

# Unravelling the Tangled Web: Continuity, Internalism, Uniqueness and Self-Locating Belief

Christopher J. G. Meacham

## Abstract

A number of cases involving self-locating beliefs have been discussed in the Bayesian literature. I suggest that many of these cases, such as the sleeping beauty case, are entangled with issues that are independent of self-locating beliefs *per se*. In light of this, I propose a division of labor: we should address each of these issues separately before we try to provide a comprehensive account of belief updating. By way of example, I sketch some ways of extending Bayesianism in order to accommodate these issues. Then, putting these other issues aside, I sketch some ways of extending Bayesianism in order to accommodate self-locating beliefs. I then propose a constraint on updating rules, the “Learning Principle”, which rules out certain kinds of troubling belief changes, and I use this principle to assess some of the available options. Finally, I discuss some of the implications of this discussion on the Many Worlds interpretation of quantum mechanics.

## 1 Introduction

The standard Bayesian theory works with something like the following model. The objects of belief are like four dimensional maps: maps of how the world might be. On one map, the tallest mountain in the world circa 2000 A.D. is in Asia, on another map it’s in North America. We have varying degrees of confidence in these maps; we think it more likely that some of the maps correctly represent our world than others. The task of Bayesianism is to describe how our degrees of confidence in these maps should change in light of new evidence. If we discover that the tallest mountain is in Asia, not North America, Bayesianism tells us how to readjust our beliefs in these maps in light of this discovery.

But we don’t just have beliefs about what the world is like—about which map is the correct one. We also have beliefs about our location on the map—beliefs about where we are on the map, when we are on the map, and who we are on the map. And the standard Bayesian account doesn’t apply to these “self-locating” beliefs.

Extending the standard Bayesian account to accommodate such beliefs is a non-trivial task. The difficulty arises because the Bayesian updating rule—conditionalization—requires certainties to be permanent: once you’re certain of something, you should always be certain of it. But when we consider self-locating beliefs, there seem to be cases where this is patently false. For example, it seems that one can reasonably change from being certain that it’s one time to being certain that it’s another.

The question of how to extend conditionalization in order to accommodate self-locating beliefs has attracted a lot of discussion in the recent literature.<sup>1</sup> Unfortunately, the project has turned out to be a tricky one, and there is little consensus regarding how to proceed.

I propose that some of the difficulty this project has encountered stems from the fact that many of these discussions are entangled with issues that are independent of self-locating belief *per se*. Three issues in particular are worth separating from the issue of self-locating beliefs.

The first issue concerns identity or continuity over time. Conditionalization is a diachronic constraint on beliefs; i.e., a constraint on how a subject's beliefs at different times should be related. But in order to impose such a constraint, it needs a way of picking out the same subject at different times. So it needs to employ something like a notion of personal identity, or some epistemic surrogate for personal identity ("epistemic continuity"). The first issue is this: what notion of personal identity or epistemic continuity should the Bayesian employ?

The second issue concerns internalism about epistemic norms. In Bayesian contexts, many people have appealed to implicitly internalist intuitions in order to support judgments about certain kinds of cases. But diachronic constraints on belief like conditionalization are in tension with internalism. Such constraints use the subject's beliefs at other times to place restrictions on what her current beliefs can be. But it seems that a subject's beliefs at other times are external to her current state. The second issue is this: how should we reconcile Bayesianism with internalism about epistemic norms, if at all?

The third issue concerns non-unique predecessors. Let's call the earlier temporal stage of a subject, the stage that held her previous beliefs, her "predecessor".<sup>2</sup> Conditionalization is a function of a subject's current evidence and prior beliefs. But conditionalization is only well-defined if the subject has a unique set of prior beliefs, i.e., a unique predecessor. And there can be cases in which subjects don't have unique predecessors, such as when a subject is the result of the fusion of two earlier agents. The third issue is this: how should we extend conditionalization in order to accommodate non-unique predecessors?

These three issues often arise in cases involving self-locating beliefs. And there are natural ways of treating these topics that lead to overlaps: one's treatment of internalism might employ self-locating beliefs of a certain kind, for example, or one might employ epistemic continuity relations in one's account of how to update self-locating beliefs. That said, these issues are separate from the issue of how to update self-locating beliefs. In light of this, I suggest a division of labor. We should attempt to address these issues separately before we try to provide a comprehensive account of belief updating.

In what follows, I will sketch some ways of adapting Bayesianism in order to accommodate each of these issues. While doing so, I'll restrict my attention to *sequential* updating rules: rules which generate what a subject's current beliefs should be from her evidence and her prior beliefs. There are, of course, other kinds of updating rules one could employ. In particular, there are also *epistemic kernel* rules: rules which pair each subject with a static "epistemic kernel", and determine what her current beliefs should be

---

<sup>1</sup>For a sampling of this literature, see Arntzenius (2002), Arntzenius (2003), Bostrom (2007), Bradley (2003), Dorr (2002), Elga (2000), Elga (2004), Halpern (2005), Hitchcock (2004), Horgan (2004), Jenkins (2005), Kierland and Monton (2005), Kim (2009), Lewis (2001), Meacham (2008), Monton (2002), Titelbaum (2008), Weintraub (2004), and White (2006).

<sup>2</sup>Throughout this paper I'll often speak as if there are temporal parts; but this is merely a matter of convenience.

using her evidence and her kernel.<sup>3</sup> Examples of such rules include formulations of conditionalization in terms of initial credence functions, ‘hypothetical priors’ or ‘ur-priors’.<sup>4</sup>

The strength of epistemic kernel rules is that they’re easy to extend to cases where there are strange and dramatic changes in one’s epistemic situation, since preserving a kind of step-wise continuity is not a concern. But a consequence of this detachment from step-wise continuity is that epistemic kernel rules have less of a diachronic feel than sequential rules do: their ability to impose diachronic constraints is effectively a side effect of how the kernel and the rule are set-up. As a result, it’s hard for such rules to capture our diachronic intuitions in certain kinds of cases, unless we impose further constraints on the epistemic kernel itself.

I think both approaches are promising. I have explored some ways of applying epistemic kernel rules to the issue of self-locating beliefs in Meacham (2008). But in this paper I will focus my attention on sequential rules.<sup>5</sup>

The rest of this paper will proceed as follows. In the next section I’ll sketch some background material. In the third section I’ll look at the three issues described above, continuity, internalism and non-unique predecessors, and sketch some natural extensions of conditionalization in light of these issues. In the fourth section I’ll examine the project of extending Bayesianism to accommodate self-locating beliefs with these other issues put aside, and I’ll sketch a natural sequential extension of conditionalization that accommodates such beliefs. I’ll then look at how to employ these different extensions in concert, and apply them to some of the standard cases discussed in the literature on self-locating beliefs. In the fifth section, I’ll turn to consider a potential desideratum for updating rules. In particular, I’ll formulate a constraint on updating rules which rules out certain troubling kinds of belief changes. Then I’ll assess the proposals I’ve discussed in light of this constraint. In the sixth and final section, I’ll briefly consider the bearing of this discussion on a debate regarding the Many Worlds interpretation of quantum mechanics.

## 2 Background

### 2.1 Belief

In what follows, I will follow David Lewis (1979) in distinguishing between two kinds of beliefs.<sup>6</sup>

First, there are beliefs which are entirely about what the world is like. We can characterize the content of such beliefs using sets of possible worlds: to have such a belief is to believe that your world is one of the worlds in that set. Call these beliefs *de dicto* beliefs,

---

<sup>3</sup>While sequential rules are clearly diachronic constraints, epistemic kernel rules seem at first glance to be synchronic constraints: they only place constraints on the subject’s belief, evidence and epistemic kernel at a single time. Epistemic kernel rules get their diachronic “grip” because the kernel is static: a subject’s beliefs at different times will all be related because they will all have been generated from the same kernel.

<sup>4</sup>For examples of such formulations, see Strevens (2004), Meacham (2008).

<sup>5</sup>I do this for convenience. The same issues arise with respect to epistemic kernel rules: in saying that a subject’s kernel is static, we implicitly employ a notion of continuity; the kernel will generally be external to the subject, which gives rise to tensions with internalism; and subjects are assumed to have a unique kernel, which makes cases of fusion between subjects with different kernels problematic. (An exception arises for objectivists who hold that there is only one rationally permissible kernel. They can arguably avoid all of these problems.)

<sup>6</sup>Which is not to say that these are the *only* two kinds of belief.

and the objects of such beliefs *de dicto propositions*. Having a *de dicto* belief entails that all of the worlds you believe might be yours—your *doxastic worlds*—are members of the set of worlds associated with that belief.

Second, there are beliefs which are both about the world and one’s place in the world. We can characterize the content of such beliefs using sets of *centered worlds* or *possible alternatives*, ordered triples consisting of a world, time and individual. To have such a belief is to believe that your world, time and identity correspond to one of the alternatives in that set. Call these beliefs *de se* beliefs, and the objects of such beliefs *centered propositions* or *de se propositions*. Having a *de se* belief entails that all of the possible alternatives you believe might be yours—your *doxastic alternatives*—are members of the set of alternatives associated with that belief.

All *de dicto* beliefs can be characterized as *de se* beliefs, but not vice versa. If we can characterize the content of a belief as a set of possible worlds, then we can characterize that content using possible alternatives just as well: simply replace each world with all of the alternatives located at that world. But most *de se* beliefs cannot be characterized as *de dicto* beliefs, since they will include some, but not all, of the alternatives at various worlds. Call such beliefs *irreducibly de se* or *self-locating* beliefs.

There are, of course, other kinds of belief besides these two. Many beliefs can’t be adequately represented in either of these ways, such as *de re* beliefs, beliefs about logical or metaphysically necessary truths, and so on. But these other kinds of belief aren’t relevant to the issues we’ll be concerned with here, so I’ll put them aside.

I’ve followed Lewis in using alternatives to model self-locating beliefs, but nothing of substance hangs on this. The problem of extending Bayesianism to accommodate self-locating beliefs persists regardless of how we choose to model the content of the beliefs in question. Consider the belief that  $w$  is a precise description of what the world is like, and that you are individual  $i$  at time  $t$ . If a method of representing the objects of belief cannot capture such beliefs, then it is too coarse to capture the kinds of beliefs we want to consider. On the other hand, if it can capture such beliefs, then we can take alternatives to correspond to the equivalence class of the beliefs that have the desired content, and translate discussion about belief in alternatives into discussion about belief in these surrogates.

## 2.2 Bayesianism

People have used the term “Bayesianism” to mean a number of different things. For the purposes of this paper, I’ll understand Bayesianism in the following way.

Bayesian theory can be divided into two parts: a description of the agents to which the theory applies, and a normative claim about what the beliefs of such agents should be like. The agents to which Bayesianism applies satisfy the following conditions:

- A1.** The agent’s belief state at a time can be represented by a probability function over a space of possibilities. These values, called *credences* or *degrees of belief*, indicate the subject’s confidence that the possibility is true, where greater values indicate greater confidence.<sup>7</sup>
- A2.** The agent’s evidential state at a time can be represented by a set of possibilities. The possibilities in the set are the possibilities compatible with the agent’s evidence.<sup>8</sup>

---

<sup>7</sup>Some drop this prerequisite, and instead take A1 to be a normative constraint (*probabilism*).

<sup>8</sup>Some take the agent’s evidence to include any proposition the agent becomes certain of (see Howson and

**A3.** The space of possibilities in question is the space of maximally specific ways the world could be, or possible worlds.<sup>9</sup>

The normative part of the Bayesian theory claims that agents who satisfy A1-A3 ought to satisfy *conditionalization*.<sup>10</sup> The sequential formulation of conditionalization is:

**Conditionalization:** If a condition-satisfying agent with credences  $cr$  receives  $e$  as her total new evidence, then her new credence function  $cr_e$  should satisfy the following constraint:

$$cr_e(\cdot) = cr(\cdot|e), \text{ if defined.} \quad (1)$$

For ease of exposition, I'll simplify the following discussion in two ways. First, I'll generally discuss things in finitary terms. Second, I'll assume that we have a way of carving up a subject's epistemic states which allows us to employ a simplified picture of what the subject's alternatives are like. According to this picture, the times that index the alternatives centered on epistemic subjects are discrete, and these times match the times at which the subject gets new evidence.<sup>11</sup> These assumptions are not entirely innocent: each obscures some deep and interesting issues. Nevertheless, for the purposes of this paper, I'll put these issues aside.

---

Urbach (1993)). I prefer to allow for a more substantive account of evidence. So I do not assume here that any proposition an agent becomes certain of should count as evidence. Likewise, I do not assume that agents are always certain of their evidence (though conditionalization will, of course, impose this as a *normative* constraint).

<sup>9</sup>This constraint is not an explicit part of the usual Bayesian package. That said, I think it's reasonable to take it to be an implicit part of the package: Bayesianism is almost always applied to beliefs about what the world is like, and the standard Bayesian account runs into difficulties when we try to apply it to broader kinds of belief.

<sup>10</sup>Some take the normative part of Bayesianism to include *probabilism*: that subjects ought to have credences which satisfy the probability axioms (see Howson and Urbach (1993)). I prefer to think of Bayesianism as a purely diachronic constraint, and to take synchronic norms like probabilism to be independent of Bayesianism proper.

<sup>11</sup>Here is one way of characterizing a subject's epistemic states which allows for this simplification. 1. Take there to be something like a "same subjective state as" relation that partitions the space of alternatives, with alternatives centered on empty spacetime points, rocks, and the like, sharing a "null" state. Identify the content of a subject's current evidence with the set of alternatives compatible with her current subjective state. 2. For agents like ourselves, and perhaps even ideally rational agents, subjective states will be something one has over a time-interval, not something one has at a particular time. To accommodate this, allow for alternatives that are ordered triples of a world, time-*interval* and individual. To believe that such an alternative is your own is to believe that your current subjective state occupies that (entire) time interval. 3. Assume that the time-intervals occupied by the alternatives of a subject with non-null subjective states are contiguous, have non-empty intersections, and are not densely ordered. 4. Require the epistemic successor relation (described in section 3.1) to hold between alternatives with non-null subjective states.

Given this picture, we can divide a subject's alternatives into those with null subjective states and those with non-null subjective states. Since the only alternatives we care about for the purposes of sequential updating rules are those the epistemic successor relation can hold between, we can ignore the alternatives with null subjective states. The alternatives we care about, those with non-null subjective states, will be located at time-intervals of the kind described above. If we "tag" these time-intervals with the earliest time in that interval, then the alternatives with non-null subjective states can be indexed by these discrete time tags, and these times will correspond to when the agent gets new evidence. Using these time "tags" to label the relevant alternatives, we get the picture of alternatives described in the text.

## 2.3 Internalism

We can distinguish between *internalists* and *externalists* about a given kind of epistemic norm. Like their cousins, the internalist and externalist positions regarding justification, these terms are not precise: one can flesh out the distinction between these positions in a number of ways. We can provide a broad characterization of these positions by framing them as generic supervenience claims of the following form:

**Internalism<sub>x</sub>:** The facts that determine whether a subject satisfies these kinds of norms supervene on *x*.

**Externalism<sub>x</sub>:** The facts that determine whether a subject satisfies these kinds of norms do not supervene on *x*.

We can set up the distinction between internalism and externalism in different ways, depending on how we fill in *x*. For example:

1. *x* = the subject's intrinsic state,
2. *x* = the subject's intrinsic mental states,
3. *x* = the intrinsic mental states the subject has access to,
4. *x* = the intrinsic mental states the subject can be held responsible for,
5. etc.

For the purposes of our discussion, the relevant kind of internalism will usually be something like 3: the facts that determine whether a subject satisfies these norms supervene on the intrinsic mental states the subject has access to.

Two comments before we proceed. First, the debate between internalists and externalists need not be a fight for hegemony. One can be an internalist about one kind of norm and an externalist about another. Likewise, if one allows for different kinds of epistemic rationality, one can simultaneously hold that the norms of one species of rationality should be internalist, while the norms of another should be externalist.

Second, I'll assume that one's current credences and evidence are "internal" features of a subject. A consequence of this assumption is that the tension between conditionalization and internalism will not arise in cases where the subject knows what her previous credences were.<sup>12</sup> In these cases, her previous credences will supervene on her current ones. Since her current credences and evidence are internal by assumption, the facts that determine whether conditionalization holds will supervene on things that are internal.

The assumption that one's current credences are internal presupposes that there is some notion of "belief" according to which beliefs have narrow content. Extreme externalists about mental content will deny this. But if we adopt this extreme position, then many of the cases discussed in the literature dissolve. So for the purposes of this paper, I'll assume that it makes sense to attribute credences with narrow content to a subject.

## 3 Continuity, Internalism and Uniqueness

As we've seen, a number of interesting issues arise concerning the canonical formulation of Bayesianism. We'll look at one of these issues—extending Bayesianism to accommodate

---

<sup>12</sup>In this paper I use "know" as shorthand for "certain of and right about". (So *a* knows *x* iff *a* is certain of *x* and *x* is true.)

self-locating beliefs—in the next section. In this section we will look at three different questions. First, what notion of continuity should Bayesianism employ? Second, how should we reconcile Bayesianism with internalism, if at all? Third, how should we extend Bayesianism to accommodate non-unique predecessors?

In order to evaluate these matters individually, I'll put the other issues aside while looking at each one. So I'll restrict my attention to cases in which subjects have unique predecessors when discussing the first and second questions. Likewise, I'll restrict my attention to cases where the tension between internalism and conditionalization doesn't arise—cases in which the subject knows what her previous credences were—when examining the first and third questions. Finally, I'll assume we have a fixed and unproblematic notion of continuity to work with when looking at the second and third questions.

### 3.1 Continuity

Diachronic credence constraints, such as updating rules, require something akin to a notion of personal identity. These constraints make claims about how a subject's credences at different times should be related. And this requires a way of picking out the same subject at two different times.

Of course, this way of picking out subjects need not mirror the relation of personal identity that metaphysicians are interested in. It just needs to track the sense of “same person as” that is relevant to these kinds of epistemic norms. I'll call this relation *epistemic continuity*, though I will not assume that it corresponds to psychological or doxastic continuity in any intuitive sense.

What notion of epistemic continuity should we employ? This is a deep and interesting question, and not one that I have a good answer to. I find myself most attracted to two views: a view which identifies epistemic continuity with the standard personal identity relation, and a view which characterizes epistemic continuity in terms of something like “psychological progression”.<sup>13</sup> But these are tentative suggestions, at best, and I will not assume either view in what follows.

Here are some features that I will take every epistemic continuity relation to have. Any notion of epistemic continuity can be characterized in terms of an *epistemic successor* relation. An epistemic successor relation is an irreflexive and anti-symmetric relation that holds between alternatives. This relation only holds between temporally adjacent alternatives located at the same world. I will not, however, assume that it only holds between alternatives centered on the same individual. So if an alternative  $c = (w, t, i)$  has a successor, then the successor must be an alternative of the form  $c' = (w, t + 1, j)$ .

Given an epistemic successor relation, we can define the *epistemic predecessor* relation as the inverse of this relation. So if one alternative is an epistemic successor of another, the latter is the epistemic predecessor of the former. We can then characterize the corresponding *epistemic continuity* relation as the symmetric relation we obtain by taking the closure of the epistemic successor and predecessor relations. So two alternatives are

---

<sup>13</sup>Or, better, an account which takes “natural psychological progression” to be a sufficient condition for succession, but which allows other considerations (physical continuity, degree of psychological similarity, etc.) to step in if no perfectly natural progressions can be found.

In either case, note that “psychological progression” need not proceed in lock step with temporal progression. When it does not, we cannot assume that successors are always temporally adjacent to and later than their predecessors (as I assume in the text and in footnote 11).

epistemically continuous *iff* one can construct a chain of epistemic successor/predecessor relations between them.<sup>14</sup>

We've cashed out epistemic continuity in terms of epistemic succession. So our notion of epistemic continuity hangs on how we characterize epistemic succession. As I noted above, I have no account of epistemic succession to offer. But it will be difficult to go through examples without some account of epistemic succession to employ. So in most of what follows, I'll adopt a default notion of succession with the following features.

In ordinary cases, if two temporally adjacent alternatives belong to the same individual, then the latter will be an epistemic successor of the former. In extraordinary cases, such as cases of fission or fusion, the subjects that result from fission or fusion will be epistemic successors of the subjects that underwent the process. So if an individual fissions into two "fissiles", each will be an epistemic successor of the original alternative. Likewise, if two individuals fuse into a single individual, the resulting alternative will be an epistemic successor of each of the original alternatives. (I'll come back to some reasons why one might *not* want to adopt this notion of continuity in section 4.3.)

In preparation for the discussion to come, it will be convenient to introduce some terminology and notation.

1. Given a *de se* proposition  $a$ , let  $es(a)$  be the set of epistemic successors of the alternatives in  $a$ . Likewise, let  $ep(a)$  be the set of epistemic predecessors of the alternatives in  $a$ .

2. Lewis employs the term "doxastic" to indicate possibilities a subject believes might be her own. We can use this terminology with respect to epistemic successors and predecessors as well. A subject's *doxastic epistemic successors* are the alternatives that she believes might be her epistemic successors; i.e., the epistemic successors of her doxastic alternatives. Likewise, her *doxastic epistemic predecessors* are the alternatives she believes might be her epistemic predecessors; i.e., the epistemic predecessors of her doxastic alternatives.<sup>15</sup>

3. Finally, it will become convenient later to have an extension of the notion of doxastic epistemic successors that includes 'dummy' successors for alternatives without successors; i.e., alternatives who will die.<sup>16</sup> I'll call this—the union of the epistemic successors of a subject's doxastic alternatives and a set of dummy successors for doxastic alternatives which don't have successors—the subject's *extended doxastic epistemic successors*.

## 3.2 Internalism

Much of the recent Bayesianism literature has implicitly relied on internalist intuitions. This is especially prevalent in the literature on self-locating belief, but it comes up in other

---

<sup>14</sup>This entails that epistemic continuity is a transitive relation. If one would like to avoid this consequence, for reasons like those discussed by Lewis (1983), then one could take the epistemic continuity relation to be primitive as well.

<sup>15</sup>Couldn't a subject have mistaken beliefs about what the successors/predecessors of her doxastic alternatives are? This would require beliefs whose content is more fine-grained than can be represented using sets of alternatives. And, as noted in section 2.1, we're restricting our attention to beliefs whose content can be represented by sets of alternatives.

<sup>16</sup>Formally, we can take the dummy successor  $c'$  of an alternative  $c = (w, t, i)$  to be the ordered triple  $(w, t+1, i)$ . Although this ordered triple is well-defined, it need not correspond to a genuine alternative since the individual  $i$  need not exist at that time and world. So we need to be sure that we don't treat dummy successors as genuine possibilities.

contexts as well.<sup>17</sup>

For example, consider the following case from Arntzenius (2003):

*Shangri La*: “There are two paths to Shangri La, the path by the Mountains, and the path by the Sea. A fair coin will be tossed by the guardians to determine which path you will take: if Heads you go by the Mountains, if Tails you go by the Sea. If you go by the Mountains, nothing strange will happen: while traveling you will see the glorious Mountain, and even after you enter Shangri La you will forever retain your memories of that Magnificent Journey. If you go by the Sea, you will revel in the Beauty of the Misty Ocean. But, just as soon as you enter Shangri La, your memory of this Beauteous Journey will be erased and be replaced by a memory of the Journey by the Mountains.”<sup>18</sup>

Arntzenius takes this case to provide a counterexample to conditionalization. He reasons as follows. Consider what our credence in heads should be at different times if the coin does, in fact, land heads. Our credence before the journey should be  $1/2$ , since the only relevant information we have is that the coin is fair. Our credence after we set out and see that we are traveling by the mountains should be 1, since this reveals the outcome of the coin toss. But once we pass through the gates of Shangri La, Arntzenius argues, our credence in heads should revert to  $1/2$ : “for you will know that you would have had the memories that you have either way, and hence you know that the only relevant information that you have is that the coin is fair”.<sup>19</sup> But this is not what conditionalization prescribes. Since conditionalization never reduces our credence in propositions we’re certain of, conditionalization will require our credence in heads to remain 1. Arntzenius concludes that since our credence in heads should be  $1/2$ , conditionalization cannot be correct.

The plausibility of Arntzenius’ counterexample hangs on internalist intuitions. Consider the reason for rejecting conditionalization’s demand that our credence remain 1: we no longer know the outcome of the coin toss, nor what our prior credences were, and we can’t require our credences to adhere with information we don’t have. But why think that what we require of our credences should be a function of information we have? This makes sense if we’re internalists, and we think the facts that determine how these norms constrain our credences should supervene on the information we have access to. But this line of thought loses its force if we assess conditionalization as externalists. If we take conditionalization to be a characterization of a kind of coherence relation between credences at different times, for example, then there’s no reason to think that our (in)ability to access past information is relevant.

Arntzenius’ assessment of the Shangri La case is motivated by applying internalist intuitions to conditionalization. But there’s a deep tension between internalism and diachronic credence constraints, like conditionalization. Diachronic credence constraints place restrictions on what our current credences can be, relative to our credences at other times. But our credences at other times are *external* to our current state, in any of the

---

<sup>17</sup>For an example from the literature on self-locating beliefs, consider a variant of the sleeping beauty case discussed by Elga (2000), where the subject is told the outcome of the coin toss before being put back to sleep. What should her credence in tails be when she wakes up Tuesday morning? It’s generally assumed that her credence on Tuesday should be the same as her credence on Monday morning, and thus that she should no longer have a credence of 1 in tails. But it’s hard to justify this assumption without appealing to internalist intuitions.

<sup>18</sup>Arntzenius (2003), p.356.

<sup>19</sup>Arntzenius (2003), p.356. “Memory” is being used in a non-factive sense here, of course.

senses relevant to internalism: they needn't supervene on what we currently have access to, our current mental or intrinsic states, and so on. So internalism and diachronic credence constraints are incompatible.

Arntzenius sets up the Shangri La case as an argument against conditionalization. But we can use it as a blueprint for generating arguments against any kind of diachronic credence constraint. Just set up a case where the relevant diachronic facts aren't determined by our intrinsic or mental states, or what we have access to. Then employ internalist intuitions to argue that this constraint is unreasonable, and conclude that this is a *reductio* of the diachronic credence constraint in question.

In light of this conflict between internalism and diachronic credence constraints, proponents of Bayesianism have two options. First, they can dismiss the internalist intuitions in question and much of the literature that accompanies it. Second, they can try to find a way of making Bayesianism, or something very much like it, compatible with these kinds of internalist intuitions. Without taking a stand on which of these options the Bayesian should adopt, let's explore how one might pursue the second option.

In order to construct an internalist version of an externalist constraint, we need to do two things. First, to make the constraint internalist, we need to replace whatever external features the original constraint appeals to with internal surrogates. Second, we need to ensure that the new constraint appropriately resembles the original one. For example, we might demand that the new constraint reduce to the original one in the appropriate cases.

In this case, a natural proposal is to replace conditionalization with the requirement that our credences equal the sum of the values conditionalization would recommend given  $cr$ , weighted by our credence that  $cr$  is our prior credence function. We can express this requirement as follows:

$$cr_e(a) = \sum_i cr_e(\langle cr = p_i \rangle) \cdot p_i(a|e), \text{ if defined,} \quad (2)$$

where  $i$  ranges over the space of probability functions, and  $\langle cr = p_i \rangle$  is the proposition that our previous credence function was  $p_i$ .<sup>20</sup> This constraint is compatible with internalism because it is a function of our current credences and our current evidence. And it reduces to standard conditionalization in cases where we know what our previous credences were.

The effect of (2) is to require our credences to be a mixture of the credences that we think conditionalization might have recommended. One might grant that (2) is a plausible

---

<sup>20</sup>This formula is reminiscent of Van Fraassen's (1995) Reflection Principle. Many have worried that the Reflection Principle is unreasonable in cases in which the subject believes her future credences are irrational (see Christensen (1991), Weisberg (2007)). Since the subject believes her future credences are epistemically defective, it seems she shouldn't take those credences to constrain her current ones, as Reflection requires. Similarly, one might doubt that (2) yields reasonable constraints in cases in which the subject believes her prior credences were irrational. Since the subject believes her prior credences were epistemically defective, it seems she shouldn't use those credences to constrain her current ones, as (2) requires.

To the extent to which this is a worry, it isn't a worry for (2) *per se*, but rather a worry for conditionalization. Suppose the subject knows what her prior credences were, so we can dispense with (2) and just use conditionalization. If the subject believes her prior credences were irrational, the same worry arises: since the subject believes her prior credences were epistemically defective, it seems she shouldn't use those credences to constrain her current ones, as conditionalization requires. (Likewise, if we employ the generalized version of (2), (6), and we replace conditionalization with an updating rule  $R$  which avoids these worries, then the problem will no longer arise.) So it is conditionalization, not the proposed extension, that is the source of the worry. (See Christensen (2000) for a discussion related to these issues.)

constraint, but take it to be too weak to take the place of conditionalization. After all, when judged by the standards of a genuinely diachronic constraint like conditionalization, it is a weak constraint: it allows you to radically shift your credences from one time to the next, so long as your credences at each time self-cohere in the manner required and are compatible with your evidence. But one can't expect more from a rule that isn't genuinely diachronic. And a rule can't be genuinely diachronic if it's going to be compatible with internalism.

To get a better feel for the strengths and weaknesses of (2), let's apply it to the Shangri La case. Let the evidence  $e$  that we get after walking through the gate be compatible with only two possibilities: the heads and tails possibilities described by Arntzenius. Let  $cr^h$  and  $cr^t$  be the credence functions you would have had before walking through the gate if the coin had landed heads or tails, respectively. According to (2), our credences should satisfy the following constraint:

$$cr_e(a) = cr_e(\langle cr = cr^h \rangle) \cdot cr^h(a|e) + cr_e(\langle cr = cr^t \rangle) \cdot cr^t(a|e). \quad (3)$$

Substituting the proposition that the coin landed heads for  $a$  yields:

$$\begin{aligned} cr_e(h) &= cr_e(\langle cr = cr^h \rangle) \cdot 1 + cr_e(\langle cr = cr^t \rangle) \cdot 0 \\ &= cr_e(\langle cr = cr^h \rangle). \end{aligned} \quad (4)$$

I.e., our credence in heads should be equal to our credence that our previous credence function was  $cr^h$ . But since our previous credence function was  $cr^h$  if and only if the coin landed heads, this is no constraint at all! So it seems that (2) is too weak to tell us anything useful.

But (2) is not as weak as it first appears. After walking through the gates of Shangri La, one presumably has many doxastic worlds at which the coin lands heads. And one's credence in heads will be distributed among these doxastic heads worlds in a particular way. (2) will require your credence in heads to be distributed among these worlds in the same way as  $cr^h$  distributes them: if  $cr^h$  assigns an equal credence to all of these worlds, for example, then you must as well. Likewise, the manner in which you distribute your credence in tails among the tails worlds must be the same as  $cr^t$  distributes them. So while (2) doesn't tell you anything about how to distribute your credence between heads and tails, that's the *only* thing it doesn't tell you: once we settle that, it will fix your credences in everything else. For example, if one thinks that the Principal Principle should require your credence in heads to be 1/2 in this case, then the conjunction of (2) and the Principal Principle will completely determine your credences.

One still might like the extension of conditionalization itself to require our credence in heads to be 1/2. But the constraint that (2) imposes is as strong as we can expect from a synchronic surrogate for conditionalization. To get the result that our credence should return to 1/2, we need to appeal to some further principle, like the Principal Principle or an Indifference Principle.

Note that the  $cr_e(\langle cr = p_i \rangle)$  term in (2)—your credence that your previous credences were  $p_i$ —is, in fact, a kind of self-locating belief, and a belief that makes implicit use of epistemic continuity. That said, this proposal is orthogonal to the question of how we should update self-locating beliefs. We can see both of these facts more clearly by reformulating the rule in a more general way. Let  $R(a, e, cr_i)$  be a generic sequential updating

rule. We can formulate the internalist version of  $R$  as:

$$cr_e(a) = \sum_{(i|c_i \in \Omega)} cr_e(es(c_i)) \cdot R(a, e, cr_i), \quad \text{if defined,} \quad (5)$$

where  $cr_i$  is the credence function of alternative  $c_i$ , and  $\Omega$  is the space of alternatives. I.e.,  $cr_e(a)$  should be the sum of: your credence that you're a successor of some alternative  $c_i$ , times the credence in  $a$  given  $e$  that  $R$  prescribes to someone with  $c_i$ 's credences.

This formulation makes the role of epistemic continuity and self-locating beliefs explicit. It also shows that this proposal is independent of how we should update self-locating beliefs, since we can plug any updating rule we like in place of  $R$ .

Although our discussion has focused on the tension between internalism and updating rules which make use of one's previous credences, there are other ways in which internalism and updating rules can conflict. For example, one might argue that there are cases in which the subject's current evidence won't be "internal" in the relevant sense either. Or one might adopt a different kind of updating rule which doesn't make use of one's previous credences, but does make use of some other arguments that are in tension with internalism. As long as we grant that one's current credences are internal, we can extend (2) in order to allow for these other kinds of cases as well. Let  $cr'$  be the subject's new credence function, and let  $R(a, x, y, \dots)$  be a generic updating rule of that employs arguments  $x, y, \dots$  to determine one's credence in  $a$ . We can formulate the internalist version of  $R$  as:

$$cr'(a) = \sum_i cr'(\langle x = x_i, y = y_i, \dots \rangle) \cdot R(a, x_i, y_i, \dots), \quad \text{if defined,} \quad (6)$$

where  $i$  ranges over possible sets of values for the arguments.

### 3.3 Uniqueness

The standard Bayesian account assumes that a subject has a single predecessor. This is implicit in the characterization of conditionalization: you take the subject's current evidence and her prior credence function, and use them to generate the subject's current credences. But this recipe breaks down if the subject has multiple predecessors. For example, if the subject is the fusion of two individuals with different credences, then there's no straightforward way to apply conditionalization. Arguably, this recipe also yields undesirable results if the subject has no predecessor, and thus no prior credence function to conditionalize on.

How should we extend conditionalization in order to accommodate these cases? For cases in which subjects don't have predecessors, I don't think conditionalization needs to be extended: the "if defined" clause of (1) does all the work required. If a subject doesn't have a predecessor, then (1) won't be defined, and it won't place any constraints on her credences. This seems to me to be the correct way to treat such cases. Conditionalization is a diachronic constraint—a constraint that ensures that the subject's current credences line up with her previous ones in the right way. If she has no predecessors, then there isn't anything that her current credences need to line up with.

What about the case in which you have multiple predecessors? A natural proposal is to require our credences to lie in the span of the credences conditionalization prescribes to our predecessors.<sup>21</sup> So if we have two predecessors, and conditionalization prescribes

---

<sup>21</sup>Another natural proposal is to take the average of the credences conditionalization prescribes. I pursue the

them a credence in  $a$  of  $1/3$  and  $1/2$ , respectively, then our credence in  $a$  should lie in the interval  $[1/3, 1/2]$ .

Let  $c_{@}$  be the subject's actual alternative after getting evidence  $e$ , and let  $cr_i$  be the credence function of alternative  $c_i$ . Since a mixture of probability functions will lie in the span of those functions, we can formulate this constraint as:<sup>22,23</sup>

$$cr_e(a) = \sum_{(i|c_i \in ep(c_{@}))} \alpha_i \cdot cr_{@}(a|e), \text{ if defined,} \quad (7)$$

for some  $\alpha_i$ 's such that  $\alpha_i \in [0, 1]$  and  $\sum_i \alpha_i = 1$ . I.e.,  $cr_e(a)$  should be a mixture of the credences conditionalization prescribes to your predecessors.

Note that, as with the internalist constraint discussed earlier, this proposal doesn't have anything in particular to do with conditionalization *per se*. We can formulate this constraint in terms of a generic sequential updating rule,  $R(a, e, cr_i)$ , as follows:

$$cr_e(a) = \sum_{(i|c_i \in ep(c_{@}))} \alpha_i \cdot R(a, e, cr_{@}), \text{ if defined,} \quad (8)$$

for some  $\alpha_i$ 's such that  $\alpha_i \in [0, 1]$  and  $\sum_i \alpha_i = 1$ .

### 3.4 Combining Proposals

We've looked at three issues that arise for sequential updating rules like conditionalization. And we've seen some proposals for how to resolve each of these issues. Now let's look at how to integrate these proposals.

Because each of these proposals is modular, integrating them is relatively straightforward. Let  $R(a, e, cr)$  be the updating rule of interest. We begin by selecting a notion of continuity. Then we identify the appropriate formulation of the internalist extension (6):<sup>24</sup>

$$cr_e(a) = \sum_{(i|c_i \in \Omega)} cr_e(c_i) \cdot R(a, c_i), \text{ if defined,} \quad (9)$$

and plug in the non-unique predecessor extension (8):<sup>25</sup>

$$R(a, c) = \sum_{(i|c_i \in ep(c))} \alpha_j \cdot R(a, e, cr_i), \text{ if defined,} \quad (10)$$

---

option described in the text instead for two reasons. First, combining the averaging approach with the internalist extension is more complicated (although not ultimately problematic). Second, the averaging approach is in tension with the Learning Principle that I advocate in section 5.

<sup>22</sup>A *mixture* (or convex combination) of probability functions  $p_1-p_n$  is a sum of the form  $\sum_i \alpha_i \cdot p_i$ , where the  $\alpha_i$ 's are coefficients such that  $\alpha_i \in [0, 1]$  and  $\sum_i \alpha_i = 1$ .

<sup>23</sup>Since (7) employs one's actual alternative, and this is something agents don't usually have access to, (7) is, of course, an externalist constraint. We'll see how to set-up an internalist version of this constraint in section 3.4.

<sup>24</sup>Why do I use formulation (6) of the internalist extension instead of formulation (5)? Because (5) requires us to plug in a rule of the form  $R(a, e, cr)$ , and the non-unique predecessors extension (8) won't yield a rule of this form. (Indeed, an extension to accommodate non-unique predecessors *can't* yield a rule of this form, because there won't be a unique prior credence function  $cr$  in cases where a subject has non-unique predecessors.) Instead, (8) yields a rule of the form  $R(a, c)$ , where  $c$  is the subject's current alternative. And if we want to get an internalist extension of this rule, we need to employ the appropriate version of (6):  $cr'(a) = \sum_i cr'(\langle c_{@} = c_i \rangle) \cdot R(a, c_i)$ .

<sup>25</sup>Where the  $e$  that appears in (10) is the evidence of alternative  $c$ .

to get a combination of all three:

$$cr_e(a) = \sum_{(i|c_i \in \Omega)} cr_e(c_i) \cdot \sum_{(j|c_j \in ep(c_i))} \alpha_j \cdot R(a, e, cr_j), \text{ if defined,} \quad (11)$$

for some  $\alpha_j$ 's that satisfy the usual constraints. Since the mixture of a mixture of probability functions is itself a mixture of these functions, we can reformulate (11) as:

$$cr_e(a) = \sum_{(i|c_i \in ep(c_j) \wedge cr_e(c_j) > 0)} \alpha_i \cdot R(a, e, cr_i), \text{ if defined,} \quad (12)$$

for some  $\alpha_i$ 's that satisfy the usual constraints. I.e.,  $cr_e(a)$  should be a mixture of the credences in  $a$  given  $e$  that  $R$  prescribes to your doxastic predecessors.

There is one hitch, however. Each of the extensions described earlier was formulated under the assumption that the issues which require the other extensions don't arise. But there may be questions which only come up when multiple issues are in play. And, indeed, when we combine these extensions, we find that a new question arises: how should we treat cases where the subject is unsure about whether she has predecessors?

There are two natural ways to answer this question. The first option is to assign alternatives without predecessors a credence of 0; i.e., to effectively ignore such possibilities. This is what (12) does. Since (12) doesn't sum over alternatives without predecessors, such possibilities are effectively ignored.

This is the natural choice if we think of the internalist extension of conditionalization in terms of a subject trying to accord her beliefs with conditionalization as best as she can. Consider a case where the subject is unsure about whether she was spontaneously created or whether she had some prior credence function  $cr$ . Given the stated goal, the subject should ignore the possibility that she might have been spontaneously created, and set her credences equal to those that conditionalization would prescribe if her prior credences were  $cr$ . If it turns out that she does have a predecessor whose credences were  $cr$ , then her beliefs will accord with conditionalization perfectly. And if it turns out that she has no predecessor, then her beliefs will also accord with conditionalization, since conditionalization won't impose any constraints on her credences. So from the perspective of trying to satisfy conditionalization as well as we can, one can argue that we're justified in ignoring no-predecessor possibilities.

The second option is to allow as much freedom in assigning credences to these possibilities as consistency allows. This is the natural choice if we think that it's unreasonable to demand that a subject be certain that she has predecessors when her evidence is equally compatible with the possibility that she doesn't.

We can implement this second option by adding a term to (12) which takes alternatives without predecessors into account. This new term can't employ  $R(a, e, cr)$ , since there's no previous credence function  $cr$  to apply it to, but we can replace it with a proxy function which effectively assigns a probability of 1 to the alternative in question. Let  $[c_j \in a]$  be a function which equals 1 if the statement in brackets is true, and 0 otherwise.<sup>26</sup> Then we can formulate the second option as:

$$\begin{aligned} cr_e(a) = & \sum_{(i|c_i \in ep(c_j) \wedge cr_e(c_j) > 0)} \alpha_i \cdot R(a, e, cr_i) \\ & + \sum_{(j|\neg \exists k(c_k \in ep(c_j)) \wedge cr_e(c_j) > 0)} \alpha_j \cdot [c_j \in a], \text{ if defined,} \end{aligned} \quad (13)$$

---

<sup>26</sup>[...] is the *Iverson bracket function*, which equals 1 if the statement in brackets is true, and 0 otherwise.

for some  $\alpha_i$ 's and  $\alpha_j$ 's such that  $\alpha_i, \alpha_j \in [0, 1]$  and  $\sum_i \alpha_i + \sum_j \alpha_j = 1$ . I.e.,  $cr_e(a)$  should be a mixture of: the credences in  $a$  given  $e$  that  $R$  prescribes to your doxastic predecessors and the proxy functions assigned to doxastic alternatives without predecessors.

Which of these two options should we adopt? Although both approaches are tenable, I think (13) better fits our intuitions about these kinds of cases. So in what follows, I'll adopt the second option.

## 4 Self-Locating Beliefs

Now let's look at self-locating beliefs. There has been a lot of discussion about how to extend Bayesianism in order to accommodate self-locating beliefs.<sup>27</sup> This discussion has been hampered, however, by the fact that many of the cases discussed raise tricky questions regarding continuity, internalism and uniqueness—issues that are independent of self-locating beliefs *per se*. For example, consider the case which has received the lion's share of the discussion in the literature, the sleeping beauty problem discussed by Elga (2000):

*Sleeping Beauty*: You know that you have been placed in the following experiment. Some researchers are going to put you to sleep for several days. They will put you to sleep on Sunday night, and then flip a fair coin. If heads comes up they will wake you up on Monday morning. If tails comes up they will wake you up on Monday morning and Tuesday morning, and in-between Monday and Tuesday, while you are sleeping, they will erase the memories of your waking. All of this will be done so as to make your evidential state upon waking the same, regardless of the day or the outcome of the coin toss.

This is an interesting case. But it is not the kind of case that we should begin our assessment of self-locating beliefs with. For the sleeping beauty case brings in a number of issues that complicate the ensuing analysis.

1. *Continuity*. How we assess the sleeping beauty case, and what kinds of issues arise, depends on the notion of continuity we employ. On most notions of continuity, issues arise regarding how to treat cases where subjects don't know what their previous credences were. And on some notions of continuity, issues arise regarding how to treat cases with multiple predecessors. So how we deal with the sleeping beauty case will depend in part on how we handle the issues concerning continuity raised in section 3.1.

2. *Internalism*. On most accounts of continuity, how we assess the sleeping beauty case will depend on how we treat cases where subjects don't know what their previous credences were. Suppose we use the default notion of epistemic continuity that I've been employing. When you wake up on Monday morning, you don't know whether it's Monday morning or Tuesday morning. As a result, you don't know whether your previous credences were those you had on Sunday night or (if the coin landed tails) those you had on Monday night.

---

<sup>27</sup>For example, see Arntzenius (2002), Arntzenius (2003), Bostrom (2007), Bradley (2003), Dorr (2002), Elga (2000), Elga (2004), Halpern (2005), Hitchcock (2004), Horgan (2004), Jenkins (2005), Kierland and Monton (2005), Kim (2009), Lewis (2001), Meacham (2008), Monton (2002), Titelbaum (2008), Weintraub (2004), and White (2006).

*Prima facie*, the credences you should adopt when you wake up on Monday morning are those prescribed by the appropriate sequential updating rule. But the prescriptions of sequential updating rules depend on what your previous credences were. And when you wake up on Monday morning, you don't know what your previous credences were, and thus don't know what credences such rules would prescribe you. From an internalist perspective, this is intolerable: given such a rule, what credences you should adopt will depend on information you don't have access to. So from an internalist perspective, we'll want to replace sequential updating rules with the appropriate synchronic surrogates. And how we do that will depend on how we decide to handle the issues concerning internalism raised in section 3.2.

3. *Non-Unique Predecessors*. On some accounts of continuity, how we assess the sleeping beauty case will depend on how we treat cases with multiple predecessors. Suppose we adopt a successor relation that tracks something like "psychological progression", such that an alternative  $c_2$  is a successor of  $c_1$  if  $c_1$  and  $c_2$  are located at the same world and  $c_2$  is in a mental state which is the "natural psychological progression" of  $c_1$ 's mental state. (Some reasons for adopting this kind of successor relation will come up in section 4.3.) Then the sleeping beauty case is a case in which you can have multiple predecessors.

To see this, suppose the coin lands tails. On Monday morning you will have a successor on both Monday night and Tuesday night, since both are in mental states that are natural psychological progressions of your mental state on Monday morning. Likewise, on Tuesday morning you will have a successor on both Tuesday night and Monday night, since both are in mental states that are natural psychological progressions of your mental state on Tuesday morning. Since both the Monday morning and Tuesday morning alternatives have the Monday night alternative as a successor, both the Monday morning and Tuesday morning alternatives are predecessors of your Monday night alternative. Thus, if the coin lands tails, you will have multiple predecessors on Monday night. And how we deal with your credence in this situation will depend in part on how we decide to handle the issues concerning multiple predecessors raised in section 3.3.

The sleeping beauty case brings in issues which are orthogonal to the question of how we should update self-locating beliefs. And unless we already know how to treat these other issues, as well as how to treat self-locating beliefs, it's unlikely that we'll be able to figure out the right way to assess the sleeping beauty case. In light of this, it seems prudent to put difficult cases like this one aside, and to begin by assessing the question of how to update self-locating beliefs in isolation.

So for most of what follows, I will restrict my attention to "pure" contexts: contexts where we're using the default notion of continuity, and where we're restricting ourselves to cases where the subject has a single predecessor and knows what her previous credences were. I will begin by discussing a natural thought about how to extend conditionalization in a sequential manner in order to accommodate self-locating beliefs. Then I'll explore some different ways of fleshing out this idea. Finally, I will come back to consider how to apply this proposal and the proposals discussed in the section 3 in concert.

## 4.1 First Steps

The standard Bayesian account provides a way to update *de dicto* beliefs, or beliefs about what the world is like. But in addition to beliefs about the what the world is like, we also have *de se* beliefs, beliefs about where we are in the world. How should we extend

Bayesianism in order to accommodate *de se* beliefs?

At a first pass, we might just try to replace the space of possible worlds with the space of possible alternatives, and leave the rest of the Bayesian account the same.<sup>28</sup> But this proposal is problematic. To see this, consider a subject who is looking at a clock that she knows to be accurate. Let  $\tau_i$  stand for the *de se* proposition that it's now time  $t_i$ . Since the subject is watching an accurate clock, any evidence  $e$  she gets at  $t_i$  will be such that  $e \Rightarrow \tau_i$ . Now suppose that at  $t_0$  her credence in  $\tau_0$  is 1; i.e.,  $cr(\tau_0) = 1$ . And suppose she gets evidence  $e$  at  $t_1$ , where  $e \Rightarrow \tau_1$ . Then it seems her credence at  $t_1$  in  $\tau_1$  should be 1. But if we apply conditionalization, this is not what we get:

$$cr_e(\tau_1) = cr(\tau_1|e) = \text{undefined}, \quad (14)$$

since  $e \Rightarrow \tau_1$ ,  $cr(\tau_1) = 0$  and thus  $cr(e) = 0$ .

More generally, the problem is that rational subjects should be able to change their credences in *de se* propositions like  $\tau$  from 0 to something greater than 0 in a systematic and well-regulated way. But conditionalization cannot provide such guidance. If  $cr(\tau) = 0$ , then conditionalization will either prescribe  $cr_e(\tau) = 0$  (if  $cr(e) > 0$ ), or offer no prescription at all (if  $cr(e) = 0$ ).

The source of this problem stems from the funny role of time. Conditionalization tries to impose a kind of conformity on the beliefs of a subject's alternatives at different times. But with respect to beliefs *about* time, this is problematic. Since these alternatives are located at different times, we don't want to require *conformity* with respect to their beliefs about what time it is. Rather, we want to allow their beliefs about time to change as time changes.

Let's see how we might do this. According to conditionalization, if my predecessor believed that  $a$  is true, then I should believe that  $a$  is true. As we've seen, this seems false when applied to *de se* beliefs, since  $\tau_0$  can be true for my predecessor, and yet false for me. But something nearby seems true. Namely, if my predecessor believed that  $a$  would be true for his successor (me), then I should believe that  $a$  is true. So if my predecessor believed that  $\tau_1$  would be true for me, then I should believe that  $\tau_1$  is true.

Now, to say that  $a$  would be true for an alternative  $x$ 's successor is just to say that  $x$  is the predecessor of an alternative for which  $a$  is true. I.e.,  $a$  is true for  $x$ 's successor *iff*  $ep(a)$  is true for  $x$ . So we can reformulate the claim given above as follows: if my predecessor believed that  $ep(a)$  is true, then I should believe that  $a$  is true.

So far we've left evidence out of it, but similar reasoning applies here. If my predecessor believed that  $a$  would be true for his successor if  $e$  was true for his successor (me), and I get  $e$  as evidence, then I should believe that  $a$  is true. Reformulating this claim in the same way as before yields: if my predecessor believed that  $ep(a)$  is true given that  $ep(e)$  is true, and I get  $e$  as evidence, then I should believe that  $a$  is true. More generally:

$$cr_e(a) = cr(ep(a)|ep(e)), \quad \text{if defined.} \quad (15)$$

Call this *predecessor conditionalization*.<sup>29</sup>

To get a feel for how predecessor conditionalization works, let's look at some examples. For ease of exposition, I'll assume that all of the alternatives in these examples have both predecessors and successors. (We'll return to relax these assumptions in a moment.)

<sup>28</sup>One finds David Lewis recommending this approach in Lewis (1979).

<sup>29</sup>A version of predecessor conditionalization is discussed in Meacham (2007). A proposal similar in spirit, but which differs in a number of interesting ways, is defended by Namjoong Kim (2009).

First example. Consider the time-changing case discussed above, where the subject is watching a clock that she knows to be accurate. Suppose she gets evidence  $e$  at  $t_1$ , and suppose that some of her doxastic alternatives are predecessors of alternatives compatible with  $e$ ; i.e.,  $cr(ep(e)) \neq 0$ . What should her credence in  $\tau_1$  be at  $t_1$ ? Applying (15) yields:

$$cr_e(\tau_1) = cr(ep(\tau_1)|ep(e)) = \frac{cr(ep(\tau_1) \wedge ep(e))}{cr(ep(e))}. \quad (16)$$

Since  $e$  is a subset of  $\tau_1$ ,  $ep(e)$  will be a subset of  $ep(\tau_1)$ , and thus:

$$cr_e(\tau_1) = cr(ep(\tau_1)|ep(e)) = \frac{cr(ep(\tau_1) \wedge ep(e))}{cr(ep(e))} = \frac{cr(ep(e))}{cr(ep(e))} = 1. \quad (17)$$

So when the subject sees the clock change to  $t_1$ , she becomes certain that it is now  $t_1$ .

Second example. Consider a subject who looks at a clock she knows to be accurate at  $t_0$ , but who then stops looking at the clock. Suppose she knows what piece of evidence,  $e$ , she will get next. So at  $t_0$  all of her doxastic alternatives are predecessors of alternatives compatible with  $e$ ; i.e.,  $cr(ep(e)) = 1$ . Further suppose that she is unsure as to how much time will pass before she gets that evidence. So let her credence at  $t_0$  in being the predecessor of an alternative located at  $t_1$  be  $1/2$ ; i.e.,  $cr(ep(\tau_1)) = 1/2$ . And suppose that the time at which she actually gets  $e$  is  $t_1$ . What will her credence in  $\tau_1$  be at  $t_1$ ? Applying (15) yields:

$$cr_e(\tau_1) = cr(ep(\tau_1)|ep(e)) = cr(ep(\tau_1)) = 1/2. \quad (18)$$

After the subject stops looking at the clock, she gets evidence that is compatible with successors at different times, and so she becomes uncertain of what time it is. That is, she loses track of the time.

Third example. Consider again the clock-watching subject from the first example. Suppose that at  $t_0$  the subject thinks that either Mt. St. Helens will erupt now, at  $t_0$ , or that Mt. St. Helens will erupt a moment from now, at  $t_1$ . And suppose that she thinks these possibilities are equally likely. I.e., letting  $m_t$  be the *de dicto* proposition that Mt. St. Helens erupts at  $t$ ,  $cr(m_{t_0}) = cr(m_{t_1}) = 1/2$ .

Now consider the *de se* proposition that Mt. St. Helens has erupted,  $d$ .<sup>30</sup> At  $t_0$ , the subject's credence in  $d$  will be equal to her credence in Mt. St. Helens erupting at or before  $t_0$ ; i.e.,  $cr(d) = \sum_{(t|t \leq t_0)} cr(m_t)$ . Since  $m_{t_0}$  is the only  $m_t$  in this sum in which she has a positive credence,  $cr(d) = \sum_{(t|t \leq t_0)} cr(m_t) = cr(m_{t_0}) = 1/2$ .

Now suppose that the subject gets her next evidence  $e$  at  $t_1$  (where  $e \Rightarrow \tau_1$ ). And suppose that she knows this, so that at  $t_0$  all of her doxastic alternatives are predecessors of alternatives compatible with  $e$ ; i.e.,  $cr(ep(e)) = 1$ . What should her credence in  $d$ , that Mt. St. Helens has erupted, be at  $t_1$ ? Applying (15) yields:

$$cr_e(d) = cr(ep(d)|ep(e)) = cr(ep(d)) = cr(d-), \quad (19)$$

where  $d-$  is the *de se* proposition that Mt. St. Helens has either erupted or will erupt in a moment.<sup>31</sup> At  $t_0$  her credence in  $d-$  will be equal to her credence in Mt. St. Helens erupting

<sup>30</sup>So  $d$  is true for alternatives which are both (i) located at time  $t$  and (ii) located at a world where Mt. St. Helens erupts at some time  $t' \leq t$ .

<sup>31</sup>So  $d-$  is true for alternatives which are both (i) located at time  $t$  and (ii) located at a world where Mt. St. Helens erupts at some time  $t' \leq t + 1$ .

at or before  $t_1$ ; i.e.,  $cr(d^-) = \sum_{(t|t \leq t_1)} cr(m_t)$ . Since  $m_{t_0}$  and  $m_{t_1}$  are the only  $m_t$ 's in this sum in which she has a positive credence,  $cr(d^-) = \sum_{(t|t \leq t_1)} cr(m_t) = cr(m_{t_0}) + cr(m_{t_1}) = 1$ . Putting this together, we get:

$$cr_e(d) = cr(ep(d)|ep(e)) = cr(ep(d)) = cr(d^-) = cr(m_{t_0}) + cr(m_{t_1}) = 1. \quad (20)$$

So when the subject sees that it's  $t_1$ , her credence that Mt. St. Helens has erupted will change from  $1/2$  to  $1$ . Since she's confident that Mt. St. Helens will erupt at either  $t_0$  or  $t_1$ , this is how it should change.

So far, we've been assuming that all of the alternatives in question have predecessors and successors. Now let's see what happens when we relax these assumptions.

First, consider alternatives without successors. These alternatives may seem to raise a worry for predecessor conditionalization. Suppose some of your previous alternatives had no successors. Since  $a$  and  $e$  won't include successors of these alternatives, they won't appear in  $ep(a)$  and  $ep(e)$ , and thus won't have any effect on what your credences are.

In a sense, this is right: such alternatives will automatically be ruled out of consideration. But that's okay. The very fact that you're around means that none of these alternatives were your predecessors. So these alternatives *should* be ruled out.

To see this, consider the same clock watching subject as before. This time, let us suppose that there will be a fair coin tossed at  $t_0$ , and that the subject will be killed at  $t_1$  if and only if the coin lands tails. Now suppose the subject is still alive at  $t_1$ , and gets evidence  $e$ , where  $e \Rightarrow \tau_1$ . What should her credence be that the coin landed heads?

For simplicity, let us assume that the only alternatives without successors are the subject's tails alternatives at  $t_0$ , and that the two outcomes of the coin toss,  $h$  and  $t$ , partition the space of possibilities. It follows that  $cr(ep(h)) = cr(h)$ , since all of the subject's doxastic alternatives at  $t_0$  in  $h$ -worlds will be predecessors of alternatives in  $h$ -worlds. Assuming that all of the subject's doxastic alternative at  $t_0$  that have successors are predecessors of alternatives compatible with  $e$ , it will be the case that  $cr(ep(e)) = cr(\text{alive at } t_1) = cr(h)$ . Plugging this into (15) yields:

$$cr_e(h) = cr(ep(h)|ep(e)) = cr(ep(h)|h) = cr(h|h) = 1. \quad (21)$$

So when the subject finds herself alive at  $t_1$ , her credence in heads will become  $1$ .

Second, consider alternatives without predecessors. These alternatives also seem to raise a worry for predecessor conditionalization. Suppose that some of the alternatives in  $a$  and  $e$  were just created, and so have no predecessors. Then  $ep(a)$  and  $ep(e)$  won't include predecessors of these alternatives, and the existence of these alternatives won't have any bearing on the credences (15) prescribes.

But recall that we're restricting our attention to "pure" cases here. And these kinds of cases are not pure, since the subject won't know whether she previously had credences, and *a fortiori* won't know what her previous credences were. To accommodate these kinds of cases, we need to extend predecessor conditionalization. We'll look at how to do this in section 4.3.

## 4.2 Fleshing It Out

The version of predecessor conditionalization given by (15) is good enough to handle ordinary cases, like the three examples considered above. But it runs into problems in cases

where we might have multiple successors, like cases of fission. For example, consider the following simple case. Suppose that, at  $t_0$ , the subject knows that her alternative is  $c$ , i.e.,  $cr(c) = 1$ . And suppose that, at  $t_1$ , she will fission into two fissiles who occupy evidentially identical situations. Let  $f_1$  and  $f_2$  be the  $t_1$  alternatives centered on the first and second fissiles, respectively.

If we apply (15) in order to determine her credences at  $t_1$  we get:

$$cr_{f_1 \vee f_2}(f_1) = cr(ep(f_1)|ep(f_1 \vee f_2)) = cr(c|c) = 1, \quad (22)$$

$$cr_{f_1 \vee f_2}(f_2) = cr(ep(f_2)|ep(f_1 \vee f_2)) = cr(c|c) = 1. \quad (23)$$

Since  $f_1$  and  $f_2$  are mutually exclusive, this assignment is probabilistically incoherent.

The natural solution is to add a normalization factor that adjusts for cases with multiple successors. Reformulating predecessor conditionalization as a sum over alternatives, this yields:

$$cr_e(a) = \sum_{(i|c_i \in a)} cr(ep(c_i)|ep(e)) \cdot N_i, \quad \text{if defined.} \quad (24)$$

All that is left to do is to decide how to normalize (24), i.e., to decide what to stick in for  $N_i$ . At this point, two options present themselves. Let's look at each of them in turn.

#### 4.2.1 Local Predecessor Conditionalization

One option is to normalize at each world. Let  $\bar{c}$  pick out the world that alternative  $c$  occupies. Then we can formulate this version of predecessor conditionalization as:

$$cr_e(a) = \sum_{(i|c_i \in a)} cr(ep(c_i)|ep(e)) \cdot N_i, \quad \text{if defined,} \quad (25)$$

where the normalization factor  $N_i$  is:

$$N_i = \frac{cr(\overline{ep(c_i)}|ep(e))}{\sum_{(j|c_j \in \bar{c}_i)} cr(ep(c_j)|ep(e))}. \quad (26)$$

Since this approach effectively normalizes at each world, I'll call the resulting version of predecessor conditionalization *local predecessor conditionalization*, or LPC.

Although the formula expressing LPC is not particularly transparent, there is a convenient way to use diagrams to determine what a subject's credences should be according to LPC:

1. Consider the subject's predecessor,  $p$ .
2. Draw a box representing  $p$ 's extended doxastic epistemic successors,  $d$ .
3. Divide the width of the box into boxes representing the worlds in  $d$ , with the width of these boxes proportional to  $p$ 's credence in those worlds.
4. Divide the height of each world-box into boxes representing the alternatives in  $d$  at that world, with the height of these boxes proportional to  $p$ 's credence in the predecessors of these alternatives.
5. Eliminate every box incompatible with the subject's new evidence.
6. The subject's new credence in a possibility is proportional to the area of the box that represents it.

Note that if we remove step 4, this method will yield the standard Bayesian prescription. Steps 1 through 3 set up the subject's previous credences in doxastic worlds, step 5 eliminates those worlds incompatible with her evidence, and step 6 renormalizes her credences in the usual way.

To get a feel for this procedure, let's apply it to some examples. We saw above that the original temporal version of the sleeping beauty case is entangled with the issues we discussed in section 3. But given the default notion of continuity we've been employing, the fission version of the sleeping beauty case avoids these entanglements. So let's apply LPC to this case:

*Sleeping Beauty (fission)*: You know that you have been placed in the following experiment. A fair coin will be flipped, out of your sight. If the coin lands heads, then nothing will happen. If the coin lands tails, then you will be fissioned at  $t_1$  into two fissiles, both in situations evidentially equivalent to the situation you would have been in at  $t_1$  had the coin landed heads.

According to LPC, what should your credences be at  $t_1$ ? At  $t_0$ , your extended doxastic epistemic successors will consist of a lone successor at the heads worlds, and a pair of successors (the two fissiles) at the tails worlds. So we can draw the space of your extended doxastic epistemic successors at  $t_0$  like this (with the evidence compatible with each possibility in parentheses):

Original, $t_1$ ( $e$ )	Fissile 2, $t_1$ ( $e$ )
	Fissile 1, $t_1$ ( $e$ )
Heads	Tails

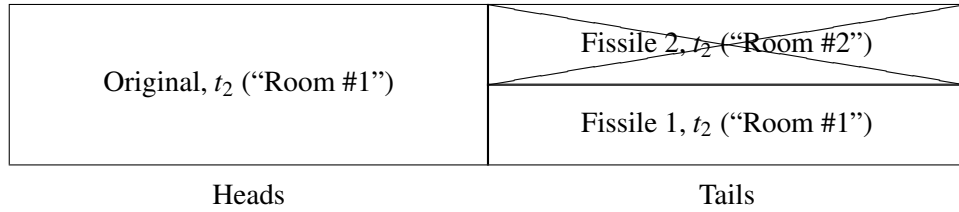
The widths of the heads and tails boxes are equal since your credence at  $t_0$  in heads and tails is equal. Likewise, the height of the Fissile 1 and Fissile 2 boxes are equal since your credence at  $t_0$  in their predecessor (the original at  $t_0$  in the tails world) is the same. Your evidence at  $t_1$  is compatible with all three possibilities, so we don't eliminate any of them. Since your credence at  $t_1$  in a possibility should be proportional to the area of the box representing it, we can conclude that your credence in the heads alternative should be  $1/2$ , and your credence in each of the fissiles should be  $1/4$ .

A second example. Take the same case as before, but now suppose that if fission is not performed at  $t_1$ , you will be placed in a particular room—room #1. If fission is performed at  $t_1$ , on the other hand, then the first fissile will be placed in room #1, and the second fissile will be placed in room #2. Finally, suppose that at  $t_2$  you will be told what room you're in. What should your credences become if you learn at  $t_2$  that you're in room #1?

At  $t_1$  your extended doxastic epistemic successors will again consist of a single alternative at the heads worlds and a pair of alternatives at the tails worlds. Your credence at  $t_1$  in the predecessors of these alternatives will be  $1/2$ ,  $1/4$ ,  $1/4$ , respectively, as we've just seen. So the box representing your doxastic successors will look much as before:



But this time your new evidence—that you’re in room #1—will be incompatible with one of the possibilities:



Assigning credences to the remaining possibilities in proportion to their area, we find that your credence at  $t_2$  that you’re at a heads world should become  $2/3$ , and your credence that you’re the first fissile at a tails world should become  $1/3$ .

#### 4.2.2 Global Predecessor Conditionalization

Another option is to normalize all of the possibilities together. We can formulate this version of predecessor conditionalization as:

$$cr_e(a) = \sum_{(i|c_i \in a)} cr(ep(c_i)|ep(e)) \cdot N_i, \text{ if defined,} \quad (27)$$

where the normalization factor  $N_i$  is:

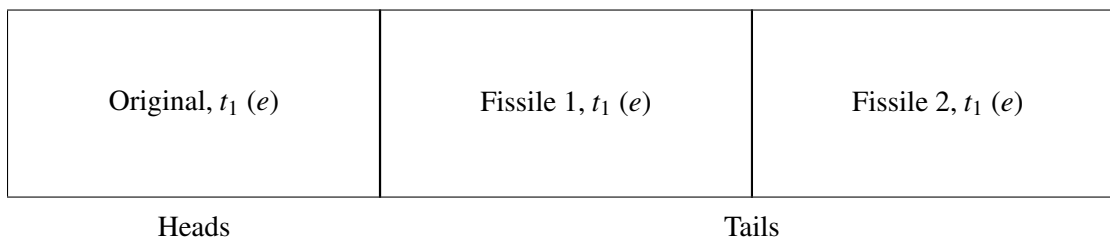
$$N_i = \frac{1}{\sum_{(j|c_j \in \Omega)} cr(ep(c_j)|ep(e))}. \quad (28)$$

Since this term normalizes all of the possibilities together, I’ll call the resulting kind of predecessor conditionalization *global predecessor conditionalization*, or GPC.

As before, there is a convenient way to use diagrams to determine what a subject’s credences should be according to GPC:

1. Consider the subject’s predecessor,  $p$ .
2. Draw a box representing  $p$ ’s extended doxastic epistemic successors,  $d$ .
3. Divide the width of the box into boxes representing the alternatives in  $d$ , with the width of these boxes proportional to  $p$ ’s credence in the predecessors of these alternatives.
4. Eliminate every box incompatible with the subject’s new evidence.
5. The subject’s new credence in a possibility is proportional to the area of the box that represents it.

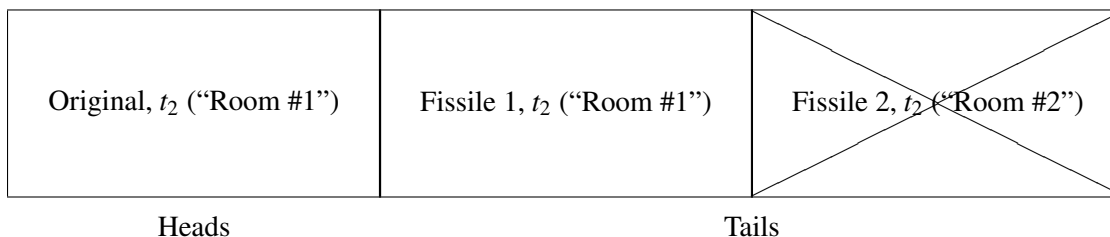
Let's apply this procedure to the same examples as before. First, consider the fission version of the sleeping beauty case. As before, at  $t_0$  your extended doxastic epistemic successors will consist of a lone successor at the heads worlds, and a pair of successors (the two fissiles) at the tails worlds. So we can draw the space of your extended doxastic epistemic successors at  $t_0$  like this (with the evidence compatible with each possibility in parentheses):



The width of the each box is equal since your credence at  $t_0$  in the predecessor of each of these alternatives is the same. Your evidence at  $t_1$  is compatible with all three possibilities, so we don't eliminate any of them. Assigning credences to the remaining possibilities in proportion to their area, we find that your credence at  $t_1$  in each possibility should be  $1/3$ .

Now consider the second example, where you'll be put in room #1 if the coin lands heads or if the coin lands tails and you're the first fissile, and you'll be put in room #2 if the coin lands tails and you're the second fissile. As before, you'll be told what room you're in at  $t_2$ . What should your credences be if you learn that you're in room #1?

Your extended doxastic epistemic successors at  $t_1$  will again consist of three alternatives, and your credence at  $t_1$  in each of the predecessors of these alternatives will be  $1/3$ . So the box representing your doxastic successors will look much as before. But this time your new evidence—that you're in room #1—will be incompatible with one of the possibilities:



Assigning credences to the remaining possibilities in proportion to their area, we find that your credence at  $t_2$  in each of the surviving possibilities should be  $1/2$ .

### 4.3 Extensions

We've looked at how LPC and GPC apply to a fission version of the sleeping beauty case. Now let's return the temporal version of this case:

*Sleeping Beauty:* You know that you have been placed in the following experiment. Some researchers are going to put you to sleep for several days. They will put you to sleep on Sunday night, and then flip a fair coin. If heads

comes up they will wake you up on Monday morning. If tails comes up they will wake you up on Monday morning and Tuesday morning, and in-between Monday and Tuesday, while you are sleeping, they will erase the memories of your waking. All of this will be done so as to make your evidential state upon waking the same, regardless of the day or the outcome of the coin toss.

What will LPC and GPC say about this case? As it stands, nothing. We've restricted LPC and GPC to "pure" contexts: contexts where we're using the default notion of continuity, and cases where you have a single predecessor and know what your previous credences were. And in this case you don't know what your previous credences were when you woke up. They were either those you had on Sunday night, or those you will have on Monday night, but you don't know which. So, as given, neither LPC nor GPC will apply to this case.

But we can plug LPC and GPC into the appropriate extension to get answers to this version of the sleeping beauty case. To make things easier, let's assume the default notion of continuity, so we can put continuity issues aside. Given this notion of continuity, the case won't involve multiple predecessors, so we can put those issues aside as well. Finally, given this notion of continuity, the case won't involve multiple successors. So LPC and GPC will yield the same results, and we needn't differentiate between them. Thus we only need to concern ourselves with one thing: the internalist extension of L/GPC specified by equation (5).

This extension tells us that your credences should be a mixture of those that "pure" L/GPC would prescribe you given each of the prior credence functions you might have had. Let's assume that the evidence you get when you wake up is compatible with only three alternatives: that it's Monday morning at the heads world, Monday morning at the tails world, and Tuesday morning at the tails world. So you have two possible prior credence functions: the credences you would have had on Sunday night at the heads and tails worlds, and the credences you would have had on Monday night at the tails world. So let's look at what L/GPC would assign you given each of these prior credence functions.

Given your Sunday night credences, your doxastic successors would consist of a Monday morning successor at the heads world and a Monday morning successor at the tails world. We can draw the space of your doxastic successors as:

Monday, ( <i>e</i> )	Monday, ( <i>e</i> )
Heads	Tails

Since the new evidence you get when you wake up doesn't eliminate either of these possibilities, L/GPC would prescribe you a credence of 1/2 in each.

Given your Monday night credences, you would have three doxastic successors: a Tuesday morning successor at the heads world, a Tuesday morning successor at the tails world, and a Wednesday morning successor at the tails world. In this case we don't need to bother working out the space of these doxastic successors, since the evidence you get when you wake up eliminates all but the Tuesday-and-Tails possibility. So L/GPC would prescribe you a credence of 1 in Tuesday-and-Tails.

Plugging these results into (5), we find that your credences when you wake up should be a mixture of  $(1/2, 1/2, 0)$  and  $(0, 0, 1)$  in Monday-and-Heads, Monday-and-Tails and Tuesday-and-Tails, respectively. This permits a number of different credence functions. For example, all of the following credences would be permitted:  $(1/2, 1/2, 0)$ ;  $(0, 0, 1)$ ;  $(1/3, 1/3, 1/3)$ , and so on. Indeed, the only substantive constraint this imposes is that your credence in Monday-and-Heads must be the same as your credence in Monday-and-Tails.<sup>32</sup>

In light of this case, we can see that the extended versions of LPC and GPC are *inegalitarian*: they treat the fission version of the sleeping beauty case differently from the temporal version. This inequality should not surprise us. LPC and GPC are diachronic credence constraints, whose prescriptions hang on a notion of continuity. As such, we should expect their prescriptions to be different when the continuity facts are different. And given the notion of continuity we're employing, the continuity facts in the fission case are different from the continuity facts in the temporal case. So it should be no surprise that LPC and GPC treat these cases differently.

In light of a given instance of inequality, we might react in one of two ways. First, we might argue that this inequality is justified because the cases in question should be treated differently. Second, we might argue that this inequality is unjustified because the cases in question should be treated in the same way. In that case, we'll want to look for a different notion of continuity, one which better captures our egalitarian intuitions. For example, we might want to adopt something like the notion of "psychological progression" discussed earlier, which treats the temporal and fission cases the same way.<sup>33</sup>

This brings us back to the first extension: changing the notion of continuity. Enacting such a change is essentially trivial. But figuring out what notion of continuity we should employ is not. And the notion of continuity we employ is key to how these cases get treated, and to whether they get treated in the same way.

In section 3.1 I mentioned that I find myself most attracted to two views: a view which identifies epistemic continuity with the standard personal identity relation, and a view which characterizes epistemic continuity in terms of something like "psychological progression".<sup>34</sup> The former view will yield an inequality treatment of these cases, while the latter view will not. Unfortunately, I have little to offer regarding which notion of continuity we should adopt. So I leave the matter open for further investigation.

---

<sup>32</sup>If we adopt something like the Indifference Principle proposed by Elga (2004), which requires a subject's credence in a world to be split evenly between her alternatives at that world, then the only credence assignment compatible with L/GPC in this case would be the  $(1/3, 1/3, 1/3)$  assignment. This is a fragile result, however, since if we change the notion of continuity we're employing (for one of the reasons discussed below, say) we'll generally get different answers.

<sup>33</sup>Strictly speaking, we can only expect these cases to be treated similarly until differences in the number of alternatives arise. If we want the two cases to be similar at all times, we need to modify the fission case slightly. I.e., we need the two fissions to fuse back together, in order to mirror the way the Monday and Tuesday night alternatives "fuse" (psychologically progress) into the same Wednesday morning alternative.

<sup>34</sup>Or, better, an account which takes natural psychological progression to be a sufficient condition for succession, but which allows other considerations to step in if no natural progressions exist.

## 5 The Learning Principle

So far, we haven't seen any reasons for preferring LPC or GPC. In this section, I present a reason for favoring LPC. I'll motivate a "Learning Principle" that places constraints on rational updating rules, and I'll show that LPC satisfies this principle. I'll end with a brief discussion about the special significance that LPC attaches to worldhood.

### 5.1 Motivating the Learning Principle

Let's return to the unextended versions of LPC and GPC. As before, we'll restrict our attention to "pure" cases, where subjects have a single predecessor and know what their previous credences were.

Consider again the fission version of the sleeping beauty case:

*Sleeping Beauty (fission)*: You know that you have been placed in the following experiment. A fair coin will be flipped, out of your sight. If the coin lands heads, then nothing will happen. If the coin lands tails, then you will be fissioned at  $t_1$  into two fissiles, both in situations evidentially equivalent to the situation you would have been in at  $t_1$  had the coin landed heads.

At  $t_0$  your credence in heads will be  $1/2$ . But what should it be at  $t_1$ ?

One natural response to this question, endorsed by Elga (2004), is that your credences in heads should be  $1/3$ . But there is something strange about this answer. There wasn't anything surprising about the evidence you got at  $t_1$ . Indeed, we can tailor the case so that the scientists will tell you at  $t_0$  precisely what you will experience when you wake up. How can evidence which you know you'll get justify this change in your credences?

To avoid this consequence, one might endorse the  $1/2$  response to the question that has been defended by Meacham (2008).<sup>35</sup> But this approach has similar consequences in other kinds of cases. Consider, for example, the fission variant of a case described in Meacham (2008):

*The Black and White Room (fission)*: Consider a case like the fission case, but with the following twist. If the coin toss comes up heads, then, as before, fission will not be performed. But the scientists will then flip a second fair coin to determine whether to put you in a black or white room at  $t_1$ . If the coin toss comes up tails, then at  $t_1$  you will be fissioned into two fissiles, the first which will be put in a black room, the second which will be put in a white room.

In this case, the account endorsed by Meacham (2008) has the consequence that once you see what color room you're in, your credence in heads should become  $1/3$ . Again, there is something strange about this answer. In this case you do learn something when you wake up: you learn whether you're in a white room or a black room. But it's still hard to see how that evidence could justify this belief change, since your credence will become  $1/3$  regardless of whether the room is black or white.

The counterintuitive aspects of these prescriptions might remind one of the Reflection Principles proposed by Van Fraassen (1995). Reflection requires an agent who knows her future credences to have the same credences now. In each of the prescriptions discussed

---

<sup>35</sup>A similar position with respect to the temporal version of the sleeping beauty case has been defended by Halpern (2005).

above, it seems something like Reflection is violated: the subject knows that her future credences will be different from her current ones.

This thought is on the right track. But Reflection isn't quite what we want. For one thing, as many people have pointed out, violations of Reflection aren't generally counter-intuitive.<sup>36</sup> It seems reasonable to be confident about the outcome of the card game you just played, for example, even if you know you won't remember the outcome ten years from now. So the source of the strangeness of these prescriptions can't just be that they violate Reflection. For another, the worry we've raised is a diachronic worry, a worry about how one's credences ought to change over time. But Reflection is a synchronic constraint, imposed on an agent's credences at a single time.<sup>37</sup> So Reflection doesn't directly bear on the kind of worry we're interested in.

The counterintuitive prescriptions described above share the following common element. In both cases, the prescribed belief change is inevitable: given any of the pieces of evidence you think you might get, your credences will change in the same way. These prescriptions seem irrational because they don't seem justified by the evidence. At a first pass, we might try to formulate this intuition as follows:

A sequential updating rule  $R$  should be such that a subject's current credences will lie in the span of the credences  $R$  prescribes to her doxastic epistemic successors.<sup>38</sup>

Rules which prescribe inevitable belief changes—changes that will occur given any of the pieces of evidence you think you might get—will violate this principle, since your current credences won't lie in the span of the credences the rule might prescribe you.

But this formulation requires two amendments. First, we only want it to apply to beliefs about what the world is like—*de dicto* beliefs—not to self-locating beliefs. After all, we don't want to reject an account for suggesting that an agent's beliefs about the time will inevitably change as time passes. Likewise, if the epistemic continuity and personal identity relations come apart, we'll want to allow an agent's beliefs about who they are to change over time.<sup>39</sup>

Second, we need to allow for cases where some of the subject's doxastic temporal successors aren't around to provide the appropriate span of credences. For example, suppose a subject knows she will be executed if and only if a fair coin toss lands tails. All of her

---

<sup>36</sup>For example, see Christensen (1991) and Weisberg (2007).

<sup>37</sup>See Weisberg (2007) for an in-depth discussion of this point.

<sup>38</sup>The restriction to *sequential* updating rules becomes important when we assess cases of memory loss and the like from an internalist perspective. Given the assumptions of section 2.3, memory loss will bring with it uncertainty regarding one's prior credences, so these are not a "pure" cases. But as we'll see in the following section, we can use the suggestion made in section 3.2 to render these "impure" cases unproblematic for the Learning Principle, even from an internalist perspective. We can only do this, however, because one's prior credences are one of the arguments of sequential rules. Epistemic kernel rules are not functions of one's prior credences, and the same method for handling these cases will not generally be available. To apply the Learning Principle to epistemic kernel rules in such cases, in a manner acceptable to the internalist, we need to add a further clause to the Learning Principle, such as that the subject must not lose *de dicto* information during the belief change.

<sup>39</sup>For example, consider someone who knows who she is, and knows that she is about to be fissioned into a pair of fissiles. And suppose the correct account of personal identity takes the original to be a different person from either of the fissiles. After fission has occurred, the fissiles should no longer believe that they are the same person as the original. And since the credences of the original will constrain the credences of the fissiles (given the default notion of epistemic continuity we've been using), a diachronic rule encoding these constraints should allow an agent's beliefs about who she is to change as time passes.

doxastic epistemic successors will be alternatives at worlds where the coin lands heads. So her current credence in heads ( $1/2$ ) won't lie in the span of those of her doxastic epistemic successors, whose credence in heads will be 1.

We can accommodate these cases by considering the span of her *extended* doxastic epistemic successors, which includes dummy successors to stand in for alternatives who get killed, and by prescribing these dummy successors a proxy credence function which assigns 1 to the world in question. So in the execution case, the dummy successors at the tails worlds will be prescribed a credence of 0 in heads, and the subject's current credence in heads will be required to lie in the interval  $[0, 1]$ , as desired.

Applying these amendments yields the following principle:

**The Learning Principle (Pure Contexts):** A sequential updating rule  $R$  should be such that a subject's current *de dicto* credences will lie in the span of the credences  $R$  prescribes to her extended doxastic epistemic successors (where dummy successors are prescribed the appropriate proxy credence function).

This principle seems to capture what went wrong with the two prescriptions we looked at above. Any updating rule which yields those prescriptions will violate the Learning Principle.<sup>40,41</sup>

How do LPC and GPC fare with respect to this principle? GPC fails to satisfy the Learning Principle, for reasons we've already seen. In the fission version of the sleeping beauty case, for example, it prescribes each of your doxastic successors a credence of  $1/3$  in heads, even though your initial credence in heads is  $1/2$ .

LPC, on the other hand, satisfies the Learning Principle. The full proof is provided in the appendix, but we can see how it satisfies the Learning Principle in each of the two cases described above. In the fission version of the sleeping beauty case, it prescribes your doxastic successors a credence of  $(1/2, 1/2)$  in heads and tails, as we saw in section 4.2.1. Since your  $t_0$  credences will be identical to the credences LPC assigns to your doxastic successors, the Learning Principle is satisfied.

What about the black and white room case? Your extended doxastic epistemic successors at  $t_0$  will consist of a single successor at the heads-and-white-room and heads-and-black-room worlds, and a pair of successors at the tails worlds. Drawing widths and heights in proportion to your  $t_0$  credences in the appropriate way, your doxastic epistemic successors at  $t_0$  will be:


---

<sup>40</sup>One might wonder whether certain counterexamples to Reflection, such as the Collin's prisoner case described in Arntzenius (2003) or the waiting for the train case described in Elga (2007), serve as counterexamples to the Learning Principle as well. They do not. Reflection gets into trouble in these cases because of how it picks out the relevant future credence functions. And the Learning Principle picks out the relevant future credence functions in a different way. Reflection picks out the credences of one's doxastic epistemic continuants at some time  $t$ , while the Learning Principle picks out the credences of one's doxastic epistemic successors, which may be located at many different times. It is only the former way of proceeding that is problematic. (See Halpern (2005) for a related discussion.)

<sup>41</sup>Why not formulate an even stronger principle, which offers more stringent constraints on the credences a rational updating rule can prescribe than just that one's credences must lie in some interval? One can; indeed, there are a number of different ways of doing so. But our intuitions about these precisifications are harder to discern, and more contentious. So I restrict myself here to the more modest goal of trying to capture the clear intuitions we have about the kinds of belief changes described above.

Original, $t_1$ ("Black Room")	Original, $t_1$ ("White Room")	Fissile 2, $t_1$ ("White Room")
		Fissile 1, $t_1$ ("Black Room")
Heads		Tails

Now suppose you were to find yourself in a black room at  $t_1$ . Then your evidence would eliminate the two white room possibilities:

Original, $t_1$ ("Black Room")	 Original, $t_1$ ("White Room")	Fissile 2, $t_1$ ("White Room")
		Fissile 1, $t_1$ ("Black Room")
Heads		Tails

Assigning credences to the remaining possibilities in proportion to their area, your credence in each of the black room possibilities would be  $1/2$ . Likewise, if you were to find yourself in a white room, your evidence would eliminate the two black room possibilities, and your credence in each of the white room possibilities would be  $1/2$ .

It follows that your credences in worlds at  $t_0$  lie in the span of the credences LPC prescribes to your doxastic successors:  $cr(\text{Heads} \wedge \text{Black}) = 1/4 \in [0, 1/2]$ ,  $cr(\text{Heads} \wedge \text{White}) = 1/4 \in [0, 1/2]$ , and  $cr(\text{Tails}) = 1/2 \in [1/2]$ . So LPC satisfies the Learning Principle in the black and white room case.

## 5.2 Impure Contexts

The Learning Principle we looked at in the previous section was restricted to ‘pure’ contexts. So when assessing whether a subject’s current credences lie in the span of those that  $R$  prescribes to her doxastic successors, we’re picking out these successors using the default notion of continuity, we’re assuming that these successors have unique predecessors, and we’re assuming that they have access to the arguments of  $R$ . But does the Learning Principle need to be restricted in these ways? Let’s see.

Let’s start with continuity. The Learning Principle should employ the same notion of continuity to pick out doxastic successors as the updating rule  $R$  it’s assessing. And there’s no reason to restrict the scope of the Learning Principle to  $R$ ’s which employ the default notion of continuity. So we can waive the restriction to the default notion of continuity.

What about the restriction to unique predecessors? At first glance, it seems like we do need this restriction. For it seems that if we allow cases in which our doxastic successors can have multiple predecessors, then we’ll let in cases in which it’s impossible for the Learning Principle to be satisfied. For example, consider a pair of subjects at  $t_0$ ,  $p_1$  and  $p_2$ , who know that they will fuse together at  $t_1$ . Suppose that at  $t_0$  their credences in some *de dicto* proposition  $a$  are  $1/3$  and  $1/2$ , respectively. Any particular credence in  $a$  that  $R$  prescribes to their “fuseile” will violate the Learning Principle. If  $R$  prescribes the fuseile a credence of  $1/2$ , then  $p_1$ ’s credence in  $a$ — $1/3$ —won’t lie in the span of those prescribed her doxastic successors. And if  $R$  prescribes the fuseile a credence that isn’t  $1/2$ , then  $p_2$ ’s credence in  $a$ — $1/2$ —won’t lie in the span of those prescribed her doxastic successor.

But this problem only arises if we assume that  $R$  must assign a precise value to  $a$ . If  $R$  permitted the subject to adopt any credence in  $a$  in  $[1/3, 1/2]$ , then the Learning Principle would be satisfied: both  $p_1$  and  $p_2$  would have credences that lie in the span of those  $R$  prescribes. Of course, the typical updating rules we're concerned with do assign precise values. But these rules are also generally restricted to cases in which the subject has a unique predecessor. And if we extend such a rule to accommodate non-unique predecessors along the lines suggested in section 3.3, the extended versions of these rules will assign imprecise values in precisely the way the Learning Principle requires. I.e., since the Learning Principle would require  $R$  to assign a credence of  $1/3$  to  $a$  if  $p_1$  were the only predecessor, and a credence of  $1/2$  in  $a$  if  $p_2$  were the only predecessor, then in the case in which they are both predecessors, the extension suggested in 3.3 will require the fuseile's credence in  $a$  to lie in the span between  $1/3$  and  $1/2$ . So the restriction to unique predecessors isn't needed either.

Finally, what about the restriction to subjects who have access to the arguments of  $R$ ? At first glance, it seems we need this restriction as well, since the Learning Principle seems to place unreasonable constraints on updating rules without it. For example, consider a rule  $R$  which takes one's previous credences as an argument. Consider a case where a subject knows that a given coin toss landed heads, that she has only one doxastic successor, and that her doxastic successor will forget the outcome of the coin toss. The Learning Principle requires  $R$  to prescribe her successor a credence of 1 in heads. But this seems like an unreasonable constraint to place on an updating rule. So it seems we should restrict the Learning Principle so that it doesn't apply in such cases.

But if we look a bit closer, we can see that we don't need this restriction. If we're externalists about this kind of norm, then the case described above is unproblematic. Prescribing a credence of 1 to the subject's successor in the case described above won't strike us as unreasonable if we're not moved by internalist intuitions.

If we're internalists about this kind of norm, on the other hand, then this prescription will strike us as unreasonable. But the fact that we find this prescription unreasonable isn't a problem for the Learning Principle. For there's already a tension between  $R$  and internalism, just in virtue of the fact that  $R$  is diachronic. In response to this tension, we'll need to replace the diachronic constraint with an appropriate synchronic surrogate. And doing this takes care of the problem: even if  $R$  makes this kind of unreasonable demand, its synchronic surrogate will not. So there's nothing problematic about requiring  $R$  to yield results like the one described above, since these results won't carry over to the rule the internalist actually cares about, the synchronic surrogate of  $R$ .

Thus, upon reflection, none of these three restrictions is needed. And we can take the Learning Principle described in section 5.1 to be fully general:

**The Learning Principle:** A sequential updating rule  $R$  should be such that the subject's current *de dicto* credences lie in the span of the credences  $R$  prescribes to her extended doxastic epistemic successors.<sup>42</sup>

---

<sup>42</sup>The Learning Principle ties into the skeptical scenarios discussed in Meacham (2008) in an interesting way. The prescriptions described in the fission versions of the 'many brains' and 'varied brains' cases discussed there can be seen as extreme violations of the Learning Principle. In conversation, several people have remarked that these two cases feel different from the third skeptical scenario discussed in the paper, the 'sadistic scientists' case. This feeling is borne out by the Learning Principle: unlike the other two prescriptions, the prescription described in the fission version of the 'sadistic scientists' case does not violate the Learning Principle.

In section 5.1 we motivated the Learning Principle by looking at “pure” cases like the fission versions of the sleeping beauty case and the black and white room case. What does the Learning Principle entail about “impure” cases, like the temporal versions of these cases?

For simplicity, let’s focus our attention on one of them: the temporal version of the sleeping beauty case. Given the default notion of continuity, the Learning Principle will require an updating rule to prescribe the subject on Monday morning a credence of  $(1/2, 1/2, 0)$  in Monday-and-Heads, Monday-and-Tails and Tuesday-and-Tails, respectively. But if we’re working in an internalist vein, as I’ve been assuming, these constraints won’t be the ones we care about. Since this is a case in which the subject doesn’t know what her previous credences were, the diachronic rule will diverge from its synchronic surrogate. For synchronic surrogates along the lines of those suggested in section 3.4, we’ll find that in this case her credences are required to be a mixture of  $(1/2, 1/2, 0)$  and  $(0, 0, 1)$ .

Some may find this unsatisfying. After all, it might seem strange that the Learning Principle permits the subject to adopt a credence of  $(1/3, 2/3)$  in heads and tails in the temporal case, even though it forbids such credences in the fission case. And some might want the Learning Principle to forbid such credence assignments in the temporal case as well.

We’ve already encountered the source of this worry, in section 4.3. The inegalitarian treatment of the fission and temporal cases stems from the default notion of continuity we’re employing. Since the continuity facts are different in these two cases, the Learning Principle treats them differently. But that’s not a problem with the Learning Principle. As before, if we want to accommodate egalitarian intuitions regarding these cases, we must adopt a different notion of continuity, one which treats temporal and fission versions of the sleeping beauty case the same way. And if we do that, then the Learning Principle will constrain both cases in the same way.

### 5.3 Worlds and Alternatives

As we saw in section 4.2.1, LPC effectively normalizes by world. As a result, beliefs about the world end up being treated differently from beliefs about time and identity. What justifies this discriminatory attitude?

Your epistemic successors can be located at different times. And your epistemic successors can be different individuals. Since the time and identity facts that hold for you and the ones that hold for your successors need not be the same, time and identity facts are unstable over time.

On the other hand, your epistemic successors cannot be located at different worlds. Since the world facts that hold for you and the ones that hold for your successors must be the same, world facts are stable over time.

So facts about the world are temporally stable in a way that facts about time and identity are not. As a result, some demands for diachronic stability that are reasonable with respect to beliefs about the world won’t be reasonable with respect to beliefs about time and identity. The difference in the stability of these beliefs supports different kinds of normative constraints. And the discriminatory manner in which LPC treats these beliefs encodes this difference.

## 6 Many Worlds

The issues we've been discussing bear in an interesting way on one of the debates surrounding the Many Worlds interpretation of quantum mechanics.

The standard interpretations of quantum mechanics posit chance events.<sup>43</sup> The Many Worlds interpretation replaces these chance events, which unpredictably yield one outcome or another, with 'branching' events, which predictably yield every outcome, with each outcome manifesting in a different 'branch'. Similarly, instead of assigning *chances* to different possible worlds where the outcomes occur, it assigns *weights* to different 'branches' at the same world where that outcome is born out.

With respect to the individuals at a world, it's natural to think of these branching events as similar to fission events: each individual in the original branch is split into a number of individuals which inhabit the new branches.<sup>44</sup> So the question of how our credences should evolve over time, with respect to which particular branch we're on, is a question about self-locating beliefs. Without taking a stand on the general viability of the Many Worlds interpretation, it's interesting to see how the previous discussion bears on this topic.

As currently formulated, neither LPC nor GPC is friendly to the Many Worlds interpretation.<sup>45</sup> Proponents of the Many Worlds interpretation want weights to play an epistemic role analogous to that of chances. Just as one's credences in the different outcomes that result from a chance event should be proportional to the chances of those events, they want one's credences in the fissiles that result from a branching event to be proportional to the weights of those branches. But neither LPC nor GPC take weights into account.

Whether we *should* try to take weights into account is a contentious issue. But put that issue aside. Can either LPC or GPC be extended to accommodate weights in the way the proponent of Many Worlds would like? Yes. Indeed, *both* admit of such extensions.<sup>46</sup>

For simplicity, let's restrict our attention to "pure" contexts. Let  $\bar{c}$  stand for the branch containing the alternative  $c$ . (At worlds where Many Worlds doesn't hold, I'll take there to be only one branch, consisting of the entire world.) Let  $w_{ep(c)}(\bar{c})$  stand for the weight of  $c$ 's branch, as assessed at the world, time and branch of  $ep(c)$ , the epistemic predecessor of  $c$ .<sup>47</sup> (Let  $w_{ep(c)}(\bar{c}) = 1$  if  $c$  is located at a world where Many Worlds doesn't hold). The desired extension of LPC is:

$$cr_e(a) = \sum_{(i|c_i \in a)} cr(ep(c_i)|ep(e)) \cdot w_{ep(c_i)}(\bar{c}_i) \cdot N_i, \quad \text{if defined,} \quad (29)$$

---

<sup>43</sup>Where I understand "chance" broadly, so as to include the non-dynamical chances posited by theories like Bohmian mechanics.

<sup>44</sup>For discussions of the Many Worlds interpretation along these lines, see Greaves (2004), (David) Lewis (2004) and (Peter) Lewis (2006). An alternative program, which explicitly rejects the analogy between branching and fission, has been defended by Saunders (1998) and Wallace (2006).

<sup>45</sup>Hilary Greaves has noted that GPC won't even be consistent in the context of the canonical 'decoherence' approaches to the Many Worlds interpretation. On these interpretations, there won't be a unique privileged way of decomposing the wave function into branches. Since GPC's assignments are sensitive to the number of fissiles (branch descendants) there are, these different decompositions will yield different credence assignments.

<sup>46</sup>Of course, whether we're justified in adopting these extensions is a different matter. For example, it's unclear whether the considerations sketched in section 5 will extend smoothly to the extended version of LPC.

<sup>47</sup>I.e., the proportion of the predecessor's current branch that will become the branch of its successor.

where the normalization factor  $N_i$  is:

$$N_i = \frac{cr(\overline{ep(c_i)}|ep(e))}{\sum_{(j|c_j \in \overline{c_i})} cr(ep(c_j)|ep(e)) \cdot w_{ep(c_j)}(\overline{c_j})}. \quad (30)$$

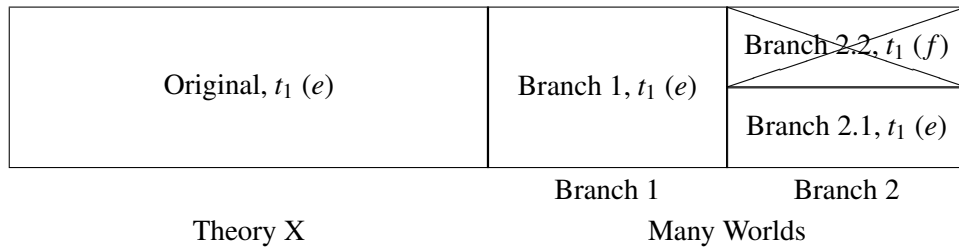
Call this the *Many Worlds Extension* of local predecessor conditionalization, or MW-LPC. Although this expression is somewhat opaque, the behavior of this rule is easy to describe. This rule will treat a split into branches of equal weight in the same way as LPC treats normal fission cases. And if the branches have unequal weights, the credence assigned them will be shifted so that it's proportional to their weight.

We can determine the prescriptions of MW-LPC using diagrams in the manner described in section 4. We simply replace the 3rd and 4th steps of the LPC procedure with:

- 3\*. Divide the width of the box into boxes representing the doxastic branches in  $d$ , with the width of these boxes proportional to  $p$ 's credence in those branches.
- 4\*. Divide the height of each box into boxes representing the alternatives in  $d$  from that branch, with the height of these boxes proportional to  $p$ 's credence in the alternative's predecessor times the weight of the alternative's branch (as evaluated at the time, world and branch of the alternative's predecessor).

Let's apply this procedure to an example. Consider a subject at  $t_0$ . Her credences are divided evenly between the Theory X and the Many World interpretation, and her credence in the Many Worlds interpretation is divided between two branches, branch 1 and branch 2. At  $t_1$  Branch 2 will split into two further branches of equal weight, branch 2.1 and branch 2.2. All of her doxastic successors get  $e$  as evidence except for the successor at branch 2.2, who gets evidence  $f$ . Finally, suppose that at  $t_1$  the subject gets  $e$  as evidence. What should her credences be, according to MW-LPC?

We can draw her space of doxastic successors at  $t_0$  as follows: divide width-wise between branches (which, at worlds where Many Worlds doesn't hold, is the entire world), and then divide heights among the successors at those branches/worlds:



The height of the branch 2 box is split evenly because her credence at  $t_0$  in the predecessor of each branch is the same, and each branch has the same weight. At  $t_1$  she will get evidence  $e$ , and eliminate one possibility: branch 2.2. Assigning credences to possibilities in proportion to their area, we find her credence at  $t_1$  in Theory X should be  $4/7$ , her credence in branch 1 should be  $2/7$ , and her credence in branch 2.1 should be  $1/7$ .

Now let's turn to GPC. The desired extension of GPC is:

$$cr_e(a) = \sum_{(i|c_i \in a)} cr(ep(c_i)|ep(e)) \cdot w_{ep(c_i)}(\overline{c_i}) \cdot N_i, \quad \text{if defined,} \quad (31)$$

where the normalization factor  $N_i$  is:

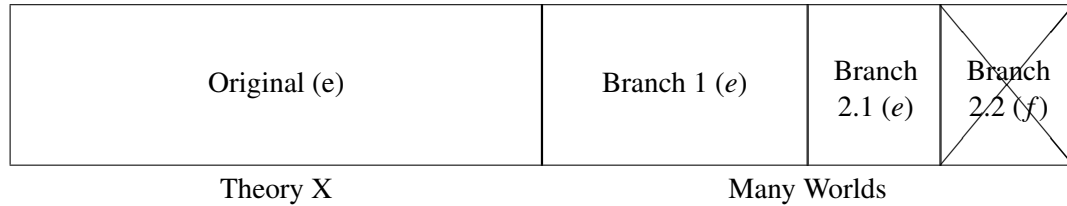
$$N_i = \frac{1}{\sum_{(j|c_j \in \Omega)} cr(ep(c_j)|ep(e)) \cdot w_{ep(c_j)}(\bar{c}_j)}. \quad (32)$$

Call this the *Many Worlds Extension* of global predecessor conditionalization, or MW-GPC.

As before, we can determine the prescriptions of MW-GPC using diagrams in the manner described in section 4. We simply replace the 3rd step of the GPC procedure with:

- 3\*. Divide the width of the box into boxes representing the alternatives in  $d$ , with the width of these boxes proportional to the product of  $p$ 's credence in the alternative's predecessors and the weight of the branch the alternative is on (as evaluated at the time, world and branch of the alternative's predecessors).

Let's apply this procedure to the example given above. For MW-GPC, we draw the space of her doxastic successors as:



When she gets evidence  $e$  at  $t_1$ , this will eliminate branch 2.2. Assigning credences to possibilities in proportion to their area, we find her credence at  $t_1$  in Theory X should be 4/7, her credence in branch 1 should be 2/7, and her credence in branch 2.1 should be 1/7.

At first glance, finding an updating rule which meshes well with the Many Worlds interpretation seems like a difficult task. And one might have reasonably expected these requirements to clash with LPC, GPC or both. As it turns out, however, this is not the case. Both can be extended to produce updating rules of the kind proponents of Many Worlds desire.

<sup>48,49</sup>

## References

Arntzenius, Frank. 2002. "Reflections on Sleeping Beauty." *Analysis* 62:53–61.

<sup>48</sup>Indeed, once we get a feel for how to make these extensions, we can see how to extend other accounts of how to update self-locating beliefs in order to accommodate Many Worlds branching. For example, one can extend the compartmentalized conditionalization described in Meacham (2008) by incorporating weights and compartmentalizing by branch instead of by world. So one's stance on how to update self-locating beliefs in canonical cases turns out to be largely independent of one's stance on Many Worlds.

<sup>49</sup>I would like to thank Phillip Bricker, Maya Eddon, Hilary Greaves, Hilary Kornblith, Jonathan Vogel, Jonathan Weisberg, an anonymous editorial board member, audience members at the 2008 Eastern APA, and members of my graduate class in the spring of 2009, for helpful comments and discussion.

- Arntzenius, Frank. 2003. "Self-locating Beliefs, Reflection, Conditionalization and Dutch Books." *Journal of Philosophy* 100:356–370.
- Bostrom, Nick. 2007. "Sleeping Beauty and Self Location: A Hybrid Model." *Synthese* 157:59–78.
- Bradley, Darren. 2003. "Sleeping Beauty: a note on Dorr's argument for 1/3." *Analysis* 63:266–268.
- Christensen, David. 1991. "Clever Bookies and Coherent Beliefs." *The Philosophical Review* 100:229–247.
- Christensen, David. 2000. "Diachronic Coherence versus Epistemic Impartiality." *The Philosophical Review* 109:349–371.
- Dorr, Cian. 2002. "Sleeping Beauty: in defense of Elga." *Analysis* 62:292–296.
- Elga, Adam. 2000. "Self-locating belief and the Sleeping Beauty problem." *Analysis* 60:143–147.
- Elga, Adam. 2004. "Defeating Dr. Evil with self-locating belief." *Philosophy and Phenomenological Research* 69:383–396.
- Elga, Adam. 2007. "Reflection and Disagreement." Forthcoming in *Nous*.
- Greaves, Hilary. 2004. "Understanding Deutch's Probability in a Deterministic Multiverse." *Studies in History and Philosophy of Modern Physics* 35:423–456.
- Halpern, Joseph Y. 2005. Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. In *Oxford Studies in Epistemology*. Vol. 1 Oxford University Press pp. 111–142.
- Hitchcock, Christopher. 2004. "Beauty and the Bets." *Synthese* 139:405–420.
- Horgan, Terrence. 2004. "Sleeping Beauty Awakened: New Odds at the Dawn of the New Day." *Analysis* 64:10–20.
- Howson, Colin and Peter Urbach. 1993. *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Open Court Publishing Company.
- Jenkins, Carrie. 2005. "Sleeping Beauty: A Wake-Up Call." *Philosophia Mathematica* 13:194–201.
- Kierland, Brian and Bradley Monton. 2005. "Minimizing Inaccuracy for Self-Locating Beliefs." 70:384–395.
- Kim, Namjoong. 2009. "Sleeping Beauty and Shifted Jeffrey Conditionalization." *Synthese* (forthcoming).
- Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *The Philosophical Review* 88:513–543.

- Lewis, David. 1983. *Survival and Identity*. In *Philosophical Papers, Vol. 1*. Oxford University Press.
- Lewis, David. 2001. "Sleeping Beauty: Reply to Elga." *Analysis* 61:171–176.
- Lewis, David. 2004. "How Many Lives Has Schrodingers Cat?" *Australasian Journal of Philosophy* 82:3–22.
- Lewis, Peter J. 2006. "Uncertainty and Probability for Branching Selves." Unpublished Manuscript.
- Meacham, Christopher J G. 2007. *Chance and the Dynamics of De Se Belief*, PhD thesis Rutgers University.
- Meacham, Christopher J G. 2008. "Sleeping Beauty and the Dynamics of De Se Beliefs." *Philosophical Studies* 138:245–269.
- Monton, Bradley. 2002. "Sleeping Beauty and the Forgetful Bayesian." *Analysis* 62:47–53.
- Saunders, Simon. 1998. "Time, Quantum Mechanics, and Probability." *Synthese* 114:373–404.
- Strevens, Michael. 2004. "Bayesian Confirmation Theory: Inductive Logic, or Mere Inductive Framework?" *Synthese* 141:365–379.
- Titelbaum, Michael G. 2008. "The Relevance of Self-Locating Beliefs." *Philosophical Review* 117:555–606.
- Van Fraassen, Bas. 1995. "Belief and the Problem of Ulysses and the Sirens." *Philosophical Studies* 77:7–37.
- Wallace, David. 2006. "Epistemology Quantized: circumstances in which we should come to believe in the Everett interpretation." *British Journal for the Philosophy of Science* 57:655–689.
- Weintraub, Ruth. 2004. "Sleeping Beauty: a simple solution." *Analysis* 64:8–9.
- Weisberg, Jonathan. 2007. "Conditionalization, Reflection, and Self-Knowledge." *Philosophical Studies* pp. 179–197.
- White, Roger. 2006. "The generalized Sleeping Beauty problem: a challenge for thirders." 66:114–119.

## Appendix

We want to show that LPC satisfies the Learning Principle: "If  $R$  is a rational updating rule, and if all of a subject's doxastic successors have unique predecessors, then the subject's current *de dicto* credences should lie in the span of the credences  $R$  assigns to her extended doxastic epistemic successors (where dummy successors are assigned the appropriate proxy credence function)."

Consider the probability function  $p$ , and a series of probability functions  $q_i$ . If we can find coefficients  $\alpha_i$  which satisfy the following three conditions:

$$(a) \quad p(\cdot) = \sum_i \alpha_i \cdot q_i(\cdot), \quad (33)$$

$$(b) \quad \sum_i \alpha_i = 1, \quad (34)$$

$$(c) \quad \forall i (\alpha_i \in [0, 1]), \quad (35)$$

then  $p$  is a mixture of  $q_i$ 's, and thus lies in the span of the  $q_i$ 's. So we can show that an updating rule  $R$  satisfies the Learning Principle if we show that, when the appropriate conditions hold,  $cr(a)$  is a mixture of the  $cr_e(a)$ 's that  $R$  assigns to her doxastic successors and the proxy functions that get assigned to dummy successors.

Let  $i$  range over both  $i'$  and  $i''$ , where  $i'$  ranges over different possible kinds of evidence, and  $i''$  ranges over alternatives without successors. Let  $\alpha_i$  be such that:

$$\begin{aligned} \alpha_{i'} &= cr(ep(e_{i'})), \\ \alpha_{i''} &= cr(c_{i''}), \end{aligned} \quad (36)$$

where  $e_{i'}$  is the set of alternatives with the evidence picked out by  $i'$ . Let  $q_i(\cdot)$  be such that:

$$\begin{aligned} q_{i'}(a) &= cr_{e_{i'}}(a), \\ q_{i''}(a) &= [c_{i''} \in a], \end{aligned} \quad (37)$$

where  $cr_{e_{i'}}(a)$  is LPC's assignment to the doxastic successor which gets evidence  $e_{i'}$ , and  $[c_{i''} \in a]$  is the Iverson bracket function (which equals 1 if the statement in brackets is true and 0 otherwise). Under the conditions specified by the Learning Principle, it will be the case that:

$$cr(a) = \sum_i \alpha_i \cdot q_i(a) = \sum_{i'} \alpha_{i'} \cdot cr_{e_{i'}}(a) + \sum_{i''} \alpha_{i''} \cdot [c_{i''} \in a], \quad (38)$$

in a manner which satisfies conditions (a)-(c). It follows that  $cr(a)$  is a mixture of the  $cr_e(a)$ 's that LPC assigns to her doxastic successors and the proxy functions that get assigned to dummy successors. Thus LPC satisfies the Learning Principle.

To see that each of the conditions (a)-(c) is satisfied:

*Condition (a).* Let's start by looking at the first half of the two part sum in (38):

$$\sum_{i'} \alpha_{i'} \cdot cr_{e_{i'}}(a). \quad (39)$$

Since we're restricting our attention to  $a$ 's which are *de dicto* propositions, we can evaluate this sum at each world, and then sum over these results for all worlds in  $a$ :

$$\sum_{i'} \alpha_{i'} \cdot cr_{e_{i'}}(a) = \sum_{i'} \alpha_{i'} \cdot \sum_{(j|w_j \in a)} cr_{e_{i'}}(w_j). \quad (40)$$

Substituting in for  $\alpha_{i'}$  and  $cr_{e_{i'}}(a)$  and simplifying:

$$= \sum_{i'} cr(ep(e_{i'})) \cdot \sum_{(j|w_j \in a)} \frac{\sum_{(k|c_k \in w_j)} cr(ep(c_k)|ep(e_{i'})) \cdot cr(\overline{ep(c_k)}|ep(e_{i'}))}{\sum_{(l|c_l \in \overline{c_k})} cr(ep(c_k)|ep(e_{i'}))}, \quad (41)$$

$$\begin{aligned}
&= \sum_{i'} cr(ep(e_{i'})) \cdot \sum_{(j|w_j \in a)} \frac{\sum_{(k|c_k \in w_j)} cr(ep(c_k)|ep(e_{i'})) \cdot cr(w_j|ep(e_{i'}))}{\sum_{(l|c_l \in \bar{c}_k)} cr(ep(c_k)|ep(e_{i'}))}, \\
&= \sum_{i'} cr(ep(e_{i'})) \cdot \sum_{(j|w_j \in a)} cr(w_j|ep(e_{i'})), \\
&= \sum_{i'} \sum_{(j|w_j \in a)} cr(w_j \wedge ep(e_{i'})), \\
&= \sum_{i'} cr(a \wedge ep(e_{i'})).
\end{aligned}$$

Finally, note that:

$$\sum_{i'} cr(a \wedge ep(e_{i'})) = \sum_{(j|\exists k(c_k=es(c_j)))} cr(a \wedge c_j), \quad (42)$$

since the only alternatives in  $a$  that  $\sum_{i'} cr(a \wedge ep(e_{i'}))$  doesn't range over are those which aren't the predecessors of anything; i.e., those which don't have successors. So we can think of  $\sum_{i'} cr(a \wedge ep(e_{i'}))$  as a sum over all the alternatives in  $a$  with successors.

Turning to the second half of the two part sum in (38), note that:

$$\sum_{i''} \alpha_{i''} \cdot [c_{i''} \in a] = \sum_{i''} cr(c_{i''}) \cdot [c_{i''} \in a] = \sum_{(j|\neg \exists k(c_k=es(c_j)))} cr(a \wedge c_j), \quad (43)$$

since  $i''$  ranges over all alternatives without successors, and  $[c_{i''} \in a]$  effectively restricts this to the alternatives without successors in  $a$ .

It follows from (42) and (43) that:

$$\begin{aligned}
\sum_i \alpha_i \cdot q_i(a) &= \sum_{i'} \alpha_{i'} \cdot cr_{e_{i'}}(a) + \sum_{i''} \alpha_{i''} \cdot [c_{i''} \in a] \\
&= \sum_{(j|\exists k(c_k=es(c_j)))} cr(a \wedge c_j) + \sum_{(j|\neg \exists k(c_k=es(c_j)))} cr(a \wedge c_j), \\
&= \sum_{(j|c_j \in \Omega)} cr(a \wedge c_j), \\
&= cr(a).
\end{aligned} \quad (44)$$

*Condition (b).* First note that:

$$\sum_{i'} \alpha_{i'} = \sum_{i'} cr(ep(e_{i'})) = \sum_{(j|\exists k(c_k=es(c_j)))} cr(c_j), \quad (45)$$

since the only alternatives in  $a$  that  $\sum_{i'} cr(a \wedge ep(e_{i'}))$  doesn't range over are those which aren't the predecessors of anything; i.e., those which don't have successors. So we can think of  $\sum_{i'} cr(a \wedge ep(e_{i'}))$  as a sum over all the alternatives in  $a$  with successors. Second, note that:

$$\sum_{i''} \alpha_{i''} = \sum_{i''} cr(c_{i''}) = \sum_{(j|\neg \exists k(c_k=es(c_j)))} cr(c_j), \quad (46)$$

since  $i''$  ranges over alternatives without successors. Thus:

$$\sum_i \alpha_i = \sum_{i'} \alpha_{i'} + \sum_{i''} \alpha_{i''} \quad (47)$$

$$\begin{aligned}
&= \sum_{(j|\exists k(c_k=es(c_j)))} cr(c_j) + \sum_{(l|\neg\exists m(c_m=es(c_l)))} cr(c_l), \\
&= \sum_{(j|c_j\in\Omega)} cr(c_j), \\
&= 1.
\end{aligned}$$

*Condition (c).* Since  $\alpha_{i'} = cr(ep(e_{i'})) \in [0, 1]$  and  $\alpha_{i''} = cr(c_{i''}) \in [0, 1]$ , it follows that  $\alpha_i \in [0, 1]$ , for all  $\alpha_i$ 's.