Accuracy and Verisimilitude: The Good, the Bad, and the Ugly

Miriam Schoenfield

It seems like we care about at least two features of our credence function: gradational-accuracy (high credences in truths, low credences in falsehoods) and verisimilitude (investing higher credence in worlds that are more similar to the actual world). Accuracy-first epistemology requires that we care about one feature of our credence function: gradational-accuracy. So if you want to be a verisimilitude-valuing accuracy-firster, you must be able to think of the value of verisimilitude as somehow built into the value of gradational-accuracy. Can this be done? In a recent article, Oddie has argued that it cannot, at least if we want the accuracy measure to be proper. I argue that it can.

1. Introduction

Consider the credences of Aggie and Vera with respect to the number of planets in our solar system.

Aggie	**Vera**
Pr(Eight planets) $= 0.5$	Pr(Seven planets) $= 1$
Pr(One planet) $= 0.5$	

Which of these two characters is getting things more right? You might think that it's hard to say. After all, Aggie invests more credence in the actual world than Vera does. On the other hand, Aggie invests a large chunk of credence in a world that is extremely dissimilar to the way things are, whereas all of Vera's credence is invested in a world that is quite similar to the actual world. These contemplations might lead you to think that Aggie and Vera are doing well on different epistemically important dimensions: Aggie, one might say, is doing well with respect to gradational-accuracy (a matter, roughly, of investing high credence in truth and low credence in falsehood), whereas Vera is doing well with respect to verisimilitude (a matter, roughly, of investing more credence in worlds that are more similar to the actual world).

The British Journal for the Philosophy of Science, volume 73, number 2, June 2022.
 https://doi.org/10.1093/bjps/axz032

You might think that getting things right is a matter of doing well on both dimensions and that we need to be epistemic-value-pluralists to explain the sense in which Aggie is doing better than Vera and the sense in which Vera is doing better than Aggie.

But epistemic utility theory (EUT), as practiced by most contemporary epistemologists, is not a pluralist endeavor. The EUT programme assumes that the only thing that's of fundamental epistemic value is gradational-accuracy. Indeed, the idea that gradational-accuracy is all that matters runs so deep that most people in the discipline call gradational-accuracy just plain old 'accuracy'. I'm going to follow this (perhaps unfortunate) convention and use 'accuracy' to refer to gradational-accuracy.[1]

As it turns out, many very attractive norms (like probabilism and conditionalization) can be derived from the assumption that accuracy is the only thing that is (fundamentally) epistemically valuable in a credence function.[2] And these arguments simply don't work if there are other dimensions of value. For if other values are important, then the fact that, for instance, every non-probabilistic credence function is accuracy-dominated by a probabilistic one won't necessarily imply that we should be probabilistic. After all, the non-probabilistic ones might be doing well in other respects. So thinking that not only accuracy, but also verisimilitude is epistemically valuable has the potential to wreak havoc to 'accuracy-first' epistemology.

It's not just a programme that some formal epistemologists are interested in that is threatened by thinking that verisimilitude, in addition to accuracy, is valuable. It would be strange if it turned out that agents need to be trading off these values when forming their opinions. Suppose *S*'s evidence supports a credence of 0.9 that there are eight planets and a credence of 0.1 that there is one planet. Her evidence decisively rules out every other number of planets. It would be odd if she should think: 'It's highly likely that there are eight planets. If there are eight planets, I'm better off investing whatever credence I don't invest the eight-planet world in the seven-planet world than I am investing it in the one-planet world (for verisimilitude reasons). So perhaps I should move some of my credence from the one-planet world to the seven-planet world'. Somehow, whatever norms we derive from the fact that we value accuracy and verisimilitude should rule out reasoning in this way.

One way to be a verisimilitude-valuing accuracy-firster, and to assure that reasoning in the way described above is unwarranted is to show that, in fact, valuing verisimilitude just amounts to valuing accuracy in a particular way. According to this picture, all we need to do in our inquiry is keep caring about accuracy, and verisimilitude will take care of itself. (Exactly how this works will be explained below.) But this hope appears to have been dashed by Oddie ([2019]). Oddie argues that given a plausible constraint on accuracy measures called 'propriety' (proper measures are those according to which every probability function maximizes expected accuracy

[1] This is just terminology—I don't mean to be taking a stand (yet!) on anything substantive about what's epistemically valuable.

[2] For a comprehensive presentation of the programme, see (Pettigrew [2016]).

relative to itself)[3] there simply is no way of valuing accuracy that can capture the value of verisimilitude. The accuracy-first epistemology programme requires that the accuracy measures we use are proper, so if the value of verisimilitude can only be captured by improper measures, accuracy-first is in trouble.

The first goal of this article is to argue, contra Oddie, that the value of verisimilar credences can indeed be captured by proper accuracy measures. So the intuition that verisimilitude is important doesn't threaten accuracy-first epistemology. We can maintain epistemic-value-monism and still capture the sense in which Vera is doing better than Aggie. However, as we'll see, given propriety, the way verisimilitude fits into the picture is a bit complicated. The second goal of this article is to present some results aimed at mapping out some of the contours of the complex relationship between accuracy and verisimilitude.

One upshot of the results that follow is this: both proper and improper measures can capture the value of verisimilitude. The difference between them is that there are improper measures that care only about how much credence is invested in the actual world, and how verisimilar the credence function is. Proper measures care about at least one additional feature: how evenly credence is distributed amongst certain false propositions. In fact, as we'll see, a version of the Brier score (a much-loved proper accuracy measure in accuracy-first circles) cares about exactly three things: the amount of credence invested in the actual world, verisimilitude, and evenness of distribution amongst (a certain class of) falsehoods. At the end, I'll appeal to results by Schervish ([1989]) and Levinstein ([2017]) to argue that these three features are exactly the ones that are important from a practical perspective when we don't know what sorts of decisions we'll face in the future.

2. Some Nuts and Bolts

Before delving in, I want to present the basics of the accuracy framework. (Familiar readers may wish to skip this section.)

Let Ω be a finite set of possible worlds (mutually exclusive possibilities) and let $P(\Omega)$ be the power set of Ω—the set of sets of worlds in Ω. We'll be thinking of propositions as sets of worlds. Credence functions defined on Ω will be assignments of numbers in [0, 1] (credences) to the propositions in $P(\Omega)$.[4] We can measure how accurate someone's credences are at a world using an accuracy measure. There are two types of accuracy measures we'll be looking at: local and global.

A local accuracy measure takes as input some credence c and a truth value (one for true, zero for false), and outputs a number (also in [0, 1]) that represents how accurate someone is who has credence c in a proposition, given the proposition's truth value.

[3] For discussion of propriety and its motivations, see, for instance, (Greaves and Wallace [2006]; Joyce [2009]; Pettigrew [2016]).

[4] For simplicity, I'll usually be assuming that the credence functions in questions are probabilistic.

It's easier for certain purposes to use measures of inaccuracy rather than accuracy. So we'll say that 0.9 in a truth is less inaccurate than 0.5 in a truth. We'll use 'I_{local}' to represent a local inaccuracy measure:

$$I_{local} : \underset{\text{(credence)}}{[0,1]} \times \underset{\text{(truth value)}}{\{0,1\}} \rightarrow \underset{\text{(inaccuracy score)}}{[0,1]}.$$

Global inaccuracy measures take as input an entire credence function and a world, and output a number representing how inaccurate the credence function is in that world. We'll use 'I_{global}' to represent a global inaccuracy measure. If c is the set of credence functions defined over a set over worlds Ω:

$$I_{global} : c \times \Omega \rightarrow [0,1].$$

We'll think of the global inaccuracy of a credence function at world w as a sum of the local inaccuracy scores that the credence function gets in w for each proposition it assigns credence to.[5]

3. First Attempts

At first glance it seems like accuracy and verisimilitude are two completely different beasts. Accuracy is based on similarity relations between credence functions: in particular, accuracy at world w is concerned with the similarity between any particular credence function, and the omniscient one at w (the one that assigns one to truths and zero to falsehoods). Verisimilitude, in contrast, is based on similarity relations between worlds. The person who thinks that there are seven planets in our solar system is doing better than the person who thinks that there is one planet, because the world in which there are seven planets is more similar to the actual world than is the world in which there is just one.

But at second glance it seems that the seven-planeter is also, in certain respects, more accurate—not merely more versimilar—than the one-planeter. The seven-planeter, for example, is accurate with respect to the following propositions: There are at least seven planets, there are at least six planets, there are between four and nine planets, and so on. The one-planeter is wrong about all of those things. So, you might think, we can explain what's better about the seven-planeter in terms of accuracy alone.

Sadly, things are not so simple. For the one-planeter is right about many things that the seven-planeter is wrong about: that there are either one or eight planets, that there are seven planets if and only if there are two, that the number of planets is greater than seven or less than five, and so forth. In fact, the one-planeter and the seven-planeter are right about exactly the same number of propositions (Oddie [2019]).

[5] Note that the sum can be weighted; more on this later.

Still, you might think, the propositions that the seven-planeter is right about are in some sense 'better' than the propositions that the one-planeter is right about. That there are at least six planets is a very respectable proposition. That there are two planets if and only if there are seven is a weird ugly one. Greaves and Wallace ([2006]) propose that we can incorporate the value of verisimilitude into the value of accuracy by assigning different weights to different propositions. Here's the thought: recall that how accurate a person is overall is a function of how accurate they are with respect to individual propositions. Instead of just adding up a person's accuracy score for each proposition, we can weight those scores to reflect the fact that we think that there are certain propositions it's more important to be right about than others. If we weight more heavily propositions like 'there are at least six planets' than propositions like 'there are two or seven planets', perhaps we can get the result that the seven-planeter is more (globally) accurate than the one-planeter.[6]

The Greaves and Wallace proposal seems promising. But Oddie argues that if one meets a constraint that he takes to be an extremely minimal requirement on being a verisimilitude valuer, there is no way of assigning weights to propositions that is consistent with the inaccuracy measure being proper. So Oddie thinks we have a choice: give up on the inaccuracy measure being proper (and so give up on accuracy-first epistemology), or give up valuing verisimilitude. Both options are unattractive.

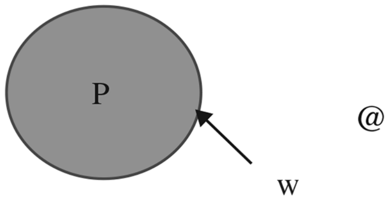
Here's my plan for bringing accuracy and verisimilitude back into harmony: first, I'll argue that Oddie's constraint is too strong. There are plenty of ways one can value verisimilitude without meeting Oddie's constraint (here I'll be echoing some considerations raised by Dunn ([2018])). Second, I'll offer some alternative constraints for what it takes to value verisimilitude and show that proper inaccuracy measures can satisfy them.

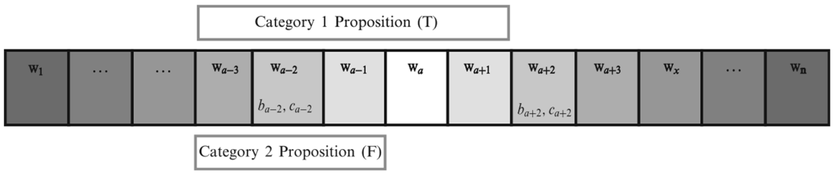
4. Oddie's Constraint

Oddie thinks that if an inaccuracy measure cares about verisimilitude, then it must satisfy the following constraint, which he calls 'weak proximity' (I'm going to argue that it's too strong, so I'll take the liberty of renaming it 'Oddie's proximity'):

> Oddie's Proximity: Consider a proposition P, thought of as a set of worlds in Ω. For any set of worlds, P, we can consider the members of that set that are closest to the actual world. We'll call each such world 'a most accurate P-world'. Let @ be the actual world and let w be a most accurate P-world (see Figure 4). Let b be a credence function that assigns a credence of one to w. Let c be a credence function that assigns non-zero credence to all and only the members of P. For any such b and c, b is at least as accurate as c at @.

[6] I'm going to be understanding this proposal as one on which the weighting of the propositions does not depend on which world is actual. Both Oddie ([2019]) and Levinstein ([2019]) provide compelling arguments for the claim that reconciling accuracy and verisimilitude on the assumption that the distance between worlds depends on which world is actual is hopeless.





Oddie's constraint is too strong. To see why, let's look at an example: Suppose that there are five possible worlds $w_1 \ldots w_5$ where w_i is a world containing exactly *i* planets, and suppose that the actual world is a world with three planets. We'll let the distance between any two worlds be the difference in number of planets between those worlds. Now consider two credence functions, *o* for 'opinionated' and *a* for 'ambivalent' (see Table 1):

(i) *o* assigns a credence of one to w_4;

(ii) *a* assigns a credence of 0.5 to w_4 and 0.5 to w_2.

Oddie's proximity constraint entails that *a* can't be doing better than *o* accuracy-wise.

But why does caring about verisimilitude mean that we can't think of *a* as doing better than *o* accuracy-wise? Oddie's thought, I take it, is this: *o* and *a* are doing equally well with respect to verisimilitude—they both invest all of their credence in worlds that are one unit away from the actual world. They also each invest the same amount of credence in the actual world (zero). So there are simply no grounds for thinking that *a* is doing better than *o*. But this line of thought requires more than thinking that verisimilitude is important. It essentially requires thinking that the accuracy of a credence function can't depend on any features of the credence function other than:

(i) how much credence is invested in the actual world;

(ii) how well it fares with respect to verisimilitude.

Table 1. The space of worlds where the distance between worlds is given by the difference in the number of planets in those worlds. *a* and *o* represent two credence functions defined over this space of worlds.

w_1: 1 planet	w_2: 2 planets $a(w_2) = 0.5$	$w_3 = @$: 3 planets	w_4: 4 planets $o(w_4) = 1$ $a(w_2) = 0.5$	w_5: 5 planets

But one might think that verisimilitude is valuable even if one thinks that additional features of the credence function are relevant to its accuracy. In particular, you might think that another relevant feature is:

> The evenness with which the credence function distributes its credence amongst non-actual worlds.

o and *a* differ with respect to this feature: *a* distributes her non-actual-worldly credence more evenly than *o*.

Whether it's plausible that evenness of distribution is relevant to accuracy is a question we'll come back to (I'll argue that it is; Dunn ([2018]) does so as well). But the thing to note for now is that many popular inaccuracy measures prefer credence functions that distribute credence more evenly amongst non-actual worlds. Valuing verisimilitude is compatible with also valuing evenness of distribution.

5. The Good

Oddie argues for the claim that verisimilitude can't be captured by proper inaccuracy measures by choosing a particular space of worlds and showing that there is no global inaccuracy measure that is proper and that satisfies his proximity principle for that space. Since Oddie's proof makes use of a particular space of worlds, Oddie hasn't shown that there are no spaces of worlds for which there are verisimilitude-valuing proper inaccuracy measures. However, Oddie chose a very natural space for his proof, which suffices to illustrate his point that there is a tension between propriety and his proximity constraint. To keep things tractable, I'm going to follow Oddie in this respect. I too will be looking at particular spaces of worlds (the same ones as Oddie) and I'll show that there are proper inaccuracy measures that value verisimilitude over those spaces. Since, like Oddie, my proofs will make use of particular spaces of worlds, I will not have shown that for every space of worlds, there is a proper inaccuracy measure that values verisimilitude. But my hope is that the results that follows will suffice to show that there is no reason to expect otherwise: there is no inherent tension between proper accuracy and verisimilitude.

5.1. Proximity over the disagreement metric (Result 1)

A minimal constraint on what's involved in valuing verisimilitude doesn't say that whenever one credence function is more verisimilar, it's more accurate. A minimal constraint says that, all else equal, the more verisimilar credence function is more accurate. This is the first sort of constraint we'll be looking for. As we proceed, we will generalize, and by the end we'll see exactly what needs to be held fixed to guarantee that the more verisimilar credence function is more accurate. (Roughly, the answer will be: we need to hold fixed how much credence is invested in the actual world and how evenly credence is distributed amongst certain falsehoods).

In this section I'll present a proximity constraint over spaces of worlds where the distance between worlds is given by what I'll call 'the disagreement metric'.

To get a sense of the metric, let's start with Oddie's example: We're wondering about the weather tomorrow. In particular, we're wondering whether it will be hot (H) and whether it will be rainy (R). The set of worlds we're considering consists of the four possible answers to these questions: $H\&R, H\& \sim R, \sim H\&R, \sim H\& \sim R$, and the distance between any two worlds is given by the number of disagreements between those worlds concerning the propositions H and R.[7] If the actual world (@) is $H\&R$, and we let D stand for the distance from @, we can represent the space as shown in Table 2.

The first replacement for Oddie's proximity that I will offer, which I'll call 'Proximity 1', will apply to any space of worlds for which there is some set of propositions (which we'll call 'the atomic propositions') such that the distance between any two worlds is given by the number of disagreements between those worlds with respect to the atomic propositions.

Proximity 1 says roughly this: Suppose we have a finite set of worlds Ω that b and c distribute their credence over, where the distance between worlds is given by the disagreement metric. Now let's hold everything fixed between b and c, except for verisimilitude. In particular, we'll suppose that b and c are identical distributions over Ω except for the fact that there's at least one pair of worlds such that b and c's credences are swapped between these worlds, with b investing the larger credence in the closer world, and c investing the larger credence in the further world. Also, in the interest of holding everything except for the increased distance of c's world fixed, we'll assume that the further world that c invests the larger credence in disagrees with the actual world about all the atomic propositions that the closer world does, in addition to at least one other atomic proposition (hence making it further).[8] Then, what our constraint will require is that b is more accurate than c.

Before the official formulation, an example: suppose b distributes its credence as in Table 3. Now, consider c, which is just like b except that the credences in w_2 and w_4 are swapped, so that c is investing the larger credence (0.5) in the more distant world (w_4), and the smaller credence (0.1) in the closer world (w_2) (see Table 4). Proximity 1 requires that b is more accurate than c at @.

> Proximity 1: Let b and c be credence functions defined over a finite set of worlds Ω, where the distance between worlds in Ω is given by the disagreement metric. Let w_a

[7] Two worlds disagree about a proposition if one is a member of that proposition and the other is not. In many cases, multiple sets of propositions will yield essentially the same metric. In the case above, for example, we can use $\{H, R\}$, $\{\sim H, \sim R\}$, or $\{H, \sim H, R, \sim R\}$ (in which case, the distance between any two worlds will be doubled). There is no need, for our purposes, to choose 'the' set of atomic propositions corresponding to a metric. The fact that multiple sets will yield the same metric just shows that there can be multiple ways of assigning importance to propositions that all correspond to the same way of thinking about the distance between worlds.

[8] The necessity of and motivation for this condition will be discussed further in Sections 6 and 7.

Table 2. A space of worlds given by the disagreement metric. In this case, the distance between two worlds is given by the number of disagreements between those worlds with respect to propositions H and R.

$H\&R$ ($w_1 = @$) $D = 0$	$H\& \sim R$ (w_2) $D = 1$
$\sim H\&R$ (w_3) $D = 1$	$\sim H\& \sim R$ (w_4) $D = 2$

be any world in Ω and suppose that the multiset $\{b(w)|w \in \Omega\}$ can be mapped one-to-one onto the multiset $\{c(w)|w \in \Omega\}$ by the function F as follows:

(i) If $b(w) = c(w)$, then $F(b(w)) = c(w)$
(ii) If $b(w) \neq c(w)$, then for some world $w^*, F(b(w)) = c(w^*)$ and $F(b(w^*)) = c(w)$ and the following conditions are satisfied:

(a) The distance between w^* and w_a differs from the distance between w and w_a.
(b) b and c's credences are swapped between w and w^*, with b investing the larger credence in the closer world (to w_a) and the smaller credence in the further world.
(c) The further of the two worlds (w and w^*) from w_a disagrees with w_a about all the atomic propositions that the closer of the two worlds disagrees with w_a about, in addition to disagreeing with w_a about at least one other atomic proposition (hence making it further).

Then b is at least as accurate as c at w_a, and if (ii) holds for at least one $w \in \Omega$, b is more accurate than c at w_a.

A strictly proper inaccuracy measure based on the much-loved 'Brier score' satisfies Proximity 1.

To measure the local Brier inaccuracy of a credence c in a proposition P, we take the difference between c and the truth value of P and square it.

Table 3. A distribution, b, over four worlds where distance is given by the disagreement metric.

$H\&R(w_1 = @)$ $b(w_1) = 0.2$	$H\& \sim R(w_2)$ $b(w_2) = 0.5$
$\sim H\&R(w_3)$ $b(w_3) = 0.2$	$\sim H\& \sim R(w_4)$ $b(w_4) = 0.1$

Table 4. b and c swap credences between w_2 and w_4 with b investing the larger credence in the closer world and c investing the larger credence in the further world.

$H\&R(w_1 = @)$ $b(w_1) = 0.2$ $c(w_1) = 0.2$	$H\& \sim R(w_2)$ $b(w_2) = 0.5$ $c(w_2) = 0.1$
$\sim H\&R(w_3)$ $b(w_3) = 0.2$ $c(w_3) = 0.2$	$\sim H\& \sim R(w_4)$ $b(w_4) = 0.1$ $c(w_4) = 0.5$

$$I_{l-brier}(c, 1) = (1 - c)^2,$$
$$I_{l-brier}(c, 0) = (0 - c)^2 = c^2.$$

Recall that the global inaccuracy of a credence function c is a weighted sum of the local inaccuracy scores that c receives for each proposition that c assigns credence to. Let's call the weights λ_i. Where $w(P_i)$ is the truth value of a proposition P_i at world w, the global Brier inaccuracy of a credence function c at a world w is:

$$I_{g-brier}(c, w) = \sum_i \lambda_i I_{l-brier}(c(P_i), w(P_i)).$$

For a space of worlds where the distance between worlds is given by the disagreement metric, what I'll call the 'atomic Brier score' is the global Brier score that gives equal weight to all the atomic propositions and no weight to any other propositions.

Lemma
The atomic Brier score satisfies Proximity 1. The proof is in the appendix.

The more general result applies not just to the atomic Brier but to any atomic inaccuracy measure (an inaccuracy measure that gives equal weight to all the atomic propositions and no weight to any other propositions) derived from a local inaccuracy measure which satisfies the following constraints:

> Truth-Directedness: For all $r \in [0, 1]$, $I_{local}(r, 0)$ is a strictly increasing function of r.
>
> Symmetry: For all $r \in [0, 1]$, $I_{local}(r, 0) = I_{local}(1 - r, 1)$ and $I_{local}(r, 1) = I_{local}(1 - r, 0)$.

Result 1
Every global atomic inaccuracy measure derived from a local inaccuracy measure that satisfies truth-directedness and symmetry satisfies Proximity 1. The proof is in the appendix.

At this point, what we have is a verisimilitude-valuing constraint across spaces of worlds where distance between worlds is given by the disagreement metric that proper inaccuracy measures can satisfy. We get this result by letting our inaccuracy measure privilege exactly those propositions that are distance-determining: the atomic ones.

The atomic Brier score assigns all of its weight to the distance-determining propositions. In the real world, though, we presumably care at least a bit about accuracy with respect to every proposition.[9] So the more general thought I'm advocating for is this: the more weight we give to the distance-determining propositions, the more important it will be to be verisimilar. The results in this article support this thought by showing that, in the extreme case, when the privileged propositions get all the weight (and various other things are held fixed), the more verisimilar credence function is guaranteed to be more accurate. As the weights of other propositions get closer to the weights of the distance-determining ones, verisimilitude will still be a factor in determining accuracy, but it will be competing with another factor: the importance of being right with respect to other matters we think are important.[10] For simplicity, for the remainder of the article I'll continue to focus on accuracy measures which assign all their weight to the distance-determining propositions. But this should be thought of as an idealization.

5.2. Proximity over the magnitude metric (Result 2)

Let's now consider a different sort of metric, what we'll call the 'magnitude metric'. On the magnitude metric, the distance between two worlds is given by the difference in the magnitude of some quantity between those worlds. For example, the quantity might be the number of planets. Pretending that the actual world is one in which there are three planets, and letting *d* represent distance to the actual world, our space of worlds might look like Table 5.

We can apply Proximity 1 to the magnitude metric by noting that the magnitude metric is also a disagreement metric. Let w_i be the world where the value of the quantity in question is *i*. The distance between w_i and w_j on the magnitude metric is equal to the number of disagreements between w_i and w_j with respect to a certain class of propositions. Which propositions? We can use what I'll call the 'at-most' propositions, what I'll call the 'at-least' propositions, or both. The at-most propositions are propositions of the form 'there are at-most *m* of quantity *Q*' and the at-least propositions

[9] Another reason it's important to assign positive weight to all propositions is that, as Dunn ([2018], Footnote 27) points out, we won't be able to apply the accuracy arguments for probabilism if certain propositions are given zero weight. The good news is that because the arguments for probabilism are dominance arguments, all that's needed for them to work is that each proposition gets some positive weight: the amount doesn't matter.

[10] My suspicion is that as the weight of various 'non-privileged' propositions increases, the strength of our intuitions about which worlds are more similar to which will start to diminish (we'll say things like, 'well sure, these two worlds are similar in these important respects, but they're quite different in these other important respects'). If the increase is extreme enough, we might start favouring another metric altogether. So the strength of our intuitive judgements concerning the importance of verisimilitude with respect to some particular metric will go hand in hand with how important we think it is to be right with respect to the propositions that are distance-determining with respect to that metric.

Table 5. A space of worlds in which distance between worlds is given by the difference in the number of planets between them. D represents distance from the actual world.

w_1: 1 planet	w_2: 2 planets	$w_3 = @$: 3 planets	w_4: 4 planets	w_5: 5 planets
$D = 2$	$D = 1$		$D = 1$	$D = 2$

are propositions of the form 'there are at least m of quantity Q'.[11] By using these as our privileged propositions, the magnitude metric will satisfy a constraint very similar to Proximity 1. Once again, the idea is to offer a constraint that says, holding all else fixed, the more verisimilar credence function is more accurate. In this case our constraint says roughly that if the only difference between b and c is that there's a credal swap between two worlds, with b investing the larger credence in the closer world, and both worlds are 'on the same side' of the actual world (the importance of 'same-sidedness' will be discussed later) b will be more accurate. So, for example, suppose b and c are as in Table 6. Our constraint will require that b is more accurate than c.

> Proximity 2: Let b and c be credence functions defined over a finite set of worlds Ω, where the distance between worlds in Ω is given by the magnitude metric. Let w_a be any world in Ω and suppose that the multiset $\{b(w_i)|w_i \in \Omega\}$ can be mapped one-to-one onto the multiset $\{c(w_i)|w_i \in \Omega\}$ by the function F as follows:
>
> (i) If $b(w_i) = c(w_i)$, then $F(b(w_i)) = c(w_i)$.
> (ii) If $b(w_i) \neq c(w_i)$ then for some world w_j, $F(b(w_i)) = c(w_j)$ and $F(b(w_j)) = c(w_i)$ and the following conditions are satisfied:
>
> (a) The distance between w_j and w_a differs from the distance between w_i and w_a.
> (b) b and c's credences are swapped between the two worlds (w_i and w_j), with b investing the larger credence in the closer world (to w_a) and the smaller credence in the further world (from w_a).
> (c) i and j are both greater than a, or i and j are both less than a.
>
> Then b is at least as accurate as c at w_a, and if (ii) holds for at least one $w_i \in \Omega$, b is more accurate than c at w_a.

Result 2

Every global inaccuracy measure which assigns equal weight to all the at-most propositions and no other propositions, all the at-least propositions and no other propositions

[11] Here's why this works: On the magnitude metric, for any x and y, the distance between w_x and w_y is $|y - x|$. The propositions amongst the at-most propositions that w_x and w_y will disagree about are all and only propositions of the form 'there are at most i of quantity Q' with $x \leq i < y$. There are $|y - x|$ such propositions. So the distance between any two worlds on the magnitude metric is just the distance between any two of those worlds on the disagreement metric, where the atomic propositions are the at-most propositions. A similar argument can be made for at least propositions. If we wanted to use both types of propositions, that would work as well, so long as we're happy with a variant according to which the distance between w_x and w_y is $2(|y - x|)$. I think we should regard the magnitude metric and this variant as equivalent for the purposes at hand.

Table 6. b and c are identical except for the fact that they swap credences between w_4 and w_5 with b investing the larger credence in the closer world and c investing the larger credence in the further world. This makes b a more verisimilar credence function than c.

w_1: 1 planet	w_2: 2 planets	$w_3 = @$: 3 planets	w_4: 4 planets	w_5: 5 planets
$b(w_1) = 0.4$	$b(w_2) = 0.2$		$b(w_4) = 0.3$	$b(w_5) = 0.1$
$c(w_1) = 0.4$	$c(w_2) = 0.2$		$c(w_4) = 0.1$	$c(w_5) = 0.3$

or both the at-most and at-least propositions and no other propositions, and which is derived from a local inaccuracy measure that satisfies truth-directness and symmetry satisfies Proximity 2. The proof is in the appendix.

6. The Bad and the Ugly (Result 3)

So far we've established that for at least certain metrics (the ones Oddie considers), we can construct proper inaccuracy measures with the feature that, holding all else fixed, the more verisimilar credence function will be more accurate. But you might have wanted more. Consider once again the space of worlds given by the magnitude metric, with b and c distributed as in Table 7. Notice that the only difference between b and c is that b and c swap credences between w_2 and w_5, with b investing the larger credence in the closer world. It seems like we're holding an awful lot fixed, and just varying verisimilitude, and so it seems like a verisimilitude-valuing accuracy measure should rate b as more accurate than c. However, on the weighted Brier score which assigns equal weight to the at-most and at-least propositions and no weight to any other propositions, c is actually more accurate than b. (Note that this isn't a counterexample to Proximity 1 or Proximity 2 because condition (ii) (c) of both constraints is not satisfied for w_2 or w_5).

Why is this happening? This is where things get a bit ugly. Even though b invests more credence in closer worlds, what c has going for it is that c has clumped the largest and the smallest credences together on one side of the actual world (0.91 and 0.01), and the remaining two medium credences (0.03, 0.05) on the other.

Why on earth should this difference in clumping be relevant? The real answer comes later, but here's the quick answer: If you care about certain propositions more than others, then you should expect clumping to matter: If being wrong about some

Table 7. This time b and c swap credences between w_2 and w_4 with b investing the larger credence in the closer world and c investing the larger credence in the further world.

w_1: 1 planet	w_2: 2 planets	$w_3 = @$: 3 planets	w_4: 4 planets	w_5: 5 planets
$b(w_1) = 0.91$	$b(w_2) = 0.03$		$b(w_4) = 0.05$	$b(w_5) = 0.01$
$c(w_1) = 0.91$	$c(w_2) = 0.01$		$c(w_4) = 0.05$	$c(w_5) = 0.03$

proposition A is worse than being wrong about some proposition B, then credence functions that clump a bunch of credence in wrong-about-A worlds, are going to do worse than credence functions that divide up that same amount of credence between wrong-about-A worlds and wrong-about-B worlds. So privileging certain propositions means caring about clumping.

That being said, a feature of certain improper scoring rules (scoring rules that aren't proper) is that, holding fixed the amount of credence invested in the actual world, the more verisimilar credence function is always more accurate. In other words: if b and c invest equal credence in the actual world and b is more verisimilar, there is no other feature of the credal clumping that could make c as accurate or more accurate than b. The following constraint formalizes this idea:

> Proximity 3: Suppose that b and c are credence functions defined over a finite set of worlds Ω, that $w_a \in \Omega$, and that b and c invest equal amount of credence in w_a. For distance d, let X_d be the proposition (set of worlds) consisting of all and only worlds that are at least d units away from w_a: $X_d = \{w \in \Omega | D(w, w_a) \geq d\}$. If for all propositions X_d, $b(X_d) \leq c(X_d)$, but for some X_d, $b(X_d) < c(X_d)$, then b is more accurate than c at w_a.[12]

No version of the weighted Brier score satisfies Proximity 3.[13] But there are improper measures that satisfy Proximity 3. One is based on what's sometimes called the 'absolute value' score: The local absolute-value-inaccuracy of credence c in proposition P is the absolute value of the difference between c and the truth value of P:

$$I_{l-abv}(c, 1) = |1 - c| = 1 - c,$$
$$I_{l-abv}(c, 0) = |0 - c| = c.$$

The weighted absolute value score over a space of worlds Ω is a global inaccuracy measure that is a weighted sum of the local absolute-value-inaccuracy scores that a credence function gets for each proposition.

Result 3
When distance between worlds in Ω is given by the disagreement metric, the weighted absolute value score that assigns equal weight to all the atomic propositions and no other propositions satisfies Proximity 3. The proof is in the appendix.

Corollary
When distance between worlds in Ω is given by the magnitude metric, the weighted absolute value score that assigns equal weight to all the at-most propositions and no

12 Thanks to Kevin Dorst and Jack Spencer for help formulating this constraint.

13 In an unpublished version of his article (Oddie [2019]), Oddie's proof of Theorem 1 establishes that no weighted Brier score can satisfy his proximity principle, but the proof can also be used to show that no weighted Brier score can satisfy Proximity 3. I have not established that no proper score will satisfy Proximity 3, but I'm not optimistic that we'll find a Proximity-3-satisfying proper inaccuracy measure. Unlike Oddie's constraint, Proximity 3 does allow the inaccuracy measure to care about features of a distribution other than verisimilitude and the amount of credence invested in the actual world, but these other features will only get to play a role as a tie-breaker: when the two credence functions in question invest equal credence in the actual world and are doing equally well with respect to verisimilitude, then some other feature might make a difference. Proper measures, however, tend to care about features of a distribution (like its evenness) in a stronger-than-tie-breaking fashion.

other propositions, all the at-least propositions and no other propositions, or both the at-most and at-least propositions and no other propositions satisfies Proximity 3.

Proof of Corollary
We noted earlier that the magnitude metric is just an instance of the disagreement metric with the atomic propositions being the at-least propositions, the at-most propositions, or both. Thus, the corollary follows immediately from Result 3.[14]

7. Some More Good: The Role of Evenness of Distribution (Result 4)

We saw above that the weighted Brier score doesn't satisfy Proximity 3. Worse than that—as illustrated by the example in the previous section—on the weighted Brier score we've been working with, there are cases in which two credence functions distribute the very same multiset of credences amongst a set of non-actual worlds, b invests more credence in closer worlds, yet c is more accurate. This can happen because the weighted Brier score is sensitive to how these credences are clumped amongst the propositions that we care about. And unlike the absolute value score, verisimilitude is not the only clumping feature that's important. The good news is that we can articulate exactly what other clumping feature our Brier score cares about, and show that, holding that feature fixed, the more verisimilar credence function is more accurate.

What is this special feature? I hinted at it in an earlier section: in general, proper scoring rules care about how evenly credence is distributed amongst falsehoods. Even before we get into the business of privileging certain propositions over others, we can note that on a proper inaccuracy measure, a credence function that assigns 0.5 to non-actual-world-1, and 0.5 to non-actual-world-2, will be more accurate than a credence function that assigns a credence of one to a single non-actual world. According to proper inaccuracy measures, it's better to hedge.

But notice that we can't explain why c is more accurate in the example above by appealing to the claim that c distributes its credence in (non-actual) worlds more evenly than b does. The distributions over non-actual worlds for both b and c contain exactly the same credences! However, once we privilege certain propositions, what matters is

[14] Recall that one motivation for an account in which the value of verisimilitude is built into (rather than distinct from) the value of accuracy is that we don't want agents to be manipulating their credences in an attempt to trade these values off one another. For example, we don't want someone with

$$Pr_1(\text{eight planets}) = 0.9, \quad Pr_1(\text{one planet}) = 0.1$$

to move to

$$Pr_2(\text{eight planets}) = 0.9, \quad Pr_2(\text{seven planets}) = 0.1$$

in an attempt to be more verisimilar. Using an improper measure like the absolute value score, however, does not help with this problem. Someone with Pr_1 will, in fact, regard Pr_2 as having higher expected accuracy using the weighted absolute-value score. So we only get around this problem by using a proper measure. Although it's true that on verisimilitude-valuing proper measures (like the weighted Brier), Pr_2 is in fact more accurate than Pr_1, from the perspective of Pr_1, Pr_2 is less expectedly accurate.

how evenly credence is distributed in the falsehoods concerning the privileged propositions. And, in fact, c does distribute its credence more evenly than b in the falsehoods concerning the privileged propositions. But to convince you of this, and of the fact that evenness of distribution explains why c does better, we need to formalize the notion of 'evenness of distribution'.

We can follow information-theorists here and think of the evenness of a credal distribution as the entropy of that distribution, where entropy is a measurement of how much information the distribution contains. Intuitively, the more evenly credence is distributed over a set of worlds, the less informative that distribution is (you have no idea how things will turn out) whereas if one world gets assigned maximal credence, that distribution is maximally informative.

The most common measure of entropy in information theory is Shannon entropy. Here's how it works: consider some credence function c and consider the most fine-grained propositions (the worlds) that c is distributed over: $\{w_1 \ldots w_n\}$. Let $c(w_i) = c_i$. The Shannon entropy of c is given by the following formula:

$$E_{shannon}(c) = -\sum_{i=1}^{n} c_i log(c_i).$$

There are some interesting connections between entropy and accuracy that will be important for our purposes. Let c^f be c restricted to the most fine-grained propositions that c is defined over. (So while c might assign some credence to the proposition $\{w_1, w_2\}$, c^f would only assign credence to $\{w_1\}$ and $\{w_2\}$ individually). The first thing to note is that $E_{shannon}(c)$ also represents how inaccurate c^f expects itself to be using the log inaccuracy measure: an inaccuracy measure that assigns the score log(c) to a credence of c in a falsehood (Grunwald and Dawid [2004]). So we can think of the Shannon entropy of c as c^f's expected inaccuracy relative to itself (where the global inaccuracy of c^f is the average of the local inaccuracy scores of c^f). Intuitively, this notion corresponds with 'evenness' because the more opinionated a credence function is (the more uneven), the more accurate it will expect itself to be. (In the extreme case, the maximally opinionated distribution is certain that it will get the maximal accuracy score.)

We can define different entropy ('evenness') measures using different inaccuracy measures. So, for example, we can define the Brier entropy of c as the expected inaccuracy of c^f relative to itself, according to the (evenly weighted) Brier score. This equals:

$$E_{brier}(c) = \sum_{i=1}^{n} c_i I_{g-brier}\left(c^f, w_i\right) = 1 - \left[\frac{\sum_{i=1}^{n} c_i^2}{\left(\sum_{i=1}^{n} c_i\right)^2}\right].[15]$$

[15] Thanks to Richard Pettigrew for pointing this out to me. Here's the proof he provided:

$$\begin{aligned}\sum c_i I_{g-brier}(c^f, w_i) &= \sum c_i(c_1^2 + c_2^2 + \ldots c_{i-1}^2 + (1 - c_i)^2 + c_{i+1}^2 + \ldots c_n^2]\\ &= \sum c_i(1 - 2c_i + \sum c_i^2) = \sum c_i - 2\sum c_i^2 + \sum c_i \sum c_i^2.\end{aligned}$$

Here's a neat fact about Brier-entropy: using Brier-entropy to measure evenness is equivalent to using Jain's fairness index, a measure used in the distributive justice literature to measure the fairness of a distribution of goods (Jain et al. [1984]).

So far this has all concerned evenness of a distribution amongst worlds. But once we start regarding certain propositions as more important than others, what's relevant is how evenly credence is distributed amongst the propositions that contribute to a credence function's inaccuracy. Suppose we have a weighted-Brier score that assigns equal weight to the set of propositions P_i and no weight to any other propositions. Then we'll let:

$$E_{weighted-brier}(c) = 1 - \left[\frac{\sum c(P_i)^2}{\left(\sum c(P_i)\right)^2}\right].$$

Call the propositions P_i and their negations the 'privileged propositions'. Let F^{w} be the subset of the privileged propositions that are false at w. Call these the 'inaccuracy-determining propositions at w'. We can now state another proximity constraint that a weighted Brier score can satisfy. It will say that holding fixed evenness of distribution (as defined by Brier-entropy) amongst the inaccuracy-determining propositions at w, the more verisimilar credence function is more accurate at w.[16]

> Proximity 4: Suppose that b and c are credence functions defined over a finite set of worlds Ω, that $\mathrm{w}_a \in \Omega$, and that b and c invest equal amount of credence in w_a. Suppose also that b distributes its credence amongst the inaccuracy-determining propositions at w_a at least as evenly as c as defined by Brier-entropy. For distance d, let X_d be the proposition (set of worlds) consisting of all and only worlds that are at least d units away from w_a: $X_d = \{\mathrm{w} \in \Omega | D(\mathrm{w}, \mathrm{w}_a) \geq d\}$. If for all propositions X_d, $b(X_d) \leq c(X_d)$, but for some X_d, $b(X_d) < c(X_d)$ then b is more accurate than c at w_a.

Suppose that b and c are credence functions defined over a finite set of worlds Ω, that $\mathrm{w}_a \in \Omega$, and that b and c invest equal amount of credence in w_a. Suppose also that b distributes its credence amongst the inaccuracy-determining propositions at w_a at least as evenly as c as defined by Brier-entropy. For distance d, let X_d be the proposition (set of worlds) consisting of all and only worlds that are at least d units away

Since $\Sigma c_i = 1$, this equals $1 - 2\Sigma c_i^2 + \Sigma c_i^2 = 1 - \Sigma c_i^2$. Since $(\Sigma c_i)^2 = 1$, this equals $1 - [\Sigma c_i^2/(\Sigma c_i)^2]$. The equality also holds for credences distributed over a proper subset of Ω and so not summing to one. Simply let $K = \Sigma c_i$. Then we can define $c^+_{\mathrm{w}_i}$ as c_i/K to get a normalized version of c and calculate the Brier entropy of c^+. This will also yield $1 - [\Sigma c_i^2/(\Sigma c_i)^2]$.

16 Note that thinking of the evenness of a multiset of credences invested in the inaccuracy-determining propositions as the Brier entropy with respect to those propositions can't be motivated directly by appeal to the credence function's expected inaccuracy. Since the privileged propositions need not be exclusive, the expected inaccuracy of a credence function with respect to these propositions will not equal the Brier-entropy of this set of credences. That's okay. The idea is that one way to motivate thinking of the evenness of a multiset of credences as the Brier-entropy of those credences is to note that we'd expect evenness and expected inaccuracy to go together when the set of possibilities is exclusive. But once we have our measure of evenness we can apply it to credences in propositions that aren't exclusive as well. (And indeed, as Jain *et al.* do, we can use it to measure the evenness of a distribution of goods which aren't credences at all).

from $w_a : X_d = \{w \in \Omega | D(w, w_a) \geq d\}$. If for all propositions X_d, $b(X_d) \leq c(X_d)$, but for some X_d, $b(X_d) < c(X_d)$ then b is more accurate than c at w_a.

Result 4
When distance between worlds in Ω is given by the disagreement metric, the weighted Brier score that assigns equal weight to all the atomic propositions and no weight to any other propositions satisfies Proximity 4. When distance between worlds in Ω is given by the magnitude metric, the weighted Brier score that assigns equal weight to all the at most propositions and no other propositions, all the at least propositions and no other propositions, or both the at-most and at-least propositions and no other propositions satisfies Proximity 4. The proof is in the appendix.

Here's where we are so far: we've seen that on two natural metrics, the weighted Brier score can satisfy the following constraint: holding fixed evenness of distribution across the privileged propositions, and the amount of credence invested in the actual world, the more verisimilar credence function is more accurate.[17] Is this good enough? I address a potential worry in the next section.

8. Some More Bad: Which Propositions to Privilege? (Result 5)

At this point you might think the following: it's obvious to me that I care about verisimilitude, and it's obvious to me that I care about accuracy. But I doubt any amount of soul-searching will reveal in me a special fondness for at-most propositions. So even if technically we can squish the value of verisimilitude into the value of accuracy, this maneuver misrepresents the phenomenon. The phenomenon is that I care about two things: being accurate and being verisimilar. The phenomenon is not that I care about one thing: being accurate about at-most propositions.

In fact, you might think, a much more natural class of propositions to privilege on the magnitude metric would be the set of convex propositions: sets of worlds with no 'gaps' like, for example, 'there are between five and eight planets'. So one question to ask is: what sort of proximity constraints are satisfied if the privileged propositions are the convex propositions rather than the at-least or at-most propositions? There is at least one such constraint:

> Proximity 5: Let Ω be a finite space of worlds where distance between worlds is given by the magnitude metric. Let w_a be a world in Ω and let w_{a+d} and w_{a-d} be two worlds that are

[17] The reader may be wondering whether an analogue of Result 4 will hold for accuracy measures other than Brier by substituting corresponding measures of entropy (for example, would Result 4 hold for the log score if we use Shannon entropy?). No. What's special about the Brier score is that in the simplest case (where all propositions get equal weight), the Brier score cares about only two things: Brier entropy and the amount of credence invested in the actual world. The log score, in contrast, cares about more than Shannon entropy and the amount of credence invested in the actual world (two credence functions with the same Shannon entropy, and the same credence invested in the actual world can have different log scores). It's difficult to articulate exactly what other features the log score cares about and so it's hard to formulate a constraint that says, 'holding X, Y, and Z fixed, the more verisimilar credence function is more accurate'. For as far as I know, nobody has articulated what *X*, *Y*, and *Z* for the log score are. (Thanks to Richard Pettigrew for helpful discussion on this point.)

d units away from w_a. Suppose that b and c are credence functions that invest equal amounts of credence in w_a and which are such that b distributes its credence at least as evenly among non-w_a worlds as c does. If b invests all of its non-w_a credence in worlds that are d units away from w_a, c invests all of its non-w_a credence in worlds that are at least d units from w_a, and there is some world in which c invests credence that is more than d units away from w_a, b is less inaccurate than c at w_a.

Result 5
The global Brier score that assigns equal weight to all convex propositions and no weight to any other propositions satisfies Proximity 5. The proof is in the appendix.

This is a nice result about convex propositions. But it's a bit misleading. There are some examples that illustrate that the convex-proposition-lover positively defies verisimilitude. Take the case in Table 8. If all the convex propositions are weighted equally, b and c are equally accurate on the convexly weighted Brier score, even though b is clearly doing better with respect to verisimilitude. Furthermore, it's hard to see what other difference between the two credence functions could explain why c's weakness on the verisimilitude front doesn't make it less accurate. But note that this (for once!) is not propriety's fault! The convexly weighted absolute-value score also yields the result that b and c are equally accurate. The problem is not the type of inaccuracy measure we use, but the choice to privilege the convex propositions.

Here's the problem with convex propositions: worlds near the middle of the space show up in more convex propositions than worlds near the edges. In the example above, w_1 and w_4 are each members of four convex propositions, whereas w_2 and w_3 are each members of six convex propositions. This means that if the actual world is near one edge, and you care about convex propositions, then you want to invest your non-actual-worldly credence in worlds that are either close to the actual world (for verisimilitude reasons) or near the opposite edge—very far from the actual world (because those worlds won't show up in many propositions). In this particular case, the fact that w_3 is more verisimilar perfectly balances the fact that w_4 shows up in fewer propositions and this explains why b and c are equally accurate.[18]

If you still find yourself attached to convex propositions, let me make two further potentially assuaging remarks. First, if you like convex propositions, you can take some comfort in the fact that every convex proposition is a conjunction of an at-least and at-most proposition. It's just that taking the at-least and at-most propositions instead of the convex ones as privileged doesn't have the unfortunate consequence that worlds near the middle get counted more than worlds near the edge. At-least and at-most propositions are egalitarian: each world shows up in the same number of propositions.

Second, I imagine that the temptation to think that convex propositions are especially important comes from the natural thought that propositions of the form 'there are about x planets' are important. And you might think that the convex propositions

[18] I believe that the reason Proximity 5 gets around this problem is that the conditions under which the constraint applies have the feature the closer the actual world is to one edge, the further from the middle are the worlds in which b invests its credence.

Table 8. Two distributions over a space of world given by the magnitude metric. *b* and *c* swap credences between w_3 and w_4 with *b* investing the larger credence in the closer world and *c* investing the larger credence in the further world.

$w_1 = @$: 1 planet	w_2: 2 planets	$w_3 = @$: 3 planets	w_4: 4 planets
		$b(w_3) = 0.9$	$b(w_4) = 0.1$
		$c(w_3) = 0.1$	$c(w_4) = 0.9$

are simply propositions expressing different 'there are about *x* planets' propositions. But that's not right. For it's natural to think of 'there are about *x* planets' propositions as propositions of the form 'there are *x*, plus or minus *y* planets'. This is not the set of convex propositions: the convex proposition $\{w_2, w_3\}$, for example, is not a member of this set. And *b* is, in fact, more accurate than *c* with respect to the 'there are *x*, plus or minus *y* planets' propositions on the weighted Brier, as we'd expect.

Here's the more general methodological point: The project is to see whether we can find some propositions, which intuitively are important, and some distance relation between worlds, which intuitively seems right, and show that privileging those propositions will amount to caring about verisimilitude as defined by that distance relation. It may turn out that we need to do a bit of massaging to get this to work. So insofar as soul-searching revealed in you a preference for convex propositions, my proposal is that your soul-searching ever so slightly missed the mark in terms of what features of credence functions you value. You don't presumably think it's great to invest your non-actual-worldly credence in a very distant world just because that world is near an edge, so, it turns out, it's not actually being right about convex propositions that you care about.

One final note: Which propositions we care about will plausibly vary from context to context. If you're hosting a party and trying to figure out how many chairs you need, you may care a lot about propositions like 'at most fifteen people will come' or 'about fifteen people will come'. If, however, you're trying to figure out whether to play a game at your party that will require people to be paired up, you may care more about propositions like 'an even number of people will come'. Relative to your game-planning goal, the metric shouldn't be the magnitude metric: the world in which fifteen people come should be regarded as more similar to the world in which thirteen people come, than to the world in which fourteen people come. We can still use a disagreement metric for these purposes (let the atomic proposition be 'an even number of people will come) and so the results in this article will still apply. So there's no problem incorporating the thought that different propositions matter to different degrees in different contexts—the relevant metric will change as the importance of propositions changes. So long as we can capture the resulting metric as a disagreement or magnitude metric, everything I've said so far will apply. If the metric is more complicated, which in many real-world cases it likely is, I have not proven that propriety and verisimilitude valuing are compatible (nor has Oddie shown them to be incompatible). But as I said earlier, my hope is to have convinced you that there's no reason to expect otherwise. There's no inherent tension between propriety and verisimilitude.

9. Concluding Thoughts: Accuracy and Practical Value

I've argued that proper measures can capture the value of verisimilitude by privileging certain propositions over others. The key difference between proper and improper measures on the verisimilitude front is that proper measures care about features of the way credence is distributed amongst non-actual worlds that aren't just verisimilitude related. They also care about evenness of distribution and, in particular, they care about how evenly credence is distributed amongst the inaccuracy-determining propositions. Indeed, holding fixed the amount of credence invested in the actual world, the weighted Brier score cares about only two things: verisimilitude and evenness across the inaccuracy-determining propositions (this follows from Result 4), whereas, holding fixed the amount of credence invested in the actual world, the weighted absolute value score cares only about verisimilitude.

At this point the question naturally becomes: are there good reasons to care about evenness of distribution? Dunn ([2018]) has argued that there are. I agree, and want to add one more reason to the list of reasons we might care about evenness: from a practical perspective, you're better off having your credences in falsehoods more evenly distributed. More precisely: suppose S has a certain amount of credence that she's going to invest in the false propositions P and Q that she may have to bet on. Suppose we (who know that P and Q are both false) don't know what sorts of bets S will face: we have a uniform distribution over the possible odds. A result from (Schervish [1989]) (which has been explicated in extremely helpful ways by Levinstein ([2017])) implies that S is better off dividing whatever sum of credence she is investing in these falsehoods as evenly as possible. Why? Because (given our ignorance of which bet she'll face) the expected amount of money she will lose from having credence x in P (a false proposition) is $x^2/2$ (see Levinstein [2017]) and the expected amount of money she will lose from having credence y in P is $y^2/2$. Thus, holding $x + y$ fixed, the quantity that S will want to minimize is $x^2 + y^2$. $x^2 + y^2$ is minimized when x and y are maximally evenly distributed, as defined by Brier entropy.[19] So it's no coincidence that we care about evenness: we'll expect agents that distribute more evenly to do better practically.

But agents who are more verisimilar will also do better practically so long as we assume that the privileged propositions are the ones that agents are going to bet on (in the broadest sense of 'betting'). Take the weather example we started with: if you go to stores that sells umbrellas (which are useful when it rains) but you don't

[19] For recall, Brier entropy in this case would equal $1 - [(x^2 + y^2)/(x + y)^2]$, so holding fixed the value of $x + y$, maximizing evenness amounts to the same thing as minimizing $x^2 + y^2$.

go to stores that sell objects that are useful in circumstances in which 'it rains if and only if it's not hot', then, all else equal, we'll expect the more verisimilar agent to do better practically.

The weighted Brier score (amongst its many other virtues!) does a good job at capturing the features that will matter to us practically. For holding fixed the amount of credence invested in the actual world, the weighted Brier score cares about only two other things: verisimilitude, and evenness with respect to the privileged propositions (as defined by Brier entropy). Evenness and verisimilitude are also exactly the features of a distribution in non-actual worlds that are important practically, when we're completely ignorant about which sorts of decisions we'll face.[20]

Appendix

Note that the Proximity constraints referred to in the results get stated in the course of the proofs.

Lemma
The atomic Brier score satisfies Proximity 1.

Proof of Lemma
Let b and c be credence functions defined over a finite set of worlds Ω, where the distance between worlds in Ω is given by the disagreement metric. Let w_a be any world in Ω and suppose that the multiset $\{b(\mathrm{w})|\mathrm{w} \in \Omega\}$ can be mapped one-to-one onto the multiset $\{c(\mathrm{w})|\mathrm{w} \in \Omega\}$ by the function F as follows:

(i) If $b(\mathrm{w}) = c(\mathrm{w})$, then $F(b(\mathrm{w})) = c(\mathrm{w})$.

(ii) If $b(\mathrm{w}) \neq c(\mathrm{w})$, then for some world $\mathrm{w}^*, F(b(\mathrm{w})) = c(\mathrm{w}^*)$ and $F(b(\mathrm{w}^*)) = c(\mathrm{w})$ and the following conditions are satisfied:

  (a) The distance between w^* and w_a differs from the distance between w and w_a.

  (b) b and c's credences are swapped between w and w^*, with b investing the larger credence in the closer world (to w_a) and the smaller credence in the further world.

  (c) The further of the two worlds (w and w^*) from w_a disagrees with w_a about all the atomic propositions that the closer of the two worlds disagrees with w_a about, in addition to disagreeing with w_a about at least one other atomic proposition (hence making it further).

[20] This assumes that we're sure to only be betting on the privileged propositions. Insofar as you might bet on non-privileged propositions like, 'it's hot if and only if it doesn't rain', but are less likely to do so, we need to say something more complicated: that being verisimilar will be a pro-tanto practical benefit of a credence function, but one that could, potentially, be outweighed by being sufficiently accurate about alternative propositions (see also the last paragraph of Section 4.1).

We'll show that on the weighted Brier score that assigns equal weight to all the atomic propositions and no weight to any other propositions, b is at least as accurate as c at w_a, and if (ii) holds for at least one $w \in \Omega$, b is more accurate than c at w_a.

Let the falsehoods concerning the atomic propositions at w_a be $\{P_1 \ldots P_m\}$. (In other words, if the atomic propositions are $\{A_1 \ldots A_m\}$, then if $w_a \in A_i, P_i = \sim A_i$, and if $w_a \notin A_i$ then $P_i = A_i$.) The inaccuracy of b at w_a on the weighted Brier score is just the sum of the inaccuracy of b with respect to the P_i.[21]

Now, for any such proposition P_i, consider those worlds w in P_i such that b invests a larger credence in w than c does. Call these worlds $w_{i1} \ldots w_{im}$ (the i is just a reminder that we're listing worlds that are members of P_i).

So we have that for $j \in \{1 \ldots m\}$, $b(w_{ij}) > c(w_{ij})$. We know that for any such w_{ij}, there exists a partner world, which we'll call w^*_{ij}, such that w^*_{ij} is further from w than w_{ij} is, and which is such that $b(w_{ij}) = c(w^*_{ij})$, and $c(w_{ij}) = b(w^*_{ij})$. Now note that w^*_{ij} is also a member of P_i. Why? Because w^*_{ij} is the further of the two partners from w_a, and we've stipulated that the further world disagrees with w_a about all the atomic propositions that the closer world disagrees with w_a about. Since $w_{ij} \in P_i$ but $w_a \notin P_i$, w^*_{ij} must be a member of P_i as well.

Now we'll order the worlds that are members of P_i as follows: First will come the pairs of worlds, w_{ij}, w^*_{ij}, where $b(w_{ij}) > c(w_{ij})$, and then will come all the remaining worlds which will be such that $b(w_{ij}) \leq c(w_{ij})$.

So,

$$\begin{aligned} I_{l-brier}(b(P_i), w_a(P_i)) = {} & [b(w_{i1}) + b(w^*_{i1}) + \ldots b(w_{im}) + b(w^*_{im}) \\ & + b(w_{i(m+1)}) + \ldots b(w_{i(m+n)})]^2 \end{aligned}$$

For $j \in \{1 \ldots m\}$ we can swap $b(w_{ij})$ with $c(w^*_{ij})$ and $b(w^*_{ij})$ with $c(w_{ij})$. So b's inaccuracy with respect to P_i:

$$= [c(w^*_{i1}) + c(w_{i1}) + \ldots c(w^*_{im}) + c(w_{im}) + b(w_{i(m+1)}) + \ldots b(w_{i(m+n)})]^2.$$

And recalling that by construction of our ordering, for all $j \in \{m+1 \ldots m+n\}$, $b(w_{ij}) \leq c(w_{ij})$, we have that the inaccuracy of $b(P_i)$ at w_a is

$$\begin{aligned} & \leq [c(w^*_{i1}) + c(w_{i1}) + \ldots c(w^*_{im}) + c(w_{im}) + c(w_{i(m+1)}) + \ldots c(w_{i(m+n)})]^2 \\ & = I(c(P_i), w_a(P_i)). \end{aligned}$$

So, for all P_i, the inaccuracy of b is less than or equal to the inaccuracy of c on the weighted Brier.

Suppose now that for some world $w \in \Omega$ condition (ii) obtains and $b(w) \neq c(w)$. Then there exists a partner world for w, w^*, such that b and c swap credences between these worlds with b investing the larger credence in the closer world. Without loss of

[21] Note that I'm relying here on the fact that the Brier score is symmetric.

generality, suppose w* is the further world. Then w* disagrees with w_a about all the atomic propositions that w disagrees with w_a about in addition to at least one other atomic proposition. So let P_z be a falsehood concerning an atomic proposition such that $w^* \in P_z$ but $w \notin P_z$.

Let's now think about *b*'s inaccuracy with respect to P_z. As before we can express *b*'s P_z-inaccuracy as:

$$I_{l-brier}(b(P_z), w_a(P_z) = [b(w_{z1}) + b(w^*_{z1}) + \dots b(w_{zm}) + b(w^*_{zm}) + b(w_{z(m+1)}) + \dots b(w_{z(m+n)})]^2.$$

Since $w^* \in P_z$,

$$w^* \in \{w_{z1}, w^*_{z1} \dots w_{zm}, w^*_{zm}, w_{z(m+1)} \dots w_{z(m+n)}\}.$$

Note that $w^* \notin \{w_{z1}, w_{z2} \dots w_{zm}\}$. This is because, for all w_{zj} where $j \in \{1 \dots m\}$, $b(w_{zj}) > c(w_{zj})$. However, since w* is the further world of w, w*, and *b* invests the smaller credence in the further world $b(w^*) < c(w^*)$.

Note also that $w^* \notin \{w^*_{z1}, w^*_{z2} \dots w^*_{zm}\}$. For the w^*_{zj} are all partners of the w_{zj} worlds. This means that if $w^* = w^*_{zj}$ for some $j \in \{1 \dots m\}$, then w**s* partner would be w_{zj} for some $j \in \{1 \dots m\}$. But w*'s partner is w, and since w (by assumption) is not a member of P_z, w can't equal any such w_{zj}.

It follows that $w^* \in \{w_{z(m+1)} \dots w_{z(m+m)}\}$.

Since we know that $b(w^*) < c(w^*)$ (w* is the further world), it follows that for some $w_{zj} \in \{w_{z(m+1)} \dots w_{z(m+m)}\}$, $b(w_{zj}) < c(w_{zj})$.

Since for all $j \in \{m + 1 \dots m + n\}$, $b(w_{zj}) \le c(w_{zj})$ and for some $j \in \{m + 1 \dots m + n\}$, $b(w_{zj}) < c(w_{zj})$, it follows that

$$b(w_{z(m+1)}) + b(w_{z(m+2)}) + \dots b(w_{z(m+n)}) < c(w_{z(m+1)}) + c(w_{z(m+2)}) + \dots c(w_{z(m+n)}) .$$

Returning to *b*'s inaccuracy with respect to P_z, we have:

$$\begin{aligned} I(b(P_z), w_a(P_z) &= [b(w_z1) + b(w^*_{z1}) + \dots b(w_zm) + b(w^*_{zm}) + b(w_{z(m+1)}) \\ &\quad + \dots\ b(w_{z(m+n)})]^2 = [c(w^*_{z1}) + c(w_{(z1)} + \dots c(w^*_{zm}) + c(w_{zm}) \\ &\quad + b(w_{z(m+1)}) + \dots b(w_{z(m+m)})]^2 < [c(w^*_{z1}) + c(w_{z1}) + \dots c(w^*_{zm}) \\ &\quad + c(w_{zm}) + c(w_{z(m+1)}) + \dots c(w_{z(m+m)})]^2 = I(c(P_z), w_a(P_z)). \end{aligned}$$

Since for all P_i, *b*'s inaccuracy with respect to P_i is less than or equal to *c*'s, but for some P_i, *b*'s inaccuracy is less than *c*'s, it follows that *b* is less inaccurate than *c* at w_a on the weighted-Brier.

Result 1
Every global atomic inaccuracy measure derived from a local inaccuracy measure that satisfies truth-directedness and symmetry satisfies Proximity 1.

Proof of Result 1

Let g be a strictly increasing function representing the local inaccuracy of a credence in a falsehood and let this local inaccuracy measure satisfy symmetry.[22] Then, in the proof of the lemma above, simply substitute any expression of the form $[\ldots]^2$ with $g[\ldots]$. □

Result 2

Every global inaccuracy measure that assigns equal weight to all the at-most propositions and no other propositions, all the at-least propositions and no other propositions, or both the at-most and at-least propositions and no other propositions, and that is derived from a local inaccuracy measure that satisfies truth-directedness and symmetry satisfies Proximity 2.

Proof of Result 2

Let b and c be credence functions defined over a finite set of worlds Ω, where the distance between worlds in Ω is given by the magnitude metric. Let w_a be any world in Ω and suppose that the multiset $\{b(\mathrm{w}_i)|\mathrm{w}_i \in \Omega\}$ can be mapped one-to-one onto the multiset $\{c(\mathrm{w}_i)|\mathrm{w}_i \in \Omega\}$ by the function F as follows:

(i) If $b(\mathrm{w}_i) = c(\mathrm{w}_i)$ then $F(b(\mathrm{w}_i)) = c(\mathrm{w}_i)$.

(ii) $b(\mathrm{w}_i) \neq c(\mathrm{w}_i)$ and there is some world w_j, such that the following conditions are satisfied:

- (a) The distance between w_j and w_a differs from the distance between w_i and w_a.
- (b) b and c's credences are swapped between the two worlds (w_i and w_j), with b investing the larger credence in the closer world (to w_a) and the smaller credence in the further world (from w_a).
- (c) i and j are both greater than a, or i and j are both less than a.

We'll show that on a weighted global score that assigns equal weight to the at-most propositions (propositions of the form 'there are most m of quantity Q') and no others, the at-least propositions (propositions of the form 'there are at least m of quantity Q') and no others, or both, and that is derived from a local score that satisfies truth-directness and symmetry, b is at least as accurate as c at w_a. If condition (ii) holds for some $\mathrm{w}_i \in \Omega$, b is more accurate than c at w_a. This will follow from Result 1, and from the fact that we can think of the magnitude metric as a kind of disagreement metric.

First, note that the distance between any two worlds on the magnitude metric is equal to the distance between any two worlds on a disagreement metric on which the atomic propositions are the at-most propositions (see Footnote 9 in the main text).

Second, note is that if x and y are both greater than a or both less than a (graphically: w_x and w_y are both to the right or both to the left of w_a), then the further of $\{\mathrm{w}_x, \mathrm{w}_y\}$ to w_a disagrees with w_a about all of the at-most propositions that the closer world disagrees with w_a about. Why? Suppose x and y are both greater than a and, without loss of generality, let $x < y$. Then w_x and w_a disagree with one another

[22] Footnote 19 explains the role that symmetry plays in the proof.

about all propositions of the form 'there are at most i of quantity Q' when $a \leq i < x$. Similarly, w_y and w_a disagree with one another about all propositions of the form 'there are at most i of quantity Q' when $a \leq i < y$. Because $y > x$, for every i such that $a \leq i < x$, it is also true that $a \leq i < y$. Thus, if x and y are both greater than a, every at-most proposition that w_a and w_x disagree about is a proposition that w_a and w_y disagree about. Parallel reasoning shows that the same holds if w_x and w_y are both less than a.

Because the magnitude metric is equivalent to the disagreement metric with the atomic propositions being the at-most propositions, it follows from Result 1 that any global inaccuracy measure that satisfies truth-directedness and symmetry, and that takes as privileged the at-most propositions, satisfies Proximity 2. Similar reasoning applies to measures that take as privileged the at-least propositions, and measures that take both the at-most and the at-least propositions as privileged. □

Result 3
When distance between worlds in Ω is given by the disagreement metric, the weighted absolute value score that assigns equal weight to all the atomic propositions and no other propositions satisfies Proximity 3.

Proof of Result 3
Suppose that b and c are credence functions defined over a finite set of worlds Ω: $\{w_1 \ldots w_n\}$ where distance between worlds is given by the disagreement metric and suppose that b and c invest equal amounts of credence in w_a (a world in Ω). For distance d, let X_d be the proposition consisting of all and only worlds that are at least d units away from w_a:

$$X_d = \{w \in \Omega | D(w, w_a) \geq d\}.$$

We'll show that on the weighted absolute value score, which assigns equal weight to all the atomic propositions and no weight to any other propositions, the following holds:

If for all propositions X_d, $b(X_d) \leq c(X_d)$, but for some X_d, $b(X_d) < c(X_d)$ then b is more accurate than c at w_a.

Let the falsehoods concerning the atomic propositions at w_a be $\{P_1 \ldots P_m\}$. (In other words, if the atomic propositions are $\{A_1 \ldots A_m\}$, then if $w_a \in A_i$, $P_i = \sim A_i$, and if $w_a \notin A_i$ then $P_i = A_i$.) The inaccuracy of b at w_a on the weighted absolute value score is just the sum of the inaccuracy of b with respect to the P_i:

$$I_{weighted-abv}(b, w_a) = \sum_{i=1}^{m} b(P_i).$$

Since $b(P_i) = \Sigma_{w \in P_i} b(w)$,

$$I_{weighted-abv}(b, w_a) = \sum_{i=1}^{m} b(P_i) = \sum_{i=1}^{m} \sum_{w \in P_i} b(w).$$

Now take any world $w \in \Omega$. Let $D(w, w_a) = d_w$ Since w is d_w units away from w_a, w disagrees with w_a about d_w atomic propositions. This means that $b(w)$ will show up d_w times in

$$\sum_{i=1}^{m} \sum_{w \in P_i} b(w),$$

once for each P_i that w is a member of. More generally, then, we can say that

$$I_{weighted-abv}(b, \mathrm{w}_a) = \sum_{\mathrm{w}\in\Omega} b(\mathrm{w})D(\mathrm{w}, \mathrm{w}_a).$$

Now recall that we're assuming that for all X_d (where X_d is the proposition consisting of worlds d or more units away from w_a) $b(X_d) \leq c(X_d)$. So, where Δ is the furthest distance any world is from w_a, we know that

$$\sum_{i=1}^{\Delta} b(X_d) \leq \sum_{i=1}^{\Delta} c(X_d).$$

Now note that if a world is one unit away from w_a, it will show up in exactly one X_d proposition (where d ranges between 1 and Δ), namely, X_1—the proposition consisting of worlds one or more units away. A world two units away from w_a will show up in exactly two such X_d propositions, namely, the proposition consisting of worlds that are at least one unit away (X_1), and the proposition consisting of worlds that are at least two units away (X_2). In general, for any w, if $D(\mathrm{w}, \mathrm{w}_a) = \delta$, then b(w) will show up in δ of the X_d propositions, with d ranging between one and Δ.

So,

$$\sum_{i=1}^{\Delta} b(X_d) = \sum_{\mathrm{w}\in\Omega} b(\mathrm{w})D(\mathrm{w}, \mathrm{w}_a) = I_{weighted-abv}(b, \mathrm{w}_a).$$

For the same reason,

$$\sum_{i=1}^{\Delta} c(X_d) = \sum_{\mathrm{w}\in\Omega} c(\mathrm{w})D(\mathrm{w}, \mathrm{w}_a) = I_{weighted-abv}(b, \mathrm{w}_a).$$

Since

$$\sum_{i=1}^{\Delta} b(X_d) \leq \sum_{i=1}^{\Delta} c(X_d),$$

it follows that

$$I_{weighted-abv}(b, \mathrm{w}_a) \leq I_{weighted-abv}(c, \mathrm{w}_a).$$

If the inequality is strict, strict inequality follows. □

Result 4

When distance between worlds in Ω is given by the disagreement metric, the weighted Brier score that assigns equal weight to all the atomic propositions and no weight to any other propositions satisfies Proximity 4. When distance between worlds in Ω is given by the magnitude metric, the weighted Brier score that assigns equal weight to all the at most propositions and no other propositions, all the at least propositions and no other propositions, or both the at-most and at-least propositions and no other propositions satisfies Proximity 4.

Proof of Result 4

Suppose that b and c are probability distributions over a finite set of worlds Ω, that $\mathrm{w}_a \in \Omega$ and that b and c invest equal amount of credence in w_a. Suppose also that b distributes its credence amongst the inaccuracy-determining propositions at w_a at least as evenly as c as defined by Brier-entropy.

For, distance d, let X_d be the proposition (set of worlds) consisting of all and only worlds that are at least d units away from $\mathrm{w}_a : X_d = \{\mathrm{w} \in \Omega | D(\mathrm{w}, \mathrm{w}_a) \geq d\}$. We'll show that if for all propositions X_d, $b(X_d) \leq c(X_d)$, but for some X_d, $b(X_d) < c(X_d)$ then b is more accurate than c at w_a.

Let the F_i be the inaccuracy-determining propositions at world w_a. By Result 3 and its corollary, we know that b is more accurate than c on the weighted absolute value score. Thus,

$$\sum b(F_i) < \sum c(F_i).$$

And so

$$\left(\sum b(F_i)\right)^2 < \left(\sum c(F_i)\right)^2.$$

If b is at least as evenly distributed as c amongst F_i, then

$$\left[\frac{\sum b(F_i)^2}{\left(\sum b(F_i)\right)^2}\right] \geq \left[\frac{\sum c(F_i)^2}{\left(\sum c(F_i)\right)^2}\right].$$

Since we've established that the denominator on the left is less than the denominator on the right, for b to be at least as evenly distributed as c, the numerator on the right must be greater than the numerator on the left:

$$\sum c(F_i)^2 > \sum b(F_i)^2.$$

But the terms on either side of the inequality are just the inaccuracy scores of c and b, respectively, according to the weighted Brier score at w_a. So b is more accurate than c at w_a on the weighted Brier score. □

Result 5
The global Brier score that assigns equal weight to all convex propositions and no weight to any other propositions satisfies Proximity 5.

Proof of Result 5
Let Ω be a finite space of worlds where distance between worlds is given by the magnitude metric. Let w_a be a world in Ω and let w_{a+d} and w_{a-d} be two worlds that are d units away from w_a. Suppose b and c are credence functions that invest equal amounts of credence in w_a and that are such that b distributes its credence at least as evenly among non-w_a worlds as c does. We'll show that if b invests all of its non-w_a credence in the worlds that are d units away from w_a and c invests all of its non-w_a credence in worlds that are at least d units from w_a, and there is some world in which c invests positive credence that is more than d units away from w_a, b is more accurate than c at w_a. Let $b(\mathrm{w}_i) = b_i$ and $c(\mathrm{w}_i) = c_i$. Then Ω looks like Table 9.

Table 9. Two credence functions, b and c, distributed over a space of worlds where distance is given by the magnitude metric. Note that b distributes all of its credence in worlds that are d units away, whereas c distributes some of its credence in a further world.

$\mathbf{w_1}$	...	...	...	$\mathbf{w_{a-d}}$	...	$\mathbf{w_a}$	...	$\mathbf{w_{a+d}}$	...	$\mathbf{w_x}$	...	$\mathbf{w_x}$
				b_{a-d},				b_{a+d},		c_x		
				c_{a-d}				c_{a-d}				

We'll now consider all the convex propositions to which b assigns a non-extreme credence (that is, the convex propositions to which b does not assign zero or one). Each such proposition belongs to one of the following four categories:

Category 1: A true proposition to which b assigns b_{a-d}.
Category 2: A false proposition to which b assigns b_{a-d}.
Category 3: A true proposition to which b assigns b_{a+d}.
Category 4: A false proposition to which b assigns b_{a+d}.

We'll first show that there is a one-to-one mapping between convex propositions in Categories 1 and 2, as well as a one-to-one mapping between convex propositions in Categories 3 and 4.

Take any proposition in Category 1. Such a proposition will be a set of worlds $[\mathrm{w}_{a-d-j}, \mathrm{w}_{a+k}]$ for some $0 \leq j < a - d$ and for some $0 \leq k < d$. Each such proposition gets mapped to a convex proposition in Category 2. Which one? The proposition with the same left-hand border as the Category 1 proposition, but to turn the proposition from a true one into a false one, the right-hand border, instead of being k units to the right of w_a, is k units to the left of w_a. For example, see Table 10. In other words, the Category 1 proposition $[\mathrm{w}_{a-d-j}, \mathrm{w}_{a+k}]$ gets mapped on to the Category 2 proposition: $[\mathrm{w}_{a-d-j}, \mathrm{w}_{a-k}]$. We'll call these two propositions 'partners'. The partners of the propositions in Category 1 exhaust the propositions in Category 2.

Now take any proposition in Category 3: A true proposition to which b assigns b_{a+d}. Such a proposition will be a set of worlds $[\mathrm{w}_{a-j}, \mathrm{w}_{a+d+k}]$ for some $0 \leq j < d$, and for some $0 \leq k \leq n - (a + d)$. Each such proposition gets mapped to a convex proposition in Category 4. Which one? One with the same right-hand border as the Category 3 proposition, but to turn the proposition from true to false, the left-hand border, instead of being j units to the left of w_a, will be j units to the right of w_a (see Table 11). In other words, the Category 3 proposition $[\mathrm{w}_{a-j}, \mathrm{w}_{a+d+k}]$ gets mapped on to the Category 4 proposition: $[\mathrm{w}_{a+j}, \mathrm{w}_{a+d+k}]$. We'll call these two propositions 'partners'. The partners of the propositions in Category 3 exhaust the propositions in Category 4.

Now, if we take any Category 1 proposition, P (which, recall, is true), b's inaccuracy with respect to P is ${b_{a+d}}^2$. Why? Because the only world that is not in the true proposition P that b assigns positive credence to is w_{a+d}.

If we take the partner proposition of P, which is the false proposition P', b's inaccuracy with respect to P' is ${b_{a-d}}^2$. Why? Because the only world in the false proposition P' that b assigns positive credence to is w_{a-d}.

Similarly, for Categories 3 and 4, b's inaccuracy with respect to Category 3 proposition, P, is ${b_{a-d}}^2$. And b's inaccuracy with respect to its partner in Category 4 is ${b_{a+d}}^2$.

So for all propositions P in Categories 1–4:

$$I_{l-brier}(b(P), \mathrm{w}_a(P)) + I_{l-brier}(b(P'), \mathrm{w}_a(P')) = {b_{a-d}}^2 + {b_{a+d}}^2.$$

Table 10. An example of a true Category 1 proposition and its false partner.

Category 1 Proposition (T)

w_1	$\ldots$	$\ldots$	w_{a-3}	w_{a-2}	w_{a-1}	w_a	w_{a+1}	w_{a+2}	w_{a+3}	w_x	$\ldots$	w_x
				b_{a-2}, c_{a-2}				b_{a+2}, c_{a+2}		c_x		

Category 2 Proposition (F)

Table 11. An example of a true Category 3 proposition and its false partner.

Category 3 Proposition (T)

w_1	$\ldots$	$\ldots$	w_{a-3}	w_{a-2}	w_{a-1}	w_a	w_{a+1}	w_{a+2}	w_{a+3}	w_x	$\ldots$	w_x
				b_{a-2}, c_{a-2}				b_{a+2}, c_{a+2}		c_x		

Category 4 Proposition (F)

Because of all the n worlds in the space, b only invests positive credence in w_{a-d} and w_{a+d}

$$b_{a-d}^2 + b_{a+d}^2 = \sum_{i=1}^{n} b_i^2.$$

From the above two equations, it follows that

$$I_{l-brier}(b(P), \mathrm{w}_a(P)) + I_{l-brier}(b(P'), \mathrm{w}_a(P')) = \sum_{i=1}^{n} b_i^2.$$

Let's now consider c's inaccuracy score with respect to convex propositions. Take a proposition P in Category 1: $[\mathrm{w}_{a-d-j}, \mathrm{w}_{a+k}]$ for some $0 \leq j < a - d$, and for some $0 \leq k < d$. Since this proposition is true at w_a, the credences that contribute to c's inaccuracy with respect to this proposition are all the positive credences invested in worlds that are not members of this set: worlds in $[\mathrm{w}_1, \mathrm{w}_{a-d-j-1}]$ as well as in the worlds in $[\mathrm{w}_{a+k+1}, \mathrm{w}_n]$. However, since we're assuming that $k < d$ and that c invests no positive credence in worlds that are fewer than d units away from w_a, it follows that c invests no positive credence in $[\mathrm{w}_{a+k+1}, \mathrm{w}_{a+d-1}]$. Thus the worlds in $[\mathrm{w}_{a+k+1}, \mathrm{w}_n]$ that contribute to c's inaccuracy are all members of $[\mathrm{w}_{a+d}, \mathrm{w}_n]$, and so the worlds that contribute to c's inaccuracy in Ω are those in: $[\mathrm{w}_1, \mathrm{w}_{a-d-j-1}]$ and $[\mathrm{w}_{a+d}, \mathrm{w}_n]$.

Now consider this proposition's partner P' (in Category 2): $[\mathrm{w}_{a-d-j}, \mathrm{w}_{a-k}]$. Since this proposition is false, the credences that contribute to c's inaccuracy with respect to that proposition are the positive credences invested in worlds that are members of this set. Since c doesn't invest any credence in worlds that are less than d units away from w, it doesn't invest any credence in worlds in $[\mathrm{w}_{a-d+1}, \mathrm{w}_{a-k}]$. Thus, the credences that contribute to c's inaccuracy with respect to this proposition will be the worlds in $[\mathrm{w}_{a-d-j}, \mathrm{w}_{a-d}]$. So looking at the inaccuracy of P and P' we have:

$$I_{l-brier}(c(P), \mathrm{w}(P)) = \left(\sum_{i=1}^{a-d-j-1} c_i + \sum_{i=a+d}^{n} c_i \right)^2 > \left(\sum_{i=1}^{a-d-j-1} c_i^2 + \sum_{i=a+d}^{n} c_i^2 \right)$$

$$I_{l-brier}(c(P'), \mathrm{w}(P')) = \left(\sum_{i=a-d-j}^{a-d} c_i \right)^2.$$

It follows that

$$I_{l-brier}(c(P), \mathrm{w}(P)) + I_{l-brier}(c(P'), \mathrm{w}(P')) > \sum_{i=a}^{a-d-j-1} c_i^2 + \sum_{i=a+d}^{n} c_i^2 + \sum_{i=a-d-j}^{a-d} c_i^2.$$

Reordering (and noting that for all $i \in (a - d, a + d), c_i = 0$):

$$= \sum_{i=1}^{a-d-j-1} c_i^2 + \sum_{i=a-d-j}^{a-d} c_i^2 + \sum_{i=a+d}^{n} c_i^2 = \sum_{i=1}^{n} c_i^2.$$

Thus,

$$I_{l-brier}(c(P), \mathrm{w}(P)) + I_{l-brier}(c(P'), \mathrm{w}(P')) > \sum_{i-1}^{n} c_i^2.$$

So here's where we are: For each proposition P, in our first two categories, the sum of the inaccuracy scores of b with respect to P, and with respect to its partner P' is the sum of the b_i^2, whereas the sum of the inaccuracy scores of c with respect to P and with respect to its partner P' is greater than the sum of the ci^2. Since b is at least as evenly distributed as c, we know that

$$1 - \left[\frac{\sum_{i=1}^{n} b(\mathrm{w}_i)^2}{\left(\sum_{i=1}^{n} b(\mathrm{w}_i)\right)^2}\right] \geq 1 - \left[\frac{\sum_{i=1}^{n} c(\mathrm{w}_i)^2}{\left(\sum_{i=1}^{n} c(\mathrm{w}_i)\right)^2}\right].$$

Since $\Sigma b(\mathrm{w}_i) = \Sigma c(\mathrm{w}_i) = 1$, it follows that $\Sigma b_i^2 \leq \Sigma c_i^2$.

Since the inaccuracy of b with respect to P and $P' = \Sigma b_i^2$, and the inaccuracy of c with respect to P and P' is greater than Σc_i^2, it follows from the fact that $\Sigma b_i^2 \leq \Sigma c_i^2$, that b's inaccuracy with respect to these two propositions is greater than c's inaccuracy with respect to these two propositions.

An analogous argument applies to propositions in the Categories 3 and 4. Thus, b's total inaccuracy with respect to all the propositions in these four categories is less than c's total inaccuracy with respect to all the propositions in these four categories.

It remains to consider convex propositions to which b assigns a credence of one or zero. There are three categories:

Category 5: True convex propositions to which b assigns a credence of one.
Category 6: True convex propositions to which b assigns a credence of zero.
Category 7: False convex propositions to which b assigns a credence of zero.

Note that there are no false convex propositions to which b assigns a credence of one.

Let's compare b and c's inaccuracy with respect to propositions in each of these three categories. Every proposition in Category 5 is a true proposition to which b assigns a credence of one, and so b gets inaccuracy a score of zero with respect to these propositions. Some of these propositions will be ones such that c assigns a credence of one to them as well. But there will be at least one proposition to which b assigns a credence of one and which is such that c assigns credence less than one. For example: the proposition $[\mathrm{w}_{a-d}, \mathrm{w}_{a+d}]$ is one to which b assigns a credence of one, but c assigns credence less than one (since c invests at least some positive credence in worlds that more than d units away from w_a).

Every proposition in Category 6 is a true proposition to which b assigns a credence of zero, and so b gets inaccuracy score one with respect to these propositions. Each such proposition is one that c also assigns a credence of zero to (since c doesn't invest any credence in worlds that are less than d units away from w_a). Thus, b and c tie with respect to each proposition in this category.

Finally, let's consider Category 7: false propositions to which b assigns a credence of zero, and so gets inaccuracy a score of zero. Some of these propositions may be ones to which c also assigns zero. But there will be at least one false proposition such that b assigns a credence of zero to it, and to which c assigns positive credence. Consider a world in which c assigns positive credence that is more than d units away from w_a, and call it 'w_{a+d+m}'. The proposition $\{w_{a+d+m}\}$ will be a false convex proposition to which b assigns a credence of zero and c assigns positive credence.

Since across each of Categories 5 and 7, b is less inaccurate than c at w_a and across Category 6, b and c are equally accurate, if we consider b's inaccuracy across propositions in Categories 5–7, b will be less inaccurate than c at w_a. We already established that b is less inaccurate than c across Categories 1–4. Since Categories 1–7 exhaust all the convex propositions, b is less inaccurate than c at w_a on the weighted Brier score that assigns equal weight to all the convex propositions and no weight to any other propositions. □

Acknowledgements

Many thanks to Kevin Dorst, Jason Konek, Branden Fitelson, Stefan Lukits, Graham Oddie, Teddy Seidenfeld, Jack Spencer, Kevin Zollman, and especially Richard Pettigrew for helpful comments and discussion. I am also grateful to audiences at UC-Boulder's Workshop on Cognitive Value, Oxford University, Carnegie Mellon University, and the 2019 Formal Epistemology Workshop for their feedback, and to the two BJPS referees for their thoughtful and thorough comments.

Department of Linguistics and Philosophy
Massachusetts Institute of Technology
Cambridge, MA, USA
mschoenfieldphil@gmail.com

References

Dunn, J. [2018]: 'Accuracy, Verisimilitude, and Scoring Rules', *Australasian Journal of Philosophy*, **97**, pp. 151–66.

Greaves, H. and Wallace, D. [2006]: 'Justifying Conditionalisation: Conditionalisation Maximizes Expected Epistemic Utility', *Mind*, **115**, pp. 607–32.

Grunwald, P. D. and Dawid, A. P. [2004]: 'Game Theory, Maximum Entropy, Minimum Discrepancy, and Robust Bayesian Statistics', *The Annals of Statistics*, **32**, pp. 1367–433.

Jain, R. K., Chiu, D. W. and Hawe, W. R. [1984]: 'A Quantitative Measure for Fairness and Discrimination for Resource Allocation in Shared Computer Systems', Eastern Research Laboratory DEC-TR-301.

Joyce, J. [2009]: 'Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief', in F. Huber and C. Schmidt-Petri (*eds*), *Degrees of Belief*, Berlin: Springer, pp. 263–97.

Levinstein, B. [2017]: 'A Pragmatist's Guide to Epistemic Utility', *Philosophy of Science*, **84**, pp. 613–38.

Levinstein, B. [2019]: 'An Objection of Varying Importance to Epistemic Utility Theory', *Philosophical Studies*, **176**, pp. 2919–31.

Oddie, G. [2019]: 'What Accuracy Could Not Be', *British Journal for the Philosophy of Science*, **70**, pp. 551–80.

Pettigrew, R. [2016]: *Accuracy and the Laws of Credence*, Oxford: Oxford University Press.

Schervish, M. [1989]: 'A General Method for Comparing Probability Assessors', *Annals of Statistics*, **17**, pp. 1856–79.