Approximate coherence: a new accuracy-centered account

Giacomo Molinari

September 17, 2022

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 argue that, even if you can't be coherent, you should try to 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 has normative bite, i.e. that non-ideal agents ought to approximate coherence.

Here is the essay plan. I start in Section 2 by briefly summarising two popular arguments for the coherence norm: one pragmatic (Dutch Book argument) and one purely epistemic (accuracy argument). This essay will mostly focus on the purely epistemic perspective, 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 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 accuracy-centered defense of the coherence norm, which follows a similar 2-step strategy. I criticise this proposal, arguing that it 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 definition of approximate coherence in terms of accuracy, and show that this definition extends De Bona and Staffel's. Taking advantage of this new accuracy definition, in Section 5 I show why non-ideal agents should approximate coherence. Here I argue that the evaluative meaning of comparative incoherence judgements can depend on contextual features, including features of the non-ideal agents we are interested in. Hence the normative relevance of approximate coherence is also dependent on these features. A proper defense of the coherence norm for human-like agents requires a better descriptions of these agents than we can provide. Still, we can show coherence has normative force over some (abstract, simplified) non-ideal agents, which at least sometimes resemble us. I end the essay by summing up the main conclusions in Section 6.

2 Ideal norms and non-ideal agents

In this section I introduce the coherence norm, together with two popular arguments used to support it. Then I present the objection that coherence is inadequate as a rationality norm due to its excessively idealised nature. I discuss two versions of this objection, and outline the response strategy I will use to defend the normative role of coherence.

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 $\mathcal{F}^*$. Many epistemologists argue that coherence has the status of a rationality norm, meaning that a rational agent *ought* to have coherent credences. One way to defend this claim is to show that, if your credences are incoherent, then 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, 1937).

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*

[1] There is some debate around what the objects of credence should be (Chalmers, 2011; Fitts, 2014). I pick sentences here (instead of sets of possible worlds, as suggested by Lewis (1986)) 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.

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 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 expression:

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

is uniquely minimised at $x = p$. 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 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 inaccuracy 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.

For the rest of the essay, I will assume inaccuracy is measured by a strictly proper measure.

Note that, although both 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 *cannot* be coherent, it's not clear these arguments have anything to say about you, besides that you cannot be rational. This has led some to doubt that coherence can be an epistemic 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's language includes the necessary arithmetic, then exactly one of the ϕ_i will be a logical truth, and the others will be logical falsehoods.

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

So 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 we cannot compute the digit, it would be odd to have extremal credence in some ϕ_i. Thus we are irrational, we know we are irrational, and we can't do anything about it.

The example naturally extends to other complex logical relations, and to complex bundles of 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 incoherencies 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 force over us only if we can comply with it. 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.

OIC assumptions are somewhat controversial in epistemology.[3] Some defenders of coherence contend that constraints may play a normative role for you even if you cannot strictly comply with them. Zynda (1996) uses the analogy of a machinist who works on a lathe, and receives a specification for a cylindrical component with certain dimensions. Clearly, no matter how skilled the machinist and how technically advanced the lathe, the component will never be perfectly cylindrical. And yet, there is reason to think that the ideal constraint still has normative force over the machinist. This is because: (i) it allows the machinist to determine that certain components (actual or hypothetical) are better than others, since they are more approximately cylindrical; (ii) if the available manufacturing technology were to improve, the machinist would be subject to an obligation to produce more perfectly cylindrical parts; and (iii) the requirement to approximate the ideal is what 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.

To show that the analogy holds, however, we need to say a bit more. First, we

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

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. In the same way, we can explain why the machinist has a conditional obligation to more accurately approximate the cylinder as his abilities improve, and indeed why she has 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.

Our 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: (ci) that it produces evaluatively meaningful comparative judgements, i.e. that approximately coherent credences are in some sense better than wildly incoherent ones; (cii) that, conditional on her abilities improving, the agent is required to better approximate coherence; and (ciii) that this gives the agent a motivation for improving her abilities.

Since you may use pragmatic or accuracy-centered considerations to justify the (ideal) coherence norm, you may want appeal to the same considerations to explain why agents ought to approximate coherence (Step 2). I will discuss both kinds of proposal in the next sections, although my main focus will be on the accuracy-centered one. First however I should mention a related but different objection against the coherence norm, which is less commonly discussed in the literature.[4]

Consider again an agent betting on some digit of π. Above we focused on the case where the agent has no hope of ever being able to calculate this digit. But imagine the digit is not so remote; let's say it would require the agent one year of (possibly computer-aided) calculations to reach it. It seems odd to say that this agent is irrational for not

[4]Hacking (1967) outlines this version of the objection in his response to Savage's example of an agent betting on the digits of π. In the same paper, Hacking responds to both objections against coherence by relaxing the coherence constraints, so that it's possible for non-ideal agents to respect them. For him, probabilism is a descriptive theory, and hence we should weaken coherence if we want to describe non-ideal reasoners. Since I am interested in defending the normative role of coherence, I will not consider Hacking's response strategy in this essay.

having credence 1 in the correct digit, given that it would take so much effort to find out which one it is. Once again the coherence norm goes against our ordinary judgements of rationality. I call this the **Coherence Impractical** objection.

Unlike Coherence Impossible, Coherence Impractical does not rest on any epistemic OIC assumption. The agent *can* compute the required digit of π, she just chooses not to. At first pass, this makes Coherence Impractical even more worrying: there is a real sense in which the agent is choosing to be incoherent here! But one could respond by saying that, in cases such as these, the agents are simply trading off some of their epistemic virtue for other, more practical benefits. 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. Arguably, dealing with these trade-offs is not the job of epistemology. Epistemology is concerned with rationality, but there is more than rationality to human life.

While this response gets something right, it's not entirely satisfying. It's true that we shouldn't expect epistemology to tell us *how* to trade off epistemic virtue for pragmatic benefits. But we would at least like to know *what* we are trading away when we deviate from coherence. Recall the earlier example of my credences in statements about the election and various economic indicators. I could spend a whole week thinking about the probabilistic relationships between my credences in various statements, looking for coherence violations and resolving them as best I can. Even if a week is not enough to reach coherence, it is plausible that this effort would improve my epistemic situation. This is something epistemology should be able to establish. If we are constantly trading off epistemic virtue for pragmatic benefits, epistemology should be able to tell us how much epistemic virtue we are giving up at a given time. Otherwise, it's not clear it has anything useful to say about our intellectual lives.

Although it is a separate objection, the above discussion gives us reason to think that Coherence Impractical can be addressed in a similar way as we plan to address Coherence Impossible: by specifying a notion approximate coherence, and giving some sense of the epistemic benefits involved in approximating the coherence ideal.

3 Approximate coherence for accuracy-firsters

This section looks at De Bona and Staffel (2017) proposal for an accuracy-centered defense of the coherence norm for non-ideal agents. This proposal is further expanded by Staffel (2020), so I will refer to both sources throughout this section. 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. I will criticise their proposal, showing that their argument for why agents ought to be approximately coherent only gives normative strength 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 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 set of propositions, we should be able to determine whether one is more incoherent than the other.

We want Incompleteness to hold because we are interested in comparing 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 seem 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:

$$I_d(c) = \min\{d(c, y) : y \text{ coherent}\} \tag{3}$$

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:

$$S(c, w) = \sum_{\phi \in \mathcal{F}} s(c(\phi), w(\phi)) \tag{4}$$

where s is a strictly proper scoring rule. To each inaccuracy measure of this kind we can associate a corresponding divergence:

$$d_S(c, c') = S(c, c') - S(c', c') = \sum_{w \in \mathcal{W}_{\mathcal{F}}} c'(w)S(c, w) - \sum_{w \in \mathcal{W}_{\mathcal{F}}} c'(w)S(c', w) \tag{5}$$

When c' is coherent, $d_S(c, c')$ is just the expected value under c' of the difference between the score of c and the score of c'.

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.

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, which is assumed to be their main goal. 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 $c^* : \mathcal{F} \to \mathbb{R}$ is the d_S-closest coherent credence function. 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),

[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 $S(\lambda c + (1 - \lambda)y, w) \leq \lambda S(c, w) + (1 - \lambda)S(c', w)$.

[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 parts separate.

[7] This fact is aknowledged in De Bona and Staffel (2017).

the argument above only shows that a small subsets of these comparisons are evaluatively meaningful.

But we need our notion of approximate coherence to make evaluatively meaningful comparisons in order to explain why it is normatively relevant (ci). 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 notion of approximate coherence, which can only compare pairs of credences when one accuracy-dominates the other:

- **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).

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, it is so weak that it might fail to satisfy the Judgement Preservation (D1) desideratum itself! Consider for example 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.

[FIGURE GOES HERE]

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

My view is that there are contexts where we can only speak of approximate coherence in such weak terms. But there are also cases where, from an accuracy perspective, a stronger notion of approximate coherence than (WAC) is normatively relevant, 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) incoherence measures formally respect the Comparability (D3) desideratum, allowing us to compare the degree of incoherence of any two credence functions, I have argued that many of these comparisons lack evaluative force. This compromises the normative relevance of her notion of approximate coherence. In this section, I put forward a new definition of approximate coherence which respects (D3), and which makes explicit the accuracy implications of every incoherence comparison. Interestingly, these incoherence measures turn out to be equivalent to (a slight extension) of the ones defined by De Bona and Staffel. In the next section I will use these accuracy implications to show that, in a number of different contexts, every incoherence comparison induced by these incoherence measures is evaluatively meaningful. I will thus show 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 epistemic utility or practical guidance value. Nor does it mean that an agent with credence function c' ought to adopt credence function 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, and guaranteed to be Dutch-bookable, whereas coherent ones are not. 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 et al., 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 k£, 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 $2k$£. 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'.

However, this is not enough to show that incoherence comparisons are evaluatively meaningful. To do so, one needs to give some motivation for why it's better to be vulnerable to a smaller, rather than larger, normalised sure loss. As far as I know, this claim has not been explicitly argued for. In the next section, I will defend 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 normalization. What can we say of two credence functions if one is more incoherent than the other under a given normalization, but the opposite is true for some other normalization?[8] Another objection is that the normalised sure loss of a credence function is not a good measure of that function's pragmatic value (Konek, forthcoming). Finally, at the most fundamental level, 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.

For all these reasons, it makes sense to ask whether a similar strategy is available from the accuracy perspective. Here 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* k iff c' is guaranteed to be more accurate than c of at least k, come what may. We want to say that a credence function is more faulty the larger the amount by which some other credence function

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

dominates it. This is captured by the following measure of incoherence:[9]

$$\underline{I}_S(c) = \sup\{\min_{w\in W}\{S(c,w) - S(f,w)\} \ : \ f \text{ is coherent}\} \tag{6}$$

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 $\underline{I}_S(c) = k$. Then:*

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

$$S(c,w) - S(c',w) \leq k \tag{7}$$

2. *There is a $c' : \mathcal{F} \to \mathbb{R}$ such that $S(c,w) - S(c',w) \geq k$ 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.[10]

It's easy to see that $\underline{I}_S$ satisfies the most clear-cut cases of Judgement Preservation (D1). For example, if c is coherent and c' is incoherent, we always have $\underline{I}_S(c) < \underline{I}_S(c')$. This is also the case when c accuracy-dominates c' under S, as we would expect. Incompleteness (D2) is also satisfied, since we can score the accuracy of credence functions even when their domains do not form a complete algebra. Comparability (D3) is not a problem: we can compare any two credence functions defined on the same domain by just comparing their incoherence measure.

Non-Inundation (D4) is somewhat more delicate. Unlike Staffel (2020), I think there might be reasons to reject (D4). Say my credence function is defined on statements about election results and economic indicators, like in the earlier example, but also on statements about different possible causes of my death. My credence function is incoherent overall, although it would be coherent if restricted to the statements about my death. Now imagine a political bookmaker's credence function is the same as mine on this domain. According to the previous definition, our incoherence measure $\underline{I}_S$ will be the same. But should it be? We can expect the bookmaker to care about the truth of political and economic statements far more than I do, and to care about my cause of

[9]This extends a similar definition given by Shervish, Seidenfeld and Kadane (1998, pp.20-21) for the Brier inaccuracy measure.

[10]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)).

death far less than I do. So there is a sense in which he is doing worse than me; I am epistemically virtuous where it matters, whereas the bookmaker is not.

We can capture this intuition is by *weighting* different propositions differently when we measure a credence function's inaccuracy. That is, in our definition of the inaccuracy measure S we assign to each statement θ a weight $m(\theta)$, such that $m(\theta) > 0$ for every θ and $\sum_{\theta \in \mathcal{F}} m(\theta) = 1$:

$$S_m(c, w) = \sum_{\theta \in \mathcal{F}} m(\theta) s(c(\theta), w(\theta)) \tag{8}$$

I give far greater weight to the statements about my death than I do to those about politics or economics, whereas the converse will be true of the bookmaker. Each statement's contribution to my overall inaccuracy will be greater or smaller based on its weight. Thus $\underline{I}_{S_m}$ may violate Non-Inundation: statements regarding my health have far greater impact on my final incoherence measure than those about the elections and the economy.

We can give an accuracy argument for the coherence norm even if we weight different propositions differently. This is because strict propriety is not affected by weighting. That is, if s is a strictly proper score, then S_m defined as in (8) is still a strictly proper measure of inaccuracy. Also, all the features of $\underline{I}_S$ discussed up to this point, and in particular those of Proposition 2, are not affected by adopting a weighted inaccuracy measure. So although the notion of approximate coherence captured by $\underline{I}_S$ can respect the Non-Inundation (D4) desideratum, it can also accommodate the view that different statements have different impact on how well the agent is doing at approximating coherence.

We can fairly easily extend De Bona and Staffel's measures to allow for the same flexibility with regards to (D4). If S_m is a weghted strictly proper inaccuracy measure, then the function defined as:

$$d_{S_m}(x, y) = S_m(x, y) - S_m(y, y) \tag{9}$$

is still a Bregman divergence. We can use this divergence to define:

$$I_{S_m}(c) = \min\{d_{S_m}(c, y) : y \text{ coherent}\} \tag{10}$$

thus extending (3) to cases where the accuracy measure is weighted. The fact that d_{S_m} is a Bregman divergence ensures that all the nice properties of the incoherence measure are preserved, including Proposition 1, while allowing for violations of No-Inundation (D4).

The last formal result of this section shows that my incoherence measures are formally equivalent to those defined by De Bona and Staffel:

Proposition 3. *Let S be an additive (possibly weighted) 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:*

$$\underline{I}_S(c) = I_S(c). \tag{11}$$

where $I_S(c)$ is the incoherence measure defined from divergence d_S as in 10.

Proof. See Appendix. □

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 mine 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 identical.

This is a first step towards answering my criticism in Section 3 that many comparative incoherence judgements made by De Bona and Staffel's measures have no 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 in the Dutch-book case, this does not yet show that c is epistemically better than c'. To reach this conclusion we still need to show that it is epistemically good to be accuracy-dominated by a smaller rather than larger amount. I tackle this problem in the next section.

5 Approximate coherence in context

Both the Dutch Book and the accuracy-based incoherence measures quantify the extent to which a credence function is faulty, whether pragmatically or in terms of accuracy. But it's not obvious that it's better to be subject to these faults to a lesser, rather than greater, degree. In this section, I will argue that this is the case in a number of different scenarios. A stronger notion of approximate coherence than (WAC) is normatively relevant in these scenarios, in the sense that: (ci) less incoherent credences are epistemically better than more incoherent ones (regardless of whether the former dominate the latter); (cii) conditional on her abilities improving, agents ought to better approximate coherence; and (ciii) this motivates agents to improve their abilities. However, the specifics of the various scenarios and the assumptions about the agent look a little artificial. I will end this section by discussing what this means for our project of defending the normative role of coherence for non-ideal evidence.

As in the rest of the essay, I will focus on the accuracy perspective. But given that my notion of incoherence is analogous to the one defined by the Dutch Book incoherence measures, it should be clear enough how these ideas can be adapted to the pragmatic case.

5.1 Trading epistemic virtue for pragmatic costs

Let me start from the bird's eye view. Non-ideal agents are engaged in trade-offs between epistemic virtue and pragmatic costs. I use the term *pragmatic costs* instead of pragmatic utility, to avoid confusion with the kind of utility that figures in the bets of the Dutch Book argument. The two exist on different levels: pragmatic utility comes from the (betting) decisions which you are disposed to make based on your credences in the statements in $\mathcal{F}$. Pragmatic costs, on the other hand, are the cost of meta-decisions which involve your own credences: the cost of checking whether you are incoherent, or of solving a coherence violation.

Focusing on the accuracy perspective alleviates the risk of confusion, but it's still important that pragmatic costs operate at the meta-level. I want to avoid comparing these two kinds of utility within the theory. As mentioned in Section 2, I do not think it is epistemology's job to determine how to trade epistemic virtue for pragmatic costs. In fact, measuring pragmatic cost already falls outside of epistemology's scope. The costs of an agent's meta-decisions, and how she trades them against epistemic virtue, are inputs which our epistemic theory will use to evaluate that agent, not something determined by the theory itself.

I want to argue that the strength of our notion of approximate coherence might vary depending on the context. For this reason, I need to make a further distinction among scenarios with incoherent agents, orthogonal to the one between Coherence Impossible and Coherence Impractical. This new distinction is based on whether the agent has any way to trade increasing amounts of pragmatic costs for increasing amounts of accuracy.

The first kind of scenario is best exemplified by mathematical statements: say ϕ is some very complex statement that has not yet been proven, and assume for simplicity the domain of the agent's credence function is restricted to ϕ. Since ϕ is either logically true or logically false, coherence requires the agent to have credence 0 or 1, respectively, in ϕ. Depending on how hard it is to prove the statement (and on whether its truth-value is decidable), this failure is either due to Coherence Impossible or Coherence Impractical. Furthermore, unless the agent is willing to pay sufficient pragmatic costs to prove (or find a counterexample to) ϕ, it seems that she has no way of paying pragmatic costs to gain accuracy.[11] This is because, in order to ensure a gain in accuracy, she needs to find a dominating credence function, but she can only do so once she knows whether ϕ is true, in which case she could be maximally accurate and coherent. I call these *All-or-nothing scenarios.*

The second kind of scenario is best exemplified by agents with credences over a large domain of interrelated statements. Here the agent fails to be coherent because she cannot keep track of all the logical and probabilistic constraints imposed by the laws of probability. But unlike in the previous case, it is at least plausible that with time and effort the agent should be able to eliminate some inconsistencies in her credences, and improve her accuracy in every possible world, without necessarily reaching coherence. I

[11] I don't mean to say here that you can have no evidence for any mathematical fact unless you have a full proof. It's enough for the discussion that there are some statements about which we are genuinely ignorant until we prove or disprove them.

call these *Gradational scenarios*.

The distinction is blurry, but at least sometimes we have a good idea of what scenario the agent is in. Together with the Coherence Impossible and Coherence Impractical distinction (which is also blurry), this leads to the four kinds of scenarios in Table 1, where I use capital letters as labels.

Table 1: Classification of different incoherence scenarios.

	Coherence Impossible	Coherence Impractical
All-or-Nothing	A	B
Gradational	C	D

Let's look at each of these cases in turn, and see how normatively relevant the notion of incoherence introduced in the previous section (and therefore, the one introduced by Staffel) is in each case.

5.2 Case (A): Coherence Impossible, All-or-nothing

In case (A) coherence is practically impossible and the agent cannot guarantee increasing improvements in accuracy by paying increasing pragmatic costs. In these scenarios, it seems that a strong notion of approximate coherence is generally not normatively relevant.

The examples that best fit this category involve logical relations which are either in principle undecidable, or practically impossible to decide for the agent in question. For instance, consider two statements ϕ, ψ which are logically equivalent, but so complex that it is practically impossible for the agent to realize this, so the agent's credence function c assigns different values to ϕ and ψ. A credence function c' is more approximately coherent than c, under my notion of approximate coherence, if and only if the maximum amount by which c' is accuracy-dominated is smaller than the maximum amount by which c is accuracy-dominated.

It might be obvious to you that leaving more accuracy on the table is epistemically worse than leaving less; or you might have some reason to think that, in this contexts, this fact implies that c' is epistemically better than c. But this is not obvious to me, and I cannot think of such a reason. Hence I think that, in these scenarios, we should fall back on a weake notion of approximate coherence, of the sort motivated by De Bona and Staffel's Proposition 1. If a credence function is guaranteed to be more accurate than another, it is less incoherent, and epistemically better. Furthermore, if the score is convex, there is a path from any incoherent credence function to any dominating one, such that moving along the path decreases incoherence while simultaneously increasing accuracy. But if c' does not accuracy-dominate c, the comparison $\underline{I}_S(c') < \underline{I}_S(c)$ lacks evaluative meaning.

5.3 Case (B): Coherence Impractical, All-or-nothing

In case (B), coherence is possible but impractical to achieve, and the agent cannot guarantee improvements in accuracy by paying increasing pragmatic costs. We can again use as an example two statements ϕ, ψ which are logically equivalent. The agent is capable of figuring out the equivalence, but it would cost her considerable effort and time, and anything short of this realisation would not help decrease her incoherence.

Let c be an incoherent credence function, and c' an alternative credence function which is more approximately coherent under my incoherence measure. We clearly have a reason to say c' is better than c if the former happens to dominate the latter in accuracy, as in the previous case. But even if c' does not dominate c in accuracy, we might still have reason to say it is epistemically better. Assume it costs an agent some amount x of pragmatic cost to determine that ψ and ϕ are logically equivalent, and some further amount y to find the maximal dominating credence function (i.e. the one that dominates hers by the largest amount). Let's also assume these costs don't depend on the agent's credence function. This is a strong assumption, but does not seem entirely unreasonable in these kinds of scenarios. The agent can find out that ψ and ϕ are logically equivalent by reflecting on the meaning of these statements, or by coming up with a proof, and the costs associated with doing so seem largely independent of her starting credences.

If the agent had credence function c, she could gain $\underline{I}_S(c)$ in guaranteed accuracy by paying $x+y$, whereas if she had credence function c', she could pay the same cost to gain up to $\underline{I}_S(c')$ in guaranteed accuracy. Depending on the rate at which the agent exchanges pragmatic costs against epistemic virtue, she could find it undesirable to pay $x + y$ to improve her accuracy of $\underline{I}_S(c')$, and desirable to pay $x + y$ to improve her accuracy of $\underline{I}_S(c)$, since $\underline{I}_S(c) > \underline{I}_S(c')$. Similarly, holding the exchange rate fixed, we can find some value of $x + y$ which makes the trade-off desirable at c and undesirable at c'.

As I said earlier, I don't think it is the epistemologist's job to fix one exchange rate as "the right one". But the above considerations show a sense in which less incoherent credences are epistemically better. No matter what exchange rate one picks, we can find a range of values of x and y such that any agent with such exchange rate and such pragmatic costs ought not to have credence function c: it is better for such an agent, according to their own standards, to invest $x + y$ effort into improving their credences. Yet the same agent may have credence function c', because the cost $x + y$ is not worth the smaller amount of guaranteed accuracy she can gain from there. For this class of agents, credence function c is inadmissible, whereas c' is not. On the other hand, for no such class of agents is c' inadmissible and c admissible. 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 a few things before moving to the other cases. First, I am *not* claiming that an incoherent credence function is rational when held by an agent whose pragmatic costs and exchange rate are such that no coherence credence function is worth the cost of finding it. The rationality or irrationality 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 have those credences. However, we can use classes of (abstract) non-ideal agents (i.e. imperfect calculators) to further distinguish between incoherent credence functions in terms of their relative irrationality.

If $\underline{I}_S(c') < \underline{I}_S(c)$, then for any exchange rate between pragmatic costs and accuracy, there is a class of non-ideal agents (identified by their pragmatic costs, i.e. their computational abilities) for which c' admissible whereas c is not, while for no such class is c' admissible and c inadmissible. Being inadmissible 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 a reason to claim that (ci) less incoherent credence functions are epistemically better than more incoherent ones.

Furthermore, regardless of your rate of exchange, if you have an incoherent credence function, then either it is inadmissible for agents like you, or reducing your pragmatic costs of a large enough amount (i.e. becoming a better calculator) would make it inadmissible, whereas any credence function with a sufficiently low incoherence measure would remain admissible under this change in costs. This supports the claim that coherence plays a normative role for non-ideal agents, much like the ideal specification of a cylinder in the machinist analogy. Like the machinist, (cii) conditional on her having somewhat improved (but not necessarily *perfect*) abilities, the believer ought to more closely approximate her ideal constraint. And (ciii) this gives the believer a reason to improve her ability, since doing so opens her up to advantageous trade-offs between (pragmatic) costs and (epistemic) virtue.

Taken together, these considerations show that a stronger notion of approximate coherence is normatively relevant in scenarios (B), at least under the assumption that the pragmatic costs of becoming coherent are roughly the same for all incoherent credence functions.

5.4 Case (C): Coherence Impossible, Gradational

In scenarios (C) coherence is impossible to achieve, but the agent can pay increasing costs to improve her accuracy of increasing amounts. An example is an agent with credences over such a large set of statements that it is practically impossible for her to figure out all logical and probabilistic relations required by coherence. The agent is incoherent, but she might still be able to spot and fix local violations of the probability axioms without ever achieving full coherence.

In these scenarios, the normative force of my incoherence measures depends on the specific way the agent is unable to reach coherence. If it's impossible to reach full coherence, but it's possible (although impractical) to achieve a very small incoherence measure, then this case approximates case (D). This is because the agent could reach a slightly incoherent credence function c^* which is very close, in terms of accuracy at each world, to the maximal dominator c' (which is coherent). Hence $I_S(c)$, which is the amount by which c' accuracy-dominates c, is a good approximation of the amount by which c^* accuracy-dominates c, i.e. a good approximation of the guaranteed accuracy the agent could achieve by paying the necessary costs, just like in cases where coherence is impractical but achievable (D). I will deal with such cases in the next subsection.

At the other extreme are cases where the possible improvements get the agent nowhere close to coherence. These scenarios look a lot like case (A), where the agent could not really improve their situation. In both cases, the problem is that the measure $\underline{I}_S(c)$ tells us something about the maximal amount of sure accuracy an agent with credence function c could gain, and this involves moving to a coherent credence function. But this trade-off is unavailable under Coherence Impossible, so it's not clear that comparisons induced by this measure are in any way meaningful. So you might worry that, like in case (A), we have to fall back to a very weak notion of approximate coherence.

However, at least under one popular way of measuring inaccuracy, a stronger notion of approximate coherence can still be normatively relevant. Even though the agent cannot reach its maximal dominator (or any other coherent credence function), the measure $\underline{I}_S$ still says something interesting about her epistemic situation, something which can give evaluative meaning to incoherence comparisons. If $\underline{I}_S(c') < \underline{I}_S(c)$, then the maximum accuracy the agent can gain by small nudges of her credence function (which she is capable of making, in these cases) at c', is smaller that the maximum accuracy she can gain by similar nudges at c. This is the content of the following proposition:

Proposition 4. *Let S be the Brier inaccuracy measure, $S(x, w_i) = \sum_{j=1}^{n}(x(A_j) - w(A_j))^2$. Let c, c' credences such that $\underline{I}_S(c') < \underline{I}_S(c)$. Then the agent can gain more guaranteed accuracy by nudging c of some small amount than by nudging c' of some equally small amount. More precisely, let k the amount of let $\overline{\delta}'$ and $\epsilon > 0$ sufficiently small, there is a unit vector $\overline{\delta}$ such that:*

$$\underline{I}_S(\overline{c}) - \underline{I}_S(\overline{c} + \epsilon\overline{\delta}) > \underline{I}_S(\overline{c}') - \underline{I}_S(\overline{c}' + \epsilon\overline{\delta}') \tag{12}$$

Proof. See Appendix □

If we assume that the pragmatic costs of small incoherence improvements don't depend on the starting credence, this result leads to a similar conclusion as the one we had in case (B). If $\underline{I}_S(c) > \underline{I}_S(c')$, and the cost of reducing incoherence by a small nudge in credence is x, then fixing any exchange rate between epistemic utility and pragmatic costs, there is a range of values of x such that nudging your credence is desirable at c but not at c'. So there is a range of abstract, non-ideal agents for which c' is an admissible doxastic state, while c' is inadmissible, and this is a sense in which c is more epistemically faulty of c', regardless of whether the former accuracy-dominates the latter.

As in case (B), conditional on your abilities improving of a sufficient amount, you have an obligation to reduce your incoherence measure accordingly, since the trade-offs between pragmatic costs and guaranteed accuracy involved in some small shift in credence become desirable. And since becoming a better calculator opens you up to such favourable trade-offs, you have a reason to improve your abilities.

Thus a stronger notion of approximate coherence is normatively relevant in scenarios (C), under the assumption that the cost of small incoherence reductions does not decrease with such shifts, and that accuracy is measured by the Brier inaccuracy measure. This measure is well-known among the strictly proper ones due to its many nice properties.[12] And while it's not hard to find a strictly proper measure of inaccuracy for which Proposition 7 fails (see Appendix), it remains an open question whether similar local results hold for some larger class of measures.

5.5 Case (D): Coherence Impractical, Gradational

In this scenario, coherence is achievable but practically demanding, and also, the agent can improve her accuracy of increasingly large amounts by paying increasing pragmatic costs. This seems a more natural way to think of an agent with credences over a large domain of related statements. Although she can spot local incoherences and fix them for some pragmatic cost, the cost required to fix every violation of the probability axioms is too large to be practical.

These cases are potentially problematic for my view. In case (B) we could show that, if $I_S(c') < I_S(c)$, then for a whole family of abstract agents (identified with their pragmatic costs), c' is inadmissible, while c is admissible, since the price of coherence is worth paying at c but not at c'. Yet this only worked because we assumed that the price of becoming coherent remained roughly constant across credence functions, whereas the amount of guaranteed accuracy to be gained decreases with decreasing incoherence, making the trade-off desirable at c but not at c'. This made sense when reaching coherence required learning logical truths, but it's far less obvious once we allow piecemeal improvements.

One might worry here that if c' is less incoherent than c, then less pragmatic costs are required to reach coherence for an agent who holds c' than for an agent who holds c, the abilities of that agent being constant, because c' requires "less improvements". This would mean that whenever an agent finds the trade-off between pragmatic costs and accuracy is desirable at c, the same agent finds it desirable at any less-incoherent c' also; i.e. for any choice of non-ideal agent, either all incoherent credence functions are admissible or they are all inadmissible. Even worse, if pragmatic costs decrease at a fast enough rate with incoherence measure, we could find classes of non-ideal agents for

[12] Joyce (2010, pp.290-293) gives a good summary these nice properties.

whom more-incoherent credence functions are admissible, while less-incoherent ones are inadmissible![13]

This worry reflects an implicit assumption about how non-ideal agents reduce their incoherence. If you think that each improvement step involves resolving inconsistencies on some subset of the domain $\mathcal{F}_k \subseteq \mathcal{F}$, *without altering* the credences assigned by previous improvement steps to other subsets of the domain, then it makes sense to think that a less incoherent agent will have "less steps" or "less changes" left in order to achieve full coherence. Each step independently achieves coherence on some subset $\mathcal{F}_k$, and crystallizes the agent's credences there. So it's reasonable to think that the closer you are to coherence, the less steps you have left to perform, and the less effort this will take. I call this the *piecewise view* of improvements.

There is an alternative, *holistic view* of improvements. This holds that, even if the agent can reach something like local coherence on some subset of the domain, this does not crystallize her credences there. As the agent resolves a violation of the axioms on some subset of the domain, she may want to (or have to) change her credence on other subsets too, even if she has considered them before. Every improvement could bring into question, and must therefore take into account, the agent's credences across the whole domain. Under this view it seems reasonable that, if not constant, the pragmatic costs associated with reaching coherence do not decrease dramatically with decreasing incoherence. So the situation is roughly the same as in case (B): if c is more incoherent than c', then for some class of agents the trade-off between costs and accuracy is desirable at c but not at c'.

What about the piecewise view, then? I think that when improvements work this way (if they ever do), then a strong notion of approximate coherence can at least be normatively relevant at the local level. No matter how small the $\mathcal{F}_k$ involved in each step, the task of reaching coherence on $\mathcal{F}_k$ is not broken down into that of reaching coherence on even smaller subsets. Every improvement on $\mathcal{F}_k$ requires the agent to take into account, and possibly adjust, her credences over the whole $\mathcal{F}_k$. As above, this gives us some reason to think that the cost of becoming coherent on $\mathcal{F}_k$ does not radically decrease as the agent becomes less incoherent on this subset: so a strong notion of approximate coherence is normatively relevant in this local sense.

Neither view of how agents reduce their incoherence is very convincing in general. A slightly better view would be that both strategies are employed at the same time: we focus on addressing a local violation of the probability axioms, but sometimes this clashes with our credences elsewhere in the domain, so we decide whether to change those or to find an alternative resolution, and so on. Still, I doubt that even this mixed strategy would exhaust all the ways in which agents do, or ought to, reason about and then improve their credence. Supporters of the coherence norm have pointed out how

[13] Note that this remains a problem even if we assume, as in case (C), that the cost of *small* incoherence improvements does not decrease with such improvements

difficult it is to give either a normative or descriptive picture of the way incoherent agents improve their epistemic situation (Savage, 1967, pp.57-58), and I don't know of any general, widely accepted theory that provides such a picture.

Until such a theory is available, the best we can do is show that a strong notion of approximate coherence can be normatively relevant for some abstract non-ideal agent like the ones I considered above. This shows that an agent does not need to be a perfect calculator for coherence to play the role of a rationality norm for them. And at least in some contexts, we are similar enough to some of these abstract non-ideal agents. Thus coherence can be a norm for us even when it is impossible or impractical to achieve.

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 are unable to respect the constraints imposed by coherence. Even when it's possible for them to respect these constraints, often it 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 this notion is formally equivalent to the one defended by De Bona and Staffel (2017, 2018); Staffel (2020) when we allow the inaccuracy measure to give different weight to different propositions. Taking advantage of this definition in terms of accuracy, I argued that a strong notion of approximate coherence is normatively relevant for non-ideal agents in a number of contexts, under some fairly demanding assumptions about how these agents deal with incoherence.

Our defence of the normative role of coherence for human-like agents can only ever be as convincing and detailed as our best description of these agents' abilities and how they are employed in reasoning. As it stands, we can only give rough and unconvincing descriptions. So the best we can do is show that ideal coherence can have normative force over simplified, non-ideal agents for whom it is impossible or impractical to satisfy its constraints. This is the sort of defence I have given in this essay. It is enough to dispel the worry that the coherence cannot be a norm for imperfect calculators, and that our being imperfect calculators implies coherence is irrelevant for us.

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

$$\underline{I}_S(c) = \sup\{ \min_{w \in \mathcal{W}_\mathcal{F}} \{S(c,w) - S(f,w)\} \,:\, f \in \mathcal{P}_\mathcal{F} \} \tag{13}$$

where S is a (possibly weighted) strictly proper additive inaccuracy measure. To prove results about this incoherence measure, it is helpful to show that the sup 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, $w_k \in \mathcal{W}_{\mathcal{F}}$, and $\underline{I}_S(c) = S(c, w_k) - S(g, w_k)$.

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

$$\underline{I}_S(c) = S(c, w_k) - S(g, w_k) \tag{14}$$

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

$$m_c(f) = \min_{w \in \mathcal{W}_{\mathcal{F}}} \{S(c, w) - S(f, w)\} \tag{15}$$

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. But the set of coherent credences over $\mathcal{F}$ is a compact subset of $\mathbb{R}^{\mathcal{F}}$, so its image under m_c is a compact subset of $\mathbb{R}$, that is, a closed and bounded subset. Hence:

$$\sup\{m_c(f) \,:\, f \in \mathcal{P}_{\mathcal{F}}\} \in \{m_c(f) \,:\, f \in \mathcal{P}_{\mathcal{F}}\} \tag{16}$$

and thus there is some g coherent such that:

$$\underline{I}_S(c) = \sup\{m_c(f) \,:\, f \in \mathcal{P}_{\mathcal{F}}\} = m_c(g) \tag{17}$$

But then $m_c(g) = S(c, w_k) - S(g, w_k)$ for some $w_k \in \mathcal{W}_{\mathcal{F}}$. So the pair (g, w_k) satisfies (14). □

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 5. *Let S a (possibly weighted) additive, strictly proper inaccuracy measure. Let $c : \mathcal{F} \to \mathbb{R}$ a credence such that $\underline{I}_S(c) = k$. Then:*

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

$$S(c, w') - S(c', w') \leq k \tag{18}$$

2. *There is a $c' : \mathcal{F} \to \mathbb{R}$ such that $S(c, w) - S(c', w) \geq k$ for every w.*

Proof. (*i*): this is immediate from the definition of $\underline{I}_S$. Let $c : \mathcal{F} \to \mathbb{R}$ such that $\underline{I}_S(c) = k$. Assume by way of contradiction that for some $c' : \mathcal{F} \to \mathbb{R}$ we have $S(c, w) - S(c', w) > k$ for every $w \in \mathcal{W}_\mathcal{F}$. Let $w' \in \mathcal{W}_\mathcal{F}$ (one of) the state(s) which minimise this difference (it will always exist, since $\mathcal{W}_\mathcal{F}$ finite). Then:

$$\underline{I}(c) \geq S(c, \omega') - S(c', \omega) > k \tag{19}$$

contradiction.

(*ii*) Let (c', w') instancing the sup and min in the incoherence measure of c (Lemma 1). Clearly the condition holds. □

This shows that if $I_S(c) < I_S(c') = k$ then some credence function is more accurate than c' of at least k in every world, whereas this cannot be the case for c.

Let's introduce a few more definitions before moving to the proof of Propositions 3 and 7. 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 $\bar{c} = (c(A_1), ..., c(A_m))$. If y is a coherent credence on $\mathcal{F}$, then it is the convex combination of 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:

$$y = \sum_{i=1}^{n} y_i^* w_i \tag{20}$$

Definition .1 (Projection vertices, projection space, projection hyperplane)**.** Let y be the d_S-closest coherent credence to c.

1. The *projection vertices* of c under S are the worlds $w_i \in \mathcal{W}_F$ such that $y_i^* \neq 0$. Denote by $\mathcal{K}(c)$ the set of projection vertices of c.

2. The *projection space* of c under S is the convex hull $CH(\mathcal{K}(c))$ of its projection vertices (clearly $y \in CH(\mathcal{K}(c))$).

3. The *projection hyperplane* of c is the smallest-dimensional hyperplane containing the projection space.

As an example, if the d_S-closest coherent credence y to c corresponds to the truth-value assignment of a world w_i, then the set of projection vertices $\mathcal{K}(c)$, the projection space, and the projection hyperplane are the singleton set $\{w_i\}$. If y is positive at only two worlds w_i, w_j, then $\mathcal{K}(c) = \{w_i, w_j\}$, the projection space is the segment connecting w_i to w_j, and the projection hyperplane is the line through those two points. Note that, because the elements of $\mathcal{W}_\mathcal{F}$ are extremes of the set of coherent credences, and y lays on the boundary of the set of coherent credences, if w_k is not a projection vertex then it will not be contained in the projection hyperplane.

The proof of Proosition 3 will use the following lemma:

Lemma 2. *Let S be a (possibly weighted) additive, strictly proper inaccuracy measure, let $c : \mathcal{F} \to \mathbb{R}$ an incoherent credence, and let y be the d_S-closest coherent credence to c. Then 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. Let $\mathcal{F} = \{A_1, ..., A_n\}$ and $\mathcal{W}_{\mathcal{F}} = \{w_1, ..., w_n\}$. Denote by d_S the Bregman divergence associated with the inaccuracy measure S, and let z be the d_S-closest credence to c. If $z = w_i$ for some i, then $\mathcal{K}(c) = \{w_i\}$ and the lemma is trivially true. So assume $z \neq w_i$. We know z minimises:

$$f(\overline{x}) = S(\overline{c}, \overline{x}) - S(\overline{x}, \overline{x}) \tag{21}$$

where $\overline{x} = (x_1, ..., x_m) = (x(A_1), ..., x(A_m))$ for any credence x. Because x is a coherent credence, we can think of the m-dimensional vector $\overline{x}$ as a function of the $(n-1)$-dimensional vector $\overline{x}^* = (x_1^*, ..., x_{n-1}^*)$ (the value x_n^* is determined by the first $n-1$ values). That is, we think of each x_j as the function of $(x_1^*, ..., x_{n-1}^*)$ defined as follows:

$$x(A_j) = x_j = \sum_{w_i : w_i(A_j)=1} x_i^*. \tag{22}$$

where we write x_n^* as shorthand for $(1 - x_1^* - ... - x_{n-1}^*)$.

Write f^* to denote the function mapping $\overline{x}^*$ to $f(x)$. So $\overline{z}$ minimises f among the coherent credence vectors on $\mathcal{F}$ iff $\overline{z}^*$ minimises f^* among the probability vectors on $\mathcal{W}_{\mathcal{F}}$. For this to happen it is necessary that:

$$\frac{\partial}{\partial x_i^*} f^*(\overline{z}^*) = 0 \tag{23}$$

whenever this partial derivative is defined at $\overline{z}^*$. If the partial derivative with respect to x_i^* is not defined at $\overline{z}^*$, that means that z_i^* is either 0 or 1. If no partial derivatives are defined at $\overline{z}^*$, then $z = w_i$ for some i, which is a case we have already dealt with. So z_i^* is never 1, meaning the partial derivative with respect to x_i^* is defined iff $z_i^* \neq 0$, that is, iff w_i is a projection vertex.

Let $w_i \in K(c)$ a projection vertex. We can expand the left-hand side of (23) to obtain:

$$\sum_{j=1}^{m} \frac{\partial}{\partial x_i^*} \left[x_j(s(c_j, 1) - s(x_j, 1)) - (1 - x_j)(s(c_j, 0) - s(x_j, 0))\right] = 0 \tag{24}$$

where x_j is a function of $(x_1^*, ..., x_{n-1}^*)$ as in (22). Each j in the sum corresponds to one A_j, and for each such A_j exactly one of the following will be true

(i) $w_i(A_j) = 1$ & $w_n(A_j) = 1$,

(ii) $w_i(A_j) = 1$ & $w_n(A_j) = 0$,

(iii) $w_i(A_j) = 0$ & $w_n(A_j) = 1$,

(iv) $w_i(A_j) = 0$ & $w_n(A_j) = 0$.

So we can split the sum into four parts, each corresponding to one of these conditions, and add them all up. Starting with part (i), we have:

$$\sum_{j:w_i(A_j)=1\,\&\,w_n(A_j)=1} \frac{\partial}{\partial x_i^*} \left[x_j(s(c_j,1) - s(x_j,1)) - (1-x_j)(s(c_j,0) - s(x_j,0))\right] \tag{25}$$

Here x_j is the linear function $x_{k_1}^* + ... + x_i^* + ... + x_n^*$, since $w_i(A_j) = 1 = w_n(A_j)$. But recall that x_n^* was just shorthand for $(1 - x_1^* - ... - x_n^*)$. So the two occurrences of x_i^* cancel out, and x_j is constant in x_i^*, meaning its partial derivative w.r.t. x_i will be zero. In the same way, the composite function $s(x_j, 1)$ will also be constant in x_i^*, and hence have partial derivative zero. So part (i) of the sum will vanish. Now let's look at part (ii):

$$\sum_{j:w_i(A_j)=1\,\&\,w_n(A_j)=0} \frac{\partial}{\partial x_i^*} \left[x_j(s(c_j,1) - s(x_j,1)) - (1-x_j)(s(c_j,0) - s(x_j,0))\right] \tag{26}$$

here $x_j = x_{k_1}^* + ... + x_i^* + ... + x_{k_l}^*$ is a linear function of x_i^*, so its derivative is 1. Similarly, $\frac{\partial}{\partial x_i^*} s(x_j, 1) = s'(x_j, 1)$. So the partial derivative in (26) gives:

$$\sum_{j:w_i(A_j)=1\,\&\,w_n(A_j)=0} s(c_j,1) - s(x_j,1) - x_j s'(x_j,1) + (1-x_j)s'(x_j,0) - s(c_j,0) + s(x_j,0) \tag{27}$$

But for each sentence A_j, when s is a proper score we have that (Shuford et al., 1966)[14]:

$$-x_j s'(x_j,1) + (1-x_j)s'(x_j,0) = 0 \tag{29}$$

So the expression above simplifies to:

$$\sum_{j:w_i(A_j)=1\,\&\,w_n(A_j)=0} s(c_j,1) - s(x_j,1) - s(c_j,0) + s(x_j,0) \tag{30}$$

[14]The result is used by Schervish (1989) to characterise proper scoring rules. In the present notation, it states that if $s(t,0)$ and $s(t,1)$ (seen as distinct functions from $\mathbb{R}$ to $\mathbb{R}$) are both differentiable on $[0,1]$, $s(t,0)$ is non-increasing on $[0,1]$, and $s(t,1)$ is non-decreasing on $[0,1]$, then s is a proper score iff

$$-t\frac{d}{dt}s(t,1) = (1-t)\frac{d}{dt}s(t,0) \tag{28}$$

which we can further rewrite as:

$$\sum_{j:w_i(A_j)=1\,\&\,w_n(A_j)=0} s(c_j, w_i(A_j)) - s(x_j, w_i(A_j)) - s(c_j, w_n(A_j)) + s(x_j, w_n(A_j)) \tag{31}$$

Proceeding in the same way, part (iii) of the sum becomes:

$$\sum_{j:w_i(A_j)=0\,\&\,w_n(A_j)=1} s(c_j, w_i(A_j)) - s(x_j, w_i(A_j)) - s(c_j, w_n(A_j)) + s(x_j, w_n(A_j)) \tag{32}$$

and part (iv) vanishes in the same way as part (i). Note that the expressions in the sum in (31) and (32) are the same. So the condition that the partial derivative w.r.t. x_i^* of f^* is zero amounts to:

$$\sum_{j:w_i(A_j)\neq w_n(A_j)} s(c_j, w_i(A_j)) - s(x_j, w_i(A_j)) - s(c_j, w_n(A_j)) + s(x_j, w_n(A_j)) = 0 \tag{33}$$

we can separate the terms with w_n and move them to the right to obtain:

$$\sum_{j:w_i(A_j)\neq w_n(A_j)} s(c_j, w_i(A_j)) - s(x_j, w_i(A_j)) = \sum_{j:w_i(A_j)\neq w_n(A_j)} s(c_j, w_n(A_j)) - s(x_j, w_n(A_j)) \tag{34}$$

Then note that, for the A_j such that $w_i(A_j) = w_n(A_j)$, we have $s(t, w_i(A_j)) = s(t, w_n(A_j))$ for any t. So by adding appropriate terms on both sides, we can rewrite the condition as:

$$\sum_{j=1}^{m} s(c_j, w_i(A_j)) - s(x_j, w_i(A_j)) = \sum_{j=1}^{m} s(c_j, w_n(A_j)) - s(x_j, w_n(A_j)) \tag{35}$$

which, by definition of the inaccuracy measure S, is just:

$$S(c, w_i) - S(x, w_i) = S(c, w_n) - S(x, w_n) \tag{36}$$

Hence the necessary condition (23) tells us that $S(c, w_i) - S(z, w_i)$ is constant for all $w_i \in \mathcal{K}(c)$. □

Proposition 6. *Let S be a (possibly weighted) additive, strictly proper inaccuracy measure which assigns the same weight to each sentence in $\mathcal{F}$. Then for every credence function $c : \mathcal{F} \to \mathbb{R}$, we have $\underline{I}_S(c) = I_S(c)$.*

Proof. Let $\mathcal{F} = \{A_1, ..., A_n\}$ and $\mathcal{W}_\mathcal{F} = \{w_1, ..., w_n\}$. Denote by d_S the Bregman divergence associated with the inaccuracy measure S. Let $c : \mathcal{F} \to \mathbb{R}$ and incoherent credence, and let z be the d_S-closest credence to c. 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, 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 $w \in \mathcal{W}_\mathcal{F}$:

$$S(c, w_y) - S(y, w_y) \leq S(c, w) - S(y, w) \tag{37}$$

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

$$S(c, w_y) - S(y, w_y) \geq \min_{w \in \mathcal{W}_\mathcal{F}} \{S(c, w) - S(z, w)\} \tag{38}$$

We will now appeal to the following fact, proven later:

- **Fact 1** If $w_i \in \mathcal{K}(c)$ and $w_k \in \mathcal{W}_\mathcal{F} \setminus \mathcal{K}(c)$, then $S(c, w_j) - S(z, w_j) \geq S(c, w_i) - S(z, w_i)$.

Combining Fact 1 with (37) (38), we obtain that whenever w_i is a projection vertex the following inquality holds:

$$S(c, w_i) - S(y, w_i) \geq S(c, w_y) - S(y, w_y) \geq S(c, w_i) - S(z, w_i) \tag{39}$$

since $S(c, w_i) - S(z, w_i)$ is constant for all such w_i, by Lemma 2. Equation (39) simplifies to $S(y, \omega_i) \leq S(z, \omega_i)$. But then:

$$\begin{aligned}
Exp_z(S(y, \cdot) - S(z, \cdot)) &= \sum_{i=1}^{n} z_i^*(S(y, w_i) - S(z, w_i)) && (40)\\
&= \sum_{i: z_i^* \neq 0} z_i^*(S(y, w_i) - S(z, w_i)) && (41)\\
&= S(y, w_i) - S(z, w_i) \quad \text{(since this is constant for } w_i \in \mathcal{K}(c) && (42)\\
&\leq 0 \quad \text{by (39)} && (43)
\end{aligned}$$

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:

$$\begin{aligned}
\underline{I}_S(c) &= S(c, w_y) - S(y, w_y) && (44)\\
&= S(c, w_y) - S(z, w_y) \quad \text{(because } z = y) && (45)\\
&= S(c, w_i) - S(z, w_i) \quad \text{for a } w_i \in \mathcal{K}(c) && (46)\\
&= \sum_{i=1}^{n} z_i^*(S(c, w_i) - S(z, w_i)) \quad \text{(because } z_i^* \neq 0 \text{ iff } w_i \in \mathcal{K}(c)) && (47)\\
&= Exp_z(S(c, \cdot) - S(z, \cdot)) && (48)\\
&= d_S(c, z) && (49)\\
&= I_S(c) \quad \text{(by construction of } z.) && (50)
\end{aligned}$$

So the proof is complete if we can prove **Fact 1**.

Fact 1 Proof: Because the space of coherent credences on $\mathcal{F}$ is a closed convex set, and $\overline{z}$ is the d_S-closest vector to $\overline{c}$, $\overline{z}$ is the d_S-projection of $\overline{c}$ on the space. Hence the generalised Pythagorean theorem for Bregman divergences holds (Predd et al., 2009):

$$d_S(c, w_j) \geq d(c, z) + d(z, w_j) \tag{51}$$

from this we derive:

$$\begin{aligned}
d_S(c, w_j) - d(z, w_j) &\geq d(c, z) && (52)\\
S(c, w_j) - S(z, w_j) &\geq Exp_z\left[S(c, \cdot) - S(z, \cdot)\right] && (53)\\
S(c, w_j) - S(z, w_j) &\geq \sum_{t:z(w_t)\neq 0} S(c, w_t) - S(z, w_t) && (54)\\
S(c, w_j) - S(z, w_j) &\geq S(c, w_i) - S(z, w_i) && (55)
\end{aligned}$$

which proves Fact 1. □

Finally, I show that when S is the Brier score, the corresponding incoherence measure $\underline{I}_S$ carries some information about the small local improvements available to the agent (Proposition 4). First I need a further definition. From this point onwards I assume that S is the Brier inaccuracy measure.

Definition .2 (Optimal direction and optimal line)**.** A unit vector $\overline{\delta} \subseteq \mathbb{R}^m$ is an *optimal direction* for c iff for any unit vector $\overline{\zeta}$:

$$-\max_{w\in\mathcal{W}_\mathcal{F}}\{\nabla S(\overline{c}, w)\cdot\overline{\delta}\} \geq -\max_{w\in\mathcal{W}_\mathcal{F}}\{\nabla S(\overline{c}, w)\cdot\overline{\zeta}\} \tag{56}$$

In other words, $\overline{\delta}$ is optimal when nudging c in its direction guarantees the greatest amount of accuracy compared to any other direction. If $\overline{\delta}$ is an optimal direction, the line in $\mathbb{R}^m$ with direction $\overline{\delta}$ is called an *optimal line* for c.

I prove the result in two steps: first I show that for every c, the optimal direction is towards the d_S-closest coherent credence (Lemma 1); secondly I show that the rate of increase of guaranteed accuracy in this direction is a monotone increasing function of $\underline{I}_S(c)$ (Lemma 2), so $\underline{I}_S(c) < \underline{I}_S(c')$ implies that the optimal nudge at c guarantees greater accuracy than the optimal at c'. In these proofs I will refer to two more technical lemmas which are proved afterwards (Lemma 3 and Lemma 4).

Lemma 3. *Let c an incoherent credence defined over $\mathcal{F} = \{A_1, ..., A_m\}$, and S the Brier inaccuracy measure. The unique optimal direction for c is the unit vector:*

$$\frac{c - y}{||c - y||} \tag{57}$$

where y is the d_S-closest coherent credence to c.

Proof. For $\epsilon > 0$ small, let $\overline{c}_y = \overline{c} + \epsilon \frac{\overline{c}-\overline{y}}{\|c-y\|}$ and let $\overline{c}' = \overline{c} + \epsilon\overline{\zeta}$, where ζ is a different unit vector. I need to show that a shift in credence from c to c_y guarantees greater accuracy than one from c to c', i.e:

$$\min_{w \in \mathcal{W}_{\mathcal{F}}} \{S(c,w) - S(c_y,w)\} > \min_{w \in \mathcal{W}_{\mathcal{F}}} \{S(c,w) - S(c',w)\} \tag{58}$$

Let y' be the d_S-closest coherent credence to c'. Note that because d_S is just the squared Euclidean distance, and c_y is on the straight line between c and the d_S-closest point y on the convex set $\mathcal{P}_{\mathcal{F}}$, the squared Euclidean distance between c_y and y is smaller than the one between c' and y'. That is:

$$d_S(c_y, y) < d_S(c', y') \tag{59}$$

By propriety of S, since y is coherent we know:

$$Exp_y[S(y,\cdot) - S(y',\cdot)] < 0 \tag{60}$$

$$\sum_{i=1}^{n} y_i^*(S(y,w_i) - S(y',w_i)) < 0 \tag{61}$$

$$\sum_{i:y_i^*>0} y_i^*(S(y,w_i) - S(y',w_i)) < 0 \tag{62}$$

So there must be some $j \in \{1, ..., n\}$ such that $y_j^* > 0$ and:

$$S(y,w_j) - S(y',w_j) < 0. \tag{63}$$

that is, there is some projection vertex w_j such that y is more accurate than y' at w_j. Then we have:

$$S(c,w_j) - S(c',w_j) - (S(c,w_j) - S(c_y,w_j)) \tag{64}$$
$$= S(c_y,w_j) - S(c',w_j) \tag{65}$$
$$= d_S(c,w_j) - d_S(c',w_j) \qquad (\text{since } S(w_j,w_j) = 0) \tag{66}$$
$$= d_S(c_y,y) + d_S(y,w_j) - d_S(c',y') - d_S(y',w_j) \tag{67}$$
$$= d_S(c_y,y) - d_S(c',y') + S(y,w_j) - S(y',w_j) \tag{68}$$
$$< 0 \qquad (\text{by inequalities (59) and (63)}) \tag{69}$$

And thus:

$$\min_{w \in \mathcal{W}_{\mathcal{F}}} \{S(c,w) - S(c',w)\} \leq S(c,w_j) - S(c',w_j) \tag{70}$$
$$< S(c,w_j) - S(c_y,w_j) \tag{71}$$

But it turns out that, when nudging c towards the d_S-closest coherent credence y, the world w which achieves the minimum in $\min_{w\in\mathcal{W}_\mathcal{F}}\{S(c,w)-S(c_y,w)\}$ is always a projection vertex (Lemma 6). Furthermore, the increase in accuracy for a nudge in this direction is the same for all projection vertices, since for every point x on the optimal line (so for both c and c_y), $S(c,w_i)-S(x,w_i)$ is the same for all the $w_i\in\mathcal{K}(c)$ by Pythagoras' theorem (this is shown in more detail in the proof of Lemma 4). Therefore we can write:

$$\min_{w\in\mathcal{W}_\mathcal{F}}\{S(c,w)-S(c',w)\}<S(c,w_j)-S(c_y,w_j)=\min_{w\in\mathcal{W}_\mathcal{F}}\{S(c,w)-S(c_y,w)\} \tag{72}$$

This proves that the guaranteed accuracy gained by nudging c in the direction of y is greater than the guaranteed accuracy gained by nudging c in any other direction. □

Lemma 4. *Let c an incoherent credence defined over $\mathcal{F}=\{A_1,...,A_m\}$, S the Brier inaccuracy measure, and y the d_S-closest coherent credence to c. The rate at which we obtain guaranteed accuracy by nudging c towards y:*

$$-\max_{w\in\mathcal{W}_\mathcal{F}}\left\{\frac{d}{dt}S((1-t)c+ty,w)\bigg|_{t=0}\right\} \tag{73}$$

is a monotone growing function of $d_S(c,y)=\underline{I}_S(c)$.

Proof. By Lemma 6, which is proved later, we can get rid of the maximum in (73). Lemma 6 shows the maximum is achieved at a projection vertex $w_i\in\mathcal{K}(c)$. But for all such w_i the derivative is the same, because:

$$\begin{aligned}
\frac{d}{dt}S((1-t)c+ty,w_i)&=\frac{d}{dt}d_S((1-t)c+ty,w_i) &(74)\\
&=\frac{d}{dt}\left(d_S((1-t)c+ty,y)+d_S(y,w_i)\right) &(75)\\
&=\frac{d}{dt}d_S((1-t)c+ty,y) &(76)
\end{aligned}$$

which does not depend on the choice of w_i. Furthermore:

$$\begin{aligned}
\frac{d}{dt}d_S((1-t)c+ty,y)&=\frac{d}{dt}(S((1-t)c+ty,y)-S(y,y)) &(77)\\
&=\frac{d}{dt}S((1-t)c+ty,y) &(78)\\
&=\frac{d}{dt}\sum_{j=1}^m((1-t)c+ty-y)^2 &(79)\\
&=-2\sum_{j=1}^m(c_j-y_j)(c_j-tc_j+ty_j-y_j) &(80)
\end{aligned}$$

At $t = 0$, this becomes:

$$-2\sum_{j=1}^{m}(c_j - y_j)^2 = -d_S(c, y) \tag{81}$$

So the rate at which accuracy increases by nudging c towards y is a function of $d_S(c, y)$, and is greater the larger $d_S(c, y) = \underline{I}_S(c)$. □

Lemma 3 and 4 allow us to prove Proposition **??**:

Proposition 7. *Let c, c' credences such that $\underline{I}(c') < \underline{I}(c)$. Then the agent can gain more guaranteed accuracy by nudging c by some small amount than by nudging c' by some equally small amount. More precisely, for any unit vector $\overline{\delta}'$, there is a unit vector $\overline{\delta}$ such that:*

$$-\max_{w\in\mathcal{W}_{\mathcal{F}}}\{\nabla S(\overline{c}, w)\cdot\overline{\delta}\} > -\max_{w\in\mathcal{W}_{\mathcal{F}}}\{\nabla S(\overline{c'}, w)\cdot\overline{\delta}'\} \tag{82}$$

Proof. By Lemma 3, the maximal rate at which accuracy increases by nudging c is obtained by nudging c in the direction of its d_S-closest coherent credence. By Lemma 4, this rate is $\underline{I}_S(c)$, which is smaller than $\underline{I}_S(c')$, the rate at which accuracy increases by nudging c' in the direction of its d_S-closest credence. □

Now let's tie together the ends left loose in the proofs of Lemma 3 and 4. A further technical lemma will help with the proof of Lemma 6.

Lemma 5. *Let c an incoherent credence defined over $\mathcal{F} = \{A_1, ..., A_m\}$, S the Brier inaccuracy measure, and y the d_S-closest coherent credence to c. If $w_k \in \mathcal{W}_{\mathcal{F}} \setminus \mathcal{K}(c)$, then nudging y in the optimal direction for c increases accuracy at w_k.*

Proof. Denote by y' the d_S-projection of w_k on the optimal line for c (the line which passes through c and y, whose direction is the optimal direction for c). Then y must lay between y' and c. To see this, assume it does not, say y' lays between y and c instead (see Figure 1 for an example of the following construction). The optimal line is orthogonal to the projection space $CH(\mathcal{K}(c))$, and also orthogonal to the line connecting y' and w_k. Hence the line connecting y' and w_k is parallel to the projection hyperplane. We know w_k cannot be on the projection hyperplane. And by construction of y, we know $d_S(c, w_k) > d_S(c, y)$. Hence if we draw the line connecting w_k and y, and project c on this line, we get a point y'' which lays between y and w_k and so is coherent, since it is a convex combination of coherent credences y and w_k. Then by Pythagoras' theorem:

$$d_S(c, y'') = d_S(c, y) - d_S(y, y'') \leq d_S(c, y) \tag{83}$$

which contradicts the assumption that y uniquely minimises $d_S(c, x)$ among coherent credences. The same reasoning can be applied when c lays between y and y'. So y must lay between y' and c, as needed.

Now let $z(x, y, y')$ be the credence obtained by nudging y towards y' by some small amount x. We can write:

$$S(z(x,y,y'), w_k) = d_S(z(x,y,y'), w_k) \tag{84}$$
$$= d_S(z(x,y,y'), y')) + d_S(y', w_k) \tag{85}$$

and hence:

$$\frac{d}{dx} S(z(x,y,y'), w_k) = \frac{d}{dx} d_S(z(x,y,y'), y') \tag{86}$$

which is negative, as increasing x reduces the squared distance d_S between $z(x, y, y')$ and y'. Hence inaccuracy decreases at w_k by nudging y in the optimal direction for c. □

(FIGURE HERE)

Figure 2: Example of the construction in the proof of Lemma 3. The projection vertices are w_1, w_2, the projection hyperplane is the line passing through these two vertices, and the projection space is the segment between w_1 and w_2. The optimal line for c is highlighted in magenta.

Draft

Our last lemma shows that the maximum in (73) is achieved at a projection vertex $w_i \in \mathcal{K}(c)$.

Lemma 6. *Let c an incoherent credence defined over $\mathcal{F} = \{A_1, ..., A_m\}$, and S the Brier inaccuracy measure. Let y the d_S-closest credence to c. Then if $w_k \in \mathcal{W}_{\mathcal{F}} \setminus \mathcal{K}(c)$, there is a projection vertex $w_i \in \mathcal{K}(c)$ such that:*

$$\frac{d}{dt} S((1-t)c + ty, w_i)|_{t=0} \geq \frac{d}{dt} S((1-t)c + ty, w_k)|_{t=0} \tag{87}$$

Hence the maximum in (73) is always achieved at a projection vertex.

Proof. Let $w_k \in \mathcal{W}_{\mathcal{F}} \setminus \mathcal{K}(c)$ and $w_i \in \mathcal{K}(c)$. As shown in the proof of Lemma 2:

$$\frac{d}{dt} S((1-t)c + ty, w_i) = -2 \sum_{j=1}^{m} (c_j - y_j)(c_j - tc_j + ty_j - y_j) \tag{88}$$

this is strictly negative for $0 \leq t < 1$, equals 0 at $t = 1$, and is strictly positive at $t + \epsilon$ for any positive ϵ. Thus shifting c in the optimal direction increases accuracy at w_i (i.e. decreases $S(x, w_i)$) at a progressively smaller rate, and once we reach y, further shifts in the same direction decrease accuracy at w_i (i.e. increase $S(x, w_i)$).

For an arbitrary $w \in W_{\mathcal{F}}$ we can compute the second derivative of $S(x, w)$ on the optimal line by:

$$\frac{d^2}{dt^2} S((1-t)c + ty, w) = \sum_{j=1}^{m} \frac{d^2}{dt^2} s((1-t)c_j + ty_j, w(A_j)) \tag{89}$$

$$= \frac{d}{dt} 2 \sum_{j=1}^{m} (y_j - c_j)(c_j - tc_j + ty_j - w(A_j)) \tag{90}$$

$$= 2 \sum_{j=1}^{m} (y_j - c_j)^2 \tag{91}$$

$$= 2d_S(y, c) \tag{92}$$

Hence the rate at which $\frac{d}{dt} S((1-t)c + ty, w)$ changes is constant in t, strictly positive, and does not depend on w.

Now assume by way of contradiction that for every $w_i \in \mathcal{K}(c)$,

$$\frac{d}{dt} S((1-t)c + ty, w_i)|_{t=0} < \frac{d}{dt} S((1-t)c + ty, w_k)|_{t=0} \tag{93}$$

As t goes from 0 to 1, both sides of the inequality increase at the same rate, given that $\frac{d^2}{dt^2} S((1-t)c + ty, w)$ is positive, constant in t, and constant across all $w \in \mathcal{W}_{\mathcal{F}}$. Hence:

$$\frac{d}{dt} S((1-t)c + ty, w_i)|_{t=1} < \frac{d}{dt} S((1-t)c + ty, w_k)|_{t=1} \tag{94}$$

$$0 < \frac{d}{dt} S((1-t)c + ty, w_k)|_{t=1} \tag{95}$$

This means that shifting y in the optimal direction decreases accuracy at w_k, contradicting Lemma 3.

□

Draft

References

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

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

David K Lewis. *On the plurality of worlds*, volume 322. Blackwell Oxford, 1986.

Richard Pettigrew. *Dutch book arguments*. Cambridge University Press, 2020.

Bruno de Finetti. *Foresight*. Publisher name, 1937.

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

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

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

Jennifer Rose Carr. Why ideal epistemology? *Mind*, 2021.

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

Ian Hacking. Slightly more realistic personal probability. *Philosophy of Science*, 34(4): 311–325, 1967.

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

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

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

Mark J Schervish, Teddy Seidenfeld, and Joseph B Kadane. Two measures of incoherence: How not to gamble if you must. 1998.

Mark J Schervish, Teddy Seidenfeld, and Joseph B Kadane. Measuring incoherence. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 561–587, 2002.

Richard Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, 2016.

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

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

Emir H Shuford, Arthur Albert, and H Edward Massengill. Admissible probability measurement procedures. *Psychometrika*, 31(2):125–145, 1966.

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

Joel B Predd, Robert Seiringer, Elliott H Lieb, Daniel N Osherson, H Vincent Poor, and Sanjeev R Kulkarni. Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792, 2009.