

“What Accuracy Could Not Be”

Graham Oddie

Forthcoming in *The British Journal for the Philosophy of Science*

What Accuracy Could Not Be

Abstract

Two different programs are in the business of explicating accuracy—the truthlikeness program and the epistemic utility program. Both assume that truth is the goal of inquiry, and that among inquiries that fall short of realizing the goal some get closer to it than others. TL theorists have been searching for an account of the accuracy of propositions. Epistemic utility theorists have been searching for an account of the accuracy of credal states. Both assume we can make cognitive progress in an inquiry even while falling short of the target. I show that the prospects for combining these two programs are bleak. A core accuracy principle, Proximity, that is universally embraced within the Truthlikeness program turns out to be incompatible with a central principle within the Epistemic Utility program, namely Propriety.

1. *Truthlikeness and Epistemic Utility*
2. *Inquiries*
3. *Accuracy For Propositions*
4. *Proximity For Propositions*
5. *Accuracy For Credal States*
6. *Propriety*
7. *Proximity For Credal States*
8. *Extensionality*
9. *Admissible Weightings*
10. *Propriety Violates Proximity*
11. *Possible Responses*
 - 11.1 *Retreat to convexity*
 - 11.2 *Reject boundedness*
 - 11.3 *Reject additivity*
12. *The Upshot*

1 Truthlikeness and Epistemic Utility

Accuracy is not a only pervasive notion but a central one. In our epistemic adventures we aspire to the whole truth of some matter, but we often fail to reach it. Not all need be in vain, however, if some attempts to capture the truth, while not completely accurate, are more accurate than others. Two different programs are in the business of trying to pin accuracy down— the truthlikeness (TL) program and the epistemic utility (EU) program. The former was inspired by

Karl Popper, the latter by Carl Hempel.¹ While these initial protagonists were pursuing slightly different epistemologies (probabilism and falsificationism respectively) they weren't so far apart. Both assumed that truth is the goal of inquiry, but that among inquiries that fall short of realizing the goal some get closer to it than others. TL theorists are after an account of the accuracy of propositions—of closeness of propositions to the fully accurate target proposition. EU theorists are after an account of the accuracy of probabilistic credal states—of the closeness of credal states to the fully accurate credal state. Both assume we can make cognitive progress in an inquiry even while falling short of the target. Both ground the possibility of such progress in the value of accuracy. Several protagonists have suggested that, since both programs are zeroing in on accuracy, consilience might be in the offing.² But it turns out that the prospects for consilience are bleak. A compelling principle governing truthlikeness—*Proximity*—is incompatible with a central principle of epistemic utility theory—*Propriety*. And the incompatibility runs deep.

2 Inquiries

The goal of an inquiry is to arrive at the truth. Not the whole truth about the universe, of course, but the truth of some particular matter—such as the number of the planets, the state of the weather in Boulder tomorrow, the laws of motion, or the causes of the fire. In pursuing an inquiry we are not trying to identify which world is actual—that would be hubristic—but to locate actuality in some more coarse-grained partition of the logical space.

Consider the inquiry (I^N) that seeks an answer to the question: What is number of the planets? Let N_j be the set of worlds in which the value of the magnitude at issue, N , is j . The question effects a partition $I^N = \{N_0, N_1, N_2, \dots\}$ of the logical space. Each cell of I^N gives a complete answer to the question I^N . The cell that contains the actual world is the truth concerning I^N . Assume that N_9 is the truth about I^N .³

Another question: What will the weather in Boulder be tomorrow? This involves a vague notion (the weather). We can sharpen the question by identifying factors that constitute the state of the weather (say, temperature and precipitation) and specifying values for each factor (say, *hot/not-hot*, and *rainy/not-rainy* for simplicity). The cells of the weather question, $I^W = \{A_1, A_2, A_3, A_4\}$, can be represented by couples of 1s and 0s.

¹Hempel [1960], Popper [1963].

²Oddie [1997], Joyce [1998] and [2009], and Greaves and Wallace [2006]. Leitgeb and Pettigrew [2010] cite this as one of five open problems within the EU program. Its still open.

³With the discovery of Pluto N_9 seemed to be the truth, but with Pluto's demotion N_8 returned. However, with intimations of a distant dark planet N_9 is on the cards again.

Not every answer to a question need be complete. That \mathbf{N} is *between seven and ten* is an answer to $I^{\mathbf{N}}$, albeit an incomplete one. That *it is hot* is also an incomplete answer to $I^{\mathbf{W}}$, the weather question. Each incomplete answer is compatible with a range of complete answers—and will be represented as a class or disjunction of such.

The simplest kind of credal state—simple belief in, or acceptance of, proposition A —can be represented by the object of that belief. At each stage of an inquiry some proposition A is the strongest answer to the question that is accepted. The course of an inquiry is a path through the space of possible answers. The task of the TL program is to specify when a step in an inquiry constitutes progress towards the truth.

3 Accuracy For Propositions

Let $I = \{C_1, C_2, \dots\}$ be some question. We are after an accuracy ordering \succeq on the set of answers over I . $A \succeq^i B$ just in case A is at least as accurate as B when C_i is the truth about I . (As usual $A \approx^i B$ if $A \succeq^i B$ and $B \succeq^i A$, and $A >^i B$ if $A \succeq^i B$ and not $A \approx^i B$.) Some answers may turn out to be incomparable, but for an account to be materially adequate it must at least deliver the intuitively compelling comparisons.

Popper was the first to attempt a general definition of \succeq . Where C_i is true, let $Tr_i(A)$ be the true answers entailed by A and $Fa_i(A)$ the set of false answers entailed by A . Popper's proposal is this: $A \succeq^i B =_{df} Tr_i(B) \subseteq Tr_i(A)$ and $Fa_i(A) \subseteq Fa_i(B)$. It follows that $A >^i B$ if $A \succeq^i B$ and one of these inclusions is strict. Unfortunately, for false A and B , if $Tr_i(B) \subseteq Tr_i(A)$ then not $Fa_i(A) \subset Fa_i(B)$, and if $Fa_i(A) \subseteq Fa_i(B)$ then not $Tr_i(B) \subset Tr_i(A)$. So Popper's account deems no falsehood more accurate than any other.⁴ This trivialization of the concept of truthlikeness is not only counterintuitive (some falsehoods seem more accurate than others), but it undermines Popper's explicit aim of vindicating the possibility of progress through a series of falsehoods.

In contrast to Popper's very gappy proposal there are rival proposals that yield complete orderings. Suppose we have a natural candidate for a measure of distance, δ , between cells of $I = \{C_1, C_2, \dots\}$: δ_{ij} is the distance between C_i and C_j . Such a measure gives us an accuracy ordering for the complete answers at least: $C_j \succeq^i C_k$ just in case $\delta_{ji} \leq \delta_{ki}$. For some inquiries a plausible δ is rather obvious. The distance between cells \mathbf{N}_i and \mathbf{N}_j of the partition $I^{\mathbf{N}}$ is naturally taken to be the distance between the values they assign to magnitude \mathbf{N} : $\delta_{ij} = |i - j|$. In other cases distances between cells can be naturally constructed. The weather question involves two 0-1 factors. The cells of $I^{\mathbf{W}}$ are couples of 0s and 1s. Assuming the two weather factors are

⁴ If A and B are false neither condition holds. One cannot add truths to a false theory without adding falsehoods, or remove falsehoods without removing truths (Tichý [1974] and Miller [1974]).

equally weighty, one measure of distance—the city block measure—is given by the number of disagreements between them.

Table 1: Distance between cells in the weather framework on the city block measure

Some atomic states may more weight than others. Such asymmetries are quite possible but there is nothing to prevent incorporating different weightings. Perhaps the temperature factor counts twice as much as the precipitation factor in accuracy assessments. In that case, the distance of $\neg\mathbf{h}\wedge\mathbf{r}$ from $\mathbf{h}\wedge\mathbf{r}$ would be twice that of the distance of $\mathbf{h}\wedge\neg\mathbf{r}$ from $\mathbf{h}\wedge\mathbf{r}$.

Figure 1: The weather space

Suppose, then, that we have an accuracy ordering \succeq of the complete answers, whether or not it derives from a distance measure δ . What we want is an accuracy ordering of the incomplete answers as well. We need to extend \succeq to the incomplete answers.

\mathcal{I}^W has four complete answers and eleven incomplete answers, each of which equivalent to a disjunction of complete answers.⁵

Table 2: Complete and incomplete answers to the weather question

There are various proposals for extending a measure of distance δ or an ordering \succeq from the cells to the incomplete answers. The simplest deem the inaccuracy of proposition A to be determined in the obvious way by some representative A -cell: such as the A -cell closest to C_i (*min*) or the A -cell furthest from C_i (*max*).⁶ *min* and *max* rank some false propositions, and they both rank some correctly, but they are crude. *min* deems all truths to be equidistant from the truth (the tautology is no less accurate than the whole truth). *max* corrects this, but at the cost of deeming no false theory less accurate than the tautology. Clearly a single representative A -cell won't do. Averaging *min* and *max* (*min-max*) might suggest itself.⁷ But why privilege just two A -cells? Why not take the distances of all A -cells into account? We could do that by averaging (*ave*).⁸ This gets a lot but it doesn't deliver a principle highly favored by Popper—the value of content for truths—that the logically stronger of two truths is closer to the truth.⁹ One extension that strongly penalizes weakening is the sum of distances of A -cells from the true cell (*sum*). *sum*

⁵ The contradiction, which rules out all cells, is on this characterization, no answer at all.

⁶ *min* is proposed by Weston [1992] and Teller [2001].

⁷ Hilpinen [1976] proposed a qualitative version and Niiniluoto [1987] proposed the numerical version.

⁸ Tichý [1974] and Oddie [2013].

⁹ This has been considered sufficient grounds by some to reject *ave*. For a defense, see Oddie [2013].

yields the value of content for truths but also delivers the value of content for falsehoods—that the stronger of two falsehoods is closer to the truth. That is as counterintuitive as it gets. We can mitigate this by incorporating a factor sensitive to proximity to the true cell (*min*, say). Taking a suitably weighted average of *min* and *sum* (call it *min-sum*) yields the value of content for truths without entailing the value of content for falsehoods.¹⁰ While strengthening a falsehood decreases *sum* it can also increase *min*, by eliminating closest cells. A different measure is based on the symmetric difference operation Δ on classes of cells.¹¹ Let *sym* be $P(A\Delta C_i)$ where P is a normalized measure of the size of A . For propositions with the same truth value *sym*, like *sum*, diminishes with increasing logical content, delivering the value of content for falsehoods as well as truths. Like *sum*, *sym* can also be ameliorated by averaging with *min* (*min-sym*).¹²

This overview is not exhaustive—we could consider other mixtures, like *min-sum-max*—but already this may give the impression that accuracy principles are all just too contested for there to be any coherent notion of propositional accuracy. This would be a mistake. Not all principles of propositional accuracy are contested. Below I give a couple of compelling desiderata which together I call *Strict Proximity*.

One last caveat. The desiderata proposed apply with equal force to any qualitative accuracy ordering, partial or complete, whether or not it is derived from a distance measure. All we really need in what follows it that there is at least one question with four complete answers that bear the accuracy relations to one another exhibited in *Figure 1*.

4 *Proximity*

The accuracy of A depends on the accuracy of the A -cells. Other things being equal, the more accurate the A -cells are, the more accurate A is. This underwrites two principles, one governing propositions of different logical strength and the other governing propositions of the same logical strength. Let $Closest_i(A)$ be the set of most accurate A -cells. *Table 3* lists some examples:

Table 3 Closest A-cells

The first principle is this: if C is a closest A -cells then C is no less accurate than A itself:

Weak Proximity

If $C \in Closest_i(A)$ then $C \geq_i A$.

¹⁰ *min-sum* is Niiniluoto's preferred measure. See his [1987].

¹¹ Miller [1977].

¹² See Kuipers [1992] p. 332 for a proposal like *min-sum* on a structuralist account of theories

If A is true then the sole member of $Closest_i(A)$ is C_i which is of course no less accurate than A . If A is false, *Weak Proximity* requires only that the move from A to a member of $Closest_i(A)$ doesn't involve a decrease in accuracy. *Weak Proximity* is satisfied by all accounts of truthlikeness considered above, which lends it strong independent support.

A striking example of *Weak Proximity* is worth noting. Where T is the truth, $\neg T$ is the weakest falsehood on offer (*Figure 2*). If there is anything to the idea of propositional accuracy it is that the cells in $\neg T$ closest to the truth are not further from the truth than $\neg T$.

Figure 2 The relative accuracy of $\neg T$ and $Closest_i(\neg T)$

Now consider answers with the same number of cells. Suppose A contains C_j and excludes C_k . Let A^{kj} be the result of substituting C_k for C_j in A . A and A^{kj} possess the same number of cells and so by one measure of the same logical strength. For example:

- (i) $(\mathbf{A}_1 \vee \mathbf{A}_3)$ is the result of substituting \mathbf{A}_3 for \mathbf{A}_2 in $(\mathbf{A}_1 \vee \mathbf{A}_2)$;
- (ii) $(\mathbf{A}_1 \vee \mathbf{A}_2)$ is the result of substituting \mathbf{A}_1 for \mathbf{A}_4 in $(\mathbf{A}_2 \vee \mathbf{A}_4)$.

Let C_j and C_k be equally accurate. Then the substitution of C_k for C_j neither decreases nor increases accuracy: A^{kj} and A are equally accurate. Suppose that C_k is more accurate than C_j . Then the substitution of C_k for C_j clearly improves accuracy.

- (iii) Since $\mathbf{A}_3 \approx^1 \mathbf{A}_2$, $(\mathbf{A}_1 \vee \mathbf{A}_3) \approx^1 (\mathbf{A}_1 \vee \mathbf{A}_2)$;
- (iv) Since $\mathbf{A}_2 >^1 \mathbf{A}_4$, $(\mathbf{A}_1 \vee \mathbf{A}_2) >^1 (\mathbf{A}_1 \vee \mathbf{A}_4)$.

We thus have:

Substitution

If $C_k \approx^i C_j$ then $A^{kj} \approx^i A$, and if $C_k >^i C_j$ then $A^{kj} >^i A$.

The first component of *Substitution* is satisfied by every account mentioned above. The second is a little more demanding. It is violated by those accounts—like *min*, *max*, and *min-max*—that are insensitive to substitutions of cells between the extreme cells. On the other hand *ave* and *min-sum*, amongst others, are appropriately sensitive to such substitutions.

The combination of *Weak Proximity* and *Substitution* I will call *Strict Proximity*. The reason is that *Weak Proximity* cannot deliver anything other than a weak inequality when sometimes a strict inequality in order. For example, by *Weak Proximity*:

- (v) $\mathbf{A}_1 \approx^1 (\mathbf{A}_1 \vee \mathbf{A}_4)$.

Intuitively, however, A_1 is more accurate than $(A_1 \vee A_4)$. But by *Weak Proximity*:

$$(vi) \quad A_1 \geq^1 (A_1 \vee A_2).$$

Since by *Substitution* $A_2 >^1 A_4$ yields $(A_1 \vee A_2) >^1 (A_1 \vee A_4)$, *Strict Proximity* yields $A_1 >^1 (A_1 \vee A_4)$.

5 Accuracy For Credal States

Not all credal states consist in the acceptance of propositions and not all developments in an inquiry are best represented by changes in propositions accepted. Suppose one begins an inquiry ignorant as to which cell of a partition contains the actual world. After gathering evidence (by tapping the barometer say) it may be that no particular cell is definitively ruled out, but that uncertainty is reduced. Some cells now seem more likely than others. Even though no new proposition has been accepted, one's credal state has changed, and such a change may constitute progress towards the truth.

Credal states can be represented as probability distributions over the cells of the inquiry. The aim of the inquiry is to end up embracing the perfectly accurate probability distribution over I . Where C_i is the true cell, the target credal state is the opinionated state, P^i , that assigns maximal probability to C_i and zero to the rest. If we had a good measure of the distance of credal states from each of the opinionated states, inaccuracy could be captured as distance from the target credal state. Ascribing accuracy value directly to credal states, rather than to propositions, or to actions such as accepting propositions, marked a crucial development in the EU program.¹³

There are well-known pragmatic justifications for probabilism (e.g. the Dutch Book), for conditionalization (e.g. the diachronic Dutch Book), and for the value of experimenting (e.g. the Ramsey-Good theorem), but are there comparable purely *cognitive* justifications. In a recent survey Pettigrew writes:

In his 1998 paper, Joyce introduced an entirely novel style of argument for epistemic norms: characterize the legitimate measures of the epistemic utility of an epistemic state at a world; then apply decision principles to this function to derive epistemic norms that govern these states. The strategy is to prove that, given undeniable features of accuracy, credal states that do not satisfy probabilism are *dominated* by credal states that do.¹⁴

The EU vindication of probabilism relies on the following claims:

¹³ Pettigrew [2015] p. 2 dates this development to Oddie [1997] and Joyce [1998].

¹⁴ Pettigrew [2013], p. 907. In Oddie [1997] the strategy was employed to vindicate the norms of seeking out new information and updating by conditionalization.

- (1a) If all legitimate measures of accuracy of credal states satisfy conditions S then probabilism is true.
- (1b) All legitimate measures of accuracy of credal states (and there are some) satisfy S .

I argue for the the following conditional:

- (2) If probabilism is true no legitimate measure of the accuracy of credal states satisfies S .

It follows from (1a) and (2) that no measure satisfying S is a legitimate measure of accuracy, depriving the vindication of probabilism of premise (1b). I assume the antecedent of (2) and demonstrate the consequent.

Like EU theorists, I am interested in the accuracy value *simpliciter* of credal states. There may be evaluable aspects of an inquiry that are not built into an inquirer's credal state at a time. Her current state might have been arrived at through wishful thinking, logical errors, dumb luck, or by judicious modification in the light of new evidence. These different epistemic *procedures* may differ in their cognitive value, as may the various epistemic trajectories they engender. As elsewhere, the value of a journey may diverge from the value of its destination.

I begin with plausible conditions on accuracy that EU theorists have found compelling. The conjunction of these principles I dub the *Core* of the EU program.¹⁵

Suppose C_i is true. Then P^i 's assignments of probabilities match the truth values of all answers: $P^i(A)=1$ if A is true, $P^i(A)=0$ if A is false. Any other credal state P , whether opinionated or not, assigns different values to some of the answers. $P(A)$ can be thought of as a more or less accurate *estimate* of the truth value of A , or of $P^i(A)$. The closer an estimate of a magnitude is to the actual value of the magnitude, the more accurate the estimate. So P is more accurate than Q with respect to A if $P(A)$ is closer to $P^i(A)$ than is $Q(A)$. A *local inaccuracy measure* $d(p,t)$ is thus a function of p (the probability P assigns to A) and t (the truth value of A), satisfying:

Truth Directedness

For all $p \in [0,1]$, $d(p,1)$ is a strictly decreasing function of p and $d(p,0)$ is a strictly increasing function of p .

Any increasing function of the absolute difference $|t-p|$ satisfies *Truth Directedness*, and the simplest is $|t-p|$ itself (α^1). α^1 seems natural enough, but any exponential function of $|t-p|$ is also truth directed: $\alpha^z(p,t) = |t-p|^z$ for $z > 0$.¹⁶ Of course, many other functions satisfy the constraint—like this logarithmic measure:

¹⁵The *Core* is not inviolable. If it leads to counterintuitive results then there will be reason to tweak it.

¹⁶Absolute difference, α^1 , has been proposed by Horwich [1982], Goldman [1999] and Maher [2002].

$$d(p,1) = -\ln(p), \text{ and } d(p,0) = -\ln(1-p).$$

These functions all have a property that arguably every inaccuracy measure should have. If local accuracy depends *only* on truth value and probability then since, in a coherent believer, believing A to degree p just is believing $\neg A$ to degree $(1-p)$, the inaccuracy of believing A to degree p should be the same as the inaccuracy of believing $\neg A$ to degree $(1-p)$.

Symmetry

For all p and t : $d(p,1) = d(1-p,0)$.¹⁷

Symmetry is not always assumed at the outset in EU vindications of probabilism, for the reason that credal states are not assumed at the outset to be probabilistic. (This is what they want to prove.) For non-probabilistic states the justification for *Symmetry* above would break down. But given that we are assuming probabilism for the purposes of establishing (2) above, *Symmetry* seems irresistible.

By *Truth Directedness*, P is minimally inaccurate with respect to A if and only if $P(A)=P^i(A)$. *Minimal* inaccuracy is naturally taken to be *zero* inaccuracy—a harmless scaling convention—so $d(0,0)=d(1,1)=0$. Another feature of the exponential functions, though not the logarithmic function, is that inaccuracy possesses a finite upper bound for $0 \leq p \leq 1$. Again, setting the upper bound at 1 is a scaling convention, but it is by no means a mere convention to set an upper bound on inaccuracy. Some have endorsed unbounded measures, like the logarithmic function. For ease of exposition I collapse the two assumptions here but we will revisit that:

Boundedness

0 is the greatest lower bound on inaccuracy and $d(p,t) = 0$ if and only if $|t-p| = 0$.

1 is the least upper bound on inaccuracy and $d(p,t) = 1$ if and only if $|t-p| = 1$.

Let $D_i(P)$ be the global inaccuracy of P given C_i . A plausible assumption is that $D_i(P)$ is an increasing function of local inaccuracies. If $D_i(P)$ were the simple sum of local inaccuracies this would clearly be satisfied. But perhaps summing is too simplistic. It treats all answers as equally important, and some answers might be more important to overall accuracy than others. Let λ_j be a weight assigned to answer A_j (in some enumeration of all answers) in the evaluation of the inaccuracy of P . The more important it is to get A_j right, the more P 's error with respect to A_j counts in overall global inaccuracy. Not every assignment of weights is admissible—about which a lot more will be said—but the basic proposal is that the global inaccuracy of a credal state is the weighted sum of local inaccuracies, using some admissible λ -assignment:

Additivity

Global inaccuracy is the λ -weighted sum of local inaccuracies:

¹⁷ This condition is called “0/1 Symmetry” in Joyce [2009].

$$D_i^\lambda(P) = \sum_j \lambda_j d(P(A_j), P^i(A))$$

where λ is an admissible assignment of weights.

I will interpret *Additivity* as a constraint on accuracy that works in both directions.¹⁸ Any legitimate measure of local inaccuracy d together with an admissible λ -assignment generates a legitimate measure of global inaccuracy D^λ . Further, any legitimate global inaccuracy measure is generated by a legitimate local accuracy measure d and an admissible λ by weighted summing.

When is λ admissible? Getting a particular answer wrong cannot uniformly *enhance* global accuracy. So all admissible λ -assignments have to be non-negative. Further, if all answers were assigned zero weight, no credal state would be more accurate than any other—total trivialization. So some λ -assignments have to be positive. Requiring that the λ -assignments sum to one entails this, and is otherwise a scaling convention. Finally, some positive weight has to be assigned to answers other than the tautology. Every coherent credal state assigns the tautology its true value, so accuracy with respect to the tautology never sets credal states apart. If the tautology hogged all the weight then again total trivialization would ensue. Summing this up:

Admissibility

λ is *admissible* only if: (i) for all j , $\lambda_j \geq 0$; (ii) $\sum_i \lambda_i = 1$; and (iii) $\lambda_{Taut} < 1$.

This is a necessary condition on admissible λ s. Further constraints will be in order depending on the nature of the λ -weighting at issue.

This initially plausible looking *Core* (*Truth Directedness*, *Symmetry*, *Boundedness*, *Additivity*, and *Admissibility*) leaves open a large pool of local and global inaccuracy functions.

Satisfying the *Core* core isn't a sufficient condition for D to count as a legitimate measure of accuracy. Neither does it constitute a set of conditions sufficient to underwrite an accuracy argument for probabilism. As noted, absolute difference, α^1 satisfies the *Core*, but it doesn't underwrite a dominance argument for probabilism. Further desiderata are needed, and several have been proposed, including *Extensionality*, *Convexity* and *Propriety*. One prominent accuracy measure that satisfies the *Core*, along with these other features, is the quadratic measure: α^2 . Plugging α^2 into *Additivity* we get the Brier measures. Let *Brier* be the thesis that some member of this class is an adequate measure of accuracy.

Brier

There is an admissible λ such that $Br_i^\lambda(P) = \sum_j \lambda_j \alpha^2(P(A_j), P^i(A))$ is a legitimate measure of inaccuracy.

Simple Brier is the special case in which the λ -coefficients are equal (λ^{equal}).

¹⁸ See Pettigrew [2015], pp. 39-40 and 52-4 for this kind of interpretation and defense of additivity.

The Brier measures, particularly Simple Brier, have played a special role in the EU program and are widely regarded as obviously legitimate accuracy measures, or even as the only legitimate measures.¹⁹ However, it would be advantageous not to have to rely on the legitimacy of these measures, with weaker conditions on legitimacy that can do the work. The central condition currently favored for this role is *Propriety*.²⁰

6 Propriety

$D_i^\lambda(P)$ is a measure of the global inaccuracy of P given that C_i is true. If one is not in possession of the truth, then typically one is also ignorant of the actual global accuracy of one's credal state. Nevertheless one can still *estimate* the inaccuracy of a credal state. The expected inaccuracy of Q , from the perspective of P , is P 's best estimate of the inaccuracy of Q :

$$P\text{'s estimate of } Q\text{'s inaccuracy} = ED_p^\lambda(Q) = \sum_j P(C_j)D_j^\lambda(Q).$$

P is *self-recommending* if P 's estimate of its own inaccuracy is less than its estimate of the inaccuracy of any other state Q .²¹ If P is self-recommending then staying in P would minimize expected inaccuracy. Suppose P is not self-recommending. Perhaps this is because P estimates the inaccuracy of some rival state Q to be less than that of P . If one were in state P then P recommends that to minimize inaccuracy one switch to Q straight away without benefit of additional evidence. Call such a P *self-undermining*. At first blush there is something very odd about a self-undermining credal state. This suggests the following constraint on accuracy:

Weak Propriety

For every P and Q , $ED_p^\lambda(P) \leq ED_p^\lambda(Q)$.

Suppose P is neither self-recommending nor self-undermining. P estimates its own inaccuracy to be no greater than that of any rival, but it does put some rivals on an equal footing. Call such states *self-deprecating*. Self-deprecating states are not as bad as self-undermining states, but they still seem odd. This suggests:

Strict Propriety

For every P and $Q \neq P$, $ED_p^\lambda(P) < ED_p^\lambda(Q)$.

¹⁹ di Finetti [1974] and Savage [1971], Joyce [1998] and [2009], Oddie [1997], Greaves and Wallace 2007, Leitgeb and Pettigrew [2010] and Pettigrew [2015].

²⁰ In his [1998] Joyce uses *convexity* as a central condition on legitimacy but in light of criticisms in Maher [2002] and Pettigrew [2015], convexity has been largely abandoned in favor of conditions like *Propriety*.

²¹ Oddie [1997].

Given *Additivity*, a global function D^λ satisfies these conditions if and only if the local accuracy function d that generates D^λ satisfies the following:

Weakly proper d: for all $q \neq p$: $pd(p,1) + (1-p)d(p,0) \leq pd(q,1) + (1-p)d(q,0)$.

Strictly proper d: for all $q \neq p$: $pd(p,1) + (1-p)d(p,0) < pd(q,1) + (1-p)d(q,0)$.

There are a wide range of proper measures, including the exponential functions and the logarithmic function, amongst many others.²²

7 Proximity For Credal States

Proximity is a constraint on propositional accuracy, but there is an obvious analogue for credal state accuracy. Call P an *A-state* if it assigns non-zero credences to all and only *A*-cells. Let C be a most accurate *A*-cell. P^C is the opinionated *C*-state. In moving from P to P^C one transfers credences from less accurate *A*-cells and concentrates them all on the optimally accurate *A*-cell, C . The credal correlate of *Weak Proximity* is this:

Weak Proximity for Credal States

If P is any *A*-state and $C \in \text{Closest}_i(A)$ then $D_i^\lambda(P^C) \leq D_i^\lambda(P)$.

Call P *flat* if it distributes probabilities evenly over some cells, and let P^A be the flat *A*-state. P^{-T} , the flat $\neg T$ -state, is not a terribly accurate credal state. Let $C \in \text{Closest}_i(\neg T)$. All that *Weak Proximity* requires is that the step from P^{-T} to P^C does not increase inaccuracy.

The generalization of *Substitution* for credal states is also compelling. Suppose $P(C_j) > 0$ and $P(C_k) = 0$. Let $P^{k/j}$ be just like P except that $P^{k/j}(C_j) = 0$ and $P^{k/j}(C_k) = P(C_j)$.

Substitution for Credal States

$C_k \approx^i C_j$ then $D_i^\lambda(P^{k/j}) = D_i^\lambda(P)$, and if $C_k >^i C_j$ then $D_i^\lambda(P^{k/j}) < D_i^\lambda(P)$.

Strict Proximity for credal states is the combination of *Substitution* and *Weak Proximity*.

In I^W , let \mathbf{P}^i be the flat state that assigns an equal non-zero credence to each cell compatible with \mathbf{A}_i and 0 to the rest. $\mathbf{A}_2 \approx^1 \mathbf{A}_3$ and $\mathbf{A}_1 >^1 \mathbf{A}_4$. By *Substitution* it follows that $\mathbf{P}_5 \approx^1 \mathbf{P}_6$ and $\mathbf{P}_5 >^1 \mathbf{P}_{10}$.

A special case of *Substitution* involves opinionated states. Let P^j assign 1 to C_j . If C_j and C_k are equally accurate then, by *Substitution*, so too are the opinionated states, P^j and $P^k (= (P^j)^{k/j})$. If C_k is more accurate than C_j then P^k is more accurate than P^j .

Opinionation

If $C_k \approx^i C_j$ then $D_i^\lambda(P^k) = D_i^\lambda(P^j)$ and if $C_k >^i C_j$ then $D_i^\lambda(P^k) < D_i^\lambda(P^j)$.

²²Joyce [2009], pp. 276-8, gives a recipe for generating proper measures but they do not all satisfy the core.

Since $A_2 \approx^1 A_3$ and $A_1 >^1 A_4$, *Opinionation* yields $P^2 \approx^1 P^3$ and $P^1 >^1 P^4$.

8 Extensionality

In his [1998] Joyce introduces *Extensionality* as a possible condition on an accuracy measure.

The "facts" which a person's partial beliefs must "fit" are exhausted by the truth-values of the propositions believed, and the only aspect of her opinions that matter is their strengths.²³

Joyce sees that propositional accuracy might well conflict with *Extensionality*:

A[n] objection to *Extensionality* is that it does not take verisimilitude into account. Here is how the complaint might go:

Copernicus (let us suppose) was exactly as confident that the earth's orbit is circular as Kepler was that it is elliptical. However, both were wrong since the gravitational attraction of the moon and the other planets causes the earth to deviate slightly from its largely elliptical path. *Extensionality* rates the two thinkers as equally inaccurate since both believed a falsehood to the same high degree. Still Kepler was obviously nearer the mark, which suggests that evaluations of accuracy must be sensitive not only to the truth-values of the propositions involved, but also to how close false propositions come to being true.²⁴

Joyce's reply:

An agent who strongly believes that the earth's orbit is elliptical will also strongly believe many more truths than a person who believes that it is circular (e.g., that the average distance from the earth to the sun is different in different seasons). This means that the overall effect of Kepler's inaccurate belief was to improve the extensional accuracy of his system of beliefs as a whole. ...I suspect that most intuitions about falsehoods being "close to the truth" can be explained in this way, and that they therefore pose no real threat to *Extensionality*.

If Joyce were right then an equal distribution of λ -weights would capture the desired ranking and *Extensionality* would be preserved. We can test this with opinionated states in the weather framework. In *Table 4*, λ^{equal} assigns equal weights to all fifteen answers. Any two false complete answers C_j and C_k entail the same number of true answers (8) and the same number of false answers (7), and as a result the associated opinionated states end up with the same score. This is a quantitative version of the problem which crippled Popper's qualitative account, and it

²³ Joyce [1998], p. 591.

²⁴ Joyce [1998], p. 592.

generates a similar trivialization result here: namely, that no false opinionated state is more accurate than any other false opinionated state.²⁵

Table 4 $D^{\lambda^{equal}}$ violates *Opinionation*

Since λ^{equal} assigns equal weights, and \mathbf{P}^4 and \mathbf{P}^2 both deliver 8 answers with maximal inaccuracy and 7 answers with minimal inaccuracy, we have $D_T^{\lambda^{equal}}(\mathbf{P}^4) = D_T^{\lambda^{equal}}(\mathbf{P}^2) = 8/15$. This suggests the EU theorist needs to abandon equal weights. Some answers must count more than others in overall accuracy. If so the distances between complete answers might be packed into a suitable selection of λ . This idea is floated in Joyce [2009], and by Greaves and Wallace 2007. The latter explicitly advocate distributing λ -weights unequally in order to capture truthlikeness:

...often, we will want to judge one credence distribution as epistemically better than another even when both assign the same degree of belief to the true state, on the grounds that the first concentrates its remaining credence among (false) states that are closer to the truth than does the second. Our sample schema [*viz Brier*] takes account of the value of verisimilitude, by a judicious choice of the coefficients λ : we simply assign high λ when A is a set of ‘close’ states.²⁶

Note that the judgement endorsed in the first sentence of the paragraph hints strongly at *Strict Proximity*. Greaves and Wallace intimate that such considerations can be captured by assigning a high λ -value to “a set of ‘close’ states”. They don’t spell what they mean by “a set of ‘close’ states”, and there are two possible interpretations

On the first interpretation a set of ‘close’ states is a set of cells close to the true cell. That is, λ_j^i —the weight assigned to A_j when C_i is true—depends on how close A_j is to C_i . λ_j^i could be simply identified with the truthlikeness of A_j given C_i . More liberally, λ_j^i could be any strictly increasing function of truthlikeness. It would be surprising if there were no such likeness-sensitive world-dependent λ -weightings compatible with *Proximity*.

There are, however, a couple of problems with this world-dependent interpretation. The authors call the λ s “constant coefficients”. A natural interpretation of this is that λ_j does not vary with the location of the true cell. Rather, it is a world-independent affair. This reading is bolstered by a sketch of a proof of the *Propriety* of the Brier measures (in their footnote 9). The proof is valid only if the λ -coefficients are world-independent. If the the λ -weights are world independent and the local inaccuracy function d is weakly/strictly proper, then the so too is the

²⁵A more general trivialization result: if A and B are false answers with the same number of cells then the number of truths (falsehoods) that A entails is the same as the number that B entails.

²⁶ Greaves and Wallace [2006], p. 628.

global inaccuracy function obtained by summing weighted local inaccuracies. There is no such guarantee if the weights are world-dependent.

The main problem with the world-independent interpretation is that it is not immediately obvious what is supposed to count as a set of 'close' states. There is, however, one rather plausible construal. Suppose we assign high λ -coefficients to those answers in which the cells are close, not to the true cell, but to one another. We can flesh out this idea using the notion of a *convex* region. Suppose we have a notion of betweenness on cells.²⁷ A convex set of cells is one that contains all the cells located between any two cells in the set. For example: $\mathbf{A}_5 = \mathbf{h} = \{\mathbf{A}_1, \mathbf{A}_2\}$ and $\mathbf{A}_6 = \mathbf{r} = \{\mathbf{A}_1, \mathbf{A}_3\}$ are both convex since \mathbf{A}_1 is close to \mathbf{A}_2 , and \mathbf{A}_1 is close to \mathbf{A}_3 (they are both adjacent pairs); whereas $\mathbf{A}_7 = \mathbf{h}\equiv\mathbf{r} = \{\mathbf{A}_1, \mathbf{A}_4\}$ and $\mathbf{A}_8 = \neg\mathbf{h}\equiv\mathbf{r} = \{\mathbf{A}_2, \mathbf{A}_3\}$ are not, since \mathbf{A}_1 is maximally distant from \mathbf{A}_4 , as \mathbf{A}_2 is from \mathbf{A}_3 . One λ -weighting compatible with this interpretation of 'close' states (λ^{convex}) assigns equal weights to all the convex regions of the weather space and zero weight to the non-convex regions. Another (λ^{atomic}) assigns equal weights to certain special convex regions—namely, the atomic propositions and their negations, and zero to the non-atomic states. Table 5 gives the local and global inaccuracies using λ^{convex} and λ^{atomic} for weights, and a d satisfying the *Core*.

Table 5: λ^{convex} and λ^{atomic} yield *Opinionation*

As is easily checked summing local errors weighted by λ^{convex} and λ^{atomic} delivers the right global ordering of opinionated states:

$$D_T^{\lambda^{convex}}(\mathbf{P}^4) = 6/9 > D_T^{\lambda^{convex}}(\mathbf{P}^2) = 4/9, \quad D_T^{\lambda^{atomic}}(\mathbf{P}^4) = 1 > D_T^{\lambda^{atomic}}(\mathbf{P}^2) = 1/2.$$

Thus with appropriate world-independent λ -weightings, the *Core* yields *Opinionation*. The question now is whether, with admissible λ -weightings, we can recover the intuitively correct accuracy ordering on credal states generally, not just the opinionated states. To settle that we first have to specify which λ -weightings are admissible on each of the two interpretations.²⁸

²⁷The notion of convexity has proved very rich in recent work in property theory (Gärdenfors [2000]). See Oddie [1987] for an earlier application of convexity in the TL program that shows the connection between convex states and atomic states. Interestingly Popper also intimated that convexity might provide an escape hatch for his beleaguered qualitative account of truthlikeness.

²⁸In an early draft I ignored world-dependent weightings, in part because one of the co-authors told me that this is what they had had in mind. An anonymous referee persuaded me that whatever the authors' intention world-dependent assignments are a more promising way to go to accommodate *Proximity*.

9 Admissible Weightings

On the world-dependent interpretation, an admissible λ -assignment is an increasing function of accuracy. A local error—like underestimating the probability of a true proposition—is thus more egregious the closer A is to the true cell C_i .²⁹

WD-Admissibility (for world-dependent λ)

λ is *WD-admissible* if and only if it is admissible, and for all A, B, i :

if $A_j \approx^i A_i$ then $\lambda_j^i = \lambda_k^i$, and if $A_j >^i A_k$ then $\lambda_j^i > \lambda_k^i$.

On the world-independent construal, a λ -assignment can assign different weights, subject to the constraint that the assignment not assign different weights to propositions with the same world-independent *distance structure*. Distance structure (like convexity) is an internal feature of a proposition which depends only the relationships among the cells of the proposition and their relationships to cells outside the proposition. Looking at figure 1, for example, it is clear that the edges of the square all share the same distance structure, as do the cells and the negations of the cells. To be maximally charitable to the EU program, we will allow as WI-admissible all assignments that supervene on distance structure. First, we define having the same distance structure: A and B over I with distance function δ have the *same distance structure* ($A \approx_\delta B$) if and only there is an automorphism f on $\langle I, \delta \rangle$ such that $f(A) = B$.³⁰

WI-Admissibility

λ is *WI-admissible* if and only if it is admissible, and

for all A_j and A_k , if $A_j \approx_\delta A_k$ then $\lambda_j = \lambda_k$.

The \approx_δ -equivalence classes in I^W under the city block measure are:

- G₁** Complete answers: **A₁, A₂, A₃, A₄**;
- G₂** Atomic propositions and their negations: **A₅, A₆, A₉, A₁₀**;
- G₃** Biconditionals of atomic propositions and their negations: **A₇, A₈**;
- G₄** Negations of complete answers: **A₁₁, A₁₂, A₁₃, A₁₄**.
- G₅** The tautology: **A₁₅**.

Propositions in **G₁**, **G₂**, **G₅** are all convex, those in **G₃** and **G₄** are not. Propositions in **G₂** are atomic, those in the other classes are not. Hence both convexity and atomicity supervene on distance structure, as do the associated equal weightings of the convex and the atomic propositions— λ^{convex} and λ^{atomic} . Different metrics typically produce different equivalence classes. For example, suppose all cells are equidistant from one another (*viz.* a miss is as good

²⁹ I owe this suggestion to an anonymous referee.

³⁰ An analogous definition works for inquiries which have an accuracy ordering on the cells but no metric.

as a mile). Then answers belong to the same equivalence class just in case they have the same number of cells (i.e. \mathbf{G}_2 and \mathbf{G}_3 combine into a single equivalence class).

Is WI-admissibility too restrictive? For example, it has been suggested that answers might count more toward accuracy the more informative they are.³¹ Or there may be a reason to assign different weights to, say, different atomic propositions.

A proposition is stronger the more cells it rules out. If $A \approx^\delta B$ then A and B rule out the same number of cells. Thus strength, like atomicity and convexity, supervenes on distance structure, so there are WI-admissible λ s that favor strength, as well as those that favor weakness, or any combination of factors that supervene on distance structure.

Differences in the weights of atomic propositions are also possible but go hand in hand with asymmetries in distance structure. Suppose temperature is twice as important as precipitation in accuracy assessments. It should receive double the λ -weighting. In that case the distance of $\neg \mathbf{h} \wedge \mathbf{r}$ from $\mathbf{h} \wedge \mathbf{r}$ should be double the distance of $\mathbf{h} \wedge \neg \mathbf{r}$ from $\mathbf{h} \wedge \mathbf{r}$. \mathbf{h} and \mathbf{r} would thus have different distance structures in such a space.

10 Propriety Violates Proximity

We will use the example in *Table 6*. \mathbf{A}_1 is true. \mathbf{P}^i is the flat \mathbf{A}_i -state. \mathbf{P}^1 , \mathbf{P}^2 , and \mathbf{P}^4 are opinionated states, and \mathbf{P}^1 is the target. \mathbf{P}^8 assigns 1/2 to each of \mathbf{A}_2 and \mathbf{A}_3 . \mathbf{P}^9 assigns 1/2 to each of \mathbf{A}_2 and \mathbf{A}_4 . \mathbf{P}^{14} assigns 1/3 to each of \mathbf{A}_2 , \mathbf{A}_3 and \mathbf{A}_4 . d is any local inaccuracy function satisfying the *Core*. Rows 1-5 specify the probabilities assigned by the five credal states, and 6-9 specify the local inaccuracies of the false credal states. *Symmetry* entails $d(1/2,0)=d(1/2,1)=i$, $d(1/3,0)=d(2/3,1)=j$, and $d(2/3,0)=d(1/3,1)=k$, for some i, j, k . The global inaccuracy of each credal state is obtained by the weighted summing of local inaccuracies, where the weight assigned to \mathbf{A}_i is of course λ_i . The global inaccuracy $D_T^\lambda(P)$ of each of the credal states under consideration can be read easily off the table:

- (i) $D_T^\lambda(\mathbf{P}^2) = \lambda_1 + \lambda_2 + \lambda_6 + \lambda_7 + \lambda_8 + \lambda_9 + \lambda_{13} + \lambda_{14} .$
- (ii) $D_T^\lambda(\mathbf{P}^8) = \lambda_1 + i\lambda_2 + i\lambda_3 + i\lambda_5 + i\lambda_6 + \lambda_7 + \lambda_8 + i\lambda_9 + i\lambda_{10} + i\lambda_{12} + i\lambda_{13} + \lambda_{14} .$
- (iii) $D_T^\lambda(\mathbf{P}^9) = \lambda_1 + i\lambda_2 + i\lambda_4 + i\lambda_5 + \lambda_6 + i\lambda_7 + i\lambda_8 + \lambda_9 + i\lambda_{10} + i\lambda_{11} + i\lambda_{13} + \lambda_{14} .$
- (iv) $D_T^\lambda(\mathbf{P}^{14}) = \lambda_1 + j\lambda_2 + j\lambda_3 + j\lambda_4 + k\lambda_5 + k\lambda_6 + k\lambda_7 + k\lambda_8 + k\lambda_9 + k\lambda_{10} + i\lambda_{11} + i\lambda_{12} + i\lambda_{13} + \lambda_{14} .$

Table 6 Example

³¹ For example, Joyce [2009], p. 293. "Some legitimate modes of epistemic evaluation will surely focus on things other than pure accuracy, e.g., some will require us to weight propositions by their informativeness". By "pure accuracy" Joyce means the closeness of a probability assignment to the actual truth value. It would of course beg the question to assume that that is all there is to genuine accuracy.

We show first that simple Brier comes into direct and immediate conflict with *Weak Proximity* which entails $D_T^\lambda(\mathbf{P}^2) \leq D_T^\lambda(\mathbf{P}^{14})$. On the quadratic measure α^2 : $i=1/4, j=1/9$ and $k=4/9$. From this, (i) and (iv), the inaccuracies assigned by simple Brier are:

$$Br_T^{\lambda^{equal}}(\mathbf{P}^2) = 0.53 > Br_T^{\lambda^{equal}}(\mathbf{P}^{14}) = 0.35 .$$

This is a striking example of the failure of simple Brier to capture the most elementary feature of accuracy—that the closest $\neg T$ -cells are no less accurate than $\neg T$. By itself, however, this does not impugn the EU approach. There are many other measures of local inaccuracy that satisfy the *Core* and *Propriety*, and there many admissible non-equal weightings, both world-dependent and world-dependent. What we need to show is that for *any* legitimate measure of local accuracy combined with *any* admissible weighting, if *Propriety* holds *Proximity* fails. And if we can show this even for *Weak Propriety* and *Weak Proximity* so much the better.³²

The following lemma will prove useful.

Lemma For any weakly proper local accuracy function d that satisfies the core, $i < 1/2$.

If d is weakly proper: for all p, q : $pd(p,1) + (1-p)d(p,0) \leq pd(q,1) + (1-p)d(q,0)$. (a)

Let $q=0$ in (a): for all p : $pd(p,1) + (1-p)d(p,0) \leq pd(0,1) + (1-p)d(0,0)$.

Boundedness: $d(0,1)=1$ and $d(0,0)=0$.

So: for all p , $pd(p,1) + (1-p)d(p,0) \leq p$. (b)

Let $p=1/2$ in (b): $d(1/2,1) + d(1/2,0) \leq 1$.

By *Symmetry*: $d(1/2,1) = d(1/2,0)$.

So: $i = d(1/2,1) \leq 1/2$.

Suppose for a reductio that: $i = 1/2$.

Let $p=1/2$ in (a): for all q , $1/2d(1/2,1) + 1/2d(1/2,0) \leq 1/2d(q,1) + 1/2d(q,0)$.

That is: for all p , $1 \leq d(p,1) + d(p,0)$. (c)

From (b), (c): for all p , $d(p,1) + \frac{(1-p)}{p}d(p,0) \leq 1 \leq d(p,1) + d(p,0)$.

Which entails: for all p , $\frac{(1-p)}{p}d(p,0) \leq d(p,0)$. (d)

By *Truth Directedness* and *Boundedness*: if $0 < p < 1/2$ then $d(p,0) > 0$ and $\frac{(1-p)}{p} > 1$. (e)

(e) contradicts (d), so: $i < 1/2$.

Now for the main result.

³²Fallis and Lewis [2015] voice a rare criticism of simple Brier, citing a Brier ranking of two credal states they deem counterintuitive. Their argument, which is different from and independent of mine here, is based on a monotonicity principle, M4. I will argue elsewhere that M4 leads to counterintuitive results once not every false cell is deemed equally inaccurate.

Theorem **Weak Propriety violates Proximity.**

There are two cases:

(I) λ is *WI-Admissible*: *Weak Propriety* violates *Weak Proximity*.

Where $\mathbf{A}_1=T$ we have: $\mathbf{A}_2 \in \text{Closest}_T(\mathbf{A}_8)$, $\mathbf{A}_2 \in \text{Closest}_T(\mathbf{A}_9)$. (a)

(a), *Weak Proximity*: $D_T^\lambda(\mathbf{P}^2) \leq D_T^\lambda(\mathbf{P}^8)$ and $D_T^\lambda(\mathbf{P}^2) \leq D_T^\lambda(\mathbf{P}^9)$. (b)

By *WI-Admissibility*: $\lambda_1=\lambda_2=\lambda_3=\lambda_4$; $\lambda_5=\lambda_6=\lambda_9=\lambda_{10}$; $\lambda_7=\lambda_8$; $\lambda_{11}=\lambda_{12}=\lambda_{13}=\lambda_{14}$. (c)

From (i) and (c): $D_T^\lambda(\mathbf{P}^2) = 2(\lambda_1 + \lambda_5 + \lambda_7 + \lambda_{11})$. (d)

From (ii) and (c): $D_T^\lambda(\mathbf{P}^8) = (1+2i)\lambda_1 + 4i\lambda_5 + 2\lambda_7 + (1+2i)\lambda_{11}$. (e)

From (iii) and (c): $D_T^\lambda(\mathbf{P}^9) = (1+2i)\lambda_1 + (2+2i)\lambda_5 + 2i\lambda_7 + (1+2i)\lambda_{11}$. (f)

From (b), (d), (e): $\lambda_1 + 2\lambda_5 + \lambda_{11} \leq 2i\lambda_1 + 4i\lambda_5 + 2i\lambda_{11}$. (g)

Lemma, *Weak Propriety*: $i < 1/2$. (h)

Admissibility: $\lambda_i \geq 0$ (for all i). (i)

From (g), (h), (i): $\lambda_1 = \lambda_5 = \lambda_{11} = 0$. (j)

From (d), (f), (j): $D_T^\lambda(\mathbf{P}^2) = 2\lambda_7$, $D_T^\lambda(\mathbf{P}^9) = 2i\lambda_7$. (k)

By *Admissibility* $\sum_i \lambda_j = 1$ and $\lambda_{15} < 1$. (l)

From (c), (j), (l): $\lambda_7 > 0$. (m)

From (h), (k), (m): $D_T^\lambda(\mathbf{P}^9) < D_T^\lambda(\mathbf{P}^2)$. (n)

(b) contradicts (n).

(II) λ is *WD Admissible*: *Weak Propriety* violates *Strict Proximity*

(Recall that *Strict Proximity* is *Weak Proximity* plus *Substitution*. Abbreviate λ_j^1 to λ_j .)

From the I^W model we have: $\mathbf{A}_2 \approx^1 \mathbf{A}_3$ and $\mathbf{A}_1 >^1 \mathbf{A}_4$.

By *Substitution* & $\mathbf{A}_2 \approx^1 \mathbf{A}_3$: $\mathbf{A}_2 \approx^1 \mathbf{A}_3$, $\mathbf{A}_5 \approx^1 \mathbf{A}_6$, $\mathbf{A}_9 \approx^1 \mathbf{A}_{10}$, $\mathbf{A}_{12} \approx^1 \mathbf{A}_{13}$. (a)

By *WD-Admissibility*, (a): $\lambda_2=\lambda_3$; $\lambda_5=\lambda_6$; $\lambda_9=\lambda_{10}$; $\lambda_{12}=\lambda_{13}$. (b)

Weak Proximity: $D_T^{\lambda^1}(\mathbf{P}^2) - D_T^{\lambda^1}(\mathbf{P}^8) \leq 0$. (c)

By (i), (ii), (c): $(1-i)\lambda_2 - i\lambda_3 - i\lambda_5 + (1-i)\lambda_6 + (1-i)\lambda_9 - i\lambda_{10} - i\lambda_{12} + (1-i)\lambda_{13} \leq 0$. (d)

From (b), (d): $(1-2i)\lambda_2 + (1-2i)\lambda_5 + (1-2i)\lambda_9 + (1-2i)\lambda_{12} \leq 0$. (e)

Weak Propriety, Lemma: $i < 1/2$, i.e. $(1-2i) > 0$. (f)

Admissibility: For all j : $\lambda_j \geq 0$. (g)

(b), (e), (f), (g): $\lambda_2=\lambda_3=\lambda_5=\lambda_6=\lambda_9=\lambda_{10}=\lambda_{12}=\lambda_{13}=0$. (h)

By *Substitution*, $\mathbf{A}_1 >^1 \mathbf{A}_4$: $(\mathbf{A}_1 \vee \mathbf{A}_2) >^1 (\mathbf{A}_2 \vee \mathbf{A}_4)$ (i)

That is: $\mathbf{A}_5 >^1 \mathbf{A}_{10}$. (j)

(j), *WD-Admissibility*: $\lambda_5 > \lambda_{10}$. (k)

(k) contradicts (h).

Thus all WI-admissible core measures of inaccuracy that satisfy *Weak Propriety* violate *Weak Proximity*, and all WD-admissible core measures of inaccuracy that satisfy *Weak Propriety* violate *Strict Proximity*.

11 Possible Responses

11.1 Retreat to convexity

In his [1998] Joyce used the convexity of the inaccuracy measure, rather than *Propriety*, as a constraint on inaccuracy functions and he gave an interesting argument that any inaccuracy function should indeed be convex.³³ Any strictly convex d that satisfies *Core* also satisfies $d(1/2,1) < 1/2$, and the above incompatibility result follows. If, however, d is merely weakly convex then we only have $d(1/2,1) \leq 1/2$, and that is insufficient to generate the incompatibility result. However the weak convexity of d is also too weak for the vindication of probabilism. The absolute difference measure α^1 is weakly convex and does not underwrite the required dominance argument. So strict convexity is too strong for a legitimate inaccuracy measure (it is incompatible with *Weak Proximity*), while weak convexity is too weak (it cannot underwrite the vindication of probabilism). Of course, even though the weak convexity of inaccuracy will not serve this purpose, it may nevertheless be a feature that a genuine measure of accuracy should possess. The weak convexity of D^λ , even in conjunction with the *Core*, poses no threat to *Weak Proximity* or *Substitution*. It is thus compatible with *Strict Proximity*.

11.2 Reject boundedness

Perhaps we can have both *Proximity* and *Propriety* by revising the *Core*. *Truth Directedness* is indispensable to the EU approach. *Symmetry* is unimpeachable for probabilistic credal states. That leaves *Boundedness* and *Additivity*.³⁴

Let d be unbounded: $d(1,0)=d(0,1)=\infty$. Suppose \mathbf{P}^1 is the target state let P be any other opinionated state ($P(\mathbf{A}_1)=0$).

Let λ be WD-admissible. Then $\lambda_1 > \lambda_j$ for all $j \neq 1$, so $\lambda_1 > 0$. $P(\mathbf{A}_1)=0$, $\mathbf{P}^1(\mathbf{A}_1)=1$, and so $D_T^\lambda(P) = \infty$. All false opinionated states are thus equally inaccurate (violating *Opinionation* and hence *Substitution* which entails it) and no more accurate than any non-opinionated state (violating *Strong Proximity*).

Let λ be WI-admissible. There are four possibilities which are exhaustive, but not necessarily exclusive:

³³Joyce calls this *Weak Convexity* in his [1998] and *Convexity* in his [2009]. But the condition he used is strict convexity of d .

³⁴Note that the assumption that the upper bound, if there is one, is 1 is merely a scaling convention. Suppose the upper bound is some $u > 1$. Then we can revise the Lemma: for any weakly proper local accuracy function d that satisfies the core, $i < 1/2u$. With systematic adjustments throughout the main result follows.

- (i) Cells have positive λ -weight. So $\lambda_1 > 0$. $P(\mathbf{A}_1)=0$, $\mathbf{P}^1(\mathbf{A}_1)=1$, so $D_T^\lambda(P) = \infty$.
- (ii) Negations of cells have positive λ -weight. $\mathbf{P}^1(\neg\mathbf{A}_1)=0$ while $P(\neg\mathbf{A}_1)=1$, so $D_T^\lambda(P) = \infty$.
- (iii) Atomic propositions and their negations have positive λ -weight. For at least one such state A , $P(A)=1$ and $\mathbf{P}^1(A)=0$. So again $D_T^\lambda(P) = \infty$.

In (i)-(iii) *Substitution* and *Strong Proximity* are violated. Suppose none of (i)-(iii) holds. Then:

(iv) $\mathbf{A}_7 (= \mathbf{A}_1 \vee \mathbf{A}_4)$ and $\mathbf{A}_8 (= \mathbf{A}_2 \vee \mathbf{A}_3)$ alone have positive and equal λ -weight ($=1/2$). \mathbf{A}_7 and \mathbf{A}_8 alone contribute to inaccuracy. Note: $\mathbf{P}^1(\mathbf{A}_7)=1$ and $\mathbf{P}^1(\mathbf{A}_8)=0$.

- (a) $\mathbf{P}^4(\mathbf{A}_7)=1$; $\mathbf{P}^4(\mathbf{A}_8)=0$. So $D_T^\lambda(\mathbf{P}^4) = 0 = D_T^\lambda(\mathbf{P}^1)$.
- (b) $\mathbf{P}^2(\mathbf{A}_7)=0$, $\mathbf{P}^2(\mathbf{A}_8)=1$. So $D_T^\lambda(\mathbf{P}^2) = \infty$.
- (c) $\mathbf{P}^9(\mathbf{A}_7)=1/2$, $\mathbf{P}^9(\mathbf{A}_8)=1/2$. So $D_T^\lambda(\mathbf{P}^9) = \lambda_7 d(1/2, 1) + \lambda_8 d(1/2, 0) = d(1/2, 1) < \infty$.

By *Substitution*, $D_T^\lambda(\mathbf{P}^4) > D_T^\lambda(\mathbf{P}^2) > D_T^\lambda(\mathbf{P}^1)$ contradicting (a) and (b).

By *Weak Proximity*, $D_T^\lambda(\mathbf{P}^2) \leq D_T^\lambda(\mathbf{P}^9)$, contradicting (b) and (c).

11.3 Reject additivity

The last element one might consider jettisoning is *Additivity*. If P and Q assign the same probability to A then both make the same local error with respect to A and, given *Additivity*, that common local error makes the same contribution to their global scores. But perhaps other features of P and Q should affect the contribution of that local error—such as what P and Q assign to T . If $P(T) > 0$ and $Q(T) = 0$, then maybe the error over A should against Q more than it does against P . For all I have shown, there may be some non-additive, strictly proper, global inaccuracy measure for credal states that also coheres with *Proximity*. This possibility has never been suggested, perhaps because EU theorists have not paid any attention to propositional accuracy and in part because the incompatibility of *Propriety* and *Proximity* has hitherto gone unnoticed. This avenue may be worth exploring although prospects look far from promising.

12 The Upshot

The accuracy of a credal state clearly depends not only on probabilities it assigns to various answers and the actual truth values of those. It also depends on the closeness of cells to the actual cell. Measures that ignore this dimension (like simple *Brier* and any other measure that obeys *Extensionality*) cannot deliver fundamental principles of accuracy. This deficit can be mitigated by allowing propositional accuracy a role in determining the accuracy of credal states—either directly (via world-dependent weights that give a greater weight to accurate propositions), or indirectly (via world-independent weights that reflect proximities). But, as the main result shows, no measure satisfying the *Core* can accommodate both *Propriety* and *Proximity*. The search for an adequate measure of the accuracy of credal states is by no means

over, and a serious question mark now hovers over the standard non-pragmatic vindications of probabilism, conditionalization, the Principal Principle and the value of experimenting.

Acknowledgements

I want to acknowledge helpful comments and criticisms I received from two anonymous referees for the journal, as well as colloquia participants at the University of Stockholm, the Ludwig Maximilian University Munich, and the Institute for Futures Studies—including, in particular, Gustaf Aarhenius, Krister Bykvist, Anandi Hattigadi and Orri Stefansson. For reading and commenting on earlier drafts I would like to thank Dominic Bailey, Don Fallis, Branden Fitelson, Alan Hajék, James Joyce, Cherie Braden and Brian Talbot. I am especially grateful to Richard Pettigrew for generous responses to the queries I lobbed at him over email throughout 2016.

References

- de Finetti, B. [1974]: *Theory of Probability*, vol. 1. New York: John Wiley and Sons.
- Fallis, Don & Lewis, Peter J. [2015]: “The Brier Rule Is not a Good Measure of Epistemic Utility (and Other Useful Facts about Epistemic Betterness)”, *Australasian Journal of Philosophy*, DOI: 10.1080/00048402.2015.1123741
- Gärdenfors, P. [2000]: *Conceptual Spaces: The Geometry of Thought*. Cambridge, Mass.: MIT Press.
- Goldman, A. [1999]: *Knowledge in a Social World*. Oxford: Oxford University Press.
- Greaves, H. and Wallace, D. [2006]: “Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility,” *Mind*, **115**: pp. 607–632.
- Horwich, P. [1982]: *Probability and Evidence*. New York: Cambridge University Press.
- Hempel, C. G. [1960]: “Inductive inconsistencies”, *Synthese*, **12**(4): pp. 439–69.
- Hilpinen, R. [1976]: “Approximate truth and truthlikeness”, in Przelecki, *et al.* (eds.), *Formal Methods in the Methodology of the Empirical Sciences*, Dordrecht: Reidel, pp. 19–42
- Joyce, J. M. [1998]: “A Nonpragmatic Vindication of Probabilism,” *Philosophy of Science*, **65**: pp. 575–603.
- Joyce, J. M. [2009]: “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief”, in F. Huber and C. Schmidt-Petri (eds.), *Degrees of Belief*, Dordrecht: Springer, pp. 263–297.

Kuipers, T. [1987]: “A structuralist approach to truthlikeness”, in *What is Closer-to-the-Truth?*, T. Kuipers (ed.), Amsterdam: Rodopi, pp. 79–99.

—, [1992]: “Naive and refined truth approximation”, *Synthese*, **93**: pp. 299–341.

Leitgeb, H. & Pettigrew, R. [2010]: “An Objective Justification of Bayesianism I: Measuring Inaccuracy,” *Philosophy of Science* **77** (2): pp. 201–235.

Maher, P., [2002]: “Joyce’s Argument for Probabilism,” *Philosophy of Science* **96**: 73–81.

Miller, D. [1974]: “Popper's Qualitative Theory of Verisimilitude”, *The British Journal for the Philosophy of Science*, **25**: 166–177.

Miller, D. [1977]: “On Distance from the Truth as a True Distance”, *Bulletin of the Section of Logic* **6** (1):15–23.

Niiniluoto, I. [1987]: *Truthlikeness*. Dordrecht: Reidel.

Oddie, G. [1987]: “Truthlikeness and the convexity of propositions”, in Kuipers, T. (ed.) *What is Closer-to-the-Truth?* Amsterdam: Rodopi. 197–217.

Oddie, G. [1997]: “Conditionalization, Cogency, and Cognitive Value,” *British Journal for the Philosophy of Science*, **48**: pp. 533–41.

Oddie, G. [2013]: “The content, consequence and likeness approaches to verisimilitude: compatibility, trivialization, and underdetermination,” *Synthese* **190** (9): pp. 1647–1687.

Popper, K. R. [1963]: *Conjectures and Refutations*, London: Routledge.

Pettigrew, R. [2013]: “Epistemic utility and norms for credences,” *Philosophy Compass* **8/10**: pp. 897–908.

Pettigrew, R. [2015]: *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Savage, L. [1971]: “Elicitation of Personal Probabilities”, *Journal of the American Statistical Association* **66**: pp. 783–801.

Teller, P. [2001]: “The Twilight of the Perfect Model Model”, *Erkenntnis*, **55**: pp. 393–415.

Tichý, P. [1974]: “On Popper's definitions of verisimilitude”, *The British Journal for the Philosophy of Science*, **25**: pp. 155–160.

Weston, T. [1992]: “Approximate truth and scientific realism”, *Philosophy of Science*, **59**: pp. 53–74.

I^w	<i>hot</i>	<i>rainy</i>		δ_{ij}	A_1	A_2	A_3	A_4
A_1 ($\mathbf{h} \wedge \mathbf{r}$)	1	1		A_1	0	1	1	2
A_2 ($\mathbf{h} \wedge \neg \mathbf{r}$)	1	0		A_2	1	0	2	1
A_3 ($\neg \mathbf{h} \wedge \mathbf{r}$)	0	1		A_3	1	2	0	1
A_4 ($\neg \mathbf{h} \wedge \neg \mathbf{r}$)	0	0		A_4	2	1	1	0

Table 1 Distance between cells in the weather framework on the city block measure

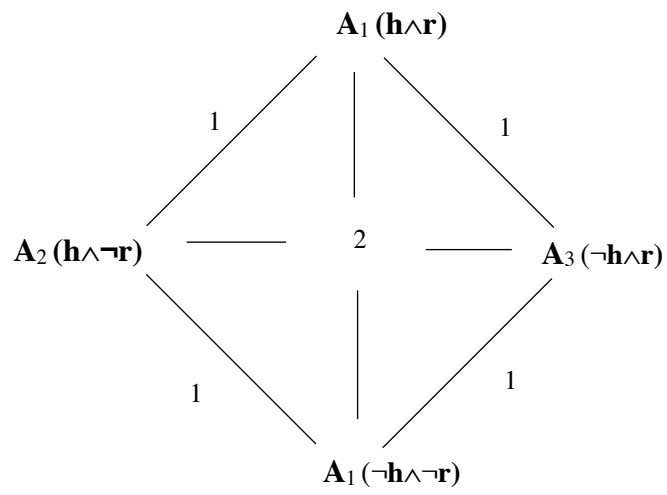


Figure 1 The weather space

A₁	A₂	A₃	A₄	A₅	A₆	A₇	A₈	A₉	A₁₀	A₁₁	A₁₂	A₁₃	A₁₄	A₁₅
$h \wedge r$	$h \wedge \neg r$	$\neg h \wedge r$	$\neg h \wedge \neg r$	h	r	$h \equiv r$	$\neg(h \equiv r)$	$\neg r$	$\neg h$	$h \vee r$	$h \vee \neg r$	$\neg h \vee r$	$\neg h \vee \neg r$	$h \vee \neg h$
A₁	A₂	A₃	A₄	$A_1 \vee A_2$	$A_1 \vee A_3$	$A_1 \vee A_4$	$A_2 \vee A_3$	$A_2 \vee A_4$	$A_3 \vee A_4$	$A_1 \vee A_2$ $\vee A_3$	$A_1 \vee A_2$ $\vee A_4$	$A_1 \vee A_3$ $\vee A_4$	$A_2 \vee A_3$ $\vee A_4$	$A_1 \vee A_2$ $\vee A_3 \vee A_4$

Table 2 Complete and incomplete answers to the weather question

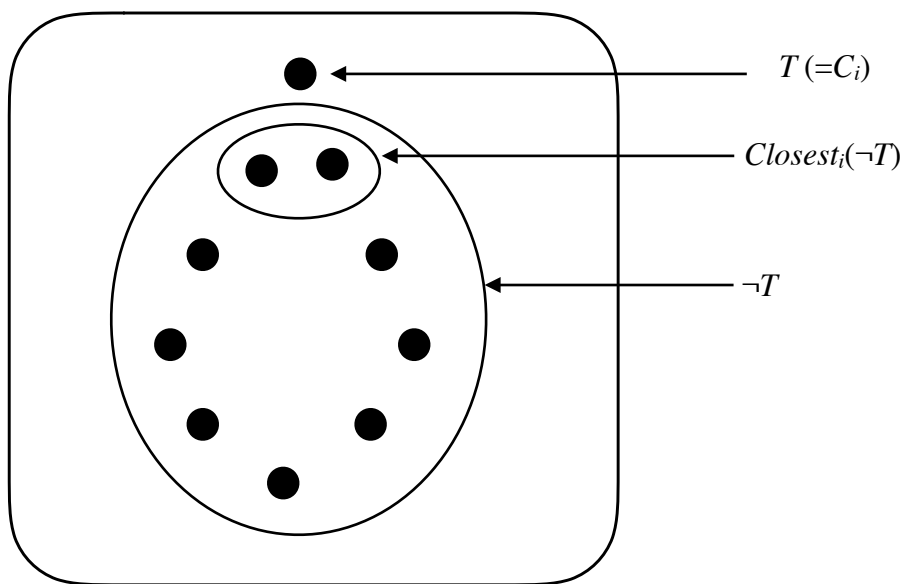


Figure 2 The relative accuracy of $\neg T$ and $Closest_i(\neg T)$

	A	C_i	A -cells	$Closest_i(A)$
1	$N < 9$	N_9	N_0, N_1, \dots, N_8	N_8
2	$N \leq 9$	N_9	N_0, N_1, \dots, N_9	N_9
3	$N > 9$	N_9	$N_{10}, N_{11}, \dots, N_{1000}, \dots$	N_{10}
4	$N \neq 9$	N_9	$N_0, \dots, N_8, N_{10}, \dots, N_{1000}, \dots$	N_8, N_{10}
6	\mathbf{h}	\mathbf{A}_1	$\mathbf{A}_1, \mathbf{A}_2$	\mathbf{A}_1
7	$\neg \mathbf{h}$	\mathbf{A}_1	$\mathbf{A}_3, \mathbf{A}_4$	\mathbf{A}_3
8	$\mathbf{h} \equiv \mathbf{r}$	\mathbf{A}_1	$\mathbf{A}_1, \mathbf{A}_4$	\mathbf{A}_1
9	$\mathbf{h} \equiv \neg \mathbf{r}$	\mathbf{A}_1	$\mathbf{A}_2, \mathbf{A}_3$	$\mathbf{A}_2, \mathbf{A}_3$
10	$\neg(\mathbf{h} \wedge \mathbf{r})$	\mathbf{A}_1	$\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4$	$\mathbf{A}_2, \mathbf{A}_3$

Table 3 Closest A-cells

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	Global inaccuracy D^λ
λ^{equal}	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	
P¹	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1	
P²	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1	
P⁴	0	0	0	1	0	0	1	0	1	1	1	0	1	1	1	
<i>local inacc</i> $d(P^2, P^1)$	1	1	0	0	0	1	1	1	1	0	0	0	1	1	0	
<i>local inacc</i> $d(P^4, P^1)$	1	0	0	1	1	1	0	0	1	1	0	1	0	1	0	
$\lambda^{equal} \times$ $d(P^2, P^1)$	1/15	1/15	0	0	0	1/15	1/15	1/15	1/15	0	0	0	1/15	1/15	0	8/15
$\lambda^{equal} \times$ $d(P^4, P^1)$	1/15	0	0	1/15	1/15	1/15	0	0	1/15	1/15	0	1/15	0	1/15	0	8/15

Table 4 $D^{\lambda^{equal}}$ violates *Opinionation*

	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈	A ₉	A ₁₀	A ₁₁	A ₁₂	A ₁₃	A ₁₄	A ₁₅	Global inaccuracy $D=\sum\lambda \times d$
P^1 (Truth values)	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1	
local inacc $d(P^2, P^1)$	1	1	0	0	0	1	1	1	1	0	0	0	1	1		
local inacc $d(P^4, P^1)$	1	0	0	1	1	1	0	0	1	1	0	1	0	1	0	
λ^{convex}	1/9	1/9	1/9	1/9	1/9	1/9	0	0	1/9	1/9	0	0	0	0	1/9	
$\lambda^{convex} \times$ $d(P^2, P^1)$	1/9	1/9	0	0	0	1/9	0	0	1/9	0	0	0	0	0	0	4/9
$\lambda^{convex} \times$ $d(P^4, P^1)$	1/9	0	0	1/9	1/9	1/9	0	0	1/9	1/9	0	0	0	0	0	6/9
λ^{atomic}	0	0	0	0	1/4	1/4	0	0	1/4	1/4	0	0	0	0	0	
$\lambda^{atomic} \times$ $d(P^2, P^1)$	0	0	0	0	0	1/4	0	0	1/4	0	0	0	0	0	0	2/4
$\lambda^{atomic} \times$ $d(P^4, P^1)$	0	0	0	0	1/4	1/4	0	0	1/4	1/4	0	0	0	0	0	4/4

Table 5 λ^{convex} and λ^{atomic} yield Opinionation

		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}
1	P^1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
2	P^2	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
3	P^8	0	1/2	1/2	0	1/2	1/2	0	1	1/2	1/2	1	1/2	1/2	1	1
4	P^9	0	1/2	0	1/2	1/2	0	1/2	1/2	1	1/2	1/2	1	1/2	1	1
5	P^{14}	0	1/3	1/3	1/3	1/3	1/3	1/3	2/3	2/3	2/3	2/3	2/3	2/3	1	1
6	<i>local inacc</i> $d(P^2, P^1)$	1	1	0	0	0	1	1	1	1	0	0	0	1	1	0
7	<i>local inacc</i> $d(P^8, P^1)$	1	<i>i</i>	<i>i</i>	0	<i>i</i>	<i>i</i>	1	1	<i>i</i>	<i>i</i>	0	<i>i</i>	<i>i</i>	1	0
8	<i>local inacc</i> $d(P^9, P^1)$	1	<i>i</i>	0	<i>i</i>	<i>i</i>	1	<i>i</i>	<i>i</i>	1	<i>i</i>	<i>i</i>	0	<i>i</i>	1	0
9	<i>local inacc</i> $d(P^{14}, P^1)$	1	<i>j</i>	<i>j</i>	<i>j</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>k</i>	<i>j</i>	<i>j</i>	<i>j</i>	1	0

Table 6 Example