overlap measures from the literature (Meijs, 2005, 2006; Koscholke, Schippers, & Stegmann, 2019) we propose the following measure which combines considerations of overlap and dependence, as measured e.g. by the Shogenji measure:

$$\begin{aligned} \mathrm{Coh_{OG^+}}(\mathbf{S}) &:= \frac{\mathrm{Coh}^{(P)}_{\mathrm{OG}}(\mathbf{S})}{\mathrm{Coh}^{(\tilde{P})}_{\mathrm{OG}}(\mathbf{S})} \qquad (5) \\ &= \frac{1 - P(\neg\mathrm{A}_1)\cdots P(\neg\mathrm{A}_\mathrm{n}) \cdot P(\neg\mathrm{C})}{1 - P(\neg\mathrm{A}_1, \ldots, \neg\mathrm{A}_\mathrm{n}, \neg\mathrm{C})} \cdot \mathrm{Coh_{Sh}}(\mathbf{S}), \end{aligned}$$

where $\tilde{P}$ is the probability distribution which is associated with P according to which all propositions are probabilistically independent and have the same marginals as under P, i.e. $\tilde{P}(\mathrm{A_i}) = P(\mathrm{A_i})$ (for all i) and $\tilde{P}(\mathrm{C}) = P(\mathrm{C})$.

We can show that using the measure $\mathrm{Coh_{OG^+}}$ in a coherentist measure of argument strength (Definition 1) helps us avoid the above identified problems of confirmation-based measures. First, let's explore what happens if we add an irrelevant premise. To keep things simple, we focus on arguments where we add an irrelevant premise to an argument that originally only had a single (relevant) premise. Then the following holds:

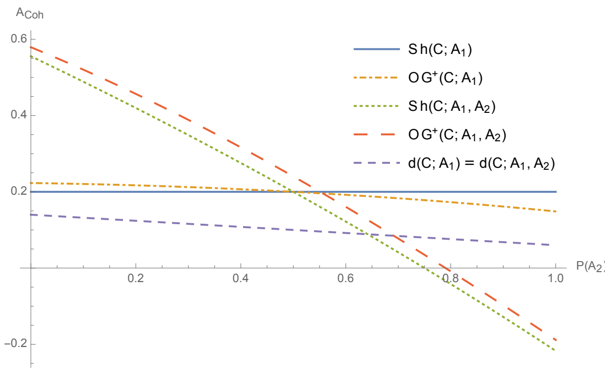

Figure 2: Argument strength as a function of $P(\mathrm{A}_2)$ for arguments from A_1 to C, and from A_2 and A_1 to C when we assume a Bayesian chain network $A_2 \to A_1 \to C$ and use $\mathrm{Coh_{Sh}}$ and $\mathrm{Coh_{OG^+}}$ as the base measures in $\mathcal{A}_{\mathrm{Coh}}$. The probability distribution is defined by $P(\mathrm{A}_1|\mathrm{A}_2) = .7$, $P(\mathrm{A}_1|\neg\mathrm{A}_2) = .3$, $P(\mathrm{C}|\mathrm{A}_1) = .6$ and $P(\mathrm{C}|\neg\mathrm{A}_1) = .4$. The assessment $d(\mathrm{C}; \mathrm{A}_1) = d(\mathrm{C}; \mathrm{A}_1, \mathrm{A}_2)$ is included for comparison.

Proposition 2. *Let P be a probability distribution defined over the propositional variables A_1, A_2 and C such that A_2 is probabilistically independent of A_1 and C. Then $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \mathrm{A}_2) \lesseqgtr \mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1)$ if $P(\mathrm{C}|\mathrm{A}_1) \gtreqless P(\mathrm{C}|\neg\mathrm{A}_1)$ and one uses the OG^+ measure of coherence.*

Hence, if one uses the OG^+ measure in our coherence-based account of argument strength, adding an irrelevant premise makes a good argument worse and a bad argument better. Both observations seem plausible as adding an irrelevant premise "dilutes" the original argument. However, further normative and empirical work needs to be done to substantiate these conclusions.

Furthermore, recall the problem of additional information. If we have a Bayesian chain network $A_2 \to A_1 \to C$ (Figure 1) such that A_2 is positively relevant for A_1 (that is, $P(\mathrm{A}_1|\mathrm{A}_2) > P(\mathrm{A}_1|\neg\mathrm{A}_2)$), then adding A_2 to an argument from A_1 to C typically improves the argument. The difference-based confirmation measure does not account for this. A coherentist measure of argument strength $\mathcal{A}_{\mathrm{Coh}}$, on the other hand, resolves this problem as may be seen from the numeric example in Figure 2. Specifically, we observe that when evaluating the argument strength by $\mathcal{A}_{\mathrm{Coh}}$, the additional information from the premise A_2 increases the argument strength unless A_2 is very probable (i.e., not informative). This seems correct as a very probable additional premise distracts from the force of the original premise and therefore weakens the argument.

Finally, let's revisit the problem of many premises and the way they are related. To keep things simple, consider the arguments (i) "A_1 and A_2, therefore C" and (ii) "A_1 and A_3, therefore C". Suppose that the two premises are in concert in (i) and in conflict in (ii) and that the conclusion is equally confirmed by the premises in both cases. Formally, $P(A_2|A_1) > P(A_2|\neg A_1), P(A_3|A_1) < P(A_3|\neg A_1)$ and $P(C|A_1,A_2) = P(C|A_1,A_3)$. Therefore, confirmation-based measures cannot set the two arguments apart. Coherence-based measures, on the other hand, can, and they typically consider the examples where premises are positively correlated as the stronger arguments. Consider the following two concrete examples with the designated probabilities. Clearly:

A_1: It will snow heavily in our city in March.

$P(A_1) = .31$

A_2: The streets will be icy.

$P(A_2|A_1) = .72; P(A_2|\neg A_1) = .03$

C: People will need extra time to get to work.

$P(C|A_1,A_2) = .81; P(C|A_1,\neg A_2) = .91,$

$P(C|\neg A_1,A_2) = .87, P(C|\neg A_1,\neg A_2) = .03$

is a better argument than:

A_1: It will snow heavily in our city in March.

$P(A_1) = .31$

A_3: The streets will be free of ice.

$P(A_3|A_1) = .03; P(A_3|\neg A_1) = .72$

C: People will (still) need extra time to get to work.

$P(C|A_1,A_3) = .81; P(C|A_1,\neg A_3) = .54,$

$P(C|\neg A_1,A_3) = .02, P(C|\neg A_1,\neg A_3) = .61$

The coherence-based measures give the expected assessment: $\mathcal{A}_{Coh}(C;A_1,A_2) > \mathcal{A}_{Coh}(C;A_1,A_3)$ if one uses the Shogenji measure or the improved Olsson-Glass measure. Specifically, we get argument strengths of .98 vs. -.77 and .96 vs. -.76 using the two measures, respectively. So the first argument is very good on both coherentist measures and the second very bad. However, the two arguments are equally strong on the difference measure d because $P(C|A_1,A_2) = P(C|A_1,A_3)$. [2]

[2] An objection may be raised that the probabilities used in this example do not fully correspond to the two arguments and were picked just so that $P(C|A_1,A_2) = P(C|A_1,A_3)$. Note, though, that one can also pick other numbers and still get a good argument in the first case and a bad argument in the second, while both will be evaluated as similarly good on the confirmation-based account (because the conclusion will be supported by the premises, even if not to the exact same degree). Consider, e.g., the following probability assignments: $P(A_1) = .3$, $P(A_2|A_1) = .8 = P(A_3|\neg A_1)$, $P(A_2|\neg A_1) = .05 = P(A_3|A_1)$, $P(C|A_1,A_2) = .9 = P(C|A_1,\neg A_3)$, $P(C|A_1,\neg A_2) = .8 = P(C|A_1,A_3)$, $P(C|\neg A_1,A_2) = .8 = P(C|\neg A_1,\neg A_3)$ and $P(C|\neg A_1,\neg A_2) = .05 = P(C|\neg A_1,A_3)$. In this case, the first argument is very good on a coherentist measure and the second bad (approx. .95 vs. $-.7$), while the measure d gives a similar and positive assessment for both (approx. .55 vs. .4).

This concludes our demonstration that a coherentist measure of argument strength successfully addresses the three identified weaknesses of confirmation-based measures of argument strength: (i) irrelevant premises, (ii) additional information, and (iii) the relation among many premises. Note that using Coh_{Sh} in $\mathcal{A}_{\text{Coh}}$ fails to take care of (i). Using Coh_{OG^+}, however, masters all three challenges.

Discussion

Let us now discuss some possible objections and open questions. First, it is interesting to note that the two approaches are conceptually quite different. In the confirmation-based approach that forms the core of the current theory of Bayesian argumentation, argumentation is reconstructed as a two-stage process. In the first step, the agent determines a prior probability distribution over the premises and the conclusion. In the second step, the agent *learns* the premises of the argument (i.e., they shift their marginal probabilities "by hand") and then updates them using, for example, Bayes' theorem. The idea behind this procedure is that we can distinguish between two classes of propositions: Hypotheses, which are uncertain and require confirmation, and evidence, which can become certain (e.g., through experiments or observations) and which can then be used to confirm hypotheses.

In the coherence-based approach, on the other hand, one considers the information set, consisting of the premises and the conclusion, as a whole, with the strength of the argument related to the coherence of the set. It is important to note that determining the argument strength also requires the agent to consider the situation where the premises are all false and the conclusion is true. All this, however, follows from the prior probability distribution. In a sense, then, the coherence-based approach evaluates the argument from the prior perspective, while the confirmation-based approach considers it from the posterior perspective (after *learning* the premises).

The following example highlights the difference:[3] Suppose there is a boxing match between boxer A and boxer B. A is the heavy favorite, but almost never wins by knockout. Now compare the following two arguments (1) KO: A knocks out B, hence WA: A wins, and (2) 6R: A wins most of the first 6 rounds, hence WA: A wins. From the posterior perspective (i.e., after learning KO or 6R), WA is necessarily true in (1) but only probable in (2), so confirmation-based measures evaluate (1) as the better argument. However, from the prior perspective (i.e., before learning of KO or 6R), (1) is a weaker argument than (2) because if A does not knock out B, it is still very likely that A wins. However, if A does not win most of the first 6 rounds, winning becomes much less likely. Argument (2) is therefore stronger in the absence of further evidence.

Second, it turns out that when the argument under consideration is deductive, the argument strength in the coherence-based representation is always 1 (i.e., maximal). This seems to contradict what we said earlier in our discussion of the key

[3] Thanks to an anonymous referee for mentioning this example.

example of Oaksford and Chater. However, a closer look at this example shows that the first premise ("When I turn the key, the car starts") is incomplete: it requires the additional assumption that everything else works normally. Since this cannot be assumed, the prior probability of this premise is not 1, and therefore the seemingly inevitable conclusion that "I did not turn the key" does not follow.

Third, while Eva and Hartmann (2018) have shown how conditionals that may not be propositions can be integrated into the confirmation-based approach to argument strength, the coherence-based approach is apparently based on the assumption that all premises are propositional. Therefore, in its current formulation, it is not applicable when indicative conditionals are present in the premises of the argument. However, a simple way to include conditionals is to represent them by a material conditional. This proposal is controversial (see Douven, 2015), but there is also some recent work defending material conditionals (see, e.g., Williamson, 2020). Moreover, the material conditional might be suitable for the present purposes, even if it does not provide a complete philosophical analysis of conditionals (see Eva & Hartmann, 2018).

Fourth, our approach is normative because it provides rational principles for what makes an argument good or bad. These principles motivate the measures that help us evaluate arguments. However, since abstract principles are not sacrosanct, it is an interesting question to examine whether our top-down reasoning holds up in descriptive reality. In future work, we plan to use the results of the present study to construct arguments of different strength. Then, we can compare experimental participants' assessments with our theory. If nothing else, it will be interesting to see how our coherence-theoretic proposal relates to existing experimental Bayesian research on the strength of arguments, which relies on the confirmation-based approach (see, e.g., Hahn, 2020).

Fifth, we still need to better understand why one should care about the strength of arguments at all, when all that seems to matter in practical decisions is posterior probability. Our coherence-based approach suggests at least one good reason to do so: knowing which premises are actually responsible for persuading an arguer allows one to argue more efficiently, and conclusions supported by good arguments hopefully lead to better decisions.

Appendix

Proof of Proposition 1

Let $\mathbf{S} := \{\mathrm{A}_1, \ldots, \mathrm{A_n}, \mathrm{C}\}$ and $\mathbf{S}' := \{\mathrm{A}_1, \ldots, \mathrm{A_{n+1}}, \mathrm{C}\}$. Then, using the probabilistic independence of A_{n+1}, we obtain:

$$\begin{aligned}
\mathrm{Coh_{Sh}}(\mathbf{S}') &= \frac{P(\mathrm{A}_1, \ldots, \mathrm{A_n}, \mathrm{A_{n+1}}, \mathrm{C})}{P(\mathrm{A}_1)\cdots P(\mathrm{A}_n)P(\mathrm{A}_{n+1})P(\mathrm{C})} \\
&= \frac{P(\mathrm{A}_{n+1})P(\mathrm{A}_1, \ldots, \mathrm{A_n}, \mathrm{C})}{P(\mathrm{A}_1)\cdots P(\mathrm{A}_n)P(\mathrm{A}_{n+1})P(\mathrm{C})} \\
&= \frac{P(\mathrm{A}_1, \ldots, \mathrm{A_n}, \mathrm{C})}{P(\mathrm{A}_1)\cdots P(\mathrm{A}_n)P(\mathrm{C})} \\
&= \mathrm{Coh_{Sh}}(\mathbf{S})
\end{aligned}$$

Similarly, we find that $\mathrm{Coh_{Sh}}(\mathbf{S}'^{\dagger}) = \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger})$ where $\mathbf{S}^{\dagger} := \{\neg\mathrm{A}_1, \ldots, \neg\mathrm{A_n}, \mathrm{C}\}$ and $\mathbf{S}'^{\dagger} := \{\neg\mathrm{A}_1, \ldots, \neg\mathrm{A_{n+1}}, \mathrm{C}\}$. Hence, given $\mathrm{Coh_{Sh}}$ as the measure of coherence, $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \ldots, \mathrm{A_{n+1}}) = \mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \ldots, \mathrm{A_n})$. □

Proof of Proposition 2

Let $\mathbf{S} := \{\mathrm{A}_1, \mathrm{C}\}$, $\mathbf{S}^{\dagger} := \{\neg\mathrm{A}_1, \mathrm{C}\}$, $\mathbf{S}' := \{\mathrm{A}_1, \mathrm{A}_2, \mathrm{C}\}$ and $\mathbf{S}'^{\dagger} := \{\neg\mathrm{A}_1, \neg\mathrm{A}_2, \mathrm{C}\}$. As A_2 is probabilistically independent of A_1 and C, we parameterize the probability distribution over A_1, A_2 and C as follows: $P(\mathrm{A}_1) = a$, $P(\mathrm{C}|\mathrm{A}_1) = p$, $P(\mathrm{C}|\neg\mathrm{A}_1) = q$ and $P(\mathrm{A}_2) = a'$. Then, using Proposition 1 and $P(\mathrm{C}) = ap + \overline{a}q$ with the shorthand $\overline{x} := 1 - x$, we obtain:

$$\begin{aligned}
\mathrm{Coh_{OG^+}}(\mathbf{S}) &:= \frac{1-\overline{a}\,\overline{c}}{1-\overline{a}\,\overline{q}} \cdot \mathrm{Coh_{Sh}}(\mathbf{S}) \\
\mathrm{Coh_{OG^+}}(\mathbf{S}^{\dagger}) &:= \frac{1-a\,\overline{c}}{1-a\,\overline{p}} \cdot \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger}) \\
\mathrm{Coh_{OG^+}}(\mathbf{S}') &:= \frac{1-\overline{a'}\,\overline{a}\,\overline{c}}{1-\overline{a'}\,\overline{a}\,\overline{q}} \cdot \mathrm{Coh_{Sh}}(\mathbf{S}) \\
\mathrm{Coh_{OG^+}}(\mathbf{S}'^{\dagger}) &:= \frac{1-a'\,a\,\overline{c}}{1-a'\,a\,\overline{p}} \cdot \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger})
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1) &= \frac{z \cdot \mathrm{Coh_{Sh}}(\mathbf{S}) - \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger})}{z \cdot \mathrm{Coh_{Sh}}(\mathbf{S}) + \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger})} \\
\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \mathrm{A}_2) &= \frac{z' \cdot \mathrm{Coh_{Sh}}(\mathbf{S}) - \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger})}{z' \cdot \mathrm{Coh_{Sh}}(\mathbf{S}) + \mathrm{Coh_{Sh}}(\mathbf{S}^{\dagger})}
\end{aligned}$$

with

$$\begin{aligned}
z &:= \frac{1-\overline{a}\,\overline{c}}{1-\overline{a}\,\overline{q}} \cdot \frac{1-a\,\overline{p}}{1-a\,\overline{c}} \\
z' &:= \frac{1-\overline{a'}\,\overline{a}\,\overline{c}}{1-\overline{a'}\,\overline{a}\,\overline{q}} \cdot \frac{1-a'\,a\,\overline{p}}{1-a'\,a\,\overline{c}}.
\end{aligned}$$

As the function $f(x) := (x-\alpha)/(x+\alpha)$ is strictly monotonically increasing in x for $\alpha > 0$, we conclude that $\mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1) > \mathcal{A}_{\mathrm{Coh}}(\mathrm{C}; \mathrm{A}_1, \mathrm{A}_2)$ iff $z > z'$.

Let's now define the function $g(x) := (1-\alpha x)/(1-\beta x)$ which is strictly monotonically increasing in x for $\beta > \alpha$. Hence, the first fraction in the expression of z is greater than the first fraction in the expression of z' iff $\overline{q} > \overline{c}$ which is equivalent to $p > q$. Similarly, the second fraction in the expression of z is greater than the second fraction in the expression of z' iff $\overline{c} > \overline{p}$ which is also equivalent to $p > q$. □

References

Adams, E. (1975). *The Logic of Conditionals*. Boston: Reidel.

Adams, E. (1996). *A Primer of Probability Logic*. Chicago: The University of Chicago Press.

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

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

Dawid, R., Hartmann, S., & Sprenger, J. (2015). The no alternatives argument. *The British Journal for the Philosophy of Science*, *66*(1), 213–234.

Douven, I. (2012). Learning conditional information. *Mind & Language*, *27*(3), 239–263.

Douven, I. (2015). *The Epistemology of Indicative Conditionals: Formal and Empirical Approaches*. Cambridge: Cambridge University Press.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235–329.

Eells, E. (1982). *Rational Decision and Causality*. Cambridge: Cambridge University Press.

Eva, B., & Hartmann, S. (2018). Bayesian argumentation and the value of logical validity. *Psychological Review*, *125*(5), 806.

Eva, B., Hartmann, S., & Rad, S. R. (2020). Learning from conditionals. *Mind*, *129*(514), 461–508.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, *66*(S3), S362–S378.

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

Gillies, D. (1986). In defense of the Popper-Miller argument. *Philosophy of Science*, *53*(1), 110–113.

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* (p. 177–82). Berlin: Springer.

Günther, M., & Trpin, B. (2022). Bayesians still don't learn from conditionals. *Acta Analytica*, 1–13.

Hahn, U. (2020). Argument quality in real world argumentation. *Trends in Cognitive Sciences*, *24*(5), 363–374.

Hahn, U., & Oaksford, M. (2006). A normative theory of argument strength. *Informal Logic*, *26*(1), 1–24.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychological Review*, *114*(3), 704.

Jeffrey, R. (1992). *Probability and the Art of Judgment*. Cambridge University Press.

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.

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

Oaksford, M., & Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Olsson, E. J. (2002). What is the problem of coherence and truth? *Journal of Philosophy*, *99*(5), 246–72.

Olsson, E. J. (2021). Coherentist theories of epistemic justification. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University.

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

Rosenkrantz, R. (1994). Bayesian confirmation: Paradise regained. *The British Journal for the Philosophy of Science*, *45*(2), 467–476.

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

Williamson, T. (2020). *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford: Oxford University Press.