ACCURACY, SELF-ACCURACY, AND CHOICE

August 17, 2013
Bristol Summer School
Bristol, UK

Jim Joyce
Department of Philosophy
University of Michigan
jjoyce@umich.edu

ACCURACY-CENTERED EPISTEMOLOGY

- The cardinal epistemic good is the holding of beliefs that accurately reflect the facts. Believers have a duty to rationally pursue doxastic accuracy.

- An **inaccuracy score** $\mathfrak{I}$ associates each credal state ***b*** and world ω with a non-negative real number, $\mathfrak{I}(\boldsymbol{b}, \omega)$, which measures ***b*'s** overall inaccuracy when ω is actual (0 = perfection).

  Truth-Directedness. Moving credences closer to truth-values improves accuracy.

  Extensionality. ***b***'s inaccuracy at $\boldsymbol{\omega}$ is solely a function of the credences ***b*** assigns and the truth-values $\boldsymbol{\omega}$ assigns.

  Propriety. If ***b*** is a probability, then ***b*** itself uniquely minimizes expected inaccuracy when expectations are calculated using ***b***.

- Such an $\mathcal{I}$ captures ***a consistent way of valuing closeness to the truth***.

A USEFUL EXAMPLES OF AN ACCURACY SCORES

$\langle x, y \rangle$ is the state in which a believer assigns credence x to X and y to $\sim X$.

- *Brier*: $\mathcal{B}_1(x, y) = ½ [(1 – x)^2 + y^2]$

  $\mathcal{B}_0(x, y) = ½ [x^2 + (1 – y)^2]$

We will use the Brier Score in what follows.

ACCURACY-NONDOMINANCE

Nondominance. If $\mathfrak{I}(\boldsymbol{c}, \omega) > \mathfrak{I}(\boldsymbol{b}, \omega)$ for all worlds ω, then, ***whatever one's evidence might be***, one is required to have an inaccuracy estimate for ***c*** that exceeds one's estimate of ***b***'s inaccuracy.

- In the accuracy-centered picture believers are required to hold beliefs that minimize estimated inaccuracy. It is *categorically* forbidden to hold accuracy-dominated credences.

- Joyce (1998) and (2009) uses this as the central premise in the accuracy argument for probabilism.

Theorem: Every incoherent credence function is accuracy-dominated (according to any score that meets the above conditions), but no coherent credence function is.

An Alleged Counterexample to Accuracy Dominance

Caie (2013): "Considerations of accuracy support the claim that an agent may rationally fail to have probabilistically coherent credences."

A *Caie proposition* is any claim *U* such that

- The believer knows that *U* is true either if $\boldsymbol{b}_t(U) < ½$ or if she has no determinate degree of belief for *U* at *t*.
- The believer knows that *U* is false if $\boldsymbol{b}_t(U) \geq ½$.

Example:

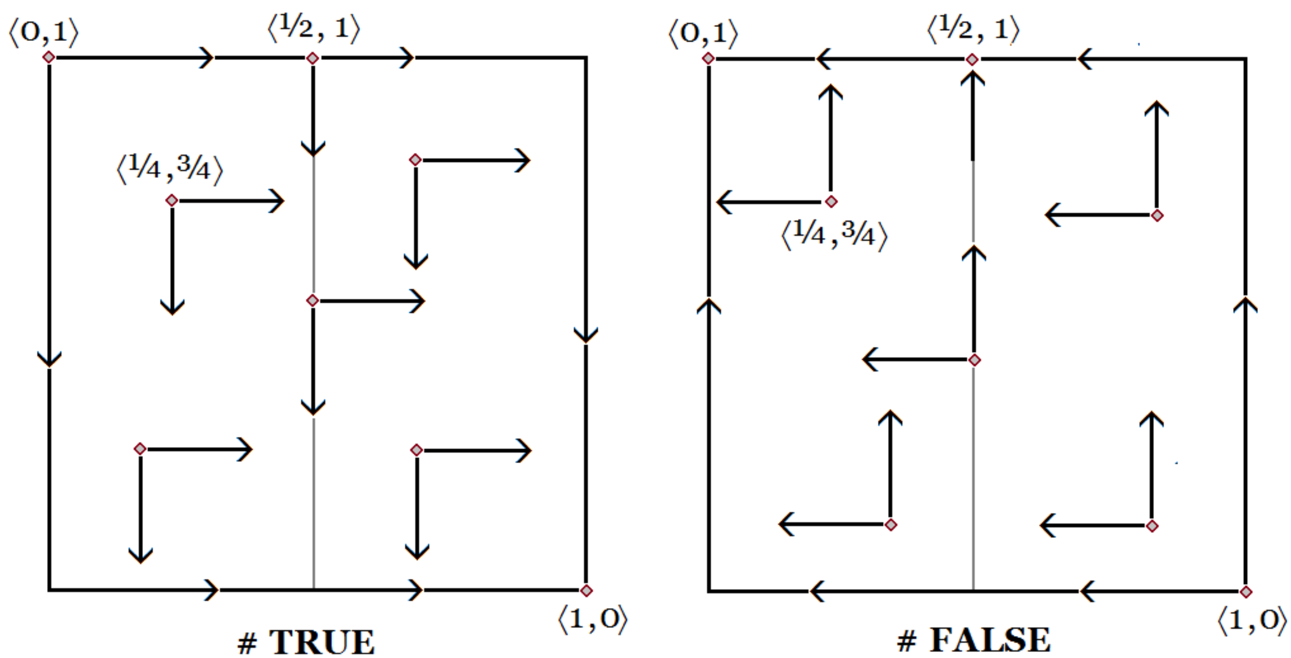
(#) My current credence in the proposition # expressed by the sentence (#) is not greater than ½.

Caie's Claim: Considerations of accuracy dictate that rational believers should assign credence ½ to *U* and credence 1 to ~*U*.

De Facto Accuracy

Let prospective credences at t be given by pairs $\langle \boldsymbol{b}_t(U), \boldsymbol{b}_t(\sim U)\rangle = \langle x, y\rangle$.

***De facto* accuracies,** with arrows pointing toward greater accuracy



$\boldsymbol{b}(U) < ½$ or $\boldsymbol{b}(U)$ not defined $\qquad \boldsymbol{b}(U) \geq ½$

- Here, we suppose a believer is in some definite credal state $\langle u_1, \tilde{u}_1\rangle$ *in re U* at time $t = 1$, and ask how accurate other credal states are *on that supposition*.

FACTS ABOUT DE FACTO ACCURACY

- If we measure inaccuracy using the Brier score, then ⟨¼, ¾⟩ has a lower *de facto* inaccuracy than ⟨½, 1⟩ whether *U* is true or false.

This suggests a **dominance argument**:

> - If I have the credences ⟨½, 1⟩, I will see one or both of *U* and ~*U* as live epistemic possibilities.
> - If *U* is true, then ⟨¼, ¾⟩ is more accurate than ⟨½, 1⟩.
> - If *U* is false, then ⟨¼, ¾⟩ is more accurate than ⟨½, 1⟩.

So, holding ⟨½, 1⟩ commits me to credences that are less accurate than ⟨¼, ¾⟩ in every world I regard as possible, which is supposed to be irrational.

Caie's Claim: This dominance reasoning is invalid!

AN EPISTEMIC DECISION

At time *t* = 0 a believer will *choose* her time *t* = 1 credences. (How?) She faces a kind of epistemic decision problem, with payoffs in Brier scores (lower = better):

	U	*~U*	
Opt for ⟨¼, ¾⟩	$^{9}/_{16}$	$^{1}/_{16}$	
Opt for ⟨½, 1⟩	$^{10}/_{16}$	$^{2}/_{16}$	← **Dominated**

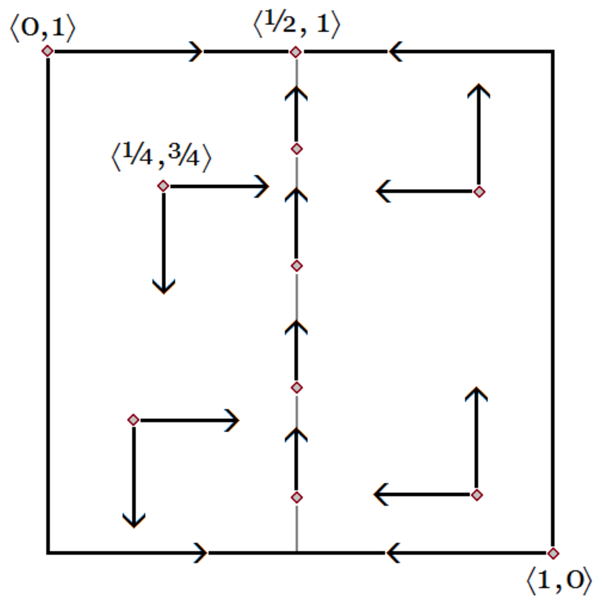
- Dominance arguments are ***not*** valid when the choice of an act affects the state of the world.

- Caie: Since adopting ⟨½, 1⟩ *makes U* false while adopting ⟨¼, ¾⟩ *makes U* true, the shaded boxes are not real possibilities. So, despite the dominance argument, ⟨½, 1⟩ is the rational choice.

- This is sound decision theory. A chooser should always select the option that is likely produce the best outcome *as a result of being chosen*. Here, this means choosing the act with the lowest ***self-inaccuracy***.

SELF-ACCURACY

- Caie asks: How accurate *would* $\langle u, \tilde{u}\rangle$ be if they *were* my credences?



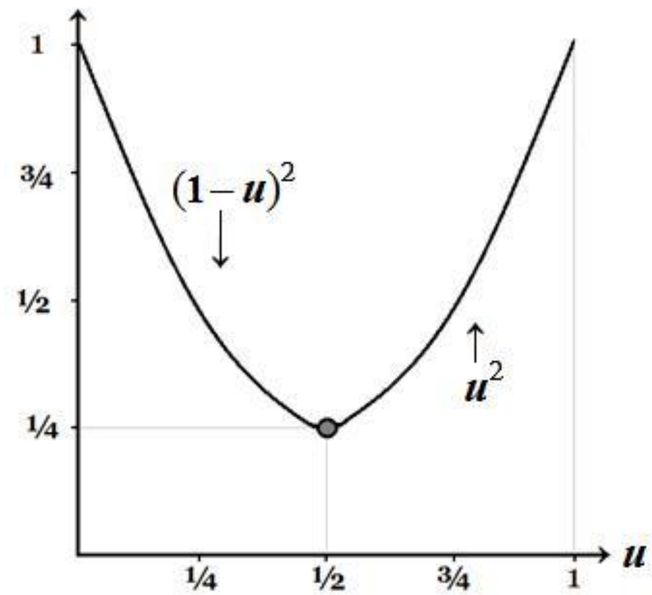
$$\mathcal{I}^{self}(u, \tilde{u}) = \mathcal{I}_1(u, \tilde{u}) \text{ when } u < ½$$

$$\mathcal{I}^{self}(u, \tilde{u}) = \mathcal{I}_0(u, \tilde{u}) \text{ when } u \geq ½.$$

FACT: For any reasonable measure of inaccuracy, minimum self-inaccuracy is (uniquely) attained at $\langle ½, 1\rangle$.

CAIE'S JUSTIFICATION FOR ⟨½, 1⟩

Step-1: If $\boldsymbol{b}_t(U) = u < ½$, the Brier inaccuracy of the *U* credence is $(1 – u)^2$.
If $u \geq ½$, the Brier accuracy is u^2.



- So, "the most accurate credence [one] can have in the proposition *U* is 0.5."

Step-2: Since an inaccuracy minimizer will choose $u = ½$, and since this makes *U* false, she will choose $\tilde{u} = 1$ as her credence for ~*U*.

THE STRUCTURE OF CAIE'S ARGUMENT

(a) Accuracy-centered epistemology is committed to saying that, at $t = 0$, the believer should choose the credal state that maximizes self-accuracy at t.

(b) The choice $\langle \boldsymbol{b}_1(U), \boldsymbol{b}_1(\sim U)\rangle = \langle ½, 1\rangle$ uniquely maximizes the self-accuracy of the time-t credences.

(c) As a result, an accuracy-centered epistemology will sanction $\langle ½, 1\rangle$ as the uniquely rational time $t = 1$ credal state to choose at time $t = 0$.

(d) If accuracy-centered epistemology sanctions some $t = 1$ credal state as the uniquely rational one to *choose* at $t = 0$, then it sanctions the holding of that state at $t = 1$.

Thus, the accuracy-centered must sanction $\langle ½, 1\rangle$ as the uniquely rational credal state to hold at t.

- I have doubts about (a), but won't press them. The real problem is (d)!

Ex Ante and Ex Post

We must be careful to distinguish the “act” of **adopting** credences at $t = 0$ from the subsequent state of **holding** those credences at $t = 1$.

- Imagine the believer taking a pill marked $\langle u, \tilde{u} \rangle$ at time $t = 0$, whose sole effect will be to cause her to have the credences $\boldsymbol{b}_1(U) = u$ and $\boldsymbol{b}_1(\sim U) = \tilde{u}$.
  - When she takes the pill at $t = 0$ we assume that she will know what she is doing, so that her $t = 0$ credences are $\boldsymbol{b}_0(U) = 0$ and $\boldsymbol{b}_0(\sim U) = 1$.
  - I assume that the agent knows what pill she has taken up through time $t = 1$. (She does not forget evidence.)
  - Following Caie, I assume that nothing prevents her from knowing her credences for U at $t = 1$ (though I don’t believe this).

- **Note for Future Reference**: If the believer takes the $\langle ½, 1 \rangle$ pill at $t = 0$, then at the moment of choice she sees herself as choosing future credences that are ***less*** accurate than her current credences!

AN INSTRUCTIVE PARALLEL: DE FINETTI'S PREVISION GAME

A proposition A is specified. The agent is given \$1 on the understanding that at $t = 0$ she must publically announce two real-valued "previsions" $a, \tilde{a} \in [0, 1]$, and must repay $\$\tfrac{1}{2}[(1 - a)^2 + \tilde{a}^2]$ if A is true and $\$\tfrac{1}{2}[a^2 + (1 - \tilde{a})^2]$ if A is false.

The decision she faces concerns how to *minimize* her penalty (= the Brier score of her previsions):

	A	$\sim A$
Announce $\langle \tfrac{1}{4}, \tfrac{3}{4} \rangle$	$^{9}/_{16}$	$^{1}/_{16}$
Announce $\langle \tfrac{1}{2}, 1 \rangle$	$^{10}/_{16}$	$^{2}/_{16}$
Announce $\langle a, \tilde{a} \rangle$	$\tfrac{1}{2} - a + \tfrac{1}{2}(a^2 + \tilde{a}^2)$	$\tfrac{1}{2} - \tilde{a} + \tfrac{1}{2}(a^2 + \tilde{a}^2)$

- De Finetti saw this as a method of *belief elicitation* that reveals the credences of coherent expected utility maximizers.
- Since Brier is *proper*, coherent agents minimize their expected penalty by reporting previsions that reveal their credences, so that the only permissible choices are $\boldsymbol{b}(A) = a$ and $\boldsymbol{b}(\sim A) = \tilde{a}$.

A FLY IN THE OINTMENT?

- In garden-variety cases, where the choice of previsions does ***not*** affect *A*'s truth-value,
  - A dominance argument can be invoked rule out the previsions ⟨½, 1⟩.
  - Announced previsions reveal credences (for expected profit maximizers).

- BUT, in cases, where the choice of previsions ***does*** affect *A*'s truth-value,
  - A dominance argument can**not** be invoked rule out the previsions ⟨½, 1⟩.
  - Announced previsions need **not** reveal credences.

- Consider this target proposition:

(*U**) The *u**-component of the reported prevision pair ⟨*u**, *ũ**⟩ will *not* be ½ or greater.

The decision:

	U^*	$\sim U^*$
Announce $\langle 1/4, 3/4 \rangle$	$^9/_{16}$	
Announce $\langle 1/2, 1 \rangle$		$^2/_{16}$
Announce $\langle u^*, \tilde{u}^* \rangle$, $a < 1/2$	$1/2 - u^2 + 1/2(u^2 + \tilde{u}^2)$	
Announce $\langle u^*, \tilde{u}^* \rangle$, $a \geq 1/2$		$1/2 - \tilde{u} + 1/2(u^2 + \tilde{u}^2)$

- The dominance argument for $\langle 1/4, 3/4 \rangle$ no longer applies.
- Whatever the agent's beliefs, her penalty is minimized by $u = 1/2$ and $\tilde{u} = 1$!
- The $u = 1/2$ prevision does ***not*** reflect her credence for U^*, but $\tilde{u} = 1$ *does* reflect her credence for $\sim U^*$!
  - Since the agent cannot to get $\tilde{u}$ any closer than a distance of ½ to it, she has an incentive to manipulate U^*'s truth-value by stating a u that does ***not*** reflect her credence.

Previsions only reveal credences when an agent can*not* influence the target proposition's truth-value by her choices.

The Prevision Game with a Pill

Suppose the agent must take a pill that *changes* her credences to match the previsions she thinks it best to announce.

- Choosing the ⟨½, 1⟩ pill produces the largest immediate payoff.
  - But, does this mean that ⟨½, 1⟩ is the best credal state to *inhabit* for purposes of decision making?
- No! ⟨½, 1⟩ is an awful credal state, practically speaking, since it leaves one open to easy exploitation.
  - One will buy the bet on the left for $5 and sell the bet on the right for any positive sum (say, $1), thereby ensuring herself of a loss (of $4).

*U** true	*U** false
$10	–$10

~*U** true	~*U** false
$0	–$10

The Practical Analogue of (d) is False

- A rational agent should only be willing to take the ⟨½, 1⟩ pill if she will revert to the ⟨0, 1⟩ credences before being offered other bets.
  - If she ends up betting on the basis of ⟨½, 1⟩, or any other incoherent credal state, she leaves herself open to a 'Dutch book'.
  - Even if she bets using coherent credences ⟨u, 1 – u⟩ with $u > 0$, she will still make decisions that are suboptimal in light of her knowledge.

Moral. The following principle of *practical* rationality is false!

> (d*) If some $t = 1$ credal state is the uniquely practically rational one for an agent to *choose* at $t = 0$, then the agent is rationally permitted to occupy that state at t.

Even if taking the ⟨½, 1⟩ pill is the right choice at $t = 0$, the $t = 1$ credences it leads to are defective from the perspective of practical irrationality.

PREVISIONS :: $t = 0$ CREDENCES (IN DE FINETTI) AS $t = 1$ CREDENCES :: $t = 0$ CREDENCES (IN CAIE).

- In each scenario a decision situation that usually produces a pair ⟨*a*, *ã*⟩ that gives the chooser's time $t = 0$ credences for *A* and ~*A* cannot properly do its job because it is infected by strategic considerations that stem from

  i. the chooser's ability to manipulate the target proposition's truth-value by her choice of a value for *a*, and

  ii. her inability to choose a value for *a* that is closer than a distance of ½ to the target's truth-value.

- In De Finetti's prevision game these factors conspire to ensure that an agent who takes the ⟨½, 1⟩ pill cannot endorse the resulting credences for use in practical decision making.

- In Caie's example they conspire to ensure that a believer who chooses ⟨½, 1⟩ cannot endorse these credences for use in representing the world.

THE USUAL CASE (FOR EXAMPLE ~*U*)

- As long as an epistemically rational believer's choice of $\boldsymbol{b}_t(A)$ does *not* affect *A*'s truth-value, she will select a credence for *A* that minimizes estimated inaccuracy in light of her evidence.

- Since $\mathcal{I}$-scores are proper, this means that a coherent believer will select a $t = 1$credence that agrees with her $t = 0$ credence, so that $\boldsymbol{b}_1(A) = \boldsymbol{b}_0(A)$.

- She is then in a position to *rationally endors*e the credence selected, i.e., she can affirm, based on her $t = 0$ evidence (which includes her choice), that her estimate for the inaccuracy of $\boldsymbol{b}_1(A) = \boldsymbol{b}_0(A)$ will be lower than her estimate for the inaccuracy of $\boldsymbol{b}_1(A) = a$ for any other *a*.

- Here premise (d) is uncontroversial: since the believer acquires no new evidence between $t = 0$ and $t = 1$, if her beliefs are rationally permissible at the former time then they will be rationally permissible at the later time too.

THE PROBLEM CASE (FOR EXAMPLE *U*)

- Things break down when the believer ***can*** influence the target proposition's truth-value by her choice of a credence, as with *U*.

- If she chooses at $t = 0$ to set $\boldsymbol{b}_1(U) = ½$, then at $t = 0$ she knows *U* is false.
  - So, at $t = 0$ she is manifestly *not* choosing what she takes to be the most accurate $t = 1$ credence in light of her evidence at $t = 0$.
  - NOTE: I am speaking here of her *actual* evidence, not the evidence she would have had had she chosen $\langle 0, 1\rangle$!

The General Point:

★ An epistemically rational believer who (at $t = 0$) chooses $\langle ½, 1\rangle$ as her $t = 1$ credences, makes it the case that $\langle 0, 1\rangle$ is the most accurate credal state to hold at *any* time. Since she knows this (through $t = 1$), she will see every credal state other than $\langle 0, 1\rangle$ as having suboptimal estimated inaccuracy.

ONE SYMPTOM OF ⟨½, 1⟩'S DEFECTIVE NATURE: INSTABILITY

- When the believer finds herself at $t = 1$ with credences ⟨½, 1⟩ she will know that she could be strictly more accurate by switching her credence for U to 0, and so should immediately shift to $\langle \boldsymbol{b}_{1+\varepsilon}(U), \boldsymbol{b}_{1+\varepsilon}(\sim U) \rangle = \langle 0, 1 \rangle$.

- In fact, if she is serious about having accurate beliefs then she should choose ⟨½, 1⟩ only if she is sure she will revert to ⟨0, 1⟩ immediately after $t = 1$.

  - If she gets stuck at $\langle \boldsymbol{b}_{t+\varepsilon}(U), \boldsymbol{b}_{t+\varepsilon}(\sim U) \rangle = \langle ½, 1 \rangle$ for any interval $x > \varepsilon > 0$, then any momentary advantage in accuracy that might have been secured by the choice of ⟨½, 1⟩ will be negated by the subsequent inaccuracy of her subsequent credences.

ANOTHER SYMPTOM OF ⟨½, 1⟩'S DEFECTIVE NATURE: LOGICAL TENSION

- Since she knows she can choose her credences, and since she will know what credences she chooses, the believer's probability for U conditional on the event 'I choose ⟨½, 1⟩' will be 1 at every time, including $t = 1$, and her $t = 1$ credences will look like this:

$$\boldsymbol{b}_1(U \mid \text{I choose } \langle ½, 1\rangle) = 0, \quad \boldsymbol{b}_1(\text{I choose } \langle ½, 1\rangle) = 1, \quad \boldsymbol{b}_1(U) = ½$$

  - Taken together, the first two identities amount to a probabilistic *modus ponens* to the conclusion $\boldsymbol{b}_1(U) = 1$.
  - The conflict between this (undrawn) conclusion and the third identity shows that the believer has a kind of unresolved tension among her credences

- The effect of the $t = 0$ choice of ⟨½, 1⟩ is to momentarily cause the believer to hold a credence that conflicts with her evidence about the accuracy of that credence, a kind of temporary irrationality.
  - When faced with this the believer should immediately revert to ⟨0, 1⟩.

A Bad Objection

Objection: (d) is a conceptual truth about rationality. If a theory of rationality requires you to choose to be in some future state, then that state must be a rationally permissible one for you to occupy.

Counterexample: Newcomb Problem with Precommitment

At $t = 0$ you can take a pill that turns you into a "one boxer". The pill will lead the mad scientist to predict (at $t = ½$) that you will take one box, and so will cause you to receive £1,000,000 but will also cause you to leave a free £1,000 on the table at $t = 1$.

- **CDT** tells you to take the pill at $t = 0$: it is obviously the act, among those available at $t = 0$, that has the best overall causal consequences.
- Even so, you act irrationally at $t = 1$ when you leave £1,000 on the table.
- Sometimes it is rational to choose an option that will make your future self behave irrationally.

Another Bad Objection

Objection: We cannot exclude ⟨½, 1⟩ on the basis of a comparison with ⟨0, 1⟩ (as ★ does) because the latter would be even more inaccurate if it were held.

- This would have bite if ★ were being used to advocate the choice of ⟨0, 1⟩, but it's not. ★ is merely used to show that ⟨½, 1⟩ defective at $t = 1$ *under the assumption* that ⟨½, 1⟩ is the correct $t = 1$ credal state to choose at $t = 0$.
    - Indeed, ★ is quite consistent with the idea that it would be impermissible to hold ⟨0, 1⟩ or to choose it.

Key Point. If one can show that ***b'*** must have a higher estimated inaccuracy than ***b*** in evidential situation *E*, then one has shown that it is impermissible to hold ***b*** when one is in *E*.

- This is true even if (i) it is rationally impermissible to hold ***b'*** in *E*, or (ii) that ***b'*** will have a higher estimated inaccuracy than ***b*** in some other evidential situation ***E'*** (e.g., the one a believer would inhabit if she were to choose ***b'***).

The Dominance Argument Revisited

- While Caie is right that we cannot appeal to the fact that ⟨¼, ¾⟩ accuracy dominates ⟨½, 1⟩ to rule out the *choice* of ⟨½, 1⟩ at $t = 0$, we *can* appeal to it to show that ⟨½, 1⟩ is an impermissible credal state to ***hold*** at $t = 1$.

- We do not *need* this additional consideration, since we already have ★, but it highlights a different problem with seeing ⟨½, 1⟩ as a permissible credal state to occupy.

  - ★ shows that the state is self-undermining in the sense that anyone who knows he is in it will have conclusive evidence which shows that ⟨0, 1⟩ is strictly more accurate.

  - The dominance argument shows that, independent of what evidence the occupier of ⟨½, 1⟩ might have, she cannot see herself as minimizing estimated inaccuracy while inhabiting that credal state.

THE DEEP MESSAGE

In accuracy-centered epistemology, the assessment of credal states is not like the assessment of choices.

- When is *choosing* among undesirable options, considerations of decision-theoretic rationality require one to 'make the best of a bad lot' even when the best is not very good at all.
  - If we offer a believer a choice among defective credal states, she acts wisely by choosing the least defective. But, it's still a defective credal state!

- In Caie's setup, every $t = 1$ credal state is defective in the very same way: a person in that state will have evidence which conclusively shows that another state is strictly more accurate.
  - This shows that *every* choice the believer makes leaves her with defective credences. There is no way to be epistemically rational at $t = 1$.
  - Even if $\langle ½, 1\rangle$ is the rational choice at $t = 0$, it does not follow that it can be rationally occupied. Even if it is the best of a bad lot, it is not very good!

References

Brier, G. W. (1950) “Verification of Forecasts Expressed in Terms of Probability,”

Caie, M (2013) “Rational Probabilistic Incoherence,” *The Philosophical Review* (forthcoming)

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

Joyce, J. (2009) “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief,” in Franz Huber & Christoph Schmidt-Petri, eds., *Degrees of Belief*. Amsterdam: Springer (Synthese Library, 342)