Forthcoming in *Philosophy of Science*

# ON THE BEST ACCURACY ARGUMENTS FOR PROBABILISM

MICHAEL NIELSEN

ABSTRACT. In a recent paper, Pettigrew (2021) reports a generalization of the celebrated accuracy-dominance theorem due to Predd et al. (2009). But Pettigrew's proof is incorrect. I will explain the mistakes and provide a correct proof.

## 1. Introduction

Accuracy arguments for probabilism have three components: (1) a specification of the properties that (in)accuracy measures must have; (2) a normative premise stating that a credence function is irrational if it is accuracy-dominated; (3) a mathematical theorem showing that every non-probabilistic credence function is accuracy-dominated by a probabilistic one. The best accuracy arguments for probabilism, according to Pettigrew (2021), require inaccuracy measures to be strictly proper, continuous, and additive, and they appeal to the well known accuracy-dominance theorem of Predd et al. (2009). Pettigrew seeks to improve these arguments by relaxing additivity and generalizing Predd et al.'s theorem. This project is of great philosophical and mathematical interest. On the one hand, by relaxing additivity one makes a stronger case for probabilism as a prinicple of rationality. On the other hand, there is a rich body of mathematical work, going back to de Finetti (1974), that establishes connections between probability and accuracy, and it is important to understand the most general conditions under which these connections can be made.

In this note, I wish to do two things. First, I will show that Pettigrew's proof of his theorem is flawed in a number of ways. I will present counterexamples to some of his key claims. Second, and more importantly, I will give a correct proof of the theorem that Pettigrew reports (in fact, I will prove a slightly more general theorem). Although the theorem that Pettigrew states is correct in the end, the tools needed to prove it are quite different from the ones that Pettigrew uses. My hope is that with the correct tools in hand, others will generalize even further. Without the right tools, the accuracy approach to justifying probabilism risks running astray.

## 2. Preliminaries

As this is a technical note, I will be quick with the formal preliminaries. Further discussion can be found in Pettigrew's (2021) paper as well as his book (Pettigrew, 2016).

Let $W$ be a finite set of worlds. A *credence function* $c : 2^W \to [0, 1]$ is a function from the powerset of $W$ into the unit interval. Let $\mathcal{C}$ be the set of all credence functions. If $(c_n)$ is a sequence of credence functions and $c \in \mathcal{C}$, we write $c_n \to c$ iff for all $A \subseteq W$ the sequence $(c_n(A))$ of real numbers converges to $c(A)$. Another way of putting this is that we are viewing $\mathcal{C} = [0, 1]^{2^W}$ as a subset of $|2^W|$-dimensional Euclidean space. A credence function $c$ is *probabilistic* iff it obeys the usual axioms of probability. Let $\mathcal{P}$ be the set of all probabilistic credence functions. Since both $\mathcal{P}$ and $\mathcal{C}$ are closed and bounded, they are compact.

---

*Date*: June 9, 2021.

If $w \in W$, then let

$$v_\omega(A) = \begin{cases} 1, & w \in A \\ 0, & w \notin A \end{cases}$$

for all $A \subseteq W$. The (probabilistic) credence function $v_w$ is the *omniscient* or *ideal* one at the world $w$.

An *inaccuracy measure* $\mathcal{J} : \mathcal{C} \times W \to [0, \infty]$ is a function from pairs of credence functions and worlds into the extended half-line. Intuitively, $\mathcal{J}(c, w)$ represents how inaccurate the credence function $c$ is in world $w$. If $p \in \mathcal{P}$, let us write

$$\mathcal{J}(c, p) = \sum_{w \in W} \mathcal{J}(c, w) p\{w\}.^{1} \tag{1}$$

Then, $\mathcal{J}(c, p)$ is the expected inaccuracy of $c$ according to $p$. Under this definition, we have $\mathcal{J}(c, w) = \mathcal{J}(c, v_w)$ for all credence functions $c$ and worlds $w$.

Pettigrew focusses on inaccuracy measures with the following properties.

**Strict Propriety:** An inaccuracy measure $\mathcal{J}$ is *strictly proper* iff $\mathcal{J}(p, p) < \mathcal{J}(c, p)$ for all $p \in \mathcal{P}$ and all $c \in \mathcal{C} \setminus \{p\}$.

**Continuity:** An inaccuracy measure $\mathcal{J}$ is *continuous* iff for all $p \in \mathcal{P}$ and $c, c_1, c_2, ... \in \mathcal{C}$ the sequence $(\mathcal{J}(c_n, p))$ of extended-real numbers converges to $\mathcal{J}(c, p)$ whenever $c_n \to c$.[2]

The Brier score is perhaps the most well known strictly proper and continuous inaccuracy measure, but we won't have need of it here. We will, however, make use of the *enhanced log score*, defined by

$$\mathfrak{L}^*(c, w) = \sum_{A \subseteq W} (-v_w(A) \log c(A) + c(A)), \tag{2}$$

which is also strictly proper and continuous. It is worth pointing out now that if $w \in A$ and $c(A) = 0$, then $-v_w(A) \log c(A) = \infty$. So, according to the enhanced log score, any credence function that assigns zero credence to a non-empty set of worlds is infinitely inaccurate at some worlds. The fact that inaccuracy measures can take the value $\infty$ plays an important role in understanding the mistakes in Pettigrew's proof.

The final definition we will need is

**Accuracy-Dominance:** Let $c, c' \in \mathcal{C}$, and let $\mathcal{J}$ be an inaccuracy measure. We say that $c$ *accuracy-dominates* $c'$ *(according to* $\mathcal{J}$*)* iff $\mathcal{J}(c, w) < \mathcal{J}(c', w)$ for all $w \in W$.

As mentioned above, accuracy arguments for probabilism have a normative premise stating that credence functions are irrational if they are accuracy-dominated. We won't question that premise here, as our focus is on the third component of accuracy arguments: the accuracy-dominance theorems.

Here is the result that Pettigrew reports.

**Theorem 1.** *Let $\mathcal{J}$ be a strictly proper and continuous inaccuracy measure. If $c$ is a non-probabilistic credence function, then there is a probabilistic credence function that accuracy-dominates $c$ according to $\mathcal{J}$.*

In other words, every non-probabilistic credence function is accuracy-dominated by a probabilistic one. If we accept the accuracy argument's normative premise, then we can use Theorem 1 to infer that rationality forbids violations of the probability axioms.

[1] In order for this to make sense, we adopt the standard convention that $\infty \cdot 0 = 0$.

[2] This definition of continuity is equivalent to the one that Pettigrew uses: for all $w \in W$ and $c, c_1, c_2, ... \in \mathcal{C}$ the sequence $(\mathcal{J}(c_n, w))$ of extended-real numbers converges to $\mathcal{J}(c, w)$ whenever $c_n \to c$.

## 3. Pettigrew's Proof

Theorem 1 is true, but Pettigrew's proof of it is incorrect. In this section, I will explain the mistakes and lay the groundwork for a correct proof. We assume throughout this section that $\mathcal{J}$ is strictly proper and continuous.

The main technique for proving accuracy-dominance theorems involves analyzing the *divergences* associated with inaccuracy measures. Given $\mathcal{J}$, the associated divergence $\mathcal{D}_{\mathcal{J}} : \mathcal{C} \times \mathcal{P} \to [0, \infty]$ is defined by

$$\mathcal{D}_{\mathcal{J}}(c,p) = \mathcal{J}(c,p) - \mathcal{J}(p,p). \tag{3}$$

The equation in (3) is well defined—never of the form $\infty - \infty$—because $\mathcal{J}$ is strictly proper, and this implies that $\mathcal{J}(p,p) < \infty$ for all $p \in \mathcal{P}$. We can think of $\mathcal{D}_{\mathcal{J}}$ as measuring the distance between the $p$-expected inaccuracy of $c$ and the minimal $p$-expected inaccuracy, which is $\mathcal{J}(p,p)$ by strict propriety.

Now, let $c \in \mathcal{C} \setminus \mathcal{P}$ be a non-probabilistic credence function. Then, $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is a function from the set of probability functions $\mathcal{P}$ into the extended half-line. The proof of Theorem 1 has two steps. Step 1: one argues that $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ can be *minimized*. That is, one argues that there is some $p^* \in \mathcal{P}$ such that $\mathcal{D}_{\mathcal{J}}(c,p^*) \leq D_{\mathcal{J}}(c,p)$ for all $p \in \mathcal{P}$. Step 2: one shows that the minimizer $p^*$ of $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ can be used to construct a probability function that accuracy-dominates $c$.

Pettigrew's proof has mistakes in both steps. It will be easier to describe the mistake he makes in the second step in the next section (Remark 1). In this section, we will focus on the mistakes in the first step.

Why is it possible to minimize $\mathcal{D}_{\mathcal{J}}(c,\cdot)$? Pettigrew gives two arguments, both contained in the proof of his Lemma 5. Here is the first argument:[3] "Suppose $c$ is not in $\mathcal{P}$. Then, since $\mathcal{P}$ is a closed convex set and since Lemma 4(iii) shows that $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is strictly convex there is a unique $p^* \in \mathcal{P}$ that minimizes $\mathcal{D}_{\mathcal{J}}(c,\cdot)$."

There are two problems with this argument. (1) The assumptions in Theorem 1, namely that $\mathcal{J}$ is strictly proper and continuous, do *not* imply that $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is strictly convex (that is, Pettigrew's Lemma 4(iii) is false). (2) Even if $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ were strictly convex, this would not imply the existence of a minimizer. In particular, it is *not* the case that every strictly convex function on a closed convex set attains a minimum on that set.

I will demonstrate (1) and (2) in a moment. But first, let us take a look at the beginning of Pettigrew's second argument. He writes: "We now briefly sketch an alternative proof...that is available if $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is not only continuous...but also differentiable."

The problem with this is: (3) The assumptions in Theorem 1 do *not* imply that $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is continuous and therefore don't imply that $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is differentiable.[4] This is important to realize because if $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ *were* continuous, then it would attain a minimum on $\mathcal{P}$ due to the basic topological fact that continuous functions attain minima on compact sets. In other words, if $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ were continuous, then Pettigrew's argument would be easy to repair.[5] As we will soon see, proving Pettigrew's theorem requires a bit more argumentation.

---

[3] I have changed the notation a bit here and in the next quotation. Pettigrew states Lemma 5 for *any* closed convex subset of $\mathcal{P}$, but we need only think about $\mathcal{P}$ itself.

[4] To be fair, Pettigrew does not seem to be claiming that $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ *is* differentiable in general, though he does seem to think that the continuity of $\mathcal{D}_{\mathcal{J}}(c,\cdot)$ is a consequences of his theorem's assumptions.

[5] If $\mathcal{J}$ happens to be real-valued (e.g. the Brier score), and therefore bounded, then $\mathcal{D}_{\mathcal{J}}$ is continuous, and the easy repair of Pettigrew's theorem works. The reader should note, however, that Pettigrew's proof is still incorrect for bounded $\mathcal{J}$ because of problem (2) mentioned above.

Now let us establish the three claims made in response to Pettigrew's arguments. Claims (1) and (3) are established by the following counterexample.

**Example 1.** *Of a credence function $c$ and a strictly proper and continuous inaccuracy measure $\mathcal{J}$ such that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is neither continuous nor strictly convex.*

Let $W = \{1, 2\}$, and let the inaccuracy measure be the enhanced log score $\mathfrak{L}^*$, defined in (2). Let $c$ be the probabilistic credence function defined by $c\{1\} = 0$. Then $\mathfrak{L}^*(c, 1) = \infty$. Let $(p_n)$ be the sequence of probability functions defined by $p_n\{1\} = n^{-1}$. Then, $\mathfrak{L}^*(c, p_n) = \infty$ for all $n$, which implies $\mathcal{D}_{\mathfrak{L}^*}(c, p_n) = \infty$ for all $n$. But $p_n \to c$. So,

$$\mathcal{D}_{\mathfrak{L}^*}(c, p_n) \to \infty \neq 0 = \mathcal{D}_{\mathfrak{L}^*}(c, c)$$

shows that $\mathcal{D}_{\mathfrak{L}^*}(c, \cdot)$ is not continuous.

To see that $\mathcal{D}_{\mathfrak{L}^*}(c, \cdot)$ is not strictly convex, note that $p_n = n^{-1}v_1 + (1 - n^{-1})c$. But then,

$$\mathcal{D}_{\mathfrak{L}^*}(c, p_n) = \infty \geq n^{-1}\mathcal{D}_{\mathfrak{L}^*}(c, v_1) + (1 - n^{-1})\mathcal{D}_{\mathfrak{L}^*}(c, c),$$

so $\mathcal{D}_{\mathfrak{L}^*}(c, \cdot)$ is not strictly convex. △

And claim (2) is established by the following counterexample.

**Example 2.** *Of a closed convex set and a strictly convex function on that set that attains no minimum.*

Consider the function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} (x-1)^2, & x \in [0, 1) \\ 1, & x = 1. \end{cases}$$

The set $[0, 1]$ is closed (indeed, compact) and convex, and $f$ is strictly convex. But $f$ does not attain a minimum on $[0, 1]$. △

The key points to take away from the examples are: (1) In general, a divergence is *not* strictly convex in its second argument; (2) Strict convexity does *not* guarantee the existence of minimizers; (3) In general, a divergence is *not* continuous in its second argument.[6] We now move on to providing an alternative proof of Theorem 1.

[6]The reader might find the following speculation about the source of Pettigrew's mistakes helpful. I conjecture that Pettigrew misread and subsequently misapplied an argument in Predd et al. The relevant portion of Predd et al.'s proof reads (changing notation), "the function $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is strictly convex, and hence achieves a unique minimum at a point $p$" (p. 4789, in the proof of their Proposition 3). The context of this remark is important, however. In the quoted passage, Predd et al. are working under the assumption that $\mathcal{D}_{\mathcal{J}}$ is a *Bregman divergence*, and this guarantees that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is strictly convex as well as *continuous* and *real-valued* (see their Definition 6). As mentioned above, this implies that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ attains a minimum on $\mathcal{P}$ because $\mathcal{P}$ is *compact*. This argument has nothing to do with *convexity* but rather *topology*. What strict convexity adds is the guarantee that minima are attained *uniquely* on convex sets: that is, a strictly convex function attains a minimum at *no more* than one point in a convex set (although it might not attain a minimum at any point in the set, as Example 2 shows). The most charitable reading of the quoted passage from Predd et al. takes it as a point about uniqueness. It seems possible that Pettigrew has read this passage as expressing an (incorrect) inference about minimization, which is why he writes (changing notation), "since $\mathcal{P}$ is a closed convex set and since...$\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is strictly convex...there is a unique $p$ in $\mathcal{P}$ that minimizes $\mathcal{D}_{\mathcal{J}}(c, \cdot)$." Again, the strict convexity of $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ would imply that there is *at most* one $p$ that minimizes $\mathcal{P}$, but it doesn't imply that a minimizer exists.

## 4. An Accuracy-Dominance Theorem

In fact, we will generalize Theorem 1 a bit by weakening both strict propriety and continuity. I do not know whether the properties that I am about to introduce are more philosophically compelling than strict propriety and continuity. My aim here is simply to lay bare as much as possible the mathematical arguments needed to support accuracy-based approaches to justifying probabilism.

**Quasi-Strict Propriety:** An inaccuracy measure $\mathcal{J}$ is *quasi-strictly proper* iff $\mathcal{J}(p,p) \le \mathcal{J}(c,p)$ for all $p \in \mathcal{P}$ and all $c \in \mathcal{C}$, with strict inequality if $c \in \mathcal{C} \setminus \mathcal{P}$.

**Continuity on $\mathcal{P}$:** An inaccuracy measure $\mathcal{J}$ is *continuous on* $\mathcal{P}$ iff for all $p, q, q_1, q_2, \ldots \in \mathcal{P}$ the sequence $(\mathcal{J}(q_n, p))$ of extended-real numbers converges to $\mathcal{J}(q,p)$ whenever $q_n \to q$.

Unlike strict propriety, quasi-strict propriety allows for $\mathcal{J}(p,p) = \mathcal{J}(q,p)$ provided $q$ is probabilistic. A strict inequality is required only for non-probabilistic credence functions. Continuity on $\mathcal{P}$ differs from continuity by allowing the inaccuracy measure to be discontinuous on non-probabilistic credence functions.

Here is the result that we will prove.[7]

**Theorem 2.** *Let $\mathcal{J}$ be a quasi-strictly proper inaccuracy measure that is continuous on $\mathcal{P}$. If $c$ is a non-probabilistic credence function, then there is a probabilistic credence function that accuracy-dominates $c$ according to $\mathcal{J}$.*

Throughout the rest of this section, let $\mathcal{J}$ be a quasi-strictly proper inaccuracy measure that is continuous on $\mathcal{P}$, and let $c \in \mathcal{C} \setminus \mathcal{P}$. As outlined in the previous section, the proof of Theorem 2 has two steps that involve analyzing the divergence $\mathcal{D}_{\mathcal{J}}(c, \cdot)$. The first thing we should note is that $\mathcal{D}_{\mathcal{J}}$ continues to be well defined under quasi-strict propriety because we still have $\mathcal{J}(p,p) < \infty$ for all $p \in \mathcal{P}$.

Another consequence of quasi-strict propriety worth noting right away is that we can assume without loss of generality that $\mathcal{J}(c,w) < \infty$ for some $w \in W$. Otherwise any *regular* probability—any probability such that $p\{w\} > 0$ for all $w \in W$—would accuracy-dominate $c$. That's because if $p$ is regular, then $\mathcal{J}(p,p) < \infty$ implies $\mathcal{J}(p,w) < \infty$ for all $w \in W$.

Recall the two steps of the proof. Step 1: show that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is minimized at some $p^* \in \mathcal{P}$. Step 2: use $p^*$ to construct a probabilistic credence function that accuracy-dominates $c$. I will dispense with Step 2 first, as this will allow us to see the final mistake in Pettigrew's proof. The proof of the following lemma is similar to arguments given by Pettigrew, but special care has to be taken to handle cases where $\mathcal{J}(c,w) = \infty$.

**Lemma 1.** *If $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ attains a minimum on $\mathcal{P}$, then there is some probabilistic credence function that accuracy-dominates $c$.*

*Proof.* Let $p^* \in \mathcal{P}$ minimize $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ on $\mathcal{P}$. Since it is safe to assume that $\mathcal{J}(c,w) < \infty$ for some $w \in W$, it follows that $\mathcal{D}_{\mathcal{J}}(c, p^*) < \infty$ because

$$\mathcal{D}_{\mathcal{J}}(c, p^*) \le \mathcal{D}_{\mathcal{J}}(c, v_w) = \mathcal{J}(c,w) - \mathcal{J}(v_w, w) < \infty.$$

This, in turn, implies that $\mathcal{J}(c, p^*) < \infty$.

Now let $w \in W$ be arbitrary. If $n \in \mathbb{N}$, let $p_n = n^{-1} v_w + (1 - n^{-1}) p^*$. Note that because $\mathcal{J}(p_n, p_n) < \infty$ and

$$\mathcal{J}(p_n, p_n) = n^{-1} \mathcal{J}(p_n, w) + (1 - n^{-1}) \mathcal{J}(p_n, p^*),$$

[7] Alex Pruss independently discovered a completely different proof of this result, using geometric techniques, around the same time that I discovered the proof in this paper.

we must have both $\mathcal{J}(p_n, w) < \infty$ and $\mathcal{J}(p_n, p^*) < \infty$ for all $n$. Using the fact that $p^*$ is a minimizer of $\mathcal{D}_{\mathcal{J}}(c, \cdot)$, we have, for all $n$,

$$\begin{aligned}
0 &\leq n\big(\mathcal{D}_{\mathcal{J}}(c, p_n) - \mathcal{D}_{\mathcal{J}}(c, p^*)\big) \\
&= n\big(\mathcal{J}(c, p_n) - \mathcal{J}(p_n, p_n) - \mathcal{J}(c, p^*) + \mathcal{J}(p^*, p^*)\big) \\
&= n\big(n^{-1}\mathcal{J}(c, w) + (1 - n^{-1})\mathcal{J}(c, p^*) - n^{-1}\mathcal{J}(p_n, w) - (1 - n^{-1})\mathcal{J}(p_n, p^*) - \mathcal{J}(c, p^*) + \mathcal{J}(p^*, p^*)\big) \\
&= \big(\mathcal{J}(c, w) - \mathcal{J}(c, p^*) - \mathcal{J}(p_n, w) + \mathcal{J}(p_n, p^*)\big) + n\big(\mathcal{J}(p^*, p^*) - \mathcal{J}(p_n, p^*)\big) \\
&\leq \mathcal{J}(c, w) - \mathcal{J}(c, p^*) - \mathcal{J}(p_n, w) + \mathcal{J}(p_n, p^*). \qquad (4)
\end{aligned}$$

where the final inequality uses quasi-strict propriety. Next, (4) rearranges to

$$\mathcal{J}(p_n, w) \leq \mathcal{J}(c, w) - \mathcal{J}(c, p^*) + \mathcal{J}(p_n, p^*), \qquad (5)$$

and $p_n \to p^*$ implies

$$\mathcal{J}(p^*, w) \leq \mathcal{J}(c, w) - \mathcal{D}_{\mathcal{J}}(c, p^*). \qquad (6)$$

Quasi-strict propriety implies $\mathcal{D}_{\mathcal{J}}(c, p^*) > 0$. So from (6), we have $\mathcal{J}(p^*, w) < \mathcal{J}(c, w)$ for all $w \in W$ such that $\mathcal{J}(c, w) < \infty$. But if $\mathcal{J}(c, w) = \infty$, then (6) implies only the weak inequality $\mathcal{J}(p^*, w) \leq \mathcal{J}(c, w)$. To deal with this, we can choose any regular probability $q$ and let $q_n = n^{-1}q + (1 - n^{-1})p^*$. Then, $q_n \to p^*$. So by continuity, for large enough $n$, we have $\mathcal{J}(q_n, w) < \mathcal{J}(c, w)$ whenever $\mathcal{J}(c, w) < \infty$. And we also have $\mathcal{J}(q_n, w) < \mathcal{J}(c, w)$ whenever $\mathcal{J}(c, w) = \infty$, simply in virtue of the fact that the regularity of $q_n$ implies $\mathcal{J}(q_n, w) < \infty$ for all $w \in W$. Thus, for large enough $n$, $q_n$ accuracy-dominates $c$. This concludes the proof of the lemma.[8] □

**Remark 1.** Pettigrew's final mistake concerns equation (6). He infers (in the proof of his Theorem 1) directly from this equation and the fact that $\mathcal{D}_{\mathcal{J}}(c, p^*) > 0$ that $\mathcal{J}(p^*, w) < \mathcal{J}(c, w)$ for all $w \in W$. But that doesn't follow if $\mathcal{J}(c, w) = \infty$.

So much for Step 2. In order to conclude the proof of Theorem 2, we must now show that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ actually attains a minimum.

To do this, we will argue that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ has a property that is *weaker* than continuity but still strong enough to imply the existence of a minimizer on compact sets. Here is that property.

**Lower Semicontinuity:** Let $X$ be a topological space, and let $f : X \to [-\infty, \infty]$. We say that $f$ is *lower semicontinuous (lsc)* iff for every $r \in \mathbb{R}$ the set $\{x \in X : f(x) \leq r\}$ is closed.

Our aim is to argue that $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is lsc, and use the following lemma to conclude.

**Lemma 2.** *Let $X$ be a compact topological space, and let $f : X \to [-\infty, \infty]$ be a lsc function. Then, $f$ attains a minimum on $X$.*

*Proof.* Let $A = f(X)$. If $-\infty \in A$, then the result is immediate, so assume this is not the case. For all $r \in A$, let $F_r = \{x \in X : f(x) \leq r\}$. If $r \in \mathbb{R}$, then $F_r$ is closed by lower semicontinuity, and if $r = \infty$, then $F_r = X$ is closed as well. If $\{r_1, ..., r_n\} \subseteq A$, then

[8] Predd et al. (2009) prove their theorem for bounded $\mathcal{J}$ first and then generalize to unbounded $\mathcal{J}$ with an inductive argument. One nice feature of the proof in this paper is that it tackles the general, unbounded case directly, and thereby simplifies Predd et al.'s approach a bit. However, readers of Predd et al. will recognize that the argument at the end of the proof of Lemma 1 is similar to an argument that they use in the proof of their unbounded case (p. 4790).

$\bigcap_{i=1}^n F_{r_i} = F_{\min_i r_i}$ is non-empty. Since $X$ is compact, this implies that $\bigcap_{r\in A} F_r$ is non-empty. But $f$ attains a minimum at any point in $\bigcap_{r\in A} F_r$.[9] □

We will need the following basic facts about lsc functions.

**Lemma 3.** *Let $X$ be a topological space, and let $f : X \to [-\infty, \infty]$. (a) If $f$ is real-valued and lsc, and $-f$ is also lsc, then $f$ is continuous. (b) If $(f_\alpha)$ is a family of lsc functions on $X$ and $f(x) = \sup_\alpha f_\alpha(x)$ for all $x \in X$, then $f$ is lsc.*

*Proof.* We omit the proof of (a). For (b), see Lemma 2.41 in Aliprantis and Border (2006). □

Finally, it is worth isolating the following key fact as a separate lemma.

**Lemma 4.** *Let $f : \mathcal{P} \to \mathbb{R}$ be the function defined by $f(p) = \mathcal{J}(p,p)$. Then, $f$ is continuous on $\mathcal{P}$.*

*Proof.* By Lemma 3(a), it suffices to show that both $f$ and $-f$ are lsc. Beginning with $f$, let $r \in \mathbb{R}$ and let $F_r = \{p \in \mathcal{P} : \mathcal{J}(p,p) \le r\}$. Let $(p_n)$ be a sequence in $F_r$ that converges to $p \in \mathcal{P}$. We aim to show that $p \in F_r$. Let $A = \{w \in W : \mathcal{J}(p,w) < \infty\}$. Then, $p(A) = 1$ by quasi-strict propriety, and by continuity on $\mathcal{P}$

$$r \ge \mathcal{J}(p_n,p_n) \ge \sum_{w\in A} \mathcal{J}(p_n,w)p_n\{w\} \to \sum_{w\in A} \mathcal{J}(p,w)p\{w\} = \mathcal{J}(p,p)$$

This proves that $p \in F_r$. So $f$ is lsc.

Next, begin by observing that quasi-strict propriety implies that $f(p) = \inf_{c\in\mathcal{C}} \mathcal{J}(c,p)$ for all $p \in \mathcal{P}$. We claim that the infimum can be taken over only those $c \in \mathcal{C}$ for which $\mathcal{J}(c,\cdot)$ is real-valued. Let the set of such $c$ be denoted by $\mathcal{C}_\mathbb{R}$, and recall that if $p$ is a *regular* probability, then $p \in \mathcal{C}_\mathbb{R}$. Now, any $p \in \mathcal{P}$ is the limit of a sequence $(p_n)$ of regular probabilities. Continuity on $\mathcal{P}$ implies that $\mathcal{J}(p_n,p) \to \mathcal{J}(p,p)$. Thus, for all $\epsilon > 0$, there is some $p_n \in \mathcal{C}_\mathbb{R}$ such that

$$\inf_{c\in\mathcal{C}_\mathbb{R}} \mathcal{J}(c,p) \le \mathcal{J}(p_n,p) \le \mathcal{J}(p,p) + \epsilon = \inf_{c\in\mathcal{C}} \mathcal{J}(c,p) + \epsilon.$$

This proves the claim, i.e. $f(p) = \inf_{c\in\mathcal{C}_\mathbb{R}} \mathcal{J}(c,p)$ for all $p \in \mathcal{P}$.

Now, if $c \in \mathcal{C}_\mathbb{R}$, then the function $\mathcal{J}(c,\cdot)$ is continuous on $\mathcal{P}$. Indeed, if $p_n \to p$, then

$$\mathcal{J}(c,p_n) = \sum_{w\in W} \mathcal{J}(c,w)p_n\{w\} \to \sum_{w\in W} \mathcal{J}(c,w)p\{w\} = \mathcal{J}(p,p).$$

Then, by the duality

$$-f(p) = \sup_{c\in\mathcal{C}_\mathbb{R}} -\mathcal{J}(c,p)$$

we see that $-f$ is the pointwise supremum of a family of continuous (and therefore lsc) functions. It follows from Lemma 3(b) that $-f$ is lsc, and the proof is complete. □

We can finally show that divergences are lsc in their second argument.

**Lemma 5.** *The function $\mathcal{D}_\mathcal{J}(c,\cdot)$ is lsc on $\mathcal{P}$.*

[9] This proof is an easy modification of the proof of Lemma 2.43 in Aliprantis and Border (2006), which is stated for real-valued functions.

*Proof.* Let $r \in \mathbb{R}$ be given and write $F_r = \{p \in \mathcal{P} : \mathcal{D}_{\mathcal{J}}(c, p) \leq r\}$. Let $(p_n)$ be a sequence in $F_r$ that converges to $p \in \mathcal{P}$. Let $A = \{w \in W : \mathcal{J}(c, w) < \infty\}$. From $p_n \in F_r$ it follows that $p_n(A) = 1$, which implies $p(A) = 1$ because $p_n \to p$. Then, using Lemma 4, we have

$$\begin{aligned} r &\geq \mathcal{D}_{\mathcal{J}}(c, p_n) \\ &= \mathcal{J}(c, p_n) - \mathcal{J}(p_n, p_n) \\ &= \sum_{w \in A} \mathcal{J}(c, w) p_n\{w\} - \mathcal{J}(p_n, p_n) \\ &\to \sum_{w \in A} \mathcal{J}(c, w) p\{w\} - \mathcal{J}(p, p) \\ &= \mathcal{J}(c, p) - \mathcal{J}(p, p) \\ &= \mathcal{D}_{\mathcal{J}}(c, p). \end{aligned}$$

Thus, $p \in F_r$, and the lemma is proved. □

Let us conclude by summarizing how all of the pieces fit together.

*Proof of Theorem 2.* Step 1: $\mathcal{D}_{\mathcal{J}}(c, \cdot)$ is lsc on $\mathcal{P}$ (Lemma 5) and therefore attains a minimum on $\mathcal{P}$ (Lemma 2). Step 2: The existence of a minimizer implies that $c$ is accuracy-dominated by some probabilistic credence function (Lemma 1). □

## References

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag, Berlin.

de Finetti, B. (1974). *Theory of Probability*, Volume 1. John Wiley & Sons.

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

Pettigrew, R. (2021). Accuracy-first epistemology without the additivity axiom. *Philosophy of Science*, Forthcoming.

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