Rational Probabilistic Incoherence

Michael Caie

1 Introduction

The following is a plausible principle of rationality:

PROBABILISM A rational agent's credences should always be probabilistically coherent.

To say that an agent's credences are **probabilistically coherent** is to say that such credences can be represented by a function $Cr(\cdot)$ satisfying the following constraints:

NORMALIZATION For any tautology $\top$, $Cr(\top) = 1$

NON-NEGATIVITY For any proposition ϕ, $0 \leq Cr(\phi)$

FINITE ADDITIVITY If ϕ and ψ are incompatible propositions, then $Cr(\phi \vee \psi) = Cr(\phi) + Cr(\psi)$

It has been argued that PROBABILISM follows given the plausible assumption that our primary epistemic goal is to represent the world as accurately as possible. Joyce [1998] and Joyce [2009] argue that for any probabilistically incoherent credal state C that an agent might have, there is a probabilistically coherent credal state C^* that is guaranteed to be more accurate than C no matter what the world is like, while the reverse is never true. Call this the **accuracy-dominance argument** for PROBABILISM. Since it is plausible that a rational agent should try to have as accurate a credal state as possible, the accuracy-dominance argument, if successful, would seem to provide good reason to endorse PROBABILISM.

PROBABILISM, however, has some surprising consequences. In particular, it can be shown that there are cases in which it is impossible for an agent with moderately good access to her own credal state and a high credence in an obvious truth to be probabilistically coherent. If PROBABILISM is true, it follows that in certain cases rationality requires that an agent either be ignorant of her own credences or be ignorant of an obvious truth.

We might simply accept this consequence of PROBABILISM, despite its prima facie implausibility. In this paper, I'll argue that this isn't the right response. To this end, I'll show that the cases in which probabilistic coherence demands either ignorance of one's own credal state or ignorance of an obvious truth can

be used to expose flaws in the accuracy-dominance argument for PROBABILISM. Once these flaws are exposed, we can see that considerations of accuracy, instead of motivating PROBABILISM, support the claim that in certain cases a rational agent's credences ought to be probabilistically *incoherent*.

The paper proceeds as follows.

In §2, I present a case in which an agent is guaranteed to be probabilistically incoherent given that she is moderately sensitive to her own credences and has high credence in an obvious truth.

In §3, I consider the bearing that this case has on the accuracy-dominance argument for PROBABILISM. I first outline the accuracy-dominance argument. I show, by appeal to our earlier case, that there are crucial steps in the argument that are invalid. We can grant, as the argument assumes, that a credal state is defective insofar as it is guaranteed to be less accurate than some other credal state. It doesn't follow, however, that probabilistic coherence is rationally required. For there are cases in which the most accurate credal state that an agent can have is one that is probabilistically incoherent. Assuming that an agent ought to try to be as accurate as possible, it follows that in these cases an agent ought to be probabilistically incoherent.

In §4, I present the accuracy-dominance argument for PROBABILISM in more explicit decision-theoretic terms. In decision theory, we can sometimes argue that an act or act-type is rationally required by showing that the act, or act-type, is better than all the alternatives no matter what the state of the world. In this type of case, we say that the act or act-type **dominates** its alternatives. It is well known, however, that in order to apply dominance reasoning in this way, the acts and states must be related in a particular way. We call this relation **independence**. In this section, I show that the accuracy-dominance argument for probabilism, framed in explicit decision-theoretic terms, fails when it is applied to the case discussed in §2 because the acts and states appealed to in this argument are not independent. I show, further, that when this defect is corrected, we can provide an accuracy-dominance argument for the conclusion that in this case it is rationally required that the agent be probabilistically incoherent.

The case that I appeal to §§2-4 crucially involves a proposition such that necessarily the truth-value of this proposition depends on a particular agent's credence in that very propositions. This raises the question of whether there is some suitably restricted version of PROBABILISM that might still be made to work. Is there some large interesting class of propositions such that a rational agent's credences over those propositions must always be probabilistically coherent?

In §5, I take up this question and argue that the answer is: no. In principle, almost any proposition is such that an agent may rationally fail to have probabilistically coherent credences in that proposition and its negation. Nonetheless, as a matter of fact, the conditions that allow for this are almost certainly extremely rare. For any actual agent, there will be a very large class of propositions such that the agent's credences in that class of propositions ought to be probabilistically coherent. But, in principle, this need not be so.

2 A Prima Facie Problem for Probabilism

In this section, I'll show that PROBABILISM has the following surprising consequence. In certain cases, PROBABILISM requires a rational agent to satisfy the following disjunctive obligation: either the agent must have low credence in an obvious truth or the agent must be insensitive to its own credal state. This disjunctive obligation is, I'll argue, at least prima facie implausible, and so this case provides us with prima facie reason to be skeptical of PROBABILISM.

In §§3-4, I'll argue that this skepticism is warranted. This case can be used to show that in certain situations a probabilistically *incoherent* credal state maximizes credal accuracy. This, I claim, gives us good reason to reject PROBABILISM.

Consider an agent who we'll call 'Yuko'. Let $(*)$ refer to the following sentence:

Yuko's credence that $(*)$ is true isn't greater than or equal to 0.5.[1]

We'll use 'Cr_y' to abbreviate 'Yuko's credence that...'. The above can, then, be represented as:

$(*)$ $\neg Cr_y T(*) \geq 0.5$

As an instance of the T-schema we have:

(1) $T(*) \leftrightarrow \neg Cr_y T(*) \geq 0.5$

If classical logic is correct (and I'll assume here that it is), then we shouldn't accept *every* instance of the T-schema.[2] As is well known, there are certain instances of this schema, e.g., instances involving liar sentences, that are inconsistent given classical logic. We should certainly reject *these* biconditionals. In the vast majority of cases, however, there is no conflict with classical logic. Given the intuitive plausibility of these biconditionals, if there is no logical reason to reject an instance of this schema, we should, I think, endorse it.[3] (1) is perfectly consistent with classical logic. We should, therefore, accept this claim.[4]

We make the following assumptions about Yuko:

[1]Here we achieve sentential self-reference via stipulation as in Kripke [1975]. This could also be achieved by a coding technique such as Gödel-numbering.

[2]It is worth noting that if we give up the assumption that classical logic is correct, then there are interesting ways of treating the types of cases we'll be looking at in this section. See [Author Suppressed] for a discussion of how to treat similar cases that arise for qualitative belief using non-classical resources.

[3]There are perfectly general non-ad-hoc treatments of the truth predicate for which this holds. This will, e.g., hold according to the theory KF. See, e.g., Field [2008] for a description of this theory.

[4]For those who are worried about the truth of (1), let me note that this case could be easily run without appeal to a truth-predicate. What is required for this case is that there be some proposition ϕ for which we have: $\phi \leftrightarrow \neg Cr_y(\phi) \geq 0.5$. The claim that $(*)$ is true is a particularly simple case, but there are other propositions that could work. For example, we might assume that Yuko is constituted so that whether she will make a free-throw in basketball depends on how confident she is that she will make the shot. We can assume that she will make

(2) $Cr_y(T(*) \leftrightarrow \neg Cr_y T(*) \geq 0.5) \geq 1 - \epsilon$

(3) $[Cr_y T(*) \geq 0.5] \rightarrow [Cr_y(Cr_y T(*) \geq 0.5)) > 0.5 + \epsilon]$

(4) $[\neg Cr_y T(*) \geq 0.5] \rightarrow [Cr_y(\neg Cr_y T(*) \geq 0.5)) > 0.5 + \epsilon]$

If $\epsilon = 0$, then Yuko is completely certain of the truth expressed in (1), and is at least mildly sensitive to her own credence in the truth of $(*)$. As ϵ increases, Yuko's credence in (1) decreases and her assumed sensitivity to her credence in the truth of $(*)$ increases. These assumptions are jointly satisfiable as long as: $0 \leq \epsilon \leq 0.5$.[5]

We can show:

From (2) - (4), it follows that Yuko is probabilistically incoherent.

To see this, first assume: $Cr_y T(*) \geq 0.5$. By (3), we have:

$Cr_y(Cr_y T(*) \geq 0.5) > 0.5 + \epsilon$. From (2), we know that if Yuko is probabilistically coherent, then $|Cr_y \neg T(*) - Cr_y(Cr_y T(*) \geq 0.5)| \leq \epsilon$. Thus, assuming that Yuko is probabilistically coherent, we have: $Cr_y \neg T(*) > 0.5$. Given our original assumption, this guarantees that $Cr_y T(*) + Cr_y \neg T(*) > 1$. But probabilistic coherence requires, in general, that $Cr_y(\phi) + Cr_y(\neg\phi) = 1$. On the assumption that Yuko has credence greater or equal to 0.5 in the truth of $(*)$, it follows that Yuko is probabilistically incoherent.

Next, assume: $\neg Cr_y T(*) \geq 0.5$. By (4), we have: $Cr_y(\neg Cr_y T(*) \geq 0.5) > 0.5 + \epsilon$. From (2), we know that if Yuko is probabilistically coherent, then $|Cr_y T(*) - Cr_y(\neg Cr_y T(*) \geq 0.5)| \leq \epsilon$. Thus, assuming that Yuko is probabilistically coherent we have: $Cr_y T(*) > 0.5$. But this is incompatible with our assumption that $\neg Cr_y T(*) \geq 0.5$. It follows that on the assumption that Yuko doesn't have credence greater than or equal to 0.5 in the truth of $(*)$ that Yuko is probabilistically incoherent.

Since it follows that Yuko is probabilistically incoherent both on the assumption that $Cr_y T(*) \geq 0.5$ and on the assumption that $\neg Cr_y T(*) \geq 0.5$, we can conclude that Yuko is probabilistically incoherent.

---

the shot, but only if she is less confident that she will make it than that she will miss. This is no doubt an unusual situation, but there doesn't seem to be anything impossible about things being this way. Instead of setting up the case using the claim that $(*)$ is true, then, we could use the claim that Yuko will make the relevant free-throw. We'll look in more detail at this type of case in §5. For now, however, I'll stick with (1).

[5] If $\epsilon > 0.5$, then by (3) and (4) we would either have to have $Cr_y(Cr_y T(*) \geq 0.5)) > 1$ or $Cr_y(\neg Cr_y T(*) \geq 0.5)) > 1$. But, I assume, it is impossible for an agent to have credence greater than 1 in any proposition. If one rejects this assumption, then one will hold that there are more values for ϵ for which (2)-(4) are jointly satisfiable.

Rational obligations, I assume, are closed under logical consequence. That is, we have: $\phi \models \psi \Rightarrow O\phi \models O\psi$. We have seen that Yuko will satisfy the requirements imposed by PROBABILISM only if either (2) fails to hold, (3) fails to hold, or (4) fails to hold. If PROBABILISM is true, then it follows that Yuko is rationally obligated to be such that one of (2)-(4) fail.

For (2) to fail, Yuko must have credence in (1) below $1 - \epsilon$. For (3) to fail, Yuko must have credence greater than or equal to 0.5 in the truth of $(*)$, but have at best $0.5 + \epsilon$ credence that her credence in the truth of $(*)$ falls in this range. For (4) to fail, Yuko must have credence greater than or equal to 0.5 in the truth of $(*)$, but have at best $0.5 + \epsilon$ credence that her credence in the truth of $(*)$ fails to fall in this range. PROBABILISM thus prohibits Yuko from both being highly confident that (1) expresses a truth and being somewhat sensitive to her own credence in the truth of $(*)$.

This should, I think, strike you as rather surprising. (1), after all, is an obvious truth. As long as Yuko is aware of what sentence $(*)$ refers to, it would seem that she should be rationally permitted in having a high credence in this proposition. And it is, I think, prima facie implausible that rationality may require an agent to be ignorant of her own credences. Given the implausibility of the claim that Yuko is rationally required to either have a low credence in (1) or be insensitive to her credence in the truth of $(*)$, we have some reason to doubt that PROBABILISM is true.

This, of course, doesn't provide anything like a conclusive argument against PROBABILISM. For while it is certainly prima facie plausible that Yuko could have high credence in (1) and be sensitive to her own credal state without thereby being irrational, PROBABILISM is also prima facie plausible. At present, then, all we have are a set of prima facie plausible claims that are jointly incompatible. We don't yet have clear reason for rejecting PROBABILISM in particular.[6] In the next section, however, I'll show that on closer inspection this case provides us with the materials to develop a strong argument against PROBABILISM.

3 Probabilism and Accuracy

We can assess a qualitative doxastic state in terms of how accurate it is. Consider an agent's attitude towards a single proposition, ϕ. If ϕ is true, we can say:

- Believing ϕ is more accurate than being agnostic about ϕ, and being agnostic about ϕ is more accurate than disbelieving ϕ.

While, if ϕ is false, we can say:

[6] Cases similar to this have been used by Andy Egan and Adam Elga to argue that there are certain surprising restrictions on what claims a rational agent may have high credence in. See Elga and Egan [2005]. According to Egan and Elga, the conclusion that we should draw from this case is that it is impermissible for Yuko to have a high credence in (1). I'll discuss Egan and Elga's argument in §5.

- Disbelieving ϕ is more accurate then being agnostic about ϕ, and being agnostic about ϕ is more accurate than believing ϕ.

It's plausible to think that our primary epistemic goal in forming beliefs is to represent matters as accurately as we can. In forming beliefs we aim to have true beliefs and avoid having false beliefs.

We may appeal to the fact that accuracy is our primary epistemic goal to justify certain claims about doxastic rationality. For example, we may argue that it is never rational to believe $\phi \wedge \neg\phi$. Since $\phi \wedge \neg\phi$ is guaranteed to be false, we are guaranteed to be more accurate if we don't believe $\phi \wedge \neg\phi$ than if we do. Since we ought to try to be as accurate as possible in our judgments, and since this goal is best achieved by never believing $\phi \wedge \neg\phi$, we ought not believe $\phi \wedge \neg\phi$.

Just as we can assess a qualitative doxastic state for accuracy, so too can we assess a quantitative doxastic state, i.e., a credal state. Consider an agent's credence in single proposition ϕ. If ϕ is true, we can say:

- A higher credence in ϕ is more accurate than a lower credence.

While, if ϕ is false, we can say:

- A lower credence in ϕ is more accurate than a higher credence.

Just as it is plausible to think that our primary goal in forming qualitative doxastic attitudes is to be as accurate as we can in our judgments of truth value, so too is it plausible that our primary goal in forming quantitative doxastic attitudes is to be as accurate as we can in our estimation of truth values.

It is a tricky question exactly how credal accuracy should be measured. There are numerous ways of measuring the accuracy of credences in particular propositions that meet the above constraints. And there are numerous ways of measuring the accuracy of a total credal state given the accuracy of particular credences.

It has been argued, however, in [Joyce, 1998] and [Joyce, 2009], that for any reasonable way of measuring accuracy the following hold:

PCA 1 For any probabilistically incoherent credal state C, there is a probabilistically coherent credal state C^*, such that C^* would be more accurate than C, no matter what the actual world is like.

PCA 2 For any probabilistically coherent credal state C^*, there is no probabilistically incoherent credal state C, such that (i) C would be at least as accurate as C^* no matter what the actual world is like, and (ii) C would be more accurate than C^* given at least one possible state of the world.

Given PCA 1 and 2, a powerful argument can be given for PROBABILISM. By PCA 1, if an agent has a probabilistically incoherent credal state C, there is some probabilistically coherent credal state C^* that would have been more accurate than C no matter what the actual world is like. Assuming that accuracy is our

primary epistemic goal, it follows that from an epistemic perspective the agent should see C^* as being preferable to C. By PCA 2, there is no countervailing reason to find any probabilistically incoherent credal state preferable to C^*. Thus, from an epistemic perspective, an agent should always prefer being probabilistically coherent to being probabilistically incoherent.[7]

What PCA 1 and PCA 2 show, if they're correct, is that the goal of credal accuracy is best achieved by being probabilistically coherent. Assuming that one ought to try to have credences that are as accurate as possible, it follows that one ought to be probabilistically coherent.

I'm happy to say that accuracy is our primary epistemic goal. Indeed, I'll assume that this is so throughout this paper. But this idea doesn't support PROBABILISM. For, both PCA 1 and PCA 2 are false. To show this, I'll show that there are cases in which:

There is a probabilistically incoherent credal state C such that, for any probabilistically coherent credal state C^*, an agent would be less accurate were her credal state to be C^* instead of C, no matter what the actual world is like.

In certain cases, the goal of credal accuracy is best achieved by being probabilistically *incoherent*. Since one ought to try to have credences that are as accurate as possible, in such cases one ought to have probabilistically incoherent credences.

In what follows, we'll consider an agent who has credences defined over a finite algebra of propositions $\mathcal{P}$.[8] To say that $\mathcal{P}$ is an algebra is to say that membership in the set is closed under negation and finite disjunction. We'll represent the agent's credal state by the function $Cr(\cdot)$.

In arguing for PCA 1 and PCA 2 , Joyce goes to great lengths to try to show that these claims will hold for a large number of possible ways of measuring the accuracy of credences. For the sake of simplicity, I will focus on one of these measures, but none of the points that follow turn essentially on any idiosyncratic features of this measure.

We assume, then, the following:

BRIER ACCURACY Given an agent with credences $Cr(\cdot)$, located in a world w, the *accuracy* of the agent's credences is given by:

$$1 - [(1/n) \sum_{\phi \in \mathcal{P}} (Cr(\phi) - w(\phi))^2]$$

[7] I find this argument quite convincing. But see Easwaran and Fitelson [forthcoming] for an interesting argument that other epistemic goods may in certain cases rule out accuracy dominating credal states.

[8] In order to ensure finiteness, I'll assume that a proposition is identical to the set of worlds in which it is true. Of course, for certain purposes we may want think of propositions in a more fine-grained way, but for our purposes here nothing will be lost by taking this coarse-grained approach. I should also note that nothing essential turns on our assumption that the algebra over which the credences are defined is *finite*, but this will help simplify the presentation at certain points.

Here $w(\phi)$ is the truth-value of the proposition ϕ at the world w. This will be 1, if ϕ is true at w, and 0, if it is false at w.

$$(1/n) \sum_{\phi \in \mathcal{P}} (Cr(\phi) - w(\phi))^2$$

is the so-called Brier score. Amongst those who think that PROBABILISM is supported by the doxastic goal of accuracy, this is often taken to be the best measure of a credal state's *inaccuracy*.[9] Proponents of this measure of inaccuracy will take

$$1 - [(1/n) \sum_{\phi \in \mathcal{P}} (Cr(\phi) - w(\phi))^2]$$

to provide our measure of accuracy. Given the popularity and plausibility of this view, it is a nice case to focus on.

Assume that there are n propositions in $\mathcal{P}$. We can represent possible credal states as points in the space $\mathbb{R}^n$. A point in this space is specified by an n-tuple $< x_1, x_2, ..., x_n >$, such that every $x_i \in \mathbb{R}$. Pick some arbitrary bijection F from $\{x : 1 \leq x \leq n\}$ onto $\mathcal{P}$. We can then view the point $< x_1, x_2, ..., x_n >$ as representing a credal state $Cr(\cdot)$, such that $Cr(F(i)) = x_i$. That is, $< x_1, x_2, ..., x_n >$ represents a credal state in which the agent has credence x_i in the proposition represented by the i-th variable under the mapping F.

We can also represent possible-worlds in such a space. A point $< x_1, x_2, ..., x_n >$ represents a possible world w just in case for every i such that $1 \leq i \leq n$, $w(F(i)) = x_i$.[10] A point representing a possible world will be such that each $x_i \in \{0, 1\}$; although not every distribution of 0s and 1s will necessarily represent a genuine possibility. Let's label the set of points in $\mathbb{R}^n$ representing possible-worlds $\mathcal{W}$.

The set of probabilistically coherent credal states can be identified as the convex hull of $\mathcal{W}$. This is the set of points in $\mathbb{R}^n$ that can be written as weighted sums of member of $\mathcal{W}$, with the weightings summing to 1.[11] Let's label this set $\mathcal{C}$.

Finally, we can define the following measure on $\mathbb{R}^n$. Let $x = < x_1, x_2, ..., x_n >$, and $y = < y_1, y_2, ..., y_n >$. We say:

$$B(x, y) =_{df} 1 - [(1/n) \sum_{i=1}^{n} (x_i - y_i)^2]$$

[9] See Joyce [2009] for a catalog of the virtues of this measure. See also Leitgeb and Pettigrew [2010a], Leitgeb and Pettigrew [2010b].

[10] What I'm calling "possible-worlds" are, of course, not maximally specific metaphysical possibilities. Instead they are sets of such possibilities that agree on the members of $\mathcal{P}$.

[11] A little more pedantically: Let W be a function listing the members of $\mathcal{W}$, i.e., a bijective function from some interval [1, n] of $\mathbb{N}^+$ onto $\mathcal{W}$. The convex hull of $\mathcal{W}$ is the set of points x such that there is some set Λ of non-negative numbers such that $\sum_{\lambda_i \in \Lambda} \lambda_i = 1$ for which $x = \sum_{i=1}^{n} \lambda_i W(i)$.

We're now in a position to state the arguments for PCA 1 and PCA 2. The arguments for these claims rely on the following mathematical results:

THEOREM 1 Given any point in $x \in \mathbb{R}^n - \mathcal{C}$, there is a point $y \in \mathcal{C}$, such that for every $w \in \mathcal{W}$, $B(y, w) > B(x, w)$.[12]

THEOREM 2 Given any point $x \in \mathcal{C}$, there is no point $y \in \mathbb{R}^n - \mathcal{C}$ such that (i) for every $w \in \mathcal{W}$, $B(y, w) \geq B(x, w)$, and (ii) for some $w \in \mathcal{W}$, $B(y, w) > B(x, w)$.[13]

Given THEOREM 1 and BRIER ACCURACY, it is tempting to argue for PCA 1 as follows:

(1a) By THEOREM 1, for any point x—representing a probabilistically incoherent credal state C—there is a point y—representing a probabilistically coherent credal state C^*—such that for every possible world w, $B(y, w) > B(x, w)$.

(1b) By BRIER ACCURACY, $B(x, w)$ is a measure of the accuracy of having the credal state represented by x in world w.

(1c) Thus, by (1a) and (1b), it follows that for any probabilistically incoherent credal state C, there is some probabilistically coherent credal state C^*, such that, for any world w, one is more accurate if one has credence C^* in w, than if one has credence C in w.

PCA 1 Thus, from (1c), it follows that for any probabilistically incoherent credal state C, there is a probabilistically coherent credal state C^*, such that C^* would be more accurate than C, no matter what the actual world is like.

Similarly, given THEOREM 2 and BRIER ACCURACY, it is tempting to argue for PCA 2 as follows:

[12]For a simple proof, see, e.g., Williams [forthcoming].

[13]Proof sketch: Let $x \in \mathcal{C}$ and $y \in \mathbb{R}^n - \mathcal{C}$. We can show that there is some $w \in \mathcal{W}$ such that $B(x, w) > B(y, w)$.

Let $H = \{z : z \cdot (x - y) = 1/2((x \cdot x) - (y \cdot y))\}$. This is the hyperplane that runs perpendicular to the vector $x - y$ containing the point $1/2(x + y)$. Let S be the half space such that $S = \{z : z \cdot (x - y) \geq 1/2((x \cdot x) - (y \cdot y))\}$. Let $\mathring{S}$ denote the interior of S. Note that $x \in \mathring{S}$ and for every $z \in \mathring{S}$, $B(x, z) > B(y, z)$.

Since $x \in \mathcal{C}$, there is a function W listing the members of $\mathcal{W}$, i.e., a bijective function from some interval [1, n] of $\mathbb{N}^+$ onto $\mathcal{W}$ and a set Λ of non-negative numbers such that $\sum_{\lambda_i \in \Lambda} \lambda_i = 1$, such that $x = \sum_{i=1}^{n} \lambda_i W(i)$. Since $x \cdot (x - y) > 1/2((x \cdot x) - (y \cdot y))$, it follows that $\sum_{i=1}^{n} \lambda_i (W(i) \cdot (x - y)) > 1/2((x \cdot x) - (y \cdot y))$. This guarantees that there is some $w_i \in \mathcal{W}$ such that $w_i \in \mathring{S}$. And so there is some $w_i \in \mathcal{W}$, such that $B(x, w) > B(y, w)$.

(2a) By THEOREM 2, for any point x—representing a probabilistically coherent credal state C^*—there is no point y—representing a probabilistically incoherent credal state C—such that (i) for every possible world w, $B(y, w) \geq B(x, w)$, and (ii) for some possible world w, $B(y, w) > B(x, w)$.

(2b) By BRIER ACCURACY, $B(x, w)$ is a measure of the accuracy of having the credal state represented by x in world w.

(2c) Thus, by (2a) and (2b), it follows that for any probabilistically coherent credal state C^*, there is no probabilistically incoherent credal state C, such that, (i) for any world w, one is at least as accurate if one has credal state C in w, as one is if one has credal C^* in w, and (ii) for some world w, one is more accurate if one has credal state C in w, than if one has credal state C^* in w.

PCA 2 Thus, from (2c), it follows that for any probabilistically coherent credal state C^*, there is no probabilistically incoherent credal state C, such that (i) C would be at least as accurate as C^*, no matter what the actual world is like, and (ii) C would be more accurate than C^*, given at least one possible state of the world.

Both of these arguments, though prima facie plausible, are ultimately flawed. There are two problems with each argument.

The first problem is with the inference, cited in (1b) and (2b), from BRIER ACCURACY to the claim that, in our geometric model, $B(x, w)$ measures the accuracy of having a credal state x in a world w. Although this may seem completely obvious, there are good reasons to reject this inference.[14] I'm going to bracket this worry until the next section, however, since there is a more fundamental problem. For now, then, I'll assume that (1b) and (2b) are correct.

The more fundamental problem with this argument is that PCA 1 doesn't follow from (1a)-(1b), and PCA 2 doesn't follow from (2a)-(2b). Indeed, while we may accept both (1a)-(1b) and (2a)-(2b), both PCA 1 and PCA 2 are false. On the way to showing this, I'll first show that the inference from (1c) to PCA 1, and the inference from (2c) to PCA 2, are invalid.

Recall the case of Yuko. This case featured a proposition, viz., the proposition that $(*)$ is true, that was true just in case a certain agent did not have credence at or above 0.5 in that proposition.[15] Consider the smallest algebra containing the proposition expressed by $(*)$, i.e., the algebra consisting of this proposition, the negation of this proposition, the logical truth $\top$, and the contradiction $\bot$. Since $\top$ is true no matter what, and $\bot$ false no matter what, credal accuracy will

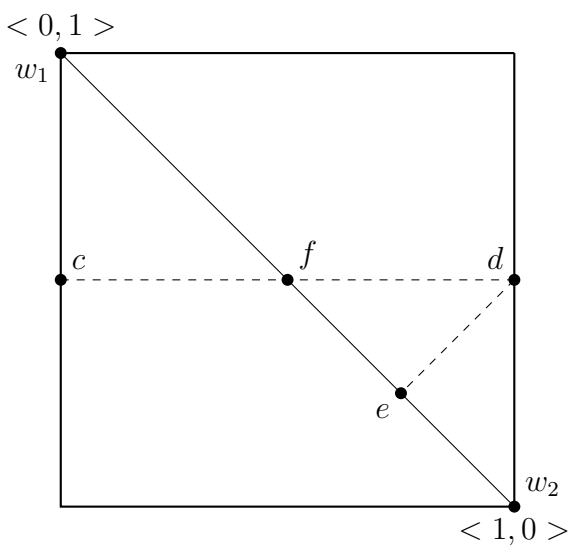
[14] See the discussion of (3a) in the following section.

[15] This claim is only correct on the assumption that $(*)$ refers rigidly to an *interpreted* sentence. If we took $(*)$ to simply refer to a string of graphemes, then, despite the fact that, in the actual world, $(*)$ is true just in case Yuko doesn't have credence at or above 0.5 that $(*)$ is true, this need not hold at some other world in which those graphemes have a different meaning. So, let's make that assumption.

be maximized by having credence 1 in $\top$ and credence 0 in $\bot$. I'll assume that Yuko has these credences in these propositions. Yuko's possible credal states, then, will differ in what credences are assigned to the proposition expressed by $(*)$, and to its negation.

We can represent these credal states as points in $\mathbb{R}^2$. We'll let x_1 represent the negation of the proposition expressed by $(*)$, and x_2 represent the proposition expressed by $(*)$. The point $w_1 = < 0, 1 >$, then, represents the possible world in which the proposition expressed by $(*)$ is true and its negation is false, while the point $w_2 = < 1, 0 >$ represents the possible world in which this proposition is false and its negation is true.

Let's focus on the credal states in $[0, 1]^2$. We can represent these states graphically. In referring to the following graph, be sure to keep in mind that $B(x, y)$ will be greater the smaller the Euclidean distance between x and y.



The probabilistically coherent states are represented by the points on the line-segment between w_1 and w_2. Consider the points $d = < 1, 0.5 >$ and $e = < 0.75, 0.25 >$. Point d represents a probabilistically incoherent credal state, while point e represents a probabilistically coherent credal state. e is, in fact, one of the points that accuracy-dominates d in the manner characterized by THEOREM 1. Thus we have: $\forall w\ B(e, w) > B(d, w)$.[16]

In this case, however, we can see that it doesn't follow from the fact that $\forall w\ B(e, w) > B(d, w)$ that were Yuko to have credal state e she would be more accurate than if she were to have credal state d. The reason for this is that in this case which of w_1 or w_2 is actual depends on what Yuko's credal state is. In particular:

- If Yuko were to have credal state d, then w_2 would be actual.

[16] A quick calculation will verify that: $B(d, w_1) = 0.375 < 0.4375 = B(e, w_1)$, and $B(d, w_2) = 0.875 < 0.9375 = B(e, w_2)$.

- If Yuko were to have credal state e, then w_1 would be actual.

When asking whether Yuko would be more accurate were she to have credal state e or credal state d, the only values that we need to compare, then, are $B(d, w_2)$ and $B(e, w_1)$. And here we see: $B(d, w_2) = 0.875 > 0.4375 = B(e, w_1)$. In this case, then, despite the fact that we have: $\forall w\ B(e, w) > B(d, w)$, it's nonetheless true that:

Yuko would be more accurate were she to have the probabilistically incoherent credal state d, than were she to have the probabilistically coherent credal state e.

This case shows us that the inference from (1c) to PCA 1 isn't valid. While from: $\forall w\ B(e, w) > B(d, w)$, we may infer (at least if we grant (1b)) that, for every world w, e is more accurate as evaluated at w than d, we can't infer from this fact that Yuko would be more accurate were she to have credal state e instead of credal state d, since which world is actual is different depending on whether Yuko has credal state d or credal state e.

This case similarly shows us that the inference from (2c) to PCA 2 isn't valid. In accordance with THEOREM 2, we have that there is no point $x \in \mathbb{R}^n - \mathcal{C}$ such that (i) $\forall w\ B(x, w) \geq B(e, w)$, and (ii) $\exists w\ B(x, w) > B(e, w)$. We may infer from this (if we grant (2b)) that there is no probabilistically incoherent credal state x that is at least as accurate as e as evaluated at every possible world, and more accurate than e as evaluated at some possible world. But we can't infer from this fact that there is no probabilistically incoherent credal state x such that (i) no matter what the state of the world, were x to be Yuko's credal state, Yuko would be at least as accurate as she would be were her credal state to be e, and (ii) for some state of the world, were x to be Yuko's credal state, Yuko would be more accurate than she would were her credal state to be e. For, as we've seen, Yuko would be more accurate were she to have the probabilistically incoherent credal state d, than the probabilistically coherent credal state e. And this subjunctive claim holds no matter which world is actual.

Having demonstrated that the inferences from (1c) and (2c) to PCA 1 and PCA 2 are invalid, we now turn to showing that the latter claims are, in fact, false. To do this, we'll show:

The most accurate credal state that Yuko could have is represented by d.

By BRIER ACCURACY, we can measure the accuracy of a credal state $Cr(\cdot)$, located in a world w, by:

$$1 - [(1/n) \sum_{\phi \in \mathcal{P}} (Cr(\phi) - w(\phi))^2]$$

If we think of $1 - (Cr(\phi) - w(\phi))^2$ as a measure of the accuracy of having a particular credence in the proposition ϕ at a world w, then we can think of the

accuracy of a credal state as simply being the average accuracy of the credences determined by that state in particular propositions.

The first point to note is that, with respect to the proposition expressed by $(*)$, the most accurate credence that Yuko can have is 0.5. To see this, refer back to our earlier graph. Let l be the line segment connecting points c and d. Let X be the set of points on the graph at or above l. Let Y be the set of points below l. We know the following facts:

(5) For every point $x \in X$, were Yuko to have credal state x, then w_2 would be actual.

(6) For every point $y \in Y$, were Yuko to have credal state y, then w_1 would be actual.

From (6), it follows that, for the members of Y, accuracy with respect to $(*)$ increases as the value of the x_2 co-ordinate (i.e., the vertical coordinate) increases, with this value always being $< .75$.

From (5), it follows that, for the members of X, accuracy with respect to $(*)$ increases as the value of the x_2 co-ordinate decreases, with maximal accuracy being 0.75. This is reached when the x_2 coordinate is 0.5. This shows that the most accurate credence that Yuko can have in the proposition expressed by $(*)$ is 0.5. The credal states with this property are those located on the line l.

If Yuko has a credal state located on l, then we know that w_2 is actual. In w_2, the negation of $(*)$ is true. Given that w_2 is actual, the most accurate that Yuko can be with respect to the negation of $(*)$ is to have credence 1 in that proposition. Indeed, if this is the case, Yuko will be maximally accurate with respect to the negation of $(*)$, i.e., there is no other possible credal state that Yuko could have which would make her more accurate with respect to the negation of $(*)$. Amongst the credal states on l, d is the only credal state in which Yuko has credence 1 in the negation of $(*)$. This establishes the following:

d is the unique credal state that has the highest possible accuracy with respect to both the proposition expressed by $(*)$ and its negation.

It follows that were Yuko to have some credal state other than d, she would be less accurate with respect to at least one of these propositions without there being any corresponding gain in her accuracy with respect to the other. If, for example, Yuko were to have some other credal state on l, she would be less accurate with respect to the negation of $(*)$ without any corresponding gain in accuracy with respect to $(*)$. And if Yuko were to have some other credal state not on l, she would be less accurate with respect to $(*)$ without any corresponding gain in accuracy with respect to the negation of $(*)$.

Since the accuracy of a credal state is simply the average of the accuracy of the particular credences sanctioned by that state in particular propositions, and since credal state d is the unique credal state that maximizes accuracy with

respect to both $(*)$ and its negation, it follows that d is the most accurate credal state that Yuko could have. Were Yuko to have any other credal state, Yuko would be less accurate.

Since the probabilistically incoherent credal state represented by d is the most accurate credal state that Yuko could have, it follows that both PCA 1 and PCA 2 are false. Thus, the argument for PROBABILISM outlined earlier fails.

Indeed, we're now in a position to see that an appeal to the epistemic goal of credal accuracy actually motivates the rejection of PROBABILISM. For, given that one ought to try to have as accurate a credal state as one can, and given that the most accurate credal state that Yuko can have is one that is probabilistically incoherent, it follows that Yuko ought be probabilistically incoherent.

4 Accuracy and Decision Theory

In this section, I'll present the accuracy-dominance argument for PROBABILISM in more explicit decision-theoretic terms.[17] Doing so helps highlight where the dominance argument for PROBABILISM goes wrong, and how dominance reasoning may be used to argue against PROBABILISM.

Call a quadruple: $\mathbf{D} = < \mathbf{A}, \mathbf{S}, \mathbf{U}, \mathbf{C} >$, **a decision problem**. Both $\mathbf{A}$ and $\mathbf{S}$ are sets of propositions. We call $\mathbf{A}$ the set of **acts** and $\mathbf{S}$ the set of **states**. Think of the members of $\mathbf{A}$ as propositions describing various acts that an agent may undertake.[18] Think of the members of $\mathbf{S}$ as propositions describing various ways the world might be that are relevant to the outcomes that would obtain were the acts described by the members of $\mathbf{A}$ to be performed. We assume that both $\mathbf{S}$ and $\mathbf{A}$ form partitions of the space of possible-worlds. $\mathbf{U}$ is a **utility function** that assigns to propositions of the form $A_i \wedge S_j$ a number that measures of the utility that would result for the agent were the act described by A_i to be performed in state S_j. Finally, $\mathbf{C}$ is a credence function that is defined on an algebra containing all propositions of the form $A_i \wedge S_j$.[19]

Given a decision problem $\mathbf{D}$, we say:

- An act A_1 **strongly dominates** an act A_2 (in $\mathbf{D}$) just in case for every $S_i \in \mathbf{S}$, $\mathbf{U}(A_1 \wedge S_i) > \mathbf{U}(A_2 \wedge S_i)$.
- An act A_1 **weakly dominates** an act A_2 (in $\mathbf{D}$) just in case (i) for every $S_i \in \mathbf{S}$, $\mathbf{U}(A_1 \wedge S_i) \geq \mathbf{U}(A_2 \wedge S_i)$, and (ii) for some $S_i \in \mathbf{S}$, $\mathbf{U}(A_1 \wedge S_i) > \mathbf{U}(A_2 \wedge S_i)$.

[17] See Pettigrew [2011a] and Pettigrew [2011b] for a helpful survey of some of the uses of decision theoretic machinery in epistemology.

[18] We'll be somewhat promiscuous with what we consider an act. In particular, we'll count an agent's coming to have a particular credal state as an act. This shouldn't, though, be seen as an endorsement of a questionable doxastic voluntarism.

[19] As we'll see, there are further constraints that we will want to put on decision problems. For now, however, it is useful to simply think about decision problems as having this minimal structure.

Let $\mathcal{A}$ and $\mathcal{B}$ be sets of acts. We say:

- $\mathcal{A}$ **act-dominates** $\mathcal{B}$ just in case (i) for every $B \in \mathcal{B}$, there is some $A \in \mathcal{A}$ such that A strongly dominates B, and (ii) there is no $B \in \mathcal{B}$ such that, for some $A \in \mathcal{A}$, B weakly dominates A.

Here are two putative norms that we might appeal to, given a decision problem, to single out a certain option or set of options as rationally obligatory.

DOMINANCE 1 If A_i strongly dominates all other members of $\mathbf{A}$, then A_i is rationally required.

DOMINANCE 2 If $\mathcal{A}$ and $\mathcal{B}$ partition $\mathbf{A}$, and $\mathcal{A}$ act-dominates $\mathcal{B}$, then it is rationally required that one choose some option in $\mathcal{A}$.

(Note that it's crucial that for DOMINANCE 1 we assume that A_i strongly dominates *every other member* of $\mathbf{A}$, and that for DOMINANCE 2 we assume that $\mathcal{A}$ and $\mathcal{B}$ *partition* $\mathbf{A}$.)

We can now present Joyce's argument for PROBABILISM using this decision theoretic framework. We can represent an agent's epistemic situation as a decision problem $\mathbf{D}$. We let $\mathbf{A}$, the set of "acts" available to an agent, be the set of propositions describing possible credal states, given an algebra $\mathcal{P}$, that a particular agent could have. We let $\mathbf{S}$, the set of states, be the set of propositions describing possible distributions of truth-values for the members of $\mathcal{P}$. We let $\mathbf{U}$ be a measure of the agent's epistemic utility, which we take to be measured by the accuracy of an agent's credal state given a particular distribution of truth-values. We, then, argue as follows:

(3a) If A is the credal state represented by x, and S the state represented by w, then, by BRIER ACCURACY, $\mathbf{U}(A \wedge S) = B(x, w)$.

(3b) By (3a), THEOREM 1 and THEOREM 2, it follows that, relative to $\mathbf{D}$, the set of probabilistically incoherent credal states are act-dominated by the set of probabilistically coherent credal states.

(3c) By (3b) and DOMINANCE 2, it follows that an agent is rationally required to have a probabilistically coherent credal state.

As with the argument in the previous section, we can locate two problems with this argument for PROBABILISM. Again, it will help in getting clear on where the argument breaks down to focus on the case of Yuko.

In accordance with the above argument, we can think of Yuko as facing the following epistemic decision problem, $\mathbf{D}^1_Y$. We let $\mathbf{A}^1_Y$ be the set of propositions describing possible credences that Yuko could have in the proposition expressed by $(*)$, and in its negation.[20] We let $\mathbf{S}^1_Y$ be the set of propositions describing

[20] We'll continue to assume that Yuko has credence 1 in $\top$ and credence 0 in $\bot$.

possible distributions of truth-values for these propositions. $\mathbf{S}^1_Y$ will, of course, have two members: S_1, in which the proposition expressed by $(*)$ is true and its negation is false, and S_2, in which these truth-values are reversed. Finally, we assume that $\mathbf{U}^1_Y(A_i \wedge S_i) = B(x, w)$, where x is the point in $\mathbb{R}^2$ representing A_i and w is the point representing S_i.

The first problem in the above argument is with (3a).[21] Grant BRIER ACCURACY. That is, grant that given an agent with credences $Cr(\cdot)$, located in a world w, the accuracy of the agent's credences is given by:

$$1 - [(1/n) \sum_{\phi \in \mathcal{P}} (Cr(\phi) - w(\phi))^2]$$

Still, it doesn't follow that, in the decision problem at hand, if x represents credal state A, and w represents a world state S, the epistemic utility of $A \wedge S$ is given by $B(x, w)$.

The reason for this is that in this decision problem not all conjunctions of the form $A \wedge S$ describe genuine possibilities, i.e, possible situations in which Yuko has credal state A in state S. For example, let A_e be the proposition that Yuko has the credal state represented by point e in our earlier graph, and let S_2 be the state represented by point w_2. We know that the conjunction $A_e \wedge S_2$ is impossible. Of course, we can assign a number to this conjunction by using the measure B defined on $\mathbb{R}^2$. But this number does not represent the epistemic utility of the possible situation in which Yuko has the credal state represented by A_e in state S_2; for there simply is no possible situation in which Yuko has this credal state and S_2 obtains.

One way of bringing out the problem here is to note that if we were to say that $B(x, w)$ always measures the epistemic utility of $A \wedge S$ (where A is the credal state represented by x, and S the state represented by w), then we would be committed to inconsistent assignments of epistemic utility to sets of possible worlds. To see this, let A_d be the proposition that Yuko has the credal state represented by point d in our earlier graph and let S_1 be the state represented by w_1. Since both $A_e \wedge S_2$ and $A_d \wedge S_1$ are impossible, they both describe the same set of possible worlds, viz., the null set. But it's easy to verify that $B(e, w_2) \neq B(d, w_1)$. Even if we could make sense of an assignment of utilities to the null set of worlds (and I doubt we can) we should surely want to hold that this utility is unique. Taking $B(x, w)$ to measure epistemic utility wouldn't allow for this.

We can draw a lesson from this first problem:

If we want to model an agent's epistemic position as a decision problem, we should make sure that we choose our states so that they are compatible with each of the agent's possible credal states.

[21] The problem that arises here is the same problem that arises with premisses (1b) and (2b) in the argument in the previous section. I earlier noted that there is a problem with these premises but deferred in depth discussion. The points that follow should make it clear why appeal to (1b) and (2b) in our earlier argument is problematic.

In a moment we'll see how to do this, but first let's look at the second problem with the above argument.

The second problem can be located in the appeal to DOMINANCE 2. Dominance reasoning is certainly plausible. After all, if some option (or set of options) is better than the alternatives *no matter what the world is like* how could it not be better *tout court*? It is well known, however, that one needs to be careful in how one sets up a decision problem if dominance reasoning is not to lead us astray.[22] Consider the following situation:

> **Bounty:** A large sum of money has been stolen from a local crime boss and you've been framed. There's a bounty on your head paying an exorbitant sum of money in return for your death. You can either flee to the mountains or stay home. If you stay at home you're very likely to be shot, and you know this. If you flee, though, there's a decent chance you'll escape alive, and you know this. You would, however, prefer to live at your house than in the mountains. You'd also prefer, somewhat, to be killed at home than to be killed in the mountains. Of course, you strongly prefer living to dying (whether in the mountains or at home). What should you do?

We might represent this situation using the following decision problem we'll label $\mathbf{D}_B$. In $\mathbf{D}_B$ there are two acts available to you: staying home, and fleeing to the mountains. And there are two possible states: in one state you are killed, in another state you live. The utilities can be represented by the following matrix:

	Die	Live
Stay	1	5
Flee	0	3

Applying either DOMINANCE 1 or DOMINANCE 2 to this decision problem yields the verdict that the rational thing to do is to stay at home.

But this is clearly wrong. If you stay at home then you're almost certain to be killed, while if you flee there's a good chance you may escape with your life; and you know these facts. Since you'd prefer to live rather than die, you should flee.

Pretty much everyone agrees that in this type of case dominance reasoning leads us astray. It turns out, however, to be a matter of some controversy exactly why this reasoning fails. Everyone agrees that in order to apply dominance reasoning to a decision problem the acts and states must in some sense be **independent**. However, there is disagreement about exactly what this condition of act-state independence amounts to.

According to **evidential decision theorists**, in order to apply dominance

[22] See Jeffrey [1983] and Joyce [1999] for discussion of some ways in which dominance reasoning may fail.

reasoning to a decision problem the acts and states must *probabilistically independent*.[23] That is, for each act A and state S, we must have: $\mathbf{C}(S|A) = \mathbf{C}(S)$.

According to **causal decision theorists**, in order to apply dominance reasoning to a decision problem, the acts and states must be *causally independent*.[24] That is, for each act A and state S, A must neither causally promote nor hinder S.

In the decision problem we've used to model **Bounty**, our acts and states are neither probabilistically nor causally independent. Given this fact, causal and evidential decision theorists will agree that dominance reasoning shouldn't be sanctioned in this case.

In the case of $\mathbf{D}^1_Y$, the decision problem we've used to model Yuko's epistemic situation, one should certainly reject the appeal to dominance reasoning if one is a causal decision theorist. For it's clear that the acts and states in this decision problem are not causally independent. Recall that the possible states are truth-value distributions for the proposition expressed by $(*)$ and its negation, while the acts are possible credence distributions in these propositions. Since which state is actual depends on what Yuko's credence in $(*)$ is, Yuko's acts will causally influence which state obtains. It follows that if one is a causal decision theorist, then one should reject the appeal to DOMINANCE 2 in (3c).

In the case of evidential decision theory, matters are a bit more subtle, since, in order to know whether we can apply dominance reasoning, we need to make some assumptions about the agent's credal state. We can show, however, that in a large class of reasonable cases the appeal to dominance reasoning will be illicit by the lights of evidential decision theory. And the reason for this is that there are a large number of reasonable credal states that Yuko could have that would make the acts and states in $\mathbf{D}^1_Y$ probabilistically dependent.

For example, assume that Yuko is aware of the way in which the state of the world is dependent on her credal state. In particular, assume that Yuko's credences are such that: $Cr_y(S_1|A_e) = 1$ and $Cr_y(S_1|A_d) = 0$. Since Yuko can't have both credence 1 and credence 0 in S_1, it follows that the acts and states in this decision problem will not be probabilistically independent. In such cases, then, if one is an evidential decision theorist, one should reject the appeal to DOMINANCE 2 in (3c).

The lesson to be drawn here is the following:

> If we want to model an agent's epistemic position as a decision problem and apply dominance reasoning, we should choose our states so that they are independent of the agent's possible credal states.

There is a simple way of reformulating the decision problem representing Yuko's epistemic situation that let's us address both of the defects in the preceding argument. Instead of representing our states as possible distributions of

[23] See Jeffrey [1983] for the canonical development of evidential decision theory.

[24] For developments of causal decision theories see, e.g., Joyce [1999] and Lewis [1981].

truth-values for the proposition expressed by $(*)$ and its negation, as we did in $\mathbf{D}^1_Y$, we should instead take our states to be **dependence hypotheses**.

A dependence hypothesis is a proposition that states for each possible act, what utility the agent would gain were that act to be performed. Let $[U = u]$ be a proposition specifying that an agent's utility is u. We may think of a dependence hypothesis as a (possibly infinite) conjunction of non-backtracking counterfactuals of the form: $A_i \mathrel{\Box\!\!\rightarrow} [U = u]$, containing exactly one conjunct for each act A_i.

In $\mathbf{D}^1_Y$ there were two states representing the two possible distributions of truth-values for $(*)$ and its negation. If, instead, we carve up the space of possible worlds by grouping together worlds that make true the same counterfactuals connecting credal states and epistemic utilities, then there will only be one state in our decision problem. For both of our possible worlds w_1 and w_2 agree about which world would be actual were Yuko to have a particular credal state. For example, both w_1 and w_2 agree that were Yuko to have credal state e, w_1 would be actual. Thus both both w_1 and w_2 will agree about what epistemic value Yuko would have were she to have a particular credal state.

Instead of representing Yuko's epistemic situation by the decision problem $\mathbf{D}^1_Y$, we should represent it by the following alternative decision problem, $\mathbf{D}^2_Y$. Let S^2_y be the dependency hypothesis specifying how Yuko's epistemic utility counterfactually depends on her credal state. We let $\mathbf{S}^2_Y$ be the singleton set consisting of S^2_y. We let $\mathbf{A}^2_Y$ be the set of possible credence distributions that Yuko can have in the proposition expressed by $(*)$ and its negation. Finally, we let $\mathbf{U}^2_Y(A \wedge S^2_y) = u \leftrightarrow (S^2_y \models A \mathrel{\Box\!\!\rightarrow} [U = u])$, i.e., i.e., just in case $A \mathrel{\Box\!\!\rightarrow} [U = u]$ is one of the conjuncts of S^2_y.

Note that the members of $\mathbf{A}^2_Y$ are all compatible with S^2_y, and are all causally independent of S^2_y. This is guaranteed, since S^2_y is true in every possible world.

Moreover, S^2_y will be probabilistically independent of each member of $\mathbf{A}$, given the assumption that Yuko gives credence 1 to S^2_y both unconditionally and conditional on each member of $\mathbf{A}$. Given that S^2_y is true in every possible world, it's reasonable to assume that Yuko is not rationally precluded from having a credal state that satisfies this constraint. We'll assume that Yuko's credal state does satisfy this constraint.

Since the acts and states in $\mathbf{D}^2_Y$ are independent (in either of the relevant senses), we can apply dominance reasoning to this decision problem to draw conclusions about what sort of credal state Yuko ought to have. And what dominance reasoning tells us here is that Yuko ought to be probabilistically incoherent.

To see this, recall that in the previous section we showed that the credal state represented by d maximizes accuracy in the following sense: were Yuko to have any other credal state, she would be less accurate than she would be if she were to have the credal state represented by d. There is, then, a non-backtracking counterfactual $A_d \mathrel{\Box\!\!\rightarrow} [U = u]$, such that $S^2_y \models A_d \mathrel{\Box\!\!\rightarrow} [U = u]$, and such that for any other $A \in \mathbf{A}^2_Y$, if $S^2_y \models A \mathrel{\Box\!\!\rightarrow} [U = u*]$, then $u* < u$. It follows that for every $A \in \mathbf{A}^2_Y$ such that $A \neq A_d$, $\mathbf{U}^2_Y(A_d \wedge S^2_y) > \mathbf{U}^2_Y(A \wedge S^2_y)$. By DOMINANCE

1, then, it follows that Yuko ought to have the credal state represented by A_d. And since A_d is probabilistically incoherent, it follows that Yuko ought to be probabilistically incoherent.

5 How Far Does The Argument Extend?

I've argued that there are cases in which an agent ought to have probabilistically incoherent credences—at least assuming that credal accuracy is our primary epistemic goal. Instead of supporting PROBABILISM, considerations of credal accuracy give us to good reason to reject this principle.

The case that we focused on, however, is unusual. This case involves a proposition ϕ such that necessarily: ϕ is true just in case a certain agent's credence in ϕ is less than a particular value. Most propositions aren't like this. Most propositions are such that their truth-values aren't tied in this way to what our credences are in those propositions. This raises the question whether there is an interesting restricted version of PROBABILISM that we may still endorse. Is there some large algebra $\mathcal{P}$ such that, if we restrict our attention to $\mathcal{P}$, it is true that a rational agent must always have probabilistically coherent credences in those propositions?

I'll argue that the answer to this question is: no. For almost any proposition ϕ, there is some possible situation in which an agent's credences in ϕ and $\neg\phi$ may be rationally probabilistically incoherent. As a matter of fact, the conditions that allow for this are, I think, extremely rare. For actual agents, then, there will be a large algebra—perhaps the whole algebra over which their credences are defined—such that it is true that those agents ought to have probabilistically coherent credences in those propositions. But this is a contingent fact.

We've seen that an agent's epistemic situation with respect to an algebra of propositions $\mathcal{P}$ can be represented as a decision problem $\mathbf{D}$ in which the set of states $\mathbf{S}$ consist of maximal specifications of how the agent's epistemic utility counterfactually depends on her credences in the members of $\mathcal{P}$. Given such a decision problem, we can appeal to principles of rational decision making, such as DOMINANCE 1 and DOMINANCE 2, to argue that certain credal states, or sets of credal states, are rationally obligatory. Dominance reasoning, however, can only be applied to a limited range of decision problems. Where there is no act that dominates all its competitors, DOMINANCE 1 falls silent. Where there is no set of acts that act-dominates all its competitors, DOMINANCE 2 falls silent. To determine the rational act(s) in these cases we need a more general principle of rational decision making.

Call a decision problem **proper**, if $\mathbf{S}$ is a set of dependency hypotheses and $\mathbf{C}$ is a credence function that is probabilistically coherent over the smallest algebra containing every proposition of the form $A_i \wedge S_j$. Given a proper decision problem $\mathbf{D}$, we can define the **causal expected utility** of an act A, $U_C(A)$, as follows:

$$U_C(A) =_{df} \sum_{S \in \mathbf{S}} \mathbf{C}(S) U(A \wedge S)$$

We can further define the **evidential expected utility** of an act A, $U_E(A)$, as follows:

$$U_E(A) =_{df} \sum_{S \in \mathbf{S}} \mathbf{C}(S|A)U(A \wedge S)$$

We can now formulate two more putative principles of rational decision making:

CUP If $\mathbf{C}$ is a rational credal state and $U_C(A_1) > U_C(A_2)$, then A_1 is rationally preferable to A_2.

EUP If $\mathbf{C}$ is a rational credal state and $U_E(A_1) > U_E(A_2)$, then A_1 is rationally preferable to A_2.

Causal decision theorists will endorse CUP, while evidential decision theorists will endorse EUP. Since evidential and causal expected utilities can come apart, CUP and EUP will sometimes give contradictory verdicts. In what follows, however, we can remain neutral on the question of which of these two principles we should endorse.

I'll now argue, by appeal to CUP and EUP, that even in cases in which there is no necessary connection between the truth of a proposition and an agent's credence in that proposition, an agent may be rationally probabilistically incoherent. Even for such propositions, there are cases in which, by a rational agent's own lights, accuracy is better achieved by being probabilistically incoherent than probabilistically coherent.

Consider the proposition that Yuko will make a particular free-throw. Name this proposition **FT**. Clearly, there is no necessary connection between the truth-value of FT and Yuko's credence in this proposition.

Consider the smallest algebra containing FT. This consists of $\top$, $\bot$, FT, and $\neg$FT. We'll assume that Yuko has credence 1 in $\top$ and credence 0 in $\bot$. Yuko's possible credal states, then, differ with respect to this algebra just over the credences assigned to FT and $\neg$FT.

We may think of Yuko's epistemic situation regarding this class of propositions as a decision problem $\mathbf{D}^3_Y$. The set of acts, $\mathbf{A}^3_Y$, is the set of possible credences that Yuko could have in FT and $\neg$FT. Unlike with $\mathbf{D}^2_Y$, however, there will be more than one dependence hypothesis stating how Yuko's epistemic utility depends counterfactually on members of $\mathbf{A}^3_Y$. The reason for this is that, unlike with $(*)$, there is no necessary connection between the truth-value of FT and Yuko's credence in FT.

Assume that as a matter of contingent fact Yuko is an extremely accurate free-throw shooter, but only when her credence is less than 0.5 that she will make the free-throw. In particular, assume that the following counterfactual claims hold:

(7) If Yuko were to have credence less than 0.5 in FT, then FT would be true. $[(Cr_y(\text{FT}) < 0.5) \mathrel{\Box\!\!\rightarrow} \text{FT}]$

(8) If Yuko were to have credence greater than or equal to 0.5 in FT, then FT would be false. $[(Cr_y(\text{FT}) \geq 0.5) \mathrel{\Box\!\!\rightarrow} \neg\text{FT}]$

It's easy to see that (7) and (8) together entail, for any $A \in \mathbf{A}^3_Y$, a counterfactual: $A \mathrel{\Box\!\!\rightarrow} [U = u]$, where $[U = u]$ specifies the agent's epistemic utility given the credences specified in A.[25] (7) and (8), thus, jointly entail a particular dependence hypothesis and are jointly incompatible with all other dependence hypotheses. Since they both obtain, they settle which dependence hypothesis obtains. Call this dependence hypothesis S^3_y. According to S^3_y, Yuko's epistemic utility depends on her credences in FT and its negation in exactly the same manner that her epistemic utility depends on her credences in $(*)$ and its negation according to S^2_y.

To show that Yuko may rationally fail to have probabilistically coherent credences in FT and its negation, we make the following assumptions:

(9) Yuko's credence in S^3_y is 0.9.

(10) Yuko's credal state is probabilistically coherent over the smallest algebra containing every proposition of the form $A_i \wedge S_j$, for $A_i \in \mathbf{A}^3_Y$ and $S_j \in \mathbf{S}^3_Y$.

(11) Yuko's credences are such that the members of $\mathbf{A}^3_Y$ and the members of $\mathbf{S}^3_Y$ are probabilistically independent.

(12) The credal profile ascribed in (9)-(11) is rational.

There are numerous credal states that satisfy (9)-(11). Shortly, I will discuss the further assumption that this credal profile may be rational. First, however, I'll show that from (9)-(12) it follows by either CUP or EUP that Yuko may rationally fail to have probabilistically coherent credences in FT and ¬FT.

Let A_d be the member of $\mathbf{A}^3_Y$ according to which Yuko has credence 0.5 in FT and credence 1 in its negation. Let $\mathbf{A}_P$ be the set of probabilistically coherent credences in FT and its negation. We first show:

$\boxed{\text{From (9)-(10), it follows that, for every } A \in \mathbf{A}_P,\ U_C(A_d) > U_C(A).}$

> We assume that $\mathbf{C}$ is a probabilistically coherent credence distribution over the smallest algebra containing every proposition of the form $A_i \wedge S_j$, for $A_i \in \mathbf{A}^3_Y$ and $S_j \in \mathbf{S}^3_Y$.

[25] Justification: For every $A \in \mathbf{A}^3_Y$, either $A \models Cr_y(\text{FT}) < 0.5$ or $A \models Cr_y(\text{FT}) \geq 0.5$. This ensures that $(Cr_y(\text{FT}) < 0.5) \mathrel{\Box\!\!\rightarrow} \text{FT}$ and $(Cr_y(\text{FT}) \geq 0.5) \mathrel{\Box\!\!\rightarrow} \neg\text{FT}$ jointly entail, for every $A \in \mathbf{A}^3_Y$, either $A \mathrel{\Box\!\!\rightarrow} \text{FT}$ or $A \mathrel{\Box\!\!\rightarrow} \neg\text{FT}$. The accuracy of members of $\mathbf{A}^3_Y$ is determined solely by whether or not FT is true or false. We have, then, for every $A \in \mathbf{A}^3_Y$, $A \wedge \text{FT} \models [U = u]$ and $A \wedge \neg\text{FT} \models [U = u*]$, for some propositions $[U = u]$ and $[U = u*]$. It follows, then, from the fact that $(Cr_y(\text{FT}) < 0.5) \mathrel{\Box\!\!\rightarrow} \text{FT}$ and $(Cr_y(\text{FT}) \geq 0.5) \mathrel{\Box\!\!\rightarrow} \neg\text{FT}$ entail, for every $A \in \mathbf{A}^3_Y$, either $A \mathrel{\Box\!\!\rightarrow} \text{FT}$ or $A \mathrel{\Box\!\!\rightarrow} \neg\text{FT}$, that they entail, for every $A \in \mathbf{A}^3_Y$, some proposition of the form: $A \mathrel{\Box\!\!\rightarrow} [U = u]$.

The epistemic utility of having the credal state represented by A_d in state S_y^3 is 0.875. This gives us a lower bound on the causal expected utility of A_d. If $C(S_y^3) = x$, then $U_C(A_d) \geq x(0.875)$.

Amongst the probabilistically coherent credal states, the most accurate state in S_y^3 will be the state represented by point f in our geometric model. If Yuko has this credal state, she will have credence 0.5 in FT and credence 0.5 in its negation. Let A_f be the proposition according to which Yuko has this credal profile. The epistemic utility of having the credal state represented by A_f in state S_y^3 is 0.75. This gives us an upper bound on the causal expected utility for $A \in \mathbf{A}_P$. If $C(S_y^3) = x$, then for any $A \in \mathbf{A}_P$, $U_C(A) \leq x(0.75) + (1-x)1$.

Given this lower bound on the expected utility of A_d, and this upper bound on the expected utility of members of $\mathbf{A}_P$, we can show that there are values of x, such that if $\mathbf{C}(S_y^3) = x$, then for every $A \in \mathbf{A}_P$, $U_C(A_d) > U_C(A)$.

To show this, we first calculate the value for x at which the lower bound for A_d equals the upper bound for members of $\mathbf{A}_P$. To do this we set $x(0.875) = x(0.75) + (1-x)1$, and solve for x. A quick calculation shows that this equality holds when $x = \frac{1}{1.125} \approx 0.89$. It follows that whenever $x > 0.89$, the lower bound for A_d will be greater than the upper bound for the members of $\mathbf{A}_P$. Thus, if $\mathbf{C}(S_y^3) > 0.89$, then for every $A \in \mathbf{A}_P$, $U_C(A_d) > U_C(A)$.

From (9)-(10), it therefore follows that for every $A \in \mathbf{A}_P$, $U_C(A_d) > U_C(A)$.

Given this fact and our assumption that the credal profile ascribed by (9)-(10) is rational, it follows from CUP that A_d is rationally preferable to every $A \in \mathbf{A}_P$. If one endorses CUP, then one should hold that Yuko is not rationally required to have probabilistically coherent credences.

From (11), it follows that for every $A \in \mathbf{A}$, $U_E(A) = U_C(A)$. Thus, we have that $U_E(A_d) > U_E(A)$, for every probabilistically coherent credal state $A \in \mathbf{A}_P$. Given this fact and (12), it follows from EUP that A_d is rationally preferable to every $A \in \mathbf{A}_P$. If one endorses EUP, then one should hold that Yuko is not rationally required to have probabilistically coherent credences.

Nothing in this argument turns on any features of FT that couldn't, at least in principle, be shared by almost any other contingent proposition. Assuming that this argument works, we have, then, a fairly general recipe for generating cases in which an agent may rationally fail to have probabilistically coherent credences in a contingent proposition and its negation.

The one substantive assumption that a proponent of CUP or EUP may question in order to block this argument is (12), i.e, the assumption that the credal profile imposed by (9)-(11) is rational. If one rejects this claim, then one must hold that it is impossible for Yuko to have a credal state satisfying (9)-(11) without being guilty of a rational failure. Why would this be?

There are three options:

- One could hold that Yuko would be irrational in virtue of having a credal state that is probabilistically coherent over the smallest algebra containing every proposition of the form $A_i \wedge S_j$, for $A_i \in \mathbf{A}^3_Y$ and $S_j \in \mathbf{S}^3_Y$.

- One could hold that Yuko would be irrational in virtue of such a credence distribution making the members of $\mathbf{A}^3_Y$ and $\mathbf{S}^3_Y$ probabilistically independent.

- One could hold that Yuko would be irrational in virtue of having a high credence in S^3_y.

I can't see any good reason for endorsing either of the first two options. The third option, however, has been endorsed by some authors. Let me now explain why some have thought that it would be irrational for Yuko to have a high credence in S^3_y and why we shouldn't agree.

Call **an anti-expert** about ϕ one who is reliably mistaken in their judgments about ϕ.[26] S^3_y entails that Yuko is an anti-expert about FT. According to Andy Egan and Adam Elga, it is never rational for an agent to have a high credence in a proposition that entails that she is an anti-expert about some proposition ϕ.[27] If this were right, then our argument for the possibility of rational probabilistic incoherence for propositions such as FT could be blocked by rejecting (12).

Egan and Elga's argument that it is never rational to have a high-credence that one is an anti-expert about some proposition ϕ hinges on the following fact:

ANTI-EXPERTISE If an agent has a high credence that she is an anti-expert about ϕ and, in addition, is sensitive to her own credence in ϕ then she will be probabilistically incoherent.[28]

We've already seen a simple case illustrating this. (1) ascribes to Yuko anti-expertise about the truth of $(*)$. And in §2 we demonstrate that if Yuko has a high credence in (1) and is somewhat sensitive to her own credence in the truth of $(*)$, then she is guaranteed to be probabilistically incoherent.

Egan and Elga endorse PROBABILISM. In addition, they endorse the following principle:

[26]Following Sorensen [1988], we can distinguish two types of anti-expertise. Focus, for the moment, on qualitative beliefs. We say that an agent is a *commissive anti-expert* about the proposition ϕ, just in case either it's the case that $\neg\phi$ and the agent believes ϕ, or it's the case that ϕ and the agent believes $\neg\phi$, i.e., just in case $(\neg\phi \wedge B(\phi)) \vee (\phi \wedge B(\neg\phi))$. We say that an agent is an *omissive anti-expert* just in case either it's the case that $\neg\phi$ and the agent believes ϕ, or it's the case that ϕ and the agent doesn't believe ϕ, i.e., just in case $\phi \leftrightarrow \neg B\phi$. If we switch to talking about credences, we can then distinguish varying degrees of commissive and omissive anti-expertise.

[27]See Elga and Egan [2005]. This same claim is defended in the case of qualitative belief in Sorensen [1988].

[28]See Elga and Egan [2005] for the general argument.

RATIONAL INTROSPECTION A rational agent must be responsive to its own credal state.

PROBABILISM, RATIONAL INTROSPECTION and ANTI-EXPERTISE together entail that it is rationally impermissible for an agent to have high credence in her own anti-expertise.[29] Thus, according to Egan and Elga: "It is never rational to count oneself as an anti-expert because doing so must involve either [probabilisitic] incoherence or poor access to one's own beliefs."[30]

The conclusion of this argument is, it must be admitted, quite surprising. In the case of (1), we noted that as long as Yuko is aware of which sentence $(*)$ refers to, it would seem, prima facie, that she should be able to have a high rational credence in this proposition. After all, it's obvious that (1) is true, given which sentence $(*)$ refers to. Similarly, in the case of S_y^3. As we set up the case, this proposition is true. Moreover, we could assume that Yuko has excellent evidence for its truth. Perhaps she has been the subject of extensive testing, and it has been determined that every time she has shot a free-throw and was at least 0.5 confident that she would make it, she has missed, and that every time she has shot a free-throw and has been less than 0.5 confident that she would make it, she has made it. Given enough evidence of this type, it seems hard to deny that it could be rational for Yuko to be highly confident in the truth of S_y^3.

Of course, these prima facie considerations could be outweighed if we had strong reason to endorse both PROBABILISM and RATIONAL INTROSPECTION. However, we've already seen that there is good reason to reject PROBABILISM. Given this, it seems gratuitous to hold, despite its prima facie implausibility, that Yuko can never rationally have high credence in S_y^3.

There is another reason to reject Egan and Elga's argument which is, I think, of independent interest. Egan and Elga infer from the fact that PROBABILISM and RATIONAL INTROSPECTION are jointly incompatible with the claim that it is rational to believe that one is an anti-expert that it must be irrational to believe that one is an anti-expert. There is, however, good reason to think that PROBABILISM and RATIONAL INTROSPECTION are themselves jointly unacceptable. Even if one is not antecedently convinced of the falsity of PROBABILISM, one should still not be persuaded by Egan and Elga's argument.

Here's why we shouldn't accept both PROBABILISM and RATIONAL INTROSPECTION. We can show that if PROBABILISM is true, then in certain cases it is rationally required that an agent have *poor* introspective access to its credences. (In a moment, we'll see how this works in detail.) Given this, if we were to endorse both PROBABILISM and RATIONAL INTROSPECTION, then we would be committed to the existence of a rational dilemma. In particular, we would be committed both to the claim that rationality requires of a certain agent that the agent have good access to its credences and that the agent have poor access to its own credences. However, the following is a plausible general constraint on principles of

[29] This follows given a plausible multi-premiss closure principle for rational obligations: $\phi, \psi \models \gamma \Rightarrow O\phi, O\psi \models O\gamma$.

[30] Elga and Egan [2005] p. 83.

rationality:

OUGHT-CAN It must always be possible for an agent to meet the requirements imposed by rationality.

Since there are situations in which it is impossible to meet all of the requirements imposed by PROBABILISM and RATIONAL INTROSPECTION, it follows, given OUGHT-CAN, that we shouldn't endorse both of these principles.

Of course, one might simply bite the bullet here and accept that an agent may sometimes be faced with a rational dilemma. But this seems to me to be poorly motivated. I think we do better if we let OUGHT-CAN guide our judgments in this case, and infer that PROBABILISM and RATIONAL INTROSPECTION aren't both correct.

To see how an agent may be doomed to probabilistic incoherence just given a moderate sensitivity to its own credal state, let us consider another agent who we'll call 'Hiro'.

Let (#) name the following sentence:

> **Hiro's credence in the proposition expressed by (#) isn't greater than or equal to 0.5.**

We'll use 'Cr_h' to abbreviate 'Hiro's credence in...' and we'll use 'ρ' to abbreviate 'the proposition expressed by'. The above can, then, be represented as:

(#) $\neg Cr_h\rho(\#) \geq 0.5$

Note that since both '(#)' and ' '$\neg Cr_h\rho(\#) \geq 0.5$' ' refer to the same sentence, we have:

(13) $\rho(\#) = \rho$'$\neg Cr_h\rho(\#) \geq 0.5$'

We'll assume the following facts about Hiro's introspective powers. We'll assume that if Hiro has credence greater than or equal to 0.5 in the proposition expressed by (#), then Hiro has credence greater than 0.5 in the proposition that he has credence greater than or equal to 0.5 in the proposition expressed by (#). We'll also assume that if Hiro does not have credence greater than or equal to 0.5 in the proposition expressed by (#), then Hiro has credence greater than 0.5 in the proposition that he does not have credence greater than or equal to 0.5 in the proposition expressed by (#). We can represent these assumptions as follows:

(14) $[Cr_h\rho(\#) \geq 0.5] \rightarrow [Cr_h(\rho$'$Cr_h\rho(\#) \geq 0.5$') $> 0.5]$

(15) $[\neg Cr_h\rho(\#) \geq 0.5] \rightarrow [Cr_h(\rho$'$\neg Cr_h\rho(\#) \geq 0.5$') $> 0.5]$

We can show:

$\boxed{\text{From (13) - (15), it follows that Hiro is probabilistically incoherent.}}$

Assume: $\neg Cr_h\rho(\#) \geq 0.5$. By (13), we can substitute ‘ ρ‘$\neg Cr_h\rho(\#) \geq 0.5$’ ’ for ‘$\rho(\#)$’ within attitude ascriptions *salva veritate*. Thus, from our assumption and (13), we have: $\neg Cr_h(\rho$‘$\neg Cr_h\rho(\#) \geq 0.5$’) > 0.5. But from our assumption, it follows, given (15), that we have: $Cr_h(\rho$‘$\neg Cr_h\rho(\#) \geq 0.5$’) > 0.5.

Since the assumption that $\neg Cr_h\rho(\#) \geq 0.5$ leads to a contradiction, it follows, given (13)-(15), that $Cr_h\rho(\#) \geq 0.5$, i.e., that Hiro has credence at least as great as 0.5 in the proposition expressed by (#). But now we can show that Hiro is doomed to probabilistic incoherence.

For probabilistic coherence requires that Hiro's credence in the proposition expressed by (#) and his credence in its negation sum to one, i.e., that $Cr_h(\rho$‘$\neg Cr_h\rho(\#) \geq 0.5$’) $+ Cr_h(\rho$‘$Cr_h\rho(\#) \geq 0.5$’) $= 1$. But given that Hiro has credence at least as great as 0.5 in the proposition expressed by (#), we can show that it follows that $Cr_h(\rho$‘$\neg Cr_h\rho(\#) \geq 0.5$’) $+ Cr_h(\rho$‘$Cr_h\rho(\#) \geq 0.5$’) > 1.

From $Cr_h\rho(\#) \geq 0.5$, it follows, given (13), that: $Cr_h(\rho$‘$\neg Cr_h\rho(\#) \geq 0.5$’) > 0.5. But from $Cr_h\rho(\#) \geq 0.5$ and (15) it follows that: $Cr_h(\rho$‘$Cr_h\rho(\#) \geq 0.5$’) > 0.5. Thus, we have: $Cr_h(\rho$‘$\neg Cr_h\rho(\#) \geq 0.5$’) $+ Cr_h(\rho$‘$Cr_h\rho(\#) \geq 0.5$’) > 1.

We have seen that Hiro will satisfy the requirements imposed by PROBABILISM only if either (14) or (15) fail to hold. If PROBABILISM is true, it follows that it is a requirement of rationality that Hiro be such that either (i) he has credence greater than or equal to 0.5 in the proposition expressed by (#), but has at best 0.5 credence, i.e., is at best agnostic, that his credence in this proposition is in this range, or (ii) he fails to have credence greater than or equal to 0.5 in the proposition expressed by (#), but has at best 0.5 credence, i.e., is at best agnostic, that his credence in this proposition fails to be in this range. PROBABILISM, thus, demands that Hiro be insensitive to his own credal state.

One natural worry about this case is the appeal to propositions. Why should we assume that (#) does in fact express a proposition that could serve as the object of Hiro's doxastic attitudes? I take it that the worry here stems from the self-referential nature of (#). In response, let me say the following.

First, we should fix on some diagnostic tests for whether a sentence ϕ expresses a proposition. I take it that a sufficient condition for ϕ to express a proposition is if ϕ can be embedded under metaphysical or doxastic operators in a way that results in a true sentence. For the resultant sentence could be true only if it expressed a proposition; and such a sentence could express a proposition only if its component sentences expressed propositions. A sentence's failure to express a proposition is something that infects any sentence of which it is a part.

Given this, we can show that a sentence is not, in general, precluded from expressing a proposition in virtue of the fact that it contains a term that purports to refer to the proposition expressed by that sentence.

One way to achieve sentential self-reference is via stipulation, as in the case of (#). Another is via a definite description that picks out the sentence in which the definite description occurs. Imagine, for example, that in room 301 there is a single blackboard, and on that blackboard is written the following sentence: 'The proposition expressed by the sentence on the blackboard in room 301 is not true.' In this case, the definite description: 'the sentence on the blackboard in room 301', refers to the very sentence of which that definite description is a constituent. And so the definite description: 'the proposition expressed by the sentence on the blackboard in room 301' purports to refer to the proposition expressed by that sentence.

To argue that this sentence does indeed express a proposition it suffices to argue that this sentence can embed under metaphysical and doxastic operators in a way that results in a true sentence.

It seems fairly obvious that this sentence can embed under doxastic operators and yield a true sentence. For example, let John be someone who believes that there is just one sentence written on the blackboard in 301 and that that sentence is: '$2 + 2 = 5$'. Let John further believe that the proposition expressed by the sentence written on the blackboard in 301 is the proposition that $2 + 2 = 5$ and that this is not true. Given these beliefs it would seem that John believes that the proposition expressed by the sentence written on the blackboard in 301 is not true. It would seem, then, that we can perfectly well embed: 'The proposition expressed by the sentence on the blackboard in room 301 is not true.', under the operator 'John believes that...' and get a true sentence. But if that's the case then it must be that 'The proposition expressed by the sentence on the blackboard in room 301 is not true.' expresses a proposition.

Similarly, it seems clear that this sentence can embed under metaphysical modal operators and produce a true sentence. Consider a possible world in which the sentence written on the blackboard in 301 is '$2 + 2 = 5$'. We assume that in this world the proposition expressed by the sentence written on the blackboard in 301 is just the proposition that $2 + 2 = 5$. In this world, then, the proposition expressed by the sentence written on the blackboard in room 301 is not true. But then it follows that it is possible that the proposition expressed by the sentence written on the blackboard in room 301 is not true.[31] It would seem, then, that we can embed: 'The proposition expressed by the sentence on the blackboard in room 301 is not true.', under the operator 'It is possible that...' and get a true sentence. And if that's the case, then 'The proposition expressed by the sentence on the blackboard in room 301 is not true.' must express a proposition.

The above reflections show that a sentence is not barred from expressing a proposition simply in virtue of containing a term that purports to refer to the proposition expressed by that sentence. The fact that (#) contains such a term

[31] Of course, this is only true on the *de dicto* reading of the above sentence.

does not, then, provide us with good reason to deny that this sentence expresses a proposition. In the absence a more convincing reason to the contrary, I think we should, therefore, assume that (#) does indeed express a proposition that can serve as the object of Hiro's doxastic attitudes.

Given that (#) expresses a proposition, we can see that there are going to be cases in which an agent with moderate introspective powers is doomed to probabilistic incoherence. For this reason, we shouldn't accept both PROBABILISM and RATIONAL INTROSPECTION.

We've seen two reasons that we shouldn't accept Egan and Elga's argument that it is irrational for an agent to self-ascribe anti-expertise. Although one could try to block our earlier argument for the claim it is rational for Yuko to have probabilistically incoherent credences in FT and its negation by maintaining that it is irrational for Yuko to have a high credence in S_y^3, this has been shown to be poorly motivated.

Even for a contingent proposition such as FT, then, there can be situations in which, by a rational agent's own lights, accuracy isn't maximized by having probabilistically coherent credences. Assuming that an agent ought to try to make her credences as accurate as possible, in such cases an agent may rationally fail to have probabilistically incoherent credences.

PROBABILISM, then, doesn't just fail when we look at propositions such as that expressed by ($*$) that concern an agent's own credences. In principle, almost any proposition could be such that an agent could rationally fail to have credences in that proposition and its negation that sum to 1.

Of course, in order for this argument to apply in a particular case, an agent must rationally have high confidence that there is a certain counterfactual connection between her having a low credence in a particular proposition and the proposition being true. Such counterfactual connections are rare, as are cases in which an agent has evidence supporting such connections. As a matter of contingent fact, then, I think it is plausible that for any actual agent there will be some large algebra of propositions $\mathcal{P}$, such that the agent's credences over this set of propositions ought to be probabilistically coherent. But things could have been otherwise.

6 Conclusion

I began this paper by showing that PROBABILISM has a surprising consequence. In certain cases, PROBABILISM demands that an agent either be insensitive to her own credal state or be ignorant of an obvious truth. Looking more closely at this case, we have seen that it provides us with the material to mount a strong case against PROBABILISM.

A *prima facie* compelling argument for PROBABILISM claims that probabilistic coherence is rationally required because it serves the goal of representing the world as accurately as possible. I've argued that the central premiss of this argument is false. In certain cases, credal accuracy is best served by being probabilistically

incoherent. Considerations of accuracy, instead of providing us with a reason to accept PROBABILISM, provide us with a reason to reject this principle.

References

Kenny Easwaran and Branden Fitelson. An "evidentialist" worry about Joyce's argument for probabilism. *Dialectica*, forthcoming.

Adam Elga and Andy Egan. I can't believe I'm stupid. *Philosophical Perspectives*, 19, 2005.

Hartry Field. *Saving Truth from Paradox*. Oxford University Press, 2008.

Richard Jeffrey. *The Logic of Decision*. Chicago University Press, 2nd edition, 1983.

James Joyce. A non-pragmatic vindication of probabilism. *Philosophy of Science*, 65:575–603, 1998.

James Joyce. *The Foundations of Causal Decision Theory*. Cambridge University Press, 1999.

James Joyce. Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber and C. Schmidt-Petri, editors, *Degrees of Belief*. Synthese Library, 2009.

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716, 1975.

Hannes Leitgeb and Richard Pettigrew. An objective justification of bayesianism I: Measuring inaccuracy. *Philosophy of Science*, 77, 2010a.

Hannes Leitgeb and Richard Pettigrew. An objective justification of bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77, 2010b.

David Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59: 5–30, 1981.

Richard Pettigrew. An improper introduction to epistemic utility theory. In Henk de Regt, Stephan Hartmann, and Samir Okasha, editors, *EPSA Philosophy of Science: Amsterdam 2009*. Springer, 2011a.

Richard Pettigrew. Epistemic utility arguments for probabilism. *Stanford Encyclopedia of Philosophy*, 2011b.

Roy Sorensen. *Blindspots*. Oxford University Press, 1988.

J.R.G. Williams. Generalized probabilism: dutch boks and accuracy domination. *Journal of Philosophical Logic*, forthcoming.