The Accuracy of Partial Beliefs, I and II
FEW 2004, May 21-23, 2004

Jim Joyce
Department of Philosophy
University of Michigan
e-mail: jjoyce@umich.edu

Traditional Epistemology

Traditional epistemology is *dogmatic* and *alethic*; it is organized around the concepts of *full belief* and *truth*.

- A person fully believes a proposition *x* when he accepts *x* as true, rejects *x* as false, or suspends judgment about *x*.

- Full beliefs are appropriately evaluated according to a *categorical* standard of accuracy. The extent to which a full belief about *x* fits the facts is solely a matter of the belief's "valence" (accept-*x*, reject-*x*, suspend belief), and *x*'s truth-value (no "verisimilitude").

Most of the central concepts of traditional epistemology (justification, reliability, sensitivity) are best understood in terms of the goal of having true full beliefs.

James's "Two Great Commandments"

William James: The "two great commandments" of epistemology are that "we must know the truth, and we must avoid error."

As James notes, there is a tension here: when one accepts a proposition in an attempt to reap the gains of true belief, one risks believing falsely.

Extremes: One can maximize true beliefs by believing everything, but at the cost of logical inconsistency. One can avoid error by suspending judgment about everything, but at the cost of having no beliefs.

- What is the right way to balance off the benefits of holding true beliefs with the costs of holding false beliefs?
- What role does the requirement of logical consistency play in this balancing act? Should we be willing to risk inconsistency to secure a large number of true beliefs?

THE JAMES/CLIFFORD DEBATE

Clifford: "It is wrong always, everywhere, and for everyone to believe anything upon insufficient evidence."

James: One can choose which "commandment" to emphasize, and the benefits of true belief often offset the risks of error.

> "We may regard the chase for truth as paramount, and the avoidance of error as secondary; or we may treat the avoidance of error as more imperative, and let truth take its chance… Clifford's exhortation has to my ears a fantastic sound. It is like a general informing his soldiers that it is better to keep out of battle forever than to risk a single wound."

While James focused on the practical costs and benefits of believing, and Clifford emphasized the moral and social effects, it is possible to frame the issue that divided them in terms of the purely epistemic consequences of believing truths and falsehoods.

PROBABILISTIC EPISTEMOLOGY

Probabilistic epistemology is non-dogmatic. It seeks to replace full beliefs with partial beliefs (= degrees of belief or *credences*).

> A person's degree of belief in the proposition *x* is her level of confidence in its truth. It determines the extent to which she is disposed to presuppose *x* in her reasoning and her decisions.

Dick Jeffrey: Degrees of belief are *estimates* of truth-values.
[Joyce, 1998]: Right, and full beliefs are *guesses* at truth-values.

- *Estimates* are evaluated on a gradational ("closeness counts") scale. They are distinguished from *guesses*, for which a categorical ("miss is a good as a mile") scale is appropriate.
- If credences satisfy the laws of probability, then estimation involves forming *expectations*, and credences are expectations of truth-values.

ALETHIC PROBABILISTIC EPISTEMOLOGY?

The core normative doctrine of probabilism, the *coherence requirement*, is the claim that rational credences must obey the laws of probability.

[Joyce, 1998] sought to justify this requirement on purely epistemic (not pragmatic) grounds by showing:

- The accuracy of credences can be measured on a gradational scale
- Gradational accuracy should play the role in probabilistic epistemology that truth plays in traditional epistemology.
- There are general requirements that any adequate measure of accuracy for credences must meet.
- Probabilistic coherence promotes accuracy. On any adequate measure of accuracy for partial beliefs, each incoherent system of credences will have at least one *uniformly* more accurate coherent counterpart, i.e., one that is strictly more accurate in every possible world.

Goal of Workshops: To explain this argument, consider a number of objections that have been raised against it, and improve upon it.

FORMAL FRAMEWORK

Our believer has opinions about proposition in some finite, ordered set of propositions $\boldsymbol{X} = \langle x_1, x_2,\ldots, x_N\rangle$.[i]

$\boldsymbol{B}$ is the family of all credence functions for propositions in $\boldsymbol{X}$. Elements of $\boldsymbol{B}$ are vectors in $\Re^N$, $\boldsymbol{b} = \langle b_1, b_2,\ldots, b_N\rangle = \langle \boldsymbol{b}(x_1), \boldsymbol{b}(x_2),\ldots, \boldsymbol{b}(x_N)\rangle$. Assume wlog that each $b_j \in [0, 1]$.

$\boldsymbol{V} = \{\boldsymbol{\omega}^1, \boldsymbol{\omega}^2,\ldots, \boldsymbol{\omega}^K\}$ is the subset of $\boldsymbol{B}$ that contains all logically consistent truth-value assignments to elements of $\boldsymbol{X}$. Each element of $\boldsymbol{V}$ is binary sequence $\boldsymbol{\omega} = \langle \omega_1, \omega_2,\ldots, \omega_N\rangle$ in which 0 = false, 1 = true.

The set of all coherent probability assignments over propositions in $\boldsymbol{X}$ is the *convex hull* $\boldsymbol{V}^+$ of $\boldsymbol{V}$, the set of all points in $\Re^N$ that are mixtures of elements in $\boldsymbol{V}$: $\boldsymbol{V}^+ = \{\Sigma_j \lambda_j \boldsymbol{\omega}^j : \Sigma_j \lambda_j = 1 \text{ and } j = 1, 2,\ldots, K\}$.

Our Question: How do we measure the accuracy of a system of credences $\boldsymbol{b} \in \boldsymbol{B}$ relative to a truth-value assignment $\boldsymbol{\omega} \in \boldsymbol{V}$?

MEASURES OF INACCURACY

A *measure of gradational inaccuracy* is a function $\boldsymbol{I}: \boldsymbol{B} \times \boldsymbol{V} \to \Re$ that for each $\boldsymbol{b}$ and $\boldsymbol{\omega}$ specifies $\boldsymbol{b}$'s degree of inaccuracy given the truth-values that $\boldsymbol{\omega}$ assigns to propositions in $\boldsymbol{X}$. We shall assume that $\boldsymbol{I}$ is continuous and that $\boldsymbol{I}(\boldsymbol{b}, \boldsymbol{\omega}) \geq 0$ with equality iff $\boldsymbol{b} = \boldsymbol{\omega}$.

Example ("ℓ_p-measures"): $\ell_p(\boldsymbol{b}, \boldsymbol{\omega}) = (\Sigma_j |b_j - \omega_j|^p)^{1/p}$ with $p > 0$.

Equivalent for our purposes, $\ell_p(\boldsymbol{b}, \boldsymbol{\omega}) = \Sigma_j\, 1/N\, |b_j - \omega_j|^p$.

- $\ell_1(\boldsymbol{b}, \boldsymbol{\omega}) = \Sigma_j\, 1/N\, |b_j - \omega_j|$ is the *absolute value measure*. [Maher, 2002] defends it as the correct measure.
- $\ell_2(\boldsymbol{b}, \boldsymbol{\omega}) = \Sigma_j\, 1/N |b_j - \omega_j|^2$ is the *Brier score*, which is often used to gauge the accuracy of probabilistic weather forecasts.

Example (weighted-ℓ_p): $\ell_{p,\lambda}(\boldsymbol{b}, \boldsymbol{\omega}) = \Sigma_j \lambda_j |b_j - \omega_j|^p$ with $\lambda_j \geq 0$ and $\Sigma_j \lambda_j = 1$.

- Weights indicate the relative importance of being accurate with respect to various propositions.
- The ordinary ℓ_p-measures measures are obtained when $\lambda_j = 1/N$.
- $\ell_{2,\lambda}(b, \omega)$ gives the general form of a *quadratic loss function*.

EXAMPLE: CALIBRATION AND DISCRIMINATION

The *truth-frequency* of $\boldsymbol{X}$ at ω is the proportion of the x_j true at $\boldsymbol{\omega}$. $freq(\boldsymbol{X}, \boldsymbol{\omega}) = [\boldsymbol{\omega}(x_1) +\ldots + \boldsymbol{\omega}(x_N)]/N$.

What's the right way to *estimate* $freq(\boldsymbol{X}, \boldsymbol{\omega})$ given one's credences?

De Finetti: A person with a coherent subjective probabilities will use her *average* credence for the x_j as her estimate of $\boldsymbol{X}$'s truth-frequency, so that $Est(freq(\boldsymbol{X})) = [\boldsymbol{b}(x_1) +\ldots+ \boldsymbol{b}(x_n)]/N$.

Special Case: $Est(freq(\boldsymbol{X})) = p$ when $\forall j\ \boldsymbol{b}(x_j) = p$.

This gets at something deep. What can it mean to assign degree of belief p to x if not to think something like, "Propositions like x (with respect to probable truth) are true p proportion of the time"?

CALIBRATION PRINCIPLE. A person (coherent or not) who assigns credence p to each x_j should use p as her estimate of $freq(\boldsymbol{X})$.

THE CALIBRATION INDEX

If we divide $\boldsymbol{X}$ into (finitely many) subsets $\boldsymbol{X}_p = \{x \in \boldsymbol{X}: \boldsymbol{b}(x) = p\}$, CP says that a rational person will use p to estimate the proportion of truths in $\boldsymbol{X}_p$.

It is then natural to measure the "fit" between redences and the facts at ω using the *calibration index*:
$\boldsymbol{Cal}(\boldsymbol{X}, \omega) = \Sigma_p (n_p/N)\,[freq(\boldsymbol{X}_p, \omega) - p]^2$, n_p = cardinality of $\boldsymbol{X}_p$, $freq(\boldsymbol{X}_p, \omega)$ = proportion of truths in $\boldsymbol{X}_p$ at ω.

P is *perfectly calibrated* when $Cal(\boldsymbol{X}, \omega) = 0$. Here half the elements of $\boldsymbol{X}$ 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.

P is *maximally uncalibrated* when $Cal(\boldsymbol{X}, \omega) = 1$. Here all the truths in $\boldsymbol{X}$ are assigned a value of 0, and all the falsehoods are assigned a value of 1.

THE DISCRIMINATION INDEX: $\boldsymbol{Dis}(\boldsymbol{X}, \omega) = 1/N\ \Sigma_p \Sigma_{x \in \boldsymbol{X}_p} [freq(\boldsymbol{X}_p, \omega) - \omega(x)]^2$

This measures the degree to which values of $\boldsymbol{b}$ sort elements of $\boldsymbol{X}$ into classes that are homogenous with respect to truth-value.

Perfect discrimination (= 0) means that every $\boldsymbol{X}_p$ contains only truths or only falsehoods. Perfect nondiscrimination (= 1/4) is attained when every $\boldsymbol{X}_p$ contains exactly as many truths as falsehoods.

Fact (Murphy): The Brier score is a straight sum of the calibration and discrimination indices,

$$\ell_2(\boldsymbol{b}, \omega) = 1/N\ \Sigma_{x \in \boldsymbol{X}} [\boldsymbol{b}(x) - \omega(x)]^2 = \boldsymbol{Dis}(\boldsymbol{X}, \omega) + \boldsymbol{Cal}(\boldsymbol{X}, \omega)$$

This shows that the Brier score incorporates two quantities that are clearly relevant to the accuracy of probabilistic estimates of truth-values.

THE ARGUMENT OF [JOYCE, 1998]

[Joyce, 1998] imposes a set of constraints meant to ensure that $\boldsymbol{I}$-functions measure the sort of accuracy that is appropriate for credences. (These hold for all $\omega, \omega^* \in \boldsymbol{V}$, all $\boldsymbol{b}, \boldsymbol{c} \in \boldsymbol{B}$, and all $x, y \in \boldsymbol{X}$.)

DOMINANCE. If $\boldsymbol{b}(y) = \boldsymbol{c}(y)$ for all $y \neq x$, then $I(\boldsymbol{b}, \omega) > I(\boldsymbol{c}, \omega)$ iff $\mathbf{c}(x)$ is closer than $\boldsymbol{b}(x)$ is $\omega(x)$.

NORMALITY. If $|\omega(x) - \boldsymbol{b}(x)| = |\omega^*(x) - \boldsymbol{c}(x)|$ for each $x \in \boldsymbol{X}$, then $\boldsymbol{I}(\boldsymbol{b}, \omega) = \boldsymbol{I}(\boldsymbol{c}, \omega^*)$.

WEAK CONVEXITY. If $\boldsymbol{I}(\boldsymbol{b}, \omega) = \boldsymbol{I}(\boldsymbol{c}, \omega)$, then $\boldsymbol{I}(\frac{1}{2}\boldsymbol{b} + \frac{1}{2}\boldsymbol{c}, \omega) \geq \boldsymbol{I}(\boldsymbol{m}, \omega)$ with equality only if $\boldsymbol{b} = \boldsymbol{c}$.

SYMMETRY. If $\boldsymbol{I}(\boldsymbol{b}, \omega) = \boldsymbol{I}(\mathbf{c}, \omega)$, then $\boldsymbol{I}(\lambda\boldsymbol{b} + (1-\lambda)\boldsymbol{c}, \omega) = \boldsymbol{I}((1-\lambda)\boldsymbol{b} + \lambda\boldsymbol{c}, \omega)$ for any $0 \leq \lambda \leq 1$.

MAIN THEOREM. If $\boldsymbol{I}$ satisfies these conditions, then for each incoherent set of credences $\boldsymbol{b} \in \boldsymbol{B} \sim \boldsymbol{V}^{+}$ there is a coherent $\boldsymbol{b}^* \in \boldsymbol{V}^{+}$ that is *uniformly* more accurate in the sense that $\boldsymbol{I}(\boldsymbol{b}, \omega) > \boldsymbol{I}(\boldsymbol{b}^*, \omega)$ for every $\omega \in \boldsymbol{V}^{+}$. Moreover, if $\boldsymbol{b}^* \in \boldsymbol{V}^{+}$, no set of credences in $\boldsymbol{V}$ is uniformly more accurate than $\boldsymbol{b}^*$.[ii]

DOMINANCE

- No serious question has been raised about Dominance.
- It is an instance of a more general requirement that one might reasonably impose on inaccuracy measures.

  SEPARABILITY.[iii] Let $\boldsymbol{Y}$ be a subset of $\boldsymbol{X}$ and $\boldsymbol{b}, \boldsymbol{b}^*, \boldsymbol{c}, \boldsymbol{c}^*$ be elements of $\boldsymbol{B}$ such that

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

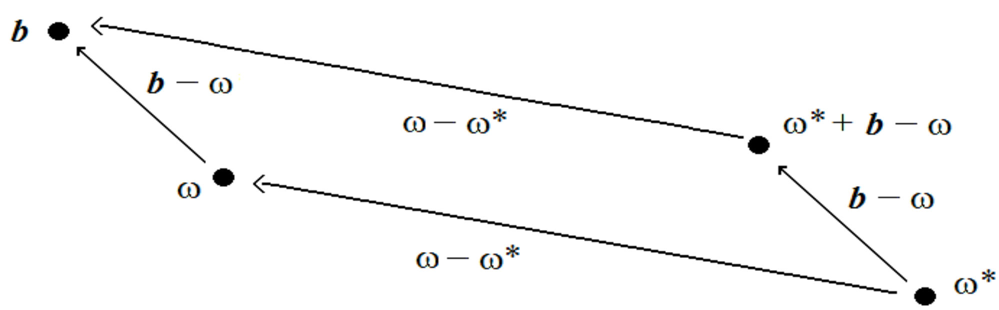
  Then $\boldsymbol{I}(\boldsymbol{b}, \omega) > \boldsymbol{I}(\boldsymbol{c}, \omega)$ if and only if $\boldsymbol{I}(\boldsymbol{b}^*, \omega) > \boldsymbol{I}(\boldsymbol{c}^*, \omega)$.

- This ensures that I has the additive form $\boldsymbol{I}(\boldsymbol{b}, \omega) = \Sigma_j \boldsymbol{g}_j(b_j, \omega_j)$ where each component function $\boldsymbol{g}_j$ measures the inaccuracy of b_j when x_j has truth-value ω_j. The $\boldsymbol{g}_j$ need *not* have the same functional form!
- Separability is an anti-holism constraint. It says that the accuracy of the credences about the propositions in $\boldsymbol{Y}$ does not depend on the credences or truth-values for propositions outside $\boldsymbol{Y}$.

NORMALITY

This strong requirement was not adequately justified in [Joyce, 1998].

It makes $\boldsymbol{I}$ *translation invariant*: $\boldsymbol{I}(\omega^* + \boldsymbol{b} - \omega, \omega^*) = \boldsymbol{I}(\boldsymbol{b}, \omega)$.

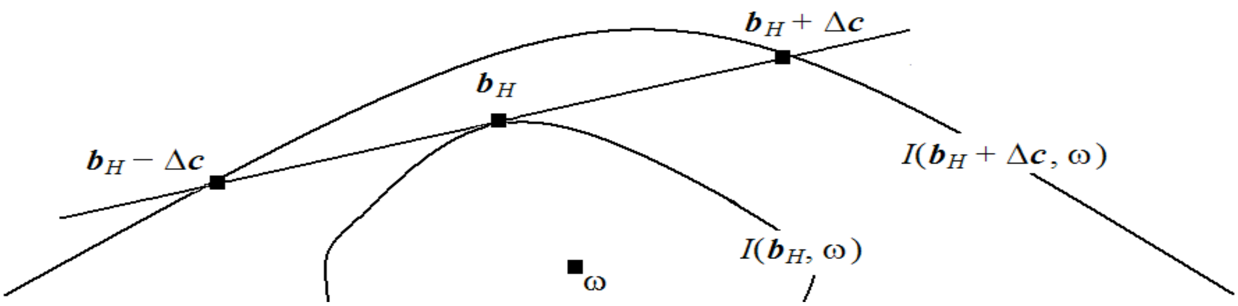


- In the presence of Separability, Normality forces each component $\boldsymbol{g}_j$ to have the form $\boldsymbol{g}_j(b_j, \omega_j) = \boldsymbol{f}_j(|b_j - \omega_j|)$ for some real-valued function f_j.
- Overestimating a proposition's truth-value by an increment ε is counted as exactly as inaccurate as underestimating its truth-value by ε.
- Allan Gibbard has argued that there is no reason to prefer a measure in which $\boldsymbol{g}_j(b_j, \omega_j) = \boldsymbol{f}_j(|b_j - \omega_j|)$ to one in which the $\boldsymbol{g}_j$ are allowed to be any arbitrary *proper scoring rules* (more below).

WEAK CONVEXITY + SYMMETRY

Consequence-1. For each $\boldsymbol{\omega} \in V$ and each hyperplane $\boldsymbol{H}$ in $\Re^N$, there is a unique $\boldsymbol{b}_H \in \boldsymbol{H}$ such that (i) $\boldsymbol{I}(\boldsymbol{b}, \boldsymbol{\omega})$ assumes its unique minimum on $\boldsymbol{H}$ at $\boldsymbol{b}_H$, and (ii) for any line $\boldsymbol{L}_c = \{\boldsymbol{b}_H + \Delta\boldsymbol{c}: \Delta \in \Re\}$ through $\boldsymbol{b}_H$ and confined to $\boldsymbol{H}$ one has $\boldsymbol{I}(\boldsymbol{b}_H + \Delta\boldsymbol{c}, \omega) = \boldsymbol{I}(\boldsymbol{b}_H - \Delta\boldsymbol{c}, \omega)$ for all $\Delta \in \Re$.

Consequence-2. For each $\boldsymbol{\omega} \in V$ and each point $\boldsymbol{b} \neq \boldsymbol{\omega}$, there is a unique hyperplane $\boldsymbol{H_b}$ of dimension $N - 1$ in $\Re^N$ such that (i) $I(\boldsymbol{b}, \boldsymbol{\omega})$ assumes its unique minimum on $\boldsymbol{H_b}$ at $\boldsymbol{b}$, and (ii) for any line $\boldsymbol{L}_C = \{\boldsymbol{b} + \Delta\boldsymbol{c}: \Delta \in \Re\}$, through $\boldsymbol{b}$ and confined to $\boldsymbol{H_b}$, $I(\boldsymbol{b} + \Delta\boldsymbol{c}, \boldsymbol{\omega}) = I(\boldsymbol{b} - \Delta\boldsymbol{c}, \boldsymbol{\omega})$ for all $\Delta \in \Re$.



SCORECARD

	Dominance	Normality	Convexity	Symmetry
$\ell_{p,\lambda}(b, \omega)$, $1 \geq p > 0$	YES	YES	NO	NO
$\ell_{p,\lambda}(b, \omega)$, $2 > p > 1$	YES	YES	YES	NO
$\ell_{2,\lambda}(b, \omega)$ Special Case: Brier	YES	YES	YES	YES
$\ell_{p,\lambda}(b, \omega)$, $p > 2$	YES	YES	YES	NO
Calibration	NO	YES	NO	NO

MAHER'S OBJECTIONS TO CONVEXITY AND SYMMETRY

[Maher, 2002] argues that the justifications in [Joyce, 1998] are weak (right, I'll do better), and that Convexity and Symmetry are implausible because $\ell_1(\boldsymbol{b}, \boldsymbol{\omega})$, which violates them, is a plausible measure of accuracy for credences (wrong, let's see why).

Maher offers two considerations to support the absolute value measure:

1. Since there are cases in which it is reasonable to use sums of absolute deviations to measure accuracy – e.g., when a teacher sets final grades by taking straight *averages* of scores on three equally important exams – it is plausible that sums of absolute deviations between credences and truth-values measure the accuracy of partial beliefs.

Response: Even if we use such a measure for grading, it does not follow that it is an appropriate way to measure the accuracy of other things.

- To determine an archer's accuracy we use a target of concentric circles rather than one of concentric squares aligned up/down and left/right. On the square target, an archer whose inaccuracy tends to be confined along one of the two principle axes is penalized less than one who is inaccurate along the diagonal. When evaluating archers there are no "preferred" directions: a deviation along any line running through the bull's eye is counts as much as a deviation along any other such line.
- Why should credence accuracy be measured like academic performance rather than archery? What are there relevant symmetries?

2. "It is natural to measure the inaccuracy of $\boldsymbol{b}$ with respect to the proposition x in possible world $\boldsymbol{\omega}$ by $|\boldsymbol{b}(x) - \boldsymbol{\omega}(x)|$. It is also natural to take the total inaccuracy of $\boldsymbol{b}$ to be the sum of its inaccuracies with respect to each proposition." (Maher 2002, p. 77)

Response: A complicated argument that begins as a question-begging *ad hominem*, but ends up providing an epistemologically illuminating, non-circular rationale for Convexity and Symmetry (assuming Separability and Normality) as well as a justification for using quadratic loss functions as measures of accuracy for degrees of belief.

OVERVIEW OF THE ARGUMENT

Step-1: Using $|\boldsymbol{b}(x) - \boldsymbol{\omega}(x)|$ to measure the inaccuracy of $\boldsymbol{b}(x)$ makes it impossible for probabilists to allow credences that are not either maximally opinionated ($\boldsymbol{b}(x) = 0$ or $= 1$) or noncommittal ($\boldsymbol{b}(x) = ½$).

This is because (a) probabilists should advocate the maximization of *expected* accuracy, (b) probabilists ought not rule out credences other than 0, ½, 1 by fiat, but (c) this is precisely what happens if expected accuracy maximization is combined with the use of $|\boldsymbol{b}(x) - \boldsymbol{\omega}(x)|$ to measure the inaccuracy of an individual credences.

Step-2. The argument in Step-1 generalizes to all improper scoring rules, which leaves the quadratic loss functions, $\ell_{2,\lambda}(\boldsymbol{b}, \boldsymbol{\omega})$, as the only Normal measures of accuracy that probabilists can endorse.

Step-3. One can avoid the question-begging appeal to expected accuracy maximization and proper scoring rules by (a) noting that many measures encourage "Jamesian" or "Cliffordian" tendencies that prevent believers from holding manifestly reasonable opinions, and (b) showing that, among Normal measures, only the quadratic loss functions balance off the costs of error with the benefits of strongly believing truths in such a way as to avoid these "Jamesian" or "Cliffordian" tendencies.

STEP-1. EXTREMENESS IN PURSUIT OF ACCURACY IS NO VIRTUE

Definition. A system of credences $\boldsymbol{b}$ is *self-deprecating with respect to* $\boldsymbol{I}$ when anyone who holds $\boldsymbol{b}$ is committed to estimating that some specific alternative set of credences $\boldsymbol{c}$ is strictly more accurate than $\boldsymbol{b}$. $\boldsymbol{b}$ is *self-deprecating* simpliciter when it is *self-deprecating* with respect to the correct measure of inaccuracy.

Premise. *Self-deprecating* credences are epistemically defective.

> Rationale: Such credences are inherently unstable. A person will not be comfortable holding them since she is thereby is committed to judging that some other system of credences provides a more accurate picture of the world than her own.[iv]

Probabilist Premise. A coherent believer's estimate of a quantity is her expected value for that quantity. In particular, if $\boldsymbol{b}$ is her system of credences, then her estimated accuracy for any other system $\boldsymbol{c}$ is $\boldsymbol{E_b}(\boldsymbol{I}(\boldsymbol{c})) = \Sigma_\omega\, \boldsymbol{b}(\omega)\boldsymbol{I}(\boldsymbol{c}, \omega)$. Thus, for a probabilist, $\boldsymbol{b}$ is self-deprecating, and so defective, whenever $\Sigma_\omega\, \boldsymbol{b}(\omega)\boldsymbol{I}(\boldsymbol{b}, \omega) > \Sigma_\omega\, \boldsymbol{b}(\omega)\boldsymbol{I}(\boldsymbol{c}, \omega)$ for some $\boldsymbol{c}$ (where $\boldsymbol{I}$ is the correct measure of inaccuracy).

Premise. If adopting $\boldsymbol{I}$ as our measure of inaccuracy makes certain clearly non-defective credences self-deprecating, then $\boldsymbol{I}$ is *not* a plausible measure of inaccuracy for partial beliefs.

Mathematical Fact. $\ell_1(\boldsymbol{b}, \omega) = \Sigma_j\, 1/N\, |b_j - \omega_j|$ makes systems of coherent credences that assigns values other than 0, ½, 1 self-deprecating. A coherent agent whose degree of belief for x is $\boldsymbol{b}(x)$ will estimate the accuracy of a credence assignment $\boldsymbol{c}$ with respect to x as $\boldsymbol{E_b}[\boldsymbol{I}(\boldsymbol{c}(x))] = \boldsymbol{b}(x)\,(1 - \boldsymbol{c}(x)) + (1 - \boldsymbol{b}(x))\,\boldsymbol{c}(x)$. This assumes its minimum at $\boldsymbol{c}(x) = 1$ when $\boldsymbol{b}(x) > ½$, at $\boldsymbol{c}(x) = ½$ when $\boldsymbol{b}(x) = ½$, and at $\boldsymbol{c}(x) = 0$ when $\boldsymbol{b}(x) < ½$. So, any degrees of belief for x, other than 0, ½, 1, are self-deprecating.

Thus, if $\ell_1(\boldsymbol{b}, \omega) = \Sigma_j\, 1/N\, |b_j - \omega_j|$ measures of inaccuracy, then a credence assignment of, say, 1/3 *to any proposition* is epistemically defective.

Premise. Some such assignments are clearly non-defective.

Conclusion. The absolute value measure must go. It is unacceptable because it encourages believers to adopt extreme opinions.

GENERALIZATION TO IMPROPER SCORING RULES

This argument can be mounted against any measure $\boldsymbol{I}(\boldsymbol{b}, \omega) = \Sigma_j\, \boldsymbol{g}_j(b_j, \omega_j)$ where $\boldsymbol{g}_j$ is an *improper* scoring rule.

> *Definition*. $\boldsymbol{g}(b, \omega)$ is a *proper scoring rule* iff, for all $b \in [0, 1]$, $\boldsymbol{E_b}[\boldsymbol{g}(c)] = b\, \boldsymbol{g}(c, 1) + (1 - b)\, \boldsymbol{g}(c, 0)$ assumes its minimum at $\boldsymbol{b}(x) = \boldsymbol{c}(x)$.
>
> Proper scoring rules were initially introduced as a way of finding incentives that would elicit the credences of coherent agents.

Premise. For any rational number $b \in [0, 1]$, there are circumstances in which it clearly rational to believe a proposition to degree b.

Mathematical Fact. If inaccuracy is given by $\boldsymbol{I}(\boldsymbol{b}, \omega) = \Sigma_j\, \boldsymbol{g}_j(b_j, \omega_j)$ with $\boldsymbol{g}_j$ improper, then there will always be at least one rational number b such that $\boldsymbol{E_b}[g_j(c)]$ assumes its minimum at a point $c \neq b$.

Conclusion. Any function that assigns credence b to any proposition will necessarily be self-deprecating with respect to $\boldsymbol{I}$. (Note: any proposition can be placed as the jth member of some set $\boldsymbol{X} = \langle x_1, x_2, \ldots, x_N\rangle$.)

EXAMPLE: THE ℓ_p MEASURES.

When $p \neq 1$, the derivative of $\boldsymbol{E}_b(c) = b(1 - c)^p + (1 - b)c^p$ has a critical point at

$$c = b^{1/(p-1)} / [(1-b)^{1/(p-1)} + b^{1/(p-1)}].$$

This is a minimum when $p > 1$, and a maximum when $p < 1$.

The general situation is this, where $c^{\#}$ is the minimum value of $\boldsymbol{E}_b(c)$:

	$b = 0$	$0 < b < ½$	$b = ½$	$½ < b < 1$	$B = 1$
$1 \geq p > 0$	$c^{\#} = 0$	$c^{\#} = 0$	$c^{\#} = ½$	$c^{\#} = 1$	$C^{\#} = 1$
$2 > p > 1$	$c^{\#} = 0$	$c^{\#} < b < ½$	$c^{\#} = ½$	$½ < b < c^{\#}$	$C^{\#} = 1$
$p = 2$	$c^{\#} = 0$	$c^{\#} = b$	$c^{\#} = ½$	$c^{\#} = b$	$C^{\#} = 1$
$p > 2$	$c^{\#} = 0$	$b < c^{\#} < ½$	$c^{\#} = ½$	$½ < c^{\#} < b$	$C^{\#} = 1$

Row-1 ("Extreme Jamesian" functions). These are not convex. They make it impossible for a believer to regard anything but 0, ½, 1 as the credence of highest expected accuracy.

- Even worse, they lead coherent believers to expect that *inconsistent* credences are more accurate than their own. Example: A fair six-sided die is about to be tossed. Two of its faces are red, two blue, and two green. Then, in terms of ℓ_1 or $\ell_{1/2}$ accuracy the uniform probability assignment

  $$\boldsymbol{b}(\text{RED}) = \boldsymbol{b}(\text{BLUE}) = \boldsymbol{b}(\text{GREEN}) = 1/3,\ \boldsymbol{b}(\sim\text{RED}) = \boldsymbol{b}(\sim\text{BLUE}) = \boldsymbol{b}(\sim\text{GREEN}) = 2/3$$

  is expected to be strictly less accurate than the *logically inconsistent* assignment

  $$\boldsymbol{b}(\text{RED}) = \boldsymbol{b}(\text{BLUE}) = \boldsymbol{b}(\text{GREEN}) = 0,\ \boldsymbol{b}(\sim\text{RED}) = \boldsymbol{b}(\sim\text{BLUE}) = \boldsymbol{b}(\sim\text{GREEN}) = 1.$$

Row-2 ("Jamesian" functions). Since these are less convex than ℓ_2 they place more weight on the commandment to believe the truth than on the commandment to avoid error. As a result, they encourage believers toward extreme opinions.

> Someone who believes x to degree 2/3 and uses $\ell_{3/2}$ will see 4/5 as the most accurate credence. If he revises his credence up to 4/5, he will then expect 16/17 to be most accurate. The process will not stabilize until he ends up believing x to degree 1. In general, all beliefs that fall short of certainty are pushed toward 1 or 0 by the $\ell_{3/2}$ measure.

Row-4 ("Cliffordian" functions). Since these are more convex than ℓ_2 they place more weight on the commandment to avoid error than on the commandment to believe the truth. As a result, they encourage believers toward more cautious, less extreme opinions.

> Someone who believes x to degree 4/5 and uses ℓ_3 will see 2/3 as the most accurate credence. If he revises his credence down to 2/3, he will then expect $\sqrt{2}/(1 + \sqrt{2}) \approx 0.586$ to be most accurate. If iterated continuously, the process will not stabilize until he ends up believing x to degree ½. In general, all beliefs that fall short of certainty are pushed toward ½ by the ℓ_3 measure.

Row-3 (Quadratic Loss). ℓ_2 is the only proper ℓ_p measure. It alone strikes the right balance between avoiding the error of strongly believing falsehoods and securing the benefits of strongly believing truths.

This balance is reflected in the symmetry $\boldsymbol{E}_b(\boldsymbol{I}(b + \Delta)) = \boldsymbol{E}_b(\boldsymbol{I}(b - \Delta))$, which holds for all small increments Δ.

Since $\boldsymbol{E}_b(\boldsymbol{I}(b + \Delta)) > \boldsymbol{E}_b(\boldsymbol{I}(b))$, this says that the believer expects equal deviations from her credences in either direction to increase her inaccuracy by the same amount.

The expected benefits of seeking to strongly believe x if it is true (by raising one's credence by Δ) are thus precisely offset by the benefits of seeking to strongly disbelieve x if it is false (by lowering one's credence by Δ).

Theorem (Savage): $\mathbf{g}(b, \omega) = |b - \omega|^2$ is the only proper scoring rule (up to a positive linear transformation) such that either (a) $\boldsymbol{g}(b, \omega) = \boldsymbol{f}(|b - \omega|)$ for some real function $\boldsymbol{f}$, or (b) $\mathbf{g}(b, \omega) = \mathbf{g}(\omega, b)$.

It follows that the quadratic loss functions $\ell_{2,\lambda}(\boldsymbol{b}, \omega) = \Sigma_j \lambda_j |b_j - \omega_j|^2$ are the only plausible measures of epistemic accuracy that someone who accepts the "Probabilist Premise" can endorse! Unfortunately, the "Probabilist Premise" begs the question!

STEP-3. A MORE GENERAL ARGUMENT

Fortunately, we can recapitulate the substance of our argument without assuming coherence or invoking the notion of expectation at all. The contrast between Jamesian and Cliffordian measures remains even when we restrict ourselves to pure dominance reasoning.

Key Premise 1. Anyone who holds credences $\boldsymbol{b}$ is committed to estimating that an alternative set of credences $\boldsymbol{c}$ is strictly more accurate than $\boldsymbol{b}$ when $\boldsymbol{c}$ is uniformly more accurate than $\boldsymbol{b}$, i.e., when $I(\boldsymbol{b}, \boldsymbol{\omega}) > I(\boldsymbol{c}, \boldsymbol{\omega})$ for all $\boldsymbol{\omega}$. Any such $\boldsymbol{b}$ is self-deprecating, and so epistemically defective.

Key Premise 2. In at some situations a uniform probability assignment, in which $\boldsymbol{b}(x_1) = \ldots = \boldsymbol{b}(x_N) = 1/N$ for $\{x_1, x_2, \ldots, x_N\}$, is non-defective.

Disclaimer. I will only present the argument for the case of $N = 3$, and will focus only on the ℓ_p measures. Both restrictions can be lifted.

FORMAL SETUP

$\{x, y, z\}$ is a partition. Write its associated algebra as a sequence $\mathcal{A} = \langle \mathsf{T},\ x \vee y,\ x \vee z,\ y \vee z,\ x,\ y,\ z,\ \mathsf{F} \rangle$.

- $\boldsymbol{\omega}_x = \langle 1, 1, 1, 0, 1, 0, 0, 0 \rangle$ is the $x = 1$ truth-value assignment.
- $\boldsymbol{b}_t = \langle 1, 1 - t, 1 - t, 1 - t, t, t, t, 0\rangle$ with $t > 0$ is the "t-flat" credence.
- $\boldsymbol{b}_{1/3} = \langle 1, 2/3, 2/3, 2/3, 1/3, 1/3, 1/3, 0\rangle$ is the *uniform* credence.

For any $p > 0$, the ℓ_p-inaccuracy of $\boldsymbol{b}$ at $\boldsymbol{\omega}_x$ is

$$\ell_p(\boldsymbol{b}, \boldsymbol{\omega}_x) = 1/8\ [(1 - \boldsymbol{b}(x \vee y))^p + (1 - \boldsymbol{b}(x \vee z))^p + \boldsymbol{b}(y \vee z)^p + (1 - \boldsymbol{b}(x))^p + \boldsymbol{b}(y)^p + \boldsymbol{b}(z)^p]$$

where $1 \geq \boldsymbol{b}(x_j) \geq 0$ for all j, $\boldsymbol{b}(\mathsf{T}) = 1$, and $\boldsymbol{b}(\mathsf{F}) = 0$.

In general, $\ell_p(\boldsymbol{b}_t, \boldsymbol{\omega}_x) = \ell_p(\boldsymbol{b}_t, \boldsymbol{\omega}_y) = \ell_p(\boldsymbol{b}_t, \boldsymbol{\omega}_z) = ½t^p + ¼(1 - t)^p$. So, the uniform distribution $\boldsymbol{b}_{1/3}$ is self-deprecating if there exists a t such that $\ell_p(\boldsymbol{b}_{1/3}, \boldsymbol{\omega}_x) > \ell_p(\boldsymbol{b}_t, \boldsymbol{\omega}_x)$. When $p \neq 2$ there is always such a t!

THE ABSOLUTE VALUE MEASURE AND ITS CONCAVE BRETHREN

$\boldsymbol{p = 1}$: Since $\ell_1(\boldsymbol{b}_t, \boldsymbol{\omega}_x) = ¼(t + 1)$ is monotonically increasing, $t = 0$ minimizes ℓ_1-inaccuracy among the $\boldsymbol{b}_t$. In particular, $\ell_1(\boldsymbol{b}_0, \boldsymbol{\omega}_x) = 1/4 < \ell_1(\boldsymbol{b}_{1/3}, \boldsymbol{\omega}_x) = 1/3$. Thus, the uniform distribution $\boldsymbol{b}_{1/3}$ is uniformly less accurate than the *logically inconsistent* $\boldsymbol{b}_1$.

> More generally, according to ℓ_1, $\boldsymbol{b}_1$ is uniformly more accurate than any coherent probability assignment $\boldsymbol{b}$ for which $\boldsymbol{b}(x_1), \boldsymbol{b}(x_2), \boldsymbol{b}(x_3) < ½$. By Key Premise 1, this means that all such credences will be self-deprecating, hence defective. But, Key Premise 2, not all such credences are defective. Thus, ℓ_1 is unacceptable as a measure of inaccuracy for partial beliefs.

$\boldsymbol{p < 1}$: The argument generalizes. If $p < 1$, $\ell_p(\boldsymbol{b}_{1/3}, \boldsymbol{\omega}_x) > 1/3 > \ell_p(\boldsymbol{b}_0, \boldsymbol{\omega}_x) = 1/4$. (In fact, $\ell_p(\boldsymbol{b}_{1/3}, \boldsymbol{\omega}_x) > \ell_p(\boldsymbol{b}_0, \boldsymbol{\omega}_x)$ for all $p < 1.394$.)

General Point: $\boldsymbol{b}_0$ will be uniformly more accurate than $\boldsymbol{b}_{1/3}$ according to any accuracy function that violates Convexity. This justifies Convexity.

THE ASYMMETRIC CONVEX MEASURES

$\boldsymbol{1 < p < 2}$: $\ell_p(\boldsymbol{b}_t, \boldsymbol{\omega}_x)$ assumes its minimum at $t^\# = 1/(1 + 2^{1/(p-1)})$, and its minimum value is[v]

$$\ell_p(\boldsymbol{b}_{t\#}, \boldsymbol{\omega}_x) \;=\; 4[1/(1 + 2^{1/(p-1)})]^p + 2(2^{1/(p-1)}/(1 + 2^{1/(p-1)}))^p = \; 4(1 + 2^{1/(p-1)})^{1-p}$$

But, $1 + 2^{1/(p-1)} > 3$ when $1 < p < 2$, from which it follows that $t^\# < 1/3$. These measures are "Jamesian" in that they rank the uniform distribution as uniformly less accurate than $\boldsymbol{b}_{t\#}$, which is more extreme.

$\boldsymbol{2 < p}$: Again, $\ell_p(\boldsymbol{b}_t, \boldsymbol{\omega}_x)$ assumes its minimum at $t^\# = 1/(1 + 2^{1/(p-1)})$, but $1 + 2^{1/(p-1)} < 3$ when $2 < p$, from which it follows that $1/2 > t^\# > 1/3$. These measures are "Cliffordian" in that they rank the uniform distribution as uniformly less accurate than the more cautious $\boldsymbol{b}_{t\#}$.

Again, by Key Premises 1 and 2, we may conclude that any ℓp-measure with $p \neq 2$ is unacceptable as a measure of inaccuracy for partial beliefs.

MORALS AND HANDWAVING

Moral. Among the ℓ_p measures, only the symmetric measure, ℓ_2, renders the uniform credence function $\boldsymbol{b}_{1/3}$ non-defective. For it, $t^\# = 1/3$.

> *Generalization*. Among the weighted $\ell_{p,\lambda}$ measures, only the symmetric measure $\ell_{2,\lambda}$ renders the uniform credence function $\boldsymbol{b}_{1/3}$ non-defective.

Handwaving. If $\boldsymbol{I}(\boldsymbol{b}, \boldsymbol{\omega}) = \Sigma_j \boldsymbol{f}_j(|b_j - \omega_j|)$ has a non-quadratic component $\boldsymbol{f}_j$, then $\boldsymbol{f}_j$ will either be more or less convex than a quadratic function on some interval $[a, b] \subset [0, 1]$. If it is more convex, it will also be more convex than some $\ell_p(\boldsymbol{b}, \boldsymbol{\omega})$ with $p > 2$, in which case some credences in $[a, b]$ are subject to Cliffordian effects that push them toward the interval's interior. If $\boldsymbol{f}_j$ is less convex than a quadratic, it will also be less convex than some $\ell_p(\boldsymbol{b}, \boldsymbol{\omega})$ with $p < 2$, in which case some credences within $[a, b]$ are subject to Jamesian effects that push them toward the interval's endpoints. Either way, there will be rational credence assignments from $[a, b]$ that are self-deprecating, and hence defective, when attached to any proposition.

Hence, the $\ell_{2,\lambda}$ functions are the only plausible inaccuracy measures since they are the only ones that do not deem some rational credences defective.

AARON BRONFMAN'S OBJECTION

- At best, this argument shows that there is an infinite family of plausible inaccuracy measures $\ell_{2,\lambda}$ (one for each set of possible weights λ). It does not single out any one true measure of inaccuracy.
- [Joyce, 1998] assumed that a credence function is epistemically defective if it has a uniformly more accurate counterpart relative to *every* plausible inaccuracy measure. (Compare "supervaluation.")
- But, for each incoherent $\boldsymbol{b} \in \boldsymbol{B}$ there is a *different* uniformly more accurate coherent counterpart $\boldsymbol{b}_\lambda$ for each $\ell_{2,\lambda}$. Moreover, even though $\boldsymbol{b}_\lambda$ is always more accurate than $\boldsymbol{b}$ relative to $\ell_{2,\lambda}$ there will always be a λ^* and a ω such that, relative to ℓ_{2,λ^*}, $\boldsymbol{b}$ is more accurate than $\boldsymbol{b}_\lambda$ at ω.
- So, no single coherent system of credences $\boldsymbol{b}^*$ is unequivocally better than $\boldsymbol{b}$. We cannot replace $\boldsymbol{b}$ with $\boldsymbol{b}^*$ without running the risk of increasing inaccuracy with respect to some ω and some acceptable accuracy measure.
- Since this is also true of coherent systems of credences, the argument just given does not provide any normatively compelling rational for preferring coherent systems over incoherent ones.

REPLY TO BRONFMAN'S OBJECTION

I now concede that the "supervaluationist" approach of [Joyce, 1998] was misguided. In any fixed context, only one of the $\ell_{2,\lambda}$ can be the correct measure of inaccuracy. It is up to epistemologists to decide which it is.

Here are two broad approaches to the question:

Monist. The Brier score ℓ_2 is the unique correct measure of inaccuracy in all contexts because (a) it treats every proposition equally and in a way that is independent of facts about its content (other than its truth-value), and (b) it decomposes into a straight sum of calibration plus discrimination, two quantities that are tied in a clear way to the accuracy of partial beliefs.

Pluralist. In each context there will be a right way to weight propositions in $\boldsymbol{X}$. The weighting coefficients will depend on facts about the contents of the propositions other than their truth values (informativeness). It is up to epistemologists to decide which systems of weights should be used in a given context.

Interesting Question: Are there analogues of ***Cal*** and ***Dis*** for $\ell_{2,\lambda}$ measures other than the Brier score?

ALLAN GIBBARD'S OBJECTION

- Normality is unmotivated and implausible. Probabilists should be happy, in principle, with any inaccuracy measure $\boldsymbol{I}(\boldsymbol{b}, \omega) = \Sigma_j \boldsymbol{g}_j(b_j, \omega_j)$ where each $\boldsymbol{g}_j$ is a proper scoring rule.
- Many proper scoring rules do not have the form $\boldsymbol{g}(b_j, \omega_j) = \boldsymbol{f}(|b_j - \omega_j|)$. See [Schervish, 1989].
- So, unless the proof of the Main Theorem can be shown to go through without Normality (it cannot), then the Theorem fails to provide any compelling reason for requiring credences to be coherent.

Response: The plausibility of the claim that non-normal proper scoring rules may be used to measure epistemic accuracy is the result of conflating considerations of accuracy with other epistemic or pragmatic issues. Also, any acceptable measure of accuracy should have this symmetry property:

> *The cost, in terms of inaccuracy, of underestimating x's truth-value by a certain amount is identical to the costs of overestimating its truth-value by that same amount.*

This intuition motivates Normalilty. I concede that it is an "additivity" intuition, but of a weak sort.

A QUESTION FOR FUTURE RESEARCH: Is it possible, using some other proof strategy, to establish an analogue of the Main Theorem for inaccuracy measures of the form $\boldsymbol{I}(\boldsymbol{b}, \boldsymbol{\omega})$
$= \Sigma_j \boldsymbol{g}_j(b_j, \omega_j)$, where each $\boldsymbol{g}_j$ an arbitrary proper scoring rule? In other words, is it possible to provide a non-pragmatic vindication of probabilism that does not rely on Normality?

REFERENCES

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

Bronfman, A. "A Gap in Joyce's Argument for Probabilism," unpublished manuscript.

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

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

Joyce, J. (1998) "A Nonpragmatic Vindication of Probabilism," *Philosophy of Science* **65**, pp. 597-603.

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

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

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

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

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

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

---

[i] We place no restriction on this set, e.g., it need not be a Boolean algebra.

[ii] [Joyce, 1998] was not sufficiently explicit about this point. The assertion is true, but it is not obvious. A proof will be included in the paper that grows out of these workshops.

[iii] Readers will recognize this as an analogue of the "sure thing principle" in decision theory.

[iv] There are analogies here with versions Moore's paradox, in which a person entertains a proposition of the form '*p* but I believe ~*p*." In much the same way that an epistemically rational believer cannot believe such a proposition, an epistemically rational agent cannot hold one system of beliefs when she expects another to be more accurate.

[v] This is a monotonically decreasing function of p, for if $p > q$ then $2^{1/(q-1)} > 2^{1/(p-1)}$ from which it follows that $1/(1 + 2^{1/(p-1)}) > 1/(1 + 2^{1/(q-1)})$ and thus that $1/(1 + 2^{1/(p-1)})^{p-1} > 1/(1 + 2^{1/(q-1)})^{q-1}$.