ACCURACY AND THE IMPS

James M. JOYCE, Brian WEATHERSON

ABSTRACT: Recently several authors have argued that accuracy-first epistemology ends up licensing problematic epistemic bribes. They charge that it is better, given the accuracy-first approach, to deliberately form one false belief if this will lead to forming many other true beliefs. We argue that this is not a consequence of the accuracy-first view. If one forms one false belief and a number of other true beliefs, then one is committed to many other false propositions, e.g., the conjunction of that false belief with any of the true beliefs. Once we properly account for all the falsehoods that are adopted by the person who takes the bribe, it turns out that the bribe does not increase accuracy.

KEYWORDS: accuracy, epistemic consequentialism, scoring rules

1.Accuracy, Bribes and Scoring Rules

Belief aims at the truth.[1] So at least in some sense, an agent is doing better at believing the closer they are to the truth. When applied to individual beliefs, this generates epistemic advice that is literally platitudinous: if you know that a change in your attitude towards *p* will make your attitude towards *p* more accurate, make that change! When applied to collective bodies of belief though, the advice turns out to be more contentious. Call **epistemic consequentialism** the view that if an agent knows that a change in their overall belief state will make their belief state more accurate, they should make that change, if they have the power to do so.

Hilary Greaves has recently argued that epistemic consequentialism is false because it licences certain epistemic 'bribes', and these should not be licenced.[2] We'll argue that the best forms of epistemic consequentialism do not licence some of these bribes after all.[3] Here is the key case Greaves uses.[4]

[1] Thanks to Alejandro Pérez Carballo, Richard Pettigrew, and the participants in the Arché Epistemology Seminar for helpful comments.

[2] Hilary Greaves, "Epistemic Decision Theory," *Mind* 122 (2013): 915–952, https://doi:10.1093/mind/fzt090.

[3] Though they do licence others; see section 2.4 for more discussion.

[4] Greaves has four other cases, but the Imps case is the only one that is a problem for all forms of consequentialism she discusses. Similar cases have suggested by Selim Berker and C. S. Jenkins, but we'll focus on Greaves's discussion since she engages more fully with the literature on scoring rules. We'll return briefly to Berker's discussion in section 2. Berker's version is in his "Epistemic Teleology and the Separateness of Propositions," *Philosophical Review* 122 (2013):

> Emily is taking a walk through the Garden of Epistemic Imps. A child plays on the grass in front of her. In a nearby summerhouse are n further children, each of whom may or may not come out to play in a minute. They are able to read Emily's mind, and their algorithm for deciding whether to play outdoors is as follows. If she forms degree of belief 0 that there is now a child before her, they will come out to play. If she forms degree of belief 1 that there is a child before her, they will roll a fair die, and come out to play iff the outcome is an even number. More generally, the summerhouse children will play with chance $(1 - \frac{q(C_0)}{2})$, where $q(C_0)$ is the degree of belief Emily adopts in the proposition C_0 that there is now a child before her. Emily's epistemic decision is the choice of credences in the proposition C_0 that there is now a child before her, and, for each $j = 1, \ldots, n$ the proposition C_j that the jth summerhouse child will be outdoors in a few minutes' time.
>
> ...if Emily can just persuade herself to ignore her evidence for C_0, and adopt (at the other extreme) credence 0 in C_0, then, by adopting degree of belief 1 in each $C_j (j = 1, \ldots, 10)$, she can guarantee a perfect match to the remaining truths. Is it epistemically rational to accept this 'epistemic bribe'?[5]

The epistemic consequentialist says that it is best to have credences that are as accurate as possible. We will focus on believers who assign probabilistically coherent credences (degrees of belief) to the propositions in some "target set" $\mathcal{X}$, and we will think of the "degree of fit" between her beliefs and the truth as being measured by a strictly proper scoring rule. This is a function $\mathbf{I}_{\mathcal{X}}$ which associates each pair $\langle \mathbf{cred}, @ \rangle$ consisting of a credence function **cred** whose domain includes $\mathcal{X}$ and a consistent truth-value assignment @ for elements of $\mathcal{X}$ with a non-negative real number $\mathbf{I}_{\mathcal{X}}(@, \mathbf{cred})$. Intuitively, $\mathbf{I}_{\mathcal{X}}$ measures the inaccuracy of the credences that cred assigns to the propositions in $\mathcal{X}$ when their truth-values are as described by @. Note that higher $\mathbf{I}_{\mathcal{X}}$-values indicate higher levels of epistemic disutility, so that lower is better from a consequentialist perspective. One popular scoring rule is the Brier score, which identifies inaccuracy with the average squared distance between credences and truth-values. (Greaves calls this the 'quadratic scoring rule', which is a useful description too.) More formally, we have:

$$\mathbf{Brier}_{\mathcal{X}}(@, \mathbf{cred}) = \frac{1}{|\mathcal{X}|} \sum_{X \in \mathcal{X}} (\mathbf{cred}(X) - @(X))^2$$

337–393, http://doi.org/10.1215/00318108-2087645, and "The Rejection of Epistemic Consequentialism," *Philosophical Issues* 23 (2013): 363–387. Jenkins's version is in her "Entitlement and Rationality," *Synthese* 157 (2007): 25–45, http://doi.org/10.1007/s11229-006-0012-2.

[5] Greaves, "Epistemic Decision Theory," 918.

where$|\mathcal{X}|$ is the number of propositions in $\mathcal{X}$ and $@(X)$ is either zero or one depending upon whether X is true or false.

Another common score is the logarithmic rule, which defines inaccuracy as:

$$\mathbf{Log}_{\mathcal{X}}(@, \mathbf{cred}) = \frac{1}{|\mathcal{X}|}\sum_{X\in\mathcal{X}} -\log(\mathbf{cred}(X)) \cdot @(X)$$

For now we will follow Greaves in assuming that our epistemic consequentialist uses the Brier score to measure epistemic disutility, but we will relax that assumption in a little while.

Now let's think about the 'bribe' that Greaves offers, from the point of view of the epistemic consequentialist. The choices are to have one of two credal states, which we'll call **cred1** and **cred2**. We'll say **cred1** is the one that best tracks the initial evidence, so $\mathbf{cred1}(C_0) = 1$, and $\mathbf{cred1}(C_i) = 0.5$ for $i \in 1,\ldots,10$. And **cred2** is the credence Emily adopts if she accepts the bribe, so $\mathbf{cred2}(C_0) = 0$, while $\mathbf{cred2}(C_i) = 1$ for $i \in 1,\ldots,10$. Which state is better?

Thinking like an epistemic consequentialist, you might ask which state is more accurate? It seems like that would be **cred2**. While **cred1** gets C_0 exactly right it does not do very well on the other propositions. In contrast, while **cred2** gets C_0 exactly wrong, it is perfect on the other ten propositions. So overall, **cred2** looks to have better epistemic consequences: when compared to being right about one proposition and off by 0.5 on ten others, being right on ten is surely worth one false belief. The Brier score seems to bear this out. If we let $\mathcal{X}$, the target set, consist of $C_0, C_1, \ldots, C_{10}$, then we have

$$\mathbf{Brier}_{\mathcal{X}}(\mathbf{cred1}, @) = \frac{1}{11}[(1 - \mathbf{cred1}(C_0))^2 + \sum_{i=1}^{10}(@(C_i) - \frac{1}{2})^2] = \frac{10}{44}$$

$$\mathbf{Brier}_{\mathcal{X}}(\mathbf{cred2}, @) = \frac{1}{11}[(1 - \mathbf{cred2}(C_0))^2 + \sum_{i=1}^{10}(@(C_i) - cred(C_i))^2] = \frac{1}{11}$$

So, it seems that a good epistemic consequentialist will take the bribe. But, doesn't that seem like the height of epistemic irresponsibility? It means choosing to believe that C_0 is certainly false when you have conclusive evidence for thinking that it is true. If you see the child on the lawn in front of you, how can you sanction believing she is not there?

As Greaves admits, intuitions are divided here. Some consequentialists might think that "epistemic bribes" are at least sometimes worth taking, while those of a more deontological bent will always find such trade-offs "beyond the pale."[6] We

[6] Berker, "Epistemic Teleology," 363.

will largely sidestep these contentious issues here, though our argument will offer comfort to epistemic consequentialists who feel queasy about accepting the bribe offered in Imps. We contend that, when inaccuracy is measured properly, the consequences of adopting the **cred2** credences are strictly worse than the consequences of adopting **cred1**.

The basic problem is that Imps cherry-picks propositions in a way no consequentialist should condone. Its persuasive force rests on the assumption that, for purposes of epistemic evaluation, nothing matters except the accuracies of the credences assigned to propositions in the target set $\mathcal{X}$. But $\mathcal{X}$ is the wrong target! By confining attention to it Greaves ignores the many other credences to which Emily becomes committed as a consequence of adopting **cred1** or **cred2**. Any (coherent) agent who invests credence zero in C_0 must also invest credence zero in any proposition $C_0 \wedge Y$, where Y is any conjunction or disjunction of elements from $\mathcal{X}$. Likewise, anyone who invests credence one in C_n must invest credence one in any proposition $C_n \vee Y$, where Y is any conjunction or disjunction from $\mathcal{X}$. In the current context (where the probabilities of the various C_i are independent), when Emily adopts a credence function over $\mathcal{X}$ she commits to having a credence for (i) every atomic proposition $\pm C_0 \wedge \pm\ C_1 \wedge \pm C_2 \wedge \ldots \wedge \pm C_{10}$, where '±' can be either an affirmation or a negation, and (ii) every disjunction of these atomic propositions. In short, she commits to having credences over the whole Boolean algebra $\mathcal{A}_{\mathcal{X}}$ generated by $\mathcal{X}$. Since each event of a child coming out is independent, adopting **cred1** will commit her to setting $\textbf{cred1}(\pm C_0 \wedge \pm\ C_1 \wedge \pm C_2 \wedge \ldots \wedge \pm C_{10}) = \frac{1}{1024}$ when C_0 is affirmed, and 0 when it is negated. While adopting **cred2** commits her to setting $\textbf{cred2}(\pm C_0 \wedge \pm\ C_1 \wedge \pm C_2 \wedge \ldots \wedge \pm C_{10})$ equal to 1 when C_0 is negated and the rest of the C_i are affirmed, and equal to 0 otherwise. In this way, each of these probability assignments over the 2048 atoms determine a definite probability for every one of the 2^{2048} propositions in $\mathcal{A}_{\mathcal{X}}$.

It is our view that consequentialists should reject any assessment of epistemic utility that fails to take the accuracies of *all* these credences into account. All are consequences of adopting **cred1** or **cred2**, and so all should be part of any consequentialist evaluation of the quality of those credal states. The right "target set" to use when computing epistemic disutility is not $\mathcal{X}$ but $\mathcal{A}_{\mathcal{X}}$. If we don't do that, we ignore most of the ways in which **cred1** and **cred2** differ in accuracy. If Emily takes the bribe, she goes from having credence 0.5 in $C_0 \leftrightarrow C_1$ to having credence 0 in it. And that's unfortunate, because the chance of $C_0 \leftrightarrow C_1$ goes from 0.5 to 1. This is another proposition, as well as C_0, that Emily acquires a false belief in by taking the bribe. Of course, there are other propositions not counted that go the other way. Originally, Emily has a credence of 0.25 in $C_1 \wedge C_2$, and its chance is

also 0.25. After taking the bribe, this has a chance of 1, and her credence in it is 1. That's an improvement in accuracy. So there are a host of both improvements and deteriorations that are as yet unaccounted for. We should account for them, and making the target set be $\mathcal{A}_X$ does that.

When seen from this broader perspective, it turns out the seeming superiority of **cred2** over **cred1** evaporates. The rest of this section (and the appendix) is dedicated to demonstrating this. We'll make the calculations a little easier on ourselves by relying on a theorem concerning Brier scores for coherent agents. Assume, as is the case here, that Emily's credences are defined over an atomic Boolean alegbra of propositions. The atoms are the 'worlds', or states that are maximially specific with respect to the puzzle at hand. In this case there are 2048 states, which we'll label $s_0$through s_{2047}. In s_k, the first child is on the lawn iff $k \leq 1023$, and summerhouse child i comes out iff the $(i + 1)$th digit in the binary expansion of k is 1. Let $\mathcal{S}_X$ be the set of all these states. That's not a terrible target set; as long as Emily is probabilistically coherent it is comprehensive. The theorem in question says that for any credence function **cred** defined over a partition of states $\mathcal{S}$, and over the algebra $\mathcal{A}$ generated by those states,

Theorem-1

$$\mathbf{Brier}_{\mathcal{A}}(\mathbf{cred}, @) = \frac{|\mathcal{S}|}{4}\mathbf{Brier}_{\mathcal{S}}(\mathbf{cred}, @)$$

(The proof of this is in the appendix.) So whichever credence function is more accurate with respect to $\mathcal{S}_X$ will be more accurate with respect to $\mathcal{A}_X$. So let's just work out $\mathbf{Brier}_{\mathcal{S}_X}$ for **cred1** and **cred2** at the actual world.

First, **cred1** will appropriately assign credence 0 to each s_k ($k \in 0,\ldots,1023$). Then it assigns credence $\frac{1}{1024}$ to every other s_k. For 1023 of these, that is off by $\frac{1}{1024}$, contributing $\frac{1}{2^{20}}$ to the Brier score. And for 1 of them, namely @, it is off by $\frac{1023}{1024}$, contributing $\frac{1023^2}{2^{20}}$. So we get:

$$\begin{aligned}
\mathbf{Brier}_{\mathcal{S}_X}(\mathbf{cred1}, @) &= \frac{1}{2048}[1024 \cdot 0 + 1023 \cdot \frac{1}{2^{20}} + \frac{1023^2}{2^{20}}] \\
&= \frac{1}{2048} \cdot \frac{1023 + 1023^2}{2^{20}} \\
&= \frac{1}{2048} \cdot \frac{1023 \cdot 1024}{2^{20}} \\
&= \frac{1}{2048} \cdot \frac{1023}{1024} \\
&= \frac{2^{10} - 1}{2^{21}}
\end{aligned}$$

It's a bit easier to work out $\mathbf{Brier}_{\mathcal{S}_\mathcal{X}}(\mathbf{cred2}, s_{2047})$. (We only need to work out the Brier score for that state, because by the setup of the problem, Emily knows that's the state she'll be in if she adopts **cred2**). There are 2048 elements in $\mathcal{S}_\mathcal{X}$. And **cred2** assigns the perfectly accurate credence to 2046 of them, and is perfectly inaccurate on 2, namely s_{1023}, which it assigns credence 1, and s_{2047} which it assigns credence 0. So we have

$$\begin{aligned}\mathbf{Brier}_{\mathcal{S}_\mathcal{X}}(\mathbf{cred2}, s_{2047}) &= \frac{1}{2048}(2046 \cdot 0 + 1 + 1) \\ &= \frac{1}{1024} \\ &= \frac{2^{11}}{2^{21}}\end{aligned}$$

In fact, it isn't even close. If Emily adopts **cred2** she becomes a little more than significantly more inaccurate.

It is tedious to calculate $\mathbf{Brier}_{\mathcal{A}_\mathcal{X}}(\mathbf{cred1}, @)$ directly, but it is enlightening to work through the calculation of $\mathbf{Brier}_{\mathcal{A}_\mathcal{X}}(\mathbf{cred2}, s_{2047})$. Note that there are two crucial states out of the 2048: s_{2047}, the actual state where all children come out, and state s_{1023} where child 0 does not come out, but the other 10 children all do. There are $2^{2^{11}-2}$ propositions in each of the following four sets:

1. $\{p: s_{2047} \vDash p \text{ and } s_{1023} \vDash p\}$
2. $\{p: s_{2047} \vDash p \text{ and } s_{1023} \nvDash p\}$
3. $\{p: s_{2047} \nvDash p \text{ and } s_{1023} \vDash p\}$
4. $\{p: s_{2047} \nvDash p \text{ and } s_{1023} \nvDash p\}$

If Emily takes the bribe, she will have perfect accuracy with respect to all the propositions in class 1 (which are correctly believed to be true), and all the propositions in class 4 (which are correctly believed to be false). But she will be perfectly inaccurate with respect to all the propositions in class 2 (which are incorrectly believed to be false), and all the propositions in class 3 (which are incorrectly believed to be true). So she is perfectly accurate on half the propositions, and perfectly inaccurate on half of them, so her average inaccuracy is $0.5 \cdot 0 + 0.5 \cdot 1 = 0.5$. And that's an enormous inaccuracy. It is, in fact, as inaccurate as one can possibly be while maintaining probabilistic coherence.

Theorem-2: When inaccuracy over $\mathcal{A}$ is measured using the Brier score, the least accurate credal states are those which assign credence 1 to some false atom of $\mathcal{A}$.

(The proof is in the appendix.) So taking the bribe is not a good deal, even by consequentialist lights. And that isn't too surprising; taking the bribe makes Emily

have maximally inaccurate credences on half of the possible propositions about the children.

So far we have followed Greaves in assuming that inaccuracy is measured by the quadratic, or Brier, rule. It turns out that we can drop that assumption. We actually only need some very weak conditions on accuracy rules to get the result that Greaves style bribes are bad deals, though the proof of this becomes a trifle more complicated.

Let $\mathcal{A}$ be an algebra of propositions generated by a partition of $2N$atoms $a_1, \ldots, a_{2N}$. Suppose a_1 is the truth, and consider two probability functions, P and Q defined in $\mathcal{A}$. Passigns all its mass to the first N atoms, so that $P(a_k) = 0$ for all $k > N$. We also assume that P assigns some positive probability to the true atom a_1. Qassigns all its mass to the false atom a_{2N}. Note that this will be a good model of any case where an agent is offered a bribe of the form: drop the positive confidence you have in proposition p_0, instead assign it credence 0, and you'll be guaranteed a maximally accurate credence in j other logically independent propositions $p_1, \ldots, p_j$. The only other assumptions needed to get the model to work are that p_0 is actually true, and $N = 2^j$.

Imagine that the accuracy of a probability function π over $\mathcal{A}$ is measured by a proper scoring rule of the form

$$\mathbf{I}(a_n, \pi) = 2^{-2N} \sum_{X \in \mathcal{A}} \mathbf{i}\,(v_n(X), \pi(X))$$

where $v_n(X)$ is Xs truth value when a_n is the true atom, and $\mathbf{i}$ is a score that gives the accuracy of $\pi(X)$ in the event that Xs truth value is $v_n(X)$. We shall assume that this score has the following properties.

Truth Directedness

The value of $\mathbf{i}(1, p)$ decreases monotonically as p increases. The value of $\mathbf{i}(0, p)$ increases monotonically as p decreases.

Extensionality

$\mathbf{i}(v_n(X), \pi(X))$ is a function only of the truth-value and the probability; the identity of the proposition does not matter.

Negation Symmetry

$\mathbf{i}(v_n(\neg X), \pi(\neg X)) = \mathbf{i}(v_n(X), \pi(X))$ for all x, n, π.

Theorem-3: Given these assumptions, *P*'s accuracy strictly exceeds *Q*'s.

Again, the proof is in the appendix.

Theorem-3 ensures that taking the deal that Greaves offers in Imps will reduce Emily's accuracy relative to any proper scoring rule satisfying Truth Directedness, Extensionality and Negation Symmetry. To see why, think of Emily's

credences as being defined over an algebra generated by the atoms $\pm C_0 \wedge \pm C_1 \wedge \pm C_2 \wedge \ldots \wedge \pm C_{10}$, where it is understood that some C_0 atom is true and all the $\neg C_0$ atoms are false. Since Emily is convinced of C_0 and believes that every other C_n has some chance of occurring, and since the various C_n are independent of one another, her credence function **cred1** will assigns a positive probability to each C_0 atom, including the true atom (whichever that might be). Now, let Q be a credence function that places all its weight on some false atom $\neg C_0 \wedge \pm C_1 \wedge \pm C_2 \wedge \ldots \wedge \pm C_{10}$. Theorem-3 tells us that Emily's **cred1** is more accurate than Q, and that this is true no matter which C_0 atom is true or which $\neg C_0$ atom Q regards as certain. By taking the bribe Emily will guarantee the truth of $C_0 \wedge C_1 \wedge \ldots \wedge C_{10}$, but the cost will be that she must adopt the **cred2** credences, which assign probability one to the false atom $\neg C_0 \wedge C_1 \wedge \ldots \wedge C_{10}$. Extensionality ensures that any two credence functions that assign probability one to a false atom will have the same inaccuracy score, and that this score will not depend on which atom happens to be the true one. The upshot is that **cred2** will have the same inaccuracy when Emily accepts the bribe as Q does when she rejects it. Thus, since **cred1** is more accurate than Q, it is also more accurate than **cred2**, which means that Emily should reject the bribe in order to promote credal accuracy.

We do not want to oversell this conclusion. Strictly speaking, we have only shown that consequentialists should reject epistemic bribes when doing so requires them to go from being confident in a truth to being certain of some maximally specific falsehood. This is a rather special situation, and there are nearby cases to which our results do not apply, and in which consequentialists may sanction bribe-taking. For example, if Emily only has to cut her credence for C_0 in half, say from $\frac{1}{2}$to $\frac{1}{4}$, to secure knowledge of $C_1 \wedge \ldots \wedge C_{10}$, then Theorem-3 offers us no useful advice. Indeed, depending on the scoring rule and the nature of the bribe, we suspect that believers will often be able to improve accuracy by changing their credences in ways not supported by their evidence, especially when these changes affect the truth-values of believed propositions. The only thing we insist upon is that, in all such cases, credal accuracy should be measured over all relevant propositions, not just over a select salient few. But that's something that is independently plausible. Perhaps it might be pragmatically justified to become more accurate on salient propositions at the expense of becoming very inaccurate over hard to state compounds of those propositions, but it is never epistemically justified.

2. Four Caveats

2.1 Greaves's Imps Argument May Work Against Some Forms of Consequentialism

We said above that no consequentialist should accept Greaves's setup of the Imps puzzle, since they should not accept an inaccuracy measure that ignores some kind of introduced inaccuracy. That means that, for all we have said, Greaves's argument works against those consequentialists who do not agree with us over the suitability of target sets that are neither algebras or partitions. And, at least outside philosophy, some theorists do seem to disagree with us.

For instance, it is common in meteorology to find theorists who measure the accuracy of rain forecasts over an n day period by just looking at the square of the distance between the probability of rain and the truth about rain on each day. To pick an example almost literally at random, Mark Roulston defends the use of the Brier score, calculated just this way, as a measure of forecast accuracy.[7] So Greaves's target, while not including all consequentialists, does include many real theorists.

That said, it seems there are more mundane reasons to not like this approach to measuring the accuracy of weather forecasts. Consider this simple case. Ankita and Bojan are issuing forecasts for the week that include probabilities of rain. They each think that there is a 0% chance of rain most days. But Ankita thinks there will be one short storm come through during the week, while Bojan issues a 0% chance of rain forecast for each day. Ankita thinks the storm is 75% likely to come on Wednesday, so there's a 75% chance of rain that day, and 25% likely to come Thursday, so there's a 25% chance of rain that day.

As it happens, the storm comes on Thursday. So over the course of the week, Bojan's forecast is more accurate than Ankita's. Bojan is perfectly accurate on 6 days, and off by 1 on Thursday. Ankita is perfectly accurate on 5 days, and gets an inaccuracy score of $0.75^2 = 0.5625$ on Wednesday and Thursday, which adds up to more than Bojan's inaccuracy. But this feels wrong. There is a crucial question that Ankita was right about and Bojan was wrong about, namely will there be a storm in the middle of the week. Ankita's forecast only looks less accurate because we aren't measuring accuracy with respect to this question. So even when we aren't concerned with magical cases like Greaves's, there is a good reason to measure accuracy comprehensively, i.e., with respect to an algebra or a partition.

[7] Mark S. Roulston, "Performance Targets and the Brier Score," *Meterological Applications* 14 (2007): 185–194, http://doi.org/10.1002/met.21.

2.2 Separateness of Propositions

There is a stronger version of the intuition behind the Imps case that we simply reject. The intuition is well expressed by Selim Berker.

> The more general point is this: when determining the epistemic status of a belief in a given proposition, it is epistemically irrelevant whether or not that belief conduces (either directly or indirectly) toward the promotion of true belief and the avoidance of false belief in *other* propositions beyond the one in question.[8]

Let's put that to the test by developing the Ankita and Bojan story a little further. They have decided to include, in the next week's forecast, a judgment on the credibility of rain. Bojan thinks the evidence is rather patchy. And he has been reading Glenn Shafer, and thinks that when the evidence is patchy, credences in propositions and their negations need not add to one.[9] So if p is the proposition *It will rain next week*, Bojan has a credence of 0.4 in both $p$and $\neg p$.

Ankita thinks that's crazy, and suggests that there must be something deeply wrong with the Shafer-based theory that Bojan is using. But Bojan is able to easily show that the common arguments against Shafer's theory are blatantly question begging.[10] So Ankita tries a new tack. She has been reading Joyce, from which she got the following idea.[11] She argues that Bojan will be better off from the point of view of accuracy in having credence 0.5 in each of p and $\neg p$ than in having credence 0.4 in each. As it stands, one of Bojan's credences will be off by 0.4, and the other by 0.6, for a Brier score of $(0.4^2 + 0.6^2)/2 = 0.26$, whereas switching would give him a Brier score of $(0.5^2 + 0.5^2)/2 = 0.25$.

But Bojan resists. He offers two arguments in reply.

First, he says, for all Ankita knows, one of his credences might be best responsive to the evidence. And it is wrong, always and everywhere, to change a credence away from one that is best supported by the evidence in order to facilitate an improvement in global accuracy. That, says Bojan, is a violation of the "separateness of propositions".[12]

---

[8] Berker, "Epistemic Teleology," 365, emphasis in original.

[9] Glenn Shafer, *A Mathematical Theory of Evidence* (Princeton: Princeton University Press, 2016).

[10] Patrick Maher, "Depragmatised Dutch Book Arguments," *Philosophy of Science* 64 (1997): 291–305, http://doi.org/10.1086/392552; Brian Weatherson, "Begging the Question and Bayesians," *Studies in the History and Philosophy of Science* Part A 30 (1999): 687–697.

[11] James M. Joyce, "A Non-Pragmatic Vindication of Probabilism," *Philosophy of Science* 65 (1998): 575–603.

[12] Berker, "Epistemic Teleology."

Second, he says, even by Ankita's accuracy-based lights, this is a bad idea. After all, he will be making one of his credences less accurate in order to make an improvement in global accuracy. And that's again a violation of the separateness of propositions. It's true that he won't be making himself more inaccurate in one respect so as to secure accuracy in another, as in the bribes case. But he will be following advice that is motivated by the aim of becoming, in total, more accurate, at the expense of accuracy for some beliefs.

We want to make two points in response. First, if the general point that Berker offers is correct, then these are perfectly sound replies by Bojan. Although Bojan is not literally in a bribe case, like Emily, he is being advised to change some credences because the change will make his overall credal state better, even if it makes it locally worse in one place. It does not seem to matter whether he can identify which credence gets made worse. Berker argues that the trade-offs that epistemic consequentialism makes the same mistake ethical consequentialism makes; it authorises inappropriate trade-offs. But in the ethical case, it doesn't matter whether the agent can identify who is harmed by the trade-off. If it is wrong to harm an identifiable person for the greater good, it is wrong to harm whoever satisfies some description in order to produce the greater good.

So if the analogy with anti-consequentialism in ethics goes through, Bojan is justified in rejecting Ankita's advice. After all there is, according to Berker, a rule against making oneself doxastically worse in one spot for the gain of an overall improvement. And that's what Bojan would do if he took Ankita's advice. But, we say, Bojan is not justified in rejecting Ankita's advice. In fact, Ankita's advice is sound advice, and Bojan would do well to take it. So Berker's general point is wrong.

Our second point is a little more contentious. We suspect that if Bojan has a good reason to resist this move of Ankita's, he has good reason to resist all attacks on his Shafer-based position. So if Berker's general point is right, it means there is nothing wrong with Bojan's anti-probabilist position. Now we haven't argued for this; to do so would require going through all the arguments for probabilism and seeing whether they can be made consistent with Berker's general point. But our suspicion is that none of them can be, since they are all arguments that turn on undesirable properties of global features of non-probabilistic credal states. So if Berker is right, probabilism is wrong, and we think it is not wrong.

2.3 Is this Consequentialism?

So far we've acquiesed with the general idea that Greaves's and Berker's target should be called *consequentialism*. But there are reasons to be unhappy with this label. In general, a consequentialist theory allows agents to make things worse in the here and now, in return for future gains. A consequentialist about prudential decision making, in the sense of Hammond, will recommend exercise and medicine taking.[13] And they won't be moved by the fact that the exercise hurts and the medicine is foul-tasting. It is worth sacrificing the welfare of the present self for the greater welfare of later selves.

Nothing like that is endorsed, as far as we can tell, by any of the existing 'epistemic consequentialists'. Certainly the argument that Ankita offers Bojan does not rely on this kind of reasoning. In particular, epistemic consequentialists do not say that it is better to make oneself doxastically worse off now in exchange for greater goods later. Something like that deal is offered to the reader of Descartes's *Meditations*, but it isn't as popular nowadays.

Rather, the rule that is endorsed is *Right now, have the credences that best track the truth!* This isn't clearly a form of consequentialism, since it really doesn't care about the *consequences* of one's beliefs. It does say that it is fine to make parts of one's doxastic state worse in order to make the whole better. That's what would happen if Bojan accepted Ankita's advice. But that's very different from doing painful exercise, or drinking unpleasant medicine. (Or, for that matter, to withdrawing belief in any number of truths.)

When Greaves tries to flesh out epistemic consequentialism, she compares it to evidential and causal versions of prudential decision theory. But it seems like the right comparison might be to something we could call *constitutive* decision theory. The core rule, remember, is that agents should form credences that constitute being maximally accurate, not that cause them to be maximally accurate.

The key point here is not the terminological one about who should be called consequentialist. Rather, it is that the distinction between causation and constitution is very significant here, and comparing epistemic utility theory to prudential utility theory can easily cause it to be lost. Put another way, we have no interest in defending someone who wants to defend a causal version of epistemic utility theory, and hence thinks it could be epistemically rational to be deliberately

[13] Peter J. Hammond, "Consequentialist Foundations for Expected Utility," *Theory and Decision* 25 (1988): 25–78, http://doi.org/10.1007/BF00129168.

inaccurate now in order to be much more accurate tomorrow. We do want to defend the view that overall accuracy right now is a prime epistemic goal.[14]

2.4 Other Bribes

As already noted, we have not offered a general purpose response to bribery based objections to epistemic consequentialism. All we've shown is that some popular examples of this form of objection misfire, because they offer bribes that are bad by the consequentialists' own lights. But there could be bribes that are immune to our objection.

For example, imagine that Ankita has, right now, with credence 0.9 in D_0, and 0.5 in D_1. These are good credences to have, since she knows those are the chances of $D_0$and D_1. She's then offered an epistemic bribe. If she changes her credence in D_0 to 0.91, the chance of D_1 will become 1, and she can have credence 1 in D_1. Taking this bribe will increase her accuracy.

We could imagine the anti-consequentialist arguing as follows.

1. If epistemic consequentialism is true, Ankita is epistemically justified in accepting this bribe.
2. Ankita is not epistemically justified in accepting this bribe.
3. So, epistemic consequentialism is not true.

We're not going to offer a reply to this argument here; that is a task for a much longer paper. There are some reasons to resist premise one. It isn't clear that it is conceptually possible to accept the bribe. (It really isn't clear that it is practically possible, but we're not sure whether that's a good reply on the part of the consequentialist.) And it isn't clear that the argument for premise one properly respects the distinction between causation and constitution we described in the last section.

Even if those arguments fail, the intuitive force of premise two is not as strong as the intuition behind Greaves's, or Berker's, anti-bribery intuitions. And that's one of the main upshots of this paper. It's commonly thought that for the consequentialist, in any field, everything has its price. The result we proved at the end of section one shows this isn't true. It turns out that no good epistemic consequentialist should accept a bribe that leads them to believing an atomic proposition they have conclusive evidence is false, no matter how strong the

[14] For further discussion of epistemic consequentialism, see James M. Joyce, "Accuracy, Ratification, and the Scope of Epistemic Consequentialism," In *Epistemic Consequentialism*, eds. H. Kristoffer Ahlstrom-Vij and Jeffrey Dunn (Oxford: Oxford University Press, 2018), 240-266. https://doi.org/10.1093/oso/9780198779681.003.0011

inducements. Maybe one day there will be a convincing bribery based case that epistemic consequentialism is unacceptably corrupting of the epistemic soul. But that case hasn't been made yet, because we've shown a limit on how corrupt the consequentialist can be.

Appendix: Proofs of Theorems 1, 2, 3

Theorem-1: $\text{Brier}_{\mathcal{A}}(\mathbf{c}, @) = \frac{N}{4}\text{Brier}_{\mathcal{S}}(\mathbf{c}, @)$ where

$$\text{Brier}_{\mathcal{S}}(\mathbf{c}, @) \quad = \frac{\sum_{s\in\mathcal{S}} (\,@(s) - c(s))^2}{N}$$

To prove this we rely on a series of lemmas.[15]

Let $\mathcal{A}$ be the algebra generated by a finite partition of states $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$. @ is a truth-value assignment for propositions in $\mathcal{A}$. For simplicity, assume s_1 is the true state, so that $@(s_1) = 1$ and $@(s_n) = 0$ for $n > 1$. The credence function $\mathbf{c}$ assigns values of $c_1, c_2, \ldots, c_{N-1}, c_N$ to the elements of $\mathcal{S}$, where $\sum_{n=1}^{N} c_n = 1$ in virtue of coherence.

It will be convenient to start by partitioning $\mathcal{A}$ into four "quadrants". Let B range over all disjunctions with disjunctions drawn from $\mathcal{B} = \{s_2, s_3, \ldots, s_{N-1}\}$ (including the empty disjunction, i.e., the logical contradition $\perp$). Then, $\mathcal{A}$ can be split into four disjoint parts:

$\mathcal{A}_1 = \{B \vee s_1 \vee s_N : B$ is a disjunction of the elements of $\mathcal{B}\}$

$\mathcal{A}_2 = \{B \vee s_1 : B$ is a disjunction of the elements of $\mathcal{B}\}$

$\mathcal{A}_3 = \{B \vee s_N : B$ is a disjunction of the elements of $\mathcal{B}\}$

$\mathcal{A}_4 = \{B : B$ is a disjunction of the elements of $\mathcal{B}\}$

Notice that:

(i) $\mathcal{A}_1 \cup \mathcal{A}_2$contains all and only the true propositions in $\mathcal{A}$.

(ii) $\mathcal{A}_3 \cup \mathcal{A}_4$contains all and only the false propositions in $\mathcal{A}$.

(iii) $\mathcal{A}_1$ and $\mathcal{A}_4$ are *complementary* sets, i.e., all elements of $\mathcal{A}_4$ are negations of elements of $\mathcal{A}_1$, and conversely.

(iv) $\mathcal{A}_2$ and $\mathcal{A}_3$ are also complementary.

(v) $\mathcal{A}_1 \cup \mathcal{A}_4$ is the subalgebra of $\mathcal{A}$ generated by $\{s_1 \vee s_N, s_2, s_3, \ldots, s_{N-1}\}$.

(vi) All four quadrants have the same cardinality of 2^{N-2}.

[15]Alejandro Pérez Carballo gives a more direct and elegant proof of this result in a recent manuscript. We have kept our inefficient proof since its structure provides a guide for the proof of Theorem-3.

For an additive scoring rule $\mathbf{I}(\mathbf{c},@) = \sum_{A\in\mathcal{A}} \mathbf{i}\,(\mathbf{c}(A),@(A))$ and $j = 1,2,3,4$, define $\mathbf{I}_j = \sum_{A\in\mathcal{A}_j} \mathbf{i}\,(\mathbf{c}(A),@(A))$, and note that $\mathbf{I}(\mathbf{c},@) = 2^{-N}(\mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3 + \mathbf{I}_4)$.

Lemma-1.1: If $\mathbf{I}$ is negation symmetric, i.e., if $\mathbf{i}(\mathbf{c}(\neg A),@(\neg A)) = \mathbf{i}(\mathbf{c}(A),@(A))$ for all A, then $\mathbf{I}_1 = \mathbf{I}_4$ and $\mathbf{I}_2 = \mathbf{I}_3$, and $\mathbf{I}(\mathbf{c},@) = 2^{1-N}(\mathbf{I}_2 + \mathbf{I}_4)$.

Proof: This is a direct consequence of the fact that $\mathcal{A}_1$ is complementary to $\mathcal{A}_4$ and that $\mathcal{A}_2$ is complementary to $\mathcal{A}_3$ since this allows us to write

$$\mathbf{I}_1(\mathbf{c},@) = \sum_{A\in\mathcal{A}_1} \mathbf{i}\,(\mathbf{c}(A),@(A)) = \sum_{A\in\mathcal{A}_1} \mathbf{i}\,(\mathbf{c}(\neg A),@(\neg A)) = \mathbf{I}_4(\mathbf{c},@).$$
$$\mathbf{I}_3(\mathbf{c},@) = \sum_{A\in\mathcal{A}_3} \mathbf{i}\,(\mathbf{c}(A),@(A)) = \sum_{A\in\mathcal{A}_3} \mathbf{i}\,(\mathbf{c}(\neg A),@(\neg A)) = \mathbf{I}_2(\mathbf{c},@).\ \text{QED}$$

Applying Lemma 1.1 with $\mathbf{I}$ = **Brier** we get

$$(\#)\quad \mathbf{Brier}_{\mathcal{A}}(\mathbf{c},@) = 2^{1-N}\sum_{A\in\mathcal{A}} (\,@(A) - c(A))^2$$
$$= 2^{1-N}\sum_{B} [\,(1-c_1)^2 - 2(1-c_1)\mathbf{c}(B) + \mathbf{c}(B)^2]$$

since

$$\mathbf{Brier}_2 = \sum_{B} [\,1 - \mathbf{c}(B \vee s_1)]^2 = \sum_{B} [\,(1-c_1) - \mathbf{c}(B)]^2$$
$$= \sum_{B} [\,(1-c_1)^2 - 2(1-c_1)\mathbf{c}(B) + \mathbf{c}(B)^2]$$

$$\mathbf{Brier}_4 = \sum_{B} \mathbf{c}\,(B)^2$$

Lemma-1.2

$$(\sum_{n=2}^{N-1} c_n)^2 = \sum_{n=2}^{N-1} {c_n}^2 + 2\sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n\, c_j$$

Proof by induction. Easy.

Lemma-1.3

$$\mathbf{Brier}_S(\mathbf{c},@) = \frac{2}{N}[(1-c_1)^2 + \sum_{n=2}^{N-1} {c_n}^2 - (1-c_1)(\sum_{n=2}^{N-1} c_n) + \sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n\, c_j]$$

Proof: Using the definition of the Brier score and the fact that s_1 is true, we have

$$\begin{aligned}
\textbf{Brier}_{\mathcal{S}}(\mathbf{c},@) &= \frac{1}{N}[(1-c_1)^2 + \sum_{n=2}^{N-1} {c_n}^2 + (1-\sum_{n=1}^{N-1} c_n)^2] \\
&= \frac{1}{N}[(1-c_1)^2 + \sum_{n=2}^{N-1} {c_n}^2 + ((1-c_1)-\sum_{n=2}^{N-1} c_n)^2] \\
&= \frac{1}{N}[(1-c_1)^2 + \sum_{n=2}^{N-1} {c_n}^2 + (1-c_1)^2 - 2(1-c_1)\sum_{n=2}^{N-1} c_n + (\sum_{n=2}^{N-1} c_n)^2] \\
&= \frac{1}{N}[(1-c_1)^2 + \sum_{n=2}^{N-1} {c_n}^2 + (1-c_1)^2 - 2(1-c_1)\sum_{n=2}^{N-1} c_n \\
&\quad + \sum_{n=2}^{N-1} {c_n}^2 + 2\sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n\, c_j] \quad (\text{Lemma} - 1.2)
\end{aligned}$$

Then grouping like terms and factoring out 2 yields the desired result. QED

Lemma-1.4

$$\sum_{n=2}^{N-1} c_n \;\; = 2^{3-N} \sum_{B\in\mathcal{B}} \mathbf{c}\,(B)$$

Proof: For each $n = 2,3,\ldots,N-1$, each s_n appears in half of the 2^{N-2} disjunctions with disjuncts drawn from $\mathcal{B}$. As a result, each c_n appears as a summand 2^{N-3} times among the sums that express the various $\mathbf{c}(B)$. So $\sum_{B\in\mathcal{B}} \mathbf{c}\,(B) = 2^{N-3}\sum_{n=2}^{N-1} c_n$. QED

Lemma-1.5

$$\sum_{B\in\mathcal{B}} \mathbf{c}\,(B)^2 \;\; = 2^{N-3}[\sum_{n=2}^{N-1} {c_n}^2 + \sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n\, c_j]$$

Proof: We proceed by induction starting with the first meaningful case of $N = 4$, where calculation shows $\sum_B \mathbf{c}\,(B)^2 = (c_2 + c_3)^2 + {c_2}^2 + {c_3}^2 = 2[{c_2}^2 + {c_3}^2 + c_2 c_3]$. Now, assume the identity holds for disjunctions B of elements of $\mathcal{B}$ and show that it holds for disjunctions A of elements of $\mathcal{B} \cup \{s_N\}$.

$$\begin{aligned}
\sum_A \mathbf{c}(A)^2 &= \sum_B \mathbf{c}(B)^2 + \sum_B \mathbf{c}(B \vee s_N)^2 \\
&= \sum_B \mathbf{c}(B)^2 + \sum_B (\mathbf{c}(B)^2 + 2c_N\mathbf{c}(B) + {c_N}^2) \\
&= 2\sum_B \mathbf{c}(B)^2 + 2c_N \sum_B \mathbf{c}(B) + \sum_B {c_N}^2 \\
&= 2 \cdot 2^{N-3}\left[\sum_{n=2}^{N-1} {c_n}^2 + \sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n c_j\right] + 2c_N \sum_B \mathbf{c}(B) + \sum_B {c_N}^2 && \binom{\text{Induction}}{\text{Hypothesis}} \\
&= 2^{N-2}\left[\sum_{n=2}^{N-1} {c_n}^2 + \sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n c_j\right] + 2^{N-2}c_N \sum_{n=2}^{N-1} c_n + \sum_B {c_N}^2 && (\text{Lemma} - 1.4) \\
&= 2^{N-2}\left[\sum_{n=2}^{N-1} {c_n}^2 + \sum_{n=2}^{N-2}\sum_{j>n}^{N-1} c_n c_j\right] + 2^{N-2}c_N \sum_{n=2}^{N-1} c_n + 2^{N-2}{c_N}^2 && \text{Since } |\mathcal{B}| = 2^{N-2} \\
&= 2^{N-2}\left[\sum_{n=2}^{N} {c_n}^2 + \sum_{n=2}^{N-1}\sum_{j>n}^{N} c_n c_j\right] && \text{QED}
\end{aligned}$$

Plugging the results of the last two lemmas into Lemma-1.3 produces a result of

$$\begin{aligned}
\mathbf{Brier}_{\mathcal{S}}(\mathbf{c}, @) &= \frac{2}{N}[(1-c_1)^2 + 2^{3-N}\sum_{B\in\mathcal{B}} \mathbf{c}(B)^2 - 2^{3-N}(1-c_1)\sum_{B\in\mathcal{B}} \mathbf{c}(B)] \\
&= \frac{2}{N}\sum_{B\in\mathcal{B}} [\,2^{2-N}(1-c_1)^2 + 2^{3-N}\mathbf{c}(B)^2 - 2^{3-N}(1-c_1)\mathbf{c}(B)] \\
&= \frac{2^{3-N}}{N}\sum_{B\in\mathcal{B}} [\,(1-c_1)^2 + 2\mathbf{c}(B)^2 - 2(1-c_1)\mathbf{c}(B)]
\end{aligned}$$

Comparing this to (#) we see that it is just $\frac{N}{4}$ times $\mathrm{Brier}_{\mathcal{S}}(\mathbf{c}, @)$, as we aimed to prove. QED.

Theorem-2. When inaccuracy over $\mathcal{A}$ is measured using the Brier score, the least accurate credal states are those which assign credence 1 to some false atom of $\mathcal{A}$.

Proof: As before, suppose that $@(s_1) = 1$, and let $\mathbf{c}$ be a credence function that assigns credence 1 to some false atom $s_2, s_3, \ldots, s_N$ of $\mathcal{A}$. In light of Theorem-1 it suffices to show that $\mathbf{Brier}_{\mathcal{S}}(\mathbf{c}, @) > \mathbf{Brier}_{\mathcal{S}}(\mathbf{b}, @)$ where $\mathbf{b}$ does not assign credence 1 to any false atom. Start by noting that for any credence function π defined on the atoms of $\mathcal{A}$ one has

$$\begin{aligned}\mathbf{Brier}_{\mathcal{S}}(\pi,@) &= \frac{1}{N}[(1-\pi_1)^2 + \sum_{n=2}^{N-1} {\pi_n}^2 + (1 - \sum_{n=1}^{N-1} \pi_n)^2] \\ &= \frac{1}{N}[1 - 2\pi_1 + \sum_{n=1}^{N-1} {\pi_n}^2 + (1 - \sum_{n=1}^{N-1} \pi_n)^2]\end{aligned}$$

But, since each $\pi_n \in [0,1]$ is non-negative, it follows that $\pi_1 \geq {\pi_1}^2, \pi_2 \geq {\pi_2}^2, \ldots, \pi_N \geq {\pi_N}^2$ with the inequality strict in each case unless π_n is either 1 or 0.

This means that the sum $\sum_{n=1}^{N-1} {\pi_n}^2 + (1 - \sum_{n=1}^{N-1} \pi_n)^2$ is less than or equal to 1, with equality if and only if exactly one of the atoms s_n is assigned probability 1 (and the rest have probability zero). As a result, $\mathbf{Brier}_{\mathcal{S}}(\pi,@) \leq \frac{2}{N}(1-\pi_1)$ with equality if and only if exactly one of the atoms s_n is assigned probability 1. So, there are three relevant cases:

(i) If π assigns some false atom probability 1, $\mathbf{Brier}_{\mathcal{S}}(\pi,@) = \frac{2}{N} \cdot (1-0) = \frac{2}{N}$.

(ii) If π assigns the true atom probability 1, $\mathbf{Brier}_{\mathcal{S}}(\pi,@) = \frac{2}{N} \cdot (1-1) = 0$.

(iii) If π does not assign any atom probability 1, $\mathbf{Brier}_{\mathcal{S}}(\pi,@) < \frac{2}{N} \cdot (1-c_1) \leq \frac{2}{N}$.

So, since **c** fits case (i) and **b** fits case (ii) or (iii) we have the desired result. QED

Theorem-3: Let $\mathcal{A}$ be an algebra of propositions generated by atoms $a_1, \ldots, a_{2N}$, where a_1 is the truth. Let P and Q be probability functions defined on $\mathcal{A}$. P assigns all its mass to the first N atoms, so that $P(a_1 \vee \ldots \vee a_N) = 1$, and it also assigns some positive probability to a_1. Q assigns all its mass to the false atom a_{2N}, so that $Q(a_{2N}) = 1$. Then, for any proper score $\mathbf{I}$ satisfying Truth-directedness, Extensionality and Negation Symmetry we have $\mathbf{I}(v_1, P) < \mathbf{I}(v_1, Q)$ where v_1 is the truth-value assignment associated with a_1 (i.e., where $v_1(X) = 1$ if and only if a_1 entails X).

Proof: We can divide the algebra $\mathcal{A}$ into four quadrants

$$\begin{aligned}\mathcal{A}^1 &= \{X \in \mathcal{A}: a_1 \vDash X \text{ and } a_{2N} \vDash X\} \\ \mathcal{A}^2 &= \{X \in \mathcal{A}: a_1 \vDash X \text{ and } a_{2N} \nvDash X\} \\ \mathcal{A}^3 &= \{X \in \mathcal{A}: a_1 \nvDash X \text{ and } a_{2N} \vDash X\} \\ \mathcal{A}^4 &= \{X \in \mathcal{A}: a_1 \nvDash X \text{ and } a_{2N} \nvDash X\}\end{aligned}$$

We know the following:

- Q is maximally accurate on $\mathcal{A}^1 \cup \mathcal{A}^4$. Every proposition in $\mathcal{A}^1$ is true, and Q assigns it a probability of 1. Every proposition in $\mathcal{A}^4$ is false, and Q assigns it a probability of 0.
- Q is maximally inaccurate on $\mathcal{A}^2 \cup \mathcal{A}^3$. Every proposition in $\mathcal{A}^2$ is true, and Q assigns it a probability of 0. Every proposition in $\mathcal{A}^3$ is false, and Q assigns it

a probability of 1.

- P is maximally accurate on $\mathcal{A}^3 \cup \mathcal{A}^4$. Every proposition in $\mathcal{A}^3 \cup \mathcal{A}^4$ is false, and P assigns it a probability of 0.
- Each quadrant has 2^{2N-2} elements.

Lemma-3.1: When a_1 is true, the accuracy score of P over the propositions in $\mathcal{A}^1$ is identical to the accuracy score of P over the propositions in $\mathcal{A}^2$.

Proof: Note first that the function $F: \mathcal{A}^1 \to \mathcal{A}^2$ that takes X to $X \wedge \neg a_{2N}$ is a bijection of $\mathcal{A}^1$onto $\mathcal{A}^2$. Since every proposition in $\mathcal{A}^1 \cup \mathcal{A}^2$ is true, we can then write the respective accuracy scores of $\mathcal{A}^1$ and $\mathcal{A}^2$ as

$$\begin{aligned} \mathbf{I}_{\mathcal{A}^1}(a_1, P) &= 2^{2-2N} \cdot \sum_{X \in \mathcal{A}^1} \mathbf{I}(1, P(X)) \\ \mathbf{I}_{\mathcal{A}^2}(a_1, P) &= 2^{2-2N} \cdot \sum_{X \in \mathcal{A}^1} \mathbf{I}(1, P(X \wedge \neg a_{2N})) \end{aligned}$$

Note: X ranges over $\mathcal{A}^1$ in both summations. But since $P(a_{2N}) = 0$ we have $P(X) = P(X \wedge a_{2N})$ for each Xin $\mathcal{A}^1$. Since $\mathbf{I}$ is extensional, this means that $\mathbf{I}(1, P(X)) = \mathbf{I}(1, P(X \wedge a_{2N}))$ for each Xin $\mathcal{A}^1$. And, it follows that $\mathbf{I}_{\mathcal{A}^1}(a_1, P)$ and $\mathbf{I}_{\mathcal{A}^2}(a_1, P)$ are identical. (Note that even if $P(a_{2N}) > 0$, Truth-directedness entails that $\mathbf{I}_{\mathcal{A}^1}(a_1, P) < \mathbf{I}_{\mathcal{A}^2}(a_1, P)$.)

Lemma-3.2: When a_1 is true, the accuracy score of Q over $\mathcal{A}^2$ is identical to the accuracy score of Q over $\mathcal{A}^3$.

Proof: To see this, note first that the function $G: \mathcal{A}^2 \to \mathcal{A}^3$ that takes X to $G(X) = \neg X$ is a bijection (i.e., the negation of everything in $\mathcal{A}^2$ is in $\mathcal{A}^3$ and vice-versa). This, together with the fact that $\mathcal{A}^2$ contains only truths and $\mathcal{A}^3$ contains only falsehoods, lets us write

$$\begin{aligned} \mathbf{I}_{\mathcal{A}^2}(a_1, Q) &= 2^{2-2N} \cdot \sum_{X \in \mathcal{A}^2} \mathbf{I}(1, Q(X)) \\ \mathbf{I}_{\mathcal{A}^3}(a_1, Q) &= 2^{2-2N} \cdot \sum_{X \in \mathcal{A}^2} \mathbf{I}(0, Q(\neg X)) \end{aligned}$$

But since $\mathbf{I}$ is negation symmetric, $\mathbf{I}(1, Q(X)) = \mathbf{I}(0, Q(\neg X))$ for every X, which means that $\mathbf{I}_{\mathcal{A}^2}(a_1, Q) = \mathbf{I}_{\mathcal{A}^3}(a_1, Q)$. (Note that this proof made no assumptions about Q except that it was a probability.)

Lemma-3.3: If $P(a_1) > 0$, the accuracy score of P over $\mathcal{A}^2$ is strictly less than the accuracy score of Q over $\mathcal{A}^2$.

Proof: Since $Q(X) = 0$ everywhere on $\mathcal{A}^2$ we have

$$\mathbf{I}_{\mathcal{A}^2}(a_1, P) = 2^{2-2N} \cdot \sum_{X \in \mathcal{A}^2} \mathbf{I}(1, P(X))$$
$$\mathbf{I}_{\mathcal{A}^2}(a_1, Q) = 2^{2-2N} \cdot \sum_{X \in \mathcal{A}^2} \mathbf{I}(1,0)$$

But, by Truth Directedness $\mathbf{I}(1,0) > \mathbf{I}(1, P(X))$ since $P(a_1) > 0$ implies that $P(X) > 0$ for all $X \in \mathcal{A}^2$. Thus $\mathbf{I}_{\mathcal{A}^2}(a_1, Q) > \mathbf{I}_{\mathcal{A}^2}(a_1, P)$.

To complete the proof of the theorem we need only note that

$$\begin{aligned}
\mathbf{I}_{\mathcal{A}}(a_1, P) &= \frac{\mathbf{I}_{\mathcal{A}^1}(a_1, P)}{4} + \frac{\mathbf{I}_{\mathcal{A}^2}(a_1, P)}{4} && \text{(since } P \text{ is perfect on } \mathcal{A}^3 \cup \mathcal{A}^4\text{)} \\
&= \frac{\mathbf{I}_{\mathcal{A}^2}(a_1, P)}{2} && \text{Lemma} - 3.1 \\
&< \frac{\mathbf{I}_{\mathcal{A}^2}(a_1, Q)}{2} && \text{Lemma} - 3.3 \\
&= \frac{\mathbf{I}_{\mathcal{A}^2}(a_1, Q)}{4} + \frac{\mathbf{I}_{\mathcal{A}^3}(a_1, Q)}{4} && \text{Lemma} - 3.2 \\
&= \mathbf{I}_{\mathcal{A}}(a_1, Q) && \text{(since } Q \text{ is perfect on } \mathcal{A}^1 \cup \mathcal{A}^4\text{)}
\end{aligned}$$