How to be an imprecise impermissivist

Seamus Bradley[*]

May 27, 2022

Rational credence should be coherent in the sense that your attitudes should not leave you open to a sure loss. Rational credence should be such that you can learn when confronted with relevant evidence. Rational credence should not be sensitive to irrelevant differences in the presentation of the epistemic situation. We explore the extent to which orthodox probabilistic approaches to rational credence can satisfy these three desiderata and find them wanting. We demonstrate that an imprecise probability approach does better. Along the way we shall demonstrate that the problem of "belief inertia" is not an issue for a large class of IP credences, and provide a solution to van Fraassen's box factory puzzle.

1. Three desiderata for rational credence

We believe some things more strongly than others, and those strengths of belief are relevant to our behaviour, for example our betting behaviour. Call these strengths of belief "credences". We want our credences to be rational, to have some good-making features. But what properties specifically do we actually want our rational credences to satisfy? Here are three candidates.

Coherence Rational credences should be coherent

Learning Rational credences should allow learning on the basis of evidence

Invariance Rational credences should be insensitive to irrelevant differences in the presentation of the epistemic situation.

Each of these candidate norms needs explanation, and that's what the following subsections will provide.[1] My main argument in this paper is that probability theory does an OK job of satisfying these three criteria, but that some imprecise probabilities (IP) do a

---

[*]www.seamusbradley.net

[1]I will use "Coherence", "Learning" and "Invariance" as generic labels for the three kinds of desiderata. As we'll see, each of these will cover several distinct properties.

better job. The remainder of this section will outline the criteria in some detail and show probability theory fares. Section 2 will introduce imprecise probabilities and one major alleged problem for them: the problem of belief inertia. Section 3 will demonstrate that this problem needn't trouble the fan of IP.

1.1. Coherence

You want your credences to have a certain amount of coherence or consistency, you want them to hang together in the right way. One way of capturing this idea is to consider betting behaviour. Let's say that you measure your strength of beliefs using real numbers, and the strength of belief you attach to event A is $\mathbf{pr}(A)$. Let's imagine that you use your credences to inform the bets you make in the following way:

> **Two-sided betting**
>
> - A bet that wins $1 - x$ if A is true, and loses x if A is false is acceptable if $\mathbf{pr}(A) > x$.
>
> - If the above bet is not acceptable then a bet that loses $1 - y$ if A and wins y otherwise is acceptable for any $y > x$.[2]

Call a bet of this form – "win $1 - x$ if A, lose x otherwise" – a "unit bet on A (with betting quotient x)".[3] If your credences inform your betting behaviour in this way, then it seems a minimal requirement of rationality is that your betting behaviour should be such that you are not subject to a sure loss. It is a bad thing to be such that, whatever happens, you lose money. As de Finetti (1964)[4] showed, this coherence condition – call it de Finetti-coherence – is enough to require that your credences satisfy the axioms of probability theory.

> **de Finetti coherence**
>
> Your prices for gambles in a two-sided betting scenario should be such that no combination of acceptable bets is guaranteed to yield a loss in every state.

Call Ω the state space: the space of all possibilities, or the set of possible worlds. A probability function $\mathbf{pr}$ is a real valued function over a Boolean algebra of propositions over Ω, $\mathcal{B}(\Omega)$ such that for all $X, Y \in \mathcal{B}(\Omega)$:[5]

---

[2] Note that these conditions together are silent about which bet is acceptable if $\mathbf{pr}(A) = x$. Note also that here and throughout we will be discussing probability functions rather than the more general framework of previsions or desirable gambles. For our purposes, this additional generality will not be necessary.

[3] The bet in the second clause here is a unit bet on $\neg A$ with betting quotient $1 - y$.

[4] The reference here is to the English translation of a paper that appeared in French in 1937. See also de Finetti (1974). Frank Ramsey has a claim to have come to almost the same conclusions somewhat earlier (Ramsey 1926) but we attribute this idea to de Finetti since the subsequent literature we shall rely on (e.g. Troffaes and de Cooman 2014; Walley 1991) is squarely in the de Finetti tradition.

[5] Note that these properties are not all logically independent: for example, monotonicity is entailed by the conjunction of the other properties.

Boundedness $0 = \mathbf{pr}(\emptyset) \leq \mathbf{pr}(X) \leq \mathbf{pr}(\Omega) = 1$

Monotonicity If $X \subseteq Y$ then $\mathbf{pr}(X) \leq \mathbf{pr}(Y)$

Superadditivity If X, Y are mutually exclusive then $\mathbf{pr}(X \vee Y) \geq \mathbf{pr}(X) + \mathbf{pr}(Y)$

Subadditivity If X, Y are mutually exclusive then $\mathbf{pr}(X \vee Y) \leq \mathbf{pr}(X) + \mathbf{pr}(Y)$

There are reasons to be sceptical that your betting dispositions are so tightly connected to your credences, but it does still seem true that being subject to a sure loss "dramatizes an inconsistency in the attitudes of an agent", as Hájek (2008) put it (see also Christensen 2004). Pettigrew (2020) provides the philosophical foundations for such betting arguments, and Troffaes and de Cooman (2014) offer a detailed and general version of the mathematical theory. I won't here be taking issue with the use of betting to ground norms of rational credence, although we will discuss alternative forms of such arguments that rely on weaker forms of coherence.

An extreme form of "permissivism" would stop here: the only form of rational constraint on your credences is that they obey the probability axioms. This is a fairly extreme form of permissivism, and most people would grant that one should also satisfy some further desideratum of the kind we shall meet in the next section.

1.2. Learning

Coherence should not just constrain attitudes at a time, but also attitudes over time. Consider your attitudes before updating on some proposition E, and after having done the update. Call these $\mathbf{pr}$ and $\mathbf{pr}'$ respectively. It's not enough for $\mathbf{pr}$ and $\mathbf{pr}'$ to both satisfy the axioms of probability individually: they should also be such that you can't incur a sure loss in moving from one to the other on learning E. In short, the two credences should be *jointly coherent* and not merely separately coherent. In practice, what this means is that $\mathbf{pr}'(X) = \mathbf{pr}(X|E) = \frac{\mathbf{pr}(XE)}{\mathbf{pr}(E)}$.[6] Now the question is: does updating in this way mean that we acquire better credences as we accrue evidence?[7] The answer turns out to be a qualified yes.

First let's note that not all probabilistic credences give rise to plausible updating behaviour. For example, consider the following "anti-inductive priors". Consider learning the outcome of several tosses of a mystery coin, a coin with an unknown bias. It is consistent with basic probability theory that credences over possible outcomes of tosses of the coin are such that, on observing the coin land heads, they become near-certain that the next toss will land *tails*. This seems to go against the intuition that, under normal circumstances, observing a coin's landing heads should make you more confident that the next toss will land heads. For example, using $H_1 = 1, H_2 = 1$ to mean the events that the first and second tosses landed heads respectively, and likewise that $H_i = 0$ means the ith toss lands tails, let $\mathbf{pr}(H_1 = 1 \wedge H_2 = 1) = \mathbf{pr}(H_1 = 0 \wedge H_2 = 0) = \frac{\varepsilon}{2}$

---

[6]For the purposes of this essay, I am interested in learning E with certainty only, so I will not be discussing Jeffrey conditioning or any other form of update.

[7]"Better" in what sense? We'll discuss this briefly at the end of this section.

and $\mathbf{pr}(H_1 = 0 \wedge H_2 = 1) = \mathbf{pr}(H_1 = 1 \wedge H_2 = 0) = \frac{1-\varepsilon}{2}$. Then $\mathbf{pr}(H_1 = 1) = \mathbf{pr}(H_2 = 1) = \frac{1}{2}$, but $\mathbf{pr}(H_2 = 1 | H_1 = 1) = \varepsilon$. That is, your credence that the second coin lands heads, conditional on having observed a head is arbitrarily low, even though your prior for heads on the second toss was a half. This seems a strange way to arrange your epsitemic life.[8] Some classes of probabilistic prior credence do intuitively allow you to learn in this simple environment, and it seems that all but the most ardent permissivists would grant that this further constrains the admissible credences.

Consider a credence function that is determined by two parameters, $\gamma \in (0,1)$ and $\lambda > 0$ where the credence that the next coin will land heads is $\mathbf{pr}(H) = \gamma$, and learning that there were h heads and t tails in the last $n = h + t$ tosses of the coin (call this evidence $S_n = h$) gives you an updated credence of $\mathbf{pr}(H | S_n = h) = \frac{\gamma \lambda + S_n}{\lambda + n}$.[9] Such a credence is probabilistically coherent. We shall call such probability functions "learning priors". For more on how such learning priors are constructed, see Appendix A. Note that Carnap's continuum of inductive methods (for one predicate) are the special case when $\gamma = \frac{1}{2}$.

Each $\mathbf{pr}$ is identified by two parameters γ and λ that capture a prior attitude to the bias of the coin or the chance of heads, and also how that prior responds to evidence. We can see this by rearranging the above expression to obtain:

$$\mathbf{pr}(H | S_n = h) = \left(\frac{n}{n + \lambda} \right) \frac{S_n}{n} + \left(\frac{\lambda}{n + \lambda} \right) \gamma \tag{1}$$

γ determines the prior attitude to the next toss' landing heads, and λ governs how much your attitude changes in response to evidence about previous tosses. So if we keep λ fixed and vary γ, we have a range of priors that start at different points, but learn at roughly the same rate (Figure 1). On the other hand, if we fix γ and vary λ, we have a range of priors that start with the same prior credence for H, but move different amounts on the basis of the same evidence (Figure 2). Note also that as these priors acquire evidence, they become less prone to change on the basis of further evidence.

Given enough evidence, any such $\mathbf{pr}$ will converge on the true value of the chance with increasing evidence. By the law of large numbers, as we increase the number of coin tosses (i.e. as n increases), the ratio $\frac{S_n}{n}$ will tend to the true chance, and as n gets larger, the first term of the right hand side of Equation (1) gets closer to $\frac{S_n}{n}$, while the second term shrinks to zero. So each learning prior will converge on the true chance of heads as evidence increases. This seems like a good result.

For the purposes of comparing this situation to others, let's clarify exactly what nice property such probability functions have. We want to know whether conditionalising on evidence makes your credences better. For our purposes, we can just take betterness to be understood as closeness to some "perfect" probability function $\mathbf{pr}^*$. We might interpret this, with Pettigrew (2016), as the credences of the perfectly informed agent, or as the chance function: "closeness" in what sense? Again, this won't be central to

---

[8]Such priors can be constructed for longer sequences of coin flips, and for more complex learning set-ups.

[9]Here and throughout H is a shorthand for $H_{n+1} = 1$: that is, the next toss of the coin will land heads, and $S_n = \sum^n H_i$.
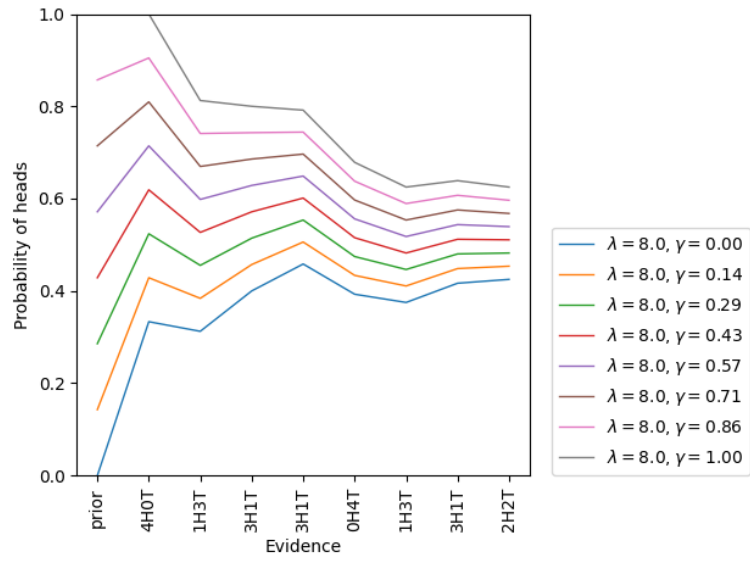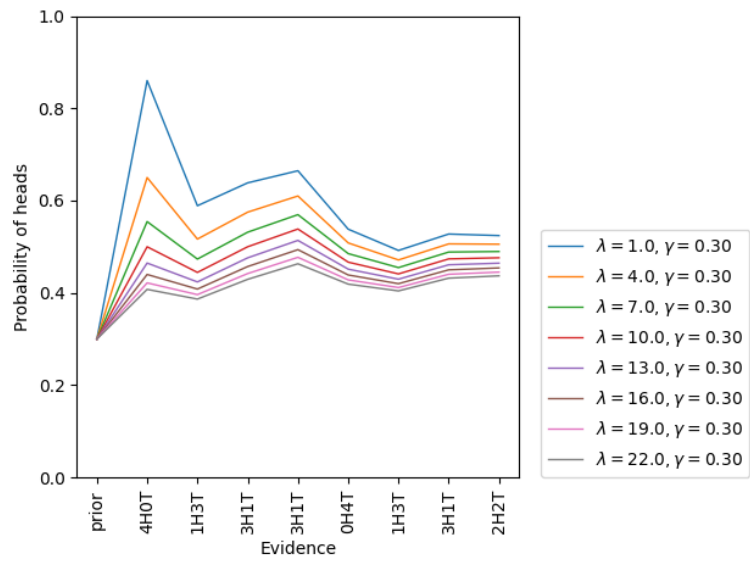
Figure 1: Learning priors for fixed λ

Figure 2: Learning priors for fixed γ

our discussion. There are many concepts of distance between probability functions and I don't think anything I say depends strongly on the choices one makes here.

So $\mathbf{pr}^*$ is whatever probability function is such that it is rationally optimal to have your credences coincide with it. And let E_n be the total evidence up to time n.

Almost sure correctness in the limit (ASCL)

With probability 1, $\lim_{n \to \infty} |\mathbf{pr}(H|E_n) - \mathbf{pr}^*(H)| = 0$

Note that this correctness is relative to a particular kind of proposition to be learned: in this case, the event H. But ASCL also depends on facts about the evidence learned E_n and in particular, about the probability of acquiring certain kinds of evidence. For example, we above made use of the fact that as n gets bigger, $\frac{S_n}{n}$ gets closer to $\mathbf{pr}^*(H)$ which we are here interpreting as the chance of heads. That's not a fact about your credences, but about what you take for granted in the chance set up. So in asserting that ASCL holds we are asserting something about the kind of evidence you will learn, as well as about your credences. If we do not take ourselves to be entitled to making these sorts of assumptions about the chance of evidence, or the structure of the epistemic situation, then it's unlikely that any sensible learning behaviour can be derived.

So, all probability functions are coherent (in de Finetti's sense), and some probability functions are such that they also satisfy our second desiderata: they can learn on the basis of evidence, at least some of the time.[10] So far so good.

1.3. Invariance

We are now moving on to introducing the third of our three plausible desiderata for rational credence. One question we might ask at this stage is "can we provide some normative reason why adopting learning priors is rationally required?". One response is to say "Well, look, if you want to learn, here's how!", but this isn't really the kind of response we're looking for.

It would be desirable to have an argument with plausible premises whose conclusion is that learning priors are the only rational ones. Arguments of this form can be given, where the "plausible premises" take the form of symmetry requirements. The idea is that it is a rational requirement that your credences treat certain propositions the same way (treat them symmetrically) and if you are required to satisfy enough such symmetries then learning priors are the only credence functions that satisfy them.

Symmetries are typically captured by requiring a credence function to be invariant under a transformation. The idea behind requiring that your credences satisfy some kind of Invariance principle is that if you have no reason to treat A and $T(A)$ differently, then you ought to have the same attitude to them. Any difference in attitude between A and $T(A)$ would not be warranted by your epistemic position.

---

[10]There are a number of other results that show that other classes of probability function can converge, for example Blackwell and Dubins (1962). See also Huttegger (2015).

> **Transformations**
>
> A transformation is a function $T : \Omega \to \Omega$ that induces a transformation on $\mathcal{B}(\Omega)$ through $T(A) = \{T(\omega) : \omega \in A\}$ for all $A \in \mathcal{B}(\Omega)$, and induces a transformation on the probability functions defined over $\mathcal{B}(\Omega)$ by $T(\mathbf{pr})(A) = \mathbf{pr}(T(A))$.

> **Precise invariance**
>
> A probability function $\mathbf{pr}$ is invariant under a collection of transformations $\mathcal{T}$ iff $T(\mathbf{pr}(A)) = \mathbf{pr}(A)$ for all $A \in \mathcal{B}(\Omega)$ and $T \in \mathcal{T}$.

The use of these kinds of invariance principles in arguing for further structure in your rational credence is motivated in large part by a desire for objectivity, by a kind of "impermissivism" about rational belief, or some form of "uniqueness thesis" (Jaynes 2003; Meacham 2014; Williamson 2010). We are searching for principles of rational credence that will narrow down the range of admissible priors, and possibly pick out the unique rational prior credence function.

1.3.1. The good news

Here's a first attempt at using invariance to produce a narrow class of permissible prior credences. Consider your prior credence over H_1 – the outcome of the first toss – and your prior credence over H_2 – the outcome of the second toss –; they should, arguably be the same.[11] After all, the process by which the coins are being tossed is such that the order of the tosses is irrelevant to you. This suggests that your credence should satisfy a principle known as "exchangeability".[12] Your credence should be invariant to permuting of the indices of the random variables.[13] In the context of the coin tossing example, the priors that are exchangeable are those that I have been calling "learning priors" (de Finetti 1974; Huttegger 2017). Satisfying some further invariance properties – the details can be found in Appendix B – entails a principle known as "Johnson's Sufficiency Postulate", which, together with Exchangeability (and Regularity[14]) entails that $\gamma = \frac{1}{2}$ (Paris and Vencovská 2015, Theorem 17.2 and Proposition 17.6). This gives us Carnap's continuum of inductive methods (for one predicate) c_λ with parameter λ (Huttegger 2017, Appendix A.1).[15] This is a pretty good outcome so far. The collection of credence functions that satisfy the above invariances all have the property that they learn on the basis of evidence. The free parameter – λ – captures how quickly you jump to conclusions. This arguably is a matter of epistemic taste, rather than something that should be controlled by basic principles of rationality.

---

[11] That is, $\mathbf{pr}(H_1 = j) = \mathbf{pr}(H_2 = j)$ for $j = 0, 1$.

[12] "Constant exchangeability" in the context of Paris and Vencovská (2015).

[13] Note that given exchangeability, learning the order of the tosses makes no difference, and thus the only thing that matters is the number of heads in n tosses. Thus it makes sense to use the shortcut of updating on S_n, rather than on an n-fold conjunction of outcomes of tosses.

[14] The principle that requires all non-contradictory propositions to have non-zero prior probability.

[15] See also Zabell (2005, Chapter 4).

Satisfying a *further* invariance condition (invariance under all structure-preserving mappings from the state space to itself)[16] forces $\lambda = 0$ (Paris and Vencovská 2015, Chapter 23). This choice of λ has some unpalatable consequences. For example before learning anything, you are certain that every observation will fall in the same category. That is, before getting any evidence, you are certain that the sequence of tosses will consist entirely of heads, or entirely of tails. So here it seems we have gone beyond the amount of invariance that we can plausibly require of a precise credence function. In a sense, we have arrived at the goal that impermissivists appear to be driving at: a set of principles on rational credence that pick out a unique credence function, c_0. On the other hand, this endpoint is undesirable, since c_0 is not a particularly palatable rational credence to adopt.[17]

1.3.2. The bad news

Consider a case of complete ignorance: a case where you know nothing about the epistemic situation. For *any* transformation of the state space T and any proposition A, it seems like you have no reason to treat A and $T(A)$ differently, and thus you ought to satisfy invariance with respect to the set of *all* transformations on the state space Ω. Call this universal precise invariance.

> **Impossibility I**
>
> No de Finetti coherent credence function satisfies universal precise invariance.

This is the first of several negative results we shall discuss in this paper.[18] We won't discuss the proof here, since it follows easily from later results. Before we move on to seeing how to escape this result, we'll take a short detour into the significance of impossibility results in general.

Impossibility theorems or "no-go" theorems are a pretty common form of formal result. They say, roughly, that some plausible set of principles are not jointly satisfiable. For example, Arrow's theorem says that no voting procedure can satisfy a plausible set of desiderata for such procedures (Gaertner 2009; Morreau 2014). Or the Kochen-Specker theorem in quantum physics establishes a contradiction between quantum mechanics and some plausible properties of the values of variables (Held 2008; Redhead 1987). The establishing of an impossibility result does not stop research dead in its tracks. Indeed, an impossibility result can be a spur to further new work. What an impossibility result can do is act as a kind of organising principle for the field. For example, voting methods can be classified by which of Arrow's axioms is weakened or denied and likewise for interpretations of quantum mechanics. I propose that the above impossibility, and the several further such results we shall discuss later, can provide the same kind of organising

---

[16] What Paris and Vencovská (2015) call "UNIV".

[17] Maher (2010) uses a different argument to arrive at the conclusion that λ should be 2. I don't discuss this further since it isn't really an argument that relies on invariance principles.

[18] See Norton (2007, 2008) for further arguments along these lines.

framework for formal epistemology. We will be able to classify theories of rational belief by which versions of the Coherence, Learning and Invariance principles they satisfy.

Note that the state space for the example of tossing the mystery coin k times has a particular structure. We have k random variables (the first toss, the second toss...the kth toss) all taking values in $\{0, 1\}$. Ω is a k-fold cross product of $\{0, 1\}$. Elements of the state space – outcomes of tossing the coin k times – are k-tuples of 0s and 1s.[19]

Impossibility I depends on an extremely strong kind of invariance; and implausibly strong form of symmetry. Consider the collection of all permutations of the 2^k k-tuples of coin tosses. As Carnap (1950, p. 565) notes, invariance with respect to that set of transformations yields a probability function that makes the coin tosses probabilistically independent of each other, and thus precludes any learning. We can draw two conclusions from this fact. First, even if an invariance principle is satisfiable, it might be undesirable because it would prevent learning. This is a phenomenon we will meet again later. Second, universal precise invariance (which entails invariance with respect to all permutations of the states) is too much invariance: we don't want to restrict ourselves to probability functions that cannot learn.

Let's discuss another problem for precise probabilistic Invariance approach. Recall in Section 1.3.1, we narrowed down the admissible priors to Carnap's continuum of inductive methods, c_λ. The problem with this is that your c_λ credences depend on arguably irrelevant features of the state space. We show this by discussing a classic puzzle case, this version is from van Fraassen (1989). Consider a factory that makes cubic boxes. Side lengths are distributed randomly, and the minimum and maximum side lengths are 0 metres and 2 metres (exclusive and inclusive respectively). What credence should you have that the next box produced will have a side length of less than one metre? Let L be the random variable for the next box's side length, and let ℓ_1 and ℓ_2 be the intervals $(0, 1]$ and $(1, 2]$ respectively. You have no reason to think it more likely that L will be in ℓ_1 than you do that it will be in ℓ_2. In the absence of any such evidence, it seems reasonable that you require your attitudes to the two propositions $L \in \ell_1$ and $L \in \ell_2$ to be the same. If we had called the interval $(1, 2]$ "ℓ_1" and called $(0, 1]$ "ℓ_2" instead – that is, if we had swapped the labels ℓ_1 and ℓ_2 around – then that should not make any difference to your credence. This is the sort of invariance that Walley (1991) calls "Symmetry" and de Cooman, De Bock, and Diniz (2015) call "category permutation invariance". So let $\mathbf{pr}(L \in \ell_i) = \frac{1}{2}$ for $i = 1, 2$. This is a well behaved de Finetti coherent credence function. So far so good.

Consider an alternative description of the box factory. The box factory produces cubic boxes of random *face area*, with face area randomly distributed between 0 metres square and 4 metres square. Let F be the random variable for the next box's face area, and consider the intervals $f_1 = (0, 1], f_2 = (1, 2], f_3 = (2, 3]$ and $f_4 = (3, 4]$. Similar

---

[19]Alternatively, we could describe the space as a space of k individuals with one predicate ("lands heads"). For the purposes of the coin toss example, this is equivalent, although the natural generalisations of the two perspectives are different: it's natural to add more possible (mutually exclusive) categories to the random variable approach, while it's natural to add more (logically independent) predicates to the predicate approach. In what follows I try to stick to the random variables perspective, but see Paris and Vencovská (2015) for the predicate perspective.

reasoning to the above suggests that each $\mathbf{pr}(F \in f_i)$ should be the same, and thus each is $\frac{1}{4}$. Again, no problem so far.

Now note that a cubic box has a side length in $(0, 1]$ metres if and only if it has a face area in $(0, 1]$ metres square. This isn't a fact about probability, it's a piece of background knowledge about geometry.[20] Given this fact, it seems that you ought to have the same attitude to the events $L \in \ell_1$ and $F \in f_1$, since they describe the same set of situations. Indeed background knowledge provides you with a mapping between the side length states and the face area states that takes "$L \in (a, b)$" to "$F \in (a^2, b^2)$", and you ought to have the same attitude to a side length proposition and its face length "translation". This is an instance of what Walley (1991) calls this "Embedding", and de Cooman, De Bock, and Diniz (2015) call "pooling invariance". This latter label makes sense, because the relation between the two descriptions is basically to "pool" the categories f_2, f_3, f_4 into one category of ℓ_2. Background knowledge appears to require that your attitudes are invariant under this mapping. This causes problems: the symmetry-derived probability for $L \in \ell_1$ is $\frac{1}{2}$ whereas the symmetry-derived probability for $F \in f_1$ is $\frac{1}{4}$. Given that the former is mapped to the latter by the translation or embedding, this violates Embedding.

That is, there is some structure to the state space you have credences over: there are a number of F_js and each of them falls into one of several categories, the f_is. In general terms, there is a set of random variables (i.e. the F_js) and they each take values in *the same* set of values (i.e. the set of f_is).[21] So the state space is a cartesian product of k "copies" of the set of categories $\{f_i\}$ for some k, and elements of the state space are k-length sequences of f_is. (This is just a generalisation of the idea we encountered earlier; that the state space for mystery coin toss outcomes is k-tuples of 0s and 1s.)

What we've seen is that in the box factory example, no probability function can be invariant with respect to a collection of transformations that includes the obvious side-length symmetries, the obvious face-area symmetries and the natural embedding of face-area propositions in the side-length proposition space. Given that each of these invariances seems plausible, this puts pressure on the idea that probability theory can provide suitably invariant priors for this kind of case. In Section 2.2, once we've introduced some necessary concepts, we shall tighten up this intuitive argument into a further impossibility result.

1.4. Impermissivism and ignorance

The basic question motivating this paper is "how constrained are your rational credences?" The fewer constraints one endorses, the more permissivist one is. The more constraints one wants to impose, the more impermissivist one is. At the far end of the impermissivist spectrum are those that endorse a "uniqueness thesis" that states that there is precisely one rational credence function for you to have. Even here there are a number of positions to distinguish, as Meacham (2014) points out. Do all agents in the

---

[20]This demonstrates that we are no longer in the realm of "Pure Inductive Logic" in the sense of Paris and Vencovská (2015), since we are appealing to non-logical facts about relations between propositions.

[21]In what follows, we will take the set of values to be finite, that is, we will ignore the fact that there's "really" an uncountable space, $(0, 4)$ in the background.

same epistemic situation have to have the same credence, or is there flexibility to allow different agents to be required to have different credences? That is, we can distinguish "Agent Uniqueness" and "Permission Parity", in the terminology of Meacham (2014). Agent Uniqueness requires that there is a unique rational credence for each agent, and Permission Parity requires that the same rational credences are permissible for all agents. The conjunction of these two properties gives us "Evidential Uniqueness". One way to satisfy Agent Uniqueness but not Permission Parity would be if different agents have different attitudes to epistemic risk, and there is a rationally mandated family of credence functions that correspond to different risk attitudes. Then, you are required to have the credence function from that family that corresponds to your risk attitude. The Carnap continuum c_λ is one such potential continuum. On the other hand, an approach like that of Williamson (2010) – where the rational credence is the most equivocal probability function consistent with your evidence – seems to satisfy Evidential Uniqueness (modulo some edge cases).

We want to find a well-motivated impermissivist epistemology, and we proceed by exploring what sort of invariance principles we can impose as requirements of rationality. The guiding principles are that the credence functions satisfying these invariances should (a) exist and (b) be able to learn. We require (a) because, as we saw, there are collections of invariance principles so strong that no probabilistic credence can satisfy them. We require (b) because the ability to learn is supposed to be (part of) what makes the "bayesian" paradigm attractive.

At this stage, it might help to distinguish two projects. The first, modest project is the project of exploring what the constraints are on your rational credences in a particular epistemic situation, where we take for granted certain features of the structure of the propositions learned. For example, in assuming that anti-inductive priors are irrational, we are implicitly making a judgment that proposition $H_1 = 1$ should be understood as evidence in favour of $H_2 = 1$.

Let's give an example. Suppose that the algebra of propositions contains $\{A, B, C, D\}$. We can now ask the question: "What should your credence be in B conditional on learning A"? With no other information about the nature of the propositions, it's unclear what kind of guidance we should expect rationality to provide. Call this a case of "complete ignorance". Now, if instead you knew that A, B, C, D are $H_1 = 1, H_2 = 1, H_1 = 0, H_2 = 0$ respectively, then you have more information about the structure of the state space: A and C are mututally exclusive, for example. And if you further knew that the outcomes are determined by tosses of a mystery coin, then perhaps certain probabilistic credences are ruled irrational, namely the anti-inductive ones. Call this a case of "structured ignorance". You are still ignorant of the bias of the coin, and that's something that you can hope to learn.

In the case of complete ignorance, arguably learning is hopeless. Rational learning involves there being some rational constraint on $\mathbf{pr}(B|A)$, and if you know nothing about what these two letters stand for, then how can you possibly hope to rationally constrain that expression? Luckily, though, we are almost never in complete ignorance, or at least, it's always at least possible to make some defeasible structural assumptions about the epistemic situation.

Let's take stock. We've seen a couple of (strong) forms of invariance that are inimical to learning (invariance with respect to all permutations of Ω, and invariance with respect to all automorphisms of the language). And we've seen a couple of (stronger?) forms of invariance that are simply unsatisfiable in a precise probability context: universal precise invariance (which is arguably too strong anyway), and invariance with respect to the acknowledged symmetries in the box factory case (which arguably is not). We also saw that weaker forms of invariance do allow us to isolate a plausible class of credence functions that can learn.

Most work responding to the box factory puzzle goes the route of modifying Invariance. One might start by noting that the transformations discussed in the box factory are not all on a par. The side-length/face area embedding invariance is a property that you have positive reason to require because it follows from your background knowledge that that embedding maps coreferring propositions onto each other. So you have some kind of *explicit invariance*. Contrast this with the side-length symmetry or face-area symmetry where the invariance is one that you merely have no reason to think doesn't hold. Call this a *default invariance*. One approach to avoiding the impossibility would be to endorse satisfying explicit invariances but deny any requirement to satisfy default invariances. This means that often there will be many probability functions consistent with your evidence and satisfying your explicit invariances. If you are comfortable with this permissivism about rational credence, then invariance has little grip on you. If however, you want to endorse some form of uniqueness thesis then the natural way to proceed is via invariance principles (Jaynes 2003; Meacham 2014; Williamson 2010). Thus you will need to search for just the right amount of invariance to pick out one probability function or one suitably small class of functions.[22]

Strangely, one popular impermissivist response to the box factory goes a different way: it says that you ought to satisfy one and only one of the default invariances, and then use the explicit invariance to determine the probabilities of the other variables. That is, you pick one of side-length or face-area as the basic variable, and satisfy the default invariance with respect to that subspace Then, you use the translation between side-length and face-area to determine probabilities for the other variable (which do not satisfy the invariance). The idea is that this is acceptable because which variables you take to be the basic ones is determined by your evidence (Williamson 2010). That is, if you are presented the box factory as a factory that produces cubes with random side length, then it's appropriate to satisfy symmetry with respect to the L description of the situation, rather than with respect to F. Exploring this line of reasoning can take us to some remarkably powerful results. For example, in Pure Inductive Logic, Paris and Vencovská (2015) start with a slightly more complex language of propositions – one that includes predicates and quantifiers – and show that a rational agent satisfying various forms of symmetry can, for example, become more confident in a universal generalisation by observing instances of it. So Invariance, judiciously applied, can yield Learning![23]

---

[22]Permissivism/impermissivism is not a binary choice but more of a spectrum of views that vary how restrictive their view of what the rational credences are. We'll return to this point later.

[23]For another example of a surprising connection between Invariance and Learning, see Grove and Halpern (1998) and van Fraassen (1989), who show that Invariance under Embedding (and surpris-

In general, the response to the above problems has typically been to keep fixed the strong (de Finetti) Coherence constraint, and explore what weaker versions of Invariance it is compatible with. In this paper, we shall be exploring the prospects for going the other way: is there an interesting but weaker notion of Coherence that is compatible with more Invariance?

2. Imprecise Probabilities

So, de Finetti coherence is incompatible with a strong form of invariance. For the remainder of this paper we will explore various attempts to weaken coherence so as to permit stronger invariance principles to be satisfied. The desire to satisfy invariance principles is driven, primarily by impermissivist, objectivist conceptions of rational credence.

Independently of the status of impermissivism or uniqueness, some have argued that picking a specific prior probability is unwise, and that we should instead work with *sets* of probability functions as the basic credence-representing object. This move to *Imprecise Probabilities* (IP) has gained some traction over the past decade, but there are several extant problems. We don't have time to discuss the motivations offered for this move to IP,[24] our focus is on the consequences of such a move.

Consider $\mathbb{P}$, a set of prior probabilities with a common algebra of events. We shall call this a *credal set*. Define $\mathbb{P}(X) = \{\mathbf{pr}(X), \mathbf{pr} \in \mathbb{P}\}$, the set-valued function defined over the same algebra as the members of $\mathbb{P}$; and define $\overline{\mathbb{P}}(X) = \sup \mathbb{P}(X)$, the "upper probability" that takes the largest value assigned to an event X by members of $\mathbb{P}$. We can also define a "lower probability": $\underline{\mathbb{P}}(X) = \inf \mathbb{P}(X)$. $\underline{\mathbb{P}}$ satisfies Boundedness, Monotonicity and Superadditivity; $\overline{\mathbb{P}}$ satisfies Boundedness, Monotonicity and Subadditivity. In this section, we shall explore IP's prospects for satisfying Coherence, Learning and Invariance.

2.1. Walley coherence

For the moment, let's consider whether the lower probability $\underline{\mathbb{P}}$ can function as a replacement for $\mathbf{pr}$ as a measure of disposition to bet as discussed earlier. That is, is $\underline{\mathbb{P}}$ de Finetti coherent as defined earlier? The answer is no. This is so since de Finetti coherence is defined in terms of two-sided betting, and the second clause of the definition of two-sided betting requires you to take the other side of any bet you don't find acceptable. So for example, since $\underline{\mathbb{P}}$ is only superadditive, you might have $\underline{\mathbb{P}}(A) = \underline{\mathbb{P}}(\neg A) \leq x < y < \frac{1}{2}$. Thus, you would find a unit bet on A for price x to be unacceptable and likewise for $\neg A$. But if you're forced to accept the other sides of both bets for any price y greater than x, you've paid $2(1 - y) > 1$ for two bets that guarantee you winnings of 1. That is, your net gain will be negative in every state of the world, which was how we characterised incoherence. As might be clear, the problem is the second clause of two sided betting that forces you to take the other side of unacceptable bets. So let's define a less restrictive betting scenario, which thus yields a less restrictive coherence requirement.

---

ingly little else) is sufficient to require that you update by conditionalisation.

[24]But see Augustin et al. (2014), Bradley (2019), Joyce (2005, 2010), Levi (1974, 1980), and Walley (1991).

> **One-sided betting**
>
> - A bet that wins $1 - x$ if A is true, and loses x if A is false is acceptable if $\mathbf{pr}(A) > x$.

And on the basis of this constraint, we define some coherence requirements.

> **Avoid Sure Loss**
>
> Your prices for gambles in a one-sided betting scenario should be such that no combination of acceptable bets is guaranteed to yield a loss in every state.

Note that in the case of de Finetti coherence, avoiding sure loss also guaranteed a further good-making feature, namely that any collection of bets that always wins you money is acceptable to you. This is what Hájek (2008) calls a "Czech book". In the case of two-sided betting this latter condition is an automatic consequence of the former condition (Troffaes and de Cooman 2014, Theorem 4.12), but for one-sided betting, the two conditions are distinct. So we might want to add in this extra clause.

> **Walley coherence**
>
> Your prices for gambles in a one-sided betting scenario should be such that no combination of acceptable bets is guaranteed to yield a loss in every state and every combination of bets that yield a gain in every state is acceptable.

So we have our weaker notion of coherence: your buying and selling prices for bets can be different, and there may be some prices such that you're not required to take either side of the bet. Call this weaker notion Walley-Coherence.[25] This is strictly stronger than Avoiding Sure Loss. We then have the following result: you are Walley coherent if and only if your maximum buying betting quotients are the lower probabilities for some credal set (Troffaes and de Cooman 2014, Theorem 4.38).[26]

2.2. Invariance weak and strong

A transformation of Ω induces a transformation on a credal set pointwise: $T(\mathbb{P}) = \{T(\mathbf{pr}) : \mathbf{pr} \in \mathbb{P}\}$. This allows us to make a distinction between two kinds of invariance (de Cooman and Miranda 2007).

> **Weak Invariance**
>
> $\mathbb{P}$ is weakly $\mathcal{T}$-invariant iff $T(\mathbb{P}) \subseteq \mathbb{P}$ for all $T \in \mathcal{T}, \mathbf{pr} \in \mathbb{P}$.

---

[25]Walley wasn't the first to discuss something like this concept (see, for example, Smith (1961)), but his influential book (Walley 1991) was the first book-length systematic treatment of the idea.

[26]Note this result is for the separate coherence of an individual lower probability, not the joint coherence of a prior and conditional lower probability. Results there are somewhat more complex.

> **Strong Invariance**
>
> $\mathbb{P}$ is strongly $\mathcal{T}$-invariant iff $T(\mathbf{pr}) = \mathbf{pr}$ for all $T \in \mathcal{T}, \mathbf{pr} \in \mathbb{P}$.

We can use the credal committee metaphor (Joyce 2010) to understand something of the difference between these concepts. Think of $\mathbb{P}$ as a committee of probability functions who, collectively, make decisions on your behalf. Weak invariance just requires that your overall credal state collectively satisfies a kind of invariance, whereas strong invariance requires each committee member to satisfy that invariance. Consider an instance of Symmetry. Within the committee there might be some member who is confident the next marble drawn from an urn will be red, and a different member who is confident it will be blue. Such a model doesn't satisfy strong invariance, since if you switch the labels "red" and "blue", the committee member's credences do change: the member who thought "red" more likely, once the labels have been swapped, now thinks "blue" is more likely. But such a model might still be weakly invariant so long as there is another committee member who is just as confident the marble will be blue as the original member is that it's red. Weak invariance seems the appropriate thing to require for default invariance, whereas strong invariance seems appropriate only when we *know* that the events concerned really do have some sort of symmetry: that is, for explicit invariance. This is what de Cooman and Miranda (2007) refer to as the distinction between "symmetry of models" (weak invariance) and "models of symmetry" (strong invariance). Weak and strong invariance are obviously the same for precise probabilism (when $\mathbb{P}$ is a singleton), but they are importantly distinct for IP models.

As before, let's consider versions of universal invariance; invariance with respect to all transformations.

> **Impossibility II**
>
> No Walley-coherent credence satisfies universal strong invariance.

This is unsurprising, given that it would be very strange to have an explicit invariance linking *any* two propositions. This is an incoherently strong kind of symmetry.

Before continuing, we shall need the following definition.

> **Vacuous prior**
>
> Let $\mathbb{V}$ be the set of all probability functions over the algebra of events.

The vacuous model – the model derived from the set of all probability functions over the algebra – assigns $\underline{\mathbb{V}}(A) = 0$ to all non-tautologous propositions (and $\overline{\mathbb{V}}(A) = 1$).

> **Possibility I**
>
> The only Walley-coherent credence that satisfies universal weak invariance is the

vacuous prior, $\mathbb{V}$.

The possibility result is proven in de Cooman and Miranda (2007, Theorem 4), which also demonstrates that impossibility I holds. From this it follows that impossibility II holds, since universal strong invariance for a credal set is essentially equivalent to universal precise invariance for every member of that set.

2.3. Belief Inertia

So with a weakened version of Coherence, we can satisfy Universal Weak Invariance. But, as we are about to see, the priors that do so – the vacuous priors – don't yield satisfying learning behaviour.

In order to discuss learning, we need to extend our discussion of conditionalisation to credal sets. Joint coherence of an unconditional and a conditional lower probability is a little trickier than the precise case[27] but for our purposes it is enough to say that Generalised Conditioning is the most informative coherent way to update: define a conditional credal set: $\mathbb{P}(-|E) = \{\mathbf{pr}(-|E), \mathbf{pr} \in \mathbb{P}, \mathbf{pr}(E) > 0\}$, the set obtained by conditioning each member of $\mathbb{P}$ on the same evidence.[28] Note the caveat that we throw out any $\mathbf{pr}$ that assign zero probability to the evidence received. We shall return to that feature shortly.

The vacuous prior $\mathbb{V}$ satisfies a reasonably strong form of Coherence, and a very strong form of Invariance. However, $\mathbb{V}$ is no good at learning. Consider gathering evidence about the mystery coin. Among the priors in $\mathbb{V}$ are the anti-inductive priors: the credence function that becomes very confident that the next toss will land tails on learning that the previous toss landed heads. Complete ignorance for IP can be really really *complete*! This is an example of the problem of *belief inertia*: Levi (1980) noticed this problem (chapter 13), and Walley (1991) was also well aware of the issue. Philosophical interest in the topic has picked up recently, with a number of papers discussing it (Castro and Hart 2019; Joyce 2010; Lassiter 2020; Lyon 2017; Moss forthcoming; Rinard 2013; Vallinder 2018). As Walley notes:

> The vacuous previsions really are rather trivial models. That seems appropriate for models of "complete ignorance" which is a rather trivial state of uncertainty. (Walley 1991, p. 93)

When it comes to the vacuous prior, it doesn't seem like belief inertia is really a problem. If you don't take there to be *any* constraints to how your past evidence constrains your attitudes about the future then you're going to struggle to make any sensible inference, just as an inductive sceptic would. If you can't rule out these crazy priors, then obviously learning will be beyond you. If we're in a situation of complete ignorance as regards some

---

[27]See Augustin et al. (2014) section 2.3, Troffaes and de Cooman (2014) chapter 13 and Zaffalon and Miranda (2013). What we call "Generalised Conditioning" is sometimes referred to as "Regular Extension".

[28]"Most informative", in the sense that for any other conditional lower probability that is jointly coherent with $\underline{\mathbb{P}}$ will be less precise than the GC update (Miranda 2009).

algebra of propositions, it's unclear why we would expect to be able to learn about one proposition by updating on a distinct proposition. And this is the correct result: if you're accommodating inductive sceptics in your credal committee, you probably will struggle to learn. The right lesson to draw here is that since satisfying universal weak invariance requires using the vacuous prior, that's too much invariance for rational learning!

The vacuous prior is the correct response to complete ignorance. As we saw in Section 1.2, even in the precise case complete ignorance seems inimical to rational learning. So, for the purposes of modelling complete ignorance, the fact that the vacuous prior suffers from belief inertia is not a bug, it's a feature.

This is not, however, the end of the trouble belief inertia causes. Let's return to the coin-tossing problem to see how credal sets get stuck, even when we think they shouldn't. Let $\mathbb{B}$ be the set of all learning priors as defined in Section 1.2. That is, $\mathbb{B}$ is a set containing priors, all of which can learn on the basis of evidence. This set of priors isn't vacuous: all the anti-inductive priors have been ruled out, for example. This means that $\mathbb{B}$ doesn't satisfy universal weak invariance, but we'll see in the next section that it does still satisfy a quite strong invariance principle. $\mathbb{B}$ is "near-ignorance for H" in the sense that $\mathbb{B}(H) = (0, 1)$.

Near ignorance

$\mathbb{P}$ is near ignorance for $\mathcal{H}$ if, for all $H \in \mathcal{H}$, $(0, 1) \subseteq \mathbb{P}(H)$.

It is not the case, however, that every event has uninformative prior probability. For example, consider the event of one head and one tail in the next two tosses. Whatever your credence for heads, the highest the credence for this event could be is $\frac{1}{2}$.[29] To put this another way, the vacuous prior is so vacuous that you can't even take any attitudes towards the structure of the space of possible outcomes: you can't take the outcomes of two coin tosses to be exchangeable.

Suppose we've gathered evidence $S_n = h$. For $\mathbb{B}$, the conditional probabilities of H given S_n are all the numbers of the form: $\frac{\gamma\lambda+h}{\lambda+n}$. Hold h and n fixed, and for any $\lambda > 0$ and $\gamma \in (0, 1)$, that fraction is a possible probability value for heads. As λ gets bigger, this fraction tends to γ, so since γ is unconstrained, the range of values for the probability of heads, given $S_n = h$, covers all of $(0, 1)$ regardless of what evidence you acquire. The problem is that, even though each prior is learning, there are always more and more recalcitrant priors that cover the whole range of possible chances. The bigger λ is the less effect the evidence has on moving that prior towards the true chance. In summary, even though every $\mathbf{pr} \in \mathbb{B}$ satisfies ASCL, it is not the case that $\underline{\mathbb{B}}$ satisfies ASCL.

So, even in quite friendly circumstances – every prior in the set can learn – certain IP prior credal sets are unable to learn from evidence. So belief inertia is not just that the vacuous prior cannot learn: the problem is that even some prima facie amenable looking prior credal sets fail to learn.

---

[29]This is so since every $\mathbf{pr} \in \mathbb{B}$ makes the coin tosses exchangeable, and thus by de Finetti's representation theorem the probability for one head and one tail is $\int 2p(1-p)dp$ for some measure dp. Since $2p(1-p)$ has a maximum value of $\frac{1}{2}$ in the range $[0, 1]$, the integral does too. And since this is true for every $\mathbf{pr} \in \mathbb{B}$, this is also a bound on $\overline{\mathbb{B}}$.

It looks like even in cases of *structured ignorance* – where we would want a rational agent to be able to learn on the basis of evidence – imprecise probabilities suffer from belief inertia. In the next section, I show that this isn't the case.

3. Learning by ignoring the stubborn

It is common to suggest that the solution to this problem of belief inertia will involve ruling out "extreme" priors, those near 0 or 1 for the chance of heads. Joyce (2010) suggests this, as do, for example, Castro and Hart (2019), Lassiter (2020), Lyon (2017), and Rinard (2013). However, as Vallinder (2018) points out, ruling out those **pr** with **pr**(H) very close to 0 or 1 is not sufficient to prevent the phenomenon of belief inertia. The very stubborn priors (the high λ priors) prevent learning, so removing the priors with "extreme" values for γ is to miss the point. The real culprit is λ.

This suggests one solution to belief inertia: just set a maximum value of λ, call it $\lambda_{\max}$, and call the resulting credal set $\mathbb{B}_{\lambda_{\max}}$. This then gives us non-trivial bounds on the conditional lower probability of heads, and, in fact, we converge on the truth as n increases, as desired.[30] In fact, the bounds on your credence are then given by:

$$\underline{\mathbb{B}}_{\lambda_{\max}}(H|S_n = h) = \frac{S_n}{n + \lambda_{\max}} \qquad \overline{\mathbb{B}}_{\lambda_{\max}}(H|S_n = h) = \frac{S_n + \lambda_{\max}}{n + \lambda_{\max}}$$

This idea was already present in Walley (1991), and has been expanded and extended since (Bernard 2005; Walley 1996).[31] The generalised model is known as the "imprecise Dirichlet model" (IDM).

The distinction between the vacuous prior – the set of all probabilities – and a near-ignorance prior allows us to clarify some somewhat imprecise claims in Moss (forthcoming): Moss states that agents with "maximally imprecise" credences (for a particular proposition H) can sometimes be such that they cannot learn (by which she appears to mean something like they don't satisfy Informativeness for H). She then concludes that rational agents should not have such maximally imprecise credences: agents should respond to evidence in a way that precludes inertia. She argues that such a constraint is a "global constraint", i.e. not a pointwise constraint on each probability in the credal set. It seems that Moss intends "maximally imprecise" credal sets to be near-ignorance, rather than specifically the vacuous prior. But as we've seen, some near-ignorance priors can indeed learn. Further, the kind of "anti-stubbornness" condition Moss endorses, in the case of learning priors, amounts to fixing a maxmimum value for λ, which is a pointwise constraint on $\mathbb{B}$. So some but not all near-ignorance priors fail to learn, and there are some near-ignorance priors where the constraint that allows learning is a pointwise constraint.

---

[30]Thus, removing "extreme" γ priors is not only not sufficient to prevent belief inertia, it is not necessary either.

[31] Bernard (2009) is the introduction to a special issue devoted to IDM. See also the references in Augustin et al. (2014, §7.4.3.2).

IDM priors also demonstrate that Castro and Hart (2019) are wrong to assert that having a $(0, 1)$ prior for H forces you to have a $(0, 1)$ posterior for every possible evidence proposition.[32]

Castro and Hart (2019) argue that imprecise impermissivists should endorse the principle that your credences cannot be arbitrary, and thus your credal set should consist of all the (precise) credences that are consistent with your evidence and the principles of rationality (p. 1628). One might think that this means that ruling out *any* learning prior, no matter how big its λ parameter is off the menu, and thus that imprecise impermissivists must use all of $\mathbb{B}$ as their credal set in the coin tossing example. One way to respond to this argument would be to suggest an additional principle of rationality that says that the lower probability for heads given a particular amount of evidence should be at least some non-zero number.[33] This would reflect the amount of epistemic risk-taking an agent is willing to accept. Any non-zero number for any particular evidence proposition would put an upper bound on λ_{max} and thus guarantee learning.[34] Such a principle would not be arbitrary: it would reflect the agent's epistemic risk profile. Even if Castro and Hart wanted to insist on Permission Parity, and argue that any particular risk attitude would be arbitrary, it is still the case that an agent with $\mathbb{B}$ as her priors has non-trivial credences in propositions about the number of heads in the next n' tosses.

IDM credal sets do not satisfy the universal weak invariance condition we outlined before since they are not vacuous. They do, however, satisfy about as much Invariance as we could expect them to.

Now consider the analogues of symmetry and embedding not for the states in the state space, but for the categories that the samples can fall into. That is, consider the invariance principles that require that your credences should be invariant under permutations of the category labels, and under the pooling of several categories into one larger category (like, for example, when f_2, f_3 and f_4 all get mapped into ℓ_2). Call weak invariance under this collection of transformations "Representation Insensitivity" (RI).[35]

We can now summarise the above discussion as follows.

> **Impossibility III**
>
> No de Finetti-coherent credence satisfies Representation Insensitivity.

That is, even if we accept that universal weak invariance is too much invariance, there is a level of invariance that seems reasonable and that no precise probability can satisfy.

---

[32] A near ignorance prior in H is not sufficient for near ignorance posterior, nor is it necessary, as the phenomenon of dilation demonstrates (Pedersen and Wheeler 2014; Seidenfeld and Wasserman 1993).

[33] Something like this approach is part of Theorem 26 of de Cooman, De Bock, and Diniz (2015), which provides a partial characterisation result for IDM priors.

[34] An alternative proposal building on Cattaneo (2008, 2014) is possible, but that must wait for future work.

[35] de Cooman, De Bock, and Diniz (2015) actually use the stronger condition that $T(\mathbb{P}) = \mathbb{P}$, rather than $T(\mathbb{P}) \subseteq \mathbb{P}$ as in weak invariance, but I am fairly sure that weak invariance with respect to transformations on category labels is sufficient for my purposes. They also include another kind of invariance in RI (renaming invariance) which I am taking for granted throughout.

This formalises the idea that a precise probability can't accommodate the two different descriptions in the box factory puzzle. The three invariances we outlined earlier for the box factory case are all instances of RI. So, since no precise probability could satisfy all three of them, it's no surprise that no precise probability can satisfy all three plus all the other invariances required by RI.

On the other hand, an IP model can satisfy a significant level of invariance and still satisfy a strong kind of Learning.

> **Possibility II**
>
> For any $\lambda_{\max}$, $\mathbb{B}_{\lambda_{\max}}$ satisfies Walley-coherence, Representation Insensitivity, Exchangeability and $\underline{\mathbb{B}}_{\lambda_{\max}}$ satisfies ASCL.

These results follow from Theorems 21 and 4 of de Cooman, De Bock, and Diniz (2015), respectively.

Let's take stock. Throughout we have been motivated to understand how we can accommodate the deep tension between Coherence, Invariance and Learning. Precise probabilities can satisfy a strong form of Coherence and Learning, but they are somewhat unsatisfactory when it comes to Invariance. Imprecise Probabilities offer an alternative approach to credence, if we allow a modest relaxation of the Coherence requirement. Within this framework we *can* satisfy a very strong Invariance requirement – using the vacuous prior – but at the cost of losing any hope of Learning. We then saw two potential avenues. First, if we move away from the "complete ignorance" idea that motivated the strong Invariance requirement, towards a case of inference under a particular set of structural assumptions – predictive inference for categorical variables under exchangeability – then we can still satisfy a strong Invariance constraint that is consistent with the structural assumptions, remain Walley-Coherent and satisfy Learning (this was the IDM approach).

Recall that we were motivated to satisfy as much invariance as was warranted by impermissivist or objectivist intuitions about rational credence. IDM priors satisfy quite a lot of invariance, but there are still uncountably many distinct IDM models, one for each choice of $\lambda_{\max}$. Note that the degree of freedom that remains in the IDM case is the $\lambda_{\max}$ parameter, which governs how quickly you learn: large $\lambda_{\max}$ means slow learning, small $\lambda_{\max}$ means fast learning. It seems plausible that this is the sort of thing that should be governed by your epistemic risk attitudes or level of epistemic conservativity,[36] rather than being imposed upon you by a norm of invariance. An impermissivism that would pick out a single level of conservativity as the uniquely rational one seems, to me, to be too much impermissivism.

4. Conclusion

We started with the problem that no approach to rational credence could satisfy de Finetti's strong version of Coherence, Learning and Invariance. We explored the pos-

---

[36]This is a different kind of "conservativity" from that discussed by Konek (forthcoming).

sibility of weakening the coherence condition. Walley-coherence (one-sided betting) allowed us to satisfy Invariance, but in doing so, we failed to satisfy Learning. This led to two options: when certain structural constraints are appropriate – which entail violations of our demanding universal weak invariance condition – it is possible to satisfy reasonably strong invariance conditions and satisfy Walley-coherence and learn from evidence (Section 3).[37] It's striking how difficult it is to do justice to these three kinds of desiderata for rational credence. This paper contributes to recognising this deep and enduring tension in mathematical models of rational belief and inference.[38]

A. Beta distributions

A beta distribution is a probability distribution over $[0, 1]$ that we interpret as a probability over possible biases of the coin or the chances of heads. The "learning priors" are the expectation of a beta distribution with parameters γ, λ, and the conditional probability is the conditional expectation.

Beta distributions are a conjugate prior for evidence drawn from a binomial distribution, meaning that if you conditionalise on a beta prior with binomial evidence, you get a posterior which is also a beta distribution, with different parameters. In particular, if you conditionalise on $S_n = h$, $\lambda_{\text{new}} = \lambda + n$ and $\gamma_{\text{new}} = \frac{\gamma\lambda+S_n}{\lambda+n}$. See Berger (1980, p.96 ff.).

In fact, beta distributions are more often parameterised by $\mu, \nu > 0$ which are such that and $\mu = \gamma\lambda$, and $\nu = (1-\gamma)\lambda$, but our preferred parameterisation (which Walley (1991, 1996) discusses) makes the discussion of "stubbornness" easier.

Figures 3 and 4 illustrate the effect of the two parameters. Each line represents a probability distribution over the chance of heads, $\mathbf{pr}(H)$ for a particular distribution is the expectation of this distribution. As we can see, holding λ fixed and varying γ (Figure 3) yields distributions that differ in terms of their expectation, but they are all moved by the evidence in broadly the same way. On the other hand, holding γ fixed and varying λ yields distributions that are centered around the same mean, but vary in how spread out they are, and also in how much they move in response to evidence.

B. Invariance and JSP

Note that this appendix is taking the "predicate perspective" rather than the random variable perspective taken in the main text (see footnote 19). We're going to describe the general case (since the case of coin tossing, where $q = 2$ has some peculiarities we return to at the end). Exchangeability – constant exchangeability in the terminology of

---

[37]The structure required to yield satisfying learning behaviour may seem quite strong. But there is scope to explore weakenings of the premises discussed here. For example, exchangeability can be weakened in a number of ways that allow us to accommodate a number of other kinds of structure (see Zabell (2005, Chapter 11), Huttegger (2017, Appendix B), Skyrms (2012, Chapter 11)), and some of these generalisations have been imported to the IP setting (De Bock et al. 2016).

[38]For further results that point to this tension, see de Cooman, De Bock, and Diniz (2015), de Cooman and Miranda (2007), Eva (forthcoming), Halpern and Koller (2004), and Norton (2007, 2008).
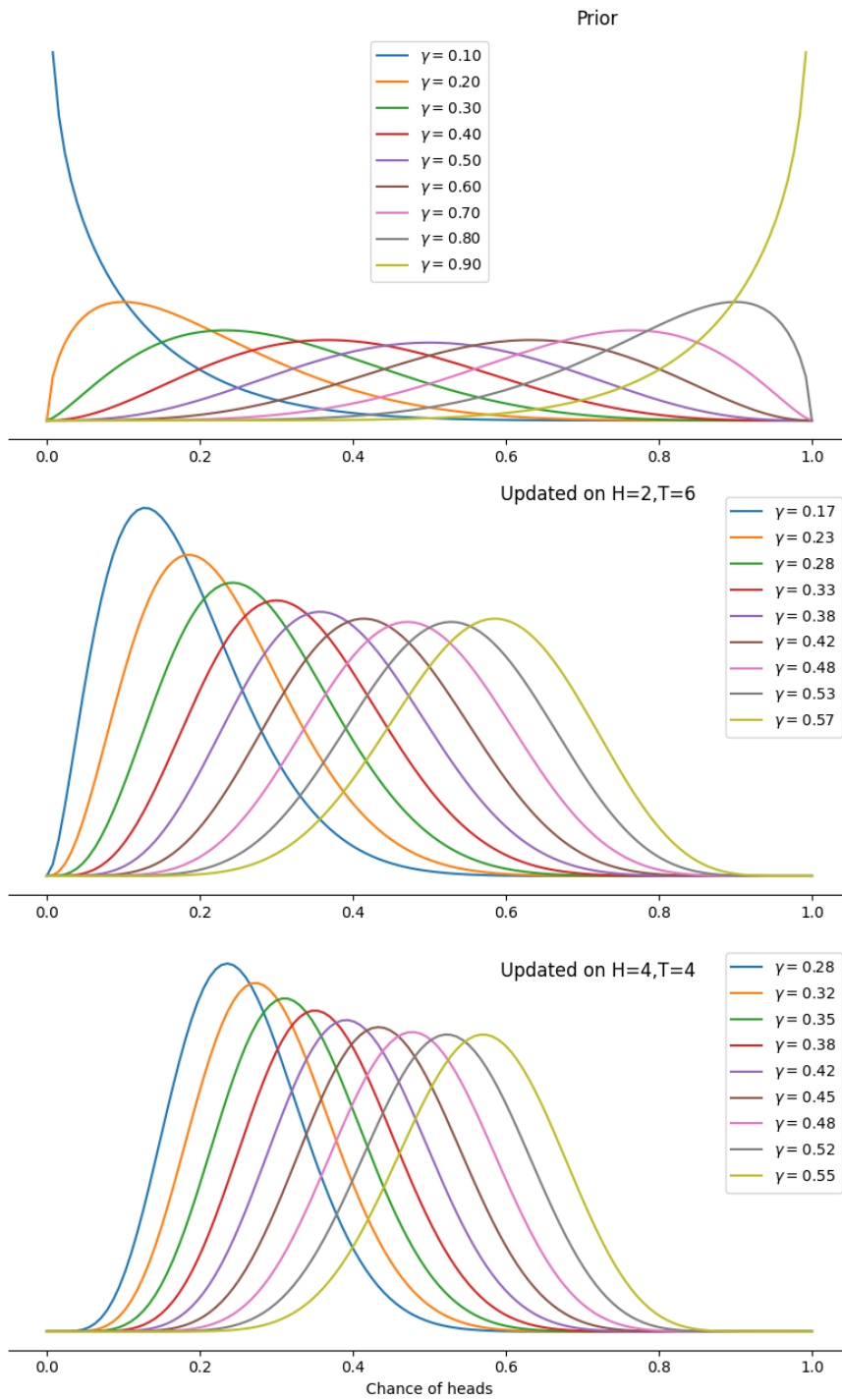
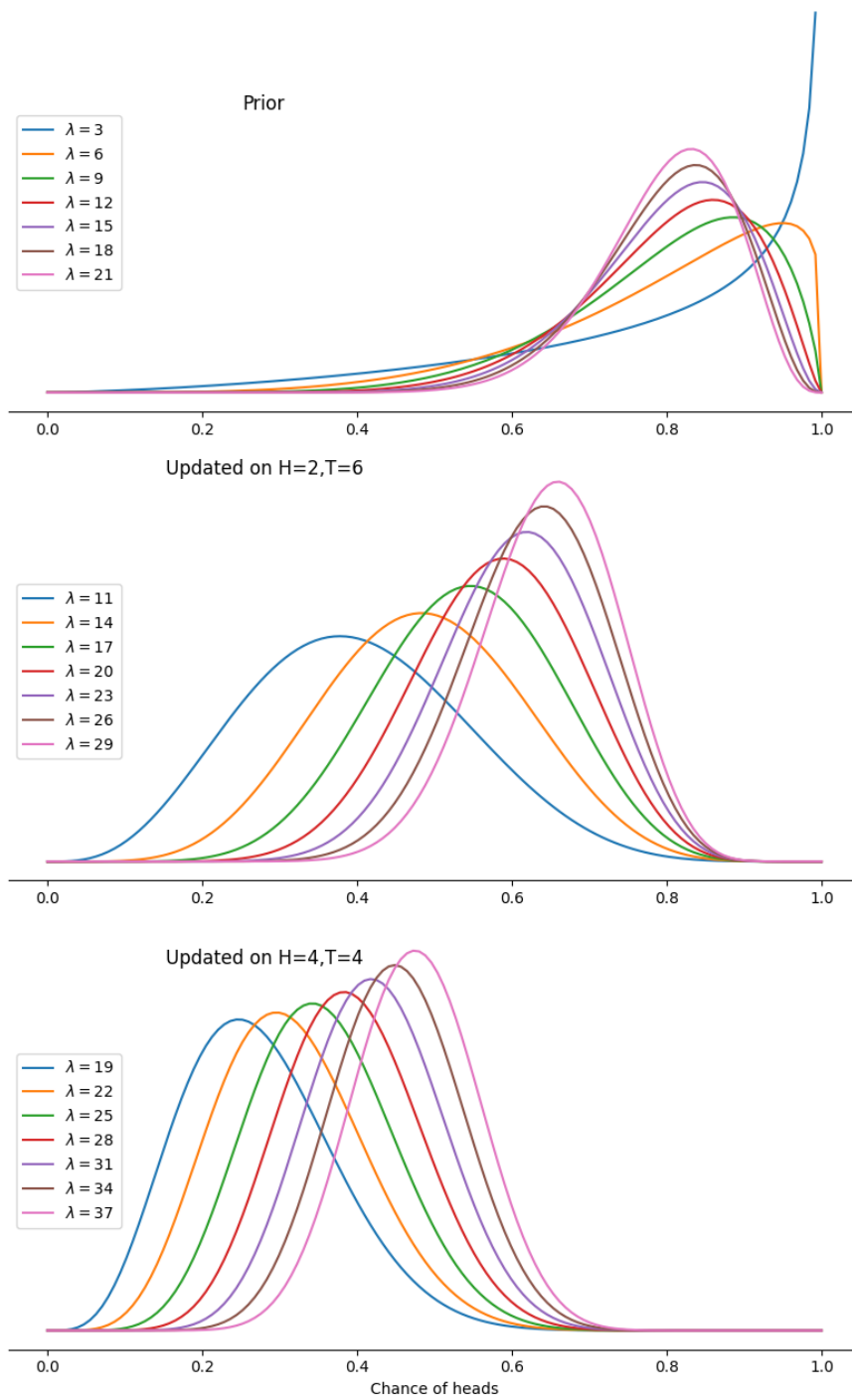Figure 3: Beta distributions holding λ fixed

Figure 4: Beta distributions holding γ fixed

Paris and Vencovská (2015) – entails that the probability that the next draw will fall in category i depends only on the sequence of numbers of draws observed for the different categories. That is, the order of observations doesn't matter.

One further property that is required is "atom exchangeability" which, essentially, requires that the probability that the next draw is i depends only on the multiset of numbers of draws obvserved for the different categories. That is, the identity of the category labels doesn't matter. (This is invariance under permutation of the category labels).

Next we also require SDCIP, which basically means that propositions that have no predicates or individuals in common with each other are probabilistically irrelevant to each other.

And finally we have ULi – a principle of language invariance – which requires your credence function to be a member of a class of credence functions (one for each predicate language) such that a credence function for a more expressive language, when restricted to a less expressive language, is equivalent to the corresponding credence function in the family.

If the class of credence functions satisfying ULi all satisfy constant exchangeability, atom exchangeability and SDCIP, then each credence function in the class satisfies Johnson's Sufficientness Postulate (Paris and Vencovská 2015, Proposition 17.6). Constant exchangeability, regularity and JSP are then sufficient to require your credence to be equivalent to a member of Carnap's continuum of inductive methods (see Zabell (2005, Chapter 4), Huttegger (2017, Appendix A.1) or Paris and Vencovská (2015, Theorem 17.2)).

This result doesn't hold for the coin tossing example, since Proposition 17.6 only holds for at least two predicates (hence four "atoms" or categories), and the result linking JSP to the continuum of inductive methods holds for at least 3 categories; although this latter result can be made to hold for the case of two categories by requiring that the probability of heads is a linear function of the number of heads observed so far.

References

Augustin, Thomas et al., eds. (2014). *Introduction to Imprecise Probabilities*. John Wiley and Sons.

Berger, James O. (1980). *Statistical Decision Theory*. Spring-Verlag.

Bernard, Jean-Marc (2005). "An introduction to the imprecise Dirichlet model for multinomial data". In: *International Journal of Approximate Reasoning* 39, pp. 123–150.

— (2009). "The imprecise Dirichlet model". In: *International Journal of Approximate Reasoning* 50.2, pp. 201–203.

Blackwell, David and Lester Dubins (1962). "Merging of Opinions with Increasing Information". In: *Annals of Mathematical Statistics* 33, pp. 882–886.

Bradley, Seamus (2019). "Imprecise Probabilities". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta.

Carnap, Rudolf (1950). *Logical Foundations of Probability*. Chicago University Press.

Castro, Clinton and Casey Hart (2019). "The Imprecise Impermissivist's Dilemma". In: *Synthese* 196.4, pp. 1623–1640.

Cattaneo, Marco (2008). "Fuzzy Probabilities based on the Likelihood Function". In: *Soft Methods for Handling Variability and Imprecision*. Ed. by D. Dubois et al. Springer, pp. 43–50.

— (2014). "A Continuous Updating Rule for Imprecise Probabilities". In: *Information Processing and Management of Uncertainty in Knowledge Based Systems*. Ed. by Anne Laurent et al. Springer, pp. 426–435.

Christensen, David (2004). *Putting logic in its place*. Oxford University Press.

De Bock, Jasper et al. (2016). "Representation theorems for partially exchangeable random variables". In: *Fuzzy Sets and Systems* 284, pp. 1–30.

de Cooman, Gert, Jasper De Bock, and Márcio Alves Diniz (2015). "Coherent Predictive Inference under Exchangeability with Imprecise Probabilities". In: *Journal of Artificial Intelligence Research* 52, pp. 1–95.

de Cooman, Gert and Enrique Miranda (2007). "Symmetry of models versus models of symmetry". In: *Probability and Inference: Essays in Honor of Henry E. Kyburg Jnr.* Ed. by William Harper and Gregory Wheeler. Kings College Publications, pp. 67–149.

de Finetti, Bruno (1964). "Foresight: Its Logical Laws, Its Subjective Sources". In: *Studies in Subjective Probability*. Ed. by Henry E. Kyburg and Howard E. Smokler. Wiley, pp. 97–158.

— (1974). *Theory of Probability*. Wiley.

Eva, Benjamin (forthcoming). "Principles of Indifference". In: *The Journal of Philosophy*.

Gaertner, Wulf (2009). *A Primer in Social Choice Theory*. Oxford University Press.

Grove, Adam and Joseph Y. Halpern (1998). "Updating sets of probabilities". In: *Proceedings of the fourteenth conference on uncertainty in AI*, pp. 173–182.

Hájek, Alan (2008). "Arguments for—or against—probabilism?" In: *British Journal for the Philosophy of Science* 59, pp. 793–819.

Halpern, Joseph Y. and Daphne Koller (2004). "Representation Dependence in Probabilistic Networks". In: *Journal of Artificial Intelligence Research*, pp. 319–356.

Held, Carsten (2008). "The Kochen-Specker Theorem". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. URL: `%5Curl%7Bhttp://plato.stanford.edu/archives/win2008/entries/kochen-specker/%7D`.

Huttegger, Simon (2015). "Merging of Opinions and Probability Kinematics". In: *Review of Symbolic Logic* 8.4, pp. 611–648.

— (2017). *The Probabilistic Foundations of Rational Learning*. Cambridge University Press.

Jaynes, Edwin T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Joyce, James M. (2005). "How Probabilities Reflect Evidence". In: *Philosophical Perspectives* 19, pp. 153–178.

— (2010). "A Defense of Imprecise Credence in Inference and Decision". In: *Philosophical Perspectives* 24, pp. 281–323.

Konek, Jason (forthcoming). "Epistemic conservativity and imprecise credence". In: *Philosophy and Phenomenlogical Research.*

Lassiter, Daniel (2020). "Representing credal imprecision: from sets of measures to hierarchical Bayesian models". In: *Philosophical studies* 177, pp. 1463–1485.

Levi, Isaac (1974). "On Indeterminate probabilities". In: *Journal of Philosophy* 71, pp. 391–418.

— (1980). *The Enterprise of Knowledge.* The MIT Press.

Lyon, Aidan (2017). "Vague Credence". In: *Synthese* 194.10, pp. 3931–3954.

Maher, Patrick (2010). "Explication of Inductive Probability". In: *Journal of Philosophical Logic* 39, pp. 593–616.

Meacham, Christopher J.G. (2014). "Impermissive Bayesianism". In: *Erkenntnis* 79, pp. 1185–1217.

Miranda, Enrique (2009). "Updating coherent previsions on finite spaces". In: *Fuzzy Sets and Systems* 160, pp. 1286–1307.

Morreau, Michael (2014). "Arrow's Theorem". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2014.

Moss, Sarah (forthcoming). "Global Constraints on Imprecise Credences: Solving Reflection Violations, Belief Inertia and Other Principles". In: *Philosophy and Phenomenological Research.*

Norton, John (2007). "Probability disassembled". In: *British Journal for the Philosophy of Science* 58, pp. 141–171.

— (2008). "Ignorance and Indifference". In: *Philosophy of Science* 75, pp. 45–68.

Paris, Jeffrey and Alena Vencovská (2015). *Pure Inductive Logic.* Cambridge University Press.

Pedersen, Arthur Paul and Gregory Wheeler (2014). "Demystifying Dilation". In: *Erkenntnis* 79, pp. 1305–1342.

Pettigrew, Richard (2016). *Accuracy and the Laws of Credence.* Oxford University Press.

— (2020). *Dutch Book Arguments.* Cambridge University Press.

Ramsey, F. P. (1926). "Truth and Probability". In: *The Foundations of Mathematics and other Logical Essays.* 1999 electronic edition. Routledge, pp. 156–198.

Redhead, M (1987). *Incompleteness, Nonlocality and Realism.* Clarendon Press.

Rinard, Susanna (2013). "Against Radical Credal Imprecision". In: *Thought* 2, pp. 157–165.

Seidenfeld, Teddy and Larry Wasserman (1993). "Dilation for sets of probabilities". In: *Annals of Statistics* 21, pp. 1139–1154.

Skyrms, Brian (2012). *From Zeno to Arbitrage.* Oxford University Press.

Smith, Cedric A.B. (1961). "Consistency in Statistical Inference and Decision". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 23, pp. 1–37.

Troffaes, Matthias and Gert de Cooman (2014). *Lower Previsions.* Wiley.

Vallinder, Aron (2018). "Imprecise Bayesianism and Global Belief Inertia". In: *The British Journal for the Philosophy of Science* 69.4, pp. 1205–1230.

van Fraassen, Bas (1989). *Laws and Symmetry.* Clarendon Press.

Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities.* Vol. 42. Monographs on Statistics and Applied Probability. Chapman and Hall.

Walley, Peter (1996). "Inferences from Multinomial Data: Learning about a Bag of Marbles". In: *Journal of the Royal Statistical Society B* 58.1, pp. 3–57.

Williamson, Jon (2010). *In Defense of Objective Bayesianism*. Oxford University Press.

Zabell, Sandy (2005). *Symmetry and its discontents*. Cambridge University Press.

Zaffalon, Marco and Enrique Miranda (2013). "Probability and Time". In: *Artificial Intelligence* 198, pp. 1–51.