

# Finding Balance in Uncertain Times

Epistemic agents are **reflective** if their opinions line up with what they expect their future rational opinions to be, and **modest** if they are uncertain what the rational opinions are. Here are two proposed norms of epistemic rationality:

1. REFLECTION: rational agents are always reflective.
2. MODESTY: rational agents are sometimes modest.

I argue for:

*Unreflective Modesty*: REFLECTION and MODESTY are incompatible.

## Part I: Modesty and Reflection

**Modesty.** Consider the following case:

*One Coin*: A fair coin is flipped out of Emily's sight. We have two colors drawn from a red-green gradient: but one is *slightly* more red and one is *slightly* more green, so that they look different but are hard to tell apart. Emily is shown the redder color if the coin lands heads, and the greener color if the coin lands tails.

Let us suppose that *One Coin* rationalizes the following update:

*Uncertain Update*: Initially, Emily should have equal credence in heads and tails. If she is shown red, she should raise her confidence in *heads* (but not to 1). If she is shown green, she should lower her confidence in *heads* (but not to 0). For concreteness, let's say Emily should raise/lower by a factor of  $1/3$ .

Let  $P$  and  $P^+$  be descriptions for the prior and posterior Emily should have (so that  $P(q) = p_h(q)$  if the coin came up heads, etc.). Then the rational prior  $P$  and posterior  $P^+$  for Emily are as follows in each possibility:

$$\begin{aligned} \text{If coin lands } \textit{heads}: [p_h(h) = 1/2, \quad p_h(t) = 1/2] &\rightarrow [p_h^+(h) = 2/3, \quad p_h^+(t) = 1/3] \\ \text{If coin lands } \textit{tails}: [p_t(h) = 1/2, \quad p_t(t) = 1/2] &\rightarrow [p_t^+(h) = 1/3, \quad p_t^+(t) = 2/3] \end{aligned}$$

Suppose that the coin in fact came up tails. Then Emily should have credence  $2/3$  in  $\langle \textit{tails} \rangle$ .<sup>1</sup> But given the setup of the case, Emily knows that

$$\langle \textit{tails} \rangle \leftrightarrow \langle \text{It's rational to have credence } 2/3 \text{ in } \textit{tails} \rangle.$$

So she should have credence  $2/3$  in  $\langle \text{It's rational to have credence } 2/3 \text{ in } \textit{tails} \rangle$ .<sup>2</sup> So if *Uncertain Update* is rational, Emily should have some opinion (credence  $2/3$  in *tails*) and also be unsure that that opinion is rational. She should be **modest**.

**Definition 1.** An agent represented by  $P$  is modest if for some world  $w$  and proposition  $q$ ,  $P$  is unsure (at  $w$ ) about the value of  $P(q)$ : for all  $t \in [0, 1]$ ,

$$p_w(\langle P(q) = t \rangle) < 1.$$

The norm of MODESTY says: rational agents are sometimes modest. The norm follows from many formulations of externalism in epistemology, as well as the thought that epistemic rationality should not be so demanding as to require that rational agents can never be unsure what is rational.

Adrian Liu (adrian.liu@rutgers.edu)  
Central APA Feb 19 2026  
This handout also at [adrianliu.net](http://adrianliu.net)  
and on the APA App.

Plan of the talk:

- **Part I** (pp.1–2): I introduce MODESTY and REFLECTION.
- **Part II** (p.2): I discuss existing arguments for *Unreflective Modesty* and argue that we should not be satisfied with them.
- **Part III** (p.3): I give my argument for *Unreflective Modesty*, via a case (*Two Coins*) and a theorem (Main Theorem).
- **Replies to Comments** (p.4).

<sup>1</sup> I use angle brackets to denote propositions.

<sup>2</sup> If the coin in fact came up tails, then  $P(t) = p_t(t) = 2/3$ . But since  $t \leftrightarrow \langle P^+ = p_t^+ \rangle \leftrightarrow \langle P^+(t) = 2/3 \rangle$ , Emily should also have credence  $P(\langle P^+(t) = 2/3 \rangle) = p_t(\langle P^+(t) = 2/3 \rangle) = p_t(t) = 2/3$ .

**Reflection.** Given a conception of Emily's rational credences in each possibility, we have a notion of the *expected* rational opinions (relative to a prior).<sup>3</sup>

For instance, from the perspective of Emily's rational prior  $p$  (whether  $p = p_h$  or  $p = p_t$ ), the expected rational posterior credence in  $t$  can be given by

$$\begin{aligned}\mathbb{E}_p[P^+(t)] &= \sum_w [p\text{'s credence in } w] \cdot [\text{value of } P^+(t) \text{ at } w] \\ &= [p(h) \cdot p_h^+(t)] + [p(t) \cdot p_t^+(t)] = [1/2 \cdot 1/3] + [1/2 \cdot 2/3] = 1/2.\end{aligned}$$

We say Emily is **reflective** because her prior credences equal, or “reflect”, her expected rational posterior credences.

**Definition 2.**  $P$  reflects  $P^+$  if for every world  $w$  and proposition  $q$ :

$$p_w(q) = \mathbb{E}_{p_w}[P^+(q)]$$

An agent represented by prior  $P$  and posterior  $P^+$  is reflective if  $P$  reflects  $P^+$ .

The norm of **REFLECTION** says: rational agents are always reflective.

Intuition: being unreflective is being “unbalanced” in a way you could expect or predict (and therefore, perhaps, prevent) beforehand. For example:

*One Coin\**:<sup>4</sup> like *One Coin*, but if the coin lands tails, Emily get *slightly stronger* evidence about the truth, warranting increasing her credence in  $t$  to  $5/6$ .

In *One Coin\**, Emily's expectation of her posterior  $P^+(t)$  is

$$\mathbb{E}_p[P^+(t)] = [p(h) \cdot p_h^+(t)] + [p(t) \cdot p_t^+(t)] = \left[\frac{1}{2} \cdot \frac{1}{3}\right] + \left[\frac{1}{2} \cdot \frac{5}{6}\right] = \frac{7}{12} > \frac{1}{2}.$$

## Part II: Some Existing Arguments for *Unreflective Modesty*

1. **The Bayesian Argument:** Suppose rational agents have non-modest priors, get propositional evidence, and update by conditionalization. Then rational agents are reflective if and only if their posteriors are not modest.<sup>5</sup>
  - *Limitation:* the Bayesian Argument assumes conditionalization. But the conditionalization norm is controversial when evidence is *nonpartitional*. Many have argued that when evidence is nonpartitional, conditionalization is not rational and have proposed alternate update rules (sometimes ones in which posteriors are modest but priors are still reflective).
  2. **The Null-Update Argument:** Consider a rational agent with a modest prior  $P$ . By a theorem of Samet (2000),  $P$  does not reflect  $P$ . So suppose an agent updates on no relevant evidence at all, so that  $P^+ = P$ . Then  $P$  does not reflect  $P^+$ . So modest priors are always unreflective.
  - *Limitation:* the null-update argument appeals to a “no-evidence update” and thus relies on a synchronic instability of modest priors. But (1) it's not at all obvious that a synchronic version of reflection is a plausible epistemic norm, (2) it only exhibits reflection failures for trivial “updates,” and (3) the argument leaves us without a specifically diachronic instability.<sup>6</sup>
- Both strategies demonstrate an incompatibility between **REFLECTION** and **MODESTY** in restricted scenarios. But neither shows a sufficiently general and thus sufficiently worrying incompatibility.

<sup>3</sup> In general,

$$\mathbb{E}_p[P^+(q)] =_{\text{df}} \sum_{w \in W} p(w) \cdot p_w^+(q).$$

<sup>4</sup> If **REFLECTION** is a rational requirement, then *One Coin\** is not possible: evidence would never warrant those shifts in credence.

<sup>5</sup> Ask me about Dorst's (2023, “Rational Polarization”) variation on the Bayesian Argument, which suffers from similar flaws.

“Quantified Beliefs and Believed Quantities.”

<sup>6</sup> Salow (2017, “The Externalist's Guide to Fishing for Compliments”) suggests a variation on **REFLECTION** to capture the “specifically diachronic” norm:  $\mathbb{E}_p(P) = \mathbb{E}_p(P^+)$ . This gets around the Null-Update argument, but not the one I will later give.

### Part III: The New Argument for *Unreflective Modesty*

Consider the following case:

*Two Coins*: Emily is shown the reddish or greenish color based on how the red coin lands, like in *One Coin*. Then a (fair) blue coin is flipped out of Emily's sight, and she is told whether or not *both coins came up heads*.

I owe *Two Coins* to Kevin Dorst.

If *Uncertain Update* is rational, then Emily's rational prior credal state before the blue coin is flipped (but after the red coin is flipped) is equal to the posterior we gave above in *One Coin* (I've condensed the table). We can fine-grain the prior to capture the blue coin (assuming it is fair).

	$P(h)$	$P(t)$	
$p_h(\cdot)$	2/3	1/3	
$p_t(\cdot)$	1/3	2/3	

fine-grain →

	$P(hh)$	$P(ht)$	$P(th)$	$P(tt)$
$p_{hh}$	1/3	1/3	1/6	1/6
$p_{ht}$	1/3	1/3	1/6	1/6
$p_{th}$	1/6	1/6	1/3	1/3
$p_{tt}$	1/6	1/6	1/3	1/3

← Each world is split into two possibilities (where the blue coin comes up heads and when it comes up tails), and Emily is 50/50 between each pair.

When Emily learns whether both coins came up heads, she learns which one of the propositions  $\{hh\}, \{ht, th, tt\}\}$  is true.

Suppose that, in each possibility Emily should conditionalize on this information.<sup>7</sup> Then her posterior at each of the four possibilities should be:

$P$	$hh$	$ht$	$th$	$tt$
$p_{hh}$	1/3	1/3	1/6	1/6
$p_{ht}$	1/3	1/3	1/6	1/6
$p_{th}$	1/6	1/6	1/3	1/3
$p_{tt}$	1/6	1/6	1/3	1/3

→

$P^+$	$hh$	$ht$	$th$	$tt$
$p_{hh}^+$	1	0	0	0
$p_{ht}^+$	0	1/2	1/4	1/4
$p_{th}^+$	0	1/5	2/5	2/5
$p_{tt}^+$	0	1/5	2/5	2/5

In *Two Coins*, Emily's rational prior is modest and unreflective:  $P$  is modest, and  $P$  does not reflect  $P^+$ .<sup>8</sup> (in fact, at every world  $w$ ,  $p_w$  does not reflect  $P^+$ ). So if Emily's update in *Two Coins* could be rational, then *Unreflective Modesty* is true: **MODESTY AND REFLECTION** are incompatible.

<sup>7</sup> i.e., in each world  $w$ , she adopts posterior  $p_w^+(\cdot) = p_w(\cdot | e_w) := p_w(\cdot \wedge e_w) / p_w(e_w)$ , where  $e_w$  is the proposition learned at  $w$ . To implement this in the table, we cross out the worlds ruled out after the information Emily learns, and renormalize so that Emily credences sum to 1.

<sup>8</sup> For example:

$p_{hh}(h_-) = 2/3$  but  $\mathbb{E}_{p_{hh}}[P^+(h_-)] \approx 0.56$   
 $p_{th}(h_-) = 1/3$  but  $\mathbb{E}_{p_{th}}[P^+(h_-)] \approx 0.38$ .

**The main theorem of this talk shows that *Two Coins* is not a special case.**

**Theorem 1** (Main Theorem). *Suppose that for any partition  $E$  of possible worlds,  $P(\cdot | E)$  is defined by  $w \mapsto p_w(\cdot | e_w)$ , and that  $W$  has at least three worlds. Now suppose  $P$  is modest and  $p_w(\langle P = p_w \rangle) > 0$  for some  $w \in W$ . Then there is some nontrivial (i.e. not just  $W$  itself) partition  $E$  such that  $P$  does not reflect  $P(\cdot | E)$ .*

This theorem says that, given minimal assumptions about rational priors and updating, modest priors are always liable to be unreflective.

Compared to previous arguments:

- The Main Theorem assumes conditionalization only when evidence is partitional, unlike the Bayesian Argument.
- The Main Theorem applies to a much wider (and less avoidable) class of possible updates than the Null-Update argument.

Thus the Main Theorem establishes *Unreflective Modesty* in general.

# Finding Replies to Comments in Uncertain Times

## Whither Externalism?

For a number of precisifications of the idea of “externalism”,<sup>9</sup> externalism and modesty are quite similar, and I take modesty to be a way of talking about rational opinions directly, without having to make formal assumptions about evidence. And then the Main Theorem establishes that externalism is incompatible with reflection in deeper ways that previously known.

## A Tale of Two Reflections

As Fang puts it, at the heart of local vs global reflection is “whether the opinion you should reflect depends on what prior you could have had.” I would put it somewhat differently: the question is whether the opinion you should reflect depends on the possible priors that (for all you know) you *should* have.

Fang argues for Local Reflection based on

**Noncontrastiveness\***: “Hold fixed your sample space and [possible] evidence; how you update should not depend on the prior you could have had.”<sup>10</sup>

Since noncontrastiveness\* is a constraint on updating rules, not ways of measuring expectation, it is neutral between local and global reflection. Indeed, the update rule  $w \mapsto p_w(\cdot | e_w)$ , i.e. “conditionalize  $p_w$  on the true cell  $e_w$  of the partitional evidence  $E$ ”, which I use for partitional evidence, satisfies noncontrastiveness\*.<sup>11</sup>

Nonetheless, the intuition of noncontrastiveness could be captured like this:

**Noncontrastive Reflection**: “Hold fixed your sample space and [possible] evidence; the posterior you should reflect depends only on the prior that is actually rational for you, and does not depend on *other* priors that (for all you know) could be rational for you.”

However, I think that noncontrastive reflection gets us reflection only by advising epistemic agents to not be modest, i.e. to assume that their own opinions are rational. After all, if you are unsure what the rational prior is, but you expect that, whatever it is, the way to update is to conditionalize, then you should not be content with reflecting a posterior that assumes your actual prior is rational and only considers variations in possible evidence. To do so would be to ignore the modesty in one’s own epistemic state. In this way, noncontrastiveness for priors preserves reflection at the expense of modesty.

So I am inclined to think that **noncontrastive reflection** does not provide an escape route to *Unreflective Modesty* because it does away with modesty, while **noncontrastiveness\*** does not provide an escape route because it does not directly take a stand on local vs global reflection.

<sup>9</sup> For example:

1. **Antiluminosity Externalism**: for (at least) many epistemically-relevant states, one can be in the state without [knowing/having some sort of access to] the fact that one is in that state.  
→ Modesty follows from antiluminosity externalism if the epistemically-relevant state is the state of having evidence that warrants a particular set of opinions, and the sort of access is being certain.
2. **Evidence Externalism**: one can get evidence  $E$  that doesn’t entail that one got  $E$ .  
→ Modesty is a version of evidence externalism. Instead of talking about the entailment relation for (propositional) evidence, it talks about warranting certainty that one got the evidence.

<sup>10</sup> For what it’s worth, I think noncontrastiveness\* is plausible, but that there are good reasons to reject noncontrastiveness within most frameworks for rational updating.

<sup>11</sup> What’s going on in local reflection is that a higher-order uncertain prior  $P$  at a world  $w$  leaves open that the rational prior might in fact be different, and so the rational posterior might in fact be different. The possibility of different priors does not change the actual rational update. Rather, the possibility of different priors makes the agent unsure what the actual rational update is.