Finding Balance in Uncertain Times

Adrian Liu

Formal Approaches to Rationality and Meaning, October 11 2025

The Reflection Principle

You **reflect** the rational opinions on a proposition q if your credence in q, conditional on the supposition that the rational credence to have in proposition q is t, is t.

- π: your actual prior credences.
- R: the rational credences to have in your situation (after gaining evidence, etc).
- $[R(q) = t]$: the proposition "the rational credence to have in proposition q is t."

$$\pi(q \mid [R(q) = t]) = t \qquad \text{(Reflection)}$$

We could swap out R for some other definite description, like "the objective chances," or "your future credences," or "the expert's opinions."

Updating

But we won't, because I'm interested in *rational updating on evidence*:

- π is your actual (rational) prior
- e is the evidence you will actually get
- ρ is the posterior credence it is rational for you to have after getting e.

$$\langle \pi, e \rangle \mapsto \rho$$

If you're not sure what evidence you'll get:

- E nonrigidly picks out the evidence you will get
- R nonrigidly picks out the posterior you should have.

$$\langle \pi, E \rangle \mapsto R$$

Consequences of Reflection

Fact 1

If a prior π *reflects* a posterior description R, then three other features also hold:

1. π **trusts** R: π expects R to always give better estimates than π.
2. π **balances** R: π expects R to have the same estimates as π on average.
3. R is **higher-order certain**: R is certain about the actual value of R.

Life Without Reflection

Suppose we don't want to require higher-order certainty. Then we cannot require reflection, and it becomes an open question whether to require value and balance: $\times$ π is rationally required to **reflect** P^+.

1. ? π is rationally required to **value** P^+.
2. ? π is rationally required to **balance** P^+.
3. $\times$ P^+ is rationally required to be **higher-order certain**.

Finding Balance in Uncertain Times

- I think **balance** is a rational requirement on *diachronic coherence* grounds: rational agents must not think they will over- or under-estimate their evidence.
- I also think it plausible that higher-order uncertainty could be rational (and thus that reflection might be too strong).
- But I now think it's hard to make balance and higher-order uncertainty play nice.

Finding Balance in Uncertain Times

In This Talk

Question: How much balance can you get in conditions of higher-order uncertainty?
Answer: Not much at all. **balance** and **higher-order uncertainty** are in great tension.
Upshot: It will be really hard to both allow higher-order uncertainty and require balance in one's theory of rational epistemic updating.

What to Expect

- §1: *One Coin*.
- §2: Balance in *One Coin*.
- §3: *Two Coins.*
- §4: The Problem for Balance in *Two Coins.*

§1. *Higher-Order Uncertainty with One Coin.*

Setup

1. Two strings on the board.
2. **Exactly one** can be made into a word by filling in the blanks.
3. **Left** if heads, **Right** if tails.

E.g.: if **Heads**:

C O _ P _ _ T _	X Y _ O F _ A R
(C O M P L E T E)	(no word completes this)

Example 1

F A _ _ I _ N

F _ A M _ L Y

Example 2

B A _ _ N _ H

F _ _ I _ I _ _

One Coin

Story:

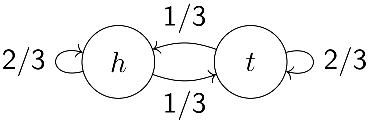
- If the coin came up heads, then the left string was completable and the right wasn't.
- Since you saw the two strings and know English, you have evidence about how the coin came up, *even if* you didn't find a word in either string (as I hope you didn't).
- So it may be rational for you to raise your credence in *heads*, perhaps due to some funny (veridical) feeling that the left word was completable.
- But it seems also rational for you to remain *unsure* how much you should raise it.

STAY WITH ME

Grant for now that such cases are possible, and that the story above is an acceptable way to model them. One possible takeaway from this talk is that this should ***not*** *be how we model the target phenomenom.*

Modeling Higher-Order Uncertainty

- The circles are (epistemically) possible worlds. Here *h* for heads and *t* for tails.
- The arrows indicate the credences you should have *in* a world if you are *at* a world. E.g.: the "1/3" arrow from *h* to *t* means that if world *h* is actual, you should have credence 1/3 in world *t*.

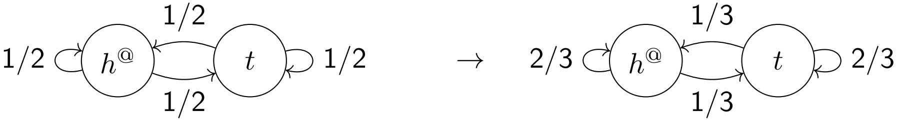


So this figure carries the information:

- If the coin came up heads, you should be 2/3 confident that the coin came up heads (and 1/3 confident that it came up tails).
- If the coin came up tails, you should be 2/3 confident in tails, and 1/3 in heads.

Modeling Higher-Order Uncertainty

(*prior to posterior*, where the @ is the actual world (unbeknownst to the agent))



Another Way to Model The Same Information

Each *row* is a credence function and each *column* of a row is that credence function's credence in the respective world.

$$\textit{prior:}\quad \left(\begin{array}{c|cc} & h & t \\ \hline \pi_h^{@} & 1/2 & 1/2 \\ \pi_t & 1/2 & 1/2 \end{array}\right) \qquad \textit{posterior:}\quad \left(\begin{array}{c|cc} & h & t \\ \hline \rho_h^{@} & 2/3 & 1/3 \\ \rho_t & 1/3 & 2/3 \end{array}\right)$$

Or we might think that, since π is the same at both worlds, we should just write it once:

$$\textit{prior:}\quad \left(\begin{array}{c|cc} & h & t \\ \hline \pi & 1/2 & 1/2 \end{array}\right) \qquad \textit{posterior:}\quad \left(\begin{array}{c|cc} & h & t \\ \hline \rho_h^{@} & 2/3 & 1/3 \\ \rho_t & 1/3 & 2/3 \end{array}\right)$$

Formalizing Higher-Order Uncertainty

To formalize these definite descriptions "the rational credences" we use functions $R : \Omega \to \mathbb{C}(\Omega)$ mapping worlds to credence functions, e.g.:

$$R : \begin{cases} h \mapsto \rho_h \\ t \mapsto \rho_t, \end{cases}$$

where $\rho_h(h) = 2/3$ and $\rho_t(h) = 1/3$. Call such functions **Credence Families.**

A credence family R is **higher-order uncertain** if, at some world w, the credence function ρ_w that R takes on at w is uncertain about the proposition "R takes on ρ_w."

Balanced Updating

The Greatest Weighting

What, if some day or night a demon were to steal after you into your loneliest loneliness and say to you:

> *"This coin as we now flip it and have flipped it, we will flip once more and innumerable times more; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unutterably small or great in our flips will have to return to us, . . ."*

what would you expect your average credence in *heads* to be?

Balance

Natural proposal: what you expect your average credence to be is just the average of the credences you think you will have, weighted by how likely you are to have them:

$$\sum_{w \in \{h,t\}} \pi(w)\rho_w(h) = \pi(h) \cdot \rho_h(h) + \pi(t) \cdot \rho_t(h)$$
$$= \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}.$$

This makes sense, and notably it matches your *prior* credence in *heads*: $\pi(h) = \frac{1}{2}$. If the coin is fair, then it should land heads half the time (on average), and it seems irrational for you to give answers that significantly diverge from that average.

Balance (Formal)

The requirement that your prior credence in a proposition and your prior expectation of your posterior credence in a proposition match, can be formalized thus:

Definition

π **balances** R on a proposition q if

$$\pi(q) = \mathbb{E}_\pi[R(q)]$$

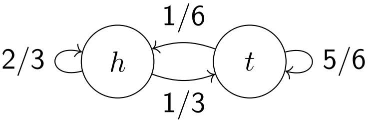
where

$$\mathbb{E}_\pi[R(q)] =_{\mathsf{def}} \sum_{w \in \Omega} \pi(w) \cdot \rho_w(q).$$

(This is also called “Reflection”, “Martingale”, and other names)

Unbalanced Updating on *One Coin*

Suppose that when the coin came up *tails*, you got *slightly stronger* evidence about the true outcome, warranting increasing your credence in t to $5/6$.



$$\left(\begin{array}{c|cc} & h & t \\ \hline \rho_h & 2/3 & 1/3 \\ \rho_t & 1/6 & 5/6 \end{array}\right)$$

Then we'd have the following expectation:

$$\mathbb{E}_\pi[R(h)] = \pi(h) \cdot \rho_h(h) + \pi(t) \cdot \rho_t(h) = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{6} = \frac{5}{12} < \frac{1}{2},$$

so over a large number of trials, you'd think that only around 40% of the coins came up heads, even though you get no evidence that the coin is tails-biased.

Life Out of Balance

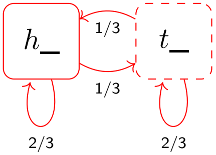
Why think balance should be a rational requirement on updating?

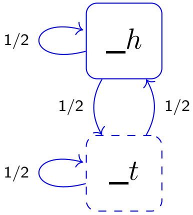
1. In long-run repeated trials like *The Greatest Weighting*, you can expect to have estimates consistent with *different* chances than you now think are the right chances. (Like thinking you'll estimate a fair coin to be biased).
2. Salow (2018) argues that anytime you have a balance failure, you can engage in *intentionally-biased inquiry*. You can set up inquiries in a way that guarantee you only learn favorable information.
3. Dorst (2023) argues that when agents have balance failures in different directions, they can *predictably* polarize: they can predict that their opinions will diverge on questions even if they know the same things about the ground truth.

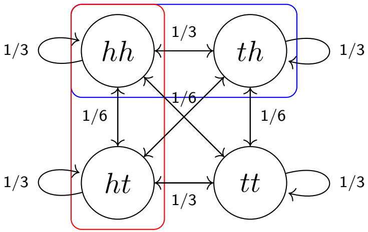
§3 Higher-Order Uncertain priors with Two Coins

Two Coins

A second coin (blue) is flipped to the right of the first (red). You receive no further evidence about the second coin.







Two Coins (Matrix Form)

A second coin is flipped to the right of the first. You receive no further evidence about the second coin.

$$\pi = \left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \pi & 1/4 & 1/4 & 1/4 & 1/4 \end{array} \right)$$

$$R = \left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho_{w_1} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{w_2} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{w_3} & 1/6 & 1/6 & 1/3 & 1/3 \\ \rho_{w_4} & 1/6 & 1/6 & 1/3 & 1/3 \end{array} \right)$$

Balanced Updating in *Two Coins*

$$\pi = \left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \pi & 1/4 & 1/4 & 1/4 & 1/4 \end{array}\right) \qquad R = \left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho_{w_1} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{w_2} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{w_3} & 1/6 & 1/6 & 1/3 & 1/3 \\ \rho_{w_4} & 1/6 & 1/6 & 1/3 & 1/3 \end{array}\right)$$

Here π balances R for every proposition: for any q,

$$\pi(q) = \mathbb{E}_\pi[R(q)].$$

So *on average*, π still thinks that R will predict that hh, ht, th, tt come up in approximately equal proportion in the long run.[1]

[1] It turns out that we can check this in the following way: π balances R on every proposition iff $\pi R = \pi$ when we interpret π as a row vector and R as a matrix.

Two Coins, Two Updates

Now you're told whether or not *both* coins came up heads. That is, you're told the true cell of the partition $\{\{hh\}, \{ht, th, tt\}\}$.

You learn this with certainty, so you conditionalize.

Supposing that the actual world is *ht*:

$$\left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho^{@}_{ht} & 1/3 & 1/3 & 1/6 & 1/6 \end{array} \right) \to \left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho^{+@}_{ht} & 0 & 1/2 & 1/4 & 1/4 \end{array} \right)$$

Representing Updating on Uncertain Priors

However, before conditionalizing, your *prior* is higher-order uncertain: you are not certain that you are at world ht and have ρ_{ht}. From your perspective, before you are told whether both coins came up heads, you leave all the worlds, and thus all of $\rho_{hh}, \rho_{ht}, \rho_{th}, \rho_{tt}$, open.

Thus you have *four* candidate credence functions that could have been the one you should update (and you don't know where the @ symbol is either):

$$\left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho_{hh} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{ht}^{@} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{th} & 1/6 & 1/6 & 1/3 & 1/3 \\ \rho_{tt} & 1/6 & 1/6 & 1/3 & 1/3 \end{array}\right)$$

Representing Updating on Uncertain Priors

So from your perspective, I claim, you should represent the update like this:

$$\left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho_{hh} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{ht}^{@} & 1/3 & 1/3 & 1/6 & 1/6 \\ \rho_{th} & 1/6 & 1/6 & 1/3 & 1/3 \\ \rho_{tt} & 1/6 & 1/6 & 1/3 & 1/3 \end{array}\right) \to \left(\begin{array}{c|cccc} & hh & ht & th & tt \\ \hline \rho_{hh}^{+} & 1 & 0 & 0 & 0 \\ \rho_{ht}^{+@} & 0 & 1/2 & 1/4 & 1/4 \\ \rho_{th}^{+} & 0 & 1/5 & 2/5 & 2/5 \\ \rho_{th}^{+} & 0 & 1/5 & 2/5 & 2/5 \end{array}\right)$$

where each ρ_w is conditionalized on the true cell of the partition $\{\{hh\}, \{ht, th, tt\}\}$.

§4. The Problem for Balance

Not Very Balanced

For families R, R^+, say that R balances R^+ if every ρ in the range of R balances R^+.

- In *Two Coins*, R does not balance R^+.
- *Also,* no matter which world is actual, the prior ρ does not balance R^+.
- *Also,* the original prior π does not balance R^+.

Not Very Balanced

Recap: π balances R, but R does not balance R^+, and π also does not balance R^+.

We thus lack forms of balance that we should want, if we were fans of balance:

1. If we don't want updates to bias/unbalance us, then we also don't want *further* updates to unbalance us. So we should want π to balance R^+.
2. For each ρ in R, we should want it to balance R, even if there is higher-order uncertainty: otherwise, there would always be some world where the agent expected to form a long-run bias.

Where Did We Go Wrong?

If these forms of balance are rational requirements, then one of these updates, $\pi \to R$ or $R \to R^+$, must be irrational.

- But surely it isn't $R \to R^+$: conditionalization on veridical, partitional evidence.
- And if $\pi \to R$ is not rational, then the prospects of rational higher-order uncertainty at all might be grim, given how simple a case it is.

The Problem is General

Two Coins is not an edge case. *Whenever* some family R has higher-order uncertainty, then, given some minimal conditions,[2] there will be some way to conditionalize R on partitional evidence into an R^+ such that some R does not balance R^+.

Theorem

Let Ω have at least three possible worlds and let R be a higher-order uncertain, self-confident[a] credence family. Then there is some partition Π of Ω with at least two nonempty cells such that R does not balance the credence family R^+ obtained by conditionalizing each R_w on the partition cell $c_w \in \Pi$ that contains w.

[a] A credence family R is **self-confident** if $\forall w \in \Omega$, $R_w(w) > 0$. I take this to be a rather minimal condition (strictly less demanding than regularity).

[2] The first being that there are at least three possible worlds; the second being self-confidence (below).

Options

Upshot: whenever an agent has higher-order uncertainty, they are always at risk of being in a state where even the update rule "conditionalize on partitional evidence" leads to expectable biases.

Options:

1. Give up higher-order uncertainty (Schoenfield 2017, standard Bayesians)
2. Give up balance (Dorst 2023, anyone who thinks you should conditionalize on nontransparent evidence, like Zendejas Medina 2024, Meehan & Zhang 2025)
3. Restrict your setting (Gallow 2021, Isaacs and Russell 2022)
4. Worry that something has gone wrong with this *framework* for higher-order uncertainty (me, maybe?)
5. Poke holes in my argument (Gallow, in one minute)

Replies

Higher-Order Uncertainty: Ideal or Nonideal?

Question: Is higher-order uncertainty possible for epistemic agents who update *infallibly*, in that we can assume that if an update rule prescribes some credences c in some situation s, the agent always forms credences c in situation s?[3] Or is higher-order uncertainty only a phenomenon for *fallible* epistemic agents who only imperfectly follow update rules?[4]

- $\rightarrow$ Those who think only fallible agents can be rationally higher-order uncertain perhaps have a more limited theory of higher-order uncertainty[5].
- $\rightarrow$ But at least they have an easier way out of (at least part of) my challenges.

[3] Williamson (2011), Dorst (2023), Zendejas Medina (2024), Meehan and Zhang (2025)
[4] Gallow (2021), Isaacs and Russell (2022), Schultheis (*forth*)
[5] e.g. Gallow's and Isaacs and Russell's theories reduce to Schoenfield's super-internalist conditionalization rule for ideal agents

Dmitri's Theory Stays Balanced, at least from the prior

Kevin Dorst would evaluate the expectation assuming agents are infallible: in each world, the agent deterministically adopts some credence function.

$$R : w \mapsto \rho \tag{1}$$

Dmitri would evaluate the expectation assuming agents are fallible: in each world, the agent adopts one of several possible credence functions, each with some probability:

$$RG : w \mapsto \mathbb{P}(\mathbb{C}r(\Omega)) \tag{2}$$

So they will have different estimations of how often the posterior *actually* ends up with particular credence functions.

Demonstration

In *Two Coins*, Kevin's:

$$\begin{aligned} \pi([R^+(h\_) = 1]) &= 1/4 \\ \pi([R^+(h\_) = \tfrac{1}{2}]) &= 1/4 \qquad \text{so } \mathbb{E}_\pi[R^+(h\_)] = 0.475 \neq 0.5 = \pi(h\_). \\ \pi([R(h\_) = \tfrac{1}{5}]) &= 1/2 \end{aligned}$$

Dmitri's:

$$\begin{aligned} \pi([RG^+(h\_) = 1]) &= 1/4 \\ \pi([RG^+(h\_) = \tfrac{1}{2}]) &= 1/3 \qquad \text{so } \mathbb{E}_\pi[RG^+(h\_)] = 0.5 = \pi(h\_). \\ \pi([RG^+(h\_) = \tfrac{1}{5}]) &= 5/12 \end{aligned}$$

At least when it comes to balance by the *initial* prior, Dmitri need not be worried.

Balance from Higher-Order Uncertain Priors

So if we want balance from an original, higher-order certain prior, Dmitri's framework can get us that, whatever worries we might have about its restricted scope.

But if we want all the ρ in the higher-order uncertain R (after the vague evidence but before learning about *hh* or $\neg$*hh*) to balance R^+, Dmitri's framework doesn't itself get us there.

This is what the rigid/nonrigid stuff is supposed to fix.

The Actual Disposition?

Dmitri says that I use the following expectation, where I take the expectation, according to the *actual* prior, of the rational posteriors resulting from the rational learning dispositions *definitely described*:

$$\mathbb{E}_{\pi_@}[R_w^w(q)],$$

and that I should instead use the following expectation, where I take the expectation, according to the *actual* prior, of the rational posteriors resulting from *actual* rational learning dispositions:

$$\mathbb{E}_{\pi_@}[R_w^@(q)].$$

My response: I *am* using the actual rational learning disposition, because the rational learning disposition doesn't vary over worlds, only its output does.

Details

Dispositions: functions from worlds to (probability distributions over) posteriors. In *Two Coins*, there's no fallibility about which of *hh*, ¬*hh* you learn, so this disposition is equivalent to a credence family. The rational disposition deterministically gives the following posterior at any world w: it conditions the rational prior at w on the true cell of the partition $hh/\neg hh$ at w:

$$R_w(q) = \begin{cases} \rho_w(q \mid hh) & w \in hh \\ \rho_w(q \mid \neg hh) & w \in \neg hh \end{cases}$$

DO NOT ENTER

4.2 Higher-Order Uncertain Priors

Higher-Order Uncertain Priors

We can interpret P^+ as itself a prior that could be further updated into a P^{++}. For instance, suppose you gain the ambiguous evidence about the first coin, updating from π to P^{++}, and then gain evidence about whether w_1 is true, thus updating into P^{++}.

Definition

Where $P, R : W \to \mathbb{P}(W)$ are credence families and $P[W]$ is the image of W under P, say that P **balances** R if for any variable X,

$$\forall \pi \in P[W] : \mathbb{E}_\pi (X) = \mathbb{E}_\pi (\mathbb{E}_R(X)) \,. \qquad \text{(Balance (for Families))}$$

Question: when does P^+ balance P^{++}?
Answer: only in a very restricted set of circumstances.

The Double Balancing Act

Theorem

If P, R are credence families, P balances R only if there is a partition $\Pi = \{T, C_1, C_2, \ldots\} \subseteq \mathcal{P}(\Omega)$ of propositions where,

1. for all $C \in \Pi$, for all $w, w' \in C, P_w = P_{w'}$.
2. for all $C \in \Pi$, for all $w \in C, R_w(C) = 1$, and thus $R_w([R_w \in R[C]]) = 1$.
3. for all $\pi \in P[\Omega]$, $\pi(T) = 0$.

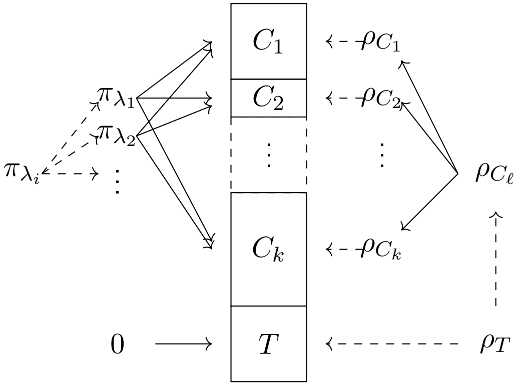
Informally, there must be a partition of conditions where,

1. Within each condition, P is constant, and thus higher-order certain. In other words, *conditional* on any condition C in the partition, P is always certain about the actual value of P_w, even though P may not know which condition is the actual condition
2. R is always certain about which condition is the actual condition.
3. There is a set of 'pathological' possibilities that P is certain will never obtain.

In this sense, the higher-order uncertainty in P and R must be 'orthogonal.'

Balancing Higher-Order Uncertain Posteriors with Higher-Order Uncertain Priors

I find the condition quite hard to visualize, but here is an attempt:



5. On Balance

Consequences for Rational Updating

1. From §4.1: When we're considering a single prior credence function and a higher-order uncertain posterior credence family, there may be no way for the prior to think the updating will be unbiased conditional on different hypotheses.
2. From §4.2: When we're considering both a higher-order uncertain posterior credence family and a higher-order uncertain prior credence family, there may be no way for the prior to even *unconditionally* balance the posterior. So updating into a higher-order uncertain state can mean giving up balance in the future.

Upshot: balance is really hard to preserve in conditions of higher-order uncertainty.

Consequences for Higher-Order Uncertainty

There may really be a tradeoff between balance and higher-order uncertainty. If the latter is rationally permissible, it's hard to see how the former, in desirable generalities, could be rationally required.

$\rightarrow$ Dorst (2023), Williamson (1997), Zendejas Medina (2024): keep higher-order uncertainty, do away with balance.

$\rightarrow$ Gallow (2021), Isaacs and Russell (2023, implicitly): keep both. (I think this talk poses problems for their position).

$\rightarrow$ Me, tentatively: Keep balance, and thus *do away with higher-order uncertainty*. Or, at least: find a *different way* to model it.