

Higher-order Probabilism

May 15, 2024

Abstract. Rational agents are often uncertain about the truth of many propositions. To represent this uncertainty, it is natural to rely on probability theory. Two options are typically on the table, precise and imprecise probabilism, but both fall short in some respect. Precise probabilism is not expressive enough, while imprecise probabilism suffers from belief inertia and the impossibility of proper scoring rules. We put forward a novel version of probabilism, higher-order probabilism, and we argue that it outperforms existing alternatives.

Keywords: Probabilism; Imprecise probabilities; Evidence; Probability; Belief inertia; Bayesian networks; Proper scores.

1 Introduction

As rational agents, we are uncertain about the truth of many propositions since our evidence about them is often fallible. To represent this uncertainty, it is natural to rely on probabilities. Two options are typically on the table: precise and imprecise probabilism. Precise probabilism models an agent’s state of uncertainty (or credal state) with a single probability measure: each proposition is assigned one probability value between 0 and 1 (a sharp credence). The problem is that a single probability measure is not expressive enough to distinguish between intuitively different states of uncertainty (§2). To avoid this problem, a *set* of probability measures can be used to represent the uncertainty of a rational agent. This approach is known as imprecise probabilism. It outperforms precise probabilism in some respects, but also runs into problems of its own, such as belief inertia and the impossibility of defining proper scoring rules (§3).

To make progress, this paper argues that a rational agent’s uncertainty should be represented neither by a single probability measure nor a set of measures, but rather by a higher-order probability measure, specifically, a probability distribution over parameter values interpreted as probabilities. The theory we propose is not mathematically novel, but addresses many of the problems that plague precise and imprecise probabilism (§?? and §??). It also fares better than existing versions of probabilism when the probability of multiple propositions, dependent or independent, is to be assessed (§?? and ??). Many of the examples in this paper are about coin tosses, but in the final two sections, we will also discuss legal examples and illustrate the broader applicability of higher-order probabilism.

2 Precise probabilism

Precise probabilism holds that a rational agent's uncertainty about a proposition is to be represented as a single, precise probability measure. Bayesian updating regulates how the prior probability measure should change in light of new evidence that the agent learns. This is an elegant and simple theory, but representing our uncertainty about a proposition with a single, precise probability measure runs into several difficulties.

Precise probabilism fails to capture an important dimension of how our fallible beliefs reflect the evidence we have (or have not) obtained. A couple of stylized examples featuring coin tosses should make the point clear.

No evidence v. fair coin You are about to toss a coin but have no evidence about its bias. You are completely ignorant. Compare this to the situation in which you know, based on overwhelming evidence, that the coin is fair.

On precise probabilism, both scenarios are represented by assigning a probability of .5 to the proposition *on the next toss, the coin comes up heads* (or *head* for short). If you are completely ignorant, the principle of insufficient evidence suggests that you assign .5. Similarly, if you know for sure the coin is fair, assigning again .5 seems the best way to quantify the uncertainty about the outcome. The agent's evidence in the two scenarios is quite different, but precise probabilities fail to capture this difference.

Learning from ignorance You toss a coin with unknown bias. You toss it 10 times and observe 5 heads. You toss it further and observe 50 heads in 100 tosses.

Since the coin initially had an unknown bias, you should presumably assign a probability of .5 to the proposition *heads*, if you stick with precise probabilism. After the 10 tosses, you again assess the probability to be .5. You must have learned something, but whatever that is, it is not modeled by precise probabilities. When you toss the coin 100 times and observe 50 heads, you learn something new as well. But your precise probability assessment will again be .5.

Precise probabilism assigns the same probability in situations in which one's evidence is quite different: when no evidence is available about a coin's bias; when there is little evidence the coin is fair (say, after 10 tosses); when there is strong evidence the coin is fair (say, after 100 tosses). Analogous problems arise for evidence that the coin is not fair. If a rational agent starts with a weak belief that the coin is .6 biased towards heads, they can strengthen that belief by tossing the coin repeatedly and observing, say, 60 heads in 100 tosses. But this improvement in their evidence is not mirrored in the .6 probability they should assign to *heads*.¹

The precise probabilist might respond that we should extend the algebra and include propositions such as *the coin has a .5 bias*, or in symbols $\theta_{.5}$. As evidence accumulates about the fairness of the coin, the (precise) probability of $\theta_{.5}$ should go up accordingly, even though the

¹Another problem for precise probabilism is known as *sweetening* (Hare, 2010). Imagine a rational agent who does not know the bias of the coin. For precise probabilism, this state of uncertainty is represented by a .5 probability assignment to heads. Next, the agent learns that the bias towards heads, whatever the bias is, has been slightly increased, say by .001. Intuitively, the new information should leave the agent equally undecided about betting on heads or tails. After sweetening, the agent still does not know much about the actual bias of the coin. But, according to precise probabilism, sweetening should make the agent bet on heads: if the probability of heads was initially .5, it should now be slightly above .5.

probability of *heads* should remain .5. How would this line of response generalize beyond cases of coin tosses? Suppose that, given a certain stock of evidence, *A* is more likely than *B*. Further, suppose that the acquisition of new evidence does not change the probabilities. Admittedly, something has changed in the agent's state of uncertainty: the quantity of evidence on which the agent can assess whether *A* is more probable than *B* has become larger. And yet, this change is not reflected in the precise probabilities assigned to *A* and *B*.² Although no proposition about the coin's bias can be added to the algebra here, the precise probabilist might recommend we extend the algebra and include the proposition *the probability that A is more likely than B is such and such*. In rough outline, we agree, if only because the added proposition contains a probability, and assigning a probability to it amounts to using higher-order probabilities. Before going higher-order, however, we should explore another view in the literature.

3 Imprecise probabilism

Imprecise probabilism holds that a rational agent's credal stance towards a hypothesis is to be represented by a set of probability measures, typically called a representor \mathbb{P} , rather than a single measure *P*. The representor should include all and only those probability measures which are compatible with the evidence (more on this point later).³ It is easy to see that modeling an agent's credal state by sets of probability measures avoids some of the shortcomings of precise probabilism. For instance, if an agent knows that the coin is fair, their credal state would be represented by the singleton set $\{P\}$, where *P* is a probability measure that assigns .5 to *heads*. If, on the other hand, the agent knows nothing about the coin's bias, their credal state would be represented by the set of all probabilistic measures, since none of them is excluded by the available evidence. Note that the set of probability measures does not represent admissible options that the agent could legitimately pick from. Rather, the agent's credal state is essentially imprecise and should be represented by means of the entire set of probability measures.

So far so good. But, just as precise probabilism fails to be appropriately evidence-responsive in certain scenarios, imprecise probabilism runs into similar difficulties in other scenarios.

Even v. uneven bias: You have two coins and you know, for sure, that the probability of getting heads is .4, if you toss one coin, and .6, if you toss the other coin. But you do not know which is which. You pick one of the two at random and toss it. Contrast this with an uneven case. You have four coins and you know that three of them have bias .4 and one of them has bias .6. You pick a coin at random and plan to toss it. You should be three times more confident that the probability

²The distinction here is sometimes formulated in terms of the *balance* of the evidence (that is, whether the evidence available tips in favor a hypothesis or another) as opposed to its *weight* (that is, the overall quantity of evidence regardless of its balance); see Keynes (1921) and Joyce (2005).

³For the development of imprecise probabilism, see Keynes (1921); Levi (1974); Gärdenfors & Sahlin (1982); Kaplan (1968); Joyce (2005); Fraassen (2006); Sturgeon (2008); Walley (1991). Bradley (2019) is a good source of further references. Imprecise probabilism is closely related to what we might call interval probabilism (Kyburg, 1961; Kyburg Jr & Teng, 2001). In interval probabilism, precise probabilities are replaced by intervals of probabilities. On imprecise probabilism, instead, precise probabilities are replaced by sets of probabilities. This makes imprecise probabilism more general since the representor set need not be an interval.

of getting heads is .4. rather than .6.

The first situation can be easily represented by imprecise probabilism. The representor would contain two probability measures, one that assigns .4. and the other that assigns .6 to the hypothesis ‘this coin lands heads’. However imprecise probabilism cannot represent the second situation. Since the probability measures in the set are all compatible with the agent’s evidence, no probability measure can be assigned a greater (higher-order) probability than any other.⁴

These examples show that imprecise probabilism is not expressive enough to model the scenario of uneven bias. Defenders of imprecise probabilism might concede this point but prefer their account for reasons of simplicity. They could also point out that imprecise probabilism models scenarios that precise probabilism cannot model, for example, a state of complete lack of evidence. In this respect, imprecise probabilism outperforms precise probabilism in expressive power, and also retains theoretical simplicity. But this is not quite true as imprecise probabilism suffers from several shortcomings that do not affect precise probabilism.

The first shortcoming we discuss has not received extensive discussion in the literature. Recall that, for imprecise probabilism, an agent’s state of uncertainty is represented by those probability measures that are *compatible* with the agent’s evidence. How should the notion of compatibility be understood here? Perhaps we can think of compatibility as the fact that the agent’s evidence is consistent with the probability measure in question. But mere consistency wouldn’t get the agent very far in excluding probability measures, as too many probability measures are consistent with most observations and data. Admittedly, there will be clear-cut cases: if you see the outcome of a coin toss to be heads, you reject the measure with $P(H) = 0$, and similarly for tails. Another class of cases might arise while randomly drawing objects from a finite set where the objective chances are known. But clear-cut cases aside, what else? Data will often be consistent with almost any probability measure.⁵

A second, related problem for imprecise probabilism is known as belief inertia. Precise probabilism offers an elegant model of learning from evidence: Bayesian updating. Imprecise probabilism, at least *prima facie*, offers an equally elegant model of learning from evidence, richer and more nuanced. It is a natural extension of the classical Bayesian approach that uses precise probabilities. When faced with new evidence E between time t_0 and t_1 , the representor set should be updated point-wise, running the standard Bayesian updating on each probability

⁴Other scenarios can be constructed in which imprecise probabilism fails to capture distinctive intuitions about evidence and uncertainty; see, for example, (Rinard, 2013). Suppose you know of two urns, GREEN and MYSTERY. You are certain GREEN contains only green marbles, but have no information about MYSTERY. A marble will be drawn at random from each. You should be certain that the marble drawn from GREEN will be green (G), and you should be more confident about this than about the proposition that the marble from MYSTERY will be green (M). For each $r \in [0, 1]$ your representor will contain a P with $P(M) = r$. But then, it also contains one with $P(M) = 1$. This means that it is not the case that for any probability measure P in the representor set, $P(G) > P(M)$, that is, it is not the case that a rational agent should be more confident of G than of M . This is counter-intuitive.

⁵Probability measures can be inconsistent with evidential constraints that agents believe to be true. Mathematically, non-trivial evidential constraints are easy to model (Bradley, 2012). They can take the form, for example, of the *evidence of chances* $\{P(X) = x\}$ or $P(X) \in [x, y]$, or *structural constraints* such as “ X and Y are independent” or “ X is more likely than Y .” These constraints are something that an agent can come to accept outright, but only if offered such information by an expert whom the agent completely defers. Besides these idealized cases, it is unclear how an agent could come to accept such structural constraints upon observation. There will usually be some degree of uncertainty about the acceptability of these constraints.

measure in the representor:

$$\mathbb{P}_{t_1} = \{P_{t_1} | \exists P_{t_0} \in \mathbb{P}_{t_0} \forall H [P_{t_1}(H) = P_{t_0}(H|E)]\}.$$

The hope is that, if we start with a range of probabilities that is not extremely wide, point-wise learning will behave appropriately. For instance, if we start with a prior probability of *heads* equal to .4 or .6, then those measures should be updated to something closer to .5 once we learn that a given coin has already been tossed ten times with the observed number of heads equal 5 (call this evidence *E*). This would mean that if the initial range of values was [.4, .6] the posterior range of values should be narrower.

Unfortunately, this narrowing of the range of values becomes impossible whenever the starting point is a complete lack of knowledge, as imprecise probabilism runs into the problem of belief inertia (Levi, 1980). This problem arises in situations in which no amount of evidence could lead the agent to change their belief state, according to a given modeling strategy. Consider a situation in which you start tossing a coin knowing nothing about its bias. The range of possibilities is [0, 1]. After a few tosses, if you observed at least one tail and one head, you can exclude the measures assigning 0 or 1 to *heads*. But what else have you learned? If you are to update your representor set point-wise, you will end up with the same representor set. For any sequence of outcomes that you can obtain and any probability value in [0, 1], there will exist a probability measure (conditional on the outcomes) that assigns that probability to *heads*. Consequently, the edges of your resulting interval will remain the same. In the end, it is not clear how you are supposed to learn anything if you start from complete ignorance.⁶

Some downplay the problem of belief inertia. After all, if you start with knowing truly nothing, then it is right to conclude that you will never learn anything. Joyce (2010) writes:

You cannot learn anything in cases of pronounced ignorance simply because a prerequisite for learning is to have prior views about how potential data should alter your beliefs. (p. 291) (Joyce, 2010)

The upshot is that uniform priors should not be used and that imprecise probabilism gives the right results when the priors are non-vacuous. Moss (2020) arrives at the same conclusion by drawing the following parallelism. If contingent propositions should not be assigned probabilities of 0 or 1 whenever these extreme values are unrevisable, then by the same token the uniform interval [0, 1] should not be used for imprecise probabilities whenever it is impervious to revision.⁷ However, as we will see in the next section, uniform priors are not necessarily

⁶Here is another example of belief inertia by Rinard (2013). Suppose all the marbles in the urn are green (H_1), or exactly one-tenth of the marbles are green (H_2). Your initial credence about these hypotheses is completely uncertain, the interval [0, 1]. Next, you learn that a marble drawn at random from the urn is green (*E*). After conditioning on this evidence, you end up with the same spread of values for H_1 that you had before learning *E*. This result holds no matter how many green marbles are drawn. This is counterintuitive: if you keep drawing green marbles, the hypothesis that all marbles are green should become more probable.

⁷Another strategy is to say that, in a state of complete ignorance, a special updating rule should be deployed. Elkin (2017) suggests the rule of *credal set replacement* that recommends that upon receiving evidence the agent should drop measures rendered implausible, and add all non-extreme plausible probability measures. This, however, is tricky. One needs a separate account of what makes a distribution plausible from a principled account of why one should use a separate special update rule when starting with complete ignorance.

unrevisable and can be a starting point for learning. This suggests that the problem lies with imprecise probabilism, not with uniform priors as such.

Finally, even setting aside belief inertia, imprecise probabilism faces a third major problem that does not arise for precise probabilism. Workable scoring rules exist for measuring the accuracy of a precise probability measure, but it is hard to define workable scoring rules for imprecise probabilities. In the precise case, scoring rules measure the distance between a rational agent's probability measure and the actual value. The Brier score is the most common scoring rule for precise probabilities.⁸ A requirement of scoring rules is *propriety*: any rational agent will expect their probability measure to be more accurate than any other. After all, if an agent thought a different probability measure was more accurate, they should switch to it. More specifically, let $I(p, w)$ be an inaccuracy score of a probability distribution p relative to the true state w . The score $I(p, w)$ is strictly proper if, for any other probability distribution q different from p , the following holds:

$$\sum_{w \in W} p(w)I(p, w) < \sum_w q(w)I(p, w).$$

That is, the expected inaccuracy of p from the perspective of p should always be smaller than the expected inaccuracy of p from the perspective of another distribution q . To calculate the expected accuracy of p , first you need to calculate its inaccuracy $I(p, w)$ at every possible true state w , and then factor in the probabilities $p(w)$ of the true states according to p . Well-known results demonstrate the strict propriety of the Brier score for precise probabilities.⁹

Can similar results be established for imprecise probabilities? The answer is likely to be negative. Several hurdles exist. To start, a plausible scoring rule for imprecise probabilities is not easy to define. Suppose a forecaster assigns the $[.8, .9]$ probability interval to the outcome that it would rain tomorrow, where the true state is 'rain'. Would the wider $[.6, .99]$ interval be more accurate since its .99 upper bound is closer to the true state? Intuitively, the wider interval should not be more accurate. If it were, the trivial interval $[0, 1]$ would always be more accurate than any other interval since either of its edges are closer to whatever the true state turns out to be. To remedy this problem, an inaccuracy score for imprecise probabilities should depend both on the closeness to the true state and the size of the interval. If, for example, the inaccuracy score were to increase as the size of the interval increases, this would block the result that the $[0, 1]$ interval is always the least inaccurate. Still, even if a well-behaved score for imprecise probabilities can be found, a further problem arises in defining its expected inaccuracy. Let $I([p_-, p_+], w)$ be an inaccuracy score for the interval $[p_-, p_+]$. What is its expected inaccuracy from the perspective of, say, the interval $[p_-, p_+]$ itself? There is no straightforward answer to this question because $I([p_-, p_+], w)$ cannot be multiplied by $[p_-, p_+]$, in the way

⁸The Brier score is defined as the squared distance between the true state and the probability forecast, or formally, $(p(x) - V(x, w))^2$, where $p(x)$ is the probability forecast and $V(x, w)$ determines if a proposition obtains at w ($V(x, w) = 1$) or not ($V(x, w) = 0$). If, for example, the proposition 'rain' obtains at w , the forecast 'rain with .9 probability' would be more accurate at w than the forecast 'rain with .8 probability'. If, on the other hand, the proposition 'not rain' obtained at w , the latter forecast would be more accurate.

⁹Besides propriety, other common requirements (which the Brier score also satisfies) are: the score $I(p, w)$ should be a function of the probability distribution p and the true state w (extensionality); and the score should be a continuous function around p (continuity).

in which $I(p, w)$ can be multiplied by $p(w)$. Perhaps the expected inaccuracy of $[p_-, p_+]$ can be evaluated from the perspective of the precise probabilities at its edges, either p_- or p_+ .¹⁰ The problem is that, if the expected inaccuracy of $[p_-, p_+]$ is evaluated from the perspective of the precise probabilities at the edges, finding a proper inaccuracy score that is also continuous turns out to be mathematically impossible (Seidenfeld, Schervish, & Kadane, 2012).¹¹

Bradley, S. (2012). *Scientific uncertainty and decision making* (PhD thesis). London School of Economics; Political Science (University of London).

Bradley, S. (2019). Imprecise Probabilities. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019). <https://plato.stanford.edu/archives/spr2019/entries/imprecise-probabilities/>; Metaphysics Research Lab, Stanford University.

Campbell-Moore, C. (2020). *Accuracy and imprecise probabilities*.

Elkin, L. (2017). *Imprecise probability in epistemology* (PhD thesis). Ludwig-Maximilians-Universität; Ludwig-Maximilians-Universität München.

Fraassen, B. C. V. (2006). Vague expectation value loss. *Philosophical Studies*, 127(3), 483–491. <https://doi.org/10.1007/s11098-004-7821-2>

Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese*, 53(3), 361–386. <https://doi.org/10.1007/bf00486156>

Hare, C. (2010). Take the sugar. *Analysis*, 70(2), 237–247.

Joyce, J. (2005). How probabilities reflect evidence. *Philosophical Perspectives*, 19(1), 153–178.

Joyce, J. (2010). A defense of imprecise credences in inference and decision Making1. *Philosophical Perspectives*, 24(1), 281–323. <https://doi.org/10.1111/j.1520-8583.2010.00194.x>

Kaplan, J. (1968). Decision theory and the fact-finding process. *Stanford Law Review*, 20(6), 1065–1092.

Keynes, J. M. (1921). *A treatise on probability, 1921*. London: Macmillan.

Kyburg, H. E. (1961). *Probability and the logic of rational belief*. Wesleyan University Press.

Kyburg Jr, H. E., & Teng, C. M. (2001). *Uncertain inference*. Cambridge University Press.

Levi, I. (1974). On indeterminate probabilities. *The Journal of Philosophy*, 71(13), 391. <https://doi.org/10.2307/2025161>

Levi, I. (1980). *The enterprise of knowledge: An essay on knowledge, credal probability, and chance*. MIT Press.

Mayo-Wilson, C., & Wheeler, G. (2016). Scoring imprecise credences: A mildly immodest proposal. *Philosophy and Phenomenological Research*, 92(1), 55–78. <https://doi.org/10.1111/phpr.12256>

Moss, S. (2020). Global constraints on imprecise credences: Solving reflection violations,

¹⁰So the expected inaccuracy would equal $\sum_{w \in W} p_-(w)I([p_-, p_+], w)$ or $\sum_{w \in W} p_+(w)I([p_-, p_+], w)$.

¹¹Proper scoring rules are often used to formulate accuracy-based arguments for precise probabilism. These arguments show (roughly) that, if your precise measure follows the axioms of probability theory, no other non-probabilistic measure is going to be more accurate than yours whatever the facts are. So, without proper scoring rules for imprecise probabilities, the prospects for an accuracy-based argument for imprecise probabilism look dim (Campbell-Moore, 2020; Mayo-Wilson & Wheeler, 2016). Moreover, as shown by Schoenfield (2017), if an accuracy measure satisfies certain plausible formal constraints, it will never strictly recommend an imprecise stance, as for any imprecise stance there will be a precise one with at least the same accuracy.

- belief inertia, and other puzzles. *Philosophy and Phenomenological Research*, 103(3), 620–638. <https://doi.org/10.1111/phpr.12703>
- Rinard, S. (2013). Against radical credal imprecision. *Thought: A Journal of Philosophy*, 2(1), 157–165. <https://doi.org/10.1002/tht3.84>
- Schoenfield, M. (2017). The accuracy and rationality of imprecise credences. *Noûs*, 51(4), 667–685. <https://doi.org/10.1111/nous.12105>
- Seidenfeld, T., Schervish, M., & Kadane, J. (2012). Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning*, 53, 1248–1261. <https://doi.org/10.1016/j.ijar.2012.06.018>
- Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, 42(1), 139–165. Retrieved from <http://www.jstor.org/stable/25177157>
- Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman; Hall London.