Approximate coherence: a new accuracy-centered account

Giacomo Molinari

February 8, 2023

1 Introduction

We say that your credences are *coherent* when they respect the laws of probability. Many epistemologists argue that coherence has the status of a rationality norm. A rational agent, they claim, ought to have coherent credences.

The coherence norm has been criticised for being excessively idealised. For example, if an agent has a credence towards a logically true statement, the laws of probability require her to have maximal confidence in its truth, no matter how complex the statement. Clearly this goes well beyond the ability of any human reasoner. And if coherence is a norm we can't respect, why should we care about it? The aim of this essay is to respond to this objection to the coherence norm. I will argue that, even if you can't be coherent, you should approximate coherence as best you can. To build such a response, I will first specify a notion of approximate coherence, and then show that this notion is normatively relevant, i.e. that non-ideal agents ought to approximate coherence.

Here is the essay plan. I start in Section 2 by summarising two popular styles of argument for the coherence norm: one pragmatic (Dutch Book argument) and one purely epistemic (accuracy argument). This essay will mostly focus on the accuracy side, although I will point out where similar ideas have been or can be applied on the pragmatic side. In Section 2 I also outline the idealisation objection to the coherence norm and the defense strategy I want to pursue: first define a notion of approximate coherence, and then show it has normative bite. Section 3 presents De Bona and Staffel's (2017) accuracy-centered defense of the coherence norm, which follows a similar two-step strategy. I will argue that this proposal fails to give evaluative meaning to many of the comparative incoherence judgements it produces, and that this compromises the normative relevance of approximate coherence. In Section 4 I give a new accuracy-based definition definition of approximate coherence, and show that this is formally equivalent to the one given by De Bona and Staffel. Taking advantage of this new definition, we can cash out every comparative incoherence judgement in terms of accuracy. I use this fact to show, in Section 5, why non-ideal agents should approximate coherence as best they can. My argument here relies on the assumption that it's better to miss out on less, rather than more, guaranteed accuracy. Since this is not obviously the case, the rest of Section 5 looks at an alternative argument for the same conclusion, based on the way non-ideal agents trade-off computational costs for accuracy. I show that, when trade-off considerations motivate comparative judgements about credence functions, these are the same judgements produced by the assumption above, and thus support my argument for the normative relevance of approximate coherence. I end the essay by summing up the main conclusions in Section 6.

2 Ideal norms and non-ideal agents

This section introduces the coherence norm, together with two arguments commonly used to support it. Then I present the objection that coherence is inadequate as a rationality norm due to its excessively idealised nature, and outline the response strategy I want to pursue.

2.1 Coherence and its justifications

Let's start by introducing some notation. I will consider agents who have opinions towards some finite set $\mathcal{F}$ of sentences in a propositional language which includes all the usual connectives. We can capture these opinions by a *credence function* which assigns to each sentence in $\mathcal{F}$ a real number in $[0,1]$, representing the agent's degree of belief in that sentence. I will denote by $\mathcal{W}_{\mathcal{F}}$ the set of "possible worlds" corresponding to all the logically possible truth-value assignments for the sentences in $\mathcal{F}$. If $w \in \mathcal{W}_{\mathcal{F}}$ and $\phi \in \mathcal{F}$, then $w(\phi)$ is the truth-value of ϕ at w.[1]

Denote by $\mathcal{F}^*$ the smallest algebra containing $\mathcal{F}$. A credence function is *coherent* iff it can be extended to a probability function on $\mathcal{F}^*$. That is, iff there is a function $p : \mathcal{F}^* \to [0,1]$ such that $p(\phi) = c(\phi)$ for every $\phi \in \mathcal{F}$, and p respects the following conditions:

$$\begin{aligned} &\vdash \phi \implies p(\phi) = 1 \\ &\phi \vdash \implies p(\phi) = 0 \end{aligned} \tag{P1}$$

$$\phi \vdash \psi \implies p(\phi) \leq p(\psi) \tag{P2}$$

$$p(\phi \wedge \psi) + p(\phi \vee \psi) = p(\phi) + p(\psi) \tag{P3}$$

[1] There is some debate around what the objects of credence should be (Chalmers, 2011; Fitts, 2014). I pick sentences here (instead of, say, sets of possible worlds) because I am interested in modeling agents who assign different credence to logically equivalent statements (i.e. statements which are true in the exact same worlds). So long as we can model such agents, any object should work.

Many epistemologists argue that coherence has the status of a rationality norm, meaning that a rational agent *ought* to have a coherent credence function. One way to defend this claim is to show that, if your credences are incoherent, they are vulnerable to a Dutch book: a bookie can construct a package of bets which all look desirable to you individually, and yet taken together they guarantee a sure loss. On the other hand, coherent credences are not vulnerable to Dutch books. Therefore, if you care about behaving (i.e. betting) rationally, you ought to be coherent (Pettigrew, 2020; de Finetti, 1964).

Starting with Joyce (1998), another style of argument has been used to justify the coherence norm. Unlike the Dutch-book argument, it does not exploit the pragmatic implications of credences. Instead, it aims to evaluate credences in terms of how accurate they are. If a credence function c is incoherent, then it is *accuracy-dominated*: there is some credence function c' which is guaranteed to be more accurate than c, come what may (i.e. in any possible world). Conversely, coherent credence functions are immune from accuracy-domination. Therefore, if you care about being accurate, you ought to be coherent.

For this argument to work we need to specify a reasonable class of functions $S(c, w)$ to measure the inaccuracy of credence function c at world w. A popular choice is the class of strictly proper, additive inaccuracy measures. To define them, first define a *strictly proper scoring rule* as a continuous function $s : [0, 1] \times \{0, 1\} \to \mathbb{R}$ such that, for any $p \in [0, 1]$, the following function of x is uniquely maximised at $x = p$:

$$ps(x, 1) + (1 - p)s(x, 0) \tag{1}$$

Intuitively, the inaccuracy of a credence x in ϕ is given by $s(x, 1)$ when ϕ is true, and by $s(x, 0)$ when ϕ is false. Then, if p is your credence that ϕ is true and $(1 - p)$ your credence that ϕ is false, you expect your own credence to be the most accurate.

Strictly proper, additive inaccuracy measures can then be defined as:

$$S(c, w) = \sum_{\phi \in \mathcal{F}} s(c(\phi), w(\phi)) \tag{2}$$

where s is a strictly proper scoring rule. These measures inherit the characteristic feature of proper scoring rules:

- **Strict Propriety**: If c is coherent, then for every $c' \neq c$, $Exp_c[S(c, \cdot)] < Exp_c[S(c', \cdot)]$. In other words, every coherent credence function expects itself to be the most accurate.

Joyce's (1998) argument for the coherence norm relies on the formal result that, when inaccuracy is measured by a strictly proper, additive measure, c is incoherent if and only if it is accuracy-dominated. That is, if and only if there is some c' that is less inaccurate than c in every world:

$$S(c, w) > S(c', w) \text{ for every } w \in \mathcal{W}. \tag{3}$$

Thus a rational agent interested in accuracy ought to have coherent credences.

Although both Dutch Book and accuracy arguments establish coherence as a rationality norm, they don't say anything about whether there is any reason for incoherent agents to approximate coherence. In particular, if you are unable to be coherent, it's not clear these arguments have anything to say about you, besides that you are unable to be rational. This has led some to doubt that coherence can be a rationality norm for non-ideal agents.

2.2 Spelling out the objection

Consider an agent interested in some remote digit of π.[2] Let ϕ_i, for $i = 0, ...9$, be the sentence that the digit is i. If the agent knows the basics of arithmetic, then coherence demands that the agent have credence 1 in the true sentence, and 0 in the others. But if we consider a digit remote enough, no human agent can ensure they respect this constraint. Indeed, given that we cannot compute the digit, it would be odd to have extremal credence in some ϕ_i. It seems that we are irrational, we know we are irrational, and we can't do anything about it.

Even if we restrict ourselves to sentences in a propositional language, as I will do in this essay, we can find analogous examples. Just think of a sentence ϕ complex enough that it would take an extremely long time to determine its truth-value, given any truth-value assignment to the atomic sentences, but which ends up being true under any such assignment. Since ϕ is a tautology, coherence again requires that we have credence 1 in it, even though finding out that ϕ is a tautology may be beyond our abilities. The example also extends to complex bundles of logical and probabilistic relations. Consider a large number of statements about the results of the next election (who wins, by how much, etc...), and a large number of statements about how various economic indicators (country's GDP, price of oil, etc...) will change before the election. I have a degree of belief for each individual statement, and in many cases, for one statement conditional on another. Yet if I put all of these opinions together I do not get a coherent credence function. Up to this point, it seems, I have been irrational. Even worse: after realising this fact, though I might adjust some opinions here and there, I soon find that correcting all incoherences is impossibly difficult, and I give up. I remain irrational from this point onwards.

The upshot is that it takes a perfect calculator, capable of working out arbitrarily complex logical and probabilistic relations, to ensure compliance with the coherence norm. This gives rise to the objection that coherence is inappropriate as a norm for human-like agents (Savage, 1967; Foley, 1993). One way to spell out this criticism employs an *ought-implies-can* assumption (OIC): intuitively, a constraint has normative

[2]This example is due to Savage (1967).

force over us only if we can ensure we comply with it.[3] If this is the case, then the above examples show that coherence cannot be a rationality norm for human-like reasoners. I call this the **Coherence Impossible** objection.

A natural response is that coherence, although not attainable, is still normatively relevant to us, serving as an ideal which we ought to approximate to the best of our ability. Zynda (1996) makes this point using the analogy of a machinist who works on a lathe, and receives a specification for a cylindrical component. Clearly, no matter how skilled the machinist and how technically advanced the lathe, the component will never be perfectly cylindrical. And yet, the constraint is still normatively relevant for the machinist. This is because: (a) it allows the machinist to determine that certain components (actual or hypothetical) are better than others, since they are more approximately cylindrical; (b) conditional on the manufacturing technology improving, the machinist is subject to an obligation to produce more approximately cylindrical parts; and (c) this ideal requirement gives the manifacturing industry a reason to improve their techniques. Zynda argues that coherence plays a similar role as the machinist's ideal constraint. Even though we cannot achieve it, it still has normative force over us.

For this response to be convincing, however, we need to say a bit more. First, we must specify *what it means* for a credence function to approximate coherence. Just like there are many ways in which a component can be said to have approximately cylidrical shape, there are many ways in which a credence function can be said to be approximately coherent. We need to specify which notion of approximate coherence is normatively relevant. And secondly, we need to show *why* approximate coherence is normatively relevant. In the machinist case, we can imagine that the component will be part of some larger mechanism. The more accurately it approximates the ideal shape and dimension, the better it will be at performing its intended function (by fitting in with the other pieces, by reducing attrition, etc...); in other words, judgements that one piece more closely approximates the ideal than another are evaluatively meaningful. Similar considerations explain why the machinist has a conditional obligation to more accurately approximate the cylinder as her abilities improve, and give her a reason to improve her abilities. To show our notion of approximate coherence is normatively relevant, we must similarly show that it leads to evaluatively meaningful comparisons, that it induces conditional obligations, and that it explains how the coherence ideal motivates agents to improve their epistemic abilities.

My response to the Coherence Impossible objection will thus involve two steps:

- **Step 1**: Specifying a notion of *approximate coherence.*
- **Step 2**: Showing that this notion is *normatively relevant.* This involves showing: (a) that it produces evaluatively meaningful comparative judgements, i.e. that

[3]For a thorough discussion of OIC assumptions and normative idealisation in epistemology, see Carr (2021).

approximately coherent credences are in some sense better than wildly incoherent ones; (b) that, conditional on her abilities improving, the agent is required to better approximate coherence; and (c) that it gives the agent a reason to improve her abilities.

Since we use Dutch Book or accuracy considerations to justify the (ideal) coherence norm, it's natural to appeal to the same considerations when explaining why agents ought to approximate coherence (Step 2). I will touch on both approaches in the next sections, although my main focus will be on the accuracy one.

3 Approximate coherence for accuracy-firsters

This section looks at De Bona and Staffel's (2017) proposal for an accuracy-centered defense of the coherence norm for non-ideal agents.[4] De Bona and Staffel follow the two-step strategy discussed in the previous section: first define a notion of approximate coherence, and then explain why rational agents ought to be approximately coherent. However, I will argue that this explanation only gives normative relevance to a weaker notion of approximate coherence than the one they define.

3.1 How to approximate coherence

De Bona and Staffel (2017) define a quantitative notion of approximate coherence: they specify an *incoherence measure* which assigns a real number to each credence function, and say that c is more approximately coherent than c' iff it has lower incoherence measure. To define this measure, Staffel (2020) first gives some desiderata that it should satisfy. I will list them here, since they will be relevant for my argument later on. The first is *judgement preservation.*

- **(D1) Judgement Preservation**
  If c and c' are credence functions such that c is clearly more coherent than c', then our measure should rank c as less incoherent than c'.

The simplest example of this desideratum is the case where c is coherent, and c' is not. In such a case, there is no doubt that c is more coherent than c'; therefore, (D1) implies that our measure should rank any coherent credence function as less incoherent than any incoherent one. Other cases may not so clear-cut, and there might be room for disagreement about whether a credence function is "clearly" more coherent than another. But insofar as such judgements can be made (more or less confidently) they

[4]This proposal is further expanded by Staffel (2020), so I will refer to both sources throughout this section.

will constrain the kind of incoherence measure which can (more or less confidently) be considered reasonable.

The second and third desiderata are *incompleteness* and *comparability*:

- **(D2) Incompleteness**
  An incoherence measure should apply even to credence functions that are not defined over a full algebra of events.

- **(D3) Comparability**
  If two credence functions c, c' are intuitively comparable in terms of their incoherence, our measure should be able to compare them. For example, if c and c' are defined on the same domain, we should be able to determine whether one is more incoherent than the other.

We want Incompleteness to hold because we are interested in assessing the credence functions of non-ideal agents, who might not have opinions towards every sentence in a (potentially very large) algebra. Comparability will be especially important to my argument in later sections. This is a key requirement for ensuring that our incoherence measures play a substantive theoretical role. In some sense it is the converse of (D1): whereas (D1) requires that we do not contradict incoherence comparisons which we (intuitively) already know to be true, Comparability (D3) requires the measure to make interesting, non-obvious comparisons. Like (D1), it's not always clear which credence functions should count as comparable: but comparing functions with the same domain seems like a good starting point.

The last requirement is *no-inundation*:

- **(D4) No inundation**
  If c is a credence function defined over a domain $\mathcal{F}$, there should not be a strict subset of $\mathcal{F}$ such that the values of c in this subset affect c's incoherence measure more strongly than the values of c on the other elements of $\mathcal{F}$.

(Staffel, 2020, p-37) motivates this desideratum by analogy with measures of the overall wealth of a country. If we measure it by taking the average income of, say, the richest 1% of the population, the results will be misleading. A good measure should take into account the wealth of every inhabitant, or rely on representative samples. The same holds for our measures of incoherence: we should evaluate c as a whole, and not focus on an unrepresentative subset of its values.

After spelling out these desiderata, Staffel (2020) defines a class of measures of incoherence which satisfy them. She proposes that we measure the incoherence of a credence function c as follows:

$$I_d(c) = \min\{d(c, y) : y \text{ is a coherent credence function}\} \tag{4}$$

where d is a continuous divergence. For their accuracy-centered defense of approximate coherence, De Bona and Staffel (2017) focus on divergences derived from additive, stricly proper inaccuracy measures. Recall that such measures can be defined as:

$$S(c, w) = \sum_{\phi \in \mathcal{F}} s(c(\phi), w(\phi)) \tag{5}$$

where s is a strictly proper scoring rule. To each score of this kind we can associate a corresponding divergence (defined for arbitrary c and coherent c'):

$$d_S(c, c') = S(c, c') - S(c', c') = \sum_{w \in \mathcal{W}_{\mathcal{F}}} c'(w)S(c, w) - \sum_{w \in \mathcal{W}_{\mathcal{F}}} c'(w)S(c', w) \tag{6}$$

where $S(c, c')$ is a shorthand for $Exp_{c'}[S(c, \cdot)]$. Note that $d_S(c, c')$ is just the expected value under c' of the difference between the score of c and the score of c'. Also, if $c' = w$ for some $w \in \mathcal{W}_{\mathcal{F}}$, then $d_S(c, w) = S(c, w)$.

To each strictly proper, additive inaccuracy measure S, we can thus associate a corresponding incoherence measure I_{d_S} defined from its divergence (I will slightly abuse the notation and write I_S instead of I_{d_S}). Staffel (2020) shows that incoherence measures constructed in this way respect the four desiderata (D1-D4). After defining a notion of approximate coherence which fits our desiderata, we need to show that this notion is normatively relevant. This involves showing that it is in some way epistemically better for agents to be less, rather than more, incoherent.

Draft

3.2 Why approximate coherence

The accuracy argument for the coherence norm shows that incoherent credences have some fault which coherent credences are immune from: they are accuracy-dominated, hence bad at pursuing accuracy. However, this argument does not justify the claim that incoherent agents should try to approximate coherence.

De Bona and Staffel (2017) set out to address this point, by showing that "there is in fact a specific way of reducing incoherence in one's credences that always constitutes an improvement of one's epistemic situation" (p.203), where the epistemic improvement amounts to a guarenteed increase in accuracy. I will argue that, although their argument does show that a certain notion of approximate coherence is normatively relevant, this is a weaker (in a sense soon to be made precise) notion than the one captured by their incoherence measures.

Their argument is centered around the following main result:

Proposition 1. *Let S be a convex,*[5] *strictly proper, additive inaccuracy measure. Let $c : \mathcal{F} \to \mathbb{R}$ be an incoherent credence function and let $c^* : \mathcal{F} \to \mathbb{R}$ be the d_S-closest*

[5] An inaccuracy measure S is (everywhere) *convex* if for every $w \in \mathcal{W}_{\mathcal{F}}$, every $c, c' : \mathcal{F} \to \mathbb{R}$, and every $\lambda \in [0, 1]$, the inequality $S(\lambda c + (1 - \lambda)y, w) \leq \lambda S(c, w) + (1 - \lambda)S(c', w)$.

coherent credence function to c. *For any* $\lambda \in (0,1]$, *define* $c_\lambda : \mathcal{F} \to \mathbb{R}$ *as* $c_\lambda(\theta) = (1-\lambda)c(\theta) + \lambda c^*(\theta)$. *Then:*

1. c_λ *strongly accuracy-dominates* c *under* S.

2. $I_S(c_\lambda) < I_S(c)$.

This result shows that moving from an incoherent credence function c to some c_λ on the path to its d_S-closest coherent credence function c^*, is guaranteed to increase accuracy.[6] Hence, for any incoherent credence function, there is a way of reducing incoherence which improves the agent's epistemic situation.

The problem is that this argument only motivates a *weaker* notion of approximate coherence than the one captured by the measure I_S; by which I mean a notion that induces strictly less incoherence comparisons. Say we measure incoherence by I_S, where S is our preferred (convex) measure of inaccuracy. Now pick credence functions c, c', such that $I_S(c') < I_S(c)$, i.e. c' is less incoherent than c. What is the evaluative meaning of this comparative judgement? Well, if c' happens to be on the direct path from c to its d_S-closest coherent credence function, then we can say c' is epistemically better than c under the accuracy-centered view of epistemic good, thanks to Proposition 1. But in general, if c' is not along this path, Proposition 1 gives us no reason to think that it is better than c under any evaluative dimension.[7] So even though I_S can compare any two credence functions defined on the same domain, as per Comparability desideratum (D3), the argument above only shows that a small subsets of these comparisons are evaluatively meaningful.

But we want our notion of approximate coherence to make evaluatively meaningful comparisons in order to explain why it is normatively relevant to non-ideal agents. In the machinist analogy, the fact that more approximately cylindrical components are better at fulfilling their goals is a key part of our story about why the ideal cylinder specification, although unreachable, still has normative force. Hence the argument based on Proposition 1 can at best show that the incoherence measure I_S is as normatively relevant as the following, qualitative notion of incoherence:

- **Weak Approximate Coherence (WAC)**: c' is less incoherent than c iff c' is guaranteed to be more accurate than c (under our preferred inaccuracy measure S).

[6] In a subsequent paper, De Bona and Staffel (2018) extend the result to also show that c_λ (defined as in Proposition 1) can be Dutch booked for a smaller (normalised) sure loss than c. So it's possible to reduce incoherence while both increasing accuracy and reducing normalized Dutch book loss. For clarity of exposition, I prefer to keep the two results separate.

[7] This fact is also aknowledged in De Bona and Staffel (2017). [Might say more here about subsequent work (De Bona and Staffel (2018); Staffel (2020)) being focused on finding ways of reducing incoherence that are beneficial from both Dutch Book and accuracy side, rather than strengthening the accuracy side]

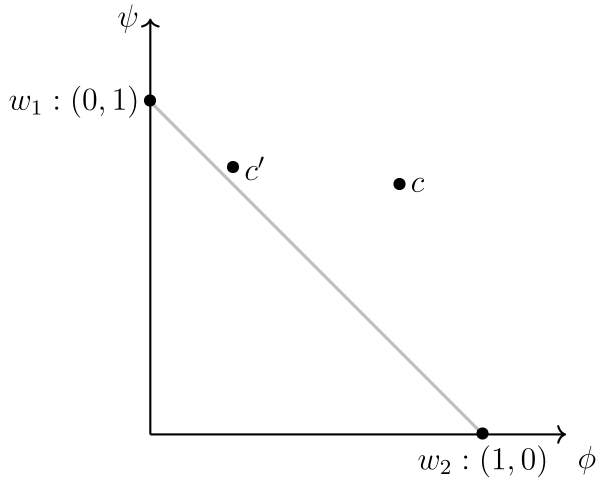

Figure 1: Here $\mathcal{F} = \{\phi, \psi\}$, where $\phi \equiv \neg\psi$, so we can draw each credence function as a point in the plane. The x coordinate is the credence in ϕ, and the y coordinate is the credence in ψ. The light gray region is that of coherent credence functions, for which $c(\phi) = 1 - c(\psi)$.

One could respond to this criticism by biting the bullet, conceding that Comparability (D3) is overly demanding. There is a useful accuracy-centered notion of approximate coherence, we might say, and a sense in which approximating coherence is good. But this notion is no stronger than (WAC). Indeed, this notion is so weak that all its comparisons might follow from the Judgement Preservation (D1) desideratum. More worryingly for this response strategy, this notion is so weak that it might fail to satisfy the Judgement Preservation (D1) desideratum itself. For example, consider the two credence functions in Figure 1. It seems that c' is doing a better job at approximating coherence than c. Yet, under most reasonable measures of inaccuracy, c' does not accuracy-dominate c. So according to this weak notion of approximate coherence, neither credence function is more incoherent than the other. My view is that a stronger notion of approximate coherence than (WAC) can be shown to be normatively relevant from the accuracy perspective, one which preserves and gives evaluative meaning to intuitive judgements like those of Figure 1. The remainder of the essay is devoted to defending this claim.

4 Rethinking approximate coherence

Although De Bona and Staffel's (2017) define incoherence measures which formally respect the Comparability (D3) desideratum, allowing us to compare the degree of incoherence of any two credence functions, they do not show the evaluative force of many of these comparisons. This compromises the normative relevance of their notion of approximate coherence. In this section, I give a new definition of incoherence measure which makes explicit the accuracy implications of every incoherence comparison. Interestingly, my incoherence measures turn out to be equivalent to the ones defined by De Bona and Staffel. The definition in terms of accuracy can then be used to show that every incoherence comparison induced by these measures is evaluatively meaningful, and therefore, that a stronger notion of approximate coherence than (WAC) can be normatively relevant from the accuracy perspective.

Let's start by looking at the evaluative force we attach to judgements of standard, bivalent coherence. If c' is coherent while c is not, there is some sense in which c' is *epistemically better* than c. But we should be careful in spelling out what this means. Crucially, it does *not* mean that c' is *guaranteed to be better* than c in terms of accuracy. Nor does it mean that an agent with credence function c ought to adopt credence function c', nor that c' is all-things-considered better than c.

So in what sense are coherent credences "better" than incoherent ones? A good way to think of this is in terms of faults. Incoherent credences are guaranteed to be accuracy-dominated, an Dutch-bookable, whereas coherent ones are guaranteed to not be accuracy-dominated, and to not be Dutch-bookable. Being accuracy-dominated and being Dutch-bookable are both *epistemic faults*, under the assumption that the goal of credences is to be accurate, and that credences guide betting behaviour, respectively. Hence, both in the pragmatic and in the purely epistemic view, all incoherent credences have a fault from which all coherent credences are immune.

If being incoherent amounts to having a certain epistemic fault, and being coherent amounts to not having that fault, we might characterise gradational incoherence by distinguishing faulty credences in terms of "how serious" their faults are. Before I use this strategy to define my incoherence measures in terms of accuracy, it's worth mentioning how the same idea has been applied from the pragmatic perspective.

4.1 Dutch Book incoherence measures

On the pragmatic side, this strategy is pursued by proponents of *Dutch-book incoherence measures* (Schervish, Seidenfeld and Kadane 1998; 2002). In their view, the epistemic fault associated with incoherence is being vulnerable to Dutch-books, i.e. being disposed to accept a package of bets which guarantees a sure loss. The idea is to distinguish between more/less faulty credence functions on the basis of the *amount of sure loss* they are vulnerable to.

There is one major obstacle for this idea. To see it, assume an agent's credence function c is incoherent. We know there is a set of bets $\mathcal{B} = \{B_1, ..., B_n\}$ which c accepts, and which guarantee a sure loss, i.e. a Dutch Book. Say that, by accepting all bets in $\mathcal{B}$, the agent is guaranteed to lose t£, come what may. But then c also accepts the set of bets $\mathcal{B}' = \{2B_1, ..., 2B_n\}$, where $2B_i$ is the same bet as B_i but with doubled payouts. And this set of bets guarantees a loss of $2t$£. Since we can do this for an arbitrary constant, any agent vulnerable to a sure loss of k£ will be vulnerable to a sure loss of arbitrary size.

If we want to compare incoherent credence functions based on the size of their guaranteed losses, we must first specify a way to *normalise* these losses. Different normalisations lead to different notions of approximate coherence. Once we pick a normalisation, the behavioural implications of comparative incoherence claims are clear: if c is more incoherent than c' under a Dutch Book incoherence measure, then c is vulnerable to a larger normalised sure loss than c'.

Proponents of Dutch Book incoherence measures then argue that facts about normalised sure loss give evaluative meaning to the incoherence comparisons induced by these measures. That is, they argue that if c' is more approximately coherent than c according to their measure, then although both credences may be vulnerable to a Dutch Book, c' is epistemically better, because it is subject to a smaller normalised sure loss. In the next section I will discuss an accuracy-analogue of this claim; but first, I need to introduce a way to define incoherence measures in terms of accuracy.

4.2 Accuracy-based incoherence measures

There are a number of reasons why you might find Dutch Book incoherence measures unsatisfying. For example, you might be worried about the dependence of these measures on the choice of normalisation. What can we say of two credence functions if one is more incoherent than the other under a given normalisation, but the opposite is true for some other normalisation?[8] Secondly, you might just not be convinced by Dutch Book arguments for coherence: any objection against this argument strategy is an objection against this way of defining approximate coherence. Finally, one may object that the normalised sure loss of a credence function is not a good measure of that function's pragmatic value. This last objection has been put forward by Konek (forthcoming), where it is argued that the pragmatic value of a credence function is better captured by the same additive, strictly proper scores we have introduced as measures of inaccuracy. Indeed, Levinstein (2017) shows that we can interpret $S(c, w)$ as a measure of the losses an agent expects to incur if they were to act as prescribed by credence c in a range of decision problems.

For all these reasons, it makes sense to ask whether we can give a defense of approximate coherence in terms of additive, strictly proper measures, rather than in terms of normalised Dutch Book losses. In answering this question I will focus on the accuracy perspective. That is, I will interpret additive, strictly proper measures as measures of inaccuracy. However, everything I will say can be translated into pragmatic terms by

[8]See (Staffel, 2020, pp.57-67) for further discussion and criticism of some specific normalisation choices.

interpreting these measures as suggested by Levinstein (2017).

From the accuracy perspective, the fault of incoherent credence functions is that they are accuracy-dominated. So we might say that a credence function is more or less faulty based on the amount by which it is accuracy-dominated. Let's make this more precise. Say c' dominates c in accuracy according to S *by amount* t iff c' is guaranteed to be more accurate than c of at least t, come what may. We want to say that a credence function is more faulty the larger the amount by which some other credence function dominates it. This is captured by the following measure of incoherence:[9]

$$\underline{I}_S(c) = \sup\{\min_{w \in \mathcal{W}}\{S(c, w) - S(f, w)\} \ : \ f \text{ is coherent}\} \tag{7}$$

The following results is an immediate consequence of the definition of $\underline{I}_S$, and shows the accuracy implications of incoherence comparisons under $\underline{I}_S$:

Proposition 2. *Let S be an additive, strictly proper inaccuracy measure, and let $c : \mathcal{F} \to \mathbb{R}$ a credence function such that $\underline{I}_S(c) = t$. Then:*

1. *For every $c' : \mathcal{F} \to \mathbb{R}$, there is a $w \in \mathcal{W}_{\mathcal{F}}$ such that:*

$$S(c, w) - S(c', w) \leq t \tag{8}$$

2. *There is a $c' : \mathcal{F} \to \mathbb{R}$ such that $S(c, w) - S(c', w) \geq t$ on every $w \in \mathcal{W}_{\mathcal{F}}$.*

Proof. See Appendix. □

This shows that, if $\underline{I}_S(c) < \underline{I}_S(c')$, then c is less faulty than c' in the sense outlined above, i.e. c is accuracy-dominated by a smaller amount than c' under S.

The main formal result of this section shows that my incoherence measures are formally equivalent to those defined by De Bona and Staffel, and hence respects desiderata (D1-D4):

Proposition 3. *Let S be an additive, strictly proper inaccuracy measure, and $\mathcal{F}$ a finite set of sentences. Then for every credence function $c : \mathcal{F} \to \mathbb{R}$, we have:*

$$\underline{I}_S(c) = I_S(c). \tag{9}$$

where $I_S(c)$ is the incoherence measure defined from divergence d_S as in (4).

Proof. See Appendix. □

[9] Note that, just like Dutch Book incoherence is defined relative to a normalisation, my notion is defined relative to an inaccuracy measure. But it's arguably easier to defend a choice of inaccuracy measure than a choice of normalisation, whether one thinks that there is only one appropriate measure of inaccuracy (Pettigrew, 2016), or that different measures are appropriate in different epistemic contexts (Joyce, 2010; Babic, 2019).

De Bona and Staffel's definition of approximate coherence captures the intuition that c approximates coherence to the extent that it is "close" to the set of coherent credence functions, whereas my definition is based on the idea of partial epistemic fault I have outlined at the start of this section. However, our incoherence measures $\underline{I}_S$ and I_S turn out to be formally equivalent.

This is a first step towards answering the criticism I outlined in Section 3 that many comparative incoherence judgements induced by De Bona and Staffel's measures seem to lack evaluative meaning. Proposition 3 shows that, if $I_S(c') < I_S(c)$, then $\underline{I}_S(c') < \underline{I}_S(c)$. And this latter comparison can be immediately expressed in terms of accuracy via Proposition 2: it means that c' is accuracy-dominated by a smaller amount than c. As for Dutch book incoherence measures, where incoherence comparisons are cashed out in terms of normalised sure loss, we can claim that this shows a sense in which c' is epistemically better than c, i.e. that this shows incoherence comparisons are evaluatively meaningful. As I show in the next section, this in turn leads to a justification for the normative relevance of approximate coherence.

5 The normative relevance of approximate coherence

In this section, I will argue that the notion of approximate coherence captured by my accuracy-based incoherence measures is normatively relevant for non-ideal agents. This involves showing that: (a) less incoherent credences are epistemically better than more incoherent ones (even when the former don't accuracy-dominate the latter); (b) conditional on their abilities improving, agents ought to better approximate coherence; and (c) approximating coherence gives agents a reason to improve their abilities.

The first justification I offer is based on the accuracy meaning of incoherence comparisons which was established in the previous section. The main problem for this proposal is that it's not obvious that cashing out incoherence comparisons in terms of accuracy suffices to give them evaluative force. To address this point, I look at whether thinking of non-ideal agents as trading-off epistemic utility for computational costs can give us an alternative motivation for the evaluative meaning of incoherence comparisons. I will give a simple model of how non-ideal agents trade-off computational efforts for epistemic goods. I then use this model to show that, when our trade-off considerations motivate comparative judgements about credence functions, these are the same comparative judgements induced by my measures.

5.1 The accuracy criterion of evaluation

Proposition 2 shows that, whenever $\underline{I}_S(c') < \underline{I}_S(c)$, then c' is accuracy-dominated by a smaller amount than c. As for the Dutch book incoherence measures, we might take this to mean that c is epistemically worse than c' in an important way. If being Dutch bookable is an epistemic fault, it seems reasonable that being vulnerable to larger normalised Dutch book is being faulty to a larger degree. Similarly, if missing out on some guaranteed accuracy is a fault, it seems reasonable that one is more faulty the more guaranteed accuracy they miss out on. This leads to the following evaluation criterion:

- **Accuracy criterion**:
  Let c, c' be two incoherent credence functions defined over the same domain. If c' is accuracy-dominated by a larger amount than c is, then c' is more epistemically faulty than c'.

The accuracy criterion establishes one way in which (a) incoherence comparisons are evaluatively meaningful. It is better to be less rather than more incoherent, because this means being accuracy-dominated by a smaller rather than a larger amount (Proposition 2), and hence being less rather than more faulty. This also shows why (b) conditional on their ability to approximate coherence improving, incoherent non-ideal agents should approximate coherence more closely: it's because this would reduce their degree of epistemic faultiness. The same considerations also (c) give non-ideal agents a reason to improve their abilities, as a way to reduce their epistemic faults. Taken together, these facts show that the notion of approximate coherence captured by my incoherence measure is normatively relevant.

This justification of the normative relevance of approximate coherence has one main problem. Even if accuracy is the primary epistemic virtue, it's not obviously better to leave less of it on the table rather than more (i.e. to be accuracy-dominated by a smaller rather than larger amount). The accuracy criterion above does not follow from the weaker assumption used to justify bivalent coherence, that leaving *some* amount of guaranteed accuracy on the table is an epistemic fault. Beyond being more intuitive, this weaker assumption can be further motivated by the fact that any ideal calculator would not leave any amount of accuracy on the table, whereas there doesn't seem to be an analogous motivation for the accuracy criterion.[10] So while the amount of guaranteed accuracy different credences leave on the table might give us a *prima facie* reason for thinking one is better than the other, it would be helpful to find other reasons to motivate the same comparative judgements.

I will now look at whether such additional reasons can be found by taking into account the computational costs required to ensure coherence. It seems natural to think of non-ideal agents as engaging in trade-offs between epistemic virtue (i.e. accuracy) and the computational costs involved in complex calculations. Everyone has some non-epistemic interests and goals in their life (even epistemologists), and pursuing them often requires one to sacrifice the pursuit of truth, to some extent. Indeed, much of the intuitive appeal of the Coherence Impossible objection comes from the fact that we ordinarily

[10] For Dutch Book incoherence measures, one can similarly question whether it's any better to be vulnerable to a smaller rather than larger normalised Dutch Book.

excuse non-ideal agents for their incoherent credences, when reaching coherence would require extremely demanding calculations. As the costs of ensuring coherence grow, it seems more and more reasonable for a non-ideal agent to remain incoherent, rather then sacrifice an excessive amount of resources to obtain some epistemic virtue.

In the remainder of this section, I provide a very simple model for describing these trade-offs between computational costs and epistemic virtue. I use this model to show that, at least in some contexts, our trade-off considerations motivate comparative judgements about incoherent credences which coincide with those sanctioned by the accuracy criterion.

5.2 Trading-off computational costs for epistemic utility

Let's start by introducing some terminology to describe the sorts of trade-offs we are interested in. I use the term *computational costs* to denote the (dis)utility the agent assigns to performing some computational task, such as deciding whether to accept a bet by using some decision rule and computing expected utility, or reasoning about her own credences.

I want to focus on a specific computational task, which I call *achieving coherence*. Achieving coherence means: (1) working out all relevant logical relationships on the domain; and (2) checking whether one's credences respect the laws of probability on the domain, and if they don't, finding the coherent credence which accuracy-dominates ones' own by the largest amount. The first step amounts to finding the set $\mathcal{W}_{\mathcal{F}}$ of possible valuations over $\mathcal{F}$. For any inaccuracy measure S, the second step is equivalent to finding the d_S-closest coherent credence function y to the agent's credence function c. This can be done by projecting c onto the convex hull of $\mathcal{W}_{\mathcal{F}}$, which is the set of coherent credences on $\mathcal{F}$. If $y = c$ the agent is coherent. Otherwise, by Proposition 3, y is the credence which accuracy-dominates c by the largest amount. Thus achieving coherence can be thought of as the task of producing, from an input pair $(\mathcal{F}, x)$ consisting of some set of sentences $\mathcal{F}$ and a credence function $x : \mathcal{F} \to \mathbb{R}$, the corresponding output pair $(\mathcal{W}_{\mathcal{F}}, y)$ consisting of a set of valuations $\mathcal{W}_{\mathcal{F}}$ and a function $y : \mathcal{F} \to \mathbb{R}$ which is the d_S-closest coherent credence function to x.[11]

Given a non-ideal agent with incoherent credence function c over domain $\mathcal{F}$, let A be the algorithm that the agent would use if they were to achieve coherence, that is, if they were to compute from some input pair $(\mathcal{F}, x)$ the corresponding output pair $(\mathcal{W}_{\mathcal{F}}, y)$ defined as above. It's reasonable to assume that the choice of algorithm will not depend on the starting pair $(\mathcal{F}, x)$ since, although different algorithms may be more efficient on

[11] Strictly speaking, both input and output credence functions are not functions to the whole $\mathbb{R}$, because non-ideal agents can only manipulate numbers up to some degree of precision. So given an input function x specified to some degree of precision, the output y will be an approximation of the d_S-closest coherent credence. But since $S(\cdot, w)$ is continuous for every w, the minimum difference in accuracy between x and y will approximate that between x and its d_S-closest coherent credence.

different inputs, one can't normally tell which algorithm is best for a given input before performing the computation.[12] The computational costs of achieving coherence for this agent should be a measure of the effort (for example, the computation time) that it would take the agent to achieve coherence using algorithm A on input $(\mathcal{F}, c)$.

Now let c, c' credence functions on some domain $\mathcal{F}$, such that c' is more approximately coherent than c under my incoherence measure, i.e. $\underline{I}_S(c') < \underline{I}_S(c)$. We clearly have a reason to say c' is better than c if the former happens to accuracy-dominate the latter, as discussed in Section 3. But even if c' does not accuracy-dominate c, trade-off considerations may reveal a way in which it is epistemically better. To see this, let k be an agent's computational costs for achieving coherence, defined as above, and assume for the moment that these are constant over all input credences. If the agent had credence function c, she could gain $\underline{I}_S(c)$ in guaranteed accuracy by paying k computational costs, whereas if she had credence function c' she could pay the same costs to gain a smaller amount $\underline{I}_S(c')$ in guaranteed accuracy. Fixing any rate of exchange between computational costs and accuracy, we can thus find an interval $(u, l) \in \mathbb{R}$ such that, for all agents with $k \in (u, l)$, it would be desirable for the agent to pay k computational costs to improve her accuracy by $\underline{I}_S(c)$, but it would not be desirable to pay the same cost to improve her accuracy by $\underline{I}_S(c')$.[13] Furthermore, for any rate of exchange, there is no value k such that trading k costs for $\underline{I}_S(c')$ accuracy is desirable, whereas trading k for $\underline{I}_S(c)$ is not.

As an example, let $\mathcal{F} = \{\phi, \psi\}$ where $\phi \equiv \neg\psi$. Let c, c' incoherent credence functions, with $c(\phi) = c(\psi) = 3/4$, $c'(\phi) = 2/10$, and $c'(\psi) = 9/10$. Let S be the Brier inaccuracy measure, i.e. $S(x, w) := \sum_{\theta \in \mathcal{F}}(x(\theta) - w(\theta))^2$. Note that c' does not accuracy-dominate c. However c' is less incoherent than c, since $\underline{I}_S(c) = 1/8$ and $\underline{I}_S(c') = 1/200$ (Figure 2). This means an agent with credence function c could gain $1/8$ units of guaranteed accuracy by achieving coherence, whereas if they had credence function c' they could gain at most $1/200$ units of guaranteed accuracy. Let the agent's rate of exchange between computational costs and units of accuracy be 1. That is, the agent is willing to pay 1 unit of computational costs (e.g. 1 hour of computation time) to increase their accuracy by 1 unit. Now assume the computational costs of achieving coherence are constant over all input credences, and denote these costs by k. All non-ideal agents with rate of exchange 1 and $1/200 < k < 1/8$ will find it desirable to pay k in order to gain $1/8$ guaranteed accuracy, and will not find it desirable to pay k in order to gain $1/200$ guaranteed accuracy.

This shows a sense in which less incoherent credences are epistemically better than

[12] Sometimes we do have good evidence that an algorithm works better than another for smaller/larger inputs. So the agent may use different algorithms depending on the size of $\mathcal{F}$. This however will not impact the following discussion, as we are keeping $\mathcal{F}$ fixed.

[13] To see this, let r be some rate of exchange between computational costs and accuracy, so that the agent would be willing to pay k computational costs to gain t accuracy iff $k < rt$. Then just pick $l = \underline{I}_S(c')r$ and $u = \underline{I}_S(c)r$.

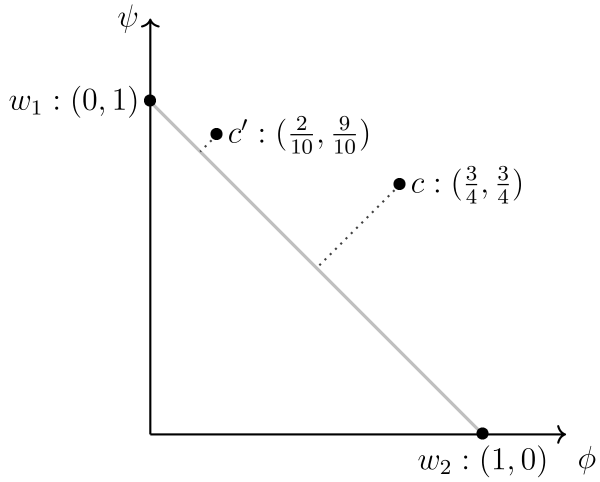

Figure 2: The light gray region is that of coherent credence functions, for which $c(\phi) = 1 - c(\psi)$. Let S be the Brier inaccuracy measure, i.e. $S(x, w) := \sum_{\theta \in \mathcal{F}} (x(\theta) - w(\theta))^2$. Then the corresponding divergence $d_S(x, y)$ is the squared Euclidean distance between x and y. Staffel's measure $I_S(x)$ is the divergence $d_S(x, y)$ between x and the d_S-closest coherent credence y. This is just the square of the Euclidean distance between x and the light gray region. By Proposition 3, $\underline{I}_S(x)$ is also the square of this distance.

more incoherent ones. If $\underline{I}_S(c) > \underline{I}_S(c')$, then no matter what exchange rate one picks, there is a range of values of k such that any reasonable non-ideal agent with such computational costs ought not to have credence function c. It would be desirable for such an agent (whether she knows it or not) to invest k effort into achieving coherence. Yet if such agent had credence function c', then the cost k would not be worth the smaller amount of guaranteed accuracy she could gain by paying it. I will say that, for this class of agents, c is *unstable*, whereas c' is *stable*. Note that, if $\underline{I}_S(c) > \underline{I}_S(c')$ then for no such class of agents is c' unstable and c stable. This gives us reason to think c' is epistemically better than c, even if the former does not accuracy-dominate the latter.

I should clarify here that I am *not* claiming that an incoherent credence function is rational when held by an agent whose computational costs and exchange rate are such achieving coherence is not worth the effort. The evaluation of a credence function is agent-independent: incoherent credences are always irrational, regardless of who holds them, because an (abstract) ideal agent (i.e. a perfect calculator) could not rationally hold them. That is, incoherent credences are inadmissible for ideal agents. My aim is to use classes of (abstract) non-ideal agents (i.e. imperfect calculators) to further distinguish amongst incoherent credences, based on whether they are stable for such agents.

I am also *not* claiming that an agent with incoherent credence c, who happens to have computational costs which make c unstable, should in fact pay the costs of achieving coherence and adopt the credence which accuracy-dominates c by the largest amount. Trade-off considerations are used to compare incoherent credences, not to determine how agents should improve. This is similar to the accuracy-argument for bivalent coherence. The fact that every incoherent credence is accuracy-dominated by a coherent one does not imply that an agent with incoherent credence function c should shift to a dominating credence function c'. It just shows a fault which every incoherent credence function shares, and every coherent credence function is immune from, thus giving us a reason to claim coherent credences are epistemically better than incoherent ones. Similarly, if $\underline{I}_S(c') < \underline{I}_S(c)$, then there is a class of non-ideal agents for which c' is stable whereas c is not, while for no such class is c stable and c' unstable. Being stable for such agents is an additional fault on top of incoherence; a fault from which all credence functions with sufficiently low incoherence measure will be immune. This gives us an additional reason to claim that less incoherent credences are epistemically better than more incoherent ones.

5.3 Robust and non-robust scenarios

The trade-off argument for the evaluative meaning of incoherence comparisosn rests on the assumption that an agent's computational costs for achieving coherence do not depend on her starting credence function. In particular, I have shown that, if $\underline{I}_S(c') < \underline{I}_S(c)$, then for a whole family of abstract agents (identified by their computational costs), c' is admissible, whereas c is inadmissible. Yet this only works because the costs needed for an agent to become coherent are roughly constant across incoherent credence functions (by assumption), whereas the amount of guaranteed accuracy to be gained decreases with decreasing incoherence measure (by definition of $\underline{I}_S$), making the trade-off desirable at c but not-desirable at c'. How can we justify this assumption?

The constant costs assumption is quite natural in scenarios where working out all logical relations on the domain accounts for most of the costs of achieving coherence. For instance, in our example involving an extremely complex tautology ϕ, once the agent finds out that this is a tautology, coherence requires them to assign credence 1 to ϕ. In these scenarios, any reasonable algorithm for achieving coherence will have equal costs, independently of the agent's starting credence.[14] Similar considerations hold approximately in all cases where the domain is comparatively small, and the logical relations are comparatively hard to work out. In these scenarios, regardless of which algorithm an agent uses to achieve coherence, it will always be the case that, if $\underline{I}_S(c') < \underline{I}_S(c)$ by a significant amount, then paying the cost of achieving coherence in exchange for the

[14] A pathological algorithm could always be devised that takes more or less time to output the value $y(\phi) = 1$ depending on the input value $c(\phi)$. But knowing a credence function is stable/unstable for someone using such an algorithm tells us very little about the degree of epistemic faultiness of that credence.

resulting accuracy is more favourable at c than at c', the costs being approximately the same and the gained accuracy being significantly greater at c. I call these *robust scenarios*, since trade-off considerations support the same comparative judgements whenever the agents use a reasonable algorithm for achieving coherence. In robust scenarios, trade-off considerations give evaluative meaning to comparative judgements about credences, independent of the agents holding those credences.

In other scenarios, the constant costs assumption is not reasonable. Different reasonable algorithms for achieving coherence may vary significantly in their execution times depending on the starting credence. Then if $\underline{I}_S(c') < \underline{I}_S(c)$, there might be some reasonable algorithm A such that running A on $(\mathcal{F}, c')$ takes much less time than running A on $(\mathcal{F}, c)$, and some different reasonable algorithm B for which the opposite is true. Paying the costs of achieving coherence would then be more favourable at c' than at c for an agent using algorithm A, while the opposite would be true for an agent using B. These scenarios are *non-robust*: looking at trade-offs between computational costs and guaranteed accuracy will motivate different comparative judgements depending on which algorithm is used by the agents involved. These comparative judgements are not just about credences, but depend on the particular choice of algorithm made by the agents holding those credences.

The upshot is that in robust scenarios, where trade-off considerations motivate comparative judgements about credence functions, these are the same judgements that are motivated by the accuracy criterion. That is, if $\underline{I}_S(c') < \underline{I}_S(c)$, trade-off considerations give us an additional reason to think that c' is epistemically better than c. This reinforces our justification of coherence, by giving us an alternative way of establishing the evaluative meaning of incoherence comparisons in these scenarios. In non-robust scenarios, trade-off considerations do not motivate comparative judgements about credence functions themselves, but only about credence function-agent (or credence function-algorithm) pairs. Such judgements neither support nor clash with those induced by accuracy criterion, since it is perfectly possible that a credence function c be epistemically better than c' even though the credence function-agent pair (c, a) is epistemically worse than the pair (c', a), due to the particular choice of algorithm made by agent a to achieve coherence.

6 Conclusion

The coherence norm states that an agent's credences ought to respect the laws of probability. Numerous examples show that human-like agents cannot ensure they respect this norm. Even when it's possible for them to respect the coherence norm, it often seems more reasonable to remain incoherent instead. These considerations lead us to question whether coherence can be a rationality norm for human-like agents.

One response to this objection is that coherence is an ideal to be approximated.

But this response only works if we can (1) specify a notion of approximate coherence, and then (2) show that this notion is normatively relevant. In this essay, I defined an accuracy-based notion of approximate coherence, measuring the degree of incoherence of a credence function as the largest amount by which it is accuracy-dominated. I showed that these incoherence measures are formally equivalent to those defended by De Bona and Staffel (2017). Taking advantage of this accuracy-based definition, I argued that the incoherence comparisons induced by these measures are evaluatively meaningful, under the assumption that it's epistemicaly better to be accuracy-dominated by a smaller rather than larger amount. This in turn allowed me to show that the notion of approximate coherence captured by these measures is normatively relevant.

My assumption that it's epistemicaly better to be accuracy-dominated by a smaller rather than larger amounts is far from self-evident. So to to further support my claim that incoherence comparisons are evaluatively meaningful, I tried to give an alternative justification, based on simple description of how non-ideal agents trade-off epistemic virtue for computational costs. I argued that in some scenarios, which I call "robust", trade-off considerations motivate genuine comparative judgements about credence functions, which coincide with those induced by my incoherence measures.

While trade-off considerations support the evaluative meaning of incoherence comparisons in robust scenarios, they are unhelpful in non-robust ones. Thus there are still many cases where my justification for the normative role of approximate coherence rests on a somewhat dubious assumption. Further work is needed to find more general ways of justifying the evaluative force of incoherence comparisons.

Appendices

A Proof of the results

Recall that the accuracy-based incoherence measure I put forward in Section 4 was defined as:

$$\underline{I}_S(c) = \sup\{\min_{w\in\mathcal{W}_\mathcal{F}}\{S(c,w) - S(f,w)\} \,:\, f \in \mathcal{P}_\mathcal{F}\,\} \tag{10}$$

where S is a strictly proper additive inaccuracy measure. To prove results about this incoherence measure, it is helpful to show that the supremum in its definition is actually a maximum, i.e. that for every credence function c there is some pair (g, w_k) where g is a coherent credence function, $w_g \in \mathcal{W}_\mathcal{F}$, and $\underline{I}_S(c) = S(c, w_g) - S(g, w_g)$.

Lemma 1. *Let $c : \mathcal{F} \to \mathbb{R}$ a credence function. Let S an additive, strictly proper inaccuracy measure. Then there is some coherent credence $g : \mathcal{F} \to \mathbb{R}$ and some $w_g \in \mathcal{W}_\mathcal{F}$ such that:*

$$\underline{I}_S(c) = S(c, w_g) - S(g, w_g) \tag{11}$$

Proof. Fix a credence function c. For every $w \in \mathcal{W}_{\mathcal{F}}$, we know $S(c,w) - S(f,w)$ a continuous function of f for each $w \in \mathcal{W}_{\mathcal{F}}$. The function:

$$m_c(f) = \min_{w \in \mathcal{W}_{\mathcal{F}}} \{S(c,w) - S(f,w)\} \tag{12}$$

seen as a function $m_c : \mathbb{R}^{\mathcal{F}} \to \mathbb{R}$, is also continuous, as it is the minimum of finitely many continuous functions. The set $\mathcal{P}_{\mathcal{F}}$ of coherent credence functions over $\mathcal{F}$ is a compact subset of $\mathbb{R}^{\mathcal{F}}$, and so its image under m_c is a compact subset of $\mathbb{R}$, that is, a closed and bounded subset. Thus there is some g coherent such that:

$$m_c(g) = \sup\{m_c(f) \ : \ f \in \mathcal{P}_{\mathcal{F}}\} = \underline{I}_S(c) \tag{13}$$

But then $m_c(g) = S(c,w_g) - S(g,w_g)$ for some $w_g \in \mathcal{W}_{\mathcal{F}}$. So the pair (g,w_g) satisfies (11). □

The advantage of this accuracy-based definition of incoherence is that it allows us to say something about what the inequality $\underline{I}_S(c') < \underline{I}_S(c)$ implies in terms of the epistemic utility of c and c'.

Proposition 2. *Let S an additive, strictly proper inaccuracy measure. Let $c : \mathcal{F} \to \mathbb{R}$ a credence such that $\underline{I}_S(c) = t$, for some $t > 0$. Then:*

(i) For every $c' : \mathcal{F} \to \mathbb{R}$, there is a $w' \in \mathcal{W}_{\mathcal{F}}$ such that:

$$S(c,w') - S(c',w') \leq t \tag{14}$$

(ii) There is a $c' : \mathcal{F} \to \mathbb{R}$ such that $S(c,w) - S(c',w) \geq t$ for every w.

Proof. (i): this is immediate from the definition of $\underline{I}_S$. Let $c : \mathcal{F} \to \mathbb{R}$ such that $\underline{I}_S(c) = t$. Assume by way of contradiction that for some $c' : \mathcal{F} \to \mathbb{R}$ we have $S(c,w) - S(c',w) > t$ for every $w \in \mathcal{W}_{\mathcal{F}}$. Let $w' \in \mathcal{W}_{\mathcal{F}}$ a world which minimise this difference (it will always exist, since $\mathcal{W}_{\mathcal{F}}$ finite). Then:

$$\underline{I}(c) \geq S(c,\omega') - S(c',\omega) > t \tag{15}$$

contradiction.

(ii) Let (c',w') be the credence function-world pair of Lemma 1. Clearly the condition holds. □

This shows that if $\underline{I}_S(c') < \underline{I}_S(c) = t$ then some credence function is more accurate than c of at least t in every world, whereas this cannot be the case for c'.

Let's introduce a few more definitions before moving to the proof of Proposition 3. Assume throughout that $\mathcal{F} = \{A_1, ..., A_m\}$ and $c : \mathcal{F} \to \mathbb{R}$ is an incoherent credence, identified with the $\mathbb{R}^m$ vector $\overline{c} = (c(A_1), ..., c(A_m))$. If y is a coherent credence on

$\mathcal{F}$, then it can be written as the convex combination of the truth-value assignments corresponding to the worlds in $\mathcal{W}_{\mathcal{F}}$. I write y_i^* to denote the weights of this convex combination, so that each coherent y can be written as:

$$y = \sum_{i=1}^{n} y_i^* w_i \tag{16}$$

where $\sum_{i=1}^{n} y_i^* = 1$ and $y_i^* \in [0,1]$ for every $i = 1, ..., n$.

Definition A.1 (d_S-projection, projection vertices)**.** Let $c : \mathcal{F} \to \mathbb{R}$ be a credence function.

1. When z is the d_S-closest coherent credence to c, we say that z is the *d_S-projection* of c onto the set of coherent credence functions.
2. Let z be the d_S-projection of c onto the set of coherent credence functions. The *projection vertices* of c are the worlds $w_i \in \mathcal{W}_{\mathcal{F}}$ such that $z_i^* \neq 0$. Denote by $\mathcal{K}(c)$ the set of projection vertices of c.

There is an important fact about d_S-projections which will be central to the following proofs.

- **Generalised Pythagorean Theorem** (Predd et al. 2009, Prop. 3):
For every credence function c there is a unique d_S-projection z onto the set of coherent credence functions. Moreover, for any coherent credence function y:

$$d_S(z, y) \leq d_S(c, y) - d_S(c, z) \tag{I}$$

The proof of Proposition 3 will use two lemmas:

Lemma 2. *Let S be an additive, strictly proper inaccuracy measure, let $c : \mathcal{F} \to \mathbb{R}$ an incoherent credence function, and let z be the d_S-projection of c onto the set of coherent credence functions. The difference in accuracy $S(c, w_i) - S(z, w_i)$ is constant for all projection vertices $w_i \in \mathcal{K}(c)$.*

Proof. [15] Let $\mathcal{F} = \{A_1, ..., A_n\}$ and $\mathcal{W}_{\mathcal{F}} = \{w_1, ..., w_n\}$. Let z be the d_S-closest credence to c. For any $w_i \in \mathcal{W}_{\mathcal{F}}$, letting $y = w_i$ in the Pythagorean inequality (I) gives:

$$d_S(z, w_i) \leq d_S(c, w_i) - d_S(c, z) \tag{17}$$
$$d_S(c, w_i) - d_S(z, w_i) \geq d_S(c, z) \tag{18}$$

[15] This version of the proof of Lemma 2 is much shorter than my original one, and was suggested by Glauber de Bona.

which, from the definition of d_S, is equivalent to:

$$S(c, w_i) - S(z, w_i) \geq d_S(c, z) \tag{19}$$

Because z is coherent, we can write it as:

$$z = \sum_{i=1}^{n} z_i^* w_i \tag{20}$$

where $\sum_{i=1}^{n} z_i^* = 1$ and $z_i^* \in [0, 1]$ for every $i = 1, ..., n$. Using the definition of $d_S(c, z)$ we have:

$$d_S(c, z) = Exp_z [S(c, \cdot) - S(z, \cdot)] \tag{21}$$
$$= \sum_{i=1}^{n} z_i^* (S(c, w_i) - S(z, w_i)) \tag{22}$$
$$= \sum_{w_i \in \mathcal{K}(c)} z_i^* (S(c, w_i) - S(z, w_i)) \tag{23}$$

But from equation (19) we know that, for any $w_i \in \mathcal{K}(c)$, the difference $S(c, w_i) - S(z, w_i)$ is at least as great as $d_S(c, z)$. So if for some $w_j, w_i \in \mathcal{K}(c)$ we had that $S(c, w_j) - S(z, w_j) > S(c, w_i) - S(z, w_i)$, then the weighted sum in (23) would be strictly greater than $d(c, z)$, violating the equality. Thus the difference $S(c, w_i) - S(z, w_i)$ must be constant for every $w_i \in \mathcal{K}(c)$. □

Lemma 3. *Let S be an additive, strictly proper inaccuracy measure, let $c : \mathcal{F} \to \mathbb{R}$ an incoherent credence, and let z be the d_S-projection of c onto the set of coherent credence functions. If $w_i \in \mathcal{K}(c)$ and $w_k \in \mathcal{W}_\mathcal{F} \setminus \mathcal{K}(c)$, then $S(c, w_k) - S(z, w_k) \geq S(c, w_i) - S(z, w_i)$.*

Proof. Fix some $w_i \in \mathcal{K}(c)$ and $w_k \in \mathcal{W}_\mathcal{F} \setminus \mathcal{K}(c)$. From the generalised Pythagorean inequality (I), letting $y = w_k$ we have:

$$d_S(z, w_k) \leq d_S(c, w_k) - d_S(c, z) \tag{24}$$
$$d_S(c, w_k) - d_S(z, w_k) \geq d_S(c, z) \tag{25}$$

Then, using the definition of $d_S(c, z)$:

$$d_S(c, w_k) - d_S(z, w_k) \geq Exp_z [S(c, \cdot) - S(z, \cdot)] \tag{26}$$
$$= \sum_{w_j \in \mathcal{K}(c)} z_j^* (S(c, w_j) - S(z, w_j)) \tag{27}$$
$$= S(c, w_i) - S(z, w_i) \tag{28}$$

where the last equality holds because, by Lemma 2, the difference $(S(c, w_j) - S(z, w_j))$ is constant for all $w_j \in \mathcal{K}(c)$. □

Proposition 3. *Let S be an additive, strictly proper inaccuracy measure. Then for every credence function $c : \mathcal{F} \to \mathbb{R}$, we have $\underline{I}_S(c) = I_S(c)$.*

Proof. Let $\mathcal{F} = \{A_1, ..., A_m\}$ and $\mathcal{W}_{\mathcal{F}} = \{w_1, ..., w_n\}$. Denote by d_S the Bregman divergence associated with the inaccuracy measure S. Let $c : \mathcal{F} \to \mathbb{R}$ an incoherent credence function, and let z be the d_S-closest coherent credence function to c. Recall that De Bona and Staffel (2017) define their incoherence measure I_S as:

$$I_S(c) = \min\{d_S(c, x) : x \text{ is a coherent credence function}\} \tag{29}$$

so having defined z as the d_S-closest coherent credence to c, we will have $I_S(c) = d_S(c, z)$.

Now let y be a coherent credence such that $\underline{I}_S(c) = S(c, w_y) - S(y, w_y)$ for some $w_y \in \mathcal{W}_{\mathcal{F}}$ (we know it exists by Lemma 1). I start by showing that y and z are the same credence function, and then use this fact to show $\underline{I}_S(c) = I_S(c)$.

From the definition of $\underline{I}_S(c)$ we know that, for every $w \in \mathcal{W}_{\mathcal{F}}$:

$$S(c, w_y) - S(y, w_y) \leq S(c, w) - S(y, w) \tag{30}$$

and also, because of the sup in the definition of $\underline{I}_S$,

$$S(c, w_y) - S(y, w_y) \geq \min_{w \in \mathcal{W}_{\mathcal{F}}} \{S(c, w) - S(z, w)\} \tag{31}$$

We know by Lemma 3 that, if $w_i \in \mathcal{K}(c)$ and $w_k \in \mathcal{W}_{\mathcal{F}} \setminus \mathcal{K}(c)$, then $S(c, w_k) - S(z, w_k) \geq S(c, w_i) - S(z, w_i)$. Hence the minimum in (31) is achieved at some $w_i \in \mathcal{K}(c)$. Combining this with (30) and (31), we obtain that whenever w_i is a projection vertex the following inquality holds:

$$S(c, w_i) - S(y, w_i) \geq S(c, w_y) - S(y, w_y) \geq S(c, w_i) - S(z, w_i) \tag{32}$$

since $S(c, w_i) - S(z, w_i)$ is constant for all $w_i \in \mathcal{K}(c)$, by Lemma 2. Equation (32) simplifies to $S(y, \omega_i) \leq S(z, \omega_i)$. So for every $w_i \in \mathcal{K}(c)$, $S(y, \omega_i)$ is no greater than $S(z, \omega_i)$. But then:

$$Exp_z(S(y, \cdot) - S(z, \cdot)) = \sum_{w_i \in \mathcal{K}(c)} z_i^*(S(y, w_i) - S(z, w_i)) \leq 0 \tag{33}$$

since each addend in the sum is no greater than zero. And because S is strictly proper, this can only be the case (with equality) if $y = z$.

From the fact that $z = y$ it follows that $\underline{I}_S(c) = I_S(c)$, because:

$$\begin{aligned}
\underline{I}_S(c) &= S(c, w_y) - S(y, w_y) && (34)\\
&= S(c, w_y) - S(z, w_y) \quad \text{(because } z = y\text{)} && (35)\\
&= S(c, w_i) - S(z, w_i) \ \text{ for some } w_i \in \mathcal{K}(c) \text{ (by Lemma 3)} && (36)\\
&= \sum_{w_i \in \mathcal{K}(c)} z_i^*(S(c, w_i) - S(z, w_i)) \quad \text{(by Lemma 2)} && (37)\\
&= Exp_z(S(c, \cdot) - S(z, \cdot)) && (38)\\
&= d_S(c, z) && (39)\\
&= I_S(c) \quad \text{(by construction of } z\text{).} && (40)
\end{aligned}$$

□

References

Babic, B. (2019). A theory of epistemic risk. *Philosophy of Science*, 86(3):522–550.

Carr, J. R. (2021). Why ideal epistemology? *Mind*.

Chalmers, D. J. (2011). Frege's puzzle and the objects of credence. *Mind*, 120(479):587–635.

De Bona, G. and Staffel, J. (2017). Graded incoherence for accuracy-firsters. *Philosophy of Science*, 84(2):189–213.

De Bona, G. and Staffel, J. (2018). Why be (approximately) coherent? *Analysis*, 78(3):405–415.

de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In Kyburg, H. and Smokler, H., editors, *Studies in Subjective Probability*, pages 53–118. Wiley New York. Translated by H. Kyburg (original published in 1937).

Fitts, J. (2014). Chalmers on the objects of credence. *Philosophical Studies*, 170(2):343–358.

Foley, R. (1993). *Working without a net: A study of egocentric epistemology.* Oxford University Press on Demand.

Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of science*, 65(4):575–603.

Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical perspectives*, 24:281–323.

Levinstein, B. A. (2017). A pragmatist's guide to epistemic utility. *Philosophy of Science*, 84(4):613–638.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Pettigrew, R. (2020). *Dutch book arguments*. Cambridge University Press.

Predd, J. B., Seiringer, R., Lieb, E. H., Osherson, D. N., Poor, H. V., and Kulkarni, S. R. (2009). Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792.

Savage, L. J. (1967). Difficulties in the theory of personal probability. *Philosophy of Science*, 34(4):305–310.

Schervish, M. J. (1989). A general method for comparing probability assessors. *The annals of statistics*, 17(4):1856–1879.

Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (1998). Two measures of incoherence: How not to gamble if you must.

Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (2002). Measuring incoherence. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 561–587.

Staffel, J. (2020). *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press, USA.

Zynda, L. (1996). Coherence as an ideal of rationality. *Synthese*, 109(2):175–216.