

# Elga 2013, New Rational Reflection, and Ambiguous Bayesianism

Kevin Dorst (kmdorst@mit.edu)

24.805, Fall 2025

## I. Elga 2013

Let's interpret  $P$  as 'the rational credence function' for now.

*The puzzle of the unmarked clock.*

Natural to say if pointing at  $n$ , should (rationally) be confident it's pointing near  $n$ .

But then find yourself reasoning, 'If 4, I'm rational to be confident of 3–5; if not, my confidence is irrationally high'.  $\rightsquigarrow$  so lower?

Analogy with Pangloss

**Certain:** if you're rational and certain of what credences it's rational to have, you'll have those credences.

If  $P_a(P = \pi) = 1$ , then  $P_a = \pi$

Natural to generalize, eg so that your credences are always a weighted average of the possible rational credences you leave open:

**Rational Reflection:** Conditional on the rational probability function being  $\pi$ , a rational person will adopt  $\pi$  as their probability function.

$P_a(q|P = \pi) = \pi(q)$   
Induces weighted average by total probability.

*Rational Reflection implies that  $P$  has clarity:*

Plug in  $\langle P = \pi \rangle$  for  $q$ . Then  $P_a(P = \pi|P = \pi) = \pi(P = \pi)$ . The left side, by definition, equals 1! So the right side must also equal 1, for any  $\pi$  on which  $\langle P = \pi \rangle$  gets positive probability.

Which is to say: if  $\pi$  might be the rational credence function, then  $\pi$  is certain that  $\pi$  is the rational credence function.

Elga thinks that's the wrong conclusion, because rational people can be *modest*, i.e. uncertain what the rational credences are.

→ Hypoxia case.

Elga's solution:

- Panel of potential experts: Cassandra, Merlin, Sherlock. They've written their credence functions down for you to inspect.
- If you learn Sherlock is the true expert, what opinions to adopt? Sherlock's?
- *No*. Make sure to inform Sherlock's credence function of anything you've learned, before adopting it.
- So conditional on Sherlock being the expert, you should adopt the opinions he *would* have were *he* to learn that he's the expert—i.e. the opinions he'd have, were his higher-order doubts removed.

Eg probability of rain given (1) Sherlock is the expert and (2) many people will use umbrellas tomorrow.

His *informed* opinions!

**New Reflection:** Conditional on the rational probability function being  $\pi$ , a rational person will adopt  $\pi(\cdot|P = \pi)$  as their probability function.

$P_a(q|P = \pi) = \pi(q|P = \pi)$ .

Elga's analysis of the clock:

- Draw it up; it satisfies New Reflection.

- Mistake was in applying Rational-Reflection-style reasoning.
- In this case, unlike Pangloss case, you have reason to think the (uninformed) rational credences are ‘falsity-guiding’.<sup>1</sup>
  - So it’s *wrong* to think, ‘Since there’s a good chance the rational credence is lower than mine, I should lower my credence’. Sometimes you should think the (uninformed) rational credence is wrong!

<sup>1</sup> In Horowitz’s (2014) terminology I.e.  $P_a(3-5)$  is high, and also  $P_a(3-5|P(3-5) = 67\%) = 1$ .

(Puzzled by how you could know that *your own evidence* (/credences) are falsity-guiding? Stay tuned for Ch. 6.)

## II. Ambiguous Bayesianism

Back to interpreting  $P$  as your reasonable but noisy credences.

Spoon frame again.

**Self-Reflection:**  $P_a(q|P(q) = x) = x$

Self-Reflection fails in this frame, for the same reason that Rational Reflection does.

- Learning your credence in  $\langle P = \pi_l \rangle$  should change it (to 0 or 1).
- So learning what  $P(d)$  is should change it (make more extreme).

New Reflection is equivalent to *Informed Reflection*:

**Informed Reflection:** Conditional on the informed credence in  $q$  being  $x$ , adopt credence  $x$ .  $P_a(q|\hat{P}(q) = x) = x$

ARGUMENT 1 for Informed Reflection: it follows from a variant of Savage’s ‘sure-thing principle’:

**No Foregone Questions (NFQ):** If you are certain that learning the true answer to a question  $Q$  would lead you to have an estimate for  $X$  above  $x$ , then you should already have an estimate for  $X$  above  $x$ .

ARGUMENT 2 for Informed Reflection: it makes ambiguity *way* easier to work with—factorization theorem.

**Standard Bayesianism:**  $(W, a, P, P^+)$  is *Standard-Bayesian* iff:

- $P$  is self-aware and **clear**; and
- There is a partition  $\mathcal{Q}$  such that your posterior is the result of conditioning on the true cell of  $\mathcal{Q}$ : for all  $w$  and  $q$ ,  $P_w^+(q) = P_w(q|Q_w)$ .

**Ambiguous Bayesianism:**  $(W, a, P, P^+)$  is *Ambiguous-Bayesian* iff:

- $P$  is self-aware and **factorable**; and
- There is a partition  $\mathcal{Q}$  such that your posterior is the result of conditioning on the true cell of  $\mathcal{Q}$ : for all  $w$  and  $q$ ,  $P_w^+(q) = P_w(q|Q_w)$ .

Choose your adventure:

- 1) Why updating by conditioning on a partition? (And how?)
- 2) Why (synchronously) stable?

### (1) Why and how partitional conditioning?

I.e.  $P$  is *factorable* into higher-order expectations of informed probabilities:  $P_a(q) = \mathbb{E}_{P_a}(\hat{P}(q))$

Oops! Ambiguous Bayesianism will sometimes violate NFQ. (But Valuable Bayesianism (Ch. 6) never will.)

Two Qs: (i) why evidence partitional, and (ii) if so, why condition?

(i) is complicated. Clocks and externalists. Without clarity, can't recover a partition. Lots of good arguments for non-partitionality.

Williamson 2000

But hard questions:

- How to reliably do them?  
If you *can*, why can't you do *better*? (Schoenfield 2017; Gallow 2021)
- How to constrain them to avoid 'rationality' absurdities.
- How to *induce* them in subjects, in a controlled way?

Contrast with partitions: update your algorithm, even if unsure of its propensities  
E.g. clock.  
Li (cf. 2025)

Suppose evidence is partitional.

Then (ii) is (arguably) vindicated by Greaves and Wallace 2006.

Proper scoring rules, update plans,  $P_w$ -expected accuracy.

The update-plan that maximizes  $P_w$ -expected accuracy is to condition  $P_w$  on the true answer to  $Q$ .

What happens if each world *implements* this plan? Each world updates from  $P_w$  to  $P_w^+(\cdot) = P_w(\cdot|Q_w)$ , as Ambiguous-Bayes says.

## 2) Why synchronically stable?

Two route to instability: (i) Self-Martingale and (ii) Informed Reflection.

(i) **Self-Martingale:**  $\mathbb{E}_{P_a}(P(q)) = P_a(q)$ .

Self-Martingale implies clarity.

Variant of Leonard Savage's argument against higher-order probabilities

Right response (I claimed): Self-Martingale is intuitive, but fails for the same reason that Self-Reflection does.

Making more intuitive:

- If can be unsure of  $X$ , can be misled to inaccurate estimates of it.
- If HOU, what happens on the edge of the range of your uncertainty?

Let  $X = P(q)$ . Then Self-Martingale says that your estimates of your credences are always certain to be accurate.

(ii) What about Informed Reflection:  $\mathbb{E}_{P_a}(\widehat{P}(q)) = P_a(q)$ .

Why can't use expectations of  $\widehat{P}(q)$  to figure out what  $P(q)$  is?

Compare:  $\mathbb{E}_{P_a}(\mathbb{1}(q)) = P_a(q)$ .

If uncertain about latter, also about former.

What about your expectations of your expectations of  $\widehat{P}(q)$ , or  $\mathbb{1}(q)$ ?

Also uncertain.

Turtle theorem, illustrated in spoon case.

Let  $w$  be a world where  $P(d) = 0.55$  and  $d$  is true;  $v$  be one where  $P(d) = 0.45$  and  $d$  is false. Then:

$$\mathbb{E}_{P_w}^0(d) = 1.$$

$$\mathbb{E}_{P_w}^1(d) = \mathbb{E}_{P_w}(\mathbb{1}(d)) = P_w(d) = 0.55.$$

$$\mathbb{E}_{P_w}^2(d) = \mathbb{E}_{P_w}(\mathbb{E}(\mathbb{1}(d))) = \mathbb{E}_{P_w}(P(d)) = 0.51$$

$$\mathbb{E}_{P_v}^0(d) = 0.$$

$$\mathbb{E}_{P_v}^1(d) = \mathbb{E}_{P_v}(\mathbb{1}(d)) = P_v(d) = 0.45.$$

$$\mathbb{E}_{P_v}^2(d) = \mathbb{E}_{P_v}(\mathbb{E}(\mathbb{1}(d))) = \mathbb{E}_{P_v}(P(d)) = 0.49$$

"If anything is to be probable, some things must be certain."

→ On one interpretation, true under clarity but false under ambiguity.

## References

- Gallow, J. Dmitri, 2021. 'Updating for Externalists'. *Noûs*, 55(3):487–516.
- Greaves, Hilary and Wallace, David, 2006. 'Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility'. *Mind*, 115(459):607–632.
- Horowitz, Sophie, 2014. 'Epistemic Akrasia'. *Noûs*, 48(4):718–744.
- Li, Jiayi, 2025. 'Vision Science as Experimental Epistemology: Evidential Asymmetry in the Lab'.
- Schoenfield, Miriam, 2017. 'Conditionalization Does Not (In General) Maximize Expected Accuracy'. *Mind*, 126(504):1155–1187.
- Williamson, Timothy, 2000. *Knowledge and its Limits*. Oxford University Press.