Bayesian Ambiguity: Imprecision or Uncertainty?

Kevin Dorst
Massachusetts Institute of Technology

October 2023

Abstract

I introduce a variety of examples and features of ambiguous judgments, the basics of standard Bayesian models, and argue that they can't properly account for it. Does that mean we must we move to *imprecise* Bayesian models? No. Due to noise in our cognitive systems, we should expect people to have *higher-order uncertainty* about what their opinions are. I show how to model this, and argue that it respects the features of ambiguous judgments better than the imprecise model.

1 Clarity and Ambiguity

Sometimes, our subjective assessments are *clear*: we know exactly which of two things is more likely, or which of two things we want more. Other times, they are *ambiguous*: we're unsure which is more likely, or which we want more—perhaps there's not even a fact of the mater. Here are some cases:

> **Dice vs. Socks**.
> *Clear:* This is a fair die; I'm going to roll it 30 times. What's your best estimate for the number of times it'll land between 1–4? That's clear: $\frac{4}{6} \cdot 30 = 20$ times. You're far from sure that this estimate is *accurate*—there's a very good chance the number is higher or lower. But you are (or should be) sure this is the rational estimate for you to have—if your peer said 21 or 19, you'd think they were confused or misinformed. By the same token, you definitely prefer to bet that it'll land 1–4 at least 19 times than that it won't.
> *Ambiguous:* I'm a runner, in my 30s, and an academic. What's your best estimate for the number of pairs of socks I own? That's ambiguous: if forced to pick, you'd no doubt come up with a number—say, 20 pairs. But not only are you far from sure that this estimate is *accurate*, you're also far from sure that it is the right (rational) estimate for you to have—you could just as well have said 21 or 19, and you wouldn't look askance at someone who did. By the same token, it's unclear whether you prefer to bet I have at least 19 pairs than that I don't.

People vs. knees

Clear: Look around you. What's the largest n such that you're sure that there are at least n people in the room? Now, this *might* be hard—if you're reading this in a hotel lobby, or a football stadium, it will be. But for those of us sitting in offices or seminar rooms, it's clear enough—5, as the case may be.

Ambiguous: Look down. What's the largest n such that you're sure that the top of your knee is at least n inches above the floor? Don't do any calculations—just try to figure it out based on how it looks. That's ambiguous. I'm sure it's at least 20 inches. I'm probably sure it's at least 25. Definitely not 30. But where does my certainty give out? It's hard to say.

Chaotic vs. enigmatic urns

Clear: I have an urn that was filled with a complicated rube-goldberg machine: a matchstick fell on a mousetrap which triggered several balls to chaotically knock down some (but not all) of the walls holding back marbles—some red, others blue. You have no idea how this works, but you know Caspar has studied it in detail, is vigilant and reasonable, and has observed the process closely many, many times before. He says he's precisely 60% confident that when I draw a marble out of the urn, it'll be red. How confident are you that it'll be red? That's clear: 60%.

Ambiguous: Here's another urn. It has 10 marbles in it total. Some are red, the rest are blue. By the way: I like the color blue. I'm about to draw a marble. How confident are you that the one I'll draw is red? That's ambiguous: since you have so little to go on, maybe 50%? But the previous urn likely had more red marbles, so maybe you are (or should be) closer to 55% or 60%? On the other hand, I just told you I like blue, so maybe that should skew you down to 45%. On the *other* other hand, maybe I'm messing with you, so that you should think red is *more* likely since I told you I like blue. But on the *other* other other hand...

Conflicting tosses vs. testimony

Clear: I rolled an 11-sided die to determine which coin to grab. They have differing biases: one of them lands heads 0% of the time, another 10%, another 20%,... and the final one 100% of the time. I've tossed it twice. On the first toss, it landed heads. On the second, it landed tails. How confident are you that it'll land heads on the next toss? That's clear: 50%. The conflicting bits of evidence perfectly balance out.

Ambiguous: You know that Roger's in his office on 50% of weekdays. You've asked two people about it. Caspar said he'd talked to him, and that Roger won't be in tomorrow. Kevin said he'd talked to him, and that Roger *would* be in tomorrow. When you told each of us about what the other said, we both simply shrugged their shoulders and said, "Huh, weird." How confident are you that Roger will be in his office tomorrow? That's unclear: maybe 50%—maybe your amount of trust in each of us balances our *perfectly.* But maybe you should trust Kevin a bit more (he's more fastidious) and so be 60%. Or maybe you should trust Caspar a bit more (he's known Roger for longer) and so be 40%.

Searching for marbles vs. words

Clear: I have two urns; one contains 2 red marbles, the other contains 1 black and 1 red marble. I flipped a coin to determine which I'm holding. You reach in, pull a marble out, and discover that it's red—you failed to find a black marble. How confident are you that the urn contains a black marble? That's clear: $\frac{1}{3}$ confident.[1]

Ambiguous: I have two word-search tasks—strings of letters with some blanks. One is *completable* (by an English word), the other is not. I flipped a coin to determine which to show you. Here it is—you have 5 seconds to try to find a word: ST_ _RE. One. Two. Three. Four. Five. Time! Supposing you failed to find a word, how confident are you there's a word that completes the string? That's unclear. Maybe 30%? Maybe 40% (maybe you're bad at this, so failing to find one isn't much evidence). Maybe 20% (you should have expected to find one, so failing to find one is strong evidence).[2]

These cases illustrate different sources of clarity or ambiguity. Dice are clear because you have knowledge of the chances; socks are not because there's so much disparate, weak evidence to go on, and you reasonably wonder whether you know anything else that's relevant. The number of people in the room is clear because it's discrete and you can count; the height of your knee is ambiguous because of the vagueness or fuzziness of your own perception. The chaotic urn is clear because you have a clear expert opinion to defer to, while the enigmatic urn is ambiguous because you have a *little* bit to go on, and reasonably wonder what to do with it. The conflicting coin tosses are clear because (again) you have chance information, while the conflicting testimony is unclear because you're not sure how to weigh my claim against Caspar's. Failing to find a black marble provides clear (though not definitive) evidence because you know the chances, while failing to find a word provides ambiguous evidence because you don't know how likely you did (or should) have thought you were to find it if there is one (not to mention the fact that there are lots of subtle bits of evidence you got beyond $\neg$*find*, e.g. how word-like it seemed and how many failed attempts you managed).

Despite these differences, across these cases, there's a common contrast between your attitudes in the clear version and the ambiguous version. Let's note some of its features:

Irresolute assessments. In the ambiguous cases, your attitudes are less robust or resolute than in the clear ones. This shows up as *stochasticity:* There is some randomness (and arbitrariness) in your judgment in the ambiguous cases. If I'd asked you in different circumstances, there's a good chance your estimate for my number of socks (or height of your knee, or probability of a red marble, or...) would've been a

[1] $P(\textit{1 red}|\textit{red}) = \frac{P(\textit{1 red})\cdot P(\textit{red}|\textit{1 red})}{P(\textit{1 red})\cdot P(\textit{red}|\textit{1 red})+P(\textit{2 red})\cdot P(\textit{red}|\textit{2 red})} = \frac{0.5\cdot 0.5}{0.5\cdot 0.5+0.5\cdot 1} = \frac{1}{3}$.

[2] $P(\textit{word}|\neg\textit{find}) = \frac{P(\textit{word})\cdot P(\neg\textit{find}|\textit{word})}{P(\textit{word})\cdot P(\neg\textit{find}|\textit{word})+P(\neg\textit{word})\cdot P(\neg\textit{find}|\neg\textit{word})}$. But what is $P(\neg\textit{find}|\textit{word})$?

bit higher or a bit lower. Not so in the clear cases: excepting situations of serious confusion, when faced with the question about die rolls you'll answer '20 times' every time.

Insensitivity. In the ambiguous cases, your comparative-confidence judgments (and resulting preferences amongst bets) seem insensitive to mild sweetening. Let *Abe's-Run* be the proposition that this penny I'm holding will land heads 10 times in a row (probability $= \frac{1}{2^{10}} = \frac{1}{1024}$). You're clearly more confident in *completable* *or* *Abe's-Run* than in *completable*, but it's not clear whether you're more confident in *completable* than in *1-red*, nor vice versa.

Fuzzy boundaries. When your judgments are ambiguous, it's usually unclear where the ambiguity gives out. We ran a fair lottery with 1000 people in it; let a_i be the claim that the ith person won, so $P(a_i) = \frac{1}{1000}$. Clearly you're more confident of $a_1 \vee a_2 \vee ... \vee a_1000$ (think of this as A++++...) than you are that I have at least 20 pairs of socks (q). Clearly you're *less* confident of a_1 than you are that I have at least 20 pairs of socks. But what's the first n such that you're more confident of $a_1 \vee a_2 \vee ... a_n$ than of q? It's hard to say.

Ambiguity aversion. You can choose between three bets:

Bet1: \$100 if *heads*, \$0 if *tails*.
Bet2: \$100 if q (I have at least 20 pairs of socks), \$0 if $\neg q$.
Bet3: \$100 if $\neg q$, \$0 if q.

Which would you prefer? If the ambiguous judgment q is chosen correctly (so that people are close to 50%-confident of it), most people prefer Bet1—they tend to be *ambiguity averse*, preferring to bet on claims for which they have clear judgments than on ones for which their opinions are ambiguous (Ellsberg 1961; Halevy 2007; Kovářík et al. 2016).

Biases and miscalibration. Suppose we repeat the 30-die-rolls hundreds of times. How confident are you that the average number of 1–4 rolls will be around 20? Pretty confident, I imagine.

Now suppose we have you estimate all sorts of things like the socks: How many fish has Caspar owned throughout his life? How many chairs has Roger bought? How many times has Sally moved? Etc. Gather up all the instances of these cases in which (as with my socks) you estimate the number to be 20. How confident are you that the average number across all these cases will be around 20? Pretty doubtful, I imagine.

Empirically, people are much more likely to exhibit various forms of biased processing of evidence when the evidence is contested, mixed, or complex (i.e. ambiguous). One result of this is *miscalibration*: of all the quantities they (ambiguously) estimate to be (say) 20, the average of these quantities if often far from 20.[3] A related finding is

[3] E.g. Lichtenstein et al. 1982; Glaser and Weber 2010; Moore et al. 2015; Peterson and Beach 1967; Koehler et al. 2002; Tetlock 2009; Brenner et al. 2005; Ortoleva and Snowberg 2015; Dorst 2023a.

selective sensitivity: when new evidence is hard to interpret, people are much more likely to let their prior beliefs or motivations influence how they interpret it.[4]

What's the best way to model, explain, and (maybe) rationalize these features of ambiguous judgments?

There are two main approaches. The *imprecise* model says that rather than modeling your beliefs and values with a single probability and utility function, we should instead model them with a *set* of probability and a *set* of utility functions—the ones that are consistent precisifications of your imprecise state. The *higher-order* model says that we can still model your beliefs and values with a single probability and utility function—all we need to is acknowledge that often you are unsure what you believe, i.e. you have a probability function which is unsure which probability function you have. I'll make a case for the higher-order model. (See Carr 2021 for a related but distinct set of arguments.)

2 Bayesian Basics

We *can't* properly model ambiguous judgments using the standard Bayesian machinery—namely, that rational people's beliefs can be represented with a single, precise probability function, their desires can be represented with a precise utility function, and they always act so as to maximize expected utility. In this section I'll say more about what this standard picture looks like, as an [Argh! More hairy than I'd hoped] set up for what's to come.

2.1 Total probability, Reflection, and Partitional Updating

Fix a set of epistemically possibly worlds, W—for simplicity, let's suppose W is finite. We'll understand it as the set of finest-grained possibilities the agent we're modeling leaves open—we can think of it as the complete answer to a (set of) question(s) (Hamblin 1976). These worlds should be fine-enough grained to settle any relevant question that the agent is unsure of. A *proposition* (or an *event* or a *claim*) is a way the world might be—a subset of $q \subseteq W$, picking out the ways q could be true. We can handle logic with set theory: q is true at world w iff $w \in q$; $\neg q$ is the complement of q in W, i.e. $W - q = \{w \in W : w \notin q\}$; given two propositions q_1 and q_2, $q_1 \& q_2$ is the intersection $q_1 \cap q_2$, $q_1 \vee q_2$ is the union $q_1 \cup q_2$, the material conditional $q_1 \to q_2$ is $\neg q_1 \cup q_2$, and so on.

A probability function π over W is an assignment of non-negative numbers to the $w_i \in W$ that sum to 1. Ordering our worlds in some canonical order $w_1, ..., w_n$, we can think of a probability function as a vector $\pi = (\pi_1, \pi_2, ..., \pi_n)$ where $\pi_i = \pi(w_i)$ is the probability assigned to world i. This gives us the probabilities of every maximally-specific proposition; other probabilities are gotten by summing over the worlds in the relevant proposition: $\pi(q) = \sum_{w \in q} \pi(w)$.

[4] E.g. Lord et al. 1979; Kelly 2008; Kahan et al. 2017; Kunda 1990; Nickerson 1998; Mercier and Sperber 2011; Ditto et al. 2019.

For instance, suppose you know the coin I'm holding is either $\frac{2}{3}$-biased toward heads or $\frac{2}{3}$ biased toward tails, and you're currently 50-50 between those possibilities. The question W is *how will it land when tossed, and is it biased?*. Let h and t be the propositions that it lands heads and tails, and let b_h and b_t be the propositions that it's biased toward heads and tails, respectively. Let's order them like so: $(h\&b_h, h\&b_t, t\&b_h, t\&b_t)$. I'll sometimes drop the & and simply smoosh conjunctions together, so we can more compactly write: (hb_h, hb_t, tb_h, tb_t). A probability function over these possibilities is simply a vector over them—for instance, you probably think it's $\frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$ likely to be biased heads *and* land heads, while it's $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$-likely to be biased heads and land tails, etc. Then your distribution over (hb_h, hb_t, tb_h, tb_t) is $\pi = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3})$. Thus, summing over the h-possibilities, you are $\frac{1}{3} + \frac{1}{6} = \frac{1}{2}$-confident that it'll land heads.

Fix an agent at a given time—say you, now. You are *probabilistic* only if your comparative confidence can be represented with a single, precise probability function π: for all q, p, you're more confident of q than p iff $\pi(q) > \pi(p)$. This implies that your comparative confidences form a total order over the propositions $q \subseteq W$ about W: for any p, q, either you're more confident of one than the other, or you're equally confident in them.

In addition to comparative confidences (and estimates), you have *conditional* comparative confidences (and estimates). You're more confident that p this fair die will land 1–4 than that q it'll land 5–6: $\pi(p) > \pi(q)$. But *conditional on* it landing 4–6 (r), you're more confident it'll land $5-6$ than $1-4$: $\pi(p|r) < \pi(q|r)$. This is a synchronic feature of your mental state now, not necessarily tied to what happens if you learned r. (Maybe learning r would send you into a mental breakdown and disrupt all your comparative confidences.) For a probability function, conditional probabilities are given by the ratio formula: $\pi(q|r) = \frac{\pi(q \cap r)}{\pi(r)}$. (When $\pi(r) > 0$. When it equals 0, things get messy. Ignore the mess.) In our vector notation, that means that conditioning on r is equivalent to zeroing out the $\neg r$-possibilities and "renormalizing" so that the entries in the new vector sum to 1.

For example, if you condition your above distribution on the coin landing heads, you start with $\pi = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3})$, zero out to get $(\frac{1}{3}, \frac{1}{6}, 0, 0)$, and then renormalize by dividing each entry by the remaining total (which equals $\frac{1}{3} + \frac{1}{6} = \frac{1}{2} = \pi(h)$) to get $\pi(\cdot|h) = (\frac{2}{3}, \frac{1}{3}, 0, 0)$. Thus after conditioning on the coin landing heads, you become $\frac{2}{3}$-confident that it's biased toward heads. (Notice: since a zeroed-out-and-renormalized vector is still a probability vector, it follows that for any r, $\pi(\cdot|r)$ is a probability function as well, so follows all the same theorems.)

Let's add to probabilism that your comparative *conditional* confidences can also be represented by a probability function: there's a π such that for all p, q, r: you're more confident of p given r than q given r iff $\pi(p|r) > \pi(q|r)$. That implies that your conditional comparative confidences form a total order.

Now, it shouldn't be called Bayesianism. It's called that because of Bayes theorem, which is important. Sometimes. Sort of. Whatever. What it *should* be called is Total-Probabilityism. Doesn't roll off the tongue, I know, but even Bayes theorem would be useless if not for the theorem of total probability. What's it say?

A **question** is a partition of W—i.e. a way of dividing logical space into mutually exclusive (only one of them can be true) and collectively exhaustive (at least one of them is true) partition-cells. The partition-cells can be thought of as complete answers to the question. *Who stole the cookies from the cookie jar?* is the partition $\{$*Caspar stole the cookies*, *Kevin stole the cookies,...*, *Bob stole the cookies*$\}$. (Since W is finite, every question will have finitely many answers.) *How many cookies were stolen?* is the partition $\{$*1 cookie was stolen*, *2 cookies were stolen,...,1032 cookies were stolen*$\}$. (It's a big jar.)

The theorem of total probability says the following. Take any proposition p. Take any question $\mathcal{Q} = \{q_1, ..., q_n\}$. The probability of p is the *(weighted) average conditional probability* of p given the true answer to $\mathcal{Q}$, with weights determined by those answers' probability:

$$\textbf{Total Probability:}\quad \pi(p) = \sum_{q_i \in \mathcal{Q}} \pi(q_i) \cdot \pi(p|q_i) \quad = \quad \pi(q_1) \cdot \pi(p|q_1) + \cdots + \pi(q_n) \cdot \pi(p|q_n)$$

Total probability is useful because if we choose our question wisely, it lets us decompose inscrutable probabilities into ones that are obvious. We've narrowed it down to Caspar and Kevin. Caspar is wily—he would never steal more than one cookie (too noticeable). Kevin is not—he needs lots of cookies to fuel that running addiction. So $\pi(\textit{Caspar}|\textit{1 cookie}) = 0.5$ while $\pi(\textit{Caspar}|\textit{2 cookies}) = 0$; meanwhile, $\pi(\textit{Kevin}|\textit{1 cookie}) = 0.5$ and $\pi(\textit{Kevin}|\textit{2 cookies}) = 1$. We're 90% confident that 2 cookies were stolen, 10% that only 1 was. So we're pretty confident that Kevin was the culprit: $\pi(\textit{Kevin}) = \pi(\textit{1 cookie}) \cdot \pi(\textit{Kevin}|\textit{1 cookie}) + \pi(\textit{2 cookies}) \cdot \pi(\textit{Kevin}|\textit{2 cookies}) = 0.1 \cdot 0.5 + 0.9 \cdot 1 = 0.95$.

Now let $\pi^{\mathcal{Q}}$ be the probability function that results from conditioning π on the true answer to $\mathcal{Q}$. (This is sometimes written $\pi(\cdot|\mathcal{Q})$—note that it's a partition in the conditioning spot there, rather than a proposition.) Notice that, mathematically, $\pi^{\mathcal{Q}}$ is not simply a probability function (an assignment of numbers to propositions), but instead is a function from worlds $w \in W$ to probability functions $\pi^{\mathcal{Q}}_w$ defined over W. After all, we need to update π in different ways in different worlds: for $w \in$ *1 cookie*, $\pi^{\mathcal{Q}}_w(\cdot) = \pi(\cdot|\textit{1 cookie})$, while in $w' \in$ *2 cookies*, $\pi^{\mathcal{Q}}_{w'}(\cdot) = \pi(\cdot|\textit{2 cookies})$.[5] So defined, it follows immediately from total-probability that $\pi(p)$ is a probability-weighted average of $\pi^{\mathcal{Q}}(p)$. Letting $\pi^{\mathcal{Q}}_q$ be the value $\pi^{\mathcal{Q}}$ takes in any q-world (for $q \in \mathcal{Q}$):

> **Question-Reflection (average):** $\pi(p) \;=\; \pi(q_1) \cdot \pi^{\mathcal{Q}}_{q_1}(p) + \cdots + \pi(q_n) \cdot \pi^{\mathcal{Q}}_{q_n}(p)$
>
> The probability of p equals a probability-weighted average of the conditional probability of p given the various possible (complete) answers to $\mathcal{Q}$.

[5] In the confusing lingo of statisticians, P is a "random" probability function; in the confusing lingo of the philosophers of language, 'P' is a *definite description* for a probability function, while 'π' is a rigid designator for a particular probability function. (Analogy: 'the tallest person in the room' vs. 'Josh'.)

This is a simple but profound claim: your opinion should always be an average of the opinions you'd have conditional on the various possible answers to some (any!) question. For example—since for any E, $\{E, \neg E\}$ is a question—if E ie evidence for p ($\pi(p|E) > \pi(p)$), then $\neg E$ is evidence against it ($\pi(p|\neg E) < \pi(p)$). (So—contrary the famous quote—absence of evidence usually *is* evidence of absence.) Or: you can't know ahead of time that the answer to a question would cause you to raise your probability of p—for if you did, you should've already raised your probability. (It can't be that for all $q \in \mathcal{Q}$, $\pi(p|q) > \pi(p)$.) Or: if you're really confident that the answer to $\mathcal{Q}$ would raise your probability for p, then there must be some chance that it'll lower it a lot. (E.g. if $\pi(p) = 0.5$ and $\pi(P(p) \geq 0.6) \geq 0.8$, then $\pi(P(p) \leq 0.1) \geq 0.2$, say.)

Relatedly, it follows that π *defers* to the opinions that would result from conditioning it on the true answer to $\mathcal{Q}$, in the sense that it's conditionally disposed to adopt them. Where 'δ' is (like π) a variable over probability functions, let $\pi^{\mathcal{Q}} = \delta$ be the claim that we're at a world w where $\pi^{\mathcal{Q}}_w = \delta$, i.e. $\{w \in W : \pi^{\mathcal{Q}}_w = \delta\}$.[6] Then it follows that:

> **Question-Reflection (global):** $\pi(\cdot|\pi^{\mathcal{Q}} = \delta) = \delta$
> The probability of any given p, conditional on the $\mathcal{Q}$-updated probability function equalling δ, equals δ.
>
> **Question-Reflection (local):** $\pi(p|\pi^{\mathcal{Q}}(p) = t) = t$
> The probability of p, conditional on the $\mathcal{Q}$-updated-probability of p being t, equals t.

So far this has all been synchronic, considering a particular probability function π at a given time and how it (at that time) is disposed to revise when conditioned on various answers to a question. In other words: all of the above are theorems of synchronic probability. But these principles are closely related to what usually go under the heading of 'Reflection' or 'Deference' principles, and involve similar relationships between a given probability function π and another descriptively-specified probability function P—for example, 'the probabilities you'll have tomorrow' or 'Caspar's probability function (today)', or 'the objective chances'.[7]

Some notational niceties are important here. I'm going to continue using lowercase Greek letters ($\pi, \delta, \rho ...$) as rigid designators for particular probability functions. But we need other ways to pick out probability functions. After all, 'your credence function (now') is something that I (or even—as we'll see—*you*) can be uncertain about. And if we're to model uncertainty about a quantity, we need to introduce a way for it to vary between worlds. Thus I'll use upper-case Roman letters ($\mathsf{P}, P, Q, R ...$) as *descriptions* for probability functions—formally they are functions from worlds w to a particular probability function P_w defined over W. When subscripted with a world, P_w, P_w, $R_{w'}$, etc. are again rigid designators for (e.g.) 'the probability function P picks out at w'.

[6] Two probability functions δ and ρ are equal when they assign the same values to all propositions: for all $q \subseteq W$, $\delta(q) = \rho(q)$.

[7] E.g. Lewis 1980; van Fraassen 1984; Skyrms 1980, 1990, 2006; Briggs 2009; Weisberg 2007; Christensen 2010; Elga 2013; Pettigrew and Titelbaum 2014; Salow 2018; Dorst et al. 2021.

Now, since we should acknowledge that all the probability functions we care about can vary from world to world, had better start using descriptions rather than rigid designators. So *your probability function (now)* will be picked out with P. Of course, P picks out different functions in different worlds—in worlds where you're sitting and reading like normal, $\mathsf{P}_w = \pi$ such that $\pi($*purple elephant in front of me*$) \approx 0$, while a world w' where a purple elephant has suddenly appeared in front of you, $\mathsf{P}_{w'} = \pi'$ such that $\pi'($*purple elephant in front of me*$) \geq 0.7$. This variation in your probability function needn't bring in any higher-oder uncertainty: in worlds w where you know exactly what your currently probability function is, that simply means that there is a π such that $\mathsf{P}_w(\mathsf{P} = \pi) = 1$. Note: when P (etc.) is unembedded and no world is specified, it'll usually use P just to pick out the *actual* value of your probability function. (If the actual world is @, then $\mathsf{P} = \mathsf{P}_@$.)

Given that, we can state the normal formulations of deference principles. For example, letting sans-serif P be your current credence function and P be the one you'll have tomorrow, then:

Reflection (global): $\mathsf{P}(\cdot|P = \pi) = \pi$
Conditional on your future self adopting a given probability function, you adopt that probability function.

Reflection (local): $\mathsf{P}(p|P(p) = t) = t$
Conditional on your future self assigning t to p, you assign t to p.

Reflection (average): $\mathsf{P}(p) = \mathsf{P}(P(p) = t_1) \cdot t_1 + \cdots + \mathsf{P}(P(p) = t_n) \cdot t_n$
Your probability for p is a weighted average of the possible future-probabilities you might assign to p.

Reflection is the cornerstone of Bayesian epistemology—it's what gives Bayesianism all the nice features you've heard about. In particular, it drives:

1) The "value of evidence" result that you should expect more evidence to make your beliefs more accurate and your decisions more effective;[8]
2) The "no wishful thinking" (no intentionally-biased-inquiry) result that Bayesians can't intentionally shift their beliefs in a desired direction;[9]
3) The "convergence to the truth" result that given enough shared evidence, Bayesian's priors will "wash out" and they will come to agree on the truth;[10] and
4) The "well-calibrated" result that Bayesians' beliefs can be expected to *calibrated* in the sense that of all the things they're (say) 80%-confident in, 80% will be true.[11]

Nothing we've said so far says whether any of these Reflection principles will hold—that depend, of course, on the relationship between your current probabilities P and (your current probabilities' beliefs about) your future probabilities P. How should these relate?

[8] Blackwell 1953; Good 1967; Ramsey 1990; Oddie 1997; Williamson 2000; Huttegger 2014, 2017.

[9] Kadane et al. 1996; Kamenica and Gentzkow 2011; Kamenica 2019; Salow 2018; Das 2020; Little 2022; Dorst 2023b.

[10] Blackwell and Dubins 1962; Schervish and Seidenfeld 1990; Huttegger 2015; Nielsen and Stewart 2021; Zaffora Blando 2022.

[11] Dawid 1982, 1983; Seidenfeld 1985; Belot 2013; Dorst 2023a.

The standard updating rule of *conditionalization* says that when the total evidence you receive between having P and having P is q, then P should be the result of conditioning P on q:

> **Conditionalization:** When q is the total evidence received, $P(\cdot) = \mathsf{P}(\cdot|q)$.

It's sometimes said that conditionalization implies Reflection. That's not right—or, at least, not without substantive assumptions about what can be your total evidence.

Here's an example. a, b, and c are trying to get on a train, but there's only one ticket left. The conductor has performed a fair lottery to determine who gets the ticket, but hasn't announced the winner yet. You are b, and you'd like to get evidence that you've gotten the ticket. Currently, $\mathsf{P}(b) = \frac{1}{3}$. But you devise a plan. You'll ask the train conductor to tell you one of the other two (a or c) who *didn't* get the ticket. He always speaks truly. If he tells you a didn't get the ticket, you'll condition on $\neg a$ and so jump to $P(b) = \mathsf{P}(b|\neg a) = \frac{\mathsf{P}(b\&\neg a)}{\mathsf{P}(\neg a)} = \frac{\mathsf{P}(b)}{\mathsf{P}(b)+\mathsf{P}(c)} = \frac{1/3}{1/3+1/3} = \frac{1}{2}$-confident that you'll get the ticket. Meanwhile, if he tells you c didn't, you'll condition on $\neg c$ and so jump to $P(b) = \mathsf{P}(b|\neg c) = \frac{1}{2}$-confident you'll get the ticket. Either way, your credence that you'll get the ticket will go up (from $\frac{1}{3}$ to $\frac{1}{2}$). So $\mathsf{P}(b) = \frac{1}{3}$ but $\mathsf{P}(P(b) = \frac{1}{2}) = 1$, violating our Reflection principles. Yet you always updated by conditioning on a truth! What went wrong?

The possible claims you might condition on don't form a partition. To see this, notice that there are four possibilities once we combine the question of who got the ticket and who the conductor tells you *didn't* get the ticket. Letting $[\overline{a}]$ and $[\overline{c}]$ be the propositions that he tells you $\neg a$ and $\neg c$, respectively, they are: $a[\overline{c}]$, $b[\overline{c}]$, $b[\overline{a}]$, $c[\overline{a}]$. (Remember: he'll never tell you that *you* didn't get the ticket.) Supposing you're 50-50 on who he'll tell you didn't get the ticket if *you* got the ticket ($\mathsf{P}([\overline{c}]|b) = \frac{1}{2}$), your prior is the following vector over these possibilities: $\mathsf{P} = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3})$.

Suppose now we want a compact way to write what your probabilities are before (P) and after (P) you ask him, in each world. We can do this with (row)*stochastic matrices.* We'll order the worlds in some canonical way—say, $(a[\overline{c}], b[\overline{c}], b[\overline{a}], c[\overline{a}])$—and then row i column j is the probability that you assign to world j when you're in world i. Thus:

$$\mathsf{P} = \left(\begin{array}{cccc} a[\overline{c}] & b[\overline{c}] & b[\overline{a}] & c[\overline{a}] \\ \hline 1/3 & 1/6 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/6 & 1/3 \end{array}\right) \qquad P = \left(\begin{array}{cccc} a[\overline{c}] & b[\overline{c}] & b[\overline{a}] & c[\overline{a}] \\ \hline 1/2 & 1/4 & 1/4 & 0 \\ 1/2 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 1/4 & 1/2 \\ 0 & 1/4 & 1/4 & 1/2 \end{array}\right)$$

Since P is the same in each row, this implies that your prior probabilities don't vary in any of the worlds we're modeling. Meanwhile, the varying rows in the P-matrix indicate P is different in different worlds. In particular, the first two rows indicate that in $a[\overline{c}]$ and $b[\overline{c}]$—i.e. the worlds where you're told $\neg c$—you update by conditioning on $\neg c$ (zeroing out and renormalizing). Likewise, rows 3 and 4 indicate that when you're told $\neg a$, you update by conditioning on $\neg a$. Notice that in all rows, $P(b) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, so your plan to raise your credence by conditioning will indeed work.[12]

Notice why. The different propositions you might condition on don't form a partition. You might condition on $\neg c = \{a[\overline{c}], b[\overline{c}], b[\overline{a}]\}$, and you might condition on $\neg a = \{b[\overline{c}], b[\overline{a}], c[\overline{a}]\}$, yet these two propositions are consistent (not mutually exclusive). Thus conditioning *can* lead to violations of Reflection—it does so whenever the possible propositions you might condition on don't form a partition.

The standard Bayesian response is to insist that the possible bits of evidence you can receive (the set of propositions you might rationally become certain of) *must* form a partition. For example, in the above case they'll point out that there's something wrong with the above model. The model says that in $a[\overline{c}]$, you assign $\frac{1}{4}$ probability to $b[\overline{a}]$—that is, you think it's 25%-likely that b (you got the ticket) and that $[\overline{a}]$ (you were told a didn't get the ticket). This *would* be appropriate if you were unsure what you were told. But you're not! At $a[\overline{c}]$, you know that you were told c didn't get the ticket. (It's not as if you're unsure about what the conductor said—if you were, you wouldn't have ruled out c.) In other words, if the conductor tells you "c didn't get the ticket", you learn not only that c didn't get the ticket but also that *the conductor told you c didn't get the ticket*, i.e. $[\overline{c}]$. Thus, in this case, they'll say that conditioning on your *total* evidence will recover partitionality (and hence Reflection)—in particular, if you update from P to P' by conditioning on the facts about what you've been told, either $[\overline{c}]$ or $[\overline{a}]$:

$$P' = \left(\begin{array}{cccc} a[\overline{c}] & b[\overline{c}] & b[\overline{a}] & c[\overline{a}] \\ \hline {}^{2}/_{3} & {}^{1}/_{3} & 0 & 0 \\ {}^{2}/_{3} & {}^{1}/_{3} & 0 & 0 \\ 0 & 0 & {}^{1}/_{3} & {}^{2}/_{3} \\ 0 & 0 & {}^{1}/_{3} & {}^{2}/_{3} \end{array}\right)$$

More generally, to avoid clearly-irrational results like the one above, in practice Bayesian modelers restrict themselves to *partitional updates*. Note that for any a partition $\mathcal{Q} = \{q_1, ..., q_n\}$, every world falls into exactly one partition-cell; let $\mathcal{Q}_w$ be the partition-cell of w. Then:

> **Partitional Updates:** The update (P, P) is *partitional* iff there is some partition $\mathcal{Q}$ such that, for all worlds w, $P_w(\cdot) = \mathsf{P}(\cdot|\mathcal{Q}_w)$.

Partitionality isn't quite enough to recover Reflection. The trouble is that we can replicate the same sort of issue if we start with a *prior* that has a similar structure—namely, is not sure of its own values. Here's the sort of example that we'll talk much more about later (Williamson 2000). You're right near the border of feeling cold. Introspecting, suppose you know you either feel slightly cold (c), feel exactly neutral (n), or feel slightly warm

[12] This is a variant on the Monty Hall problem.

(w)—but you can't tell exactly which. Then it might be natural (we'll come back to this) to describe your current (prior) state with P below: if in fact you feel cold, you're 50-50 between feeling cold and feeling neutral; if in fact you feel neutral, you are uniform over the three possibilities, and if in fact you feel warm, you are 50-50 between feeling neutral or warm. You know that in a moment, your thermometer will beep (if you feel warm) or bloop (if you don't feel warm), updating your probabilities to P by conditioning on the partition $\mathcal{Q} = \{\{c, n\}, \{w\}\}$:

$$\mathsf{P} = \left(\begin{array}{ccc} c & n & w \\ \hline {}^1\!/_2 & {}^1\!/_2 & 0 \\ {}^1\!/_3 & {}^1\!/_3 & {}^1\!/_3 \\ 0 & {}^1\!/_2 & {}^1\!/_2 \end{array}\right) \qquad P = \left(\begin{array}{ccc} c & n & w \\ \hline {}^1\!/_2 & {}^1\!/_2 & 0 \\ {}^1\!/_2 & {}^1\!/_2 & 0 \\ 0 & 0 & 1 \end{array}\right)$$

In this update, Reflection is once again violated. For consider P_w—your prior if in fact you feel warm—and what it expects about your posterior. $\mathsf{P}_w(c) = 0$: you're certain that you don't feel cold. But you leave open both that you're at w—where you'll maintain your certainty ($P_w(c) = 0$)—or at n—where you'll end up $\frac{1}{2}$-confident of w ($P_n(c) = \frac{1}{2}$). Thus you're on the edge of the range of future credences you think might be rational: $\mathsf{P}_w(c) = 0$, but $\mathsf{P}_w(P(c) \geq 0) = 1$ and $\mathsf{P}_w(P(c) > 0) > 0$.[13] It follows that your prior doesn't equal a (P_w-)probability-weighted average of your posterior in c—in particular, that average is $\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} > 0 = \mathsf{P}_w(c)$. Reflection is violated.[14]

The standard response to this problem is to require that the prior be known—that is, like the prior in the train example, it doesn't vary amongst possibilities that the prior itself leaves open: if $\mathsf{P}_w(x) > 0$, then $\mathsf{P}_w = \mathsf{P}_x$. Equivalently, it must be clear what the prior is—the prior obeys an *introspection* axiom. As we'll see below, standard examples of introspection axioms are on knowledge or belief—letting Bp be the proposition that you believe that p, positive introspection on belief says that if you believe p, you believe that you do: $Bp \to BBp$. Negative introspection says that if you don't believe p, you believe that you don't: $\neg Bp \to B\neg Bp$. (See e.g. Hintikka 1962; Stalnaker 2006.) Combined, they say that you always have correct beliefs about whether or not you believe p.

Porting that over to probabilities, introspection says that you always have (correct) certainty about which prior you have: for all p, $[\mathsf{P}(p) = t] \to [\mathsf{P}(\mathsf{P}(p) = t) = 1]$. Equivalently:

> **Prior Clarity:** P is *clear* iff, at all worlds w, if $\mathsf{P}_w = \pi$, then $\mathsf{P}_w(\mathsf{P} = \pi) = 1$.
> You're certain of what your prior probability function is.

Usually Prior Clarity flies under the radar, and isn't explicitly stated as an assumption. Instead, it's smuggled in using the notation. When statisticians and social scientists specify a prior probability function, π, the intended interpretation is (almost always) as a rigid

[13] For discussion of such "no lose investigations" (your rational credence can't go down, and it might go up), see White 2006; Titelbaum 2010; Salow 2018; Fraser 2021; Dorst 2021.

[14] Relatedly, note that $\mathsf{P}_w(c|P(c) = \frac{1}{2}) = \mathsf{P}_w(c|\{c, n\}) = \frac{0}{0+1/2} = 0$, violating local Reflection.

designator—meaning it doesn't vary between possibilities as an object of uncertainty. But as we'll see, if we want to treat a quantity (including a probability) as an object of uncertainty, we need to make it something that can vary between worlds—i.e. is a *function* from worlds to quantities.

Combining Prior Clarity with Partitional Updates yields the set of *Clear Bayes* updates:

> **Clear Bayes:** $\langle \mathsf{P}, P\rangle$ is a *clear-Bayes* update iff P is clear and there is a partition $\mathcal{Q}$ such that, for all worlds w, $P_w(\cdot) = \mathsf{P}_w(\cdot|\mathcal{Q}_w)$.

Any Clear Bayes update satisfies all of our Reflection principles—and therefore our nice results like convergences to the truth, etc. Clear Bayes (when generalized to infinite cases) is the standard operating procedure in Bayesian modeling—it is the class of models used in most every application of Bayesianism in psychology, economics, and the behavior sciences.

And for good reason! The models are flexible, tractable, and extremely useful. But, as we'll see, they can't model genuine ambiguity. That's why we'll have to break them.

2.2 Expectations and Decisions

So far we've talked about how Bayesians hold and update beliefs. What about how they make decisions? In short, they make decisions using a certain feature of their beliefs—what they "expect" or "estimate" will be best (on average). So first, how does a Bayesian estimate a quantity?

In addition to assigning probability-values to propositions, a probability function also assigns *expectations* (estimates) to *variables.* (What statisticians call "random variables", since they can take on different values according to a given probability function.) A variable (or quantity) X is a function from worlds w to numbers $X(w) \in \mathbb{R}$. We can think of them as descriptions of numbers, for example $X =$ *the number of cookies in the cookie jar* or $Y =$ *the number of pairs of socks Kevin owns.* Since there are different numbers of cookies (pairs of socks) in different possibilities, X (and Y) take on different values at different possibilities. Generally, a given probability function π will be uncertain about the value of a random variable; but it can form an *expectation* of the variable—a probability-weighted average of the variables possible values. Precisely, the expectation $\mathbb{E}_\pi(X)$ of X relative to π is:

$$\textbf{Expectations: } \mathbb{E}_\pi(X) \;=\; \sum_{t\in\mathbb{R}} \pi(X = t)\cdot t$$

For example, if you know there are either 10, 20, or 30 cookies in the jar, and you're $\frac{1}{3}$-confident of each of these possibilities, then your expectation for the number of cookies is $\pi(X = 10)\cdot 10 + \pi(X = 20)\cdot 20 + \pi(X = 30)\cdot 30 = \frac{1}{3}(10) + \frac{1}{3}(20) + \frac{1}{3}(30) = 20.$

Notice that we've implicitly been using expectations when talking of "probability-weighted averages". For example, Reflection (average) can be equivalently stated as the constraint that your prior in p equals your priors expectation of your posterior in p: $\mathsf{P}(p) = \mathbb{E}_{\mathsf{P}}(P(p))$.

In general, although π will be uncertain about the value of X, if it knows what π is, it will know what the *expected value* of X relative to π, $\mathbb{E}_\pi(X)$, is—since that is determined just by π, not by the world. Here some notational niceties are again crucial (cf. Williamson 2008). '$\mathbb{E}_\pi$' is a rigid designator for a particular function from random variables to numbers—the one specified above. But often we want to refer to the expectations not of a particular probability function π, but of a descriptively specified probability function P—e.g. *Kevin's expectations tomorrow* or *Caspar's expectations today.* For that, we'll use $\mathbb{E}_P$—which, just as P compares to π—is a function from worlds w to expectation-functions $\mathbb{E}_{P_w}$, and therefore can be an object of uncertainty. Precisely: $\mathbb{E}_P(X)$ varies from world to world; for any world w, $\mathbb{E}_P(X)(w)$—which I'll write $\mathbb{E}_{P_w}(X)$, since it's variations in P which determine the variations in $\mathbb{E}_P(X)$—we have $\mathbb{E}_{P_w}(X) = \sum_{t\in\mathbb{R}} P_w(X = t)\cdot t$. We'll come back to this later.

Back to rigidly-designated expectations. Notice that a special type of random variable is an *indicator variable* X_p which takes the value 1 if p is true and 0 if p is false. By definition, $\mathbb{E}_\pi(X_p) = \pi(X_p = 1)\cdot 1 + \pi(X_p = 0)\cdot 0 = \pi(p)$—that is, the expectation of X_p equals the probability that p is true. So probabilities of propositions are special cases of expectations. That's why it's sometimes said that *probabilities are estimates of truth-values.*

As this brings out, π's expectation of X needn't be a value it thinks X can obtain—after all, you know the truth-value *won't* be 0.5 or 0.6. (It'll be 0 or 1.) Rather (by the law of large numbers), $\mathbb{E}_\pi(X)$ is the value that π thinks a bunch of independent copies of X would average out to. Thus if $\mathbb{E}_\pi(X_p) = \pi(p) = 0.5$, this is equivalent to saying that π is confident that amongst a big set of independent claims $q_1, ..., q_n$—each exactly as likely as p—roughly 50% will be true. In other words: to be be 0.5-confident in p is to think "claims like this tend to be true about half the time." For example, let p be the claim that this (fair) die will land 1–4 when I toss it. You should be $\frac{2}{3}$-confident of this, so $\mathbb{E}_\pi(X_p) = \frac{2}{3}$. Equivalently, if I toss it 30 times, you should be confident that it'll land 1–4 roughly $\frac{2}{3}\cdot 30 = 20$ times total.

Let's add to probabilism the claim that your estimates that can be represented by some probability's expectation-function: there is a π such that for any variables X and Y, your estimate of X is higher than your estimate of Y iff $\mathbb{E}_\pi(X) > \mathbb{E}_\pi(Y)$. This implies that your estimates form a total order.

Similarly for *conditional* estimates. Just as we can condition a probability function π on a proposition q to get a new probability function $\pi(\cdot|q)$, so too we can condition an expectation function $\mathbb{E}_\pi$ on a proposition q to get a new expectation function $\mathbb{E}_\pi(\cdot|q)$. The operation is simply to condition π on q, and then take expectations relative to this updated probability function $\pi(\cdot|q)$:

$$\textbf{Conditional Expectations: } \mathbb{E}_\pi(X|q) \;=\; \sum_{t\in R}\pi(X = t|q)\cdot t \;=\; \mathbb{E}_{\pi(\cdot|q)}(X)$$

Let's add to probabilism that your comparative conditional estimates can be represented by some probability function—and hence form a total order.

Expectations are (weighted) averages, so are mathematically easy to work with. Their nicest feature is that they obey an analogue of the Total Probability theorem, which gives convenient ways to calculate them. Take any question (partition) $\mathcal{Q}$; then the expectation of X is a probability-weighted average of it's *conditional* expectations given the true answer to $\mathcal{Q}$:

Total Expectation: $\mathbb{E}_\pi(X) \;=\; \sum_{q_i \in \mathcal{Q}} \pi(q_i) \cdot \mathbb{E}_\pi(X|q_i) \;=\; \pi(q_1)\mathbb{E}_\pi(X|q_1) + \cdots + \pi(q_n)\mathbb{E}_\pi(X|q_n)$

Notice that this weighted average is *itself* an expectation. Thus—analogously to $\pi^{\mathcal{Q}}$—let's write $\mathbb{E}_\pi^{\mathcal{Q}}(X)$ as a description of π's expectation of X conditional on the true answer to $\mathcal{Q}$, whatever it is. Then we can write Total Expectation equivalently as $\mathbb{E}_\pi(X) = \mathbb{E}_\pi(\mathbb{E}_\pi^{\mathcal{Q}}(X))$. This is sometimes called the "tower law" of expectations, and (misleadingly) stated as "the expectation of the expectation equals the expectation". Notice that, as with our different versions of Reflection above—Total Expectation is a *synchronic* constraint on a *rigidly specified* probability function π: it states a relationship between π's expectations and π's expectations of π's conditional expectations. As we'll see (importantly) below, once we consider probability functions P that can be uncertain of their own values, it is *not* a theorem that $\mathbb{E}_P(X) = \mathbb{E}_P(\mathbb{E}_P(X))$: *Caspar's expectation of Caspar's expectation of* X needn't equal *Caspar's expectation of* X. (When Caspar is uncertain what he believes, generally it won't.)

Back to rigidly-specified expectations. Total expectation makes expectations easy to work with, since we can always break them down into conditional expectations that we know the values of. The simplest version of this is if we let the question $\mathcal{Q}$ be the finest-grained partition of W into its singletons: $\mathcal{Q} = \{\{w_1\}, ..., \{w_n\}\}$. Then, since $\mathbb{E}_\pi(X|\{w\}) = X(w)$, Total Expectation implies:

Expectations (by-world): $\mathbb{E}_\pi(X) \;=\; \sum_{w \in W} \pi(w) \cdot X(w) \;=\; \pi(w_1)X(w_1) + \cdots + \pi(w_n)X(w_n)$

This means that in our vector and matrix notation, expectations are easy to calculate. Again order our worlds in some canonical order $w_1, ..., w_n$. Then—just like a probability function—a random variable X can also be thought of as a vector $X = (X(w_1), ..., X(w_n))$ in which the ith entry is the value X takes at the ith world.

Let's take our above train example, with worlds $(a[\overline{c}], b[\overline{c}], b[\overline{a}], c[\overline{a}])$ and a prior $\pi = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3})$. Let X be how much money you lose, and suppose that if you don't get the ticket you'll have to buy another one for \$100, but a is nice and c is selfish—if a gets the ticket she'll give you \$50, whereas if c does, he'll give you nothing. Then your change in assets in each world is $X = (-50, 0, 0, -100)$. Your *expectation* for your change in assets is a weighted average of these values, with the weight on each determined by its probability, i.e. the corresponding entry of π's probability-vector. Thus $\mathbb{E}_\pi(X) = \pi(w_1)X(w_1) + \pi(w_2)X(w_2) +$

$\cdots + \pi(w_n)X(w_n)$. That operation is a very well-known one in linear algebra: it's the *dot product* of our two vectors π and X, wherein we multiply corresponding entries and then sum them up. (It's a pain to calculate by hand, but computers do them instantly.) If we write our vectors more compactly as $\pi = (\pi_1, ..., \pi_n)$ and $X = (X_1, ..., X_n)$, the operation is $\pi \cdot X = \pi_1 X_1 + \cdots + \pi_n X_n$. Applied to our example, that means $\mathbb{E}_\pi(X) = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{3}) \cdot (-50, 0, 0, -100) = -50$.

So much for estimates. What about *decisions*? Simple: just pick the option that you estimate will be best. More precisely, suppose you are given a (let's say, finite) set of options $\mathcal{O} = \{O_1, ..., O_m\}$. Each option would lead to different outcomes in different worlds. It's standard to assume that rational preferences are fine-grained enough to induce both a total order on and cardinal comparisons between (how much better is x than y?) these outcomes. Thus your preferences amongst outcomes can be represented with a utility function that assigns outcomes to numbers, and your preferences amongst risky options can be represented by your comparisons of *expected* utility.

Picking one of those utility functions, what this means is that we can simply model a decision problem as a set of random variables: O_i is the variable *the amount of utility you'd get from option O_i*, and $O_i(w)$ is that amount at w.[15]

For example, suppose in our train case above I offer you insurance to guard against the cost of a new ticket. You value money linearly, so if you don't buy the insurance, your utility will be X—the variable from above, with $X = (-50, 0, 0, -100)$ across worlds $(a[\overline{c}], b[\overline{c}], b[\overline{a}], c[\overline{a}])$. If you *do* buy the insurance, it'll cost \$40, but will cover the cost of the ticket if you don't get it. Thus Buying the insurance yields $B = (-40, -40, -40, -40)$.

What should you do? According to your prior, the expected value of B is $\pi \cdot B = -40$, while the expected value of X is $\pi \cdot X = -50$, so it maximizes expected value to buy the insurance.

Now suppose I give you another option: rather than deciding now, you can first hear what the conductor tells you (learn whether $[\overline{c}]$ or $[\overline{a}]$), and *then* decide. What's the expected value getting the information and then deciding? That is equivalent to following a strategy S in which you first update your beliefs on what the conductor says, and then do whatever maximizes expected value given those posterior beliefs. What will that amount to? If the conductor tells you $\neg c$ (you're in a world $w \in [\overline{c}]$), you'll condition on that and move to $P_w = (\frac{2}{3}, \frac{1}{3}, 0, 0)$, in which case the expected value of not getting the insurance (X) is $P_w \cdot X = \frac{2}{3}(-50) + \frac{1}{3}(0) \approx -33.3$, which is better than the (still) -40 value of getting the insurance—so if you learn $[\overline{c}]$, you'll take X. (You won't buy the insurance, because even if you don't get the ticket, a will help you pay for a new one.) Meanwhile, if the conductor tells you $\neg a$ (you're in a world $x \in [\overline{a}]$), you'll move to $P_x = (0, 0, \frac{1}{3}, \frac{2}{3})$, in which case the

[15] Note that we're implicitly doing "causal" decision (or, really, counterfactual) decision theory here (Lewis 1981; Hedden 2023). There are facts about which option you take at which world. Perhaps at w you take O_1. Still, $O_2(w)$ is well-defined—it's the value you *would've* gotten at w if you *had* taken O_2. We could give a semantics for this using counterfactuals (Gibbard and Harper 1978) or imaging (Lewis 1976). But for our purposes, we needn't.

expected value of not getting insurance is $P_x \cdot X = \frac{1}{3}(0) + \frac{2}{3}(-100) \approx -66.7$. That's worse than getting the insurance, so you'll get the insurance.

Thus learning-and-then-deciding is equivalent the strategy S of taking X if you're told c didn't get the ticket, and taking B if you're told a didn't get the ticket. In other worlds, we can think of S as a new random variable which outputs $S_w = \begin{cases} X(w) & \text{if } w \in [\overline{c}] \\ B(w) & \text{if } w \in [\overline{a}] \end{cases}$. In other words, $S = (-50, 0, -40, -40)$. What's the (prior) expected value of *that*? It's $\mathsf{P} \cdot S = \frac{1}{3}(-50) + \frac{1}{6}(0) + \frac{1}{6}(-40) + \frac{1}{3}(-40) \approx -36.7$. Notice that this is better (by about 3.3 utility) than the option of *not* learning-and-then-deciding, i.e. simply choosing the option (B or X) that looks best by your prior's lights, i.e. $B = -40$. As a result, you should be willing to pay a bit over $3 to get the chance to hear what the conductor has to say before making your decision.[16]

This isn't an accident. The fact that—when learning is free—the expected value of learning-and-then-deciding is always greater than the expected value of deciding-without-learning is called the **value of evidence** (or "value of information") theorem (Ramsey 1990; Blackwell 1953; Good 1967), mentioned above. In this standard setting, it is closely linked to the Reflection principle and the fact that—if you're a precise Bayesian updating by conditioning on the true answer to a question—you defer to your future self. As we'll see, the differing ways of adding ambiguity threaten to break the value-of-evidence theorem—to make it the case that sometimes we should pay to *avoid* rationally updating on free evidence. This result is (arguably) inevitable if we model ambiguity with imprecision, but avoidable if instead we model it with higher-order uncertainty.

3 What Clarity Can't Do

Clear Bayes can't capture our five features of ambiguous judgments: irresolute assessments, insensitivity, fuzzy boundaries, ambiguity aversion, and biased processing.

Why not? A quick theorem says that any Clear-Bayes model (and, therefore, any combination of such models into a long string of updates $(P^0, P^1, P^2, ...,))$ will be *clear* in the sense that you will be *introspective*: certain of exactly what your probability function is:

> **Bayesian Clarity:** If (P, P) is a clear-Bayes update, then for all worlds w:
> $\mathsf{P}_w(\mathsf{P} = \mathsf{P}_w) = 1$ and $P_w(P = P_w) = 1$.[17]

[16] Why? Expectations are "linear", meaning the expectation of the sum equals the sum of the expectation: $\mathbb{E}_\pi(X + Y) = \mathbb{E}_\pi(X) + \mathbb{E}_\pi(Y)$. The value of paying $3 to learn is equivalent to the option $S - 3$, i.e. $(-53, -3, -43, -43)$, which has expectation equal to $\mathbb{E}_\mathsf{P}(S - 3) = \mathbb{E}_\mathsf{P}(S) - \mathbb{E}_\mathsf{P}(3) \approx -36.7 - 3 = -39.7 > -40 = \mathbb{E}_\mathsf{P}(B)$.

[17] This is immediate for the prior by the definition of Prior Clarity. For the posterior: the fact that P is P updated on the true cell of a partition means that if $P_w(x) > 0$, then w and x must be in the same partition-cell: $\mathcal{Q}_w = \mathcal{Q}_x$. Since they both started out with the same prior, and they both conditioned on the same proposition ($\mathcal{Q}_w$), that means they are the same. In other words, if at w your posterior leaves open another possibility x ($P_w(x) > 0$), then that possibility is one where you have the exact same posterior $P_x = P_w$.

This might be confusing, since many (Clear-)Bayesian models involve what's sometimes called *hierarchical* uncertainty. For example, suppose you—a clear Bayesian—are uncertain of the bias of a coin. You know it tends to land heads either 10% or 20% or... or 90% of the time, but you don't know which. Let's call that the *objective chance* of heads, and pick it out with the random variable *ch(h)*. Suppose you're uniform over these 9 possibilities, so $\mathsf{P}(\textit{ch(h)} = 0.1) = \frac{1}{9}$, $\mathsf{P}(\textit{ch(h)} = 0.2) = \frac{1}{9}$, etc. So you're uncertain about a probability function—namely, the objective chances.

Still, this is not *higher-order uncertainty* in the relevant sense: you are not unsure of your own probability function. Since you're a clear Bayesian, you know that this is what your prior is: $\mathsf{P}(\mathsf{P}$ *is uniform over ch(h)* = *0.1,..,0.9*$) = 1$. Or, more cumbersomely, $\mathsf{P}\Big([\mathsf{P}(\textit{ch(h)} = \textit{0.1}) = 0.1]\&...\&[\mathsf{P}(\textit{ch(h)} = \textit{0.9}) = 0.1]\Big) = 1$. Assuming you obey the Principal Principle (Lewis 1980) (an analog of Reflection) your prior probability for the coin landing heads equals your expectation of the chance of heads, which is 0.5: $\mathsf{P}(h) = \mathbb{E}_{\mathsf{P}}(ch(h)) = 0.5$. Thus you know that your prior assigns 0.5 to h: $\mathsf{P}(\mathsf{P}(h) = 0.5) = 1$. Thus if you go on to investigate the coin (say, by flipping it), we can model you with a Clear-Bayes model—and hence you'll satisfy Reflection and obey all the good theorems.

Likewise for expectations. Baylee the Clear-Bayesian is wondering how many socks I have. I'm wondering too. She know that I have more information about that quantity—call it X—than she does; but since I haven't counted, I'm unsure as well. Letting R be my (descriptively-specified) credence function, and P be hers, let's say she defers to my expectation: $\mathbb{E}_P(X|\mathbb{E}_R(X) = t) = t$. She's unsure whether my estimate is 15 or 20 or 25: $P(\mathbb{E}_R(X) = 15) = P(\mathbb{E}_R(X) = 20) = P(\mathbb{E}_R(X) = 25) = \frac{1}{3}$, so she's uncertain about an expectation. But she's not uncertain about *her* expectation: since she know what P is, she know (P assigns probability 1 to the claim) that $\mathbb{E}_P(X) = 20$.[18]

More generally, it follows from Clear-Bayes that probabilities and estimates are introspective:

Introspection:
If $P(q) = t$, then $P(P(q) = t) = 1$.
If $\mathbb{E}_P(X) = t$, then $P(\mathbb{E}_P(X) = t) = 1$.

Similarly, a Clear Bayesian might be unsure what a *more-rational* version of themselves would believe. Let P be your actual opinions and $\widehat{P}$ be your idealized-self's opinions. Then we can let P be uncertain about $\widehat{P}$, and require that P defers to $\widehat{P}$: $P(q|\widehat{P}(q) = t) = t$. Thus there's a sense in which you can be unsure what you ought to believe—you're non-ideal self can be unsure what your ideal-self thinks. But—supposing both P and $\widehat{P}$ are Clear-Bayes—there is no uniform reading of 'ought' such that "you ought to be unsure what you ought to believe" comes out true. After all, P is certain of what P is and $\widehat{P}$ is certain of what $\widehat{P}$ is. Rather, what we have is a mixed reading: you ought (given your limitations) to match P, and so be uncertain what you ought (ideally, i.e. matching $\widehat{P}$) to believe. Thus on a uniform

[18] Since $\mathbb{E}_P(X) = P(\mathbb{E}_R(X) = 15)\cdot 15 + P(\mathbb{E}_R(X) = 20)\cdot 20 + P(\mathbb{E}_R(X) = 25)\cdot 25 = \frac{1}{3}(15) + \frac{1}{3}(20) + \frac{1}{3}(25)$.

reading, Clear-Bayes implies that if I'm believing as I ought, I can have no uncertainty about whether I'm believing as I ought.

Since Clear-Bayes assumes you have precise beliefs and values *and you know what your beliefs are*, it can't do justice to our five features of ambiguous judgments. Here's why.

First, **irresoluteness**. This was the fact that there was some randomness in your ambiguous judgments. We can of course imagine a Clear-Bayesian whose estimate about my socks is fluctuating rapidly, since they're constantly conditioning on small bits of evidence that are coming to mind. But *fixing* a given time, there should be no randomness or arbitrariness in their judgment. At a given time t, they have an estimate for my number of socks at $\mathbb{E}_P(X) = 21.43$; since they're a Clear Bayesian, at t they are certain that they have exactly that estimate. There should be no stochasticity—if you asked a thousand doxastic duplicates of them (at t), they should all say '21.43'.

Second, **insensitivity.** This was the fact that your ambiguous comparative judgments seem insensitive to small sweetenings. Let p be the proposition that this penny will land heads 10 times in a row when flipped. Let q be the claim that I have at least 20 socks and r be the claim that the die will land 1–4 on at least 20 (of the 30) tosses. Then, intuitively, it's (1) unclear whether you're more confident of q than r, (2) unclear whether you're more confident of $q \vee p$ than r, but (3) clear that you're more confident of $q \vee p$ than of q. Yet if you're a Clear Bayesian, then not only does 'more confident than' form a total order, but you're *certain* of what that order is. So for each of these comparisons, you should be able to say (in fact, be willing to bet your life on!) which is true.

Third, **fuzzy boundaries:** when your judgments are ambiguous, it's usually unclear what the edges of the ambiguity are. We ran a fair lottery with 1000 people in it; let a_i be the claim that the ith person won, so $P(a_i) = \frac{1}{1000}$. Clearly you're more confident of $a_1 \vee a_2 \vee ... \vee a_{1000}$ (think of this as A++++...) than you are that I have at least 20 pairs of socks (q). Clearly you're *less* confident of a_1 than you are that I have at least 20 pairs of socks. But what's the first n such that you're more confident of $a_1 \vee a_2 \vee ...a_n$ than of q? If you're a Clear Bayesian, there is such a number, and you know exactly what it is. In particular, you'll have some credence in q—say, $P(q) = 0.4702$—and you'll know as much: $P(P(q) = 0.4702) = 1$. You'll also have a precise (and known) credence in the a_i: $P(a_1 \vee ... \vee a_{470}) = 0.470$, while $P(a_1 \vee ... \vee a_{471}) = 0.471$. Since you'll be certain of all these facts, you can put them together to be certain that $n = 471$.

Fourth, **ambiguity aversion:** people tend to prefer to make bets on unambiguous judgments than on ambiguous ones. For example, they prefer Bet1 to both Bet2 *and* Bet3::

Bet1: \$100 if *heads*, \$0 if *tails*.

Bet2: \$100 if q (I have at least 20 pairs of socks), \$0 if $\neg q$.

Bet3: \$100 if $\neg q$, \$0 if q.

There's no way for a Clear-Bayesian to make sense of this. If $P(q) = 0.5$, then they should be indifferent and know as much. If $P(q) > 0.5$, they should prefer Bet2 and know as much.

If $P(q) < 0.5$, they should prefer Bet3 and know as much.

Fifth, **biases and miscalibration.** I'm going to have you give probabilities for a bunch of questions for which you have ambiguous judgments—*Kevin has at least 20 pairs of socks*, *Roger will be in his office tomorrow*, etc. Then I'm going gather up all the ones that you answer "0.5" on—say, I'll do this till I have 100 of those, $q_1, ..., q_{100}$. Right now, before we do this, how many of them do you think will be true? Presumably, you have little idea—*maybe* around half of them, but who can say. Letting X_{q_i} be the indicator variable for q_i, the number that are true is the variable $\sum_i X_{q_i}$, i.e. $X_{q_1} + X_{q_2} + \cdots + X_{q_{100}}$. I'm guessing you have some hesitancy in saying that your estimate for $\sum_i X_{q_i}$ is exactly 50; and I'm guessing that you're not terribly confident that $\sum_i X_{q_i}$ will be (say) between 40 and 60—it could well be much lower or much higher. After all, you know that when you're forming estimates with so little to go on (or so much to sort through) you're not great at getting things right.

Clear Bayesians can't say this. Since they obey Reflection toward their future self, they defer to these future 50% judgments: for all q_i, $\mathsf{P}(q_i|P(q_1) = \cdots = P(q_{100}) = 0.5) = 0.5$. Abbreviating that big condition to C, this means $\mathbb{E}_\mathsf{P}(X_{q_i}|C) = 0.5$, and hence (since expectations are linear) that $\mathbb{E}_\mathsf{P}(\sum_i X_{q_i}|C) = \sum_i \mathbb{E}_\mathsf{P}(X_{q_i}|C) = \sum_i^{100} 0.5 = 50$. Moreover, so long as the Bayesian treats the judgments independently (being right about my socks doesn't shift their probability for whether Roger will be in tomorrow), then they'll be quite confident that X will be near 50: $P(40 < \sum_i X_{q_i} < 60) \approx 0.94$.[19] That is, they expect to be *well-calibrated.* Yet intuitively, in a case like this, you might expect no such thing!

So Clear Bayesians can't capture our five features.

But wait! I've focused on the *epistemic* versions of these features—insensitivity in your comparative-confidence judgments, etc. And I've been looking at Clear Bayesians, who know exactly what their *beliefs* are. But nothing we've said seems committed to saying Clear Bayesians know what their *preferences* (or utilities) are. So maybe we could still capture the *practical* versions of (some of) our features, so long as we allow uncertainty about utilities. Right?

I don't think so. An analogous point to the response to "hierarchical" uncertainty applies again. Suppose our Clear Bayesian is unsure whether they have utility function U_1 or U_2. To represent such uncertainty, as always, we must introduce a variable that takes on different values at different worlds. Let U be that variable—"their utility function, whatever it is" (descriptively specified). Suppose they're 50-50 between U_1 and U_2, so $P(U = U_1) = P(U = U_2) = 0.5$. Can we capture insensitivity with these tools? I don't think so. In short: their *expected* utility is still well-defined, precise, and known. Let a be some action that can result in different outcomes which might in turn have different utilities: $U_1(a) \neq U_2(a)$ in any

[19] Relative to $\mathsf{P}(\cdot|C)$, $\sum_i X_{q_i}$ follows a binomial distribution with probability 0.5 and 100 trials.

world. Still, by total expectation,

$$\begin{aligned}\mathbb{E}_P(U(a)) &= P(U=U_1)\cdot\mathbb{E}_P(U(a)|U=U_1)+P(U=U_2)\cdot\mathbb{E}_P(U(a)|U=U_2)\\ &= P(U=U_1)\cdot\mathbb{E}_(U_1(a)|U=U_1)+P(U=U_2)\cdot\mathbb{E}_P(U_2(a)|U=U_2)\end{aligned}$$

Which will be a precise (and, by introspection on P, known!) value.

Less formally, the point is simply that if you're unsure what your utility function is, that is a way of being unsure what the world is like—unsure how much value a given option will yield for you. So long as you have precise beliefs about how likely it is to be one versus the other—and know what those beliefs are—you know exactly what your *expected* utility is, which is what you need to make your decision. You might, of course, be unsure what a more-ideal version of yourself who *knew* what your utilities are would do (as you will be in this case)—but what that more-ideal version of yourself would do is not what *you* (given your uncertainty) should do. (Compare: you might be unsure whether someone who knew the bias of this coin would bet on heads or tails; but if you're uniformly unsure whether it's $0.1-0.9$, then you know what *you* should do—namely, be indifferent between betting on heads or tails.)[20]

4 Going Imprecise?

The standard solution is to introduce imprecise probabilities and utilities. Instead of a single probability and utility function, you are best represented with a *set* of probability functions $\mathbb{P}$ and a *set* of utility functions $\mathbb{U}$. Think of them as a set of precifisied versions of your opinions and preferences who disagree on how exactly to precisifiy them—your "credal committee" and "valuation committee", respectively. You're more confident of p than q iff every $\pi\in\mathbb{P}$ is more confident of p than q; if some are more confident and some are not, then there's no fact about whether you're more confident of p than q. Similarly with utilities for your preferences between outcomes. We should now think of decision problems $\mathcal{O}=\{o^1,...,o^n\}$ as giving a set of functions from worlds to *outcomes* (rather than numbers), and then each utility function $U\in\mathbb{U}$ transforms that decision problem into a set of random variables $U(o^1),...,U(o^n)$, i.e. functions from worlds to numbers. You prefer option o to option o' iff for every probability-utility pair (π,U) in your representors, the π-expected U-utility of o is higher than that of o': $\mathbb{E}_\pi(U(o))>\mathbb{E}_\pi(U(o'))$.

Notice that since so far we've only replaced rigidly-specified probability- and utility-functions with *sets* of them, the imprecise model doesn't represent uncertainty about your credal committee. Since $\mathbb{P}$ doesn't vary from world to world, the proposition that it takes a certain value is either true everywhere or nowhere: $\{w\in W:\mathbb{P}=S\}$ either equals W (if $\mathbb{P}=S$) or $\emptyset$ (if $\mathbb{P}\neq S$). That means that either our agent has no opinions at all about

[20] [Subtle things here about representation theorems and scaling of utility functions...]

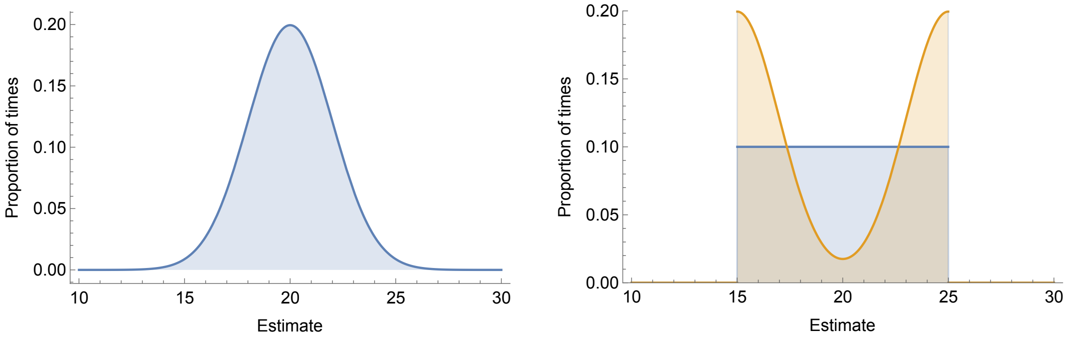
what their credal committee is (questions about it aren't in the domain of their opinions), or they're certain about what it is. Since you clearly *do* have opinions about what their credal committee is—you know that you're not more confident in *Kevin has at least 20 pairs of socks* than in *Caspar has at least 2 pairs of socks*, for instance—that means the model represents your as certain of what your credal committee is.

Can imprecision account for our features of ambiguous judgments? Sort of.

It's designed to account for **insensitivity**. Suppose your credal committee assigns estimates for my number of socks (X) ranging between 15 and 25, but all of them estimate the number of times (Y) the die lands 1–4 (of 30 rolls) at 20. Then there's no fact about whether you estimate X as higher than Y nor Y as higher than X, since some $\pi \in \mathbb{P}$ have $\mathbb{E}_\pi(X) < 20 = \mathbb{E}_\pi(Y)$, and others π' have $\mathbb{E}_{\pi'}(X) > 20 = \mathbb{E}_{\pi'}(Y)$. Suppose now Caspar gives me a pair of socks, so you know my updated number of pairs of socks is $X^+ = X + 1$. Then you determinately estimate my updated number of socks as higher than my previous number of socks, since for all $\pi \in \mathbb{P}$, $\mathbb{E}_\pi(X+1) = \mathbb{E}_\pi(X) + 1 > \mathbb{E}_\pi(X)$. But again, there's no fact about whether you estimate the updated number of socks as higher or lower than the number of 1–4-rolls.

It's also not hard to squint and see how imprecision can account for **irresoluteness**. Suppose you're credal committee's estimates for my number of socks range from 15 to 25. Then if I force you to name a specific number, you of course will—you can "identify" with one of your credal-committee's estimates, at least temporarily. But there need be no deterministic rule for how you do this: maybe sometimes when I ask you this, you'll say '20'; other times, you'll say '22', other times you'll say '18', etc.

Still, it's not clear that the imprecise model explains *all* of irresoluteness. As mentioned, it seems like in cases like this the numbers you name will be stochastic, and if we were to repeatedly probe you for your estimate (erasing your memory between each probe, to return you to your previous doxastic state), there'd be some fact about the distribution of the numbers you name. In a case like this, it seems plausible that those distributions would form a sort of bell curve centered around 20 and dropping off 15 and 25 (left):



Of course, this is *consistent* with the imprecise model. But nothing built in to the model explains why it would take that shape rather than (say) a uniform distribution between 15 and 25 or even a bimodal one (right).

What about **fuzzy boundaries**? Here the imprecise model struggles with a problem analogous to *higher-order vagueness* for supervaluationism. Since your credal committee is a set, it has precise boundaries. Remember the fair lottery between 1000 people, with a_i the claim that the ith person won. Let q be the claim that I have at least 20 pairs of socks. Suppose the range of your credal committee in q is $\mathbb{P}(q) = [0.353, 0.582]$. As we've seen above, since you know what your credal committee is, that means you know this fact: $\mathbb{P}(\mathbb{P}(q) = [0.353, 0.582]) = \{1\}$. And since you know this fact, you should know where your comparisons give out. If asked to name the smallest n such that you're more confident of $a_1 \vee ... \vee ...a_n$ than q, you should confidently say that $n = 583$.

What about **ambiguity aversion**? That is, people who are about 50%-confident of q tend to prefer Bet1 to Bet2 and to Bet3:

Bet1: \$100 if *heads*, \$0 if *tails*.
Bet2: \$100 if q (I have at least 20 pairs of socks), \$0 if $\neg q$.
Bet3: \$100 if $\neg q$, \$0 if q.

The imprecise model can explain this if it uses the right decision rule (Ellsberg 1961). There are variety of examples, but a simple one that will do (assuming you value money linearly, for simplicity) is CONSERVATIVE: prefer a bet to a sure \$x iff every member of your representor assigns it higher expected value than \$x. Since every member of your representor assigns $\mathbb{E}_\pi(Bet1) = 50$, Bet1 is valued like a sure \$50. Meanwhile, since $\mathbb{P}(q)$ spans numbers less than 50% and greater than 50%, some of your committee members value Bet2 less than \$50, and some value Bet3 less than \$50, and so you can sensibly prefer Bet1.

This result comes with a cost, though: the imprecise model can't vindicate plausible versions of the value of evidence. In particular, suppose $\mathbb{P}(q) = [0.3, 0.7]$, while you're exactly 50% that this fair coin will land heads: $\mathbb{P}(h) = \{0.5\}$. Suppose you know you're about to do a partitional update on the partition $\mathcal{Q} = \{h \leftrightarrow q, \neg h \leftrightarrow q\}$, i.e. to either learn (1) *either both h and q, or neither*, or (2) *either h and not q, or not-h and q*. Then (assuming your credal committee updates by conditioning), you know beforehand that your posterior credal committee in h will "dilate" to $\mathbb{P}'(h) = [0.3, 0.7]$ (White 2009).

This is puzzling for at least two reasons. First, it seems that you don't defer to your future self—and in particular, it's not fully explicable *why* you don't. Initially, you know that your more-informed, rational future self will have no opinion about whether the coin is more or less than 50%-likely to have landed heads. Yet right now you have exactly that opinion. Presumably the point of being (epistemically) rational and gathering information is that doing so is a good way to figure things out—to better map your beliefs onto the way the world is. So if this really is the predictable rational response, why not defer to it? (Formally, $\mathbb{P}(\mathbb{P}'(h) = [0.3, 0.7]) = 1$ and yet $\mathbb{P}(h) = 0.5$, so we have a violation of the Reflection-like principle which requires that $\mathbb{P}(h|\mathbb{P}'(h) = [0.3, 0.7]) = [0.3, 0.7]$.)

Second, and relatedly, this way of updating seems to be one in which you can be made to pay not to do the update. Suppose your decision-problem is whether to take a bet B which pays \$2 if *heads* and costs $-\$1$ if *tails*. Right now you want to take this bet: $\mathbb{E}_{\mathbb{P}}(B) = 0.5(2) - 0.5(1) = 0.5$, so you'd be willing to pay up to 50 cents to take the bet. But if you follow the above ambiguity-averse decision rule, you know that once you do the partitional update, you *won't* take the bet. Therefore right now you're willing to pay up to 50 cents to guarantee that you *do* take the bet—e.g. by covering your eyes so that you don't learn whether $h \leftrightarrow q$.

I believe this is fully general.[21] Whenever an update will predictably "dilate" your opinion about h, we will be able to find a decision-problem such that it's worth your while to pay not to do the update before making that decision.[22]

What about **biases and miscalibration**—can imprecision explain them? It's not so clear. It can definitely explain some of the intuitions. For example, suppose right now you have ambiguous judgments about a bunch of q_i; for each, $\mathbb{P}(q_i) = [0.1, 0.9]$, and you treat them all independently. Suppose what's going to happen is you're going to be asked to give point-estimates for each of them, and then we're going to gather all the ones you had a point-estimate of 0.5 in. Say we get 100 of them. What proportion do you think will be true? The imprecise model can allow for lots of uncertainty, as we want. In particular, since even if you point-estimate 0.5, you still actually have the attitude $[0.1, 0.9]$, then you have some members of your representor who expect the number that are true to be as few as 10 (since they're 0.1 in each), and others expect the number that are true to be as high at 90 (since they're 0.9 in each), so you can avoid determinately having high confidence that roughly 50 will be true.

Still, the possibility of dilation poses a variant challenge. Suppose right now you're *precise* on each of the q_i, and have exactly 50%-credence in them. You're about to get evidence that makes your opinions ambiguous in them, so you'll dilate (perhaps to different degrees) in each of them. Again, we're going to force you to make point-estimates and gather up all the ones that you estimate at 0.5 probability. Suppose we get 100 of those. What right now is your estimate for the number that will be true? Since you're still precise (predictable dilation doesn't affect your current beliefs), just like the Clear Bayesian you must estimate 50, and you must be confident that roughly 50 will be true—even though you know you're going to get highly ambiguous evidence about each of them. That, intuitively, seems wrong.

This is related to a more general challenge for the imprecise model: it's hard for it to explain how you *gain* ambiguity in your judgments. You can of course do it be linking ambiguous judgments to unambiguous ones—that's what happened above with the coin case. But in this case, the evidence you're getting is itself unambiguous: it's either $h \leftrightarrow q$, or

[21] **Q:** Is this right?

[22] It's sometimes pointed out that, on some decision-rules, it's *compatible* with them that you'll act exactly as your prior self wants your future self to (Bradley and Steele 2016). But it's also compatible with them that you don't, so as long as you can have uncertainty about what you'll do, it seems we can generate the problem. I think.

$\neg h \leftrightarrow q$, and you always know exactly what evidence you received. Yet many of our intuitive cases of ambiguity involve getting evidence that is *itself* ambiguous: the ambiguity in your judgment stems from the vagueness of the evidence you received. For example, suppose you start perfectly precise in all your opinions, including how tall the tree outside might be, and how likely it is to be various heights conditional on looking various ways. Now you glance at the tree. What exactly was the evidence you received? (How exactly did it look?) Intuitively, you can be unsure—and as a result of being unsure, you can have ambiguity in your posterior estimate of the height of the tree (Williamson 2000). But you can't explain this with a precise prior updating on a partition—if you update on a partition, you always know exactly what you updated on.

More generally still, the imprecise model is going to need higher-order uncertainty. For if I ask you what your representor is, the model implies that you should be able to tell me exactly: you have certainty about it, since it doesn't vary from world to world. But that's wrong: when you have ambiguous judgments, it's hard to say exactly how far that ambiguity extends.

Upshot: even if we like the imprecise model, we're going to need to add a way to think about higher-order uncertainty about your own (precise or imprecise) judgments. So let's see how we could do that—and, once we do, how much need there is to layer on imprecision over and above the higher-order uncertainty that we'll already need to add.

5 Higher-Order Uncertainty

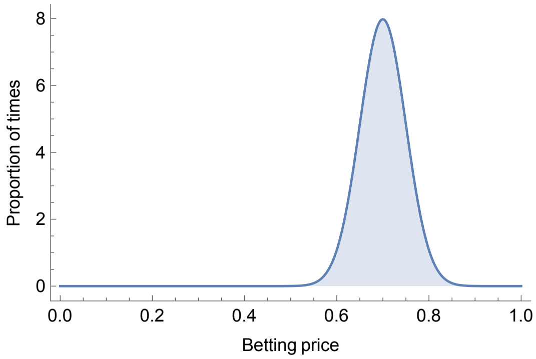
5.1 Why go higher-order?

Why introduce higher-order uncertainty? In some sense, the answer is obvious: we have it! We have uncertainty about what our own opinions are, so we need some way to represent that. But you might still like a theoretical explanation for *why* we have this uncertainty. After all, suppose your credences are simply your dispositions to bet—your fair prices on unit bets ($1 if true, $0 if false) on propositions. Then why would you have any uncertainty about your credences? Just see what it is you write down when forced to write down your fair prices—that's your credence. Right?

Intuitively, no. You might mess up. Almost certainty you'll give a round number, even if in reality your attitude isn't exactly that round number. And most likely, if I were to prompt you the exact same way many times when you're in that exact same doxastic state (say, by wiping your memory in between each; or by making a bunch of doxastic duplicates of you), you'd write slightly different numbers each time.

That is, we have *noise* or *stochasticity* in our cognitive systems. There is no deterministic function between what your attitude toward q is and exactly what number you'll write down in a given context. Psychologists have known this forever (Thurstone 1927), which is why most contemporary ways of probing people's degrees of confidence allows for the fact that there will be noise and "error" in the realization of that degree of confidence (the numbers they write down, or the actions they take) in any given probe (Erev et al. 1994; Juslin et al. 2009; Bonawitz et al. 2014; Icard 2016; Sanborn and Chater 2016).

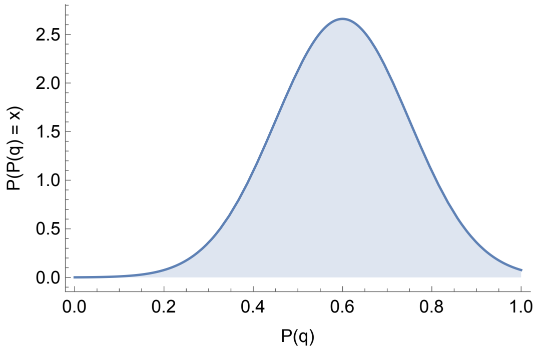
Here's a simple picture: your credence in q measures your *disposition* to bet on q, but that disposition can be masked with noise. Suppose if we were to probe your credence (fair betting price) in q a bunch of independent times, what we'd find is that the distribution of your answers is normally distributed with mean 0.7 and some variance: most of the time you write down a number between 0.6 and 0.8, but occasionally it's even further away:



Then I think there's good sense to be made to the claim that you really do have credence 0.7 in q, but that noise can mask this value. In other words: your credences are the *average* of how much you'd be willing to rely on various propositions.[23]

Suppose this is our picture of credences. Then higher-order uncertainty is going to fall out. After all, your credence is a hard-to-observe disposition which you can only probe by eliciting it (seeing how much you're willing to write down as your fair price, say). But since such elicitations are noisy, although eliciting it provides *some* evidence for what your credence is, it's not definitive.

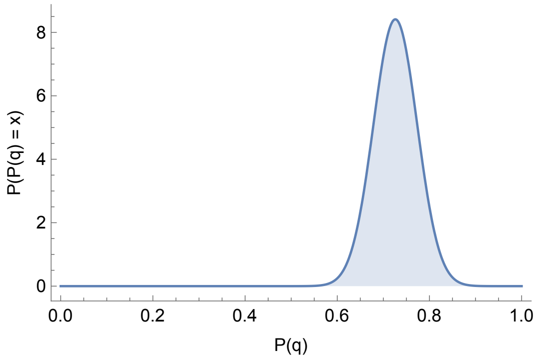
Suppose, for instance, you started quite unsure what your credence in q was: maybe it's as high as 0.9 or as low as 0.3:



[23] Something like: your credence is what the objective chances would expect you to write down if probed; $P(q) = t$ iff *probed* $\Box\!\!\rightarrow \mathbb{E}_{ch}(answer) = t$.

(Note that the previous graph showed the objective chances or likelihoods of you writing down various numbers, while this graph shows your subjective distribution over what the mean of those various objective-chance distributions are.) These various hypothesis are more and less likely to generate different numbers when you elicit your credence. $P(q) = 0.7$ makes it very likely that you'll write a number between 0.6 an 0.8, while $P(q) = 0.4$ makes this quite unlikely.

Then you go ahead and elicit your credence and find that you wrote down 0.74, This provides evidence in favor of higher credences, but that evidence is not definitive. If you know the distributions are all Gaussian, and you update by conditioning on what you wrote down, your updated distribution over what your credence actually is will be higher, but still have a fair bit of uncertainty in it:



The point? We should expect creatures with noise in their cognitive systems to have uncertainty about what their own dispositions are, and for that uncertainty to not be easily resolvable—to remain upon seeing various elicitations of those dispositions, for example. Insofar as we understand credences as dispositions to act, that means we should expect reasonable people to have uncertainty about their own credences.

5.2 How go higher-order?

How can we model people who have higher-order uncertainty? It's harder than you'd think. The core reason is that once we do, the standard components of Bayesianism start to break down. In particular, the Reflection principle rules out the possibility of higher-order uncertainty. Equivalently, any model of higher-order uncertainty must violate Reflection (Elga 2013).

Why? It's easiest to see if we look at the global version, which says that conditional on exactly what distribution (say) your future-self has, you should adopt that distribution: $\mathsf{P}(\cdot|P = \pi) = \pi(\cdot)$. Suppose that future-self has higher-order uncertainty: $P = \pi$, but $P(P = \pi) < 1$, i.e. $\pi(P = \pi) < 1$. This is inconsistent with Reflection. For what is your *prior* credence that $P = \pi$, conditional on $P = \pi$? 1, of course, by definition: $\mathsf{P}(P = \pi|P = \pi) = 1$.

But by hypothesis, π does *not* assign 1 to this proposition: $\pi(P = \pi) < 1$. Thus you don't match P upon learning that it's π: $\mathsf{P}(P = \pi | P = \pi) = 1 > \pi(P = \pi)$. Reflection fails.

In a way, this is reminiscent of the Reflection failures of imprecise credences. But in this case, there's a clear explanation for *why* Reflection fails. Suppose you learn that $P = \pi$. If P has higher-order uncertainty (so $P(P = \pi) < 1$), by hypothesis *you've learned something that P didn't know*. And if you know something that P doesn't, you shouldn't defer to P's opinions about that thing. Yet Reflection says to defer to *all* of P's opinions. That's why Reflection must fail.[24]

Because of that, lots of standard ways for reasoning about probability will break down—and it'll be very easy to find our way into paradox. We need to proceed with care, so we need to introduce a class of models for reasoning carefully about these issues.

5.2.1 Warm-up: epistemic logic

When a Bayesian is unsure what their probabilities are, how can we represent their belief-state?

Start simpler. Consider an agent who just has binary beliefs: for every proposition q, they either believe it (Bq), or disbelieve it ($B\neg q$), or suspect judgment on it ($\neg Bq \& \neg B\neg q$). If they are uncertain what they believe, how do we represent their belief-state?

The answer was famously given by Hintikka, Kripke, and others in the form of epistemic and modal logic, and the "Kripke frames" that form their backbone (Hintikka 1962; Kripke 1963). The claim that you believe q is itself a proposition—it's a way the world could be; it's something that I can be uncertain about. When you're uncertain whether you believe q, you leave open possibilities where that proposition is true, and possibilities where it's false. So to represent what you believe, we need each possibility in our model to determine what you believe. In the simplest model, what you believe is determined just by the set of possibilities you leave open: you believe q iff all of these possibilities are q-possibilities.

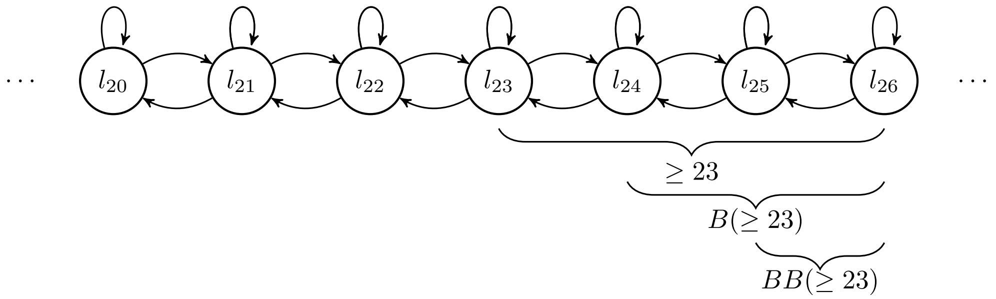
Precisely, let W be our background set of possibilities. What we need, then, is a *function* $\mathcal{B}$ from worlds w to sets of possibilities $\mathcal{B}_w$ consistent with your beliefs in w. (Note the similarity between $\mathcal{B}$ and a random variable.) Thus $\mathcal{B}$ is, effectively, a *description* of your beliefs: "the set of possibilities you leave open, whatever they are".

That pair $(W, \mathcal{B})$ is all we need. Propositions about your beliefs are defined in terms of $\mathcal{B}$: the set of worlds where you believe q, Bq, is $Bq = \{w \in W : \mathcal{B}_w \subseteq q\}$. This is a proposition like any other, so we can do all our normal set-theoretic operations on it: the proposition where you believe q and where r is true is just $Bq \cap r$, etc. The magic of this is that it allows us to "unravel" any higher-order belief claims into simply claims about what sets of possibilities you rule in and rule out. For example, it follows from the above definition that

[24] You might think that we can just restrict Reflection to other propositions q that aren't about P's opinions. But that won't do. If P is an expert about q, which opinions it has about q will be correlated with q, so its lack of knowledge about its own opinions will also make it less informed *about* q than you are once you learn what its opinions are. More on this below.

the set of worlds where you believe that you don't believe q is $B\neg Bq = \{w : \mathcal{B}_w \subseteq \neg Bq\}$, where $\neg Bq = W - \{w : \mathcal{B}_w \subseteq q\}$.

Here's a concrete example. Suppose you glance at the tree in the distance, and come to form some beliefs about how it looks. What exactly those beliefs are depends on how it looks, but you can't fully introspect how it looks so you're not fully opinionated about that. In particular, let l_i be the claim that *it looks i feet tall (to you, now).* If it looks i feet tall, then the strongest proposition you believe (the smallest set of possibilities you leave open) is that it looks between $i-1$ and $i+1$ feet tall. Then, letting an arrow from x to y indicate that at x you leave open that you're at y (i.e. $y \in B_x$), we can diagram your situation as follows:



For example, $[\geq 23]$ is the proposition that it looks at least 23 feet tall. You believe this at worlds $l_{24}, l_{25}, ...$ You believe that you believe this at worlds $l_{25}, l_{26}, ...$ Thus at world l_{24} you believe that it looks at least 23 feet tall, but you don't believe that you believe it since you leave open that you're at l_{23}, where you in turn leave open l_{22} and therefore don't believe that it looks at least 23 feet tall. (At l_{24}, $Bq\&\neg BBq$ is true.)

Generally, we can model constraints on your beliefs as constraints on the set of Kripke frames like this that are admissible. For instance suppose we want to impose the *positive introspection* constraint that if you believe q, you believe that you believe q. This means requiring that for all q, $Bq \to BBq$ is true at every world in the frame. That, in turn, is equivalent to saying that the "leaves open" relation is *transitive*: if at world x you leave open y ($y \in B_x$) and at world y you leave open z ($z \in B_y$), then at world x you leave open x ($z \in B_x$). (Note that the above frame is not transitive.)

This is how we model *categorical* higher-order uncertainty: lack of belief about what you believe.

5.2.2 Probability frames

Things are exactly analogous when we turn to modeling *quantitative* higher-order uncertainty: lack of certainty about what your degree of uncertainties are.

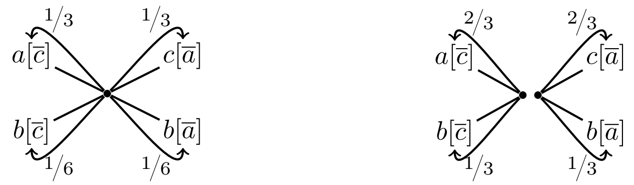
In the belief case, the basic object for representing your beliefs was a set $\mathcal{B}_w$: you believe q iff $\mathcal{B}_w \subseteq q$. In the probabilistic case, the basic object for representing your beliefs is a probability function π: you're t-confident of q iff $\pi(q) = t$. So just as in the belief case we needed a set of worlds and a function from worlds to sets of worlds, the analogous move is to take a set of worlds and a function from worlds to probability distributions over worlds.

Formally, a **probability frame** (W, P) is a (finite) set of worlds and a function P from worlds w to probability distributions P_w defined over the subsets of W. As before, if we put the worlds in some order, $W = \{w_1, ..., w_n\}$, we can think of a probability function as a vector $\pi = (\pi_1, ..., \pi_n)$ where $\pi_i = \pi(w_i)$ is the probability assigned to world i. As always, logic is done via set theory: given two propositions $p, q, \subseteq W$, $p\&q = p \cap q$, $p \to q = \neg p \cup q$, etc.

As with epistemic logic above, the crucial step is to use the function P to define claims about probabilities so that they are sets of worlds in the frame—and, therefore, automatically get assigned probabilities themselves. The strategy is the obvious one: the claim that P has some property ϕ picks out the set of worlds w where P_w has property ϕ. For example, the claim that P assigns t-probability to q is $[P(q) = t] := \{w \in W : P_w(q) = t\}$. The claim that P assigns higher credence to p than to q conditional on r is $[P(p|r) > P(q|r)] := \{w \in W : P_w(p|r) > P_w(q|r)\}$. The claim that P exactly matches the (rigidly-specified) probability function π in all it's verdicts is $[P = \pi] := \{w \in W : P_w = \pi\}$. Once we specify W and P, all of these sets are automatically well-defined. As a result, they automatically get assigned probabilities: the probability at w that P assigns probability t to q is simply $P_w(P(q) = t) = P_w(\{w : P_w(q) = t\})$. As a further result, that means claims about *probabilities about probabilities* are also well-defined as propositions: the claim that P makes it t-likely that P makes it s-likely that q is just the claim $[P(P(q) = s) = t]$; unpacking our definitions, the inner probability claim is $\{w : P_w(q) = s\}$, so the full proposition is $\{x \in W : P_x(\{w \in W : P_w(q) = s\}) = t\}$—which, again is just a set of worlds. Thus every world w will assign some probability to this claim—there will be some t' such that $P_w(P(P(q) = s) = t) = t'$—and so on all the way up.

This is all terribly abstract. Let's look at some examples. In fact, we've already seen some.

First example is our model of the train case. We can represent a probability frame in two different ways: (1) a *generalized Kripke frame*, or (2) a *stochastic matrix.* A generalized Kripke frame looks like our above Kripke frame of the tree, except the arrows are now labeled: an arrow labeled t from x to y indicates that the probability function at world x assigns t-probability to world y: $P_x(y) = t$. To avoid them getting too cluttered, it's convenient to have the arrows merge before being labeled and directed to their target when two worlds agree on the relevant probabilities, and to draw no arrow when the probability is 0. For example, here's how we're diagram your prior and posterior opinions in the (proper, where you condition on what you're told) version of the train case:



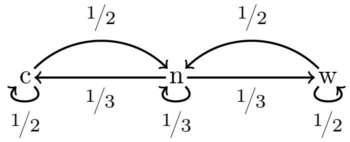
We can equivalently represent this in stochastic matrices, as we've seen above, where row i column j represents $P_i(j)$:

$$\mathsf{P} = \left(\begin{array}{cccc} a[\overline{c}] & b[\overline{c}] & b[\overline{a}] & c[\overline{a}] \\ \hline 1/3 & 1/6 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/6 & 1/3 \end{array}\right) \qquad P = \left(\begin{array}{cccc} a[\overline{c}] & b[\overline{c}] & b[\overline{a}] & c[\overline{a}] \\ \hline 2/3 & 1/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{array}\right)$$

Both of these probabilities are higher-order certain, since whenever $P_w(x) > 0$, $P_w = P_x$. This can be seen efficiently in the Kripke frame since the worlds are grouped into sets of worlds whose arrows merge; it can be seen efficiently in the stochastic-matrix notation because the matrix is "block diagonal", wherein there are blocks of nonzero values from top left to bottom right which are all equal to each other.

Let's work through some of the abstract reasoning in this case. $[P(c) = \frac{2}{3}] = \{c[\overline{a}], b[\overline{a}]\}$. Thus at the world where a got the ticket and you're told $\neg c$, you assign 0 probability to the claim that you assign $\frac{2}{3}$ probability to c getting the ticket: $P_{a[\overline{c}]}(P(c) = \frac{2}{3}) = P_{a[\overline{c}]}(\{c[\overline{a}], b[\overline{a}]\}) = 0$. (You know that you've ruled out c!). Similarly, at the world where c got the ticket and you've been told $\neg a$, you assign probability 1 to the claim that you assign $\frac{2}{3}$ probability to c getting the ticket: $P_{c[\overline{a}]}(P(c) = \frac{2}{3}) = 1$. (You know that you assign $\frac{2}{3}$ to c getting the ticket.) Meanwhile, $[P = (0, 0, \frac{1}{3}, \frac{2}{3}) = \{b[\overline{a}\, c[\overline{a}]\}$—the claim that you have exactly the probability function $(0, 0, \frac{1}{3}, \frac{2}{3})$ is the same set of worlds as the claim that you assign $\frac{2}{3}$ to c in this frame. Thus, again, $P_{a[\overline{c}]}(P = (0, 0, \frac{1}{3}, \frac{2}{3})) = 0$ and $P_{c[\overline{a}]}(P = (0, 0, \frac{1}{3}, \frac{2}{3})) = 1$.

Let's do a more interesting case, involving higher-order uncertainty. For example the Williamsonian one where you're unsure whether you feel cold; let's just focus on the prior:



$$\mathsf{P} = \left(\begin{array}{ccc} c & n & w \\ \hline 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{array}\right)$$

At the world where you feel cold (c), you're 50-50 on whether you feel cold or feel neutral, and know you're not warm. At the world where you feel neutral, you're uniform over the possibilities. Therefore, at the world where you feel cold, you're 50-50 on whether you're (at c, so) 50-50 between c and n, or (at n, so) uniform over c, n, and w. More precisely,

$[\mathsf{P} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})] = \{n\}$, so $[\mathsf{P}(\mathsf{P} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})) = \frac{1}{2}] = \{c, w\}$, i.e. at both c and w you're 50%-confident that you're uniform. Meanwhile, at n, you're only $\frac{1}{3}$-confident that you're uniform: $\mathsf{P}_n(\{n\}) = \frac{1}{3}$.

How is it that at c you're stably $\frac{1}{2}$ confident that you're at n, even though $\mathsf{P}_c \neq \mathsf{P}_n$? Because at c you don't *know* that you're $\frac{1}{2}$-confident that you're at n; you think there's a half chance you're $\frac{1}{3}$! Note therefore that at c you're only $\frac{1}{2}$-confident that you're $\frac{1}{2}$-confident that you're $\frac{1}{3}$-confident you're at c: $[\mathsf{P}(\mathsf{P}(\mathsf{P}(c) = \frac{1}{3}) = \frac{1}{2}) = \frac{1}{2}]$ is true at c, since we can "unravel" this higher-order claim to $[\mathsf{P}(\mathsf{P}(n) = \frac{1}{2}) = \frac{1}{2}]$, which in turn is $[\mathsf{P}(\{c, w\}) = \frac{1}{2}]$, which in turn is $\{c, w\}$, and so true at c.

This is fully general. In well-behaved higher-order models (ones that obey "New Reflection"—see below), whenever you have second-order uncertainty (P is uncertain what $P(q)$ is), you also have third-order uncertainty (P is uncertain what $P(P(q) = t)$ is, for various t), and so on all the way up. This is why the uncertainty is stable: if you knew what your higher-order distribution was, you could use it to infer what your lower-order distribution is; but you don't, so you can't.

Since we're considering an interpretation of credence as dispositions, i.e. as *averages* of the degrees to which you'd be willing to act on the proposition, let's look at a simple example of how averaging two introspective probabilities could lead that average-probability function to fail to be introspective. Note that the set of probability functions is closed under (weighted) averaging: if π and δ are both probability functions then so is $\rho := x \cdot \pi + (1 - x) \cdot \delta$, the weighted-average between them defined so that $\rho(q) = x \cdot \pi(q) + (1 - x) \cdot \delta(q)$. With vectors, we do this just by averaging their components, e.g. $\frac{1}{2}(\frac{1}{2}, \frac{1}{2}, 0) + \frac{1}{2}(0, \frac{1}{2}, \frac{1}{2}) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

Here's a simple game. I'm going to flip a coin twice, showing the first outcome to Caspar and the second outcome to Roger. Thus there are four possible outcomes, TT, TH, HT, HH; those outcomes determine what each of Caspar's and Roger's credence over the outcomes are. They're good (Clear) Bayesians, so they update by becoming certain of what they saw and being 50-50 on what the other person saw. Thus Caspar and Roger's probability frames, C and R, are:

$$C = \left(\begin{array}{cccc} TT & TH & HT & HH \\ \hline {}^1\!/_2 & {}^1\!/_2 & 0 & 0 \\ {}^1\!/_2 & {}^1\!/_2 & 0 & 0 \\ 0 & 0 & {}^1\!/_2 & {}^1\!/_2 \\ 0 & 0 & {}^1\!/_2 & {}^1\!/_2 \end{array}\right) \qquad R = \left(\begin{array}{cccc} TT & TH & HT & HH \\ \hline {}^1\!/_2 & 0 & {}^1\!/_2 & 0 \\ 0 & {}^1\!/_2 & 0 & {}^1\!/_2 \\ {}^1\!/_2 & 0 & {}^1\!/_2 & 0 \\ 0 & {}^1\!/_2 & 0 & {}^1\!/_2 \end{array}\right)$$

Each of these probabilities are introspective. But *the average of Caspar and Roger's opinions* is not. That average is:

$$A := \frac{1}{2}C + \frac{1}{2}R = \left(\begin{array}{cccc} TT & TH & HT & HH \\ \hline 1/2 & 1/4 & 1/4 & 0 \\ 1/4 & 1/2 & 0 & 1/4 \\ 1/4 & 0 & 1/2 & 1/4 \\ 0 & 1/4 & 1/4 & 1/2 \end{array}\right)$$

This is as it should be. Since each knows that *they* saw but wonders what the *other* person saw, they both wonder what the *average* of their opinions is. Since averages preserve features that they both share, this means the average of their opinions wonders what the average of their opinions is. For example, suppose it lands TT. Then Caspar wonders whether Roger saw a T or a H—if the former their at TT so their average is $(1/2, 1/4, 1/4, 0)$, if the latter they're at TH so it's $(1/4, 1/2, 0, 1/4)$. Meanwhile, Roger wonders whether Caspar saw a T or a H; if the former they're at TT to their average is $(1/2, 1/4, 1/4, 0)$, if the latter they're at HT so their average is $(1/4, 0, 1/2, 1/4)$. Thus the average of them (A) is itself uncertain between these three possibilities. And since the average varies across possibilities, the average opinion is uncertain what the average opinion is! For example, $A_{TT}(TT) = \frac{1}{2}$, but $A_{TT}(A(TT) = \frac{1}{2}) = \frac{1}{2}$, while $A_{TT}(A(TT) = \frac{1}{4}) = \frac{1}{2}$ as well: if both tosses land tails, the average is 50-50 on whether the other person saw tails (so they both leave open TT), or the other person saw heads (so only one of them leaves open TT).

(Perhaps it's worth considering a different case. We each form an estimate about the number of jellybeans in a jar. We each wonder whether and how far our estimate is from the average estimate, since we wonder what the other people estimated. Unbeknownst to us, by a fluke we all gave the same estimate (and, let's stipulate, have the same distribution over what the others' estimates might be)—thus each of our opinions *equals* the average opinion. But we obviously don't know that—the average person in the room can fail to know that they're the average person in the room.)

As a sanity check for the claim I made above, we can note that if we "draw samples" from this average distribution (like forcing yourself to write down a fair price), this doesn't necessarily remove higher-order uncertainty. This is obvious in the jellybeans case: if (unbeknownst to us) we all had the same estimate of 504, and then we learned that *someone*'s estimate was 504, we'd each slightly increase our confidence that 504 is the average, but continue to be quite unsure what the average is.

Similarly in the coin case. Suppose a random person is selected and that person's credence in TT is announced. This will be either 0.5 or 0. If it's 0.5, both of Caspar and Roger will condition on that. If one of them was 0.5 in TT, they're unsure whether this random announcement was *their* credence or the other person's—so it provides some evidence that the other person also saw tails. If their credence in TT was 0, that means they saw heads and they now know the other person must've seen tails.

Letting 0 mean credence 0 is announced and 5 mean 0.5 is announced, their priors and average over the expanded question of what the coins did *and* what will be announced are:

$$C = \left(\begin{array}{cccccc} TT_5 & TH_5 & TH_0 & HT_5 & HT_0 & HH_0 \\ \hline 1/2 & 1/4 & 1/4 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/2 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/2 \end{array}\right)$$

$$R = \left(\begin{array}{cccccc} TT_5 & TH_5 & TH_0 & HT_5 & HT_0 & HH_0 \\ \hline 1/2 & 0 & 0 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 1/4 & 0 & 0 & 1/2 \\ 0 & 1/4 & 1/4 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/4 & 1/4 & 0 \\ 1/2 & 0 & 0 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 1/4 & 0 & 0 & 1/2 \end{array}\right)$$

$$A = \left(\begin{array}{cccccc} TT_5 & TH_5 & TH_0 & HT_5 & HT_0 & HH_0 \\ \hline 1/2 & 1/8 & 1/8 & 1/8 & 1/8 & 0 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 1/4 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 1/4 \\ 1/4 & 0 & 0 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 1/8 & 1/8 & 1/8 & 1/8 & 1/2 \end{array}\right)$$

Updating on what's announced yields:

$$C' = \left(\begin{array}{cccccc} TT_5 & TH_5 & TH_0 & HT_5 & HT_0 & HH_0 \\ \hline 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 \end{array}\right)$$

$$R' = \left(\begin{array}{cccccc} TT_5 & TH_5 & TH_0 & HT_5 & HT_0 & HH_0 \\ \hline 2/3 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 2/3 \\ 2/3 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 2/3 \end{array}\right)$$

$$A' = \tfrac{1}{2}C' + \tfrac{1}{2}R' = \left(\begin{array}{cccccc} TT_5 & TH_5 & TH_0 & HT_5 & HT_0 & HH_0 \\ \hline {}^2/_3 & {}^1/_6 & 0 & {}^1/_6 & 0 & 0 \\ {}^1/_3 & {}^2/_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & {}^2/_3 & 0 & 0 & {}^1/_3 \\ {}^1/_3 & 0 & 0 & {}^2/_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & {}^2/_3 & {}^1/_3 \\ 0 & 0 & {}^1/_6 & 0 & {}^1/_6 & {}^2/_3 \end{array}\right)$$

Notice that the average's higher-order uncertainty remains—for example, in TT_5, it's $\frac{2}{3}$-confident that TT_5, but leaves open TH_5, where instead it's $\frac{1}{3}$-confident that TT_5. (If both Caspar and Roger saw tails, then when the announcement comes that a random one of them is 0.5 in TT, they each increase their credence that they both saw tails, but leave open that maybe the random person was themself, in which case the other person now knows how both coins landed.)

5.2.3 Constraints on probability frames

We can look at the above frames to see why Reflection must fail in contexts of higher-order uncertainty. For example, consider the average-credence cases. You're uniform over how the coin has landed: $\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Thus your estimate for the number of heads, X, equals $\mathbb{E}_\pi(X) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) \cdot (0, 1, 1, 2) = 1$. You're not sure what the average estimate $\mathbb{E}_A(X)$ is; if TT, it's $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0) \cdot (0, 1, 1, 2) = 0.5$. If TH or HT, it's 1, and if HH, it's 1.5. What's your estimate *conditional on* the *average estimate*, $\mathbb{E}_A(X)$, being 1.5? That's true only if HH, so you can infer that it landed heads twice: $\mathbb{E}_\pi(X|\mathbb{E}_A(X) = 1.5) = 2$. Thus you violate an expectations-version of the local Reflection principle: $\mathbb{E}_\pi(X|\mathbb{E}_A(X) = t) \neq t$.

Why don't you defer to the average opinion about the number of heads? Because the average opinion doesn't *know* what the average opinion is—if $\mathbb{E}_A(X) = 1.5$, that's because both Caspar and Roger think, "Well *I* saw heads, but I don't know what the other guy saw; since I'm 50-50 that he saw heads or tails, my expectation for the number of heads is $0.5 \cdot 1 + 0.5 \cdot 2 = 1.5$." Their estimates are 1.5 because they leave open that the other person didn't see heads—in which case the average estimate is *not* 1.5, but instead is 1. Thus when *you* learn that the average estimate is 1.5, you learn something that this average didn't know, and therefore can infer more—namely, that they *both* saw heads.

So Reflection fails, as it must whenever probabilities are not introspective, i.e. don't satisfy Clear Bayes. Still a weakening of Reflection called "New Reflection" does hold (Elga 2013). Here's the idea: if the reason you don't completely defer to A is that A doesn't know what A is, then when you learn what A is, what you should defer to is what A *would* do, if it were to learn what you learned. Precisely:

> **New Reflection:** $\mathsf{P}(\cdot|P = \pi) = \pi(\cdot|P = \pi)$
> Conditional on P having a particular set of opinions (π), adopt the opinions that *it* would adopt upon learning this fact.

Define the *informed* version of P, $\widehat{P}$, to be the opinions P would have if it learned what P was: at world w, $\widehat{P}_w(\cdot) = P_w(\cdot|P = P_w)$. Then New Reflection is equivalent to the claim that your beliefs equal your expectation of informed-P's beliefs: $\mathsf{P}(\cdot) = \mathbb{E}_{\mathsf{P}}(\widehat{P}(\cdot))$ (Stalnaker 2019; Dorst 2019).

π obeys New Reflection toward A, due to the fact that once A learns what A is, it knows exactly how the coin landed, due to the fact that each way the coin landed picks out a different average opinion. For example, $\pi(\cdot|A = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)) = (1, 0, 0, 0) = A_{TT}(\cdot|A = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0))$.

Notice that although π doesn't reflect A, it does obey a weakening of reflection. Conditional on $\mathbb{E}_A(X)$ being to some degree deviated from π's estimate of 1, π goes in the same direction *but more so.* In particular:

$\mathbb{E}_\pi(X) = 1$, while:
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) = 1.5) = 2$
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) = 1) = 1$
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) = 0.5) = 0$

As a consequence of this, π obeys a weaker Reflection-like principle toward A's estimate of X: it "trusts" A's binary judgments about whether X is high or not, for every threshold of what counts as 'high'; and likewise for 'low' judgments:

$\mathbb{E}_\pi(X|\mathbb{E}_A(X) \geq 1.5) \geq 1.5 \quad (= 2)$
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) \geq 1) \geq 1 \quad (= \frac{4}{3})$
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) \geq 0.5) \geq 0.5 \quad (= 1)$; and

$\mathbb{E}_\pi(X|\mathbb{E}_A(X) \leq 1.5) \leq 1.5 \quad (= 1)$
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) \leq 1) \leq 1 \quad (= \frac{2}{3})$
$\mathbb{E}_\pi(X|\mathbb{E}_A(X) \leq 0.5) \leq 0.5 \quad (= 0)$

This is a 'local' version of the Trust principle that we'll see below yields the value of evidence when applied generally (Dorst 2020; Dorst et al. 2021).

We'll see other constraints we might want to impose on probability frames later on, but we've seen enough to turn to what models of our ambiguous judgments might look like.

5.2.4 Ambigous features

Here you are, wondering how many socks I have. You're also wondering how many times the die will land 1–4 out of 30 tosses, but you know that your estimate for that should be 20. How do we model the ambiguity in your estimate about my number of socks?

The basic idea is to allow some parameter to vary which determines what your estimate is, but to allow you to be uncertain what that parameter is. The fact that the parameter determines your estimate means there will be precise facts about what your estimates are. The fact that you're unsure what the parameter is—and hence unsure what your estimate is—will explain our features of ambiguity. Hopefully.

Before giving a concrete model, let me show abstractly how we'll get each feature—or something near enough.

The basic strategy: as with the imprecise model, we now have access to a set of precise credence functions rather than just a single one. But that set comes *from* that single one—namely, it's the set of credence functions that the single credence function leaves open. That is, if P_w is your credence function at world w, we have both P_w (a rigidly-specified precise credence function), but also $S = \{\pi : P_w(P = \pi) > 0\}$ (the set of values of P that P_w leaves open, i.e. those in the *support* of P_w). Whenever P_w has higher-order uncertainty, S will be a non-singleton, and the range it covers will be the range of precise opinions that you leave open that you have.

Irresolute assessments can be gotten in at least two ways.

The first is from the interpretation: if we're interpreting your credences and estimates as *averages* of the numbers you'd write down if forced, then whenever there's any variance in the distribution of numbers you write down, there'll be some stochasticity and felt arbitrariness in your judgment.

The second is directly from the introspection failures. Suppose $\mathbb{E}_P(X) = 20$, but $P(\mathbb{E}_P(X) = 20)$ is low, and (say) in fact $P(\mathbb{E}_P(X) = 21) = 0.2$ and $P(\mathbb{E}_P(X) = 19) = 0.2$. This may lead to you sometimes mis-stating your estimate—perhaps in a way that's proportional to how likely you think your estimate is that particular mis-statement—e.g. 20% of the times you're probed, you'll say that your estimate is 21. (This is sometimes called *probability-matching* in reporting: being disposed to give an answer q in proportion to how likely you think q is the right answer.)[25]

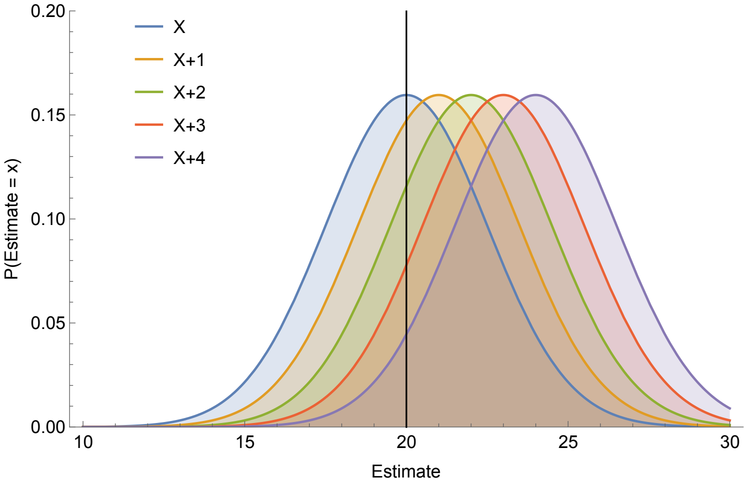
Insensitivity can't strictly-speaking be recovered, since your probabilities and estimates are precise, so if $P(q) = P(r)$, then $P(q \vee a) > P(r)$. (Where a is the proposition that the penny will land heads 10 times in a row.) But the higher-order model captures much of the intuitive judgment by saying you can be massively unsure about these facts. Recall that the imprecise model says "you confidence in q is on a par with your confidence in r, and your confident in r is on a par with your confidence in $q \vee a$, but your confidence in $q \vee a$ is greater than your confidence in q". As a substitute, the higher-order model says: "You're unsure whether you're more confident of q or of r, and you're unsure whether you're more confident of $q \vee a$ or of r, but you're certain that you're more confident of $q \vee a$ than of q.

An advantage of this is that it automatically encodes the apparent gradability of how "incomparable" propositions are (Hájek and Rabinowicz 2022). Suppose Caspar's trying to cheer me up from my latest racing debacle by giving me pairs of running socks. At t_0 you know I own X socks, and you have some ambiguous estimate for what X is—say, 20. Every minute, Caspar gives me another pair of socks, so you know that at t_1 I have $X + 1$ pairs, and t_2 I have $X + 2$ pairs, and so on. By hypothesis, it's unclear whether your estimate

[25] [NOTE TO SELF: averages of informed credences? Maybe is $P(\widehat{P} = \pi) = t$, you answer $\mathbb{E}_\pi(X)$ with probability t. Or rather, if $P(\mathbb{E}_{\widehat{P}}(X) = s) = t$, then there's probability t that you'll answer s to the estimate of X.]

for my original (X) number of pairs of socks is greater or less than your estimate for the number of 1–4 rolls of the die (20). By the time Caspar has given me 15 pairs of socks, it's clear that your estimate for $X + 15$ is higher than 20. And if we observe you struggling to make comparisons, presumably what we'll find is that at t_0, t_1, and t_2 you're very torn. By the times t_3 and t_4 roll around, you're starting to lean more reliably toward thinking that $\mathbb{E}_P(X + 4) > 20$. And so on.

This is automatic on the higher-order uncertainty picture. Supposing you start out with the below blue higher-order distribution for what your estimate of X is, it follows that your distribution for my number of socks at later times is progressively shifted to the right:



So although you're initially 50-50 between whether you have a higher estimate for my number of socks than your estimate for the number of 1–4 rolls, i.e. 20 ($P(\mathbb{E}_P(X) > 20) = 0.5$), by the time Caspar gives me 4 pairs of socks you're 95%-confident that you have a higher estimate for the number of pairs of socks I now have: $P(\mathbb{E}_P(X + 4) > 20) \approx 0.95$.

Fuzzy boundaries. Relatedly, the higher-order model is better-equipped to explain the fact that the ambiguities in your assessments have fuzzy boundaries. Recall that the problem was that since your credal committee has sharp boundaries, there are precise points where the imprecision gives out. Thus if your estimate for my number of socks is determinately higher than 15 and lower than 25 (i.e. $\mathbb{E}_{\mathbb{P}}(X) = [15, 25]$), then there's a first time when you determinately have a higher estimate for my number of socks than for the number of 1–4 rolls, namely t_6. Equivalently: at the original time there's a clear greatest lower bound on your estimate of my number of socks—namely 15. The imprecise model can try to avoid this by somehow introducing higher-order indeterminacy or imprecision, but the relevant point of contrast is that the higher-order model gets this for free.

Suppose $P(15 \leq \mathbb{E}_P(X) \leq 25) = 1$, i.e. you're sure that your estimate is between 15 and 25. Let's further suppose that this is the *strongest* thing about $\mathbb{E}_P(X)$ you're sure of—there's no narrower interval $[l, h] \subseteq [15, 25]$ such that you're sure $l \leq \mathbb{E}_P(X) \leq h$. Let's write that fact $sup_{P(E_P(X))} = [15, 25]$, for P's *support* (the range of possibilities it leaves open).[26] It follows that there's a greatest lower bound on what you think your estimate might be—namely, 15. But it *doesn't* follow that this greatest lower bound is clear. Indeed, as we've seen, when you have second-order uncertainty (P is unsure what $\mathbb{E}_P(X)$) is, this is always accompanied by *third*-order uncertainty (P is unsure what $P(\mathbb{E}_P(X) = t)$ is, for various t). As a result, most sensible models of your uncertainty will say that when the strongest thing P is certain of is that $\mathbb{E}_P(X)$ is between 15 and 25, you're not certain of this fact. For example, a Williamsonian margin-for-error principle would say that at worlds w where $\mathbb{E}_P(X) = 15$, you're not sure that $E_P(X)$ isn't a bit lower: $P_w(\mathbb{E}_P(X) < 15) > 0$. Thus even if the greatest lower bound on your estimate is 15, you leave open that your estimate *is* 15, in which case the greatest lower bound on your estimate is at least a bit lower: when $sup_{P(E_P(X))} = [15, 25]$, $P(sup_{P(E_P(X))} = [15, 25]) < 1$.[27] That means the boundaries of your ambiguity are themselves ambiguous: it's unclear whether the greatest point n that your estimate for X is higher than is 15, or 14, or 16.

Ambiguity Aversion. This one's subtle. There are definitely forms of ambiguity aversion that fall our easily from the higher-order uncertainty approach. For instance, sometimes gaining ambiguity can lead you to become predictably less accurate in your estimates, so you of course would like to avoid forming or relying on such ambiguous estimates (Ahmed and Salow 2018). For example, ambiguity can lead to updates that look like our first model of the train case, where you know the update is going to raise your credence in q—so you'd prefer not to rely on your posterior after the update.

How (or if) this can be extended to the specific version of ambiguity aversion given by our bets is something I'm still thinking through. Recall that, empirically, where q is the claim that I have at least 20 socks, people often prefer Bet1 to both Bet2 and Bet3:

Bet1: \$100 if *heads*, \$0 if *tails*.
Bet2: \$100 if q, \$0 if $\neg q$.
Bet3: \$100 if $\neg q$, \$0 if q.

On the simple theory that, given higher-order uncertainty, you should maximize expected utility given your credences (even if you're unsure what they are), it follows that either you're indifferent between all three (because $P(q) = P(\neg q)$, or you prefer one to the other: if $P(q) > 0.5$, then $\mathbb{E}_P(\mathit{Bet1})$ is highest; if $P(q) < 0.5$, $\mathbb{E}_P(\mathit{Bet2})$ is highest.

What else could we say? The analog of the CONSERVATIVE decision rule from imprecise probability that yields ambiguity aversion, ported over to higher-order uncertainty, says that

[26] $sup_{P(E_P(X))} = [l, h]$ is the proposition $\{w \in W : P_w(l \leq E_P(X) \leq h) = 1$ and $P_w(l < E_P(X)) < 1$ and $P_w(E_P(X) < h) < 1\}$.

[27] The margin-for-error argument implies that the "x leaves open y" relation ($P_x(y) > 0$) is non-transitive, which in turn implies failures of the value of evidence (Dorst 2020). But the same point can be made without transitivity failures: even if $P(15 \leq E_P(X) \leq 25) = 1$ implies that $P(P(15 \leq E_P(X) \leq 25) = 1) = 1$ (if you're sure $E_P(X)$ is between some values, then you're sure that you're sure of it), you still needn't know *exactly* what possibilities you leave open—transitive but non-Euclidean probability frames will be such that you're unsure whether you're sure of something stronger, and hence that the greatest lower-bound is *not* 15 after all.

if you leave open that the expected value of X is as low as t, then you shouldn't value it at more than a sure t. That would imply that you should prefer Bet1, since you know it's expected value is 0.5, while you leave open that the expected value of each of Bet2 and Bet3 is lower than that.

An alternative strategy is to think of the choice really as between an unambiguous estimate (Bet1), and doing one of the ambiguous ones (Bet2 or Bet3)—who knows which. Suppose you know that you'll maximize expected value if you do the latter—but you're not sure which one maximizes expected value, since you don't know what your credences are. Then if you don't trust your own credences—if $P(q|P(q) > 0.5) < 0.5$, as can happen given higher-order uncertainty (Dorst 2020, §6)—then you'll prefer to take Bet1 than to risk choosing incorrectly between Bet2 and Bet3. This allows that people who *do* trust their judgments might prefer the ambiguous options, which fits with the generalization of ambiguity-aversion known as "home bias" (Trautmann and van de Kuilen 2015): people who aren't familiar with reasoning in a certain domain tend to be ambiguity averse, while those who *are* familiar tend to be ambiguity *seeking* (preferring to bet on claims like Bet2 or Bet3 when they have some expertise about q-like propositions).

There are other strategies, too.[28] What should I say here?

Biases and miscalibration are easier, due to the fact that Reflection fails whenever you have higher-order uncertainty. Here you are, with your credences P. You're unsure what your own opinions about q are (let alone what you're going to write down if forced to name your betting price), so $P(P(q) = t) < 1$ for all t. Conditional on you being 50%-confident in q, how likely is it that q is true? Since Reflection fails, we are not obligated to say 50%: it's possible (in fact common) that $P(q|P(q) = 0.5) \neq 0.5$. For example, take the average-credence-function A from above. If the coin lands TT, then $A(TT) = 0.5$ and $A(TH \vee HT) = 0.5$. But A doesn't know that it has these values: A is 50-50 over whether $A(TT) = 0.5$ or instead $A(TT) = 0.25$. As a result, *conditional* on A assigning 0.5 to these propositions, A acts very differently. $A_{TT}(TT|A(TT) = 0.5) = 1$, while $A_{TT}(TH \vee HT|A(TH \vee HT) = 0.5) = 0$.

More generally, if you have higher-order uncertainty, you might assign 0.5 to q only because you don't know that you assign 0.5 to q—and learning that you do could send your credence up or down. So suppose we gather up 100 independent q_i, each of which you assign 0.5 to. Is your expectation of the number that are true necessarily 50? No! Suppose that for each q_i, $P(q_i|P(q_i) = 0.5) = 0.6$—then your expectation for the number that are true will be 60! More generally, learning that you assign 0.5 to each could scatter your credences in each in all sorts of different directions. As a result, even if they're all independent, you can be quite doubtful that the proportion of them that are true will be anything close to 50%—you can expect to be miscalibrated.

You can even expect to be more blatantly biased. For example, you might currently be

[28] Something about cognitive processing. You know that further reflection/thinking would change your estimate about the value of Bet2 and Bet3; if you act on them, then you feel more pressure to do that cognitive effort...?

50%-confident of q, be certain that your probability won't drop, and leave open that it might go up (Dorst 2021). (This is what's known as a "no-lose investigation".) Here's a toy case:

$$\mathsf{P} = \left(\begin{array}{cc} q & \neg q \\ \hline 0.5 & 0.5 \\ 0.5 & 0.5 \end{array}\right) \qquad\qquad P = \left(\begin{array}{cc} q & \neg q \\ \hline 1 & 0 \\ 0.5 & 0.5 \end{array}\right)$$

Here you start out 50%-confident of q, are sure that you're credence won't drop, and leave open that it might go up (to 1). Notably, it'll do this only if q is true—so you know that your (biased) posterior will either be equally accurate or more accurate than your prior.

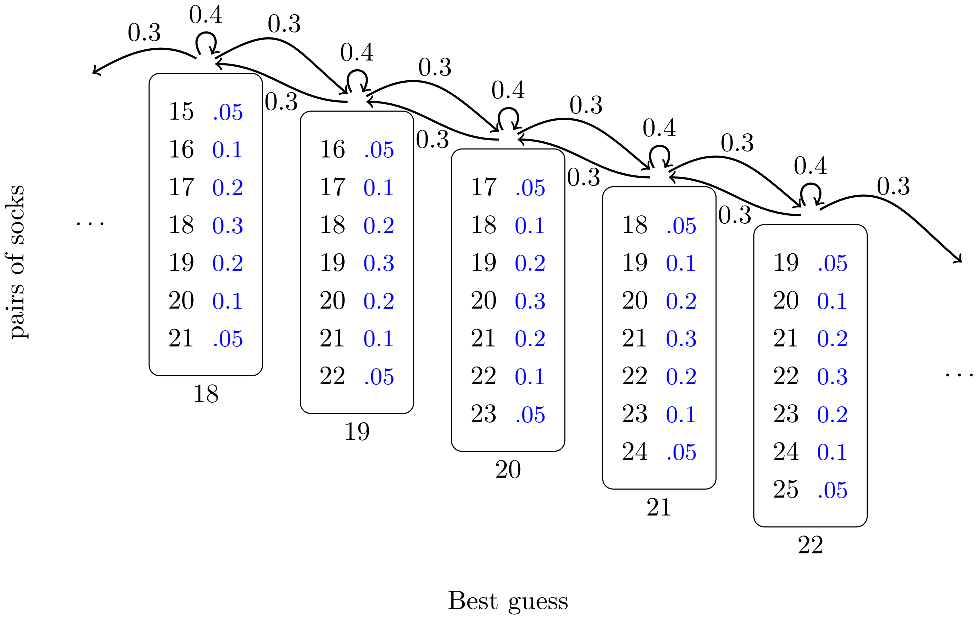
5.2.5 Models

Enough hand-waving; let's start drawing some models.

Let's give a toy model of you estimating the number of socks I own. For simplicity, let's deal with a case where in fact you're sure your estimate is between 19 and 21. Suppose you have an internal "signal"—an amalgamation of the little bits of evidence you have about the question—which fixes your best guess about the number of pairs of socks I have. Conditional on your best guess being n, you're sure that the number I have is $n \pm 3$. But you're unsure what your best guess is: if in fact it's n, you leave open that it's $n+1$ or $n-1$.

We can diagram this using the following structure. Here we slightly modify our Markov diagrams as follows. We partition the worlds into what your posterior P is, which is determined by your best guess (arrayed along the x-axis). Within each partition-cell q_i, blue numbers represent the conditional probability of each world w within that cell, $P(w|q_i)$. For instance, conditional on your best guess being 18, you have 0.05 credence that the true value is 15, 0.1 that it's 16, etc. The unconditional probabilities of each cell is then given by the labeled arrows between cells: if your best guess is 18, you're 0.4-confident that it is, and 0.3 confident that it's 19, etc. Unconditional probabilities for *worlds* are then gotten by multiplying through the probability of a cell times the conditional probability of a world given that you're in the cell—for instance, if you're best guess is 20 (you're in the middle column), the unconditional probability that *your best guess is 20 and the true value is 17* is $0.4 \cdot 0.05 = 0.02$; the unconditional probability that *your best guess is 19 and the true value is 18* is $0.3 * 0.2 = 0.06$, and so on.

(This representation of our model is possible because, in models that obey New Reflection, all worlds *agree* on these conditional probabilities; see Dorst 2019. Thus for all worlds w, $P_w(\cdot) = \mathbb{E}_{P_w}(\widehat{P}(\cdot))$, as mentioned—notice that $\widehat{P}$ at a given world is given by the blue numbers inside that world's partition-cell.)



What does this model imply? Suppose your best guess is 20, so you're in a world w in the middle column. Let b_i be the proposition that your best guess is i. Notice that $\mathbb{E}_w(X|b_{19}) = (0.05, 0.1, 0.2, 0.3, 0.2, 0.1, 0.05)\cdot(16, 17, 18, 19, 20, 21, 22) = 19$, and similarly for all b_i on which it's well-defined, $\mathbb{E}_w(X|b_i) = i$. Thus by total expectation the unconditional expectation at each b_i is also equal to i, due the symmetry of the uncertainty. For example, for $w \in b_{20}$:

$$\begin{aligned}\mathbb{E}_w(X) &= P(b_{19})\cdot\mathbb{E}_w(X|b_{19}) + P(b_{20})\cdot\mathbb{E}_w(X|b_{20}) + P(b_{21})\cdot\mathbb{E}_w(X|b_{21})\\ &= 0.3\cdot 19 + 0.4\cdot 20 + 0.3\cdot 21\\ &= 20\end{aligned}$$

How does this model do with our features?

Irresolute assessments: we haven't explicitly modeled noise here, though perhaps there's some way to. But we might think it comes directly from the higher-order uncertainty: when your estimate is 20, you're 60%-confident that it's either 19 or 21; so it's reasonable to think maybe you'd sometimes state 20, and other times state 19 or 21 as your estimates. (Perhaps when your best guess is in fact 20, you're 40%-likely to name 20 as your estimate, 30%-likely to name 19, and 30%-likely to name 21—the chances of you saying various estimates matches your probability of having those estimates.)

Insensitivity. You know that your estimate for the number of times the die rolls 1–4 is 20. Thus when your estimate for my number of socks is 19, that estimate is less than your estimate for the numbers of 1–4 rolls, but you're only 70%-confident that it's less (since

$P(\mathbb{E}_P(X) = 20) = 0.3$). Then when Caspar gives me one more pair of socks, you're now 30%-confident that your estimate for my new number of socks is less than your estimate for the number of 1–4-rolls. Note that if we make the degree of sweetening smaller relative to to the band of your uncertainty in your estimate, we can make cases where the degree to which you're uncertain before- and after- the sweetening is roughly the same. (E.g. both before and after the sweetening, you're still roughly 50-50 on which estimate is higher.)

Fuzzy boundaries. Suppose your estimate is 20. In this model, the strongest thing your sure about your estimate is that it's between 19 and 21. But you *don't* know the answer to the question, "What's the greatest lower bound on your estimate for Kevin's number of socks?". In fact, the answer to this is 19. But you're 30%-confident that your estimate *is* 19, in which case the answer to this question is 18. And you're 30%-confident that your estimate is 21, in which case the answer is 20. So although the true answer to the question is 19, you're 30%-confident that it's 18, 40% that it's 19, and 30% that it's 20. The same is true at all possibilities: you're always unsure what the greatest lower-bound on your estimate is, i.e. what the greatest number is that you're estimate is definitely higher than.

Ambiguity Aversion. Letting $q =$ *Kevin has more than 20 socks*, suppose your estimate is 20 and you're offered the following three bets:

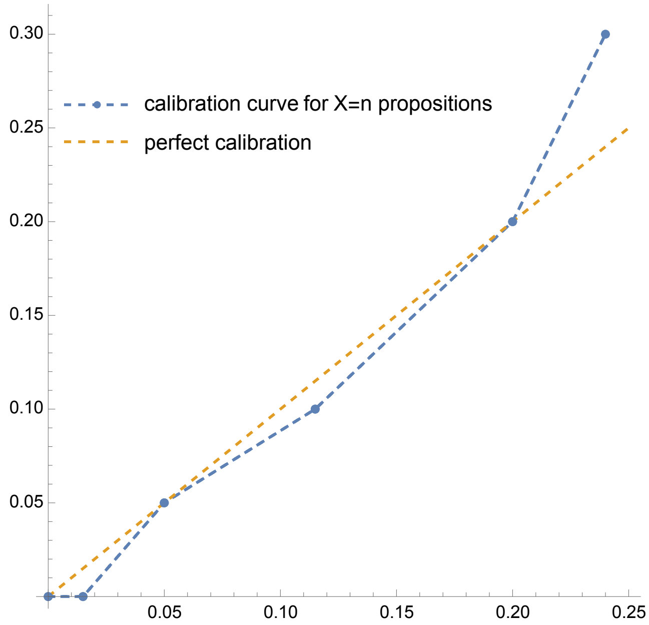
Bet1: \$100 if *heads*, \$0 if *tails*.

Bet2: \$100 if q, \$0 if $\neg q$.

Bet3: \$100 if $\neg q$, \$0 if q.

Your confidence in q is $0.3(0.1+0.05)+0.4(0.2+0.1+0.05)+0.3(0.3+0.2+0.1+0.05) = 0.38$, so your confidence in $\neg q$ equals 0.62. Thus your expectation for Bet3 is \$62, higher than that for Bet1 (=\$50). Still you're 30%-confident that you're at b_{21}, in which case your confidence in q is $P(q) = 0.62$, so $P(\neg q) = 0.38$ and your estimate for the value of Bet3 is only \$38. Thus when you think about taking Bet3, you worry that your expected value is lower than 50—an expected-value that you know you can achieve, by taking Bet1. Perhaps that can help explain ambiguity aversion. (...?)

Biases and miscalibration. Since Reflection fails in this model, you won't be perfectly calibrated. For example, notice that the highest credence you ever assign to a proposition of the form $[X = n]$ (e.g. *Kevin has exactly 20 socks*) is 0.24—e.g. at b_{20}, $P(X = 20) = 0.3(0.2) + 0.4(0.3) + 0.3(0.2) = 0.24$; everywhere also you assign lower credence to it. But in these worlds where you assign that credence, it's 30%-likely that $X = 20$ (since $P(X = 20|b_{20}) = 0.3$). Thus if we were to repeat this structure independently a bunch of times (with many different estimates), of all the times you were 24%-confident of a claim of the form $X = n$, it would be true 30% of the time. More generally, here's your calibration curve for specific propositions about the value of X, i.e. of the form $X = n$:



Note that it doesn't match the diagonal line: of all the things you're t-confident in, the expected proportion that are true is sometimes less and sometimes more than t.

There are a variety of issues with model that might move us to revise it. One of the most salient is that it leads to rather dramatic failures of the value of evidence. Let *odd* be the claim that the number of socks you'll estimate me to have will be odd, i.e. $[\mathbb{E}(X) \in \{1, 3, 5, ..., \}]$. Then you're anti-reliable about *odd*: in worlds where it's true, you assign 40% credence to it, and in worlds where it's false, you assign 60% credence to it. Thus, for instance, if you were initially 50-50 on whether your estimate would be odd (before thinking about my socks), and you had to decide on a bet about your estimate, you'd pay money not to think about the evidence and form posteriors in this way. (You'd know your prior of 50-50 on *odd* will be more accurate than your posterior.)

This is reminiscent of failures of Reflection and the value of evidence with imprecise credences. However, while (I believe) all imprecise decision theories (that don't mimic precise ones) lead to some form of a failure of the value of evidence, there *are* higher-order-uncertain updates that satisfy the value of evidence.

Even the above one satisfies it partially: for binary questions of the form *is your estimate above or below n?*, the update from the uniform prior will make your beliefs more accurate. Likewise for questions of the form *is Kevin's number of socks at least n?*

But more than that: some updates satisfy the value of evidence for *all* questions we could ask. That is, for every decision problem, the prior's expectation for the value the strategy of first doing the update and then maximizing expected value is always greater than it's expectation for ignoring the update. We've seen an example of such an update at the end of §5.2.4, where the prior was going to do a biased update but one that is guaranteed to only ever improve accuracy.

Generally, the value of evidence is satisfied when the prior obeys the following weakening of Reflection toward the posterior (Dorst et al. 2021):

> **Trusts**: For all X, t: $\mathbb{E}_{\mathsf{P}}(X|\mathbb{E}_{P}(X) \geq t) \geq t$
> Conditional on P having having a high estimate for t, have a high estimate for it.

The value of evidence will be *partially* satisfied if this holds for restricted classes of Xs. In particular, if every option X in the decision-problem satisfies this constraint, then [we know] the prior will expect the posterior to be more accurate about the values of each option on every proper accuracy scoring-rule, and [we think] the prior will expect the posterior to make a better decision.

6 Conclusion

Oof. What a journey.

I've made a preliminary, messy case for the claim that higher-order uncertainty is a better way of understanding ambiguous cases than imprecision is. Please help me figure out what's helpful and plausible, and what's not. Thanks!

References

Ahmed, Arif and Salow, Bernhard, 2018. 'Don't Look Now'. *British Journal for the Philosophy of Science*, To appear.

Belot, Gordon, 2013. 'Bayesian Orgulity'. *Philosophy of Science*, 80(4):483–503.

Blackwell, David, 1953. 'Equivalent Comparisons of Experiments'. *The Annals of Mathematical Statistics*, 24(2):265–272.

Blackwell, David and Dubins, Lester, 1962. 'Merging of opinions with increasing information'. *The Annals of Mathematical Statistics*, 33(3):882–886.

Bonawitz, Elizabeth, Denison, Stephanie, Gopnik, Alison, and Griffiths, Thomas L., 2014. 'Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference'. *Cognitive Psychology*, 74:35–65.

Bradley, Seamus and Steele, Katie, 2016. 'Can free evidence be bad? Value of information for the imprecise probabilist'. *Philosophy of Science*, 83(1):1–28.

Brenner, Lyle, Griffin, Dale, and Koehler, Derek J, 2005. 'Modeling patterns of probability calibration with random support theory: Diagnosing case-based judgment'. *Organizational Behavior and Human Decision Processes*, 97(1):64–81.

Briggs, R., 2009. 'Distorted Reflection'. *Philosophical Review*, 118(1):59–85.

Carr, Jennifer Rose, 2021. 'Why Ideal Epistemology?' *Mind*, To Appear.

Christensen, David, 2010. 'Rational Reflection'. *Philosophical Perspectives*, 24:121–140.

Das, Nilanjan, 2020. 'The Value of Biased Information'. *The British Journal for the Philosophy of Science*, To Appear.

Dawid, A P, 1982. 'The Well-Calibrated Bayesian'. *Journal of the American Statistical Association*, 77(379):605–610.

Dawid, A. P., 1983. 'Calibration-Based Empirical Inquiry'. *The Annals of Statistics*, 13(4):1251–1273.

Ditto, Peter H., Liu, Brittany S., Clark, Cory J., Wojcik, Sean P., Chen, Eric E., Grady, Rebecca H., Celniker, Jared B., and Zinger, Joanne F., 2019. 'At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives'. *Perspectives on Psychological Science*, 14(2):273–291.

Dorst, Kevin, 2019. 'Higher-Order Uncertainty'. In Mattias Skipper Rasmussen and Asbjørn Steglich-Petersen, eds., *Higher-Order Evidence: New Essays*, 35–61. Oxford University Press.

———, 2020. 'Evidence: A Guide for the Uncertain'. *Philosophy and Phenomenological Research*, 100(3):586–632.

———, 2021. 'Be Modest: You're Living on the Edge'. *Analysis*, 81(4):611—-621.

———, 2023a. 'Being Rational and Being Wrong'. *The Philosophers' Imprint*, To appear.

———, 2023b. 'Rational Polarization'. *The Philosophical Review*, To appear.

Dorst, Kevin, Levinstein, Benjamin, Salow, Bernhard, Husic, Brooke E., and Fitelson, Branden, 2021. 'Deference Done Better'. *Philosophical Perspectives*, 35(1):99–150.

Elga, Adam, 2013. 'The puzzle of the unmarked clock and the new rational reflection principle'. *Philosophical Studies*, 164(1):127–139.

Ellsberg, Daniel, 1961. 'Risk, Ambiguity, and the Savage Axioms'. *Quarterly Journal of Economics*, 75(4):643–669.

Erev, Ido, Wallsten, Thomas S, and Budescu, David V, 1994. 'Simultaneous over-and underconfidence: The role of error in judgment processes.' *Psychological review*, 101(3):519.

Fraser, Rachel, 2021. 'Mushy Akrasia'. *Philosophy and Phenomenological Research*, To Appear.

Gibbard, Allan and Harper, William L., 1978. 'Counterfactuals and Two Kinds of Expected Utility'. In *Foundations and Applications of Decision Theory*, volume 1, 125–162.

Glaser, Markus and Weber, Martin, 2010. 'Overconfidence'. *Behavioral finance: Investors, corporations, and markets*, 241–258.

Good, I J, 1967. 'On the Principle of Total Evidence'. *The British Journal for the Philosophy of Science*, 17(4):319–321.

Hájek, Alan and Rabinowicz, Wlodek, 2022. 'Degrees of commensurability and the repugnant conclusion'. *Noûs*, 56:897–919.

Halevy, Yoram, 2007. 'Ellsberg revisited: an experimental study b'. *Econometrica*, 75(2):503–536.

Hamblin, Charles L, 1976. 'Questions in montague english'. In *Montague grammar*, 247–259. Elsevier.

Hedden, Brian, 2023. 'Counterfactual Decision Theory'. *Mind*, fzac060.

Hintikka, Jaako, 1962. *Knowledge and Belief*. Cornell University Press.

Huttegger, Simon M, 2014. 'Learning experiences and the value of knowledge'. *Philosophical Studies*, 171(2):279–288.

Huttegger, Simon M., 2015. 'Merging of opinions and probability kinematics'. *Review of Symbolic Logic*, 8(4):611–648.

Huttegger, Simon M, 2017. *The probabilistic foundations of rational learning*. Cambridge University Press.

Icard, Thomas, 2016. 'Subjective Probability as Sampling Propensity'. *Review of Philosophy and Psychology*, 7(4):863–903.

Juslin, Peter, Nilsson, Håkan, and Winman, Anders, 2009. 'Probability Theory, Not the Very Guide of Life'. *Psychological Review*, 116(4):856–874.

Kadane, Joseph B., Schervish, Mark J., and Seidenfeld, Teddy, 1996. 'Reasoning to a foregone conclusion'. *Journal of the American Statistical Association*, 91(435):1228–1235.

Kahan, Dan M., Peters, Ellen, Dawson, Erica Cantrell, and Slovic, Paul, 2017. 'Motivated numeracy and enlightened self-government'. *Behavioural Public Policy*, 1:54–86.

Kamenica, Emir, 2019. 'Bayesian Persuasion and Information Design'. *Annual Review of Economics*, 11:249–272.

Kamenica, Emir and Gentzkow, Matthew, 2011. 'Bayesian persuasion'. *American Economic Review*, 101(6):2590–2615.

Kelly, Thomas, 2008. 'Disagreement, Dogmatism, and Belief Polarization'. *The Journal of Philosophy*, 105(10):611–633.

Koehler, Derek J, Brenner, Lyle, and Griffin, Dale, 2002. 'The calibration of expert judgment: Heuristics and biases beyond the laboratory'. *Heuristics and Biases: The Psychology of Intuitive Judgment*, 686–715.

Kovářík, Jaromír, Levin, Dan, and Wang, Tao, 2016. 'Ellsberg paradox: Ambiguity and complexity aversions compared'. *Journal of Risk and Uncertainty*, 52(1):47–64.

Kripke, Saul A, 1963. 'Semantical analysis of modal logic i normal modal propositional calculi'. *Mathematical Logic Quarterly*, 9(5-6):67–96.

Kunda, Ziva, 1990. 'The case for motivated reasoning'. *Psychological Bulletin*, 108(3):480–498.

Lewis, David, 1976. 'Probabilities of Conditionals and Conditional Probabilities'. *The Philosophical Review*, 85(3):297–315.

———, 1980. 'A subjectivist's guide to objective chance'. In Richard C Jeffrey, ed., *Studies in Inductive Logic and Probability*, volume 2, 263–293. University of California Press.

———, 1981. 'Causal decision theory'. *Australasian Journal of Philosophy*, 59(1):5–30.

Lichtenstein, Sarah, Fischhoff, Baruch, and Phillips, Lawrence D., 1982. 'Calibration of probabilities: The state of the art to 1980'. In Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment under Uncertainty*, 306–334. Cambridge University Press.

Little, Andrew T, 2022. 'Bayesian Explanations for Persuasion *'. (April):1–48.

Lord, Charles G., Ross, Lee, and Lepper, Mark R., 1979. 'Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence'. *Journal of Personality and Social Psychology*, 37(11):2098–2109.

Mercier, Hugo and Sperber, Dan, 2011. 'Why do humans reason ? Arguments for an argumentative theory'. 57–111.

Moore, Don A, Carter, Ashli B, and Yang, Heather H J, 2015. 'Wide of the Mark: Evidence on the Underlying Causes of Overprecision in Judgment'. *Organizational Behavior and Human Decision Processes*, 131:110–120.

Nickerson, Raymond S., 1998. 'Confirmation bias: A ubiquitous phenomenon in many guises.' *Review of General Psychology*, 2(2):175–220.

Nielsen, Michael and Stewart, Rush T, 2021. 'Persistent Disagreement and Polarization in a Bayesian Setting'. *British Journal for the Philosophy of Science*, 72(1):51–78.

Oddie, Graham, 1997. 'Conditionalization, Cogency, and Cognitive Value'. *The British Journal for the Philosophy of Science*, 48(4):533–541.

Ortoleva, Pietro and Snowberg, Erik, 2015. 'Overconfidence in political behavior'. *American Economic Review*, 105(2):504–535.

Peterson, CAMERON R. and Beach, LEE R., 1967. 'Man As an Intuitive Statistician'. *Psychological Bulletin*, 68(1):29–46.

Pettigrew, Richard and Titelbaum, Michael G, 2014. 'Deference Done Right'. *Philosopher's Imprint*, 14(35):1–19.

Ramsey, F. P., 1990. 'Weight or the value of knowledge'. *British Journal for the Philosophy of Science*, 41(1):1–4.

Salow, Bernhard, 2018. 'The Externalist's Guide to Fishing for Compliments'. *Mind*, 127(507):691–728.

Sanborn, Adam N. and Chater, Nick, 2016. 'Bayesian Brains without Probabilities'. *Trends in Cognitive Sciences*, 20(12):883–893.

Schervish, M. and Seidenfeld, T., 1990. 'An approach to consensus and certainty with increasing evidence'. *Journal of Statistical Planning and Inference*, 25(3):401–414.

Seidenfeld, Teddy, 1985. 'Calibration , Coherence , and Scoring Rules'. *Philosophy of Science*, 52:274–294.

Skyrms, Brian, 1980. 'Higher Order Degrees of Belief'. In D H Mellor, ed., *Prospects for Pragmatism*, 109–137. Cambridge University Press.

———, 1990. 'The Value of Knowledge'. *Minnesota Studies in the Philosophy of Science*, 14:245–266.

———, 2006. 'Diachronic coherence and radical probabilism'. *Philosophy of Science*, 73(5):959–968.

Stalnaker, Robert, 2006. 'On the Logics of Knowledge and Belief'. *Philosophical Studies*, 128(1):169–199.

———, 2019. 'Rational Reflection, and the Notorious Unmarked Clock'. In *Knowledge and Conditionals: Essays on the Structure of Inquiry*, 99–112. Oxford University Press.

Tetlock, Philip E, 2009. *Expert political judgment*. Princeton University Press.

Thurstone, L.L., 1927. 'A law of comparative judgement'.

Titelbaum, Michael G., 2010. 'Tell me you love me: Bootstrapping, externalism, and no-lose epistemology'. *Philosophical Studies*, 149(1):119–134.

Trautmann, Stefan T and van de Kuilen, Gijs, 2015. 'Ambiguity Attitudes'. In *The Wiley Blackwell handbook of judgment and decision making*.

van Fraassen, Bas, 1984. 'Belief and the Will'. *The Journal of Philosophy*, 81(5):235–256.

Weisberg, Jonathan, 2007. 'Conditionalization, reflection, and self-knowledge'. *Philosophical Studies*, 135(2):179–197.

White, Roger, 2006. 'Problems for Dogmatism'. *Philosophical Studies*, 131:525–557.

———, 2009. 'Evidential Symmetry and Mushy Credence'. In Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology*, volume 3, 161–186. Oxford University Press.

Williamson, Timothy, 2000. *Knowledge and its Limits*. Oxford University Press.

———, 2008. 'Why Epistemology Cannot be Operationalized'. In Quentin Smith, ed., *Epistemology: New Essays*, 277–300. Oxford University Press.

Zaffora Blando, Francesca, 2022. 'Bayesian Merging of Opinions and Algorithmic Randomness'. *British Journal for the Philosophy of Science*, To appear.