New credences clarify self-location and assume less

Eissa Haydar

July 14, 2025

1 Introduction

There is a need for a new account of credence. Witnessing this are a number of problems formal epistemologists and, in particular, Bayesians face. The main problem I will focus on in this paper is the Sleeping Beauty problem. Secondary problems which I will subject to analysis are: the distinction between probability raising and justification (including the problem of old evidence); the principle of indifference; doomsday/simulation-style arguments; and, Newcomb's problem.

The alternative account of credences I will present capitalizes on the tools of comparative probability to formalize credences as a sort of belief, subject to logical analysis.

The structure of this paper is as follows. The first major portion is dedicated to set-up. After the introduction, section 2 will introduce Bayesianism and partial belief. Section 3 introduces the main detractors from Bayesianism, anti-Bayesians and anti-probabilists, distinguished by what it is they reject from the Bayesian thesis. Two anti-Bayesian frameworks, time-slicing and imaging, will be given special attention for their ability to match the accuracy of the account I present here (though in an ad-hoc way). Section 4 introduces comparative probability relations and their axioms.

Then, the paper will move toward presenting my account of credences, and its accompanying framework. Section 5 gives a formal definition of credences as a form of belief, as opposed to Bayesian partial belief. Section 6 presents a pattern of erroneous updating, say, an *updating fallacy*. It is this updating fallacy which illustrates exactly what goes on when partial beliefs are employed in cases like the Sleeping Beauty problem, and we will see precisely how they fail to be expressive enough. Section 7 presents a game-like semantics for my credences, yielding a nice way of capturing the notion of a generating process, and revealing a generalization of Bayesian confirmation. This generalization will suffice to state the Sleeping Beauty problem, which will be analyzed in sections 8 and 9. Finally, section 10 presents much briefer analyses of those secondary problems named earlier.

2 Basics

Bayesians take partial beliefs as their account of credences.[1] *Credences* are those doxastic states which summarize relative uncertainty about exclusive and exhaustive collections of propositions. A collection of propositions is *exclusive* if the truth of any of its members rules out the truth of the remainder; it is *exhaustive* if it is guaranteed that at least one of its members is true. A collection that has both properties is a *partition*.

Before saying what partial beliefs are, I must explain what I mean by "beliefs." I will take beliefs as a primitive. Rather than ask what they *are*, the goal is to subject agents' doxastic states to explicit anaylsis, to ask questions such as: are *these* beliefs consistent? or: how should we update *these* beliefs in light of *this* evidence? So, I am presupposing the sort of agent who can write, fixing some logical language, those formulas they would say they believe on a blackboard, this blackboard being their register, or memory, or corpus of knowledge. We will always be fixing some agent. In general, the motivation for the (perhaps formal) analysis of belief is belief-talk; that we communicate meaningfully using the concept of belief, and give reasons for our actions in terms of "beliefs," etc.

Partial beliefs, then, are those beliefs which are only partially in that register. For

[1] If you have never encountered the subject matter of this paper, (Lin 2024) is a good place to start. I will try my best to keep everything self-contained.

instance, thinking of the blackboard, one may imagine there being an opacity scale, where a proposition is more believed if it is written more opaquely and less believed if it is written less opaquely. I will first give more detail about the structure and dynamics of partial belief, then I will discuss the motivations.

For simplicity, we will allow the agent to fix some analytic truth which we will call *verum* and use the symbol $\top$ for, and similarly for some analytic falsehood, *falsum*, $\perp$. The agent will write verum on the register, and they will not write falsum on the register. This implies *non-triviality*; i.e. taking PB to be the function taking a proposition to its opacity ranking: $\mathrm{PB}(\top) \neq \mathrm{PB}(\perp)$. It is called "non-triviality" because we think of the propositions as being given partial belief rankings relative to one another, and a system which has only one ranking, which all propositions will have, is said to be *trivial*. By affirming that $\mathrm{PB}(\top) \neq \mathrm{PB}(\perp)$, we guarantee the existence of more than one probability ranking. Then, we stipulate that for any proposition A: $\mathrm{PB}(\perp) \leq \mathrm{PB}(A) \leq \mathrm{PB}(\top)$; that is, nothing can be less opaque than falsum (which will be invisible, as it is not written), and nothing can be more opaque than verum by convention, since nothing can be more believed than something which is true *simpliciter*. Next, we will stipulate that for any two mutually exclusive propositions A and B, $\mathrm{PB}(A \oplus B) = \mathrm{PB}(A) + \mathrm{PB}(B)$ where $\oplus$ represents the exclusive disjunction. We say that a function which has these just-stated properties satisfies the *axioms of probability*, and we call it a *probability function*.

Bayes' rule is the updating rule, sometimes seen as a rational norm, that partial beliefs update on new information as follows. For any partition, any partition elements inconsistent with the new datum are ruled out, and those which remain have their partial beliefs rescaled so as to sum to certainty (i.e. $PB(\top)$) while retaining the same relative ratios.

For instance, if an agent believes a six-sided die to be fair, the partial beliefs model cashes this out by saying that the agent assigns equal partial beliefs to the six unique propositions stipulating that the die lands a fixed side. The axioms of probability will guarantee that the partial belief assignments sum to certainty, since the sides of the die are together exclusive and exhaustive. Thinking of the agent's partial beliefs as partially in their register: they have all six of those propositions written on their blackboard, but only partially, each with the same opacity, such that their opacities combine to be as opaque as $\top$, which we also think of as written on the register. If the agent is then able to rule out 6 (say it is *publicly announced* that $\neg 6$, i.e. the agent learns in an idealized sense where they are guaranteed the truth of what is learned, and they *only* learn what is announced[2]), Bayes' rule stipulates that the agent's partial beliefs update on this announcement by taking away the partial belief which was assigned to 6, and reassigning it to the remaining sides of the die, so that the ratios of the partial belief assignments remain the same; i.e. the agent still assigns equal partial belief, but now only to 1 through 5 (thinking of the blackboard: 6 is erased, and 1-5 are each now darkened so as to be 1/5 as opaque as $\top$).

If, instead, the agent believed 6 twice as likely as any other side, but the rest equally likely as one other, this means the agent assigns equal partial belief to 1 through 5, twice as much to 6, and these partial beliefs add up to certainty. Then, if the agent rules out 1, its assigned partial belief is taken away and redistributed to 2 through 6 such that they now sum to certainty, while 2 through 5 are still assigned the same partial belief as one another, and 6 is still assigned twice as much.

Partial belief is generally seen as an alternative framework to belief, sometimes called *full* belief, and there has been a trend of attempting to give an account of belief, and belief update, as a special case of partial belief, and partial belief update ((Lin and Kelly 2012), (Mierzewski 2020)). I will argue that this is the opposite of the correct view; we should move away from partial belief, and toward an account of credence accommodated entirely within belief.

The motivations for partial belief lie in the desire to give an account of subjective probability, as much probability- or chance-talk can only be cashed out as referring to our own uncertainty as agents. Furthermore, it is not obvious how something like betting works using

[2]and *that* it was announced, but for simplicity's sake, I am assuming the announcement happens iff the announced does

only full beliefs, especially in cases where there is no obvious *thing* for the agent to hold a full belief about (e.g. some notion of chance), since the agent cannot be said to believe any of the uncertain outcomes. So, partial beliefs are thought to add the expressivity needed to deal with situations like this. It turns out, however, that partial beliefs are no more expressive than full belief, assuming we add some relation of comparative probability (e.g. "is more likely than") to that language in which the agent's beliefs are expressed.

Normally, this result has been taken to suggest that an agent's comparative probability beliefs *fix* their partial beliefs, and that it is enough to probe them to probe the agent's partial beliefs. But it will also turn out that partial belief, as a framework for credence and credence update, is actually strictly less expressive than the account of credence I will present herein, cashed out within full belief. And so, partial beliefs will sometimes update incorrectly where full belief update will succeed.

To surmount this expressivity gap, those who stick to partial belief, as the literature has done, must do away with updating altogether, or else Bayes' rule needs to be replaced by some alternative which encodes the correct updates into some oracle which the agent has access to when updating.

I will offer the perspective that comparative probability beliefs induce probabilistic models that the agent uses in describing their uncertainty. Sometimes the agent will update a model, keeping those beliefs which induced them otherwise fixed, and sometimes the agent will update those beliefs which induce their model; these two will not always agree, and the partial belief model is constrained to the latter form of update. Thus, instead of thinking of comparative probability beliefs as fixing an agent's partial beliefs, and then working within this less expressive framework, I will be taking exclusive disjunctions together with collections of comparative probability relations about their disjuncts *as credences*.

3 Alternatives to Bayesianism

Before getting into more detail, I will summarize the main positions which detract from Bayesianism. *Anti-Bayesians* agree with partial beliefs as a model of credence, but do not use Bayes' rule to update. I will focus on two different versions of anti-Bayesianism, and I will be able to explain their success in dealing with those problems which supposedly motivate the move away from Bayes' rule. But they will have missed the point, doing away with Bayes' rule when this was not at all necessary, failing to diagnose the issue as with partial belief.

Time-slicers are those anti-Bayesians who do away with dynamic rational norms altogether, thus losing the ability to make claims about how the agent *should* update ((Moss 2015), (Hedden 2015)). There is a sense in which the view I will present in this paper time-slices, only because normal belief is time-slicey. We sometimes retract our beliefs, or find ourselves experiencing a reality where our preferences differ enough, and we are perhaps no longer at all in contact with our previous beliefs so as to be rationally bound by them (e.g. memory loss, the influence of psychedelic drugs, etc). (Orthodox) Bayesians model agents as beginning with some universal set of possibilities together with a probability distribution over them, and then only updating via Bayes' rule. It is not hard to show (and I will later, in section 7) that there can be no such universal distribution. But, *at any moment* there is some probability distribution, and thus assignment of partial beliefs, which will agree with the account I give in this paper, and so time-slicers will be able to match the account I give. But, unlike time-slicers, I have formalized that reasoning which yields those solutions they think correct.

Imagers are those anti-Bayesians who do away with Bayes' rule, replacing it with an updating rule which also takes into account some notion of distances between different possible worlds. Of course, just as time-slicers begin at the partial belief assignments they want at each moment, imagers begin with an arbitrary assignment of distances between possible worlds which already has the updating information baked in, and thus imagers, too, will always be able to match the account I will present here.

It seems that anti-Bayesians, seeing the partial belief model fail to yield the correct updates, think of themselves as either needing rid themselves of updating altogether, or needing to create an alternative way to update that yields the intuitively correct solutions to those problems which arise. But, of course, by doing so, they do not *solve* problems; they presume their solutions to the problems correct, and attempt to support their framework, as opposed to orthodox Bayesianism, via the premise that theirs matches the correct solution to the problem, while Bayesianism does not. But they have not *shown* that this solution was correct, and they can only appeal to intuition and thought experiment. And so the literature has become messy.

Anti-probabilists see similar issues facing the Bayesian framework, and, unlike the anti-Bayesians who stop short of the actual issue, do away with the use of probabilities in representing the agent's uncertainty. This sort of anti-probabilist has gone too far, and the tools of comparative probability make this easy to see; indeed, probabilities are entirely expressible within the setting of belief, fixing a relation with the right axioms (and there are many). Many of the problems which concern this sort of anti-probabilist (e.g. non-measurable spaces and infinite lotteries (Norton 2021)) are not cases the sort of agent I am interested in will be entertaining in the relevant way.

Anti-probabilists may, instead, and as I am doing here, challenge the partial belief model itself. But they then rest contently without recovering the sort of reasoning which the partial belief model is capable of in whatever framework they prefer to use as a model of an agent's doxastic states. And a model that is useful is worth using. This sort of anti-probabilist must not merely argue that perhaps people do not have, or speak of, partial beliefs; but they must argue that we formal epistemologists can do without them as a modeling assumption, while still keeping the probabilistic dynamics which make them so attractive. Comparative probability has already yielded this result for full belief. So, this sort of anti-probabilist would do even better to show that their alternative *surpasses* partial beliefs in accuracy, explaining why. I will accomplish this within the remainder of this paper.

4 Comparative probabilities

A *comparative probability relation* is, first and foremost, a subjective relation.[3] I am giving an account of a kind of uncertainty, and must always fix an agent, considering *their* attitudes about certain propositions taken to be possible. We will denote a comparative probability relation for an agent a by $\leq_a$. What exactly this means depends on the comparative probability relation we choose. And whether an agent is justified in making such judgments is the problem of induction.

The most obvious candidate for a comparative probability relation is "is no more likely than," fixing some agent a, so $A \leq_a B$ would read as "a takes A to be no more likely than B." Of course, it remains to say what it means for a to take it so that one proposition is more or less "likely" than another without being circular, so it is better to start with something else first, and define "likelihood" out of it. Common examples of what tends to be employed for $A \leq_a B$ are such judgments as "a is willing to bet no more on A than B," and "a prefers A no more than B," etc. All that matters is that the relation we choose satisfies the right axioms. I will be thinking in terms of surprisal, reading $A \leq_a B$ as "a believes that B being true would be no less surprising to them than A being true" (reading it backwards for surprisal, so that $A \leq_a B$ means that a takes A to be no more likely than B, likelihood being the obvious inverse of surprisal).

Now we move to those axioms which a probability ordering relation must satisfy. Choose your favorite probability ordering relation and agent, and fix them. I will suppress denoting the agent. Firstly, $\top > \bot$.[4] This is *non-triviality*. This obviously holds for surprisal: something which is known to be true is far less surprising than something which is known to be false. $\bot$ turning out to be true would be something like "the most" surprising (and would

[3]See (Konek 2019) and (Fishburn 1986) for good introductions to these ideas.

[4]Where we define $A \approx_a B$, *equally likely*, as $A \leq_a B \wedge B \leq_a A$; and $<$, *strictly less likely*, as $A \leq_a B \wedge \neg(B \leq_a A)$; and $>$ and $\geq$, in the obvious way.

collapse the probability ordering, rendering it *trivial*; i.e. only one probability judgment would exist), while the truth of $\top$ is not surprising at all. Next, we assume that for any proposition A: $\perp \leq A \leq \top$. This holds for surprisal: there is nothing less surprising than the agent's favorite fixed analytic truth and there is nothing more surprising than some fixed analytic falsehood (should it turn out to be true).

Finally, we must assume a *finite-cancellation axiom*. Take some two collections of propositions $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ for some natural number n. We call them *isovalent* if they are such that whenever exactly $m \leq n$ of the X_i are true, then exactly m of the Y_i are true, and vice versa. The idea is that (perhaps the agent believes that) the X_i and Y_i always contain the same number of truths, this being expressible in propositional logic. Now, if $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are isovalent, we stipulate that whenever for all $i \leq n$, $X_i \leq Y_i$, i.e. the X are uniformly no more likely than the Y, then actually for all $i \leq n$, $X_i = Y_i$. That is, the two collections cannot contain the same number of truths and while one collection is strictly more likely than the other.

These three axioms ensure (indeed they are necessary and sufficient conditions) that there is a probability function P which *agrees* with the comparative probability assignments; i.e. $\mathrm{P}(A) < \mathrm{P}(B)$ iff $A < B$. And Bayesians take this P to be their PB (i.e. their assignment of partial beliefs).

5 New credences

Now we have seen what comparative probability orderings are, and that they give rise to probabilities without the use of partial belief. It is using these that I will give an account of credence as belief.

Firstly, consider what occurs when an agent believes an exclusive disjunction over some collection of propositions $A_1, \ldots, A_n$; denote this $\oplus_{i \leq n} A_i$, and we take this to be in their register. These are the sort of situation credences are concerned with, where we have a space of propositions-judged-possible collectively forming a partition. Fix some comparative probability relation $\leq$. Imagine $n = 3$; i.e. we are concerned with a three-element partition. Imagine furthermore that the agent has not yet judged the relative probabilities of A_1, A_2, and A_3. We immediately have everything contained in the following diagram (see Figure 1), where an arrow between two vertices A and B represents $A \leq B$. We imagine it closed under *transitivity*, it being entailed by the axioms for $\leq$; i.e. if there is an arrow between A and B and between B and C, we imagine there implicitly being an arrow between A and C.

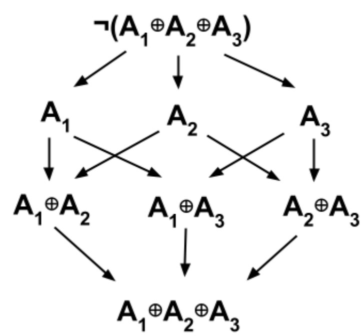

Figure 1: The model induced by the belief that $A_1 \oplus A_2 \oplus A_3$

We will call these *diagrammatic models*, or frequently just *models* when it is clear what is under discussion, and we say that beliefs which entail all of the content of a model *induce* that model. This model, in Figure 1, is induced merely by the belief that $A_1 \oplus A_2 \oplus A_3$. Note that, here, $\oplus_{i \leq 3} A_i$ is taken to be as likely as $\top$ (as the agent believes both) and $\neg(\oplus_{i \leq 3} A_i)$ is thus taken to be as likely as $\bot$.

The fact that, for instance $A_1 \leq A_1 \oplus A_2$ and $A_2 \leq A_1 \oplus A_2$, can be shown formally using *additivity* (that for some A and B which are each mutually exclusive with some C, $A \leq B$ iff $A \oplus C \leq B \oplus C$), which the axioms entail. But an intuitive way of thinking about it is that A_1 (and similarly for A_2) entails $A_1 \oplus A_2$;[5] thus we should expect that $A_1 \leq A_1 \oplus A_2$, or that $A_1 \oplus A_2$ is no less surprising to the agent than A_1, which entails it, and thus, to

[5]Because A_1 entails $A_1 \vee A_2$, and we know them to be exclusive.

the agent, A_1 is no more likely than $A_1 \oplus A_2$.[6] The point is that, given only a comparative probability relation, an agent's belief in an exclusive disjunction immediately gives rise to some structure.

I will be taking a credence to be a belief in an exclusive disjunction, together with probability ordering beliefs about its disjuncts. But I will not yet give an exact formal definition of a credence, as we will be using the term within the scope of this paper. Because, to make more fine-grained probability judgments, we need more events. For instance, looking again at Figure 1, if we drew arrows in both directions between A_1 and A_2 and between A_2 and A_3, we could stipulate a uniform distribution over them. If, instead, we drew an arrow in both directions between A_1 and $A_2 \oplus A_3$, and then also an arrow in both directions between A_2 and A_3, we could stipulate that A_1 is twice as likely as A_2 and A_3, which are equally likely, thus assigning probabilities $1/2, 1/4$, and $1/4$, respectively, taking $\mathrm{P}(\bot) = 0$ and $\mathrm{P}(\top) = 1$, as is conventional. But we cannot, for instance, stipulate that A_1 has probability $1/2$ and A_2 has probability $1/3$. In order to circumvent this issue, we will allow the agent to entertain *hypothetical fair dice rolls.*

Similarly to our choice of $\top$ and $\bot$ as arbitrary propositions with a fixed truth value, we will allow the agent to consider arbitrary propositions with a fixed comparative probability value. It does not matter what these propositions are chosen to be; imagine an agent wants to assign probabilities to A_1, A_2, and A_3, as above. Fixing some hypothetical die roll (which I will define next), our agent will not be entertaining hypotheses such as: the die lands some certain side *and* A_2. This die is purely hypothetical and the agent is *only* considering propositions about it in order to make comparative probability judgments.

One way of logically cashing out the notion of hypothetical fair dice rolls would be to have the agent literally believe hypothetical judgments. For instance, if the agent who believes $A_1 \oplus A_2 \oplus A_3$ wants to judge A_2 to have probability $1/3$, we can allow them to make such a judgment as: *if* there were some fair three-sided die, then A_2 would be as likely as one of those

[6] For instance, someone who says that they think that if it rains tomorrow, then pigs will fly, but also that they would be strictly more surprised if pigs fly than if it rains tomorrow, is being inconsistent.

sides. Logically: $\forall X_1, X_2, X_3(((X_1 \oplus X_2 \oplus X_3) \wedge (X_1 \approx X_2 \approx X_3)) \rightarrow (A_3 \approx X_1))$, where of course it does not matter which side of the die is used to assign the probability of A_3. We may also allow the agent to simply stipulate that there is some fair three-sided die, and the right probability judgments hold; e.g. $\exists X_1, X_2, X_3((X_1 \oplus X_2 \oplus X_3) \wedge (X_1 \approx X_2 \approx X_3) \wedge (A_2 \approx X_1))$. Such an existential claim is not too strong, especially for reasonably sized n. For instance, fixing some n, the agent probably believes that the number of stars in their observable universe modulo n is as likely to be each among $0, 1, \ldots, n-1$, and exactly one of these is true.

In either case, the point is that the agent is entertaining these hypothetical propositions to make more precise probability judgments than possible using $\leq$ with only whatever number of events they are actually entertaining, and this is all done without partial belief. Let us consider now what the above judgment, assigning probability 1/3 to A_2 among A_1 and A_3, looks like diagrammatically (see Figure 2).

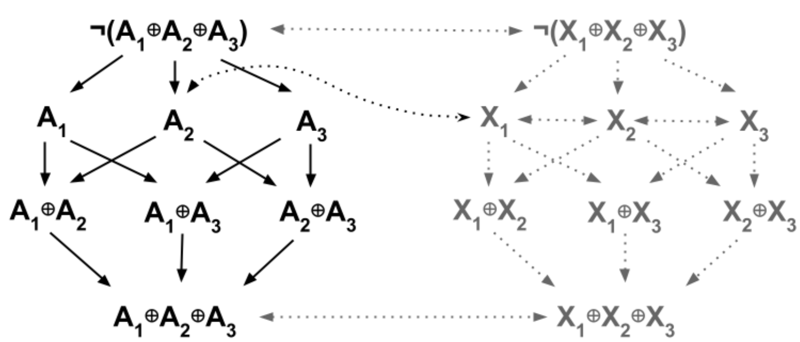

Figure 2: Assignment of probability 1/3 to A_2, with A_1, A_2, and A_3 forming a partition

Recall that the die roll is hypothetical (hence it being faded in the diagram), so we are only considering those events built out of the A_i, and those built out of the X_i (i.e. the hypothetical die's sides), the agent only relating the two of these collections via comparative probability judgments. We can also assign A_1 probability 1/2, using the same method (see Figure 3).

Now, for instance, we will expect that A_3 has probability 1/6, and indeed the axioms

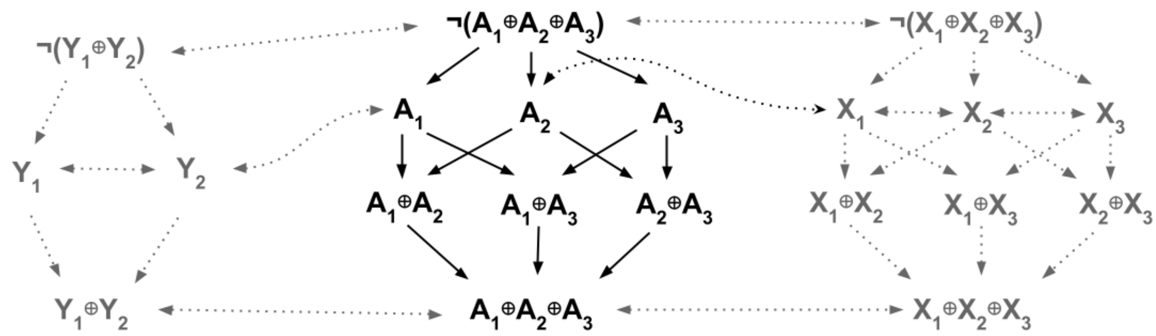

Figure 3: Assignment of A_1 probability 1/2 and A_2 probability 1/3, with A_1, A_2, and A_3 forming a partition

allows us to conclude this, should we add in a diagram for a hypothetical fair six-sided die.

The diagram in Figure 3 has two different hypothetical dice; we can think of this as the agent having two meter sticks to measure probabilities with. Of course, we can always consolidate to have one single meter stick, if we measure them both by some single metric (them both being measures of the same thing). In general, the nm-sided die is precise enough to exactly measure probability $1/n$ events and probability $1/m$ events. This is because the diagram of the nm-sided die contains both the n-sided die and the m-sided die as sub-dice.

So, in the example above, we could consolidate by using a hypothetical fair six-sided die, it containing both a fair two-sided die and a fair three-side die within. Figure 4 illustrates this by choosing two such subgraphs, with $X_1 \oplus X_2 \oplus X_3$ and $X_4 \oplus X_5 \oplus X_6$ (note that they form a partition) standing in for the sides of the fair two-sided die, and similarly for $X_1 \oplus X_2$, $X_3 \oplus X_4$, and $X_5 \oplus X_6$ for the sides of the fair three-sided die.

And Figure 5 shows $A_1 \oplus A_2 \oplus A_3$ with the A_1 having probability 1/2 and A_2 having probability 1/3, using one hypothetical six-sided fair die.

Already this diagrammatic semantics has become somewhat intractable in size, thus the abbreviation of the hypothetical fair six-sided die in both of the figures it features in. But, it is illustrative.

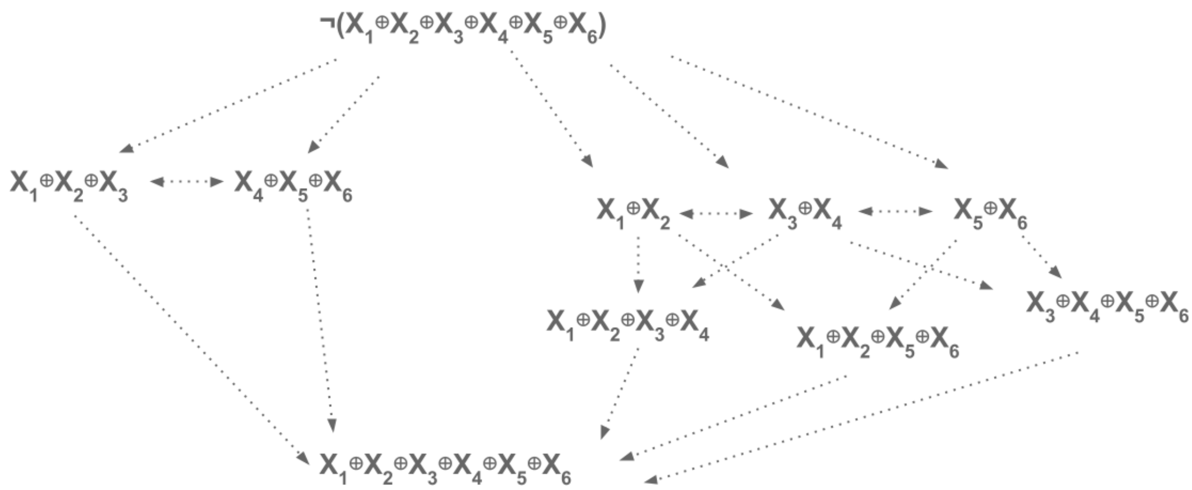

Figure 4: Hypothetical fair two- and three sided-dice within the hypothetical fair six-sided die

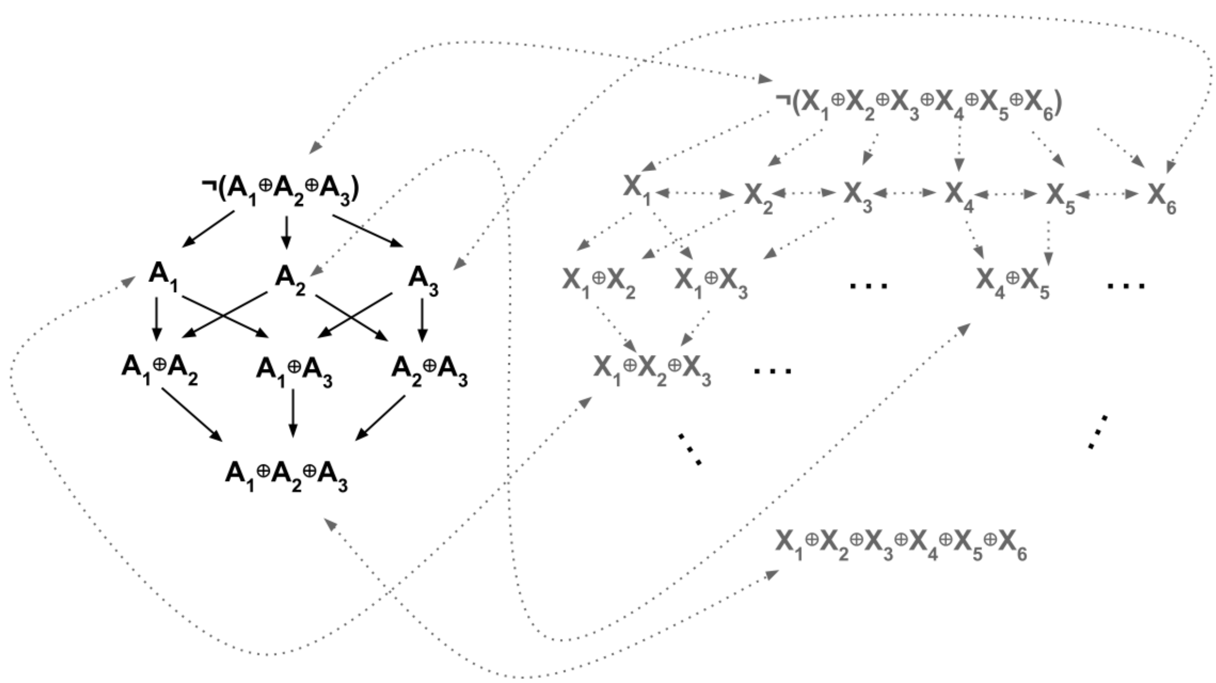

Figure 5: Canonical model for $A_1 \oplus A_2 \oplus A_3$ with A_1 assigned probability 1/2 and A_2 assigned probability 1/3.

Notice, also, that in Figure 5, no side of the die was used to represent the probability of two different A_i. A_1 is defined using X_1, X_2, and X_3, A_2 using X_4 and X_5, and A_3 using X_6.

This was intentional; we will be associating each side of the hypothetical die with exactly one of the events we are assigning probability to. The idea is to see each event we are assigning probability to as an equivalence class of hypothetical die sides associated with it. This will facilitate simple updating, as I will explain shortly. Call a model which contains (i) some $A_1, \dots, A_n$ which the agent believes to be in exclusive disjunction and (ii) exactly one hypothetical fair die roll with (iii) probability judgments assigned such that each side of the hypothetical die is associated with exactly one of the A_i a *canonical model*.

In this paper, I am defining a *credence* to be any belief, or collection of beliefs, which induces a canonical model. This is obviously not the most general definition. A more general account would take any beliefs which induce a diagrammatic model to be a credence; i.e. any exclusive disjunction, together with whatever (if any) probability judgments the agent has about the propositions, will be a more general credence. But the goal of this paper is to use this account, rather than flesh it out in the utmost detail.

With regard to credence update, ruling out occurs as follows. Suppose that an agent has some credence about $A_1, \dots, A_n$ (assumed, remember, to induce a canonical model), and it is publicly announced that A_n is false. Then the agent removes from their model A_n and all of those sides of the hypothetical die associated with it, keeping all remaining probability judgments the same. I will be calling this *model update*, to distinguish it from a more general credence update, which I will call *belief update*. The next section will present a simplified example of this distinction. It is known, and obvious, that this process of model update yields Bayes' rule.[7]

Consider the A_1, A_2, and A_3 from earlier, where A_1 has probability 1/2 and A_2 has probability 1/3. Suppose the agent learns that A_2 is false, and updates (see Figures 6 and 7). Notice that this assigns A_1 probability 3/4, and A_3 probability 1/4, thus giving them the ratio 3:1, yielding the same result as Bayes' rule: A_1 initially had probability 1/2, and

[7]Note also that I am presupposing that the agent will never assign PB($\perp$) to any proposition they think may yet be true, and thus I avoid the need to specify an updating rule in this case; the occurrence of an event so surprising ought to make the agent rethink their whole model!

A_3 1/6, the ratio between the two of these being 3:1 as well.

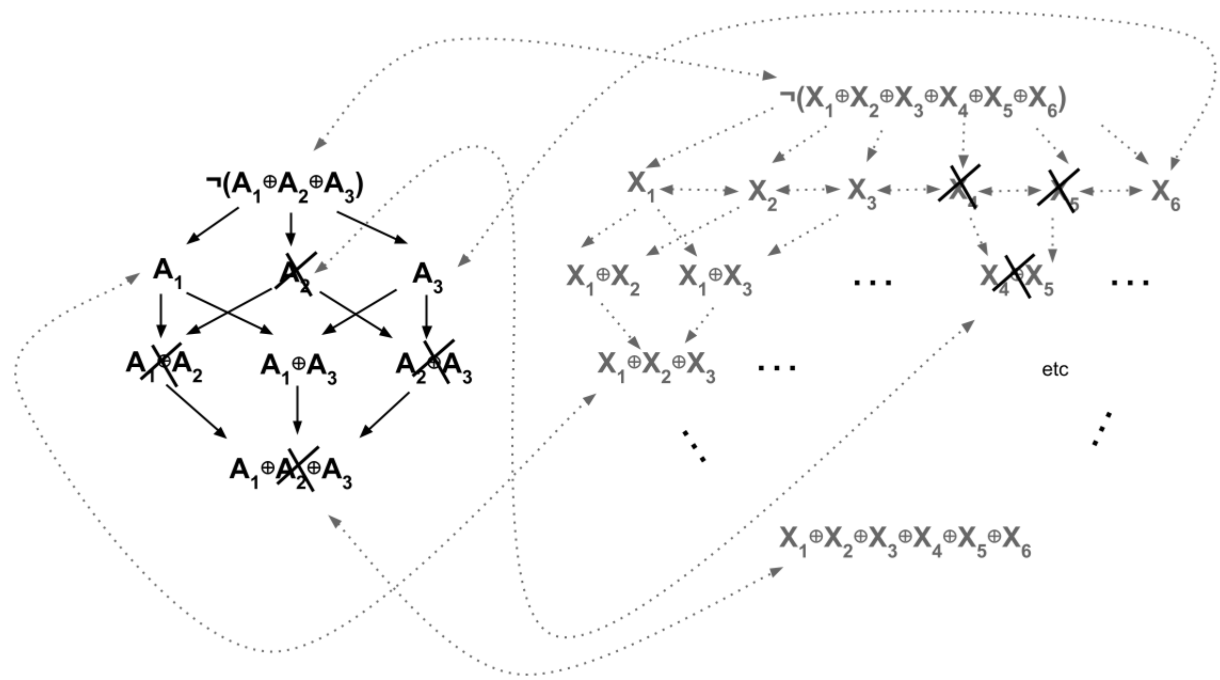

Figure 6: the update involves removing the ruled out propositions (i.e. A_2) and their associated die rolls (i.e. X_4 and X_5), and keeping everything else fixed

In general, I will denote the credence that assigns (rational, admissible) probabilities p_i to propositions in exclusive disjunction A_i by $\{A_1^{p_1}, \ldots, A_n^{p_n}\}$; we will always be assuming there are only finitely many. We do this for simplicity, but also because I am defining credences in terms of surprisal, and think this is only well-defined for agents who believe all of the outcomes under consideration may occur, and that they could come to know which. This does not mean that there cannot be interesting fruit to considering an infinitary version of this logic.

Note that this and the last section allow us to conclude that we can represent everything Bayesians can — i.e. probability functions over finite partitions, constraining ourselves to rational assignments of probabilities — without the use of partial beliefs. In section 7.2, I will give an explicit example of Bayesian confirmation between chance hypotheses done in this setting.

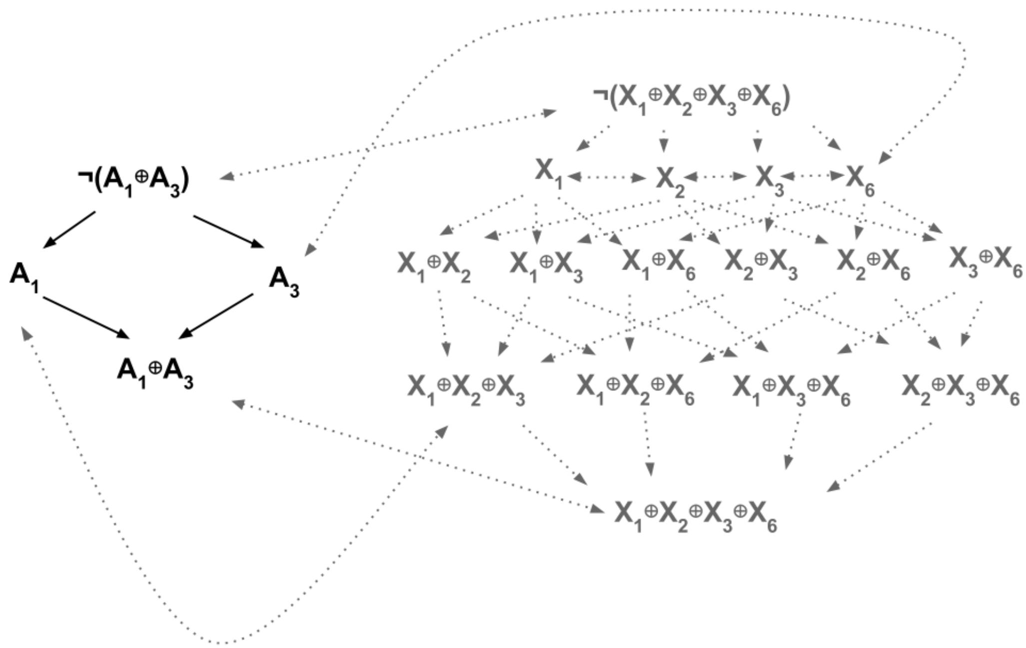

Figure 7: the model after updating

Importantly, the partial belief model constrains agents to have one single probability distribution, this being their *prior* (thinking of the register with opacity rankings: *every* proposition the agent is considering is written with some opacity, and then the only update that can occur is by erasing and redistributing, perhaps via Bayes' rule). Intuitively this constraint arises due to the fact that one can always pass to some finest partition and ask about the agent's partial belief assignments, or betting strategies, etc., *there*.[8] And so, in my terminology, they are constrained to a single credence, or a single model. But since model update and belief update will not always coincide, and they have no way of representing more general belief update, having only the model and having forgotten those beliefs which induced it, they will not always update accurately. I will get to explicit examples shortly,

[8]What I mean by this is passing to the product space. For instance: if the agent is uncertain over propositions about the sides of a die, 1 through 6, and also about some coin landing heads or tails, we can ask them about all of the pairs, e.g. *heads and 6*. And if they have partial beliefs over partitions, they will have partial beliefs on the each of these pairs, too, and the agent's assignments here will entail the rest of their relative partial belief assignments. The Bayesian ideal, then, is to accept this constraint but have some *universal* prior which encodes all of the relevant updating information only in terms of relative ratios (i.e. Bayes' rule). But this is not possible, as is shown in section 7.

but I will first explain this phenomenon in a simple setting.

6 An updating fallacy

Consider a situation in which some agent adds two propositions A and B to their register, due to some evidence; i.e. they believe A and B. Imagine also that A and B together entail some C. So, the agent believes C, now, too.

Now imagine the agent undergoes some rational update, having learned some new information. Denote this by saying the update takes A to A', B to B', and C to C'. This will be an instance of *model update*. But, suppose that A' and B' do not entail C', and instead entail some D which contradicts it.

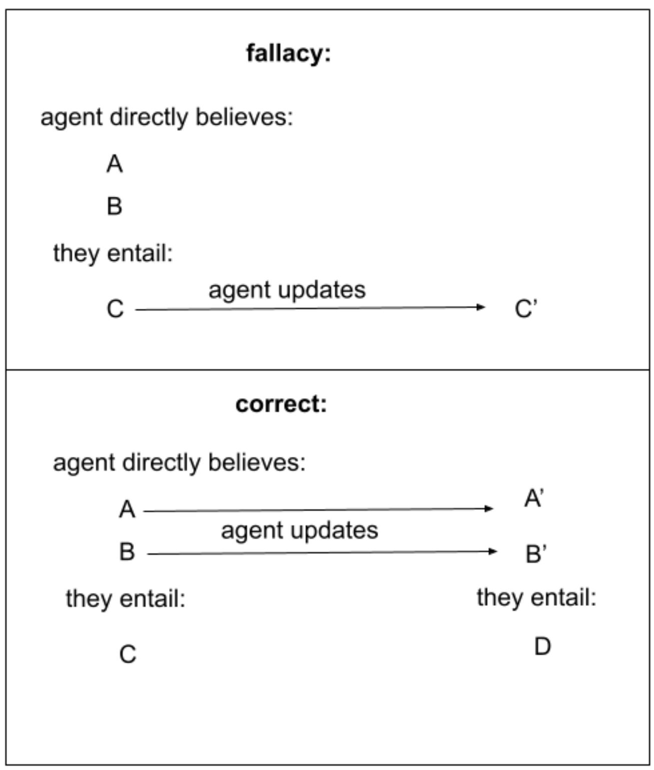

Figure 8: The fallacy I am accusing Bayesians of being susceptible to

The agent only believed C because they believed A and B, which they take to be evidenced. Now believing A' and B' to be evidenced, them being the rational update of A and B, they should consider themself as no longer having grounds for belief of C/C'. They should believe D, it being entailed by what they take to be evidenced. The idea here is that if the agent passes to C, it being entailed by their beliefs, and then they forget how they got there, they will think they have grounds to believe C' after updating, even though they do not.

Of course, I have not shown yet that it is possible to have some such A, B, C, and D, but that is the target of the next section. The point will be that partial belief necessitates the pass to a single model, as illustrated in the last section. Instead, we must also have some capacity for updating those beliefs which induce models, if the need should arise, rather than just naively updating a model, *itself not directly evidenced*, forgetting how we got to it.

7 Generating processes

In this section, I will give a semantics for my credences as describing *generating processes*, and this perspective yields, not only a natural way of representing the Sleeping Beauty problem, which I will analyze in the next section, but also, a more general theory of confirmation than Bayesianism, being able to represent a larger class of hypotheses than Bayesian priors. It is here where it will be made clear that there can be no universal prior. But I must first show that the system I have presented here can represent, not only partial beliefs, but Bayesian confirmation (which I have yet to define formally).

7.1 Urn semantics

Suppose we have some agent a who believes (i.e. it is in their register) some credence $\{A_1^{p_1}, \ldots, A_n^{p_n}\}$. Consider, now, the hypothetical agent b who is playing the following game: There is a trustworthy interlocutor who guarantees b that they are going to draw a ball from an opaque urn (which b can see) which contains balls labeled A_1 through A_n, each drawn, *for* *whatever reason*, with probabilities p_1 through p_n, respectively. Since we are working with rational probabilities only, by the assumption that our credences are as nice as stipulated earlier, we could give a mechanistic explanation.[9] But this does not matter. One does not ask how a Turing machine moves. Now, it is clear that agents a and b believe all of the same probability judgments about their respective collections of A_1 through A_n. We will take these urn games as a semantics for credences, as this will be illustrative.

Note, however, that this is an imperfect semantics; it requires one important clarification. Consider the agent who believes $\{\text{Odd}^{0.5}, \text{Even}^{0.5}\}$, meaning that they take it to be the case that some die is equally likely to land on an odd or even value. Say that agent also believes $\{\text{Prime}^{0.5}, \neg\text{Prime}^{0.5}\}$, meaning that they take it to be the case that the same die's outcome is as likely as not to be prime. We are modeling this as the agent being faced with two urns; one which draws balls labeled Even and Odd, each with probability 1/2, and one which is the same for Prime and ¬Prime. This suggests that these two processes are separate, and perhaps thus probabilistically independent. We might imagine drawing from both urns, and considering the combined result, multiplying the probabilities, but this is not correct. We will presume the urns *must* be drawn from at the same time, and we must entertain any magicky connexions between these two urns separated in space which may yet allow for counterfactual dependencies.

To see why, consider the agent who believes that a fair four-sided die is being rolled; i.e. they believe $\{1^{1/4}, 2^{1/4}, 3^{1/4}, 4^{1/4}\}$, and that Odd $\longleftrightarrow (1 \vee 3)$ and Prime $\longleftrightarrow (2 \vee 3)$, where $\longleftrightarrow$ denotes logical equivalence. This agent believes $\{\text{Odd}^{0.5}, \text{Even}^{0.5}\}$ and $\{\text{Prime}^{0.5}, \neg\text{Prime}^{0.5}\}$, as these are entailed by their beliefs about the die, and, in this case, Odd and Prime are probabilistically independent. If the agent instead believed that the die was fair and six-sided, i.e. $\{1^{1/6}, 2^{1/6}, 3^{1/6}, 4^{1/6}, 5^{1/6}, 6^{1/6}\}$, Odd $\longleftrightarrow (1 \vee 3 \vee 5)$, Prime $\longleftrightarrow (2 \vee 3 \vee 5)$, then they would still believe $\{\text{Odd}^{0.5}, \text{Even}^{0.5}\}$ and $\{\text{Prime}^{0.5}, \neg\text{Prime}^{0.5}\}$, but Odd and Prime will

[9] E.g. the urn actually being populated by many balls, each of the same size, with the ratio of A_i balls to total balls being $p_i / \sum_{i \leq n} p_n$

not be probabilistically independent.[10]

Thus, we must always imagine the urns representing credences as only conferring probabilistic relations to their elements (i.e. the labeled balls within), relative to one another, and we do not assume independence of different urns, this being logically stronger than the belief in the two credences themselves, which we are still modeling as two urns being drawn from, though leaving the possibility of some distinct third urn which refines on the two, without contradicting them.

7.2 Matching Bayesian confirmation

Now comes what I see as the main contribution of this paper. We are taking this urn to be some kind of black-box whose draws, for whatever reason, yield certain results with certain probabilities. I want to consider the perspective where the urn is thought to *generate* the balls. And so I want to consider these urns not merely as a descriptive model of credences, but also as a way to describe the content of chance hypotheses. So, imagine we had such a generating urn, but we were uncertain about the true "chances" of generation. This is the sort of situation Bayesian confirmation is concerned with. This situation is analogous to the uncertainty between two scientific theories which make different probabilistic predictions about some particle.

What I am calling *Bayesian confirmation* is statistical confirmation within some collection of chance-hypotheses. This is a dynamic thing, and *confirmation* refers to one chance-hypothesis having its probability raised in light of some evidence. So imagine the agent is playing the urn game, but they are told that the urn is either really $\{H^{2/3}, T^{1/3}\}$ or really $\{H^{1/3}, T^{2/3}\}$. We can think of these as two different theories for describing the urn; and so if the urn is really just an abstract descriptive device for discussing the bias of a coin (thus the use of H, T), the agent is perhaps told to entertain two theories about the bias and to perform confirmation among them as they see flips. It is very important to highlight

[10] Thanks to [redacted] for this example

that the agent begins with their hypotheses, and then performs confirmation over them. For simplicity, I will assume that the agent is told that a different, definitely fair coin is flipped, and this determines the real bias of the urn/coin under scrutiny (an alternative is that perhaps the agent, knowing themselves to want to perform confirmation on these two hypotheses while not yet being able to do better than random, begins by assigning them equal probability).

So, the agent believes $\{\{H^{2/3}, T^{1/3}\}^{1/2}, \{\{H^{1/3}, T^{2/3}\}^{1/2}\}$. Immediately we are faced with some ambiguity. The agent who believes this credence might instead believe that the (thinking about the urns) generating process is: one of those two sub-processes is chosen (imagine that there are urns within the urn), 50-50, and this occurs on every draw. This is distinct from the agent who believes that only one of those two sub-credences is *actually* describing the urn's probabilistic content, and they think they each has a fifty percent chance of being the true one. But the description $\{\{H^{2/3}, T^{1/3}\}^{1/2}, \{\{H^{1/3}, T^{2/3}\}^{1/2}\}$ does not distinguish between them.

Distinguishing between these two syntactically necessitates encoding what happens when a draw takes place, as this is what differentiates them as processes; i.e. we ask: what hypotheses does the agent entertain about the *next* draw? Firstly, we will distinguish the processes, for ourselves, as follows: call by A the process where the urn actually is $\{H^{2/3}, T^{1/3}\}$, by B the process where the urn is $\{H^{1/3}, T^{2/3}\}$, and by C $\{\{H^{2/3}, T^{1/2}\}^{1/3}, \{\{H^{1/3}, T^{2/3}\}^{1/2}\}$. We will denote which process generated the proposition by a subscript, so we have one agent who believes that $\{\{H_A^{2/3}, T_A^{1/3}\}^{1/2}, \{\{H_B^{1/3}, T_B^{2/3}\}^{1/2}\}$, and another who believes that $\{\{H_{C_A}^{2/3}, T_{C_A}^{1/3}\}^{1/2}, \{\{H_{C_B}^{1/3}, T_{C_B}^{2/3}\}^{1/2}\}$ (distinguishing between the sub-processes as well; it will not matter).

Figure 9 is what the first agent's credence, $\{\{H_A^{2/3}, T_A^{1/3}\}^{1/2}, \{\{H_B^{1/3}, T_B^{2/3}\}^{1/2}\}$, looks like diagrammatically.

Note that moving such a nested model to a canonical model (i.e. having only one die roll; the diagram in Figure 9 has 3, one of which, X, is rolled with probability 1, and the

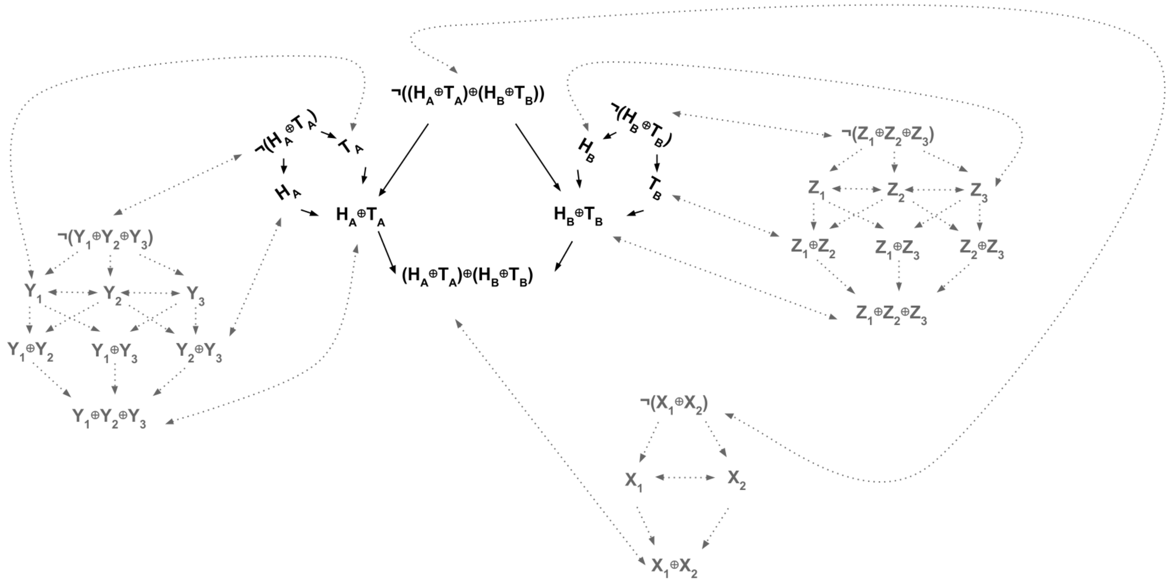

Figure 9: Diagrammatic model of $\{\{H_A^{2/3}, T_A^{1/3}\}^{1/2}, \{\{H_B^{1/3}, T_B^{2/3}\}^{1/2}\}$

other two of which are rolled with probability 1/2) is represented syntactically by collapsing the probability spaces, and indeed the collapsed space will be deductively entailed under the definition of credence I have provided. This agent, for instance, who believes that $\{\{H_A^{2/3}, T_A^{1/3}\}^{1/2}, \{\{H_B^{1/3}, T_B^{2/3}\}^{1/2}\}$, will definitely believe that $H_A \oplus T_A \oplus H_B \oplus T_B$, and all of the relevant probability judgments will be entailed. So, the agent can deduce that $\{H_A^{1/3}, T_A^{1/6}, H_B^{1/6}, T_B^{1/3}\}$.

Now, before having a ball be drawn and shown to the agent, let us consider what each agent will believe about the next draw. In particular, the agent who believes that the process really is A or B will believe that, subscripts now not only to denote the generating process but also all previous all previous draws: $H_A \longleftrightarrow \{H_{A,H_A}^{2/3}, T_{A,H_A}^{1/3}\}$, $T_A \longleftrightarrow \{H_{A,T_A}^{2/3}, T_{A,T_A}^{1/3}\}$ (the A process will continue, if it is the true one); also $H_B \longleftrightarrow \{H_{B,H_B}^{1/3}, T_{B,H_B}^{2/3}\}$, $T_B \longleftrightarrow \{H_{B,T_B}^{1/3}, T_{B,T_B}^{2/3}\}$ (the B process will continue, if it is the true one).

On the other hand, the agent who believes that the process is really C, believes that

$H_{C_A} \longleftrightarrow \{\{H^{2/3}_{C_A,H_{C_A}}, T^{1/3}_{C_A,H_{C_A}}\}^{1/2}, \{\{H^{1/3}_{C_B,H_{C_A}}, T^{2/3}_{C_B,H_{C_A}}\}^{1/2}\}$, and similarly replacing H_{C_A} with T_{C_A}, H_{C_B}, and T_{C_B}, because the process continues no matter the outcome.

I will go now through the process of Bayesian confirmation. We will imagine that the first draw from the urn is an H; i.e. that the agents learn $H_A \vee H_B$ and $H_{C_A} \vee H_{C_B}$, respectively. Then we will ask each agent about their assigned probability of A being the process which generates the next draw, having seen this H. Note firstly that Bayesian confirmation would yield that the agent who holds that the process is C will stay at 1/2 for C_A, and the agent who had A or B at 50-50 will move to $\mathrm{PB}(A|H) = \mathrm{PB}(H|A)\mathrm{PB}(A)/\mathrm{PB}(H) = (2/3)(1/2)/((1/2)(2/3) + (1/2)(1/3)) = 2/3$; i.e. A is now believed to be twice as likely as B by Bayes' rule.[11] We will match this.

I will begin with the more complicated case of that agent who believes that the process may be either A or B. Recall that they believe:

$$\{\{H^{2/3}_A, T^{1/3}_A\}^{1/2}, \{\{H^{1/3}_B, T^{2/3}_B\}^{1/2}\}$$

$$H_A \longleftrightarrow \{H^{2/3}_{A,H_A}, T^{1/3}_{A,H_A}\}$$

$$T_A \longleftrightarrow \{H^{2/3}_{A,T_A}, T^{1/3}_{A,T_A}\}$$

$$H_B \longleftrightarrow \{H^{1/3}_{B,H_B}, T^{2/3}_{B,H_B}\}$$

$$T_B \longleftrightarrow \{H^{1/3}_{B,T_B}, T^{2/3}_{B,T_B}\}$$

As above, we can deduce $\{H^{1/3}_A, T^{1/6}_A, H^{1/6}_B, T^{1/3}_B\}$ from the first line. H being drawn is equivalent to ruling out T_A and T_B, so we update that to $\{H^{2/3}_A, H^{1/3}_B\}$. Substituting logically equivalent propositions, we get:

$$\{\{H^{2/3}_{A,H_A}, T^{1/3}_{A,H_A}\}^{2/3}, \{H^{1/3}_{B,H_B}, T^{2/3}_{B,H_B}\}^{1/3}\}$$

[11] I am using the notation PB(A|B) to represent the agent's partial belief that A after updating on the information that B, as is standard for denoting Bayes' rule. An expert will note that there are other possible versions of Bayesian conditionalization we could have studied, some of them synchronic, or defined in terms of the indicator variables, but I am, as is hopefully obvious, only considering conditionalization in the sense of temporal conditionalization.

Of course, we have now arrived at the same conclusion as Bayesian confirmation; there now is a $2/3$ probability that the process is A.

As for the case where the agent believes the process is C, recall that they believe:

$$\begin{aligned}
&\{\{H_{C_A}^{2/3}, T_{C_A}^{1/3}\}^{1/2}, \{\{H_{C_B}^{1/3}, T_{C_B}^{2/3}\}^{1/2}\} \\
&H_{C_A} \longleftrightarrow \{\{H_{C_A,H_{C_A}}^{2/3}, T_{C_A,H_{C_A}}^{1/3}\}^{1/2}, \{\{H_{C_B,H_{C_A}}^{1/3}, T_{C_B,H_{C_A}}^{2/3}\}^{1/2}\} \\
&T_{C_A} \longleftrightarrow \{\{H_{C_A,T_{C_A}}^{2/3}, T_{C_A,T_{C_A}}^{1/3}\}^{1/2}, \{\{H_{C_B,T_{C_A}}^{1/3}, T_{C_B,T_{C_A}}^{2/3}\}^{1/2}\} \\
&H_{C_B} \longleftrightarrow \{\{H_{C_A,H_{C_B}}^{2/3}, T_{C_A,H_{C_B}}^{1/3}\}^{1/2}, \{\{H_{C_B,H_{C_B}}^{1/3}, T_{C_B,H_{C_B}}^{2/3}\}^{1/2}\} \\
&T_{C_B} \longleftrightarrow \{\{H_{C_A,T_{C_B}}^{2/3}, T_{C_A,T_{C_B}}^{1/3}\}^{1/2}, \{\{H_{C_B,T_{C_B}}^{1/3}, T_{C_B,T_{C_B}}^{2/3}\}^{1/2}\}
\end{aligned}$$

And, reasoning like above, we may deduce from the first line that $\{H_{C_A}^{1/3}, T_{C_A}^{1/6}, H_{C_B}^{1/6}, T_{C_B}^{1/3}\}$. We may similarly simplify the right hand sides of the above equivalences. Now, we may use the fact that H is drawn first to update it to $\{H_{C_A}^{2/3}, H_{C_B}^{1/3}\}$, and substituting in the simplified equivalences yields:

$$\{\{H_{C_A,H_{C_A}}^{1/3}, T_{C_A,H_{C_A}}^{1/6}, H_{C_B,H_{C_A}}^{1/6}, T_{C_B,H_{C_A}}^{1/3}\}^{2/3}, \{H_{C_A,H_{C_B}}^{1/3}, T_{C_A,H_{C_B}}^{1/6}, H_{C_B,H_{C_B}}^{1/6}, T_{C_B,H_{C_B}}^{1/3}\}^{1/3}\}\}$$

Now it is clear that the agent believes that there is a $2/3$ chance that there is a $1/2$ chance the next process is A (i.e. C_A), and there is a $1/3$ chance also that there is a $1/2$ chance. So this agent believes there is $1/2$ chance the next draw is generated by process A, and so disagrees after updating with that agent who thought the process either A or B.

Having demonstrated that Bayesian confirmation is representable using belief, under the perspective that there are processes being described, I will move to generalizing.

7.3 More general generating processes

We have thus far considered hypotheses which assign some probability to each member of a partition. But, there is another obvious generating process, one which is deterministic. We may imagine that some urn is perhaps guaranteed to draw A first, and B second; or, for instance it draws A or B with some probability. How are we to compare these hypotheses? But I will begin with some more motivation.

The Bayesians, being constrained to some single credence, in my sense, or prior probability function, are likewise constrained to some single urn. And they will fail to represent the following: *I did not know that this was what was happening, but now that I do know, I think it would have happened in any case.*

Imagine the following case. An agent knows that one of the two following bit sequences is being revealed to them, one bit at a time: 000, or 0001. Assume the agent has reason to think them equally likely. We may think of the agent as uncertain over two generating processes, though ones with deterministic structure. There is also the potential for a new dimension to the agent's uncertainty: uncertainty about *which* bit of the sequence is being revealed, or which draw is taking place, them not all being revealed at once. So, take that agent and imagine they space out, so that they are not certain which bit is currently being revealed, or whether any bit has been revealed at all. If they were to hear a 0, they would think that this would have happened anyways. If they hear a 0 and then nothing else, or if they hear a 1, they can distinguish between the two processes.

But, consider how the Bayesian represents the dual-uncertainty over both time and process. They pass to some single probability space, finer than both; call the 000 process A and the 0001 process B, then the Bayesian considers the set:

$$\{(A, 0_1), (A, 0_2), (A, 0_3), (B, 0_1), (B, 0_2), (B, 0_3), (B, 1)\}$$

and assigns *it* probabilities. Thus, *whenever* 0 *occurs, there was also some chance* 1 *would have occurred instead*, them all being in the same urn drawn from with chance, and so the Bayesian agent cannot update thinking: I did not know this was going to happen and now think it would have anyways. For, if the steps of the A and B each has collectively probability

$1/2$ and also $(B, 1)$ has non-zero probability, then it will be the case that $\mathrm{PB}((A, 0_1) \oplus (A, 0_2), \oplus(A, 0_3)) > \mathrm{PB}((B, 0_1) \oplus (B, 0_2), \oplus(B, 0_3))$; and this will be preserved after ruling out $(B, 1)$ and updating via Bayes' rule. For the Bayesian agent, A *must* become more likely than B after hearing 0.

Of course, before updating, they would have assigned that finer space consistent values *at that time*; but it is not their model that thinks either 0 or 1 may occur with chance which needs updating given the information that 0 occurred, since it would have happened in any case. The Bayesian agent does not remember the deterministic structure of the problem!

So, how do we go about representing an agent's uncertainty in such a situation? It is by having them believe a description of the problem; i.e. they add such a description to their register. And so we must be able to formalize our description of the problem. We imagine our agent in a discrete-time setting (I don't need time to be discrete for the agent, only that the agent measures time into discrete chunks, to be able to meaningfully say that a ball is being drawn *now*, or *next*, etc.), playing that same urn game given a semantics for credences.

I must begin by distinguishing between two different propositions, and I will just call them *now* and *ever* propositions, which will take on the role of distinguishing between the proposition (1) that a ball with a certain label is being drawn from the urn on this time-step, as opposed to (2) that a ball with a certain label is ever drawn from the urn. We will denote them ${}_{\mathrm{now}}A$ and ${}_{\mathrm{ever}}A$, respectively. Obviously, ${}_{\mathrm{now}}A$ entails ${}_{\mathrm{ever}}A$. And ever propositions will be the default, if nothing is denoted.

We introduce a relation, which I will denote $A \Rightarrow B$, which means that A and B are both drawn and in that order; i.e. both A and B are drawn, and if ${}_{\mathrm{now}}A$, then B is drawn on the next time-step (we can denote this ${}_{\mathrm{next}}B$). We will call collections of propositions connected by $\Rightarrow$ *chains* of propositions, where $A \Rightarrow B \Rightarrow C$ is shorthand for $(A \Rightarrow B) \wedge (B \Rightarrow C)$. I emphasize that the content of $A \Rightarrow B$ is exactly: $A \wedge B$ and ${}_{\mathrm{now}}A \rightarrow {}_{\mathrm{next}}B$. The agent will also be able to consider the order of the draws they see as evidence, in order to rule out hypotheses consisting of chains.

Now, we will be able to describe a wider class of generating processes. For instance, that which generates A first for sure, then B or C with odds 2-1, would be denoted $A \Rightarrow \{B^{2/3}, C^{1/3}\}$ and this is equivalent to $\{(A \Rightarrow B)^{2/3}, (A \Rightarrow C)^{1/3}\}$.

Now we can denote the statement of the problem that was described just above, with the bit sequences being revealed, still thinking in terms of the urn semantics. There are two generating processes: A, the process $0_{A,1} \Rightarrow 0_{A,2} \Rightarrow 0_{A,3}$ and B, the process $0_{B,1} \Rightarrow 0_{B,2} \Rightarrow 0_{B,3} \Rightarrow 1$, where the subscripts are to distinguish propositions/draws which the agent cannot distinguish, except by the process which generates them and their place in it. Thus, the agent has a credence (I will refer to this as their "first belief" below) about one of these two processes being the one which is generating the bits they will be hearing:

$$\{(0_{A,1} \Rightarrow 0_{A,2} \Rightarrow 0_{A,3})^{1/2}, (0_{B,1} \Rightarrow 0_{B,2} \Rightarrow 0_{B,3} \Rightarrow 1)^{1/2}\}$$

The agent is also uncertain about step of the process they are on. The way we will represent this is by considering the agent's hypothesis about how far along the process they *would* be, fixing each particular process. Perhaps they are entirely indifferent about what step is currently taking place, or perhaps they have some evidence which suggests being at a particular stage of the process, and perhaps the evidence too implies being earlier on if the process actually is A, but being later on if it's B, etc. I will assume the agent is just indifferent, but the point is that the evidence could depend on what is actually going on, and that can be accommodated. So, the agent also believes

$$(0_{A,1} \Rightarrow 0_{A,2} \Rightarrow 0_{A,3}) \longleftrightarrow \{{}_{\text{now}}0_{A,1}^{1/3}, {}_{\text{now}}0_{A,2}^{1/3}, {}_{\text{now}}0_{A,3}^{1/3}\}$$
$$(0_{B,1} \Rightarrow 0_{B,2} \Rightarrow 0_{B,3} \Rightarrow 1) \longleftrightarrow \{{}_{\text{now}}0_{B,1}^{1/4}, {}_{\text{now}}0_{B,2}^{1/4}, {}_{\text{now}}0_{B,3}^{1/4}, {}_{\text{now}}1^{1/4}\}$$

The agent thus also believes an exclusive disjunction, and thus a credence about all seven draws; that is entailed by these two beliefs together with that first belief. But they have more information, viz. that they get to this seven-member partition because really they think there are two possible processes, one with three and one with four steps.

Now consider what happens when the agent learns that the current draw is a 0, i.e. ${}_{\text{now}}0$. This entails ${}_{\text{ever}}0$, and they look at their first belief, and think that this would have occurred with probability 1 in either case (both A and B entailing that 0 is definitely drawn), and thus they do not update their first belief. However, the agent is, at this point, able to refine their models of where in either process they might be; in particular, they are able to rule out ${}_{\text{now}}1$ in case of B, and so they update that credence to:

$$B \longleftrightarrow \{{}_{\text{now}}0^{1/3}_{B,1}, {}_{\text{now}}0^{1/3}_{B,2}, {}_{\text{now}}0^{1/3}_{B,3}\}$$

Thus the hypothetical in-case-of-B model is updated. In turn, this collection of beliefs after updating will entail a model with different probabilities than that which is obtained by starting at those original beliefs, passing to the credence with seven elements, and then updating *that* after ruling out 1. In particular, A is not taken to be more likely after seeing ${}_{\text{now}}0$ in the former case, whereas ${}_{\text{now}}0$ makes A more likely in the latter case.

Thus, we now have our first example of that fallacy described in the last section. Here, the agent has evidence for believing each of:

$$\{(0_{A,1} \Rightarrow 0_{A,2} \Rightarrow 0_{A,3})^{1/2}, (0_{B,1} \Rightarrow 0_{B,2} \Rightarrow 0_{B,3} \Rightarrow 1)^{1/2}\}$$
$$(0_{A,1} \Rightarrow 0_{A,2} \Rightarrow 0_{A,3}) \longleftrightarrow \{{}_{\text{now}}0^{1/3}_{A,1}, {}_{\text{now}}0^{1/3}_{A,2}, {}_{\text{now}}0^{1/3}_{A,3}\}$$
$$(0_{B,1} \Rightarrow 0_{B,2} \Rightarrow 0_{B,3} \Rightarrow 1) \longleftrightarrow \{{}_{\text{now}}0^{1/4}_{B,1}, {}_{\text{now}}0^{1/4}_{B,2}, {}_{\text{now}}0^{1/4}_{B,3}, {}_{\text{now}}1^{1/4}\}$$

and, together, these entail:

$$\{{}_{\text{now}}0^{1/6}_{A,1}, {}_{\text{now}}0^{1/6}_{A,2}, {}_{\text{now}}0^{1/6}_{A,3}, {}_{\text{now}}0^{1/8}_{B,1}, {}_{\text{now}}0^{1/8}_{B,2}, {}_{\text{now}}0^{1/8}_{B,3}, {}_{\text{now}}1^{1/8}\}$$

If we update the entailed model upon learning that 0 is drawn, we get:

$$\{_{\text{now}}0_{A,1}^{4/21},{}_{\text{now}}0_{A,2}^{4/21},{}_{\text{now}}0_{A,3}^{4/21},{}_{\text{now}}0_{B,1}^{3/21},{}_{\text{now}}0_{B,2}^{3/21},{}_{\text{now}}0_{B,3}^{3/21}\}$$

thus yielding that A is now more likely than B, even though 0 would have been drawn for the first three draws, and then never again, in any case!

We realize the problem when we see that if we instead update those beliefs which entailed that model to begin with, we now get another model, which is:

$$\{_{\text{now}}0_{A,1}^{1/6},{}_{\text{now}}0_{A,2}^{1/6},{}_{\text{now}}0_{A,3}^{1/6},{}_{\text{now}}0_{B,1}^{1/6},{}_{\text{now}}0_{B,2}^{1/6},{}_{\text{now}}0_{B,3}^{1/6}\}$$

Thus, we see that we need some more refined belief update than merely passing to the finest partition and applying Bayes' rule, having forgotten how we got there, forgetting too any deterministic structure, etc. But this does not mean that we need a new rule for updating probabilities, nor that we need to do away with update.

8 The Sleeping Beauty problem

At last, the statement of the Sleeping Beauty problem is as follows. Sleeping Beauty goes to sleep on Sunday knowing all of the following. A fair coin is flipped (and she is not told the result); if it lands heads, she will awaken on Monday, and if she lands tails, she will awaken on Monday and Tuesday. But, on her Tuesday wake up, she will have no memories of waking up the previous day, and so all three wake ups are entirely indistinguishable experiences to her. Upon waking up, knowing she is in the game, she is asked about the probability the coin landed heads. Also, on Monday she may be then told it is Monday and then asked once more about the probability the coin landed heads.

Let us denote this. Firstly, if the coin lands heads, she undergoes one experience, and if the coin lands tails, she undergoes two experiences, all indistinguishable to her, except by the corresponding coin flip and the order. Thus, we can think of heads and tails as processes which generate wake ups, and we see that the game is analogous to a situation where the agent is 50-50 between two urns, one which draws one ball, and one which draws two balls deterministically in sequence, and each ball has the same label (thus indistinguishability). Syntactically, the agent believes the following to be her description of that process which generates those draws, where W stands for one of those indistinguishable wake-ups:

$$\{W_H^{1/2}, (W_{T,1} \Rightarrow W_{T,2})^{1/2}\}$$

Furthermore, she knows exactly what stage of the process she is currently at in case of heads, there being only one, and her knowing herself to be in the game. Thus, we may denote:

$$W_H \longleftrightarrow {}_{\text{now}}W_H$$

In case of tails, she has no reason to think either one more likely, but she is guaranteed that she is at some step of the process, so she assigns both now propositions equal probability under this hypothesis:

$$(W_{T,1} \Rightarrow W_{T,2}) \longleftrightarrow \{{}_{\text{now}}W_{T,1}^{1/2}, {}_{\text{now}}W_{T,2}^{1/2}\}$$

The three of these together is the full formal statement of the Sleeping Beauty problem.

Using the system presented in the last two sections, we may find those asked after probabilities. Firstly, knowing herself to be awake, Beauty learns ${}_{\text{now}}W_H \oplus {}_{\text{now}}W_{T,1} \oplus {}_{\text{now}}W_{T,2}$ (these are exclusive, so I use the exclusive disjunction); she thus also learns $W_H \vee W_{T,1} \vee W_{T,2}$ (these are not exclusive, so I use the inclusive disjunction). Now, of course, W_H and $W_{T,1} \Rightarrow W_{T,2}$ both entail $W_H \vee W_{T,1} \vee W_{T,2}$, so she does not update her credence about the processes. She also can do no better with respect to her hypotheses about where she is in either process, having ruled nothing out at all. Thus, she deduces from these original beliefs:

$$\{{}_{\text{now}}W_H^{1/2}, {}_{\text{now}}W_{T,1}^{1/4}, {}_{\text{now}}W_{T,2}^{1/4}\}$$

and is able to provide that the probability of ${}_{\text{now}}W_H$, coextensive with heads given being awake, is 1/2.

Next, if she is told she is in a Monday wake up, she learns ${}_{\text{now}}W_H \oplus {}_{\text{now}}W_{T,1}$; and thus also $W_H \oplus W_{T,1}$. Of course, still, W_H and $W_{T,1} \Rightarrow W_{T,2}$ both entail $W_H \oplus W_{T,1}$ (i.e. that the Monday wake up corresponding to that coin flip definitely occurs, and only that Monday wake up). She wakes up on Monday, and thinks this would have happened anyways. But, in case of tails, she can now do better, and she can rule out ${}_{\text{now}}W_{T,2}$, yielding:

$$(W_{T,1} \Rightarrow W_{T,2}) \longleftrightarrow {}_{\text{now}}W_{T,1}$$

And this, together with $\{W_H^{1/2}, (W_{T,1} \Rightarrow W_{T,2})^{1/2}\}$ and $W_H \longleftrightarrow {}_{\text{now}}W_H$, yields:

$$\{{}_{\text{now}}W_H^{1/2}, {}_{\text{now}}W_{T,1}^{1/2}\}$$

Thus, she still assigns heads probability 1/2. But she does not move between these two probability assignments via a single model update.

The Bayesians are, of course, incapable of representing this. They must make a choice of some wrong conclusion to accept. Some bite the bullet and accept one wrong conclusion, and then try to prop it up theoretically by making the assumption that individuals are distributed among those places/times they *might* be. Thus, in my terms, they simply begin by believing a credence containing each of the three wake ups, rather than thinking it merely entailed by the description of the problem, but even then they disagree about how to assign such a finest partition probabilities. I will come back to this over next two sections.

The first partial belief analysis of the problem begins by noting that given Monday, we should think heads and tails still equally likely. After all, the coin may not have been flipped yet! Thus, heads and Monday and heads and Tuesday must have a 1:1 partial belief ratio, that being maintained after ruling out the second tails wake up. But, then, notice that given tails, she thinks the first and second wake ups both equally likely. Thus, they too have the same ratio of partial belief, yielding: $\mathrm{PB}({}_{\text{now}}W_H) = \mathrm{PB}({}_{\text{now}}W_{T,1}) = \mathrm{PB}({}_{\text{now}}W_{T,2})$. And, since these are exclusive and exhaustive, that gives each of them probability 1/3 (Elga 2000).

The other option is to say: upon waking up, heads and tails are equally likely. Thus, the probability of the two tails wake ups must combine to the probability of the heads wake up. But, again, the two tails wake ups must have the same probability, Sleeping Beauty being indifferent given tails. This yields that $\mathrm{PB}({}_{\text{now}}W_H) = 1/2$ and $\mathrm{PB}({}_{\text{now}}W_{T,1}) = \mathrm{PB}({}_{\text{now}}W_{T,2}) = 1/4$. This matches my analysis at this time-slice. But, then, it means that learning that the current wake up is on Monday, she updates via Bayes' rule, yielding that heads is now twice as likely (i.e. $\mathrm{PB}({}_{\text{now}}W_H) = 2/3$) as tails (Lewis 2001).

So, in either case, we have a coin that is guaranteed to be fair changing probabilities depending on how much the agent knows about what time it is. But we can see that this is because the Bayesians start at the same description of the problem, but then try to condense it into a single distribution of partial beliefs. Thus, information is lost: Sleeping Beauty cannot think Monday would have happened anyways, if she is a Bayesian.

A prominent anti-Bayesian view within philosophy, time-slicing, has sought to avoid this by beginning at the partial belief assignments they desire at each step, and doing away with updating. So, they begin at their desired solution, and they say: even if Bayes' rule does not get me this, I will take it anyways. At each time step: go to the intuitively correct answer. And thus they can only back their solutions up with intuition, appealing to thought experiments about such spooky things as personal identity over time, which the literature has gotten itself into the habit of using, as if such questions, if they can even really be cashed out in a sensible way, are in any way within the domain of something like formal epistemology, which presupposes some well-defined notion of an agent (Builes 2020).

Thus, time-slicers will easily match my analysis. But, unlike the time-slicers, I have given a framework within which this problem can be represented. And these representations are, of course, all static, being beliefs about descriptions of processes. Thus I say to the time-slicer: time-slice by taking credences to be beliefs, and not some special other thing, some mysterious partial-beliefs-prior the agent begins with and is bound to.

A similar complaint goes for those who opt to image, using distances between possible worlds to yield the results they would intuitively like. For instance, imagers deal with the Sleeping Beauty problem by conceiving of each wake up itself as a possible world, with the two tails wake ups each being closer to the other than to the heads wake up. But there is no sense in which, say, if today is Monday, there is another "possible world" in which it is Tuesday; if today is Monday, then that means *tomorrow* is Tuesday. Perhaps I am uncertain about what day it is, but we have already seen that passing to some finest partition of possibilities loses the retrospective distinction between two different kinds of possibility, one which says this would have happened anyways, and one which says this might not have. But, still within this faulty framework, the imager encodes all of the relevant information into something their updating rule has access to, and they must then argue using intuition for these distances being sensible, let alone correct, such that their analysis of the Sleeping Beauty problem might actually yield a solution (Cozic 2011).

9 More on Sleeping Beauty-type cases

Now I will move away from the Sleeping Beauty problem itself, and toward the literature it has become situated within. I will begin by giving a static version of the Sleeping Beauty problem, one that is well-known.

Imagine an agent who believes that one of two universes has been created, and that they are equally likely: one with one planet and one being, and one with two planets, each with one being, but they are entirely separate. The agent finds themselves on such a planet, knowing themselves to be one of these beings. How likely is either universe? It is obvious that this is analogous to the Sleeping Beauty problem, though it involves less, with no need to keep track of order. If we denote the proposition that there is someone is alive on a planet A, we consider two generating processes (we may as well call them H and T), one which generates $A_{H,1}$ and one which generates $A_{T,1} \wedge A_{T,2}$. Learning that one is on a planet is learning ${}_{\text{here}}A_{H,1} \oplus {}_{\text{here}}A_{T,1} \oplus {}_{\text{here}}A_{T,2}$ (where we distinguish between *here* and *anywhere* propositions, in the place of now and ever). And then the analysis goes through just as above. The agent's evidence is that at least one being is undergoing phenomenal experience, which was guaranteed in either case.

I bring this example up because apparently the literature has taken to thinking the analysis of such problems suggests that we may live in a multiverse, as if such things as how some agent-in-the-world updates their own evolved constructions of the world, or even idealized versions of these, bears on such questions (Isaacs et al. 2022)!

In any case, I merely wish to make the following comment about this paper; they distinguish between different approaches to these sorts of problems, each as a contender to be The Rule which yields the correct assignment of partial beliefs to this finest partition, containing worlds and times/places within them. Of course there is no such rule.

One of the contenders, so-called *compartmentalized conditionalization*, updates as we do here in the Sleeping Beauty problem, but then they, constrained to the partial beliefs model which cannot distinguish between uncertain stages of deterministic and indeterministic generating processes, are forced to reason incorrectly in the case where an agent is uncertain between two urns, one drawing red with probability 1/3 and one drawing red with probability 2/3.[12] Just as Sleeping Beauty wakes up on Monday and thinks: this would have happened anyways, the compartmentalized conditioner sees a red draw and thinks: both urns *could* have produced this, and thus does not update. Of course, this is not correct: the urn which

[12]We analyzed this case in section 7.2 when I showed my system could match Bayesian confirmation.

is more likely to produce red should now be thought more likely. Unlike the Sleeping Beauty problem's Monday, red, here, is not guaranteed to occur in *either* case, let alone both.

But the alternatives, which are able to accomodate such reasoning, where we think those theories more likely to have produced some evidence themselves more likely, cannot reason correctly about the Sleeping Beauty problem. They are all bound to this partial beliefs model, and all of their best theories are something along the lines of: fix one way of analyzing a problem, and do this always. But no such way will suffice, there being some times when one analysis is correct, and some times when another is. But, in order to see when, one must represent the problems correctly, which until now has not been done.

9.1 Appealing to the long-run

One common appeal made alongside the Bayesian analysis of the Sleeping Beauty problem, that which assigns probability 1/3 to heads upon waking up, goes: in the long-run, over many plays, 1/3 of the wake ups are heads wake-ups, thus heads should be assigned probability 1/3 upon waking up.

This appeal models the game incorrectly. The wake ups are not drawn from some single pool, some single urn as the Bayesians are bound to. This appeal thinks of the game as one in which, on each heads wake up, a red ball is placed into the urn, and a green ball upon each tails wake up. Then it asks about the chance of drawing some ball from the urn, it having been populated by many runs, and of course, one will be more likely to draw a green ball/tails wake up in this case. Instead, they should ask about the chance of currently, given that one is in some single round of the game, placing red balls or green balls (and this is obviously 1/2, not changed over each iteration of the game). Thus, this naive frequentist appeal forgets that Sleeping Beauty knows she is in some single run of the game.

We can actually show that 1/3 is the correct solution when she does not know this. We do this by generalizing the game by adding coin flips. Consider a version of the Sleeping Beauty game where instead two coins are flipped; for each coin flip that lands heads, a single heads wake up is added, and for each that lands tails, two tails wake ups are added. This game considers four possible processes which may generate the wake ups: HH, HT, TH and TT, the first generating two heads wake ups, the next one heads and two tails wake ups, etc. If we represent this game syntactically, we get that Sleeping Beauty will wake up and think that heads being responsible for the current wake up has probability 5/12.[13] For a fixed n-many coin flips, the probability of heads on a given wake up is $\sum_{i=0}^{n}\binom{n}{i}\frac{n-i}{n+i}$, and if we take the limit of this as n goes to infinity, we get 1/3. What this means is: if Sleeping Beauty were instead playing infinitely many games, seamlessly, she can do no better than that the current wake up has probability 1/3 of being generated by a coin landing heads. But, for any finite-sized game, she can use the fact that she is in a run of the game to do better than 1/3, but no better than 1/2 at $n = 1$.

10 More problems for Bayesians

This section will conclude by discussing some other problems for Bayesians, with the hope of providing some perspective by using the account of credences I presented in this paper.

10.1 Justification versus probability raising

There has been a trend in the Bayesianism literature of being concerned with the distinction between probability raising, or confirmation, and justification (Moss 2025). This problem is entirely due to the use of partial beliefs.

A characteristic example is the so-called *problem of old evidence*, which takes the following form. Suppose an agent has a theory, and assigns it some subjective probability within some space of theories. Suppose now that the agent realizes that the theory entails something the agent already knew about, but which was perhaps yet unexplained. Since the evidence was known, the agent has not actually ruled any possibilities out when they learn this, so they

[13] I think that this is an instructive exercise, and so I leave it to the reader to verify this.

may not update via conditionalization (in my terminology, they perform no model update). So, no probabilities are raised, meaning, in this context, that the proposition is not believed *any more*, the beliefs being partial and updating via Bayes' rule; but we think that intuitively the agent now has some more justification for believing the theory, it explaining something which was previously unexplained.

Taking credences to be full beliefs dissolves the issue entirely. Those reasons for which we believe a credence are "justification," and we may have better or worse reasons for the same belief. This is banal. The question of what reason is sufficient is the problem of induction. But it is obvious that the agent can have better or worse reasons for using the same probabilistic model, and thus assigning the same probabilities.

10.2 Indifference

Now I will turn to another problem, which is about the so-called *principle of indifference*. This principle, to *objective Bayesians*, is a rational norm that one sets their partial beliefs to equal values for a partition about which they know nothing. Modern objective Bayesians look to the principle of maximum entropy as the generalization, and their updating rule is anti-Bayesian (Williamson 2024). A problem seems to arise with the principle of indifference.

Take the following illustrative case ((Keynes 1921), (Norton 2008)): suppose an agent awakes, knowing nothing, and then is asked: "are you in the British Isles or France?" The agent, being indifferent (and trusting their questioner would not be deceiving them) assigns them both partial belief 1/2. Then the agent is asked: "are you in Great Britain, Ireland, or France?" Note that these two sets are equivalent, the British Isles being Great Britian and Ireland.[14] But, the agent now assigns each partial belief 1/3, thus assigning France two inconsistent partial beliefs being asked equivalent questions.

So-called *subjective* Bayesians dodge this concern by saying that the principle of indifference and its generalizations are bunk, and that any prior is rationally permissible, so long

[14]along with some smaller islands which this example erases — my apologies but please play along

as the agent cannot be *Dutch-booked*; i.e. made to lose money on bets based on their partial belief assignments. This seems true; it seems unreasonable to say any belief, other than a contradiction, is rationally impermissible *simpliciter*.

But subjective Bayesians have missed the point also. I do not think any prior is *reasonable*, even if it cannot be discarded as a matter of irrationality. For instance, an agent who is asked after "A, B, or C?" knowing that one is correct but nothing else, should assign each 1/3. I think it strange for the agent to be thought perfectly reasonable to assign probabilities 98/100, 1/100, and 1/100, respectively, and then be terribly surprised if the answer is C; this seems like bad modeling practice. But I cannot rule out that the agent has a hunch, some probabilistic judgment they stand by but cannot give explicit evidence for.

So, this is the account of the principle of indifference I want to give: that it is a modeling practice. With regard to the problem with the agent asked about France and Britain, I think that the agent did not even have any conception of being in any of those places before being asked any questions. Then, they are asked about the British Isles or France, and now they believe an exclusive disjunction over these two, them being the possible *answers to the question*. They assign them equal probability because they are asked a question, they think they can do no better than random among the answers, so they think them all equally likely. Then, asked a different question, the agent no longer believes that old credence, and constructs a new one, evidenced in this new question, thus assigning each among Great Britain, Ireland, and France probability 1/3. There is no contradiction because I am not presupposing any kind of more entrenched thing as the agent's partial belief that they are in France, which must remain consistent between different questions. Those questions were why the agent considered France, or anything else, a possibility at all, their probability assignment reflecting their belief that these answers are all equally plausible, as answers to the question, in the face of the (lack of) evidence. Thus, the agent has two different models at two different times due to two different evidences, and there is no contradiction.

The agent is allowed, of course, to remain committed to the first model, and model the second question in a way consistent with it (by having the probabilities of Ireland and Great Britain sum to the probability of the British Isles). But their probability assignments reflect how likely they think each option is *to be an answer to the question*; thus, each time they are asked a question, they take all of the answers to be in some sense comparable. Indifference is the thing we name when creating a model and choosing to assign comparable things the same probability. The assignment that France and the British Isles are equally probable contains the agent's judgment that they are comparable as candidates. And so, while committed to this, the agent cannot think that Great Britain and Ireland are also comparable to France, since they are not comparable to the British Isles (them composing it), and comparability intuitively being a transitive thing. But, since, to begin with, the agent judged them comparable as answers to the question, there is no reason they should not be allowed to model a different question with a different space of answers differently (unless they commit to some more entrenched thing, like a bet).

10.3 Doomsday/simulation-style arguments

Now we come to what I am calling doomsday/simulation-style arguments. The idea is basically as follows. Imagine that there are either going to be 1,000 people or 1,000,000 people in the history of society, and that everyone knows this. One of the people, at first uncertain as to which person they were, is told that they are the 100th. Of course, being the 100th, the agent thinks "this would have happened in either case," and does not now have evidence to think one of those hypotheses much more likely than the other, had they not thought so before. Of course, the Bayesian framework cannot accommodate this for the same reasons for which they cannot accommodate the Sleeping Beauty problem. But then people take this and theorize in the direction of: well, the agent will think, what are the chance's *I* would have been the 100th. As if there is any sense of "I" beyond that which arises out of that physical process we call that agent; some chance *they* could have been *someone else*. This does not make any sense, unless we presuppose a system of souls being distributed into bodies. And they apply this faulty reasoning to all kinds of cases; with high probability, we are toward the end of human society (Leslie 1990)! with high probability, we live in a multiverse! with high probability, we live in a simulation (Bostrom 2003)! Enough!

10.4 Newcomb's problem

The final problem I will discuss is Newcomb's problem and the EDT/CDT debate. Note firstly that this debate, like many in formal epistemology, is searching for some correct and universally applicable decision rule, which takes in the right sorts of probabilities, etc. But by now there are a number of contenders (Weatherson 2024).

The statement of Newcomb's problem is generally of the form: an agent knows that they are playing a game set up by some predictor who has been historically very accurate, say 999/1000 times. There are two boxes, an opaque one and a transparent one in the room of the game. The agent is allowed to take both or only the opaque one. The day before, the predictor comes and places one thousand dollars under the transparent box. They will also place one million dollars under the opaque box if they predict that the agent will take only the opaque box (Weirich 2024).

There are so-called *causal decision-theorists* (CDT) who recommend taking both boxes, citing that the money is already in the box(es — or not). And then there are *evidential decision-theorists* who recommend taking only the opaque box, citing the fact that the predictor has been historically highly accurate.

Then, there is a version of the game where the predictor is assumed not just to have been historically accurate, but somehow robust, a prophet who can, with 99.9% accuracy, predict what you actually will do, conditional on you doing either thing! In this case, there is clearly sufficient evidence for a credence about how accurate the predictor is, and the agent should take this into account, and take only the opaque box. Otherwise, when the predictor only has claim to having been historically accurate, the correct move is to take both boxes, there being no need for probabilistic reasoning, a mere frequency not constituting sufficient evidence for a credence.

A reductio of the one-boxing position in the standard game is to consider the open-box variant, wherein both boxes are transparent. EDT, as it is normally understood, recommends double-boxing in this case, because now the agent can see the money, and no longer has any uncertainty about their payoffs. But this is too hasty an analysis. Note that they do still have uncertainty about how accurate the predictor was; and a metaphysical appeal is required to assure the agent that the money they see will not *poof* disappear before their eyes. They will say: the money is already there. But this same metaphysical appeal was already being made by those who said that "the money is already in the box" in the standard case.

The questions which should be debated around this problem do not involve some decision rule which has yet to, and will not, be found. Instead, one should see this problem and think: what is sufficient evidence for a credence? And this is the problem of induction.

10.5 The problem of induction

On the problem of induction, I will only say that I think that there is no general thing called "inductive inference." Some special cases like Bayesian enumerative induction or material-theory-of-inductions (Norton 2021) presuppose enough theoretical framework to make that "inductive inference" somehow instead deductive; and it is choosing such a theoretical framework that the problem of induction is concerned with. In general I think that agents begin to notice patterns within those models of the world evolution endowed them with, and then they actively construct more models. Then they deduce from those models also, and perhaps they fill its parameters using measurements of what those models claim to be models-of. Then perhaps a model is outdone, or somehow seriously falsified beyond reasonably considering recovering it, etc. And it may be replaced. But there is no general rule for model creation, and the existence of some thing which can be constrained as evidence is no guarantee of a model. We do not pass, inferentially, from evidence to models, but we justify our models using evidence.

References

Bostrom, Nick (2003), “Are We Living in a Computer Simulation?”, *Philosophical Quarterly*, 53, 211, pp. 243-255, DOI: 10.1111/1467-9213.00309.

Builes, David (2020), “Time-Slice Rationality and Self-Locating Belief”, *Philosophical Studies*, 177, 10, pp. 3033-3049.

Cozic, Mikael (2011), “Imagining and Sleeping Beauty: A Case for Double-Halfers”, *International Journal of Approximate Reasoning*, 52, 2, pp. 137-143.

Elga, Adam (2000), “Self-Locating Belief and the Sleeping Beauty Problem”, *Analysis*, 60, 2, pp. 143-147, DOI: 10.1111/1467-8284.00215.

Fishburn, Peter C. (1986), “The Axioms of Subjective Probability”, *Statistical Science*, 1, 3, pp. 335-358.

Hedden, Brian (2015), “Time-Slice Rationality”, *Mind*, 124, 494 (Mar. 2015), pp. 449-491.

Isaacs, Yoaav, John Hawthorne, and Jeffrey Sanford Russell (2022), “Multiple Universes and Self-Locating Evidence”, *Philosophical Review*, 131, 3, pp. 241-294, DOI: 10.1215/00318108-9743809.

Keynes, John Maynard (1921), *A Treatise on Probability*, Macmillan & co., London,

Konek, Jason (2019), “Comparative Probabilities”, in *The Open Handbook of Formal Epistemology*, ed. by Richard Pettigrew and Jonathan Weisberg, pp. 267-348.

Leslie, John (1990), “Is the End of the World Nigh?”, *The Philosophical Quarterly (1950-)*, 40, 158, pp. 65-72, ISSN: 00318094, 14679213, http://www.jstor.org/stable/2219967 (visited on 08/11/2024).

Lewis, David (2001), “Sleeping Beauty: Reply to Elga”, *Analysis*, 61, 3, pp. 171-76, DOI: 10.1111/1467-8284.00291.

Lin, Hanti (2024), “Bayesian Epistemology”, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta and Uri Nodelman, Summer 2024, Metaphysics Research Lab, Stanford University.

Lin, Hanti and Kevin T. Kelly (2012), "Propositional Reasoning that Tracks Probabilistic Reasoning", *Journal of Philosophical Logic*, 41, 6, pp. 957-981, DOI: 10.1007/s10992-012-9237-3, https://doi.org/10.1007/s10992-012-9237-3.

Mierzewski, Krzysztof (2020), "Probabilistic Stability, Agm Revision Operators and Maximum Entropy", *Review of Symbolic Logic*, pp. 1-38, DOI: 10.1017/s1755020320000404.

Moss, Sarah (2015), "Time–Slice Epistemology and Action under Indeterminacy", in *Oxford Studies in Epistemology Volume 5*, Oxford University Press, ISBN: 9780198722762.

— (2025), "Justification and Grounding", Colloquium talk presented at CMU Philosophy.

Norton, John D. (2008), "Ignorance and Indifference", *Philosophy of Science*, 75, 1, pp. 45-68, DOI: 10.1086/587822.

— (2021), *The Material Theory of Induction*, Calgary, Alberta, Canada.

Weatherson, Brian (2024), "Why Ain?T Evidentialists Rich?", *Analysis*, 84, 4, pp. 813-821, DOI: 10.1093/analys/anae038.

Weirich, Paul (2024), "Causal Decision Theory", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta and Uri Nodelman, Winter 2024, Metaphysics Research Lab, Stanford University.

Williamson, Jon (2024), "Where Do We Stand on Maximal Entropy?", in, *Perspectives on Logics for Data-driven Reasoning*, ed. by Hykel Hosni and Jürgen Landes, Springer Nature Switzerland, Cham, pp. 39-61, ISBN: 978-3-031-77892-6, DOI: 10.1007/978-3-031-77892-6_3, https://doi.org/10.1007/978-3-031-77892-6_3.