Time-Slice Rationality and Self-Locating Belief

David Builes

Forthcoming in *Philosophical Studies*

The epistemology of self-locating belief concerns itself with how rational agents ought to respond to indexical information concerning where they are located in the world. Over the past couple of decades, it has been discovered that this issue is deeply interconnected with a wide variety of unresolved questions in formal epistemology. As Titelbaum (2013) has noted, the epistemology of self-locating belief is bound up with questions about relative frequencies, objective chances, Dutch book arguments, accuracy arguments, reflection principles, and indifference principles. Debates about self-locating belief are even interconnected with many of the central questions in the metaphysics of science, including debates about the correct interpretation of quantum mechanics (Bradley (2011a), Lewis (2006), Sebens and Carroll (2018)), the existence of multiple universes (Bradley (2012)), the measure problem in cosmology (Arntzenius and Dorr (2017)), and even classic questions in the metaphysics of time (Builes (2019)).

The goal of this paper is to argue for a novel way of resolving these debates about the epistemology of self-locating belief, by appealing to the recently popular thesis of *Time-Slice Rationality*, defended by Hedden (2015a, 2015b) and Moss (2015).[1] Since much of the literature on self-locating belief has revolved around Elga's (2000) Sleeping Beauty problem, my discussion will follow suit.

Before we start, some introductory remarks on Time-Slice Rationality and Sleeping Beauty are in order. Following Hedden (2015b), Time-Slice Rationality is the conjunction of two theses:

> *Synchronicity*: What attitudes you ought to have at a time does not directly depend on what attitudes you have at other times.
>
> *Impartiality*: In determining what attitudes you ought to have at a time, your *beliefs* about what attitudes you have at other times play the same role as your beliefs about what attitudes other people have. (p. 452)

There are two main motivations for adopting Time-Slice Rationality. The first stems from the thought that the requirements of rationality should not make reference to the relation of personal

[1] For some more recent criticisms and defenses of Time-Slice Rationality, see Podgorski (2016), Dori Döring and Eker (2017), Snedegar (2017), and Hedden (2016, 2017).

identity over time, which notoriously gets very murky in certain puzzling cases (e.g. teletransportation, fission, partial brain transplants, etc.). As Hedden (2015b) writes:

> Determining what an agent ought to believe does not require first figuring out the correct theory of personal identity over time. This means that requirements of rationality should not make reference to the relation of personal identity over time; what you ought to believe does not depend on who *you* are. That is, the requirements of rationality should be *impersonal*. (p. 452)

Let us call this intuition, that the requirements of rationality should not make reference to the relation of personal identity over time, the *No Reference* intuition. This intuition will bear much of the argumentative weight in the arguments below, and since the *No Reference* intuition is so central to Time-Slice Rationality, I will be taking it as an implicit commitment of the view from now on. Both *Synchronicity* and *Impartiality* are meant to be precisifications of the idea that the requirements of rationality should be 'impersonal' in the way that *No Reference* demands. The second motivation stems from the internalist intuition that what it is rational for an agent to believe should supervene on that agent's perspective. Since, according to the internalist, an agent's perspective on the world at a time is constituted by their mental life at that time, *Synchronicity* follows.

Next, let us recall the setup of the Sleeping Beauty case:

> Sleeping Beauty: Beauty is a perfectly rational agent who is told that the following events will occur. On Sunday, she will be put to sleep. A fair coin will then be tossed. If it lands Heads, she will be awakened on Monday morning. Later, in the evening, she will be told that it is Monday, and then she will be let go. If the coin lands Tails, as before, Beauty will be awakened on Monday morning, and then she will be told that it is Monday later that evening. However, instead of being let go, she will be given a memory-loss drug that will make her forget all of her memories of Monday, and she will be put back to sleep. She will then be awakened on Tuesday, and then she will be let go. Her wakings on Monday and Tuesday will be indistinguishable. When she first awakens on Monday morning, what should her credence be that the coin landed Heads? When she is subsequently told that it is Monday, on Monday evening, what should her credence be that the coin landed Heads?

There are three main responses to the problem. Currently, the most popular answer to the problem seems to be the Thirder position, according to which Beauty's credence should be 1/3 in the morning and 1/2 in the evening.[2] There are two other less popular answers. According to Lewisian Halfers, Beauty's credences should be 1/2 in the morning and 2/3 in the evening (e.g. Lewis (2001)

[2] Authors who have defended the Thirder solution include Arntzenius (2003), Dorr (2002), Draper and Pust (2008), Elga (2000), Hitchcock (2004), Horgan (2004), Monton (2002), Stalnaker (2008), Titelbaum (2008), and Weintraub (2004). For a useful survey of the arguments for the Thirder position, see Ross (2010).

and Bradley (2011a)). According to Double Halfers, Beauty's credences should be 1/2 in the morning and evening (e.g. Bostrom (2007) and Meacham (2008)).

I will be arguing that the connection between Time-Slice Rationality and Double Halfing runs very deep, and in fact the connection is already implicit in much of the literature. To show this, I will be presenting three independent arguments for why Time-Slicers should be Double Halfers. The first argument is a positive argument in favor of Double Halfing: if one takes an accuracy-first approach to Sleeping Beauty, then Time-Slicers should be Double Halfers. This first argument will be inspired by the discussion in Kierland and Monton (2005). The second argument is an argument against the popular Thirder view: perhaps the best way for Thirders to avoid certain highly implausible consequences of their view is to reject Time-Slice Rationality. This second argument will be inspired by discussions in Leitgeb (2010), Meacham (2008), and Bradley (2011a, 2015). The third argument is a defense of Double Halfing: perhaps the main objection to Double Halfing fails if Time-Slice Rationality is true. This third argument will be inspired by the initial exchange between Elga (2000) and Lewis (2001).

I. The First Argument: Accuracy and Sleeping Beauty

It would be nice if Sleeping Beauty could be settled using accuracy arguments. After all, it's natural to think that what matters at the end of the day is getting at the truth, not avoiding clever Dutch books or best satisfying our intuitions.[3] Unfortunately, several authors have argued that accuracy arguments alone fail to settle Sleeping Beauty (e.g. Briggs (2010), Kierland and Monton (2005), and Pettigrew (MS)). I agree with these authors. However, I will argue that accuracy arguments *do* settle Sleeping Beauty when combined with the thesis of Time-Slice Rationality.

The reason why an accuracy-first approach to Sleeping Beauty fails to settle the problem is because on one precisification of 'minimizing inaccuracy', we should be Thirders, and on another precisification we should be Double Halfers. In the rest of this section, I will briefly outline why this is so.[4]

Following Lewis (1979), we will be taking the objects of credence to be sets of *centered worlds*, which are ordered pairs consisting of a world *w* together with the location of a particular time-slice of an agent within *w*. A set *X* of centered worlds is called a *de dicto* or *uncentered* proposition iff whenever it contains some centered world, it also contains all centered worlds with the same world-coordinate. A set of centered worlds is called a *de se* or *centered* proposition otherwise. Intuitively, de dicto propositions are entirely about what the world is like, while de se propositions are also about where you are in the world. Examples of de dicto beliefs include the belief that space-time is curved, the belief that laptops exist, and the belief that a Republican won the 1992 US

[3] For a book length defense of an accuracy-first approach to epistemology, see Pettigrew (2016).
[4] The following discussion will largely follow Kierland and Monton (2005).

presidential election. If two agents have the same de dicto belief, either both are right or both are wrong. Examples of de se beliefs include the belief that today is Tuesday and the belief that I am in Minnesota. If two people believe the de se content expressed by 'It is Monday', one might be right and one might be wrong.

Next, some remarks about accuracy. Just as beliefs may be true or false, credences can be more or less accurate. Intuitively, we would like a measure of how accurate a credence is based on how far away it is from the credence function that assigns 1 to all truths and 0 to all falsehoods. As is common, the particular measure I will be using to measure the inaccuracy of a credal state will be the Brier score.[5] Here's how the Brier score works. First, for any world w and any uncentered proposition X, we will let $w(X) = 1$ iff X is true at w and $w(X) = 0$ iff X is false at w.[6] Then, given some credence function c and some uncentered proposition X, we define the *Brier score* of c at w *with respect to* X as:

$$B_{c,w}(X) = (w(X) - c(X))^2$$

So, what credence should Beauty assign to the proposition that the coin landed Heads when she wakes up if she wants to minimize her expected inaccuracy? There are two plausible ways to do this. First, one may try to minimize the expected *total* inaccuracy that one will accrue across all wakings by setting one's credence in the proposition that *the coin will land Heads* to h. There's an objective chance of 1/2 that one's *total* inaccuracy for this proposition will be $(1\text{-}h)^2$ if the coin lands Heads (since there will only be one waking), and there's an objective chance of 1/2 that one's total inaccuracy will be $(0 - h)^2 + (0 - h)^2$ if the coin lands Tails (one summand for each waking). Using a bit of elementary calculus, we find that the value of h that will minimize one's expected total inaccuracy, which is $(1/2)*(1\text{-}h)^2 + 1/2*((0 - h)^2 + (0 - h)^2)$, is 1/3, vindicating Thirders.

Second, one may try to minimize the expected *average* inaccuracy that one will accrue in the following way. There's an objective chance of 1/2 that one's average inaccuracy in the proposition that *the coin will land Heads* will be $(1\text{-}h)^2$ if the coin lands Heads, and there's an objective chance of 1/2 that one's average inaccuracy will be $((0 - h)^2 + (0 - h)^2)/2 = h^2$ if the coin lands Tails, which is the average inaccuracy of both wakings. The value of h that will minimize $(1/2)*(1\text{-}h)^2 + 1/2*h^2$ is 1/2, vindicating Halfers.

Lastly, it is easy to see that both methods recommend adopting a credence of 1/2 in h on Monday evening. On Monday evening, there are two uncentered possible worlds that might be actual, the world in which the coin lands Heads and the world in which the coin lands Tails. Both worlds have an objective chance of 1/2 of being the actual world. Furthermore, since one knows that it is Monday evening on Monday evening, there is only one center in each world that is compatible with one's current evidence. So, since both uncentered worlds have an objective chance of 1/2 and

[5] For much more on different inaccuracy measures and their justifications see Leitgeb and Pettigrew (2010).
[6] We can think of a world w as a function from uncentered propositions to the truth value of that proposition at w.

both uncentered worlds only have one center compatible with one's evidence, the method of minimizing one's expected total inaccuracy and the method of minimizing one's expected average inaccuracy agree that one should have a credence of 1/2 in *h* on Monday evening.

So, minimizing expected total inaccuracy entails the Thirder position in Sleeping Beauty, and minimizing expected average inaccuracy entails the Double Halfer position in Sleeping Beauty. Briggs (2010) has generalized this result to arrive at fully general updating procedures that handle all de se cases. In Briggs terminology, minimizing expected total inaccuracy implies the 'Thirder Rule' and minimizing expected average inaccuracy implies the 'Halfer Rule'.[7]

The Halfer Rule is easy to describe. Suppose an agent's total evidence is E, and their current rational credence function is Cr. Let Cr* be the agent's rational prior credence function, and let E^* be the strongest de dicto proposition entailed by E. Then, according to the Halfer Rule, for any uncentered proposition X:

$$\textit{Halfer Rule}\text{: } \mathrm{Cr}(X) = \mathrm{Cr}^*(X \mid E^*)$$

While I will not be going over the details of Briggs' derivations of the Halfer Rule and the Thirder Rule here, it should be stressed that a successful defense of either one of these accuracy goals, given Briggs' results, amounts to a successful defense of a completely general theory of how to handle all de se and de dicto evidence.

So, how are we to decide between the competing goals of minimizing our expected average inaccuracy and minimizing our expected total inaccuracy? Prima facie, either goal looks equally permissible. Pettigrew sums up the dialectical situation between these two epistemic goals as follows:

> Are there any such arguments [that favor one goal over the other]? I haven't seen them, nor have I been able to formulate them. Moreover, I find it hard to imagine how such an argument might go. Both understandings of probability seem reasonable; both seem to give reasonable definitions of expected inaccuracy; and, most importantly, both give definitions of a quantity that one would hope, intuitively, to minimize. Thus, any such argument would have to favour one and explain why our intuitive attraction to the other is mistaken. This is a difficult task. But that is not to say that it cannot be done. (p. 16)

In the next section, I will try to take up this challenge.

[7] The Halfer Rule is also equivalent to the updating procedures defended by Meacham (2008) and Halpern (2006).

II. The First Argument: Solving the Problem

We have two epistemic goals on the table: the goal of minimizing one's total inaccuracy and the goal of minimizing one's average inaccuracy. I will now argue that the goal of minimizing one's total inaccuracy is inconsistent with Time-Slice Rationality.

This fact can be brought out by considering the following variant of Sleeping Beauty:

> Duplicating Beauty: On Sunday, Beauty will be put to sleep. She will be woken up on Monday, and then let go. A coin will then be tossed on Monday night. If it lands Heads, nothing happens. If it lands Tails, a perfect subjective duplicate of Beauty, call her Tuesday Beauty, will be created, and this duplicate will be woken up on Tuesday morning. Tuesday Beauty will then be let go. Beauty is told that her Monday waking will be subjectively indistinguishable from the Tuesday waking of Tuesday Beauty.

Note that this case is structurally identical to Sleeping Beauty. There are two uncentered possibilities, the Heads possibility and the Tails possibility, which each have an objective chance of 1/2. In the Heads possibility, there is only one epistemically possible center of Beauty, and in the Tails possibility, there are two epistemically possible centers of Beauty. When Beauty wakes up, for all she knows, it might be Monday or Tuesday. The only relevant change in this case is the personal identity facts. Beauty is not Tuesday Beauty.

However, in this case, the goal of minimizing one's total inaccuracy, as Kierland and Monton (2005) themselves note, recommends that Beauty ought to have a credence of 1/2 when she wakes up.[8] This is because the goal is to minimize one's *own* total inaccuracy. When Beauty wakes up, she knows that, no matter which uncentered possibility is actual (Heads or Tails), there is only one epistemically possible center that is *her*. Even if the coin lands Tails, she knows that she is either Beauty (and not Tuesday Beauty) or that she is Tuesday Beauty (and not Beauty). It easy to see that the credence that will minimize her expected total inaccuracy in this case is therefore 1/2.[9]

So, the goal of minimizing one's total inaccuracy gives different recommendations in Sleeping Beauty and Duplicating Beauty. Since the only relevant difference in Sleeping Beauty and Duplicating Beauty are the personal identity facts, the recommendations that the goal gives explicitly depend on the personal identity facts, contradicting the *No Reference* intuition. In particular, since the goal requires Beauty to treat the attitudes of *other* people (Tuesday Beauty)

[8] Kierland and Monton's position is that one ought to be a Halfer in Duplicating Beauty, but either Thirding or Halfing is permissible in Sleeping Beauty.

[9] If her credence in the coin landing Heads is h, her total inaccuracy will be $(1-h)^2$ if the coin lands Heads, and her total inaccuracy will be $(0-h)^2$ if the coin lands Tails (because there will only be *one* center that is her). So, her expected total inaccuracy is $(1/2)*(1-h)^2 + (1/2)*(0-h)^2 = h^2 - h + 1/2$. The value that minimizes this quantity is $h = 1/2$.

very differently than the attitudes of herself, it explicitly violates the requirement of *Impartiality*. The goal is therefore inconsistent with Time-Slice Rationality.

Next, let us turn to our second main goal: the goal of minimizing one's average inaccuracy. This goal is not only consistent with Time-Slice Rationality, but it is the goal that is best motivated by Time-Slice Rationality. Given Time-Slice Rationality, you should only care about minimizing the inaccuracy of your *current* time-slice. Since you may be unsure of which time-slice you are in the world in cases of self-locating uncertainty, your best guess for the expected inaccuracy of your current time-slice is the expected *average* inaccuracy of all the time-slices that may be you, for all you know.[10] So, the goal of minimizing one's expected average inaccuracy *just is* the goal of minimizing the expected inaccuracy of your current time-slice. For the averager, whether or not certain centers are identical to other centers is entirely irrelevant.

Here is an analogous ethical case to drive the point home. Consider the debate between total utilitarians, who want to maximize the total utility in the world, and average utilitarians, who want to maximize the average utility in the world. Let possible world w_1 contain three agents, each with utilities 10, 20, and 30 respectively, and let possible world w_2 contain two agents, each with 25 utility. The total utilitarian will think w_1 is better than w_2, and the average utilitarian will think that w_2 is better than w_1. The crucial point is that average utilitarianism can be motivated in the following way. Suppose you were behind a veil of ignorance, and you were unsure of which agent you were going to be in the world. If your goal was to selfishly maximize your own utility, you would prefer that w_2 be actual rather than w_1, because your expected utility in w_1 is the *average* utility of the three agents in w_1, namely 20, and your expected utility in w_2 is the *average* utility of the agents in w_2, namely 25. So, in the ethical case, if you selfishly want to maximize your own utility when you don't know who you are in the world, you take averages of the utilities of all the agents you might be.[11] Similarly, in the epistemic case, if you selfishly want to minimize the inaccuracy of your current time-slice when you don't know which time-slice you are in the world, you take averages of the inaccuracies of all the time-slices you might be.

So, Time-Slicers have a principled way to resolve the disagreement between the two competing epistemic goals. They should minimize expected average inaccuracy, resulting in the Double Halfer position and the Halfer Rule.

---

[10] It is worth noting that the expected inaccuracy of your current time-slice should actually be the *weighted* average of the inaccuracies of all the time-slices you may be (weighted by your credence in each of the corresponding centers). However, if one appeals to Elga's (2000) 'highly restricted principle of indifference', which says that one should assign equal credence to all the epistemically possible centers within an uncentered word, one will be required to assign equal weights to all those time-slices within the same uncentered world. While Elga's principle has proven very popular, for some pushback see Weatherson (2005). For a brief response to Weatherson (2005), see Bradley (2011a: 338-339).

[11] Here I am assuming Elga's highly restricted principle of indifference (see footnote 11).

III. The Second Argument: Thirding and Time-Slice Rationality

For many philosophers, Thirding is unattractive because its most natural generalizations lead to unacceptable consequences. Leitgeb (2010) has argued that the Thirder is committed to implausible cosmological consequences. Meacham (2008) has argued that the Thirder is committed to implausible skeptical consequences. Bradley (2011a, 2015) has argued that the Thirder is committed to implausible quantum mechanical consequences. However, *all* of these consequences may be defused if the Thirder is willing to generalize their position in a way that explicitly rejects Time-Slice Rationality (by, for example, adopting the goal of minimizing total inaccuracy discussed above). This is a powerful reason for Thirders to reject Time-Slice Rationality. In the absence of other ways to avoid these consequences, Time-Slicers who wish to avoid these implausible consequences must reject Thirding.

Leitgeb's cosmological case is as follows:

> Eternal Recurrence: Suppose astrophysicists have pinned down the actual evolution of our universe to either of two models. According to the first model, our universe is going to expand indefinitely. On the second model, our universe is expanding and contracting indefinitely, so that history repeats itself over and over again. An indeterministic quantum event shortly after the Big Bang determined whether our universe would evolve according to the first or second model. There is an objective chance of 1/2 that the quantum event went either way. What should your credence be in Eternal Recurrence?

While the prima facie obvious answer is '1/2', Leitgeb argues that Thirders should have credence 1 in Eternal Recurrence in this case. Moreover, Thirders should have credence 1 in Eternal Recurrence so long as the objective chance of Eternal Recurrence is *non-zero*! In fact, given that the objective chances are dispensable, Thirders seem to be forced to have credence 1 in Eternal Recurrence, so long as their prior probability in Eternal Recurrence is non-zero![12]

Meacham's skeptical case is as follows:

> Many Brains: Consider the proposition that you're in a world where brains in vats are constantly being constructed in states subjectively indistinguishable from your own. Let your credence in this possibility be $0 < p < 1$, and your credence that there will be no multiplication of [your subjective states] be $1 - p$.

Meacham argues that if you accept Thirding, then 'you should come to believe (if not yet, then in a little while) that these brains in vats are being created . . . as you become certain that these

[12] To my knowledge, the only Thirder who has argued that the presence of objective chance makes a crucial difference in cases of self-locating uncertainty is Wilson (2014). Wilson also argues against the case of Quantum Measurement below on the grounds that the correct interpretation of Quantum Mechanics is not decided by a chancy process. For a response to this suggestion, see Bradley (2015: 689-692).

brains in vats are being created, you should become certain that you're a brain in a vat' (p. 260). This is an unwelcome result.

Bradley's (2011a, 2015) quantum-mechanical case is as follows, where 'MWI' refers to the Many Worlds interpretation of Quantum Mechanics:

> Quantum Measurement: You are about to perform a spin measurement with possible outcomes Up and Down. Quantum mechanics says that Up and Down each has a chance of fifty percent. According to MWI, the universe will divide, so you will have two future successors, one of whom will observe Up, and one Down. According to a stochastic theory (ST), there will be only one future successor, who will observe either Up or Down, each with fifty percent probability. You are unsure of whether MWI or ST is correct, and you assign each a credence of fifty percent.

Bradley argues that Thirders are committed to the view that you ought to increase your credence in MWI after you perform the experiment, regardless of what your observation will be! If he is right, Thirders are therefore committed to 'easy confirmation' of MWI. Every time the universe branches, we gain reason to believe in MWI over ST! As Bradley writes, '. . . our everyday observations are constantly confirming MWI. On this reasoning, MWI gets enormous confirmation without the need for modern physics. The Ancients could have worked out that they have overwhelming evidence for MWI merely by realizing it was a logical possibility and observing the weather' (2011a: p. 336).

It's not hard to see why these philosophers have seen connections between these cases and Sleeping Beauty. Each of the above three cases shares a common structure with Sleeping Beauty. In Sleeping Beauty, there are two relevant uncentered possibilities: the Heads possibility and the Tails possibility. Similarly, in Eternal Recurrence there are two relevant cosmological models; in Many Brains there is one skeptical possibility and one non-skeptical possibility; in Quantum Measuerment there are two different interpretations of Quantum Mechanics. In Sleeping Beauty, Thirders deviate from their initial credence in these two possibilities by giving more 'weight' to possibilities in proportion to how more epistemically possible centers it contains. Similarly, in each of our three cases, one of the two relevant possibilities only includes one epistemically possible center while the other includes more than one epistemically possible center. So, it seems like Thirders should assign more credence to the possibilities which includes more centers (i.e. the eternal recurrence possibility, the skeptical possibility, and the MWI possibility).

Moreover, many of the standard arguments for the Thirder position in Sleeping Beauty straightforwardly apply to these cases. For example, Elga's (2000) first argument for the Thirder position is a long-run frequency argument. If we imagined the Sleeping Beauty experiment being run many different times, in the long run roughly 1/3 of Beauty's wakings will correspond to a Heads flip and roughly 2/3 of the wakings would correspond to a Tails flip. Similarly, if we imagine a 'multiverse' in which many different universes come into existence, each with chance

1/2 of being one-history worlds or eternal recurrence worlds, then almost every agent would end up being in an Eternal Recurrence world. Following the same line of reasoning leads one to have credence 1 that one is an Eternal Recurrence world!

In addition, the result that Thirders should adopt these implausible consequences is also a consequence of many of the formal, mathematically precise generalizations of the Thirder position that have been developed in the literature. While I will not rehearse these generalizations here, readers are encouraged to see Briggs (2010), Meacham (2008), and Pettigrew (MS) for generalizations of the Thirder position which all imply the counterintuitive consequences above.[13]

The Thirder who wishes to avoid these three consequences must identify some relevant difference between Sleeping Beauty and these cases. Fortunately, there is a clear difference: only in Sleeping Beauty do the relevant centers within each uncentered possibility correspond to the *same* agent.[14] The Thirder also has a principled reason for thinking that this difference is a *relevant* difference, namely that the goal of minimizing *total* inaccuracy (discussed above) is a principled Thirder position that is sensitive to this sort of difference. Because the goal of minimizing total inaccuracy gives 'Halfer' verdicts in cases where there aren't multiple centers of the *same* agent in a single possible world, it avoids giving implausible verdicts in any of our three cases above. In the absence of some other relevant difference between Sleeping Beauty and these three cases, Time-Slicers who wish to avoid these three counterintuitive consequences should reject Thirding.

IV. The Third Argument: The Diachronic Argument Against Double Halfing

In the original exchange between Elga (2000) and Lewis (2001), which introduced the Sleeping Beauty problem to philosophers, it is striking that the Double Halfer view was never even mentioned as a possible solution to the problem. In fact, the debate over the problem continued for several years under the assumption that the Thirder and Lewisian Halfer positions were the only possible solutions. The reason for this widespread presumption against the Double Halfer position is that philosophers were relying on an essentially *diachronic* constraint on any possible solution

[13] Interestingly, while Titelbaum's (2013) Thirder framework does have counterintuitive consequences in Eternal Recurrence and Many Brains, Titelbaum has argued that it avoids the charge of easy confirmation in Quantum Measurement (p. 273 – 282). However, Titelbaum only avoids the charge of easy confirmation *given a certain theory about personal identity*. If the experimenter before the experiment is not identical to *either* person after the experiment, or if the experimenter before the experiment is identical to *both* persons after the experiment, Titelbaum's framework does not generate the desired probabilities. However, given an account of personal identity defended by Lewis (1976) and supplemented by Saunders and Wallace (2008), Titelbaum's account does give the desired probabilities. On this account, there really are two observers *before* the experiment, each of which is identical to one of the observers after the experiment. So, there really isn't an increase in the number of observers before and after the experiment! While this interpretation of the personal identity facts can be questioned on independent grounds (e.g. see Tappenden (2008)), the Time-Slicer will also be wary of the fact that this framework rejects *No Reference*, since it gives different verdicts depending on the correct theory of personal identity.

[14] This assumes that in cases of fission (like Quantum Measurement), neither of the post-fission agents is identical to the pre-fission agent.

to the problem. Namely, everyone agreed that Beauty's credence in Heads, whatever it is, should *increase* between Monday morning and Monday evening. The motivation for this diachronic constraint is straightforward. On Monday morning, Beauty has three options open to her corresponding to the two possible wakings on Tails and the one possible waking on Heads. Once Beauty is told that it is Monday on Monday evening, she is able to eliminate one of the possible Tails wakings. So, it seems to be a straightforward consequence of conditionalization that Beauty's credence in Heads should increase.

Given this diachronic constraint, Elga used as a (synchronic) premise that Beauty's credence in Heads should be 1/2 on Monday evening and retroactively inferred that her credence must have been 1/3 on Monday morning, and Lewis used as a (synchronic) premise that Beauty's credence in Heads should be 1/2 on Monday morning and inferred that her credence should increase to 2/3 in the evening. So, it seemed that one simply had to pick one's poison. One could either be a Halfer on Monday morning or a Halfer on Monday evening, but one couldn't be a Halfer at both times.

Many philosophers have questioned whether this diachronic constraint should really be adhered to in this case.[15] However, the Time-Slicer has an easy response here. Time-Slicers are free to choose neither poison and endorse both of Lewis' and Elga's synchronic premises. According to Time-Slicers there *just are* no essentially diachronic constraints on rationality; the fundamental norms of rationality are all synchronic. For any given case, the Time-Slicer should only be asking what credences an agent should assign at a particular time given the agent's evidence at that time, without any heed to what credences the agent assigned at any other time.

That being said, it might well be that there are diachronic norms of rationality which can be *derived* from purely synchronic norms of rationality. So one might naturally wonder whether the diachronic premise appealed to by Lewis and Elga can be derived in this way. Unfortunately, perhaps the most promising way to attempt such a derivation, which roughly follows the way Hedden (2015a, 2015b) derives a diachronic version of conditionalization, runs into difficulties when dealing with diachronic updates on self-locating information.

Here is how one might attempt such a derivation. First, stipulate as part of the case that Beauty has the same prior probability function on Monday morning and Monday evening, which encodes what credences she would have in every (centered and uncentered) proposition in the absence of any evidence.[16] Next, propose the following synchronic norm:

> *Synchronic Conditionalization*: If an agent A at time t has total evidence E and prior credence function C, then their credence at t in each proposition H should equal $C(H \mid E)$.

[15] Some philosophers who have questioned this diachronic constraint include Bostrom (2007), Meacham (2008), Briggs (2010), and Cozic (2011).

[16] Hedden himself endorses *Uniqueness*, which, in the context of Bayesian epistemology, is the claim that there is a uniquely rational prior probability function. While *Uniqueness* would guarantee that Beauty is rationally required to have the same prior on Monday morning and Monday evening, for our purposes we need not rely on it.

Given *Synchronic Conditionalization*, then *if one's evidence grows monotonically and one retains the same prior credence function*, then one's credences should evolve in the standard way governed by diachronic versions of conditionalization.[17]

In the context of Sleeping Beauty, however, Bostrom (2007) notes that Beauty's total evidence on Monday evening is *not* strictly greater than Beauty's total evidence on Monday morning. On Monday morning, it is part of Beauty's evidence that *I have not been told that it is Monday*, but on Monday evening, Beauty 'loses' this piece of evidence and learns its negation: *I have been told that it is Monday*! Consequently, there is no straightforward way to apply conditionalization in this case (as well as other cases where self-locating evidence is at issue).

In response to Bostrom, Titelbaum (2013) has developed a modeling framework, the 'Certainty-Loss Framework' (CLF), which entails that Beauty's credence in Heads should increase from Monday morning to Monday evening, even when accounting for Beauty's evidence that *I have been told that it is Monday* (p. 217-219). However, Titelbaum's framework invokes a crucial principle called (PEP), and as Titelbaum himself says, 'I categorize (PEP) as a diachronic systematic constraint of CLF' (p. 194). Time-Slicers, however, will be skeptical of any such diachronic principles. Meacham (2008) also notes that both Thirders and Lewisian Halfers need to endorse certain (underived) diachronic Continuity principles to get the diachronic results that they want to have. Again, Time-Slicers will not find any such primitive diachronic principles attractive. While the diachronic premise invoked by Elga and Lewis seems uncontroversial at first glance, it turns out to be very difficult to justify by only appealing to uncontroversial synchronic principles.

In sum, given Time-Slice Rationality, the only objections to the Double Halfer position that have any hope of being successful are ones that rely on entirely synchronic premises. Consequently, perhaps the most powerful objection to the Double Halfer position fails to get off the ground given Time-Slice Rationality.[18]

---

[17] For suppose that at t_1 your total evidence is E_1, and at t_2 you gain evidence E_2 (which makes your total evidence $E_1 \wedge E_2$). Then, if you have prior credence function C at both times, by *Synchronic Conditionalization*, your credences at t_1 should be $C_1(\text{-}) = C(\text{-} \mid E_1)$, and your credences at t_2 should be $C_2(\text{-}) = C(\text{-} \mid E_1 \wedge E_2)$. C_2 is just the probability function that results from taking C_1 and conditionalizing on E_2.

[18] One important synchronic argument against Double Halfing is given in Titelbaum (2012). If one supplements Sleeping Beauty with the claim that a fair coin will *also* be flipped on Tuesday (which will have no effect on the rest of the experiment), then it turns out that Double Halfers must assign a credence greater than 1/2 in the proposition that 'today's coin flip will land Heads' on Monday morning. In response, I grant that this a strong objection to Double Halfers who are primarily motivated by aligning their credences with the objective chances. However, this is not the relevant motivation for the Time-Slicer. The Time-Slicer is primarily motivated by having an *impersonal* epistemology, which is encapsulated in principles like *No Reference*, *Synchronicity*, and *Impartiality*. Given that the proposition 'today's coin will land Heads' is a merely indexical proposition, the fact that the Double Halfer gives an unintuitive verdict on this indexical proposition shouldn't be much of an embarrassment for the Time-Slicer.

V. The Costs of Denying Time-Slice Rationality

So far, I have only argued for the conditional claim that if you're a Time-Slicer, you should be a Double Halfer. In response, one might think that this conditional claim isn't terribly interesting, given that Time-Slice Rationality is a relatively new thesis, and it's dubious whether most philosophers would endorse it anyway. In this section, I will try to push back on this natural thought by drawing out some implausible consequences of any epistemology of self-location that explicitly denied Time-Slice Rationality.

Consider the following variant of Sleeping Beauty:

> Sorites Beauty$_X$: On Sunday, Beauty will be put to sleep. A fair coin will then be tossed. If it lands Heads, she will be woken up on Monday and then let go. If it lands Tails, she will be woken up on Monday and then put back to sleep. An evil neurosurgeon will then replace X% of Beauty's body with qualitative duplicate parts while she sleeps, so that the resulting body will be qualitatively just like the original body. The resulting agent, call her Tuesday Beauty, will then be woken up on Tuesday in the usual room.

Again, this case is structurally similar to Sleeping Beauty. Given Heads, there is one waking, and given Tails, there are two wakings, and all three possible wakings are indistinguishable. However, in this case the personal identity facts are being distorted. When X=0, the case should clearly be treated the same as Sleeping Beauty. When X=100, Tuesday Beauty is only a qualitative duplicate of Beauty, so Beauty ≠ Tuesday Beauty.[19]

There are two different ways one can react to Sorites Beauty$_X$, given that one treats Sleeping Beauty and Duplicating Beauty differently (presumably by being a Thirder in Sleeping Beauty and a Halfer in Duplicating Beauty, as per the goal of minimizing total inaccuracy). First, there is an externalist approach. On this externalist approach, if Tuesday Beauty is *in fact* identical to Beauty (regardless of what beliefs Beauty has about whether Tuesday Beauty is Beauty), then Beauty ought to assign credence 1/3 on Monday morning, just as in Sleeping Beauty. If Tuesday Beauty is *in fact* not identical to Beauty (regardless of what beliefs Beauty has about whether Tuesday Beauty is Beauty), then Beauty ought to assign credence 1/2 on Monday morning, just as in Duplicate Beauty.

This externalist approach clearly conflicts with *No Reference*. What credence Beauty ought to assign on Monday morning crucially depends on whether Beauty = Tuesday Beauty. Moreover, it also seems to conflict with *Synchronicity*. *Synchronicity* states that what attitudes you ought to have at a time does not directly depend on what attitudes you have at other times. However, on the externalist approach, what credence Beauty ought to have on Monday morning explicitly depends on whether it is *Beauty* who is adopting a credal state about the coin on Tuesday morning. If she

[19] If you do not believe that Beauty ≠ Tuesday Beauty in Sorites Beauty$_{100}$, then pick some other analogous soritical case where Beauty = Tuesday Beauty in Sorites Beauty$_0$, and Beauty ≠ Tuesday Beauty in Sorites Beauty$_{100}$

is the one adopting attitudes on Tuesday morning, Beauty should have a 1/2 credence on Monday morning, and if she is not adopting any attitudes on Tuesday morning, Beauty should have a 1/3 credence on Monday morning. So, the credence Beauty ought to have on Monday morning depends on what Beauty's attitudes are at other times.

Even bracketing *Synchronicity*, the externalist approach seems unattractive. It faces the familiar internalist worry that it does not give followable advice, and an agent would be epistemically blameless if they failed to follow the externalist advice.[20]

A second approach is an internalist one. According to this approach, if Beauty *believes* that Tuesday Beauty is identical to Beauty (regardless of whether Beauty is in fact identical to Tuesday Beauty), then Beauty ought to assign credence 1/3 on Monday morning just as in Sleeping Beauty. If Beauty *believes* that Tuesday Beauty is not identical to Beauty (regardless of whether Tuesday Beauty is in fact identical to Beauty), then Beauty ought to assign credence 1/2 on Monday morning, just as in Duplicate Beauty.

This internalist approach explicitly goes against *Impartiality*, which is the thesis that your *beliefs* about what attitudes you have at other times should play the same role as your beliefs about what attitudes other people have. According to the internalist approach, your beliefs about what attitudes *you* have at other times plays an entirely different role than your beliefs about what attitudes *other* people have (e.g. when Tuesday Beauty is not Beauty).

It is worth pausing to draw out some implausible consequences of the internalist version of the goal, even bracketing any commitment to the thesis of Time-Slice Rationality. What should the internalist version of the goal say when Beauty is unsure about whether Tuesday Beauty is identical to Beauty? For example, if we let X=50 in Sorites Beauty$_x$, it is natural to be unsure whether Tuesday Beauty is identical to Beauty.[21] Suppose, for example, that Beauty has a credence of 1/2 that Beauty = Tuesday Beauty in this case. We know that if Beauty = Tuesday Beauty, Sorites Beauty$_{50}$ should be treated like Sleeping Beauty, and if Beauty ≠ Tuesday Beauty, Sorites Beauty$_{50}$ should be treated like Duplicating Beauty. So, it looks like if Beauty has credence 1/2 that Beauty = Tuesday Beauty, Beauty should assign a credence of (1/2)*(1/2) + (1/2)*(1/3) = 5/12 to the coin landing Heads! So, in this case, neither Thirders nor Halfers are correct, but 5/12-ers are correct! It is telling that no author has ever offered any answer other than 1/3 or 1/2 in any structurally similar variation to the Sleeping Beauty case.

A second implausible consequence has to do with what sort of evidence this internalist version of the goal deems relevant to the proposition that the coin landed Heads. Suppose Beauty wakes up

[20] Of course, there are familiar externalist reasons for thinking that *no* epistemic theory can always give followable advice having to do with Williamson's (2000) Luminosity argument. For a response to the Luminosity argument, see Berker (2008). For motivations to be an internalist that do not appeal to Luminosity, see Schoenfield (2015).
[21] If you have determinate intuitions for the case of X=50, pick some other value of X for which you lack a determinate intuition.

on Monday morning and assigns credence 1/2 that Beauty = Tuesday Beauty, and hence assigns a credence of 5/12 to the proposition that the coin lands Heads. Suppose Beauty really wants to assign the best credence she can to the coin's landing Heads, so she spends all of Monday morning reading literature on the metaphysics of personal identity (before she is told that it is Monday). After reading many journal articles, her credence that Beauty = Tuesday Beauty drops to 1/4, and hence her credence that the coin will land Heads goes from 5/12 to (1/4)*(1/2)+(3/4)*(1/3) = 3/8. Intuitively, however, all these abstract arguments in the metaphysics journals should be treated as evidentially irrelevant to the coin's landing Heads or Tails. One shouldn't be able to pursue one's curiosity about the state of a fair coin by reading metaphysics journal articles on personal identity!

VI. Concluding Remarks

Apart from these three more specific arguments, it is worth noting that the general philosophical spirit behind the Time-Slice view and the Double Halfer view are remarkably similar. Both the Double Halfer and the Time-Slicer adopt a deflationary stance on the importance of 'I' in epistemology. Time-Slicers think that facts about which time-slices across the universe count as 'I' are simply irrelevant to one's theorizing. Time-slices that count as other people should be treated in the same way as time-slices that count as oneself. Similarly, the natural generalization of the Double Halfer position, in the form of the Halfer Rule, also adopts a deflationary take on the importance of 'I' in epistemology. The Halfer Rule entails the so-called 'Relevance Limiting Thesis' introduced by Titelbaum (2008), which says that one should only revise one's credences in uncentered propositions when one learns new uncentered propositions as evidence. In other words, essentially indexical information is *always irrelevant* to one's theorizing about non-indexical matters. For example, in Sleeping Beauty, since Beauty learns no new non-indexical information on Monday morning or Monday evening, she ought to retain her credence of 1/2 in Heads. Given the Time-Slicers deflationism about the role of 'I' in epistemology, a principle like the Relevance Limiting Thesis is a natural one to adopt. Combining the Double Halfer View and the Time-Slice view results in an epistemology that takes a sort of 'view from nowhere', which deflates the importance of the subject in one's inquiry into the objective, non-indexical world.

Meacham (2010) has argued that de se puzzle cases form a 'tangled web' with other issues that are independent of self-locating beliefs *per se*, such as the internalism vs externalism debate in epistemology and debates about how to handle identity over time. I believe my discussion has supported this contention that these de se cases form such a tangled web, but at the same time I hope to have offered a principled way to untangle this web for those who endorse Time-Slice Rationality. Time-Slicers should be Double Halfers.[22]

[22] Thanks to Miriam Schoenfield, Jack Spencer, Roger White, and an anonymous referee for helpful feedback.

References

Arntzenius, Frank. 2003. Some Problems for Conditionalization and Reflection. *Journal of Philosophy* 100, no. 7: 356–370.

Arntzenius, Frank, and Cian Dorr. 2017. Self-Locating Priors and Cosmological Measures. In *The Philosophy of Cosmology*, edited by Khalil Chamcham, John Barrow, Simon Saunders, and Joe Silk, 396–428. Cambridge: Cambridge University Press.

Berker, Selim. 2008. Luminosity Regained. *Philosophers' Imprint* 8: 1-22.

Bostrom, Nick. 2007. Sleeping Beauty and Self-Location: A Hybrid Model. *Synthese* 157, no. 1: 59–78.

Bradley, Darren. 2011a. Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty. *British Journal for the Philosophy of Science* 62, no. 2: 323–342.

Bradley, Darren. 2011b. Self-Location Is No Problem for Conditionalization. *Synthese* 182, no. 3: 393–411.

Bradley, Darren. 2012. Four Problems About Self-Locating Belief. *Philosophical Review* 121, no. 2: 149-177.

Bradley, Darren. 2015. Everettian Confirmation and Sleeping Beauty: Reply to Wilson. *British Journal for the Philosophy of Science* 66, no. 3: 683–693.

Briggs, Rachael. 2010. Putting a Value on Beauty. In *Oxford Studies in Epistemology, Volume 3*, edited by Tamar Szabo Gendler and John Hawthorne, 3-34. Oxford University Press.

Builes, David. 2019. Self-Locating Evidence and the Metaphysics of Time. *Philosophy and Phenomenological Research* 99, no. 2: 478–490.

Chalmers, David. 2011. Frege's Puzzle and the Objects of Credence. *Mind* 120, no. 479: 587–635.

Cozic, Mikael. 2011. Imagining and Sleeping Beauty: A Case for Double-Halfers.' *International Journal of Approximate Reasoning* 52, no. 2: 137–143.

Draper, Kai, and Joel Pust. 2008. Diachronic Dutch Books and Sleeping Beauty. *Synthese* 164, no. 2: 281–87.

Döring, Sabine, and Bahadir Eker. 2017. Rationality, Time and Normativity: On Hedden's Time-

Slice Rationality. *Analysis* 77, no. 3: 571–585.

Dorr, Cian. 2002. Sleeping Beauty: In Defence of Elga. *Analysis* 62, no. 4: 292–296.

Elga, Adam. 2000. Self-Locating Belief and the Sleeping Beauty Problem. *Analysis* 60, no. 2: 143–147.

Halpern, Joseph. 2006. Sleeping Beauty Reconsidered: Conditioning and Reflection in Asychronous Systems. In *Oxford Studies in Epistemology Vol. 1*, edited by Tamar Gendler and John Hawthorne, 111-142. Oxford University Press.

Hedden, Brian. 2015a. *Reasons Without Persons: Rationality, Identity, and Time*. Oxford University Press UK.

Hedden, Brian. 2015b. Time-Slice Rationality. *Mind* 124, no. 494: 449–491.

Hedden, Brian. 2016. Mental Processes and Synchronicity. *Mind* 125, no. 499 (2016): 873–888.

Hedden, Brian. 2017. Replies to Döring and Eker, Snedegar and Lenman. *Analysis* 77, no. 3: 607–618.

Hitchcock, Christopher. 2004. Beauty and the Bets. *Synthese* 139, no. 3: 405–420.

Horgan, Terry. 2004. Sleeping Beauty Awakened: New Odds at the Dawn of the New Day. *Analysis* 64, no. 1: 10–21.

Kierland, Brian, and Bradley Monton. 2005. Minimizing Inaccuracy for Self-Locating Beliefs. *Philosophy and Phenomenological Research* 70, no. 2: 384–395.

Leitgeb, Hannes, and Richard Pettigrew. 2010. An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science* 77, no. 2: 201–35.

Leitgeb, Hannes. 2010. Sleeping Beauty and Eternal Recurrence. *Analysis* 70, no. 2: 203–205.

Lenman, James. 2017. Reasons Without Humans. *Analysis* 77, no. 3: 586–595.

Lewis, David. 1976. Survival and Identity. In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 17–40. University of California Press.

Lewis, David. 1979. Attitudes de Dicto and de Se. *Philosophical Review* 88, no. 4: 513–543.

Lewis, David. 2001. Sleeping Beauty: Reply to Elga. *Analysis* 61, no. 3: 171–76.

Lewis, Peter J. 2006. Quantum Sleeping Beauty. *Analysis* 67, no. 1: 59–65.

Meacham, Christopher J. G. 2008. Sleeping Beauty and the Dynamics of de Se Beliefs. *Philosophical Studies* 138, no. 2: 245–269.

Meacham, Christopher J. G. 2010. Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs. In *Oxford Studies in Epistemology, Volume 3*, edited by Tamar Szabo Gendler and John Hawthorne, 86. Oxford University Press, 2010.

Monton, Bradley. 2002. Sleeping Beauty and the Forgetful Bayesian. *Analysis* 62, no. 1: 47–53.

Moss, Sarah. 2015. Time-Slice Epistemology and Action Under Indeterminacy. In *Oxford Studies in Epistemology*, edited by Tamar Szabó Gendler and John Hawthorne, 172–94. Oxford University Press.

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

Pettigrew, Richard. Unpublished Manuscript. Self-Locating Belief and the Goal of Accuracy.

Podgorski, Abelard. 2016. A Reply to the Synchronist. *Mind* 125, no. 499: 859–871.

Ross, Jacob. 2010. Sleeping Beauty, Countable Additivity, and Rational Dilemmas. *Philosophical Review* 119: 411-447.

Saunders, Simon, and David Wallace. 2008. Branching and Uncertainty. *British Journal for the Philosophy of Science* 59, no. 3: 293–305.

Schoenfield, Miriam. 2015. Internalism Without Luminosity. *Philosophical Issues* 25, no. 1: 252–272.

Sebens, Charles T., and Sean M. Carroll. 2018. Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics. *The British Journal for the Philosophy of Science* 69, no. 1: 25–74.

Snedegar, Justin. 2017. Time-Slice Rationality and Filling in Plans. *Analysis* 77, no. 3 : 595–607.

Stalnaker, Robert. 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

Tappenden, Paul. 2008. Saunders and Wallace on Everett and Lewis. *British Journal for the Philosophy of Science* 59, no. 3: 307–314.

Titelbaum, Michael G. 2008. The Relevance of Self-Locating Beliefs. *Philosophical Review* 117, no. 4: 555–606.

Titelbaum, Michael G. 2012. An Embarrassment for Double-Halfers. *Thought: A Journal of Philosophy* 1, no. 2: 146–151.

Titelbaum, Michael G. 2013. Ten Reasons to Care About the Sleeping Beauty Problem. *Philosophy Compass* 8, no. 11: 1003–1017.

Titelbaum, Michael G. 2013. *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press.

Weatherson, Brian. 2005. Should We Respond to Evil with Indifference? *Philosophy and Phenomenological Research* 70, no. 3: 613–635.

Weintraub, R. 2004. Sleeping Beauty: A Simple Solution. *Analysis* 64, no. 1: 8–10.

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

Wilson, Alastair. 2014. Everettian Confirmation and Sleeping Beauty. *British Journal for the Philosophy of Science*, no. 3: 573–598.