Elimination Counterexamples: State of the Art

Branden Fitelson[1]

Philosophy & Religion @ Northeastern

branden@fitelson.org

[1]Draws heavily on the work of (and useful discussions with) Peter Lewis, Don Fallis, Ben Levinstein, Richard Pettigrew, and Bernhard Salow.

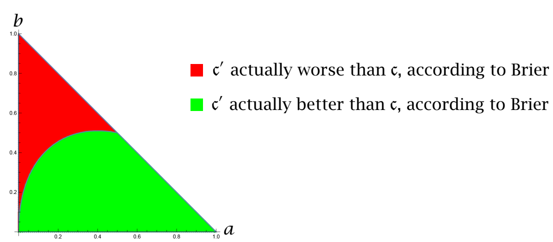
- According to *Veritistic Bayesian Epistemology* (VBE), the epistemic goodness (badness) of a credence function $\mathfrak{c}(\cdot)$ is given (entirely) by its *gradational (in)accuracy*.
- VBE-ers adopt *strictly proper scoring rules* ($\mathfrak{S}$) to measure the gradational inaccuracy of credence functions $\mathfrak{c}$ over sets of propositions $\mathcal{P}$ (we'll see various strictly proper $\mathfrak{S}$'s below).
- Suppose we are evaluating $\mathfrak{c}$ over a partition $\mathcal{P} = \{A, B, C\}$.
- Let A be true at the actual world (*i.e.*, w_A is actual); and, suppose one of the false hypotheses (wlog C) is *refuted*.
- Bayesians will think of these cases as ones in which the evidence $E = A \vee B$ is learned *via* conditionalization.
- Then, the prior $\mathfrak{c}$ and the posterior $\mathfrak{c}'$ will be given by:

$$\mathfrak{c} = \langle a, b, 1 - (a + b) \rangle$$

$$\mathfrak{c}' = \left\langle \frac{a}{a+b}, \frac{b}{a+b}, 0 \right\rangle$$

☞ Lewis & Fallis [2, 5] assume that *all elimination experiments should yield improvements in overall epistemic goodness*.

- In other words, $\mathfrak{c}'$ should always be (overall) *epistemically better* than $\mathfrak{c}$. They use this to argue against the Brier Score.
- According to Brier ($\mathfrak{B}$), there are *many* values of a, b such that $\mathfrak{c}'$ is *more inaccurate* than $\mathfrak{c}$ at w_A. These **ECEs** — in which $\mathfrak{B}(\mathfrak{c}, w_A) < \mathfrak{B}(\mathfrak{c}', w_A)$ — are shaded in red.



- Lewis, Fallis, and I [6] argue that such cases are ubiquitous in the history of science. In this talk, I will consider four responses to ECEs that can be made on behalf of VBE.

- Here are four possible responses to ECEs.
  - (I) **Bite the bullet**, *i.e.*, *accept* the Brier Score's verdicts (rankings) regarding (all) prior/posterior pairs $\langle \mathfrak{c}, \mathfrak{c}' \rangle$.
  - (II) **Appeal to considerations of "verisimilitude"** to help explain (away) the cases [1].
  - (III) **Appeal to a notion of "misleading evidence"** to help explain (away) the cases [10].
  - (IV) **Abandon the Brier Score**, in favor of another (strictly proper) scoring rule $\mathfrak{S}$, which avoids ECEs [13].
- I will set aside response (I). [Even if (I) could be made to work, I now think that (IV) is a much better response.]
- Regarding response (II), we have argued elsewhere that it faces a Dilemma: either it involves an inadequate account of "verisimilitude" [9], or it does not handle all ECEs [6].
- In today's talk, I will focus on responses (III) and (IV).

- Intuitively, evidence E is *misleading* if E *favors some false propositions over some true propositions*.
- Formally, we'll say E is misleading — according to $\mathfrak{c}(\cdot)$, and relative to a set $\mathcal{P}$ of propositions — iff there exists a pair of propositions $\langle t, f\rangle$ from $\mathcal{P}$ such that both
  (i) t is (actually) true; and f is (actually) false, *and*
  (ii) E favors f over t, according to $\mathfrak{c}(\cdot)$.
- To make this more precise, we will need a general (and uncontroversial) Bayesian explication of *favoring*.
- Happily, the following condition is generally accepted as a Bayesian sufficient condition for favoring [4, 3].

  Weak Law of Likelihood (WLL). E favors f over t, according to $\mathfrak{c}(\cdot)$ if both of the following obtain:
  (ii.1) $\mathfrak{c}(E \mid f) > \mathfrak{c}(E \mid t)$, and
  (ii.2) $\mathfrak{c}(E \mid \neg f) \leq \mathfrak{c}(E \mid \neg t)$.

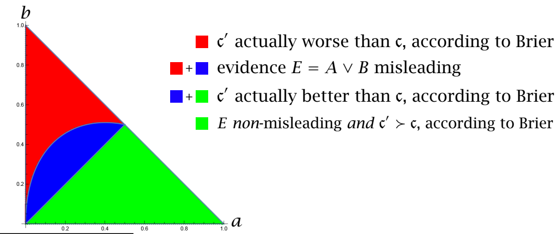
☞ Replacing (ii) with (ii.1) & (ii.2), yields our official Bayesian explication of "E is *misleading*, relative to $\mathfrak{c}(\cdot)$ and $\mathcal{P}$."

Theorem 1. The evidence in the elimination counterexamples is *never* misleading, according to *any* regular credence function, relative to the *partition* $\mathcal{P} = \{A, B, C\}$.

- Because $\Pr(A \vee B \mid C) = 0$, and $\Pr(A \vee B \mid A) = \Pr(A \vee B \mid B) = 1$.

Theorem 2. The evidence in the elimination counterexamples is *always* misleading, according to *every* regular credence function, relative to the *algebra* $\mathcal{A} = \{A, B, C, A \vee B, A \vee C, B \vee C\}$.[2]

- Because in all elimination counterexamples $A \vee B$ ($\neg C$) favors $B \vee C$ ($\neg A$) over $A \vee C$ ($\neg B$), according to *every* regular $\mathfrak{c}$.



[2]The Brier Score is *perspective invariant* here. Its $\mathcal{A}$-ranking agrees with its $\mathcal{P}$-ranking (from both perspectives, it implies $\mathfrak{c} \succ \mathfrak{c}'$ *iff* in the red region).

- I think Theorems 1 and 2 are useful for understanding various conflicting intuitions one might have regarding the elimination counterexamples, and the evidence they involve.
- If one focuses only on the partition (coarse-grained evaluation), then updating on the elimination evidence E seems (*clearly*) to lead us to be *better off, epistemically*.
  - *E rules-out a falsehood*; *and*, E is *not* misleading (rel. to $\mathcal{P}$).
- But, if one looks at the entire algebra (fine-grained evaluation), then there is some *ambiguity* in the epistemic ramifications of updating on the elimination evidence E.
  - *E rules-out a falsehood*; *but*, E *is* misleading (rel. to $\mathcal{A}$).
- This suggests an interesting reply on behalf of the (or, at least, one type of) defender of the Brier Score.

  ☞ **Idea**. The Brier Score correctly ranks prior/posterior pairs $\langle \mathfrak{c}, \mathfrak{c}'\rangle$ — in terms of *overall* epistemic goodness — ***if*** *the evidence yielding the posterior is non-misleading (rel. to $\mathcal{A}$)*.

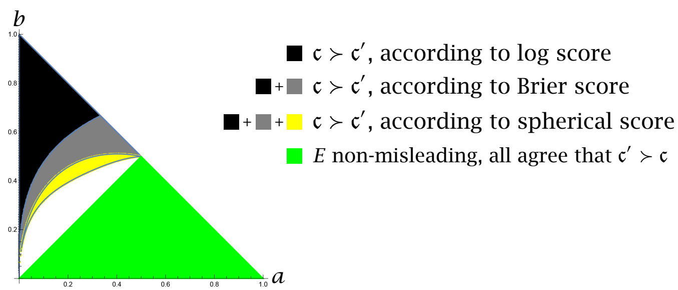
- *Formally*, this "non-misleading evidence caveat" allows a defender of the Brier Score to avoid the ECEs. But, *this stratagem seems (to me) to be unavailable to VBE.*
- VBE is committed to a specific form of *monism* about epistemic value — *epistemic value supervenes on accuracy*.

☞ Misleadingness of evidence does *not* supervene on accuracy.

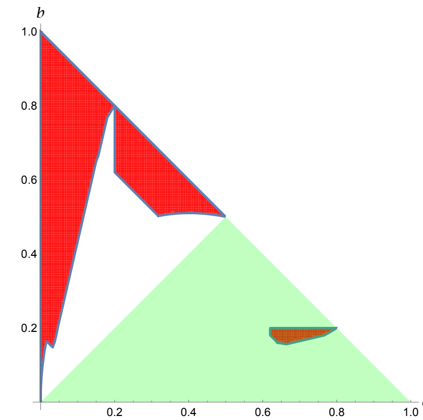
- This is because *favoring* does *not* supervene on accuracy.
- Differences in favoring do not entail differences in accuracy.
  - In all elimination counterexamples, E favors $B \vee C$ over $A \vee C$ — according to *all* (regular) $\mathfrak{c}$'s; but, E does *not* favor B over A — according to *any* (regular) $\mathfrak{c}$'s.
  - So, pick any regular credence function $\mathfrak{c}$. Inevitably, $\mathfrak{c}$ will exhibit the above difference in favoring relations, but $\mathfrak{c}$'s accuracy (*viz.*, Brier Score at the actual world) is constant.
- Because I'm a pluralist about epistemic value, I kinda like this defense of Brier. And, it *seems* to generalize…

- *Additive* strictly proper scoring rules $\mathfrak{S}$ (which support dominance arguments for probabilism & conditionalization) are defined in terms of *local* scores $\mathfrak{s}(x, 0)$ and $\mathfrak{s}(x, 1)$.
- For the Brier Score: $\mathfrak{s}(x, 0) = x^2$; $\mathfrak{s}(x, 1) = (x - 1)^2$.
- For the log score: $\mathfrak{s}(x, 0) = -ln(1 - x)$; $\mathfrak{s}(x, 1) = -ln(x)$.
- For the spherical score:

$$\mathfrak{s}(x, 0) = -\frac{1 - x}{\sqrt{x^2 + (1 - x)^2}} \qquad \mathfrak{s}(x, 1) = -\frac{x}{\sqrt{x^2 + (1 - x)^2}}$$



- I was surprised when Levinstein [8] used Schervish's technique [14] to generate an additive, strictly proper scoring rule $\mathfrak{L}$ with some *non*-misleading ECEs.
- I won't write down $\mathfrak{L}$ here (it's monstrous). But, here is a plot of $\mathfrak{L}$'s ECEs (non-misleading evidence cases in green).



- Lewis and Fallis [5] showed that all *symmetric*, additive, strictly proper scoring rules admit ECEs.
- Using Schervish's technique [14], Levinstein [7] generalized this result to all additive, strictly proper scoring rules.
- Levinstein's result implies that — if response (IV) is going to work — then it will require the adoption of a non-additive (*viz.*, non-local) strictly proper scoring rule.

☞ Richard Pettigrew [12] and Bernhard Salow [13] have recently shown that there are non-additive (*viz.*, non-local) strictly proper scoring rules that avoid all ECEs.

- I will be discussing Salow's scoring rule, since it is somewhat easier to grok (mathematically).
- Salow's score $\mathfrak{S}$ is not defined in terms of a local scoring rule $\mathfrak{s}$. Instead, it is defined globally over the relevant set of propositions $\mathcal{P}$. I will focus on the 3-partition $\{A, B, C\}$.

- Salow gives his scoring rule a pragmatic gloss (as will I). [One of the questions I have about this strategy is: what is the best way to give these rules an *alethic*/VBE gloss?]
- An agent with credence function $\mathfrak{c}$ is to face a decision problem involving three options $\langle o_1, o_2, o_3 \rangle$.
- The utility table for the problem is of the following form.

	A	B	C
o_1	x_1	0	0
o_2	0	x_2	0
o_3	0	0	x_3

- The agent doesn't know what the utilities $x_i \in (0, 1)$ are, but they know (a) they will choose the option that maximizes expected utility (according to $\mathfrak{c}$), and (b) the x_i have been sampled iid from a continuous $(0, 1)$-distribution π.
- Salow then defines $\mathfrak{S}^{\pi}(\mathfrak{c}, w)$ — the π-expected payoff of using $\mathfrak{c}$ to choose among the o_i, at world $w \in \{w_A, w_B, w_C\}$.

- Suppose π is the uniform distribution on $(0, 1)$. Then, we (*viz.*, *Mathematica*) can compute closed-form algebraic expressions for the $\mathfrak{S}^{\pi}(\mathfrak{c}, w)$, as functions of a, b (Extras).
- We can use *Mathematica* to verify $\mathfrak{S}^{\pi}(\mathfrak{c}, w)$ admits no ECEs.[3]
- Here is a general argument that $\mathfrak{S}^{\pi}(\mathfrak{c}, H_i)$ admits no ECEs, for any continuous π, and any partition $\{H_1, \ldots, H_n\}$.
- Suppose H_i is true. Then a subject using $\mathfrak{c}$ to choose wins x_i iff $\frac{\mathfrak{c}(H_i)}{\mathfrak{c}(H_j)} \cdot x_i > x_j$, for all $j \neq i$ (and they win nothing o.w.).
- Similarly, a subject using $\mathfrak{c}'$ to choose wins x_i iff $\frac{\mathfrak{c}'(H_i)}{\mathfrak{c}'(H_j)} \cdot x_i > x_j$, for all $j \neq i$ (and they win nothing o.w.).

☞ Since conditionalization preserves ratios of probabilities of uneliminated H's, $\mathfrak{S}^{\pi}(\mathfrak{c}', w) \geq \mathfrak{S}^{\pi}(\mathfrak{c}, w)$; and, $\mathfrak{c}'$ yields more winnings in some cases to which π assigns non-zero prob.

[3] Verifying that $\mathfrak{S}^{\pi}(\mathfrak{c}, w)$ is strictly proper (in this way) is very difficult.

$$\mathfrak{S}^{\pi}(\mathfrak{c}, w_A) = \begin{cases} \frac{1}{4} & 3a = 1 \wedge 3b = 1 \\ -\frac{a^2}{4b(a+b-1)} & 3a < 1 \wedge a \leq b \wedge 2a + b \leq 1 \\ \frac{b(7b-8)+2}{4a^2} & a + 2b = 1 \wedge ((2a + b > 1 \wedge ((2a < 1 \wedge a > b \wedge 3a > 1) \vee (2a \geq 1 \wedge 4b \leq 1))) \vee 2a > 1) \\ \frac{4a^3-6a(b-1)^2-b((b-6)b+6)+2}{12a^2(a+b-1)} & a + 2b < 1 \wedge ((2a + b > 1 \wedge ((2a < 1 \wedge a > b \wedge 3a > 1) \vee (2a \geq 1 \wedge 4b \leq 1))) \vee 2a > 1) \\ \frac{3(a-1)b^2+3(a(3a-2)+1)b+(a-1)^3-b^3}{12a^2b} & a > b \wedge a + 2b > 1 \wedge 2a \neq 1 \\ \frac{b^3-4a^3}{12a^2(a+b-1)} & 2a + b \leq 1 \wedge a > b \wedge 3a \neq 1 \\ \frac{5a^3+3a^2(b-1)+3a(b-1)^2+(b-1)^3}{12a^2b} & a \leq b \wedge 2a + b > 1 \wedge 3a \neq 1 \\ \frac{4-27b^3}{24-36b} & 3a = 1 \wedge 3b < 1 \\ \frac{3}{4}(b-2)b - \frac{1}{9b} + 1 & 3a = 1 \wedge 3b > 1 \\ \frac{1}{24}(-(8b^2) - 12b - \frac{1}{b} + 18) & \text{o.w.} \end{cases}$$

$$\mathfrak{S}^{\pi}(\mathfrak{c}, w_B) = \begin{cases} \frac{1}{4} & 3a = 1 \wedge 3b = 1 \\ \frac{b}{4a} & a + 2b = 1 \wedge \left((a > b \wedge \frac{1}{3} < a < \frac{1}{2}) \vee (2a \geq 1 \wedge 4b \leq 1) \vee 2a > 1\right) \\ -\frac{b^2}{4a(a+b-1)} & \left(a + 2b < 1 \wedge \left((a > b \wedge \frac{1}{3} < a < \frac{1}{2}) \vee (2a \geq 1 \wedge 4b \leq 1) \vee 2a > 1\right)\right) \vee (3a < 1 \wedge a > b) \\ \frac{b}{4-8b} & a = b \wedge 2a + b < 1 \\ \frac{9b^2}{8-12b} & 3a = 1 \wedge 3b < 1 \\ \frac{a^3-4b^3}{12b^2(a+b-1)} & a < b \wedge a + 2b \leq 1 \\ \frac{-a^3+3a^2(b-1)+a(9b^2-6b+3)+(b-1)^3}{12ab^2} & a + 2b > 1 \wedge a < b \wedge (3a > 1 \vee (2a + b > 1 \wedge 3a < 1)) \\ -\frac{a^3-6a^2+6(a-1)^2b+6a-4b^3-2}{12b^2(a+b-1)} & a + 2b > 1 \wedge 2a + b < 1 \\ \frac{3(a-1)b^2+3(a-1)^2b+(a-1)^3+5b^3}{12ab^2} & a + 2b > 1 \wedge \left((a > b \wedge \frac{1}{3} < a < \frac{1}{2}) \vee 2a > 1\right) \\ \frac{1}{48}\left(-\frac{1}{b^2} + 40b + \frac{6}{b} - 12\right) & 2a = 1 \wedge a + 2b > 1 \\ \frac{27b^3+36b-10}{108b^2} & 3a = 1 \wedge 3b > 1 \\ \frac{6b(2(b-1)b+1)-1}{12b^3} & a = b \wedge a + 2b > 1 \\ \frac{b(7b+2)-1}{16b^2} & 2a + b = 1 \wedge a < b \end{cases}$$

$$\mathfrak{S}^{\pi}(\mathfrak{c}, w_C) = \begin{cases} \frac{1}{4} & 3b = 1 \wedge a + 2b \geq 1 \wedge a \leq b \\ \frac{1}{8}\left(\frac{1}{b} - 1\right) & 2a + b = 1 \wedge a < b \\ \frac{b}{4a} & a + 2b = 1 \wedge ((3a > 1 \wedge 2a + b > 1) \vee 2a \geq 1) \\ \frac{(a+b-1)^2}{4ab} & a + 2b > 1 \wedge ((2a + b > 1 \wedge ((a < b \wedge 3a < 1) \vee 3a > 1)) \vee 2a \geq 1) \\ \frac{3b}{4} + \frac{1}{3b} - 1 & 3a = 1 \wedge 3b > 1 \\ -\frac{12(a-1)b^2+12(a-1)^2b+4(a-1)^3+5b^3}{12a(a+b-1)^2} & a + 2b < 1 \wedge ((3a > 1 \wedge 2a + b > 1) \vee 2a > 1) \\ -\frac{5a^3+12a^2(b-1)+12a(b-1)^2+4(b-1)^3}{12b(a+b-1)^2} & a + 2b > 1 \wedge 2a + b < 1 \\ \frac{4a^3+12a^2(b-1)+6a(b-1)^2-b^3}{12a(a+b-1)^2} & 2a + b < 1 \wedge a + 2b < 1 \wedge (3a > 1 \vee (a > b \wedge 3a < 1)) \\ -\frac{a^3-12(a-1)b^2-6(a-1)^2b-4b^3}{12b(a+b-1)^2} & a < b \wedge a + 2b < 1 \\ \frac{22-9b(3(b-2)b+8)}{12(2-3b)^2} & 3a = 1 \wedge a > b \\ \frac{6b(2(b-1)b+1)-1}{12b^3} & a + 2b = 1 \wedge 2a + b < 1 \\ \frac{3b((b-1)b+1)-1}{3(b-1)^3} & 2a + b = 1 \wedge a + 2b < 1 \\ \frac{b(7b-8)+2}{4(1-2b)^2} & a = b \wedge a + 2b < 1 \end{cases}$$

$$\mathfrak{S}^{\pi}(\mathfrak{c}', w_A) = \begin{cases} \frac{a}{3b} & a \leq b \\ \frac{1}{2} - \frac{b^2}{6a^2} & \text{o.w.} \end{cases}$$

[1] J. Dunn, "Accuracy, verisimilitude, and scoring rules," 2019.

[2] D. Fallis and P. Lewis, "The Brier rule is not a good measure of epistemic utility (and other useful facts about epistemic betterness)," 2016.

[3] B. Fitelson, "Contrastive Bayesianism," 2013.

[4] ______, "Likelihoodism, Bayesianism, and relational confirmation," 2007.

[5] P. Lewis and D. Fallis, "Accuracy, conditionalization, and probabilism," 2021.

[6] P. Lewis, D. Fallis, and B. Fitelson, "Accuracy-First Epistemology and Scientific Progress," 2024.

[7] B. Levinstein, "All rules suffer ECEs," 2024.

[8] ______, personal communication (*Mathematica* notebook), 2023.

[9] G. Oddie, "What accuracy could not be," 2019.

[10] I. Park, "Evidence and the epistemic betterness," 2023.

[11] R. Pettigrew, *Accuracy and the Laws of Credence*, 2016.

[12] ______, "Lewis, Fallis, and Fitelson on accuracy and scientific progress", 2024.

[13] B. Salow, "A little learning," 2024.

[14] M. Schervish, "A general method for comparing probability assessors," 1989.