Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief

James M. Joyce

Traditional epistemology is both *dogmatic* and *alethic*. It is dogmatic in the sense that it takes the fundamental doxastic attitude to be *full belief*, the state in which a person categorically accepts some proposition as true. It is alethic in the sense that it evaluates such categorical beliefs on the basis of what William James calls the 'two great commandments' of epistemology: Believe the truth! Avoid error! Other central concepts of dogmatic epistemology – knowledge, justification, reliability, sensitivity, and so on – are understood in terms of their relationships to this ultimate standard of truth or accuracy.

Some epistemologists, inspired by Bayesian approaches in decision theory and statistics, have sought to replace the dogmatic model with a probabilistic one in which partial beliefs, or *credences*, play the leading role. A person's credence in a proposition X is her level of confidence in its truth. This corresponds, roughly, to the degree to which she is disposed to presuppose X in her theoretical and practical reasoning. Credences are inherently gradational: the strength of a partial belief in X can range from certainty of truth, through maximal uncertainty (in which X and its negation $\sim X$ are believed equally strongly), to complete certainty of falsehood. These variations in confidence are warranted by differing states of evidence, and they rationalize different choices among options whose outcomes depend on X.

It is a central normative doctrine of probabilistic epistemology that rational credences should obey the laws of probability. In the idealized case where a believer has a numerically precise credence $b(X)$ for every proposition X in some Boolean algebra of propositions,[1] these laws are as follows:

J.M. Joyce (✉)
Department of Philosophy, University of Michigan, Ann Arbor, MI, USA
e-mail: jjoyce@umich.edu

[1] These are the laws of *finitely* additive probability. The results discussed below extend to the countably additive case, and weaker versions of these principles apply to subjects who lack precise credences. Also, this formulation assumes a framework in which there are no distinctions in probability among logically equivalent propositions.

Non-triviality. $b(\sim T) < b(T)$, where T is the logical truth.
Boundedness. $b(X)$ is in the closed interval with endpoints $b(\sim T)$ and $b(T)$.
Additivity. $b(X \vee Y) + b(X \wedge Y) = b(X) + b(Y)$.[2]

Philosophers have offered a number of justifications for this requirement of probabilistic coherence. Some, following Ramsey (1931), de Finetti (1937) and Savage (1972), have advanced *pragmatic* arguments to show that believers with credences that violate the laws of probability are disposed to make self-defeating choices. Others, like Howson and Urbach (1989) and Christensen (1996), argue that incoherence generates inconsistencies in value judgments. Still others, notably van Fraassen (1983) and Shimony (1988), seek to tie probabilistic coherence to rules governing the estimation of relative frequencies. Finally, Joyce (1998) hoped to clarify the normative status coherence, and to establish an alethic foundation for probabilistic epistemology, by showing coherence is conducive to accuracy. The central claims of that article were as follows:

1. Partial beliefs should be evaluated on the basis of a *gradational* conception of accuracy, according to which the accuracy of a belief in a true/false proposition is an increasing/decreasing function of the belief's strength.
2. One can identify a small set of constraints that any reasonable measure of gradational accuracy should satisfy.
3. Relative to any measure of gradational accuracy that satisfies the constraints, it can be show that: (3a) each incoherent system of credences is *strictly inadmissible* in the sense that there is a coherent system that is strictly more accurate in every possible world; and (3b) coherent credences are always *admissible*.
4. Inadmissibility relative to all reasonable measure of gradational accuracy renders incoherent credences defective from a purely epistemic perspective.

This essay will clarify and reevaluate these claims. As it happens, the constraints on accuracy measures imposed in Joyce (1998) are not all well justified. They are also much stronger than needed to obtain the desired result. Moreover, neither (3b) nor (4) where adequately defended. Finally, the focus on accuracy measures is unduly restrictive: a broader focus on epistemic utility (which has accuracy as a central component) would make the results more general. These deficiencies will

---

[2] While this formulation may seem unfamiliar, one can secure the standard laws of probability via the convention that logical truths have probability 1 and that contradictions have probability 0. I have chosen this formulation to emphasize that $b(\sim T) = 0$ and $b(T) = 1$ are mere conventions. Non-triviality also has a conventional element. Conjoining Additivity with the stipulation $b(\sim T) > b(T)$ produces an *anti-probability*, a function whose *complement* $b^-(X) = 1 - b(X)$ is a probability. The difference between representing degrees of belief using probabilities or anti-probabilities is entirely a matter of taste: the two ways of speaking are entirely equivalent. It is crucial, however, to recognize that when b measures strengths of beliefs, adopting the conventional $b(\sim T) > b(T)$ requires us to interpret larger b-values as signaling *less* confidence in the *truth* of a proposition and *more* in confidence in its *falsehood*. If one mistakenly tries to retaining the idea that b measures confidence in truths while setting $b(\sim T) > b(T)$, one ends ups with nonsense. This mistake seems to be the basis of the worries raised in Howson (2008, pp. 20–21).

be corrected here, and the prospects for 'non-pragmatic vindication of probabilism' along the lines of 1–4 will be reassessed.

1 Formal Framework

We image an (idealized) believer with sharp degrees of confidence in propositions contained in an ordered set $X = \langle X_1, X_2, \ldots, X_N \rangle$. For simplicity we will assume that X is finite, and that its elements form a *partition*, so that, as a matter of logic, exactly one X_n is true. Our subject's degrees of belief can then be represented by a *credence function* b that assigns a real number $b(X)$ between zero and one (inclusive) to each $X \in X$. We can think of b as a vector[3] $\langle b_1, b_2, \ldots, b_N \rangle$ where each $b_n = b(X_n)$ measures the subject's confidence in the truth of X_n on a scale where 1 and 0 correspond, respectively, to certainty of truth and certainty of falsehood.

The N-dimensional cube $\mathcal{B}_X = [0, 1]^N$ then contains all credence functions defined on X. Its proper subsets include both (a) the family $\mathcal{P}_X$ of all probability functions defined on X, which in turn properly includes (b) the collection $\mathcal{V}_X$ of all consistent truth-value assignments to elements of X. If we let 0 signify falsity and 1 denote truth, we can identify $\mathcal{V}_X$ with the set $\langle v^1, v^2, \ldots, v^N \rangle$ where v^n is the binary sequence that has a 1 in the nth place and 0 elsewhere. One can think of v^n as the 'possible world' in which X_n is true, and all other X_j are false.

It is easy to show that a credence function obeys the laws of probability if and only if it is a weighted average of truth-value assignments, so that $b \in \mathcal{P}_X$ exactly if $b = \Sigma_j \lambda_j \cdot v^j$, where the λ_j are non-negative real numbers summing to 1. For example, if $X = \langle X_1, X_2, X_3 \rangle$, then $\mathcal{V}_X$ contains three points $v^1 = \langle 1, 0, 0 \rangle$, $v^2 = \langle 0, 1, 0 \rangle$, $v^3 = \langle 0, 0, 1 \rangle$ in $\Re^3$. $\mathcal{P}_X$ is the triangle with these points as vertices (where $1 = b_1 + b_2 + b_3$), and the regions above ($1 < b_1 + b_2 + b_3$) and below ($1 > b_1 + b_2 + b_3$) this triangle contain credence assignments that violate the laws of probability.

2 Epistemic Utility and Scoring Rules

Part of our goal is to isolate features of credence assignments that are advantageous from a purely epistemic perspective. This task requires substantial philosophical reflection, and little more than a broad cataloging of epistemically desirable features will be attempted here. Readers will be left to decide for themselves which of the properties discussed below conform to their intuitions about what makes a system of beliefs better or worse from the purely epistemic perspective.

---

[3] Notation: (a) Vector quantities are in **bold;** their arguments are not. (b) indices ranging over integers (i, j, k, m, n) are always lower case, and their maximum value is the associated upper case letter, so that, e.g., n ranges over 1, 2, ..., N. (c) A vector $x = \langle x_1, x_2, \ldots, x_N \rangle$ will often be denoted $\langle x_n \rangle$.

To make headway, let us adopt the useful fiction that the notion of overall epistemic goodness or badness for partial beliefs can be made sufficiently precise and determinate to admit of quantification. In particular, let's assume that for each partition of propositions X there is a *scoring rule* S_X that, for each b in $\mathcal{B}_X$ and v in $\mathcal{V}_X$, assigns a real number $S_X(b, v) \geq 0$ which measures the *epistemic disutility*[4] of holding the credences b when the truth-values for elements of X are as given in v. Intuitively, $S_X(b, v)$ measures the extent to which b's credences diverge from some epistemic ideal at v. A perfect score is obtained when $S_X(b, v) = 0$, and $S_X(b, v) > S_X(b^*, v)$ means that, in terms of overall epistemic quality, the credences in b^* are better than those in b when v is actual. Thus, the choice of an epistemic scoring rule should reflect our considered views about what sorts of traits make beliefs worth holding from the purely epistemic perspective.

The term 'scoring rule' comes from economics, where values of S are seen as imposing penalties for making inaccurate probabilistic predictions. If, say, a meteorologist is paid to predict rain, her employer might seek to promote accuracy by docking her pay $\$S(b, v)$ where b_n is the predicted chance of rain on the nth day of the year and v_n is 1 or 0 depending upon whether or not it rains that day. When scoring rules are so construed, it is vital to know whether they create incentives that encourage subjects to make honest and accurate predictions. The focus here will be different. Rather, than thinking of a subject as being motivated to minimize her penalty, as economists do, we will use scoring rules to gauge those qualities of credences that have epistemological merit. So, instead of interpreting epistemic scoring rules as setting penalties that believers might suffer when they reveal their beliefs, we view them as tools of evaluation that third parties can use to assess the overall epistemic quality of opinions. The fact that one set of credences incurs a lower penalty than another at a given world should be taken to mean that, from a purely epistemic point of view, it would be better in that world to hold the first set of credences than to hold the second. It is, of course, quite consistent with this that the agent has an incentive structure that encourages her to hold beliefs that diverge greatly from the epistemic ideal.[5]

---

[4] I use epistemic *dis*utility rather than epistemic utility so as to more easily relate this investigation to the work on proper scoring rules in statistics and economics.

[5] It must be emphasized that believers do not *choose* credences with the goal of minimizing their score, as a weather forecaster would if her pay was at risk. Indeed, it is not clear that believers should ever be thought of as 'choosing' their credences. Believing is not an action. Likewise, in contrast with the scoring rule tradition, epistemic disutilities attach to credences directly, as opposed to public reports of credences (or 'previsions'). So, when we say that a person's credences score poorly relative to some scoring rule, we criticize his or her beliefs directly, but we do not thereby suggest that there is any past action that the person should have performed differently or even that there is any future action that they should perform.

3 The Principle of Admissibility

Before beginning substantive discussions of epistemic scoring rules, it is important to understand one general commitment that is involved in endorsing any such a rule. Since a scoring rule is meant to provide an overall evaluation of credence functions, when we endorse a rule as the correct measure of epistemic disutility we commit ourselves to thinking that there is something defective, from a purely epistemic perspective, about credences that score poorly according to that rule. Moreover, if these poor scores arise as a matter of necessity, then the defect is one of epistemic irrationality. To make this precise, say that one credence function b in $\mathcal{B}_X$ is *dominated* by another b^* with respect to S when $S(b, v) > S(b^*, v)$ for every truth-value assignment v. Credences that are dominated this way are *epistemically inadmissible* according to S. Endorsing S as the correct measure of epistemic utility involves, minimally, committing oneself to the view that S-inadmissible credences are epistemically irrational.

> ADMISSIBILITY. A system of credences $b \in \mathcal{B}_X$ is epistemically irrational if there is another system $b^* \in \mathcal{B}_X$ such that $S(b, v) \geq S(b^*, v)$ for every $v \in \mathcal{V}_X$ and $S(b, v) > S(b^*, v)$ for some $v \in \mathcal{V}_X$. Epistemically rational credences are never weakly dominated in this way.

This is not a substantive claim about epistemic rationality. It is, rather, a constraint on the choice of scoring rules. If one doubts ADMISSIBILITY for a given rule, the problem resides not the principle, but in the fact that the rule fails to capture one's sense of what is valuable about beliefs from a purely epistemic perspective.

4 Estimation and Accuracy

The interest of any epistemic scoring rule depends on the virtues it captures. While systems of beliefs can possess a range of laudable qualities – they might be informative, explanatory, justified, reliably produced, safe, useful for making decisions, and so on – epistemological evaluation is always concerned with the relationship between belief and truth or, as we shall say for credences, epistemic accuracy. Indeed, many of the qualities just mentioned are desirable largely in virtue of their connection to accuracy. Accuracy is the one epistemic value about which there can be no serious dispute: it *must* be reflected in any plausible epistemic scoring rule. This does not mean that accuracy is all there is to epistemic utility. Perhaps other qualities are involved, but accuracy remains an overriding epistemic value. All else equal, if one system of credences is more accurate than another, then, from a purely epistemic perspective, the first system is better than the second. Accuracy is an unalloyed epistemic good.

But, what does it mean to say that credences are accurate, and how is their accuracy assessed? As a step toward answering, we can exploit the fact that a person's

credences determine her best *estimates* of the sorts of quantities that, from a purely epistemic perspective, it is important to be right about. The accuracy of a system of credences can then be assessed by looking at how closely its estimates are to the actual values of these quantities. Here, a 'quantity' is any function that assigns real numbers to possible worlds in $\mathcal{V}_X$ (what statisticians call a random variable). Here are two natural 'epistemologically significant' quantities:[6]

- *Truth-values*: Quantities are *propositions*, thought of as indicator functions that map $\mathcal{V}_X$ into $\{0, 1\}$.[7] For each proposition Y, $Y(\nu) = 1$ means that Y is true at ν and $Y(\nu) = 0$ means that Y is false at ν.
- *Relative frequencies*: Each quantity is associated with a set of propositions $Y = \{Y_1, Y_2, \ldots, Y_K\}$. Every world ν is mapped to the proportion of Y's elements that are true in ν, so that $Freq_Y(\nu) = [Y_1(\nu) + \ldots + Y_K(\nu)]/K$.

One's choices about which quantities to focus on will be tied up with one's view of what credences are. Some accounts construe them as truth-value estimates (Jeffrey 1986, Joyce 1998), while others tie them to estimates of relative frequency (Shimony 1988, van Fraassen 1983).

Once appropriate epistemic quantities have been selected, the next step is to explain how credences fix estimates. Estimation is straightforward when b obeys the laws of probability: the correct estimate for a quantity F is its expected value computed relative to b, so that $Est_b(F) = \Sigma_n b(\nu^n) \cdot F(\nu^n)$. It is then easy to see that the estimated truth-value for any proposition is its credence $Est_b(Y) = b(Y)$. And, since expectation is additive, $Est_b(F + G) = Est_b(F) + Est_b(G)$, it follows that for any set of propositions $\{Y_k\}$ one has both

Estimated Truth-value Additivity. $Est_b(\Sigma_k Y_k) = \Sigma_k b(Y_k)$.
Estimated Frequency Additivity. $Est_b(Freq(Y)) = \Sigma_k b(Y_k)/K$.

Coherent credences can thus be summed to estimate either the number of truths or the relative frequency of truths in a set of propositions. These are universal facts: as long as b is coherent, these identities hold for any set of propositions.

When a credence function b violates the laws of probability the additivity equations fail, and it becomes less clear what estimates b sanctions. Fortunately, in the special case of truth-value estimation there is a principle that does apply to all credences, solely in virtue of what they are, whether or not they obey the laws of probability. In the probabilistic tradition, *the* defining fact about credences is that

---

[6] Some might include objective chances in this list. For current purposes, however, it is not useful to focus on the connection between credences and chances. Credences can often be portrayed as estimates of objective chances, but, unlike the cases of truth-values and frequencies, the relationship is not uniform. There are situations, namely those in which a believer has 'inadmissible' information in the sense of Lewis (1980), in which degrees of belief and estimates of objective chance diverge.

[7] The choice of 1 to represent truth and of 0 to represent falsity is pure convention; any choice with $T(\nu) > \bot(\nu)$ will do. If one sets $T(\nu) < \bot(\nu)$, the ideas developed here lead to an 'anti-probability' representation. See footnote 2.

they are used to estimate quantities that depend on truth-values. These quantities might be, e.g., values of bets, relative frequencies, outcomes of experiments, or something else. But, whatever they are, one estimates these values by estimating truth-values. In light of this, a person's credence for Y will function as a kind of 'summary statistic' that encapsulates those features of her evidential situation that are relevant to estimates of Y's truth-value. As a result, we have

> *The Alethic Principle*. A rational believer's best estimate for the truth-value of any proposition will coincide with her credence for it: $Est_b(Y) = \boldsymbol{b}(Y)$ for all credence functions $\boldsymbol{b}$ (coherent or not) and all propositions Y.

In light of this, the accuracy of any credence should be evaluated by considering the accuracy of the truth-value estimate that it sanctions.

Since accuracy in estimation involves getting as close as possible to the true value of the estimated quantities, estimates are always appropriately evaluated on a 'closeness counts' scale of accuracy.

> GRADATIONAL ACCURACY. At world $\boldsymbol{v}$, if the estimates $Est(\boldsymbol{F}_j)$ are uniformly closer to the values of quantities $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_J$ than are the estimates $Est^*(\boldsymbol{F}_j)$, so that $Est^*(\boldsymbol{F}_j) \geq Est(\boldsymbol{F}_j) \geq \boldsymbol{F}_j(\boldsymbol{v})$ or $Est^*(\boldsymbol{F}_j) \leq Est(\boldsymbol{F}_j) \leq \boldsymbol{F}_j(\boldsymbol{v})$ for all j, and $Est^*(\boldsymbol{F}_j) > EST(\boldsymbol{F}_j) \geq \boldsymbol{F}_j(\boldsymbol{v})$ or $Est^*(\boldsymbol{F}_j) < Est(\boldsymbol{F}_j) \leq \boldsymbol{F}_j(\boldsymbol{v})$ for some j, then the first set of estimates is more accurate than the second at $\boldsymbol{v}$.

When we apply this to truth-value estimation,[8] and combine it with the thought that making accurate truth-value estimates is an overriding epistemic value, the result is

> TRUTH-DIRECTEDNESS. For credence functions $\boldsymbol{b}$ and $\boldsymbol{b}^*$, if $\boldsymbol{b}$'s truth-value estimates are uniformly closer to the truth than $\boldsymbol{b}^*$'s at world $\boldsymbol{v}$, so that either $b_n^* \geq b_n \geq v_n$ or $b_n^* \leq b_n \leq v_n$ for all n and $b_n^* > b_n \geq v_n$ or $b_n^* < b_n \leq v_n$ for some n, then $S(\boldsymbol{b}^*, \boldsymbol{v}) > S(\boldsymbol{b}, \boldsymbol{v})$.

In other words, moving credences uniformly closer to the actual truth-values of propositions always increases epistemic utility. Rules that fail this test let people improve the epistemic quality of their opinions by becoming less certain of truths

---

[8] It is important not to confuse estimation with *guessing* (see Jeffrey 1986). A guess at a truth-value is evaluated solely on the basis of whether or not it gets the truth-value exactly right; nothing is gained by missing by a small rather than a large amount. In estimation, on the other hand, the goal is get as close as one can to the actual value of the estimated quantity. This is why 1/4 is a fine estimate of the truth-value of the proposition that the top card of a well-shuffled deck is a spade, whereas the only guesses that make sense are 1 and 0. (Joyce 1998) argues that full-beliefs are best evaluated as guesses, whereas credences are best evaluated as estimates.

and more certain of falsehoods. Such rules do not make accuracy a cardinal virtue, and so are not instruments of pure epistemic assessment.

The combination of TRUTH-DIRECTEDNESS and ADMISSIBILITY entails that it is epistemically irrational to have credences that are further from the truth than some other set of credences at every possible world. This suffices to show that rational credences must obey the first two laws of probability. For the first law, if $b(T) \neq 1$ then b is always uniformly further from the truth than the credence function b^* defined by $b^*(T) = [1+b(T)]/2$ and $b^*(T) = b(\sim T)$ (and similarly for $b(\sim T) \neq 0$). For the second law, if $b(Y) > 1$, then b is always uniformly further from the truth than the credence function b^* defined by $b^*(Y) = [1 + b(Y)]/2$ and $b^*(X) = b(X)$ for $X \neq Y$ (and similarly for $b(Y) < 0$).

A more substantive argument is required to establish that credences should satisfy the third law of probability. One strategy is to augment ADMISSIBILITY with further constraints on rational estimation for quantities other than truth-values, and to show that these constraints force estimation to be additive. Two such requirements are:

> *Dominance*. If $F(v) \geq G(v)$ for all $v \in \mathcal{V}_X$, then $Est(F) \geq Est(G)$.
> *Independence*. If $F(v) = G(v)$ for all v in some subset $\mathcal{W}$ of $\mathcal{V}_X$, then $Est(F) \geq Est(G)$ iff $Est(F^*) \geq Est(G^*)$ for all other quantities F^* and G^* that (i) agree with one another on $\mathcal{W}$, and (ii) agree, respectively, with F and G on $\mathcal{V}_X \sim \mathcal{W}$.

This general approach is reflected in most justifications of coherence, including Dutch-book arguments and representation theorems.

Alternatively, one can simply require that estimates be additive. Jeffrey, for example, has called the additivity of estimations, 'as obvious as the laws of logic' (1986, p. 52). This is unlikely to move anyone with doubts about the normative status of coherence, however, since the additivity of truth-value estimates is straightforwardly equivalent to the additivity of credences.

A slightly less objectionable approach would be to introduce the following principle, which does not so obviously presuppose that credences are additive.

> *Calibration*. For all credence functions b (coherent or not) and all sets of propositions Y, if $b(Y) = b$ for all $Y \in Y$, then $Est_b(Freq(Y)) = b$.

This is intermediate between ALETHIC and the idea that estimated frequencies are just summed credences. It reflects an intuition that is surely central to degrees of belief: what can it mean to assign credence $b(Y)$ to Y unless one is committed to thinking that propositions with Y's overall epistemic profile are true roughly $b(Y)$ proportion of the time? Despite this, Calibration is still too similar to additivity to serve as a premise in the latter's justification. Calibration requires every uniform distribution of credences over a partition $\{Y_k\}$ to be coherent, so that $b(Y_k) = 1/K$ for all k. This is a strong requirement, and opponents of probabilistic coherence will surely want to know what there is, other than a prior commitment to additivity, that prevents a rational person from believing both Y and $\sim Y$ to degree 0.45

and yet acknowledging the logical fact that exactly half the elements of $\{Y, \sim Y\}$ are true.

Rather than pursuing these strategies, we will presuppose only ALETHIC, TRUTH-DIRECTEDNESS, ADMISSIBILITY,[9] and will investigate the prospects for deriving the requirement of coherence directly from various further constraints on epistemic scoring rules. Our objective will be to determine the extent to which probabilistic coherence can be established as an epistemic virtue, and incoherence as an epistemic vice, by showing that incoherent credences are inadmissible on all reasonable ways of assessing epistemic utility. After spending the next few sections discussing various properties that epistemic scoring rules might possess, we will consider the significance of some theorems that purport to prove this sort of result. We will conclude by relating these results to some facts about the estimation of relative frequencies.

5 Atomism or Holism?

Let's begin by considering whether to endorse an *atomistic* or a *holistic* conception of epistemic utility. On an atomistic picture, the epistemic utility of each b_n can be ascertained independently of the values of other credences in b. On a holistic conception, individual credences have epistemic utilities only within the confines of a system – one can speak of the epistemic value of the function b but not of the various b_n – and overall utility is not simply a matter of amalgamating utilities of individual components. To illustrate, let $X = \langle X_1, X_2, X_3 \rangle$ and $v = \langle 1, 0, 0 \rangle$. Truth-directedness ensures that accuracy, and hence epistemic utility, improves when $b(X_1)$ is moved closer to 1 and $b(X_2)$ is moved closer to 0, and this is true whatever value $b(X_3)$ happens to have. If, however, $b(X_1)$ and $b(X_2)$ both shift toward 1, so that b becomes more accurate in its fist coordinate but less accurate in its second coordinate, then it is consistent with TRUTH-DIRECTEDNESS that epistemic utility increases for some values of $b(X_3)$ but decreases for others. On an atomistic conception this should not happen: the effect of changes in $b(X_1)$ and $b(X_2)$ on overall epistemic value should not depend on what credences appear elsewhere in b. To put it differently, one should be able to ignore cases where b and c agree when assessing the relative change in epistemic utility occasioned by a shift in credences from b to c.

One can enforce atomism by requiring epistemic scoring rules to obey an analogue of the decision-theoretic 'sure thing principle'.

> SEPARABILITY. Suppose Y is a subset of X, and let b, b^*, c, c^* be credence functions in $\mathcal{B}_X$ such that

---

[9] We will also assume that the ordering of the propositions in X is immaterial to the value of the scoring rule.

$$b(X) = b^*(X) \text{ and } c(X) = c^*(X) \text{ for all } X \in Y$$
$$b(X) = c(X) \text{ and } b^*(X) = c^*(X) \text{ for all } X \in X \sim Y$$

Then $S(b, v) \geq S(c, v)$ if and only if $S(b^*, v) \geq S(c^*, v)$.

This ensures that the overall epistemic utility of a person's credences over Y can be assessed in a way that is independent of what credences are assigned outside Y.

Many scoring rules are separable. Consider, for example, functions of the additive form $S(b, v) = \Sigma_n \lambda_X(X_n) \cdot s_n(b_n, v_n),$[10] where each *component function* s_n measures the epistemic utility of b_n on a scale that decreases/increases in its first coordinate when the second coordinate is one/zero, and where the *weights* $\lambda_X(X_n)$ are non-negative real numbers summing to one that reflect the degree to which the utilities of credences for X_n matter to overall epistemic utility. Such a function is separable as long as each s_n depends only on X_n, b_n and v_n (and not on X_k, b_k or v_k for $k \neq n$), and each $\lambda_X(X_n)$ depends only on X and X_n (and not on b or v).[11]

Those inclined toward a holist conception will deny SEPARABILITY. The issue, at bottom, is whether one thinks of estimation as a 'local' process in which one's estimate of X's truth-value reflects one's thinking about X taken in isolation, or whether it is a more 'global' process that forces one to take a stand on truth-values of propositions that do not entail either X or its negation. We shall leave it to readers to adjudicate these issues on their own.

6 Content Independence

Another set of issues concerns the extent to which the standard of epistemic value for credences should be allowed to vary in response to features of propositions other than truth-values. It is consistent with everything said so far that being right or wrong about one truth has a greater effect on overall epistemic utility, *ceteris paribus*, than being right or wrong about another. It might be, for example, that a shift in credence toward a proposition's truth-value matters more or less depending upon whether the proposition is more or less informative. Or, in addition to being awarded points for having credences that are near truth-values, subjects might get credit for having credences that are close to objective chances, so that a credence of 0.8 for a truth is deemed better if the proposition's objective chance is 0.7 than if it is 0.2. Or, perhaps assigning high credences to falsehoods is less of a detriment to epistemic utility when the falsehood has high 'verisimilitude'. Or, maybe assigning a high

---

[10] Many authors assume that epistemic utilities have this additive form. We shall not be making this assumption here, though many of the examples we discuss will be additive.

[11] The weights may depend on X. If X_n is an element of another set Y it can happen that $\lambda_X(X_n) < \lambda_Y(X_n)$, in which case the credence for X_n matters less to overall epistemic value in the context of X than in the context of Y.

credence to a truth is worth more or less depending on the practical costs of falsely believing it.

Those who take an austere view will maintain that nothing but credences and truth-values matter to assessments of epistemic utility. They will insist on:

> EXTENSIONALITY. Let b and b^* be credence functions defined, respectively, over $X = \langle X_1, \ldots, X_N \rangle$ and $X^* = \langle X_1^*, \ldots, X_N^* \rangle$, and let v and v^* be associated truth-value assignments. If $b_n = b_n^*$ and $v_n = v_n^*$ for all n, then $S_X(b, v) = S_{X^*}(b^*, v^*)$.

This requires the same basic yardstick to be used in evaluating all credences. It ensures that b's overall epistemic utility at v is a function only of b's credences and v's truth-values. Additional facts about propositions – their levels of justification, informativeness, chances, verisimilitude, practical importance and so on – have no influence on S except insofar as they affect credences and truth-values.

EXTENSIONALITY is a strong requirement. It entails that $S_X(v, v) = S_X(v^*, v^*) < S_X(1 - v, v) = S_X(v^*, 1 - v^*)$ for all v and v^*. This makes it possible to fix a single scale of epistemic utility using $S_X(v, v)$ as the zero and $S_X(1 - v, v)$ as the unit.

For another illustration of EXTENSIONALITY's effects, consider its impact on additive rules. In theory, the component functions in an additive rule might vary proposition by proposition, so that $s_n(b_n, v_n)$ and $s_k(b_k, v_k)$ differ even when $b_n = b_k$ and $v_n = v_k$. In addition, utility with respect to X_n might still have more impact on overall epistemic utility than utility with respect X_k because $\lambda_X(X_n) > \lambda_X(X_k)$. EXTENSIONALITY eliminates this variability: it requires that $s_n(b, v) = s_k(b, v)$ for all b and v, and $\lambda_X(X_n) = 1/N$ for all n. All additive rules then assume the simple form $S(b, v) = 1/N \cdot \Sigma_n s(b_n, v_n)$ where s represents a single standard for evaluating a credence given a truth-value.

As further indication of EXTENSIONALITY's potency, notice that applying it to $X = \langle X, \sim X \rangle$ and $X^* = \langle \sim X, X \rangle$ enforces a symmetry between propositions and their negations since $S(\langle b, c \rangle, \langle v, 1 - v \rangle) = S(\langle c, b \rangle, \langle 1 - v, v \rangle)$ for $b, c \in [0, 1]$ and $v \in \{0, 1\}$. So, assigning credences b to X and c to $\sim X$ when X has a given truth-value is the same, insofar as extensional epistemic disutility is concerned, as assigning credences c to X and b to $\sim X$ when X has the opposite truth-value. When $c = 1 - b$, this becomes $S(\langle b, 1 - b \rangle, \langle v, 1 - v \rangle) = S(\langle 1 - b, b \rangle, \langle 1 - v, v \rangle)$.

This last equation is independently plausible, and some might want to extend it to incoherent credences by requiring that $S(\langle b, c \rangle, \langle v, 1 - v \rangle) = S(\langle 1 - b, 1 - c \rangle, \langle 1 - v, v \rangle)$ for all b and c. Failures of this identity do have an odd feel. Suppose Jack sets credences of 0.8 and 0.4 for X and $\sim X$, while Mack, Jack's counterpart in another possible world, sets his at 0.2 and 0.6. Imagine that X is true in Jack's world, but false in Mack's world. It seems unfair for Jack's epistemic utility to exceed Mack's since each has a credence for X that is 0.2 units from its truth-value, and a credence for $\sim X$ that is 0.4 units from its truth-value. The reversal of these truth-values in the different worlds does not seem relevant to

assessments of Jack's or Mack's credences. This sort of reasoning suggests the requirement of

> 0/1-SYMMETRY. Let v^i and v^j be truth-value assignments for the partition $\langle X_N \rangle$ with $v^i(X_i) = v^j(X_j) = 1$. Let b and b^* be credence functions for X that are identical except that $b(X_i) = 1 - b^*(X_i)$ and $b(X_j) = 1 - b^*(X_j)$. Then, $S_X(b, v^i) = S_X(b^*, v^j)$.

The combination of 0/1-SYMMETRY and EXTENSIONALITY entails a condition that was endorsed in (Joyce 1998) as a requirement of epistemic accuracy:

> NORMALITY. Let b and b^* be defined, respectively, over $X = \langle X_1, \ldots, X_N \rangle$ and $X^* = \langle X_1^*, \ldots, X_N^* \rangle$, and let v and v^* be associated truth-value assignments. If $|b_n - v_n| = |b_n^* - v_n^*|$ for all n, then $S_X(b, v) = S_X^*(b^*, v^*)$.

This makes epistemic utility depend entirely on absolute distances from credences to truth-values, so that $S(b, v) = F(|b_1 - v_1|, \ldots, |b_N - v_N|)$ where $F(x_1, \ldots, x_n)$ is a continuous, real function that decreases monotonically in each argument. Any additive rule will then take the form $S(b, v) = \Sigma_n \lambda_n \cdot f(|b_n - v_n|)$, where $f: \Re \to \Re$ is monotonically decreasing.[12]

The appropriateness of the preceding conditions as epistemic norms is up for debate. Detractors will contend that judgments of overall epistemic quality depend on a believed proposition's informativeness, its objective chance, or on the costs of being mistaken about it. Being confident to a high degree in a specific truth is a more significant cognitive achievement than being equally confident in some less specific truth. Having a credence far from a proposition's objective chance seems like a defect even if that credence is close to the proposition's truth-value. Being highly confident in a true proposition whose truth-value matters a great deal seems 'more right' than being confident to the same degree in a true proposition whose truth-value is a matter of indifference. For all these reasons, some will argue, we need a notion of epistemic accuracy that is more nuanced than EXTENSIONALITY or NORMALITY allow.

Those on the other side will emphasize the central role of considerations of accuracy in assessments of epistemic utility. EXTENSIONALITY and NORMALITY, they will argue, are plausible when S measures accuracy. So, to the extent that considerations of pure accuracy are dominant in evaluations of epistemic value, these principles will seem reasonable. Those who think otherwise, it will be suggested, are conflating issues about what makes a belief worth holding with questions about how hard it is to arrive at a justified and accurate belief about some topic, or how important it is, for practical reasons, to hold such a belief. For example, the more informative a proposition is, *ceteris paribus*, the more evidence it takes to justify a belief in its truth. While this added 'degree of difficulty' might be relevant to

---

[12] Note that NORMALITY is weaker than the combination of 0/1-SYMMETRY and EXTENSIONALITY because it allows for the possibility that S's various component functions have different weights.

evaluations of the belief's justification, it is not germane to its accuracy or, on this view, its epistemic utility. Likewise, even though more or less might hang on being right about a proposition, this does not affect the underlying accuracy of beliefs involving that proposition. So, to the extent that we see epistemic utility as reflecting considerations of accuracy alone we will be inclined toward EXTENSIONALITY and NORMALITY.

Once again, we leave it to readers to adjudicate these issues.

7 Some Examples

It might be useful to consider some examples of additive scores that satisfy the conditions listed so far. One interesting class of examples is provided by the

Exponential Scores: $\alpha^z(\boldsymbol{b}, \boldsymbol{v}) = (1/N)\Sigma_n|v_n - b_n|^z$, for $z > 0$.

The best known of these is the *Brier score*, $\boldsymbol{Brier}(\boldsymbol{b}, \boldsymbol{v}) = \alpha^2(\boldsymbol{b}, \boldsymbol{v})$, which identifies the inaccuracy of each truth-value estimate with the squared Euclidean distance between it and the actual truth-value. Since Brier (1950), meteorologists have used this score to gauge the accuracy of probabilistic weather forecasts. Another popular exponential score is the *absolute value measure* $\alpha^1(\boldsymbol{b}, \boldsymbol{v})$, which measures epistemic utility as the linear distance between each credence and its associated truth value. α^1 has been defended by Patrick Maher (2002) and Paul Horwich (1982), among others, as the correct yardstick for measuring epistemic value.

Here are some other rules in common use:

Power Scores: $\pi^z(\boldsymbol{b}, \boldsymbol{v}) = 1/N\Sigma_n(z-1) \cdot b^z + v_n \cdot (1 - z \cdot b^{z-1})$, for $z > 1$.[13]
Logarithmic score: $\chi(\boldsymbol{b}, \boldsymbol{v}) = 1/N\Sigma_n - ln(|(1 - v_n) - b_n|)$
Spherical score: $\beta(\boldsymbol{b}, \boldsymbol{v}) = 1/N\Sigma_n 1 - [|(1 - v_n) - b_n|/(b_n^2 + (1 - b_n)^2)^{1/2}]$.

All these scores are 0/1-symmetric and the last two are Normal.

It is also possible to design scores that treat truth-values differentially, so that a credence of b for X is assessed using one standard when X is true and with another when X is false. A useful example here is the hybrid function that sets the penalty for having credence b at its distance from the truth $\boldsymbol{S}(b, 1) = 1 - b$ when X is true and at $\boldsymbol{S}(b, 0) = (1 - b) - ln(1 - b)$ when X is false.

These scores can differ along a number of dimensions. The most important differences, for current purposes, have to do with (a) the question of whether or not minimizing *expected* scores encourages coherence, (b) their convexity or concavity properties, and (c) the degree to which they permit coherence. We will begin by considering (a).

---

[13] Note the Brier score is just the $z = 2$ power score.

8 Strictly Proper Measures

Economists call a scoring rule *strictly proper* when it gives a coherent agent an incentive to announce her actual credence as her estimate of X's probability. Likewise, an epistemic utility is strictly proper when each *coherent* credence function uniquely minimizes expected value, relative to its own probability assignments.

> PROPRIETY. $Exp_b(S(b)) = \Sigma_n b(v^n) \cdot S(b, v^n) < \Sigma_n b(v^n) \cdot S(c, v^n) = Exp_b(S(c))$ for every $b \in \mathcal{P}_X$ and $c \in \mathcal{B}_X$.

It is easy to show that an additive rule is strictly proper if and only if its component functions are strictly proper in the sense that

$$b \cdot s_n(b, 1) + (1 - b) \cdot s_n(b, 0) < b \cdot s_n(c, 1) + (1 - b) \cdot s_n(c, 0)$$

for all $b, c \in [0, 1]$. When the rule satisfies EXTENSIONALITY, this becomes a constraint on a single pair of functions $s(b, 1)$ and $s(b, 0)$. When the rule is also 0/1-symmetric it becomes a constraint on the single function $s(b, 1)$.

Readers are invited to verify the following facts:

- The Brier score, α^2, is the only strictly proper exponential score.
- Every power rule π^z is strictly proper.
- The logarithmic score χ is strictly proper.
- The spherical score β is strictly proper.

Propriety places strong restrictions on accuracy measures. Indeed, as Schervish (1989) shows, a necessary and sufficient condition for PROPRIETY in additive, extensional rules is the existence of a strictly increasing, positive function h such that

> ***Schervish.*** $h'(b) = -s'(b, 1)/(1 - b) = s'(b, 0)/b.$[14]

An equivalent characterization was given earlier in Savage (1972), who showed that $s(b, 1)$ and $s(b, 0)$ define a strictly proper additive scoring rule if and only if there is some twice differentiable positive function g on [0, 1] with $g'' < 0$ on (0, 1) such that

> ***Savage.*** $s(b, v) = g(b) + (v - b) \cdot g'(b)$

One can see that the two characterizations are equivalent by setting $h(b) = -g'(b)$.

Here are the Schervish and Savage functions for the Brier, spherical and logarithmic scores.

---

[14] If we relax EXTENSIONALITY, these relationships hold for each s_n.

Score	Brier	Spherical β, $\boldsymbol{D}(b) = (b^2 + (1-b)^2)^{1/2}$	Logarithmic χ
$\boldsymbol{s}(b,1)$	$(1-b)^2$	$1 - b/\boldsymbol{D}(b)$	$-ln(b)$
$\boldsymbol{s}(b,0)$	b^2	$1 - (1-b)/\boldsymbol{D}(b)$	$-ln(1-b)$
$\boldsymbol{g}(b)$	$b \cdot (1-b)$	$1 - \boldsymbol{D}(b)$	$(1-b) \cdot ln(1-b) + b \cdot ln(1-b)$
$\boldsymbol{h'}(b)$	2	$\boldsymbol{D}(b)^{-3}$	$1/[b \cdot (1-b)]$

There are many strictly proper scores. As Gibbard (2008) emphasizes, the Schervish equations provide a recipe for constructing them: if $\boldsymbol{s}(b, 1)$ is *any* strictly decreasing differentiable function of b, then setting $\boldsymbol{s}(b, 0) = \int_0^b x \cdot h'(x)\, dx$ yields a strictly proper rule. For example, if $\boldsymbol{s}(b, 1) = 1 - b$, then $h'(b) = 1/(1 - b)$ and so $\boldsymbol{s}(b, 0) = 1 - b - ln(1 - b)$. Or, if $\boldsymbol{s}(b, 1) = (1 - b)^3$, then $h'(b) = 3 \cdot (1 - b)$ and so $\boldsymbol{s}(b, 0) = 1/2(3b^2 - 2b^3)$.

A number of authors have argued that epistemic scoring rules should be strictly proper, including Oddie (1997), Fallis (2007), Greaves and Wallace (2006), and Gibbard (2008). Here is one such argument taken, in substantially modified from, from Gibbard. Call a coherent credence function $\boldsymbol{b}$ *immodest* with respect to a scoring rule $\boldsymbol{S}$ when $\boldsymbol{b}$ uniquely minimizes expected epistemic disutility from its own perspective, so that $Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{c})) > Exp_{\boldsymbol{b}}(\boldsymbol{S}(\boldsymbol{b}))$ for all $\boldsymbol{c} \in \mathcal{B}_X$. An immodest $\boldsymbol{b}$ expects itself to be better, from a purely epistemic perspective, than any alternative set of credences. A *modest* $\boldsymbol{b}$, in contrast, assigns some other credence function a lower expected epistemic disutility than it assigns itself. Someone with modest credences is committed to expecting that she could do better, in epistemic terms, by holding opinions other than the ones she holds.

Modest credences, it can be argued, are epistemically defective because they undermine their own adoption and use. Recall that a person whose credences obey the laws of probability is committed to using the expectations derived from her credences to make estimates. These expected values represent her best judgments about the actual values of quantities. If, relative to a person's own credences, some alternative system of beliefs has a lower expected epistemic disutility, then, by her own estimation, that system is preferable from the epistemic perspective. This puts her in an untenable doxastic situation. She has a *prima facie*[15] epistemic reason, grounded in her beliefs, to think that she should not be relying on those very beliefs. This is a probabilistic version of Moore's paradox. Just as a rational person cannot fully believe 'X but I don't believe X,' so a person cannot rationally hold a set of credences that require her to estimate that some other set has higher epistemic utility. The modest person is always in this pathological position: her beliefs undermine themselves.

This sort of pathology makes it unreasonable for a modest person to rely on her beliefs when making estimates. As Gibbard argues, modest agents cannot rationally rely on their credences when estimating the prudential values of actions: modest credences lack 'guidance value,' as he puts it. Likewise, it has been suggested, see

---

[15] It does not follow that she has an all-things-considered reason to change her beliefs. Epistemic considerations are only one among many reasons that a person might have to alter one's beliefs.

Oddie (1997), one cannot rely on modest credences when deciding whether to gather new information about the world.

Given that modesty is a defect, it would be a serious flaw in an epistemic utility function if it required obviously rational credences to be modest. So, those impressed by the forgoing argument, might want to introduce the following principle.

> IMMODESTY: An epistemic scoring rule S should not render any credences modest when there are epistemic circumstances under which those credences are clearly the rational ones to hold.

This principle has no real content unless some set of 'clearly rational' credences can be identified, and the larger this set is, the more bite IMMODESTY will have. As a result, the principle's force will depend on the degree to which people can agree about which individual credences count as rational (in a given epistemic situation). If such agreement cannot be achieved, the principle is empty.

Fortunately, some credences do seem uncontroversially rational, and so IMMODESTY can be used to rule out various candidate scoring rules. Consider the exponential scores with $z \neq 2$. It seems clear that, in many epistemic situations, a believer can rationally have a credence that assumes some value other than the maximally opinionated 0 and 1 or the maximally undecided 1/2. Normally, for instance, it seems fine to align one's credences for the tosses of a fair die with the uniform probability over the six sides. However, with the exception of the Brier score, all exponential scores make such credences modest. To see why, note that a coherent agent with credence b for X will set $Exp_b(\alpha^z(c)) = b(1-c)^z + (1-b)c^z$. When $z \neq 1$, this has an extreme point at $q = b^y/[b^y + (1-b)^y]$ where $y = 1/(z-1)$, and this q is a maximum when $z < 1$ and a minimum when $z > 1$. So, while all the α^z agree that someone who is certain about X's truth-value can be immodest, only the Brier score α^2 permits immodesty across the board, since $q = b$ only at $z = 2$. When $z > 2$ doxastic conservatism is encouraged. Someone who is the least bit uncertain about X can improve expected epistemic utility by shifting her credence to q, which lies between b and 1/2. Thus, when $z > 2$ one can only be immodest about credences of 1, 1/2 or 0. When $1 < z < 2$ doxastic extremism rules: a person who is leaning, even slightly, toward thinking that X is true (or false), can improve expected epistemic utility by leaning even more strongly in that direction. Again, the message is that one should be either entirely opinionated or completely non-committal. This sort of extremism is even more pronounced for the absolute value score α^1, or any α^z with $z \leq 1$. Here a person who is more confident than not of X's truth (falsity) does best, by her own lights, by jumping to the conclusion that X is certainly true (false). To see the pernicious effects of this, imagine someone who believes that a given die is fair, and so assigns a credence of 5/6 to $X_j =$ 'face j will not come up on the next toss' for $j = 1, 2, .., 6$. If α^z measures epistemic utility for $z \leq 1$, the person will expect her beliefs to worse off, epistemically speaking, than someone who holds the *logically inconsistent* view that every X_j is certainly true! So, if modest credences are defective, then all credences within the intervals (0, 1/2) or (1/2, 1) are unsound according to every exponential rule, except the proper rule

α^2. Worse yet, this judgment is independent of both the content of the proposition believed and the believer's evidential position with respect to it.[16] Thus, to the extent that we regard intermediate credences as legitimate, IMMODESTY requires us to reject every α^z with $z \neq 2$.

IMMODESTY entails PROPRIETY provided that there are conditions under which *any* coherent credence function can be rationally held. It is plausible that there are such conditions. After all, for any assignment of probabilities $\langle p_n \rangle$ to $\langle X_n \rangle$ it seems that a believer could, in principle, have evidence that justifies her in thinking that each X_n has p_n as its objective chance.[17] Moreover, this could exhaust her information about X's truth-value. According to the 'Principal Principle' of Lewis (1980), someone who knows that the objective chance of X_n is p_n, and who does not possess any additional information that is relevant to questions about X_n's truth-value, should have p_n as her credence for X_n. Thus, $\langle p_n \rangle$ is the rational credence function for the person to hold under these conditions. In light of this, one might argue, the following restriction on scoring rules should hold:

> MINIMAL COHERENCE: An epistemic scoring rule should never preclude, *a priori*, the holding of any coherent set of credences.

This does not mandate coherence. It merely says that coherent credences should be at least *permissible* states of opinion.

MINIMAL COHERENCE and IMMODESTY suffice for PROPRIETY since together they entail any acceptable measure of epistemic disutility will make all coherent credence functions immodest by ensuring $Exp_b(S(c)) > Exp_b(S(b))$ for all $b \in \mathcal{P}_X$ and $c \in \mathcal{B}_X$. Of course, this line of reasoning will only convince those who accept the rationales given for MINIMAL COHERENCE and IMMODESTY. The former seems hard to deny: even those who are not convinced that epistemic rationality *requires* coherence should stop short of saying that any coherence is prohibited *a priori*. IMMODESTY, on the other hand, does have detractors, e.g., Maher (1900), and so it is worth exploring other constraints on epistemic scoring rules.

Before moving on, we should note that, even in the absence of IMMODESTY, MINIMAL COHERENCE imposes a substantive constraint on epistemic scoring rules. When combined with ADMISSIBILITY it requires all coherent credence functions to be admissible.

---

[16] All failures of modesty have this character if EXTENSIONALITY holds. Some credence *values* are prohibited independent of the propositions to which they attach or the believer's evidence with respect to them!

[17] Some have held objective chances are not probabilities. This seems unlikely, but explaining why would take us too far afield. In any case, nothing said here presupposes that all chance distributions are realized as probabilities. Only the converse is being assumed: for any probability distribution $\langle p_n \rangle$ over $\langle X_n \rangle$ it is possible that a believer knows that the objective chance of each X_n in p_n. This very weak assumption is especially compelling when EXTENSIONALITY is assumed. For in this case, the requirement is only that there could be some partition or other for which each X_n has p_n as its objective probability.

COHERENT ADMISSIBILITY: An epistemic scoring rule S is unreasonable if there are $b \in \mathcal{P}_X$ and $c \in \mathcal{B}_X$ such that $S(b, v) \geq S(c, v)$ for every $v \in \mathcal{V}_X$ and $S(b, v) > S(c, v)$ for some $v \in \mathcal{V}_X$.

While this allows a coherent credence function to exceed an incoherent one in epistemic disutility in some worlds, it prohibits this from happening in *every* world, for that would make the coherent credences irrational *a priori*. As we shall see, this idea has major ramifications for the prospects of justifying probabilism.

9 Convexity

While PROPRIETY delimits the range of allowable inaccuracy scores, significant variation still remains. Another restriction is provided by considering the effects of a scoring rule's *convexity* or *concavity* properties. A scoring rule is everywhere convex/flat/concave at v iff $1/2 \cdot S(b, v) + 1/2 \cdot S(c, v) > / = / < S(1/2 \cdot b + 1/2 \cdot c, v)$ for all credence functions b and c. For everywhere convex rules, the epistemic disutility of credences formed by evenly compromising between two credence functions is always lower than the average disutilities of the initial credence functions themselves. So, if Jacob and Joshua have credences b and c, and if Emily's credences m are an even compromise between the two, so that $m(X) = 1/2 \cdot b(X) + 1/2 \cdot c(X)$ for each X, then a convex/flat/concave rule will make Emily's beliefs more/as/less sound, from the epistemic perspective, than the average of those of Jacob and Joshua.

Here are two useful and general ways of stating that $S(\bullet, v)$ is convex, with each assumed to hold for all $v \in \mathcal{V}_X$ and all $b, c \in \mathcal{B}_X$.

- $S(b + \delta, v) - S(b, v) > S(b, v) - S(b - \delta, v)$ for every vector of real numbers $\delta = \langle \delta_n \rangle$ with $0 \leq b_n \pm \delta_n \leq 1$.
- For any credence functions $b^1, b^2, \ldots, b^m$, and $\mu_1, \mu_2, \ldots, \mu_m \geq 0$ with $\Sigma_k \mu_k = 1$, $\Sigma_m \mu_m \cdot S(b^m, v) > S((\Sigma_m \mu_m \cdot b^m), v)$.

It is easy to see that an additive rule $S(b, v) = \Sigma_n \lambda_X(X_n) \cdot s_n(b_n, v_n)$ is convex at b iff its components are convex at b. For a fixed truth-value v, a component $s(b, v)$ is *convex* at b just when $s(b + \delta, v) - s(b, v) < s(b, v) - s(b - \delta, v)$ for small δ. This means that any prospective gain in epistemic value that might be achieved by moving a credence incrementally closer to v is exceeded by the loss in value that would be incurred by moving the credence away from v by the same increment. $s(b, v)$ is *concave* at b when $s(b + \delta, v) - s(b, v) > s(b, v) - s(b - \delta, v)$ for small δ. It is *flat* at b when $s(b + \delta, v) - s(b, v) = s(b, v) - s(b - \delta, v)$.

As a way of getting a handle on these concepts, imagine a random process that, with equal probability, slightly raises or lowers credences of magnitude b for propositions with truth-value v. If epistemic disutility is measured using an additive rule that is convex at b, this process is, on balance, detrimental: it would be better, on average, if the credence just stayed at b. If epistemic disutility is convex, the process will be beneficial, on balance. If epistemic utility is flat it should have no average effect.

An additive score's convexity/concavity properties are reflected in the second derivatives of its components, when these exist. A positive/zero/negative value of $s''(b, v)$ signifies that the score is convex/flat/concave at b and v. The exponential scores provide a useful example. An easy calculation shows that $\alpha^{z''}(b, 1) = z \cdot (z - 1) \cdot (1 - b)^{z-2}$ and $\alpha^{z''}(b, 0) = z \cdot (z - 1) \cdot b^{z-2}$. These values are positive/zero/negative throughout $(0, 1)$ depending upon whether z is greater than/equal to/less than 1. So when $z > 1$, α^z is everywhere convex; it penalizes incremental shifts in credence away from a truth-value more than it rewards similar shifts toward that truth-value. α^1 is everywhere flat for both truth-values; its penalty for shifting away from a truth-value is equal to its reward for shifting toward it. α^z rules with $z < 1$ are everywhere concave: they reward incremental shifts toward truth-values more than they penalize similar shifts away from them.[18]

As our earlier discussion of exponential rules might suggest, the convexity properties of a scoring rule determine the degree of epistemic conservatism or adventurousness that it encourages. Altering any credence involves risking error, since one might move away from the truth, but it also carries the prospect of increased accuracy, since one might move closer to the believed proposition's truth-value. The more convex a score is at a point, the greater the emphasis it places on the avoidance of error as opposed to the pursuit of truth near that point. The more concave it is, the greater the emphasis it places on the pursuit of truth as opposed to the avoidance of error. As William James famously observed, the requirements to avoid error and to believe the truth – epistemology's two 'great commandments' – are in tension, and different epistemologies might stress one at the expense of the other. James endorsed a liberal view that accents the second commandment, while W. K. Clifford, his conservative foil, emphasized the first. This debate plays out in the current context as a dispute about convexity/concavity properties of measures of epistemic accuracy.[19] Convexity encourages (in at least a small way) Cliffordian conservatism in the evaluation of credences. It makes the epistemic costs of moving away from the truth a little higher than the benefits of comparable moves toward the truth. This makes it relatively risky to modify credences, and so discourages believers from making such changes without being compelled by evidence. In contrast, concavity fosters Jamesian liberalism by making the costs of moving away from a truth smaller than

---

[18] Here are some other examples: (i) The power scores are everywhere convex in both components for $2 \geq z > 1$; (ii) Since $-\chi''(b, 1) = 1/(1-b)^2$ and $\chi''(b, 0) = 1/b^2$ are positive everywhere between zero and one, the exponential rule is everywhere convex; (iii) the spherical score β has a convexity profile that varies across the unit interval. Its second derivative $\beta''(b, 1)$ is negative/zero/positive depending on whether b is $</=/>(7 - \sqrt{17})/8$, and $\beta''(b, 0)$ is negative/zero/positive depending on whether b is $</=/>(1 + \sqrt{17})/8$. Between $(7 - \sqrt{17})/8$ and $(1 + \sqrt{17})/8$ both its components are convex.

[19] Not every aspect of the James/Clifford debate is captured by the convexity question. For example, James held that the requirement to believe the truth can be justified on the basis of pragmatic considerations, whereas Clifford maintained that epistemic conservatism was justified on both practical and moral grounds. Also, the James/Clifford debate is bound up with issues of doxastic voluntarism. Despite the talk of credences being improved by moves toward or away from the truth, nothing said here should be taken to imply that agents ever have a hand in choosing what they believe.

the benefits of moving the same distance toward it. This can encourage believers to alter their credences even in the absence of corresponding changes in evidence. Flat measures set the costs of error and the benefits of believing the truth equal, and so it becomes a matter of indifference whether or not one makes a small change in credence.

Those with a Cliffordian outlook will suggest that epistemic utility should encourage a conservative policy by insisting on the following.

> CONVEXITY. $1/2 \cdot S(b, v) + 1/2 \cdot S(c, v) > S(1/2 \cdot b + 1/2 \cdot c, v)$ for any credence functions b and c.

Joyce (1998) defends CONVEXITY. Maher (2002) criticizes both the defense and the principle itself. A different defense will be offered below, but it will be instructive to start by seeing why Maher's criticisms go awry.

Maher's case against CONVEXITY rests ultimately on the claim that the non-convex absolute value score, $\alpha^1(b, v) = |b - v|$, is a plausible measure of epistemic disutility. Maher offers two considerations to support α^1. First, he writes, 'it is natural to measure the inaccuracy of b with respect to the proposition X in possible world v by $|b(X) - v(X)|$. It is also natural to take the total inaccuracy of b to be the sum of its inaccuracies with respect to each proposition.' (2002, p. 77) Second, he points out that there are many situations in which people measure accuracy using α^1. For instance, one naturally *averages* when calculating students' final grades, which is tantamount to thinking that the inaccuracy of their answers is best measured by the absolute value score.

Neither argument is convincing. While α^1 may be a natural scoring rule to use when grading papers, it is inappropriate in other contexts. When testing an archer's accuracy, for example, we use a target of concentric circles rather than concentric squares aligned with vertices up/down and left/right. There is a sound reason for this. With a square target, an archer whose inaccuracy is confined mainly along the vertical or horizontal axis is penalized less than one whose inaccuracy is distributed more evenly over both dimensions, e.g. an arrow that hits 9 inches below and 2 inches right of the bull's-eye is deemed more accurate than one that hits 6 inches from the bull's-eye at 45° from vertical. While one can contrive scenarios in which accuracy along the vertical or horizontal dimension is more important than accuracy along other directions, this is not the norm. There are no preferred directions for accuracy in archery; an error along any line running through the bull's eye counts for just as much as an error along any other such line. The square target uses an absolute value metric, while the circular one employs Euclidean distance, the analogue of the Brier score. Both modes of measurement can seem 'natural' in some circumstances, but unnatural in others.

Moreover, for all its 'naturalness', the absolute value measure produces absurd results if used across the board. We have already seen that α^1 is not strictly proper, but this is just the tip of the iceberg. Measuring epistemic disutility using α^1 – or any extensional, everywhere non-convex rule – lets *logically inconsistent* beliefs dominate probabilistically coherent beliefs in situations where the latter are clearly

the right ones to hold. This violates COHERENT ADMISSIBILITY. Suppose a fair die is about to be tossed, and let X_j say that it lands with j spots up. Though it is natural to set $b_j = 1/6$, the absolute value score forces one to pay an inescapable penalty, *not* just an expected penalty, for doing so. For if c is the inconsistent[20] credence assignment $c_j = 0$ for all j, then $\alpha^1(b, v) = 10/36 > \alpha^1(c, v) = 1/6$ for *every* truth-value assignment v.[21] So, no matter how the truth-values turn out, a believer does better by adopting the inconsistent $\langle 0, 0, 0, 0, 0, 0 \rangle$ over the correct consistent credence assignment $\langle 1/6, 1/6, 1/6, 1/6, 1/6, 1/6 \rangle$. Here we cross the boarder from probabilistic incoherence into logical inconsistency. The believer minimizes expected inaccuracy by being absolutely certain that every X_j is false even though logic dictates that one of them must be true. Measures of epistemic disutility that encourage this should be eschewed. This includes the absolute-value rule and every other additive rule whose components are uniformly non-convex. So, Maher's appeal to the absolute-value rule as a counterexample to CONVEXITY fails.

But, is there anything that can be said in favor of CONVEXITY? For those who hope to preserve a 'Cliffordian' picture of the relationship between belief change and evidence the answer is yes. To illustrate, suppose that a single ball will be drawn at random from an urn containing nine white balls and one black ball. On the basis of this evidence, a person might reasonably settle on a credence of $b = 0.1$ for the proposition that the black ball will be drawn and a credence of $b^- = 0.9$ for the proposition that a white ball will be drawn. Suppose that the ball is drawn, and that *we* learn that it is black. We are then asked to advise the person, without telling her which ball was drawn, whether or not to take a pill that will randomly raise or lower her credence for a black draw, with equal probability, by 0.01, while leaving her credence for a white draw at 0.9. If our only goal is to improve the person's epistemic utility, then our advice should depend on the convexity of the score for truths at credence 0.1. For a rule that is convex here, like the Brier score, the pill's disadvantages outweigh its advantages. For a rule that is concave at that point, like the spherical score β, the potential benefits are, on average, worth the risks. For a rule that is flat at 0.1, like the absolute value score, there is no advantage either way.

Concavity or flatness in scoring rules thus give rise to an epistemology in which the quality of a person's beliefs can be improved, or at least not degraded, by the employment of random belief-altering processes that vary credences independently of the truth-values of the propositions believed. Believers are then able to improve their *objective* expected epistemic utility by ignoring evidence and letting their opinions be guided by such processes. Cliffordians will see this sort of epistemic liberality as encouraging changes of opinion that are inadequately tied to corresponding changes in evidence. In any plausible epistemology, they will say, epistemic disutility should

---

[20] A credence assignment is logically inconsistent (not merely probabilistically incoherent) when it either assigns probability zero to all elements of some logical partition or when it assigns probability one to all members of some logically inconsistent set.

[21] This is because $\alpha^1(b, v^1) = 1/6 \cdot [(1 - b_1) + b_2 + \ldots + b_6] = 5/6$ when $b_n = 1/6$, whereas $\alpha^1(c, v) = 1/6 \cdot [1 + 0 + \ldots + 0] = 1/6$ when $c_n = 0$. The situation is the same for the other five truth-value assignments.

be at least slightly conservative; the penalties for belief changes that decrease accuracy should be at least a little more onerous, on average, than the penalties for staying put and forgoing a potential increase in accuracy. It might be, of course, that believers have non-epistemic reasons for altering their beliefs in the absence of changes in their evidence, but from a purely epistemic point of view this sort of behavior should not be rewarded. To the extent that one agrees with this conservative stance, one will be inclined toward CONVEXITY.

Proponents of PROPRIETY might not be moved by this argument. They can respond by noting that strictly proper scoring rules discourage the use of random belief-altering mechanisms even when these rules are not convex. If inaccuracy is measured using spherical score β, say, then a person with credences $\langle 0.1, 0.9 \rangle$ for $\langle$black, white$\rangle$ should *not* take a pill that will move her credences to $\langle 0.11, 0.9 \rangle$ or $\langle 0.09, 0.9 \rangle$ with equal probability even though β is strictly concave at $\langle 0.1, 0.9 \rangle$. Since β is proper, the person's own subjective expectations rank her credences above any other, and so taking the pill is a poor idea from her point of view. Even though the *objective* expected epistemic disutility of taking the pill is lower than that of refusing it, this is not the person's own view of things.

Defenders of CONVEXITY can counter by stressing that, whatever the person's subjective view of things, it remains true that if β measures epistemic disutility then, objectively speaking, she would be well advised to let herself be guided by a random process that has just as much chance of moving her away from the truth as it has of moving her toward it. This is objectionable, whatever the person herself may think. Moreover, her own epistemic position seems vexed. Suppose she knows how epistemic disutility is measured and explicitly aims to minimize it. If the line of reasoning explored in the previous paragraph were correct, then she should still stick to her $\langle 0.1, 0.9 \rangle$ credences. But, would this make sense? It would be one thing if the person refused the pill on the grounds that belief-alteration should not be randomly tied to truth. But, on the PROPRIETY rationale, this would not be the story. Rather, she would decline the pill because she is unsure whether or not a random beliefs forming process will raise her expected score. She thinks it probably will not, but she also recognizes that there is a one-in-ten chance that it will. Her subjective expectations rule the day, even though, objectively speaking, taking the pill is the better choice in the sense that the objective expected epistemic utility of taking it is greater than that of not taking it. Those with Cliffordian leanings will see this as intolerable. Our epistemology, they will claim, should not leave believers to wonder about whether, as an empirical matter, they would be wise to leave their opinions to the whims of random processes that are uncorrelated with the truth. The only way to avoid this, they will emphasize, is by requiring epistemic scoring rules to be everywhere convex.

Rather than trying to sort this out, we will treat CONVEXITY as an optional constraint, and turn to the question of assessing the prospects for a non-pragmatic vindication of probabilism.

10 Prospects for a Nonpragmatic Vindication of Probabilism

The idea of vindicating coherence on the basis of accuracy considerations – and without the use of Dutch book arguments or representation theorems – stems from the work of van Fraassen (1983) and Shimony (1988). These articles sought, in different ways, to show that incoherence leads believers to make poorly calibrated estimates of relative frequencies, while coherence enhances such calibration. Unfortunately, frequency calibration is a poor standard of epistemic assessment. The case against it is made in Joyce (1998), though many of the basic points were raised in (Seidenfeld 1985). The central problem is that calibration violates TRUTH-DIRECTEDNESS; my credences might be uniformly closer to the truth than yours, and you still might be better calibrated to the frequencies than I am.

Joyce (1998) sought to improve on the van Fraassen/Shimony approach by focusing on truth-values rather than frequencies, and by arguing that 'reasonable' measures of epistemic inaccuracy would make both the following true:

(I). For any incoherent credence function c there is a coherent b that is strictly more accurate than c under every logically possible assignment of truth-values, so that $S(c, v) > S(b, v)$ for all $v \in \mathcal{V}_X$.

(II). No coherent credence function b is accuracy-dominated in this way by any incoherent c: there is always a $v \in \mathcal{V}_X$ such that $S(c, v) > S(b, v)$.

In keeping with the idea that inadmissible credences are flawed, Joyce (1998) saw accuracy domination as an epistemic defect, and thus endorsed:

(III). The fact that incoherent credences are inadmissible relative to any reasonable measure of epistemic accuracy, and that coherent credences are admissible, is a strong, purely epistemic reason to prefer the latter over the former.

We will reevaluate the prospects for vindicating probabilism on the basis of (I)–(III), though with a focus on epistemic utility more broadly construed. (So, 'accuracy' in (I)–(III) should be replaced by 'epistemic utility.') The goal will be to determine whether it is possible to show that, relative to any reasonable epistemic scoring rule, all and only coherent credences are admissible, and to establish that any incoherent credence is always dominated by a coherent one.

For this sort of argument to work, we need some account of what makes a scoring rule 'reasonable'. Joyce (1998) required reasonable rules to obey TRUTH-DIRECTEDNESS, NORMALITY, SEPARABILITY and CONVEXITY, supplemented by a strong symmetry principle that forces complementary mixtures of equally accurate credences to be equally accurate.[22] Maher (2002) and Gibbard (2006) object to this latter principle, and Gibbard rejects NORMALITY. These objections have merit, and it would be best to find a vindication of probabilism that avoids such controversial premises.

---

[22] The precise requirement is that $S(\lambda b + (1-\lambda)c, v) = S((1-\lambda)b + \lambda c, v)$ when $S(b, v) = S(c, v)$ for any $0 \leq \lambda \leq 1$.

Two different sorts of arguments turn out to be feasible: one that rests on PROPRIETY, and another that relies on the weaker COHERENT ADMISSIBILITY. Results of the first sort are found in Lindley (1982) and Lieb et al. (Probabilistic Coherence and Proper Scoring Rules, unpublished). These works are in the tradition of de Finetti (1974) and Savage (1971) in that they focus on those features of scoring rules that give rational agents incentives to reveal their true credences. The underlying mathematical arguments can, however, be adapted to the task of providing a vindication of probabilism. After briefly discussing this work, a new and fairly sweeping result based on COHERENT ADMISSIBILITY will be proved.

The first PROPRIETY-based argument for probabilism is found in Lindley (1982).[23] Lindley assumes a scoring rule with the following features:

- Additive form,[24] $S(\boldsymbol{b}, \boldsymbol{v}) = \Sigma_n \lambda_X(X_n) \cdot s_n(b_n, v_n)$.
- Each $s_n(b, 1)$ and $s_n(b, 0)$ is defined for each $b \in [0, 1]$.
- Each $s_n(b, 1)$ and $s_n(b, 0)$ has a continuous first derivative that is defined everywhere on $[0, 1]$.
- These derivatives are such that $s_n{}'(b, 0) > 0$ and $s_n{}'(b, 1) < 0$ on $(0, 1)$. (This follows from TRUTH-DIRECTEDNESS given the previous conditions.)
- $s_n{}'(b, 0)/s_n{}'(b, 1)$ approaches 0 when b approaches 0 from above.
- $s_n{}'(b, 1)/s_n{}'(b, 0)$ approaches 0 when b approaches 1 from above.
- 1 is the unique admissible credence for T and 0 is the unique admissible value for $\sim$T. (Again, this follows from TRUTH-DIRECTEDNESS given the previous conditions.)

Lindley establishes the following result.

Theorem 1 (Lindley's Theorem) *Given the assumptions above, a set of credences* $\langle b_n \rangle$ *for* $\langle X_n \rangle$ *is admissible only if the values*

$$p_n = s_n{}'(b_n, 0)/[s_n{}'(b_n, 0) - s_n{}'(b_n, 1)]$$

collectively satisfy the laws of finitely additive probability. If, in addition, the mapping taking b_n to p_n is one-to-one for each n, so that $x \neq y$ only if $s_n{}'(x, 0)/[s_n{}'(x, 0) - s_n{}'(x, 1)] \neq s_n{}'(y, 0)/[s_n{}'(y, 0) - s_n{}'(y, 1)]$, then the p_n are coherent only if $\langle b_n \rangle$ is admissible.

While this does not yet show that all and only coherent credences are admissible, it does show that every set of admissible credences $\langle b_n \rangle$ has a 'known transform' $\langle p_n \rangle$ that obeys the laws of probability. And, if the $b_n \to p_n$ map is one-to-one, every set of credences whose transform is coherent is admissible.

---

[23] I have simplified Lindley's result somewhat by (a) ignoring the generalization to conditional probabilities, (b) assuming that credences fall in the unit interval (rather than some arbitrary closed interval of the real line), and (c) skipping some technicalities involving the values of the *component functions at the* endpoints of this interval.

[24] Lindley suggests (p. 6) that the additivity assumption can be relaxed, but does not give details.

Lindley remarks, almost in passing, that if one requires S (and hence each s_n) to be proper, then $b_n = p_n$ for each n. Moreover, if S is proper and truth-directed, then the last three conditions of the theorem are satisfied, and the $b_n \rightarrow p_n$ map is one-to-one. Putting this all together, we obtain the following as a straightforward consequence of Lindley's Theorem.

Corollary *If S is a truth-directed, proper scoring rule of additive form with $s_n(b, 1)$, $s_n(b, 0)$, $s_n'(b, 0)$, and $s_n'(b, 1)$ defined for each $b \in [0, 1]$, then the following are equivalent:*

- *b is incoherent (coherent)*
- *There is a (is no) credence function c such that $S(b, v) \geq S(c, v)$ for all $v \in \mathcal{V}_X$ with $S(b, v) > S(c, v)$ for at least one v.*

This is just the sort of result we seek. If the conditions on S strike one as essential to any reasonable definition of epistemic utility, then Lindley's Theorem entails that all and only coherent credences are admissible relative to any reasonable epistemic scoring rule. It does not, however, show that any incoherent credence is always strictly dominated by a coherent one, which would make the vindication of probabilism all the more convincing.

A similar success has recently been obtained by Lieb et al. (Probabilistic Coherence and Proper Scoring Rules, Unpublished)[25] Their assumptions are nearly identical to Lindley's, but they are able to prove a slightly stronger result by exploiting Savage's characterization of strictly proper scoring rules and by making use of some elegant mathematics involving a quantity called the 'Bergman divergence'. The advantages of this approach are (a) it does show that every incoherent credence is *strictly* dominated by a coherent one, (b) it does not presuppose that credences are defined over a partition, and (c) its method of proof (the author's report) generalizes to non-additive scoring rules.

Of course, these results will only be convincing to those who are already sold on the idea that epistemic disutility should be measured by a strictly proper scoring rule. Those looking for a vindication of probabilism that does not assume PROPRIETY, might be moved by the following theorem which requires only the weaker COHERENT ADMISSIBILITY, together with continuity.

Theorem 2 *Let S be a scoring rule defined on a partition $X = \langle X_n \rangle$. If S satisfies* TRUTH-DIRECTEDNESS *and* COHERENT ADMISSIBILITY*, and if $S(b, v)$ is finite and continuous for all b in $\mathcal{B}_X$ and $v \in \mathcal{V}_X$, then*

 (i). *every incoherent credence function is inadmissible relative to S and, moreover, is dominated by some coherent credence function, and*

---

[25] Though aware of Lindley's approach Lieb, et al., are not entirely clear about its application to proper scoring rules. They write (p. 3) 'the reliance on the transformation [from b_n to p_n], however, clouds the significance of Lindley's theorem,' and do not mention its application to proper rules. This is odd, given that Lieb et al. assume that scoring rules are proper, just the condition under which the transformation becomes inert (being the identity function).

(ii). *every coherent credence function is admissible relative to S.*

The proof can be found in the Appendix. Note that the theorem does *not* assume that S has additive form, nor does it require S to be proper, normal, separable or convex (though convexity will ensure S's continuity on the interior of $\mathcal{P}_X$). It is, insofar as the author knows, the least restrictive result of this sort that has yet been given. It also readily generalizes to the case where X is not a partition, though we will not carry out this exercise here. One clear limitation of the result is that it fails to address scoring rules, like the logarithmic rule, that go infinite at their extreme points. While this is a serious restriction, it may be that the result can be extended to such rules provided that (a) they are finite and continuous everywhere in the interior of $\mathcal{B}_X$, and (b) their limiting behavior near the boundary is sufficiently well-behaved. One natural conjecture is that CONVEXITY would suffice to ensure well-behaved limiting behavior, but we cannot pursue this matter here.

The theorem and its proof highlight how strong COHERENT ADMISSIBILITY really is: surprisingly, by forcing coherent credences to be admissible we go a long way toward ensuring that incoherent credences are inadmissible. In addition, it seems a likely conjecture that COHERENT ADMISSIBILITY, or something that implies it, will be essential to any result of this sort. This need not be worrisome, however, given the extreme stance that anti-probabilists would have to take in order to deny COHERENT ADMISSIBILITY.

11 Is Inadmissibility an Epistemic Defect?

The significance one assigns to these theorems will depend on whether one thinks that epistemic disutility satisfies the requirements being placed upon it, and on how plausible one finds claim (III). Since the merits of the requirements have been discussed, let's focus on (III). Aaron Bronfman (A Gap in Joyce's Argument for Probabilism, unpublished) has raised serious questions about (III)'s normative status.[26] The basic thrust of the objection, albeit not in Bronfman's terms, runs thus: (III) has a wide-scope reading and a narrow-scope reading. Read wide, it says that a credence function c is defective whenever some alternative b dominates it relative to *every* reasonable epistemic disutility. Read narrowly, it says that c is defective when, for each reasonable S, there is a b_S that dominates c relative to S. This b_S need not, however, dominate c relative to other reasonable scoring rules. Indeed, it is consistent with (I) and (II) that there might be no coherent b that dominates c with respect to *every* reasonable S. So, a narrow reading of (III) is required if (I) and (II) are to vindicate probabilism.

---

[26] The same objection was raised independently by Franz Huber and Alan Hájek (who inspired the Australia example). An excellent discussion of this, and related points, can be found in Huber (2007).

Bronfman argues that (III) is of questionable normative force when read narrowly. If no single coherent system of credences b is unequivocally better than the incoherent c, then a believer cannot move from c to b without risking increased inaccuracy relative to *some* reasonable scoring rule in some world. Since this is also true of coherent credences – for every coherent b there is an incoherent c such that $S(c, v) < S(b, v)$ for some reasonable S and some truth-value assignment v – (I) and (II) offer no compelling rationale for having credences that obey the laws of probability. The mistake in Joyce (1998), Bronfman claims, lies in assuming that a credence function that is defective according to each reasonable way of measuring epistemic disutility is thereby defective simpliciter.

To appreciate the worry, consider an analogy. Suppose ethicists and psychologists somehow decide that there are just two plausible theories of human flourishing, both of which make geographical location central to well-being. Suppose also that, on both accounts, it turns out that for every city in the U.S. there is an Australian city with the property that a person living in the former would be better off living in the latter. The first account might say that Bostonians would be better off living in Sydney, while the second says they would do better living in Coober Pedy. Does it follow that any individual Bostonian will be better off living in Australia? It surely would follow if both theories said that Bostonians will be better off living in Sydney. But, if the first theory ranks Sydney > Boston > Coober Pedy, and the second ranks Coober Pedy > Boston > Sydney, then we cannot definitively conclude that the person will be better off in Sydney, nor that she will be better off in Coober Pedy. So, while both theories say that a Bostonian would be better off living somewhere or other in Australia, it seems incorrect to conclude that she will be better off in Australia per se because the theories disagree about which places in Australia would make her better off.

While Bronfman's objection does have some intuitive force, it still seems problematic that for incoherent credences, but not coherent credences, one is in a position to know that some alternative set of credences is better solely on the basis of knowledge of the properties of reasonable epistemic scoring rules. The problem remains even if we cannot identify what the alternative credences might be. In our analogy, it seems problematic for Bostonians that they know they are not best off in Boston, and that there is no American city in which they would be best off either. The Australians have at least this advantage: they know that if there are better places for them to be then the best such places are in Australia. Likewise, when apprised of the results of the above proofs, coherent agents know that if there is an incoherent set of credences with a higher epistemic utility than their own, then there is a coherent set of credences that is sure to be even better whatever happens. The Australians at least know that they are in the right country; agents with coherent credences at least know that they are in the right region of $\mathcal{P}_X$.

More importantly, however, Bronfman's objection only applies if there is no determinate fact of the matter about which reasonable measure of inaccuracy is correct in a given context. If any reasonable scoring rule is as good as any other when it comes to measuring epistemic disutility, then (I)–(III) cannot vindicate coherence without the help of an inference from 'c is defective on every reasonable measure'

to '*c* is unqualifiedly defective'. If, on the other hand, there is some single reasonable epistemic disutility, then the wide and narrow readings of (III) collapse and Bronfman's worries become moot. It may be that the correct scoring rule varies with changes in the context of epistemic evaluation, and it may even be that we are ignorant of what the rule is, but the nonpragmatic vindications of probabilism we have been considering are untouched by Bronfman's objection as long as there is some one correct rule in any given context of epistemic evaluation. Consider two further analogies. Many philosophers claim that the standards for the truth of knowledge ascriptions vary with context, but that in any fixed context a single standard applies. Under these conditions, if every standard of evaluation has it that knowledge requires truth then knowledge requires truth per se. Similarly, even if we do not know which ethical theory is correct, as long as there is some correct theory, then the fact that every reasonable candidate theory tells us to help those in need means that we have a moral obligation to help those in need. So, the argument from (I) to (III) to the requirement of coherence goes through with the help of one further premise:

(IV). Only one scoring rule functions as the correct measure of epistemic disutility in any context of epistemic evaluation.

How plausible is this premise? It is hard to say in the abstract without some specification of the relevant epistemic context. However, there are certainly contexts in which it makes sense to single out one scoring rule as uniquely best. For example, in contexts where we are concerned about *pure accuracy* of truth-value estimation, the Brier score has properties that make it an excellent tool for assessing epistemic utility.

12 Homage to the Brier Score

There are a number of reasons for using the Brier score to assess epistemic accuracy. First, in addition to being truth-directed, strictly proper, and convex, it is continuous, separable, extensional and normal. In many contexts of evaluation – specifically those involving assessments of pure accuracy, in which questions of holistic dependence or informativeness are ignored – these are reasonable properties for a scoring rule to have.

Moreover, as Savage (1971) showed, the Brier score is the only rule with these properties that can be extended to a measure of accuracy for probability estimates generally. It is natural to think of truth-value estimation as a species of probability estimation. One can assess such estimates using an *extended scoring rule* that takes each $b \in \mathcal{B}_X$ and $p \in \mathcal{P}_X$ to a real number $S^+(b, p) \geq 0$ that gives the inaccuracy of b's values as estimates of the probabilities assigned by p. In keeping with the gradational character of estimation, if b_n is always strictly between c_n and p_n, then $S^+(b, p) < S^+(c, p)$. S^+ *extends* a truth-value based rule S when

$S^+(\boldsymbol{b}, \boldsymbol{v}) = S(\boldsymbol{b}, \boldsymbol{v})$ for every $\boldsymbol{v}$. Extended scoring rules can be strictly proper, convex, separable, additive or normal. In his (1971) Savage proved the following result (in slightly different terms):

Theorem 3 *If an extended scoring rule S^+ is strictly proper, convex, additive and normal, then it has the quadratic form $S^+(\boldsymbol{b}, \boldsymbol{p}) = \Sigma_n \lambda_n \cdot (p_n - b_n)^2$.*

So, if one thinks that accuracy evaluations for truth-values should dovetail with accuracy evaluations for probability estimates, and that the latter should be strictly proper, convex, additive and normal, then one will assess truth-value estimates using a function of the form $S^+(\boldsymbol{b}, \boldsymbol{v}) = \Sigma_n \lambda_n \cdot (v_n - b_n)^2$. If, in addition, one also accepts EXTENSIONALITY, one must use the Brier score since EXTENSIONALITY requires $\lambda_n = \lambda_m$ for all m and n.

Savage provided yet another compelling characterization of the Brier score. Instead of assuming NORMALITY, which makes the inaccuracy a $\boldsymbol{b}$ as an estimate of $\boldsymbol{p}$ a function of the absolute differences $|p_n - b_n|$, he insisted on $S^+(\boldsymbol{b}, \boldsymbol{p}) = S^+(\boldsymbol{p}, \boldsymbol{b})$ for all coherent $\boldsymbol{b}$ and $\boldsymbol{p}$. Again, the score so characterized has the quadratic form $\Sigma_n \lambda_n \cdot (v_n - b_n)^2$. Selten (1998) obtained the same result using a related symmetry property. Selten offers an argument that is compelling for both properties. He imagines a case in which we know that either $\boldsymbol{p}$ or $\boldsymbol{b}$ is the right probability, but do not know which. He writes:

> The [inaccuracy] of the wrong theory is a measure of how far it is from the truth. It is only fair to require that this measure is 'neutral' in the sense that it treats both theories equally. If $\boldsymbol{p}$ is wrong and $\boldsymbol{b}$ is right, then $\boldsymbol{p}$ should be considered to be as far from the truth as $\boldsymbol{b}$ in the opposite case that $\boldsymbol{b}$ is wrong and $\boldsymbol{p}$ is right.... A scoring rule which is not neutral [in this way] is discriminating on the basis of the location of the theories in the space of all probability distributions.... Theories in some parts of this space are treated more favorably than those in some other parts without any justification. (Selten 1998, p. 54 minor notational changes)

This defense seems correct, at least when considerations about the informativeness of propositions are being set aside.

A final desirable feature of the Brier score has to do with the relationship between truth-value estimates and frequency estimates. Let $\boldsymbol{Z}$ be an arbitrary finite set of propositions and let $\{\boldsymbol{Z}_j\}$ be any partitioning of $\boldsymbol{Z}$ into disjoint subsets. n_j is the cardinality of $\boldsymbol{Z}_j$, and $N = \Sigma_j n_j$ is the cardinality of $\boldsymbol{Z}$. Imagine a person with credences $\boldsymbol{b}$ who makes an estimate f_j for the frequency of truths in each $\boldsymbol{Z}_j$. Following Murphy (1973), we can gauge the accuracy of her estimates using an analogue of the Brier score called the *calibration index*.

$$\boldsymbol{Cal}(\{\boldsymbol{Z}_j\}, \langle f_j \rangle, \boldsymbol{v}) = \Sigma_j (n_j/N) \cdot (Freq_{Zj}(\boldsymbol{v}) - f_j)^2$$

As already noted, a coherent believer will use average credences as estimates of truth-frequencies, so that $f_j = \Sigma_{Z \in Zj} \boldsymbol{b}(Z)/n_j$. It is then possible to write:

$$\boldsymbol{Cal}(\{\boldsymbol{Z}_j\}, \boldsymbol{b}, \boldsymbol{v}) = (1/N) \cdot [\Sigma_j \alpha_{Zj}^2(\boldsymbol{b}, \boldsymbol{v}) - 2 \cdot \Sigma_j (\Sigma_{Y \neq Z \in Zj} (\boldsymbol{v}(Y) - \boldsymbol{b}(Z)) \cdot (\boldsymbol{v}(Z) - \boldsymbol{b}(Y)))]$$

This messy equation assumes a simple and illuminating form when propositions are grouped by credence. Suppose that each element of Z has a credence in $\{b_1, b_2, \ldots, b_J\}$, and let $Z_j = \{Z \in Z : b(Z) = b_j\}$. It then follows that $Cal(\{Z_j\}, b, v) = (1/N) \cdot \Sigma_j \Sigma_{Z \in Z_j} (Freq_{Z_j}(v) - b_j)^2$. So, relative to this partitioning, b produces frequency estimates that are perfectly calibrated ($Cal = 0$) when half of the propositions assigned value 1/2 are true, two-fifths of those assigned value 2/5 are true, three-fourths of those assigned value 3/4 are true, and so on. b's estimates are maximally miscalibrated ($Cal = 1$) when all truths in X are assigned credence 0, and all falsehoods are assigned credence 1.

As Murphy showed, relative to this particular partition the Brier score is a straight sum of the calibration index and the average *variance* in truth-values across the elements of $\{Z_j\}$. For a given v, the variance in truth-value across Z_j is given by $s^2(Z_j, v) = (1/n_j) \cdot \Sigma_{Z \in Z_j} (Freq_{Z_j}(v) - v(Z))^2$. To measure the average amount of variation across all the sets in $\{Z_j\}$ Murphy weighted each Z_j by its size to obtain the *discrimination index*[27]

$$Dis(\{Z_j\}, b, v) = \Sigma_j (n_j/N) \cdot s^2(Z_j, v).$$

This measures the degree to which b's values sort elements of Z into classes that are homogenous with respect to truth-value. Perfect discrimination ($Dis = 0$) occurs when each Z_j contains only truths or only falsehoods. Discrimination is minimal ($Dis = 1/4$) when every Z_j contains exactly as many truths as falsehoods.

As Murphy demonstrated, the sum of the calibration and discrimination indexes is just the Brier score.[28]

MURPHY DECOMPOSITION. $Cal(\{Z_j\}, b, v) + Dis(\{Z_j\}, b, v) = Brier(b, v)$

The Brier score thus incorporates two quantities that seem germane to assessments of epistemic accuracy. Other things equal, it enhances accuracy when credences sanction well-calibrated estimates of truth-frequency. It is likewise a good thing, *ceteris paribus*, if credences sort propositions into classes of similar truth-values.

Even so, neither calibration nor discrimination taken alone is an unalloyed good. As Murphy noted, some ways of improving one at the expense of the other harm overall accuracy. One can, for example, ensure perfect calibration over a set of propositions that is closed under negation by assigning each proposition in the set a credence of 1/2. Such credences are highly inaccurate, however, because they do not discriminate truths from falsehoods. Conversely, one achieves perfect discrimination by assigning credence one to every falsehood and credence zero to every truth, but one is then inaccurate because one is maximally miscalibrated. The moral here is that calibration and discrimination are components of accuracy that must be balanced off against one another in a fully adequate epis-

---

[27] Murphy actually broke the discrimination index into two components.
[28] For the proof, use $s^2(Z_j, v) = (1/n_j) \cdot \Sigma_{Z \in Z_j}[v(Z)^2 - n_j \cdot (Freq_{Z_j}(v))^2]$.

temic scoring rule that is designed to capture pure accuracy. The fact that the Brier score, a rule with so many other desirable properties, balances the two off in such a simple and beautiful way provides yet another compelling reason to prefer it as a measure of epistemic accuracy across a wide range of contexts of evaluation.

This is not to say that the Brier score is the right rule for every epistemic context. 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, in which case a quadratic rule $\Sigma_n \lambda_n \cdot (\nu_n - b_n)^2$ might be called for. Doubtless there are other options. Still, as long as there is one notion of epistemic disutility at play in any given context, and as long as that notion is captured by some continuous, truth-seeking scoring rule that allows all coherent credences to be rationally held, it will remain true that coherence contributes to the epistemic value of a set of credences while incoherence is a detriment.

13 Appendix: Proof of Theorem

Theorem 4 *Let **S** be a scoring rule defined on a partition **X** = $\langle X_n \rangle$. If **S** satisfies* TRUTH-DIRECTEDNESS *and* COHERENT ADMISSIBILITY, *and if **S**(**b**, **v**) is finite and continuous for all **b** in $\mathcal{B}_X$ and **v** $\in \mathcal{V}_X$, then*

  (i). *every incoherent credence function is inadmissible relative to **S** and, moreover, is dominated by some coherent credence function, and*
 (ii). *every coherent credence function is admissible relative to **S**.*

Proof (ii) is just a restatement of COHERENT ADMISSIBILITY. We establish (i) by means of a fixed point theorem (with some inspiration from the method of Fan et al. (1957)).

Fix an incoherent credence function $c = \langle c_n \rangle \in \mathcal{B}_X$. For each n, define a map $f_n(b) = S(b, v^n) - S(c, v^n)$ from the set of coherent credence functions $\mathcal{P}_X$ into the real numbers. $f_n(b)$ is the difference in **S**-score between the coherent **b** and the incoherent **c** at the world v^n. f_n is clearly continuous everywhere in the interior of $\mathcal{P}_X$ given that $S(\bullet, v^n)$ is continuous in this region.

To prove (i) it suffices to find a $b \in \mathcal{P}_X$ with $f_n(b) < 0$ for all n. Start by supposing that $\Sigma_n c_n < 1$. (The $\Sigma_n c_n > 1$ proof is a mirror image, see below.) Define N *coherent* points $b^m = \langle c_1, c_2, \ldots, c_{m-1}, (1 - \Sigma_{n \neq m} c_n), c_{m+1}, \ldots, c_N \rangle$ and notice that, in light of TRUTH DIRECTEDNESS and since $(1 - \Sigma_{n \neq m} c_n) > c_m$, we have both $S(b^m, v^m) < S(c, v^m)$ and $S(b^n, v^n) > S(c, v^n)$ for $n \neq m$. So, if we consider the N points $f(b^m) = \langle f_1(b^m), f_2(b^m), \ldots, f_N(b^m) \rangle$ of $\mathfrak{R}^N$ we will find that $f(b^1)$ is negative in the first coordinate and positive elsewhere $f(b^2)$ is negative in the second coordinate and positive elsewhere, $f(b^3)$ is negative in the third coordinate and positive elsewhere, and so on.

Now, consider B^+, the convex hull of $\{b^m\}$. This is the compact, convex subset of $\Re^n$ composed of all probability functions of form $p = \Sigma_m \mu_m \cdot b^m$ where $\mu_1, \mu_2, \ldots, \mu_N \geq 0$ and $\Sigma_m \mu_m = 1$. Since all of the b^m are in the interior of $\mathcal{P}_X$, and in virtue of the way the b^m are defined, elements of B^+ can be written as

$$p = \langle p_n \rangle = \langle \mu_n \cdot (1 - \Sigma_{k \neq n} c_k) + (1 - \mu_n) \cdot c_n \rangle$$

And, since $(1 - \Sigma_{k \neq n} c_k) > c_n$ it follows that $p_n > c_n$ when $\mu_n > 0$ and $p_n = c_n$ when $\mu_n = 0$. In virtue of this and TRUTH-DIRECTEDNESS we have the following little result, whose importance will emerge as the proof progresses.

Lemma 1 *If $p = \Sigma_m \mu_m \cdot b^m$ is in B^+, and if $\mu_m = 0$ then $f_m(p)$ is positive.*

Proof This is a straightforward dominance argument. Assume $\mu_m = 0$. Recall that v^m contains a 1 in its mth and a 0 everywhere else. As just noted, $p_k = c_k$ holds for $k = m$ and for every other coordinate at which $\mu_k = 0$. But, of course, some $\mu_n \neq \mu_m$ must be positive, and for all these we have $p_n > c_n$. So, the values of c are everywhere as close, and sometimes closer, to the values of v^m than are the values of p. TRUTH-DIRECTEDNESS then requires that $S(p, v^m) > S(c, v^m)$, i.e., $f_m(p) > 0$.

We now aim to define a function $G: B^+ \rightarrow B^+$ that is continuous on B^+ and has the property that $p = G(p)$ only if $f_n(p) < 0$ for all n. If such a function can be found, then the result we seek will be a consequence of:

Theorem 5 (Brouwer's Fixed Point Theorem) *Let $A \subset \Re^n$ be nonempty, compact, and convex. If $F: A \rightarrow A$ is continuous throughout A, then F has a fixed point, i.e., there exists $a \in A$ such that $a = F(a)$.*

To obtain G, start by defining a function $M(p) = 1/N \cdot (\Sigma_n f_n(p))$ from B^+ into $\Re$. $M(p)$ is the mean value of the $f_n(p)$. It is continuous everywhere on B^+ since each f_n is continuous on B^+. Next, for each $n = \{1, 2, \ldots, N\}$ define

$$g_n(p) = \max\{f_n(p) - M(p), 0\}$$

Each g_n is continuous and non-negative. $g_n(p) > 0$ exactly when $f_n(p)$ exceeds the mean value of the $f_m(p)$. For each $p \in B^+$ specify a set of $N + 1$ mixing coefficients:

$$\lambda_n(p) = g_n(p) / [1 + \Sigma_{1 \leq k \leq N} g_k(p)] \text{ for } n \leq N$$
$$\lambda_{N+1}(p) = 1 / [1 + \Sigma_{1 \leq k \leq N} g_k(p)]$$

Cleary, the values $\lambda_n(p)$ are all non-negative and sum to one, and $\lambda_{N+1}(p) > 0$. All the λ_n are continuous throughout B^+ as well. Finally, define

$$G(p) = (\Sigma_{1 \leq n \leq N} \lambda_n(p) \cdot b^n) + \lambda_{N+1}(p) \cdot p$$

G is continuous throughout B^+, and $G(p) \in B^+$ because $p \in B^+$ and B^+ is closed under mixing.

Since G satisfies the prerequisites for the Brouwer theorem it has a fixed point: there is a $p \in B^+$ such that $p = G(p)$.

Lemma 2 $p = G(p)$ *only if* $f_n(p) < 0$ *for all n.*

Proof The identity $p = (\Sigma_{1 \leq n \leq N} \lambda_n(p) \cdot b^n) + \lambda_{N+1}(p) \cdot p$ can hold in only two ways. First, it could be that $\lambda_n(p) = 0$ for all $n \leq N$, so that $\lambda_{N+1}(p) = 1$, which happens only if $f_n(p) - M(p) \leq 0$ for all N. But, the only way for a set of real numbers to *all* be less than or equal to their mean is for all of them to be equal to the mean (and so equal to one another). So, the $f_n(p)$ must be either all positive (if $M(p) > 0$), all zero (if $M(p) = 0$), or all negative (if $M(p) < 0$). The first two possibilities entail $S(p, v^n) \geq S(c, v^n)$ for all n, which contradicts COHERENT ADMISSIBILITY. Thus, we must have $f_n(p) < 0$ for N when $\lambda_n(p) = 0$ for all $n \leq N$.

Suppose now that $\lambda_n(p) > 0$ for some $n \leq N$ Here, $p = G(p)$ will hold iff $p \cdot \Sigma_{1 \leq n \leq N} \lambda_n(p) = \Sigma_{1 \leq n \leq N} \lambda_n(p) \cdot b^n$. Since $\Sigma_{1 \leq n \leq N} \lambda_n(p) > 0$ there are mixing coefficients $\pi_n = \lambda_n(p) / (\Sigma_{1 \leq n \leq N} \lambda_n(p))$ that allow us to write p as a mixture of the b^n, so that $p = \Sigma_{1 \leq n \leq N} \pi_n \cdot b^n$. However, if some $\lambda_n(p) > 0$ then some $f_n(p)$ strictly exceeds the mean, and this can only occur if some other $f_m(p)$ falls strictly below the mean. But, if $f_m(p) < M(p)$, then $\lambda_m(p)$ and hence π_m will be zero. Accordingly, if $\lambda_n(p) > 0$ for some n, then there exits values of $m \leq N$ for which $\lambda_m(p) = \pi_m = 0$, and all values of m for which $f_m(p) < M(p)$ have $\lambda_m(p) = \pi_m = 0$. But, in virtue of Lemma-1, $f_m(p)$ will be negative at each one of these values. But, if there are values of $f_m(p)$ that fall below the mean, and if all of these are positive, then $f_n(p)$ must be positive for *all* n. Thus, the assumption that $\lambda_n(p) > 0$ for some n yields the conclusion that $S(p, v^n) > S(c, v^n)$ for all n. Since this contradicts COHERENT ADMISSIBILITY, $\lambda_n(p) > 0$ cannot hold when p is a fixed point of G.

Thus, the only scenario in which $p = G(p)$ that is consistent with the assumptions of the proof is one in which $S(p, v^n) < S(c, v^n)$ for all n.

On the basis of Lemma-2 and the Brouwer Fixed Point Theorem, we have found what we wanted – a probability p than dominates the incoherent c – at least in the case when the c_n sum to less than one.

The proof for $\Sigma_n c_n > 1$ proceeds as a mirror image. The one subtlety concerns the definition of the b^m. Instead of subtracting from c_m and leaving the other c_n fixed (which will not work when the c_n are too large), one leaves c_m fixed and diminishes all the other c_n. Specifically, the arguments of b^m are given by $b_m = c_m$ and $b_n = c_n \cdot [(1 - c_m)/(\Sigma_{k \neq m} c_k)]$. Then one proves:

Lemma 1* *If* $p = \Sigma_m \mu_m \cdot b^m$ *is in* B^+, *and if* $\mu_k = 0$ *for* $k \neq m$ *then* $f_k(p)$ *is negative.*

Again, this is a straightforward dominance argument since all the elements of each b^m, other than the one at m, are less than their associated element of c.

The definition of the g_n needs to be modified so that

$$g_n(p) = \max\{M(p) - f_n(p), 0\}$$

With this change, $g_n(\boldsymbol{p}) > 0$ exactly when $f_n(\boldsymbol{p})$ falls *below* the mean, and $g_n(\boldsymbol{p}) = 0$ when $f_n(\boldsymbol{p}) \geq M(\boldsymbol{p})$. The coefficients $\lambda_n(\boldsymbol{p}), \ldots, \lambda_{N+1}(\boldsymbol{p})$ are defined as before, and the Brouwer theorem again guarantees the existence of a $\boldsymbol{p} \in B^+$ with $\boldsymbol{p} = G(\boldsymbol{p})$.

Lemma 2[*] $\boldsymbol{p} = G(\boldsymbol{p})$ *only if* $f_n(\boldsymbol{p}) < 0$ *for all n.*

Proof Again, the identity $\boldsymbol{p} = G(\boldsymbol{p})$ can hold in two ways. If, $\lambda_n(\boldsymbol{p}) = 0$ for all $n \leq N$, then $M(\boldsymbol{p}) - f_n(\boldsymbol{p}) \leq 0$ for all N. As, before, this means that the $f_n(\boldsymbol{p})$ must be either all positive, all zero, or all negative, and only the last possibility conforms to COHERENT ADMISSIBILITY.

If $\lambda_n(\boldsymbol{p}) > 0$ for some *n* then, as before, write $\boldsymbol{p} = \Sigma_{1 \leq n \leq N} \pi_n \cdot \boldsymbol{b}^n$. But, $\lambda_n(\boldsymbol{p}) > 0$ implies that $f_n(\boldsymbol{p})$ is strictly less than the mean, and so there must be some other $f_m(\boldsymbol{p})$ that exceeds the mean and for which $\lambda_m(\boldsymbol{p}) = \pi_m = 0$. Indeed, we have $\lambda_m(\boldsymbol{p}) = \pi_m = 0$ for all $f_m(\boldsymbol{p})$ that exceed $M(\boldsymbol{p})$. Lemma-1[*] tells us that all such $f_m(\boldsymbol{p})$ must be negative, which ensures that all the $f_n(\boldsymbol{p})$, whether above, at, or below the mean must be negative.

This completes the proof of the theorem.

References

Brier, G. W. (1950) "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review* **78**: 1–3.

Christensen, D. (1996) "Dutch Books Depragmatized: Epistemic Consistency for Partial Believers," *Journal of Philosophy* **93**: 450–79.

de Finetti, B. (1937) "La prévision : ses lois logiques, ses sources subjectives," *Annales de l'institut Henri Poincaré*, 7: 1–68. Translated as "Foresight: Its Logical Laws, Its Subjective Sources," in H. Kyburg and H. Smokler, eds., *Studies in Subjective Probability*. New York: John Wiley, 1964: 93–158.

de Finetti, B. (1974) *Theory of Probability*, vol. 1. New York: John Wiley and Sons.

Fallis, D. (2007) "Attitudes Toward Epistemic Risk and the Value of Experiments," *Studia Logica*, **86**: 215–246.

Fan, K., Glicksberg. I. and Hoffman, A. J. (1957) "Systems of Inequalities Involving Convex Functions," *Proceedings of the American Mathematical Society* **8**: 617–622.

Gibbard, A. (2008) "Rational Credence and the Value of Truth," in T. Gendler and J. Hawthorne, eds., *Oxford Studies in Epistemology* vol. 2. Oxford: Clarendon Press.

Greaves, H., Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind* **115**: 607–32.

Horwich, Paul. (1982) *Probability and Evidence*. New York: Cambridge University Press.

Howson, C. (2008) "De Finetti, Countable Additivity, Consistency and Coherence," *British Journal for the Philosophy of Science* **59:** 1–23

Howson, C. and Urbach, P. (1989) *Scientific Reasoning: The Bayesian Approach*. La Salle: Open Court.

Huber, F. (2007) "The Consistency Argument for Ranking Functions," *Studia Logica* **86**: 299–329

Lindley, D. (1982) "Scoring Rules and the Inevitability of Probability," *International Statistical Review* **50**: 1–26.

Jeffrey, R. (1986) "Probabilism and Induction," *Topoi* **5**: 51–58.

Joyce, J. (1998) "A Non-Pragmatic Vindication of Probabilism," *Philosophy of Science* **65**: 575–603.

Lewis, D. (1980) "A Subjectivist's Guide to Objective Chance," in *Studies in Inductive Logic and Probability*, edited by R. Jeffrey, vol. 2, Berkeley: University of California Press: 263–94.

Maher, P. (1990) "Why Scientists Gather Evidence," *British Journal for the Philosophy of Science* **41**:103–19.

Maher, P. (2002) "Joyce's Argument for Probabilism," *Philosophy of Science* **96**: 73–81.

Murphy, A. (1973) "A New Vector Partition of the Probability Score," *Journal of Applied Meteorology* **12**: 595–600.

Oddie, G. (1997) "Conditionalization, Cogency, and Cognitive Value," *British Journal for the Philosophy of Science* **48**: 533–41.

Ramsey, F. P. (1931) "Truth and Probability," in R. Braithwaite, ed., *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul: 156–98.

Savage, L. J. (1971) "Elicitation of Personal Probabilities," *Journal of the American Statistical Association* **66**: 783–801.

Savage, L. J. (1972) *The Foundations of Statistics*, 2nd edition New York: Dover.

Schervish, M. (1989) "A General Method for Comparing Probability Assessors," *The Annals of Statistics* **17**: 1856–1879.

Seidenfeld. T. (1985) "Calibration, Coherence, and Scoring Rules," *Philosophy of Science* **52**: 274–294.

Selten, R. (1998). "Axiomatic Characterization of the Quadratic Scoring rule," *Experimental Economics* **1**: 43–62.

Shimony, A. (1988) "An Adamite Derivation of the Calculus of Probability," in J. H. Fetzer, ed., *Probability and Causality*. Dordrecht: D. Reidel: 151–161.

van Fraassen, B. (1983) "Calibration: A Frequency Justification for Personal Probability," in R. Cohen and L. Laudan, eds., *Physics Philosophy and Psychoanalysis*. Dordrecht: D. Reidel: 295–319.