Sleeping Beauty and the Dynamics of *De Se* Beliefs

Christopher J. G. Meacham

1 Introduction

In standard possible worlds semantics, propositions are sets of possible worlds. To believe a proposition is to believe that your world is one of the worlds in that set. So the proposition that there are extraterrestrials is the set of worlds in which there are extraterrestrials, and to believe that there are extraterrestrials is to believe that your world is a member of that set.

A belief in a proposition is a belief about what the world is like. But in addition to beliefs about what the world is like, there are beliefs about where one is in the world. David Lewis (1979) has argued that these beliefs can't be expressed in terms of possible worlds. To accommodate beliefs about where we are in the world, Lewis proposed to extend standard possible worlds semantics by introducing *centered worlds*, possible worlds paired with individuals and times. A set of centered worlds is a *centered proposition*.[1] To believe a centered proposition is to believe that your current centered world is one of the centered worlds in that set. So the centered proposition that it's 9 am is the set of centered worlds at which it's 9 am, and to believe that it's 9 am is to believe that your current centered world is a member of that set.

Following Lewis, call beliefs that can be expressed in terms of possible worlds *de dicto* beliefs, and beliefs that can be expressed in terms of centered worlds *de se* beliefs. In his paper Lewis raises the question of what happens to Bayesian decision theory when we consider *de se* beliefs instead of *de dicto* beliefs. His answer is a natural one:

> "Very little. We replace the space of worlds by the space of centered worlds, or by the space of all inhabitants of worlds. All else is just as before."[2]

However, this answer is untenable. When you update your beliefs using standard Bayesian conditionalization your certainties are permanent: if you're certain a proposition is true before updating then you'll be certain it's true after updating. So on the account Lewis suggests, if you're certain that a centered proposition is true you will always remain certain that it's true. But suppose you're looking at a clock you know is accurate. If the clock reads 9 am, then you're certain of the centered proposition that it's 9 am. Given Lewis' suggestion, since you're certain of the centered proposition that

[1] Lewis himself calls them *properties*.

[2] Lewis (1979), p. 149.

it's 9 am when the clock reads 9 am, you should always remain certain that it's 9 am. So you should remain certain that it's 9 am a minute later, when the clock reads 9:01 am. Obviously, this is not how our beliefs should be updated.[3] We need a more sophisticated dynamics of *de se* beliefs.

Lewis (2001) himself employs a more sophisticated dynamics in his discussion of the sleeping beauty case. This case raises precisely the issue of how *de se* beliefs should change over time. By looking at the different treatments of the case, we can gain insight into the dynamics of *de se* beliefs. In this paper I'll look at three accounts of the sleeping beauty case: an account proposed by Adam Elga (2000), an account proposed by David Lewis (2001), and a third account I'll defend in this paper.

I'll offer two reasons for preferring my account over theirs. First, every account of what our credences should be must accommodate the temporal continuity of our credences. I'll show that the dynamics I propose accounts for this continuity naturally. This is not the case for the dynamics favored by Elga and Lewis. Second, I'll argue that Elga's and Lewis' treatments of the sleeping beauty case lead to highly counterintuitive consequences. I'll show that the account I offer also leads to consequences that some may find counterintuitive, but I'll argue that they're not as bad as those of Elga's account, and no worse than those of Lewis' account.

There are several other considerations that can be used to assess the merits of these accounts, such as betting arguments, considerations regarding reflection, etc. These issues are important, and have been addressed in a number of places.[4] They are beyond the limited scope of this paper, however, and I won't look at them here.

The rest of this paper will proceed as follows. In the next section I'll look at some natural dynamics for *de se* beliefs. In the third section I'll discuss the temporal continuity of beliefs, and I'll show that the dynamics I propose can account for this continuity naturally, while the dynamics of Elga and Lewis cannot. In the fourth section I'll present the sleeping beauty case and sketch the three responses to it. In the fifth and sixth sections I'll look at Elga's and Lewis' responses in detail, and show how they both lead to counterintuitive consequences. In the seventh section I'll critically examine my account by looking at some consequences of it that might also seem counterintuitive. In the eighth section I'll sum up my conclusions.

2 Belief Dynamics

It's standard to assume that belief is not an all-or-nothing affair, but rather admits of degrees. A subject's beliefs are represented by a probability function over the space of possibilities. The function assigns values between zero and one to regions of the space, representing the subject's confidence that some possibility in that region obtains.

[3] Arntzenius (2003), Halpern (2004) and Hitchcock (2004) have noted this problem of extending standard conditionalization to *de se* beliefs.

[4] For a sampling of this literature, see Elga (2000), Lewis (2001), Arntzenius (2002), Dorr (2002), Arntzenius (2003), Halpern (2004) and Hitchcock (2004). The article by Halpern is especially relevant to this paper, as he defends an account similar to the account defended here. As a result, much of what he says, such as his treatment of betting arguments, reflection, etc., applies to my account as well.

The values it assigns are countably additive: the value of the union of countably many non-overlapping regions of the space is the sum of the values of each of these regions. The value it assigns to the entire space of possibilities is one, representing the subject's certainty that some possibility or other obtains.

In the case of *de dicto* beliefs, the space of possibilities is the space of possible worlds. The credence function takes worlds as arguments, and assigns to each world a degree of belief, or credence. The credence assigned to a proposition is the sum of the credences assigned to each world in that proposition.[5] The worlds in which the subject has non-zero credences are the worlds she thinks might be hers, or her *doxastic worlds*.

When we generalize to *de se* beliefs, the space of possibilities becomes the space of centered worlds. The credence function takes centered worlds as arguments, and assigns to each centered world a credence. The credence assigned to a centered proposition is the sum of the credences assigned to each centered world in that centered proposition. The centered worlds in which the subject has non-zero credences are the centered worlds she thinks might be hers, or her *doxastic alternatives*.

Let's look at the dynamics of *de dicto* belief. At the core of an account of belief is an updating rule, a rule for generating new credences. The canonical updating rule is conditionalization. For simplicity let's focus on standard conditionalization, ignoring Jeffrey conditionalization and the like.[6]

On standard conditionalization you generate your current credences from your prior credences and your current evidence. To get your new credences you take your prior credences, set the credence in every world incompatible with your evidence to 0, and then normalize the credences in the remaining doxastic worlds; i.e., adjust the values such that they sum to 1, and such that the ratios between them are the same as the ratios between their prior credences. This way of updating makes certainties permanent. This is because you can only lose doxastic worlds in this process, not gain them. Being certain of a proposition P entails that all of your current doxastic worlds are compatible with P, and if you only lose doxastic worlds when you update then all of your future doxastic worlds will be compatible with P as well.

There is another version of *de dicto* conditionalization that does not have this consequence. For lack of a better name, call it *new conditionalization*. On new conditionalization, the role of prior credences in standard conditionalization is played by *hypothetical priors*. Hypothetical priors are a fixed set of values that encode a subject's epistemic norms; we can think of hypothetical priors as the credences a subject should have if she had no evidence whatsoever. On new conditionalization you generate your current credences from your hypothetical priors and your current evidence. To get your new credences you take your hypothetical priors, set the credence in every world incompatible with your evidence to 0, and then normalize the credences in the remaining doxastic worlds; i.e., adjust the values such that they sum to 1, and such that the ratios between them are the same as the ratios between their hypothetical priors.

Unlike standard conditionalization, new conditionalization does not make certainties

[5]Throughout the paper I'll ignore the complications that arise when we consider uncountably infinite numbers of worlds, and which we need measure theory to properly address.

[6]See Howson and Urbach (1993) for a description of Jeffrey conditionalization.

permanent, since subjects can lose *and* gain doxastic worlds. If your current evidence is compatible with worlds that your previous evidence was not, then you gain doxastic worlds when you update. This can happen, for example, when a subject suffers memory loss. If you have a perfect memory then your current evidence will include your memory of the previous evidence you've received, and you'll usually continue to rule out possibilities that you've ruled out in the past.[7] But if you forget evidence that ruled out certain worlds, then your current evidence will no longer rule those worlds out.

How should we generalize conditionalization to *de se* beliefs? As we saw in section 1, we cannot simply start with standard conditionalization and replace the space of worlds with the space of centered worlds. In the context of *de se* beliefs we both gain and lose possibilities, but standard conditionalization only allows the loss of possibilities.

We get a more promising account if we start with new conditionalization and replace worlds with centered worlds. Call this version of *de se* conditionalization *centered conditionalization*. On centered conditionalization you generate your current credences from your hypothetical priors and your current evidence. To get your new credences you take your hypothetical priors in centered worlds, set the credence in every centered world incompatible with your evidence to 0, and then normalize the credences in the remaining doxastic alternatives; i.e., adjust the values such that they sum to 1, and such that the ratios between them are the same as the ratios between their hypothetical priors.

Centered conditionalization is one way to modify new conditionalization in order to account for *de se* beliefs. However, centered conditionalization and unmodified new conditionalization are incompatible. To see this, consider a subject with just two doxastic worlds, A and B, with two doxastic alternatives at each world. Assume that her credences are divided equally among alternatives, so that her credence in each alternative is $\frac{1}{4}$ and her credence in each world is $\frac{1}{2}$. What should her credences in worlds A and B be if one of her alternatives at A is eliminated? According to new conditionalization her credences in A and B should remain $\frac{1}{2}/\frac{1}{2}$. Her evidence hasn't eliminated any doxastic worlds, so new conditionalization will assign the same credences. According to centered conditionalization, on the other hand, her credences in A and B should change. After the alternative at A is eliminated, centered conditionalization redistributes this credence among alternatives, so that her credence in each alternative is $\frac{1}{3}$. Since she has one alternative at A and two alternatives at B, her credence in A should now be $\frac{1}{3}$ and her credence in B should now be $\frac{2}{3}$.

There is another way to modify new conditionalization in order to accommodate *de se* beliefs that avoids this conflict. I'll call it *compartmentalized conditionalization*. On compartmentalized conditionalization you use new conditionalization to assign your credence in worlds, and then you distribute your credence in each world equally among the alternatives at that world. So on compartmentalized conditionalization, your credences are determined by your priors and your current evidence. Recall that your priors, like any probability function, are additive, so your prior in a world is the sum of your priors

[7] There can be exceptions to this if there are centered worlds that (i) you have a non-zero prior in, (ii) are subjectively identical to your current state of perfectly remembering your past evidence, and (iii) are located at worlds which you had previously eliminated.

in the centered worlds at that world. Given your priors and current evidence, you can determine your new credences in three steps. First, you take your hypothetical priors, and set the credence in every centered world incompatible with your current evidence to 0. Second, you normalize the credences in the remaining doxastic worlds; i.e., adjust the values assigned to each doxastic world such that they sum to 1, and such that the ratios between them are the same as the ratios between their priors. Finally, you distribute the credence assigned to each world equally among the remaining doxastic alternatives at that world.[8] So alternatives at the same world will always have the same credences, even if their priors are different.

In this paper I'll look at three accounts of the sleeping beauty case: my account, Elga's account and Lewis' account. An account of sleeping beauty requires an updating rule for *de se* beliefs, as well as further constraints on a subject's credences. The account I'm defending in this paper employs compartmentalized conditionalization and the Principal Principle. Elga's and Lewis' accounts both employ centered conditionalization as their updating rule, but differ on the other principles they adopt. We'll see what additional principles Elga's and Lewis' accounts employ in sections 5 and 6.

3 Continuity

3.1 Continuity and the Passage of Time

The dynamics of *de se* beliefs raise questions about belief continuity which don't arise in *de dicto* contexts. Consider again the case presented in the introduction, where a subject is watching a clock she knows to be accurate. When the clock changes from 9 am to 9:01 am, the subject discards all of her alternatives at which it's 9 am and replaces them with alternatives at which it's 9:01 am. It seems that her credence in these new alternatives should bear some relation to her credence in the alternatives they've just replaced. But nothing we've said so far requires that this be the case.

Suppose, for example, that the subject watching the clock has only two doxastic worlds, A and B, and that she has only one doxastic alternative at each world. Further suppose that she updates her beliefs using centered conditionalization and that at 9 am her priors in her two alternatives (A(9:00) and B(9:00)) are equal, so her credences in A(9:00) and B(9:00) are $\frac{1}{2}/\frac{1}{2}$. When she sees the clock register 9:01 am, what should her credences in A(9:01) and B(9:01) be? Intuitively, they should be $\frac{1}{2}/\frac{1}{2}$. But there is no reason they have to be this way. Although her priors in A(9:00) and B(9:00) are equal, at 9:01 am these are no longer her alternatives. Her alternatives are now B(9:01) and B(9:01), and nothing forces her to have equal priors in these alternatives.

[8] In extreme situations one can imagine subjects who have an infinite number of alternatives at a given world. In such cases the above prescription to divide the credence assigned to the world equally among the alternatives at that world won't do. Since Elga's and Lewis' accounts both endorse Elga's indifference principle (see section 5) this problem afflicts all three of the accounts I look at in this paper. For uncountably infinite numbers of alternatives, the natural move is to employ measure theory and adopt some suitably uniform measure over the alternatives. For countably infinite numbers of alternatives the situation is more difficult; employing a non-standard measure that rejects countable sub-additivity might be the best bet.

For subjects like us, who have a sense of time passing, every belief change will include a time changing component. As we notice time pass, we replace our old alternatives with new ones located at a later time. Since every evidential change brings an awareness that time has passed, every belief change involves the replacement of old alternatives with new ones. Nothing we've said so far entails that the beliefs of such subjects will be in any way constant—that their credences won't fluctuate wildly simply due to the passage of time—unless we impose further constraints on their credences. We intuitively think that there should be such constraints; constraints which require a rational subject's beliefs to be diachronically coordinated in the appropriate way. I'll call constraints of this kind *Continuity Principles*.

A Continuity Principle will take the following form: a subject's credences in her alternatives before and after a belief change should be diachronically coordinated when those alternatives are suitably related. For convenience, call a new and an old alternative whose credences should be diachronically coordinated *continuous alternatives*. To obtain a specific Continuity Principle we need to answer two questions. First, how should the credences of a pair of continuous alternatives be related? Second, when are a pair of alternatives continuous?

Let's start with the first question. Consider again the case of a subject watching a clock. In this case we're naturally inclined to assume that her A(9:00) and B(9:00) alternatives are continuous with her A(9:01) and B(9:01) alternatives, respectively. Intuitively what does this entail about her credences in these alternatives? It seems as if her credences in the new alternatives should be the same as her credence in the earlier alternatives they're continuous with. So if her credences in A(9:00) and B(9:00) are $\frac{1}{2}/\frac{1}{2}$, it seems her credences in A(9:01) and B(9:01) should be $\frac{1}{2}/\frac{1}{2}$ as well.

Of course, we don't want to require that credences in continuous alternatives always be the same. Suppose that at 9:01 am the subject learns ¬B, and so has only one alternative at 9:01 am, A(9:01). A(9:01) is continuous with A(9:00), but her credence in A(9:01) should be 1, not $\frac{1}{2}$. What we want is not for continuous alternatives to always have the same credences, but for continuous alternatives to have the same credences in similar evidential situations.

We can cash out this intuition more precisely as follows. Consider a pair of alternatives, A_1 and A_2, and an arbitrary set of centered worlds S not containing A_1 and A_2. Let $cr(A_1 : A_1 \vee S)$ be one's credence in A_1 if one's evidence is the centered proposition $A_1 \vee S$. Then if A_1 and A_2 are continuous, it should be the case that $cr(A_1 : A_1 \vee S) = cr(A_2 : A_2 \vee S)$. I.e., given otherwise identical evidence, A_1 and A_2 should be assigned the same credences. As we'll see in the next section, for the dynamics we're looking at this will yield the desired result that the subject's credence in A(9:01) and B(9:01) should be $\frac{1}{2}/\frac{1}{2}$.

Let's turn to the second question: when are a pair of alternatives continuous? Remember where the issue of continuity arises. For subjects like us, who have a sense of time passing, every belief change involves the replacement of old alternatives at a world with new ones. And it seems that our credences in these replacements should, if they're relevantly similar, be diachronically coordinated with our credences in the originals. So when we ask whether a pair of alternatives are relevantly similar, we're asking about

pairs where one is an old alternative, the other is a new alternative, and one replaces the other at a world. Restricting our attention to these kinds of cases gives us a necessary condition for continuity between alternatives: one of them must replace the other at a world.[9]

Is this condition sufficient as well as necessary? It isn't clear. Consider a subject about to undergo duplication. Before duplication she has one doxastic alternative at each of her doxastic worlds. After duplication she'll have two alternatives, one centered on the original individual and one centered on the duplicate. How should her credences in these two new alternatives be related to her credence in her original alternative? When we consider the original individual, it seems that the new alternative should be continuous with the old alternative she had before duplication, since it's just its temporal successor. But it's less clear what to think when we consider the duplicate individual. Should this new alternative be continuous with the old alternative of the original individual? I think it's not obvious what to say.

There are a number of other hard cases to consider, such as cases of fission, fusion, the addition and elimination of alternatives located at different times, and so on. These cases make it difficult to spell out precise necessary and sufficient conditions for the continuity of alternatives. Other than the necessary condition given above, I won't take a stand in this paper on what the criteria for continuity should be. Instead, I'll allow for a variety of Continuity Principles, differing in what standard of continuity they employ.

In the rest of this paper I'll assume that the subjects in question have a sense of time passing. As a result, several of the arguments I'll look at will require a Continuity Principle of some kind. In these places, I'll point out what standards of continuity are required for these arguments to go through.

3.2 Continuity and Dynamics

A Continuity Principle requires that we have diachronically coordinated credences in pairs of suitably related alternatives. I've left open the question of when pairs of alternatives are suitably related, but we can avoid making decisions about this by looking at a generic Continuity Principle. A Continuity Principle places a constraint on our credences. For the kinds of dynamics we looked at in section 2, this translates into a constraint on hypothetical priors. Since different dynamics generate credences from priors in different ways, how constraints on credences translate into constraints on priors will depend on the dynamics in question. Let's see what constraints on priors are imposed by a generic Continuity Principle given the two dynamics for *de se* beliefs we looked at in section 2, centered and compartmentalized conditionalization.

First, let's look at centered conditionalization. Recall that in otherwise identical evidential situations continuous alternatives should have the same credences. On centered conditionalization, a subject's credences are distributed among her doxastic alternatives in proportion to their priors. If one of two continuous alternatives has a different prior

[9]That is, (i) the new and old alternatives should be located at the same world, (ii) the subject had a non-zero credence in the old alternative before but not after the belief change, and (iii) the subject has a non-zero credence in the new alternative after but not before the belief change.

than its partner, then in otherwise identical evidential situations it will have a different proportion of the total priors, and therefore a different credence. So on centered conditionalization, the Continuity Principle requires continuous alternatives to have the same priors.[10]

What about compartmentalized conditionalization? On compartmentalized conditionalization, a subject's credences are distributed among worlds in proportion to her priors in those worlds, and then divided equally among the alternatives at each world. Now consider an evidential situation containing one of two continuous alternatives, and an otherwise identical evidential situation containing its continuous partner instead. These two evidential situations have the same doxastic worlds and the same number of alternatives at each world. So on compartmentalized conditionalization the credence assigned to an alternative at a world will be the same in both evidential situations, and since continuous alternatives are located at the same world, they'll be assigned the same credence. So if we adopt compartmentalized conditionalization we don't need to adopt a Continuity Principle; the diachronic coordination of our credences falls right out of the dynamics!

Compartmentalized conditionalization offers another advantage. We saw in section 3.1 that it's hard to give a precise characterization of when pairs of alternatives should be continuous. If we adopt compartmentalized conditionalization, we don't need to worry about this. Recall the necessary condition for a pair of alternatives being continuous: one replaces the other at a world. On compartmentalized conditionalization, any pair of alternatives that satisfies this necessary condition will have diachronically coordinated credences. So we don't need to worry about when pairs of alternatives are continuous, because any pair of alternatives that plausibly could be continuous—that satisfy the obvious necessary condition—will automatically have diachronically coordinated credences.

Let's look at an example of how compartmentalized conditionalization imposes diachronic coordination on our credences, and how centered conditionalization does not. Consider again the subject who is watching a clock she knows to be accurate, and who has two doxastic worlds, A and B. At 9 am she has one doxastic alternative at each world, A(9:00) and B(9:00), and has equal credence in each. When she sees the clock register 9:01 am she'll replace each of her 9 am alternatives with a 9:01 am alternative. What do centered and compartmentalized conditionalization require of her credences in these alternatives?

If she's a centered conditionalizer, the fact that her credences in A(9:00) and B(9:00) are equal entails that her priors in A(9:00) and B(9:00) must be equal. But this doesn't say anything about her priors, and thus her credences, in A(9:01) or B(9:01). So if she's a centered conditionalizer her credences in her 9:01 am alternatives can be completely unrelated to those of her 9 am alternatives.

If she's a compartmentalized conditionalizer, the fact that her credences in A(9:00) and B(9:00) are equal entails that her priors in worlds A and B are equal, although her

[10]With one exception: if the subject has a zero prior in every centered world except the two under consideration, then they will be assigned the same credence (one) in otherwise identical evidential situations regardless of their priors.

priors in the centered worlds A(9:00) and B(9:00) may not be. This doesn't say anything about her priors in A(9:01) or B(9:01), of course, but it doesn't matter. Her credences in A(9:01) and B(9:01) will be $\frac{1}{2}/\frac{1}{2}$ regardless of their priors, since her credences in A and B will be $\frac{1}{2}/\frac{1}{2}$, and A(9:01) and B(9:01) are the only alternatives at A and B. So if she's a compartmentalized conditionalizer she'll have diachronically coordinated credences, even though we haven't imposed any restrictions on her priors.

On centered conditionalization we need to invoke special purpose Continuity Principles in order to account for the diachronic coordination of our credences, and we need to work out an account of when pairs of alternatives are continuous. On compartmentalized conditionalization the diachronic coordination of our credences falls right out of the dynamics, and we don't need to bother providing an account of when alternatives are continuous. This is a substantial mark in favor of compartmentalized conditionalization.

4 Sleeping Beauty

An interesting case of *de se* belief change is the sleeping beauty case:

> *The Sleeping Beauty Case:* 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 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 your are sleeping, they will erase the memories of your waking.

When you wake up there is no way for you to know if it is Monday or Tuesday. If you are in the world in which the coin came up heads, then it's Monday. If you are in the world in which the coin came up tails, then it may be Monday or Tuesday. Suppose you then learn that it's Monday. Then you'll know what day it is, but you still won't know whether the coin came up heads or tails. There are two questions to ask here. First, what should your credences be when you wake up? Second, what should your credences be if you learn that it's Monday?

Let's look at what my account says. Assume the Principal Principle, that a subject's credences should line up with what she thinks the chances are (if she's not in possession of inadmissible information).[11] On compartmentalized conditionalization a subject first divides her credences among worlds, and then divides the credence of each world equally among the alternatives at that world. So a subject's credence in worlds, and thus in *de dicto* propositions, only changes when she gains or loses doxastic worlds.

On Sunday you will have a $\frac{1}{2}/\frac{1}{2}$ credence that the coin toss came up heads/tails by the Principal Principle, with one doxastic alternative at each of your doxastic worlds. When you wake up on Monday you have one alternative (Monday) at each heads world and two alternatives (Monday and Tuesday) at each tails world. But although your doxastic alternatives have changed, you have the same doxastic worlds you had on Sunday. Since your doxastic worlds have remained the same, you will have the same credence in heads/tails: $\frac{1}{2}/\frac{1}{2}$. How should your $\frac{1}{2}$ credence in tails be divided between Monday

[11] See Lewis (1980).

and Tuesday? On compartmentalized conditionalization your credence in tails is divided equally between these two alternatives, so your credence in Monday/Tuesday given tails will be $\frac{1}{4}/\frac{1}{4}$.

What if you then learn that it's Monday? This eliminates the Tuesday alternative at your tails worlds, but doesn't eliminate any doxastic worlds. So again, your credence in heads/tails will remain the same: $\frac{1}{2}/\frac{1}{2}$.

What do Elga and Lewis say about the sleeping beauty case? Elga (2000) proposes that upon waking we should have a $\frac{1}{3}$ credence in heads and a $\frac{2}{3}$ credence in tails, the latter split evenly between Monday and Tuesday. If you learn that it's Monday, you should be a centered conditionalizer and regain your original $\frac{1}{2}/\frac{1}{2}$ credence in heads/tails.

Lewis (2001) proposes that we retain our $\frac{1}{2}/\frac{1}{2}$ credence in heads/tails when we wake up, with our credence in tails split evenly between Monday and Tuesday. Lewis' account diverges from my account in what happens when you learn that it's Monday. Lewis holds that you should be a centered conditionalizer and come to have a $\frac{2}{3}$ credence in heads and a $\frac{1}{3}$ credence in tails.

Consider a subject with more than one doxastic world, who undergoes a belief change which just increases or decreases the number of alternatives at a world (to a minimum of 1). As we'll see, we can capture the flavor of these three accounts by looking at how such a belief change affects the subject's credence in that world. On my account the subject's credence remains unchanged. On Lewis' account, if the number of alternatives at that world increases then the subject's credence will remain unchanged. But if the number of alternatives at that world decreases, then the subject's credence will decrease as well. On Elga's account the subject's credence will change in both cases. If the number of alternatives at that world increases or decreases, then the subject's credence in that world will likewise increase or decrease.

In the next two sections I'll look in more detail at how Elga and Lewis treat the sleeping beauty case. Before we do that, a caveat is in order. Neither Elga nor Lewis offer an explicit account of the dynamics of *de se* beliefs they endorse. So in presenting Elga's and Lewis' arguments I've had to add implicit premises that their arguments require. That said, I take the accounts I offer on their behalf to be fair.

5 Elga's Response to Sleeping Beauty

In Elga's (2000) account of the sleeping beauty case, he proposes that after waking up our credence in heads/tails should be $\frac{1}{3}/\frac{2}{3}$, the latter split evenly between Monday and Tuesday. If we then learn it's Monday, he proposes that our credence in heads/tails should become $\frac{1}{2}/\frac{1}{2}$. Elga's proposal follows from four principles:

1. Centered Conditionalization
2. The Principal Principle
3. Elga's Indifference Principle
4. A Continuity Principle

We've already looked at centered conditionalization, and the Principal Principle is familiar. The third principle, Elga's (2004) Indifference Principle, states that your credences in doxastic alternatives at the same world should be equal.[12] The fourth principle is a Continuity Principle. As we've seen, the content of a Continuity Principle depends on when we take pairs of alternatives to be continuous. For Elga's proposal, any Continuity Principle the includes the the following sufficient condition for continuity will do: a new and old alternative are continuous if (i) both are centered at the same world and individual, (ii) the new alternative is centered at a later time than the old alternative, and (iii) there's no other new alternative satisfying (i) and (ii) that's centered at an earlier time.

Given these four principles, Elga's proposal follows. Let $\mathrm{cr}(\cdot)$ be your credence function and $\mathrm{hp}(\cdot)$ your hypothetical priors. Let H/T be the propositions that the coin came up heads/tails, and SUN/MON/TUE be the centered propositions that it's Sunday/Monday/Tuesday.

By the Principal Principle, your credences in your heads and tails alternatives on Sunday will be $\mathrm{cr}(\mathrm{H}\wedge\mathrm{SUN}) = \mathrm{cr}(\mathrm{T}\wedge\mathrm{SUN}) = \frac{1}{2}$. Given centered conditionalization, this entails that $\mathrm{hp}(\mathrm{H}\wedge\mathrm{SUN}) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{SUN})$. When you wake up on Monday, your Sunday alternatives are replaced by Monday alternatives at the heads worlds, and by Monday and Tuesday alternatives at the tails worlds. Both the Monday and the Tuesday alternatives are centered at the same worlds and individuals as the Sunday alternatives, and at later times. But the Monday alternatives are centered at an earlier time than the Tuesday alternatives. So according to the Continuity Principle given above, it's the Monday (not Tuesday) alternatives that are continuous with the Sunday alternatives. We saw in section 3 that given centered conditionalization, the Continuity Principle requires that the priors of continuous alternatives be the same. So $\mathrm{hp}(\mathrm{H}\wedge\mathrm{SUN}) = \mathrm{hp}(\mathrm{H}\wedge\mathrm{MON})$ and $\mathrm{hp}(\mathrm{T}\wedge\mathrm{SUN}) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{MON})$. Elga's Indifference Principle requires that your credences in the two alternatives at the tails worlds be equal, and given centered conditionalization this entails that $\mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{TUE})$. Putting these equalities together, we get $\mathrm{hp}(\mathrm{H}\wedge\mathrm{MON}) = \mathrm{hp}(\mathrm{H}\wedge\mathrm{SUN}) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{SUN}) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{TUE})$. When you wake up your doxastic possibilities are H∧MON, T∧MON and T∧TUE, so on centered conditionalization your credences after waking on Monday are $\mathrm{cr}(\mathrm{H}\wedge\mathrm{MON}) = \mathrm{cr}(\mathrm{T}\wedge\mathrm{MON}) = \mathrm{cr}(\mathrm{T}\wedge\mathrm{TUE}) = \frac{1}{3}$.

Now, say you're woken up at 9 am. What if at 9:01 am you learn that it's Monday? After learning it's Monday you will have one alternative at each world, H∧MON(9:01) at the heads worlds and T∧MON(9:01) at the tails worlds. By the Continuity Principle, $\mathrm{hp}(\mathrm{H}\wedge\mathrm{MON}(9{:}00)) = \mathrm{hp}(\mathrm{H}\wedge\mathrm{MON}(9{:}01))$ and $\mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}(9{:}00)) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}(9{:}01))$. We know from above that $\mathrm{hp}(\mathrm{H}\wedge\mathrm{MON}(9{:}00)) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}(9{:}00))$, so it follows that $\mathrm{hp}(\mathrm{H}\wedge\mathrm{MON}(9{:}01)) = \mathrm{hp}(\mathrm{T}\wedge\mathrm{MON}(9{:}01))$. So on centered conditionalization your credences after learning it's Monday are $\mathrm{cr}(\mathrm{H}\wedge\mathrm{MON}(9{:}01)) = \mathrm{cr}(\mathrm{T}\wedge\mathrm{MON}(9{:}01)) = \frac{1}{2}$.

[12] Elga (2004) proposes that subjectively indistinguishable alternatives at the same world should have the same credences. I'm assuming that one's current evidence includes (and is possibly exhausted by) one's current subjective state. It follows that all of one's alternatives are subjectively indistinguishable, and Elga's Indifference Principle becomes the claim that alternatives at the same world should have the same credences.

Note that the Principal Principle only plays a superficial role in the argument for Elga's proposal. The Principal Principle sets our credences in heads and tails on Sunday to $\frac{1}{2}/\frac{1}{2}$. But the argument goes through equally well given any reason for $\frac{1}{2}/\frac{1}{2}$ credences in heads and tails on Sunday. Likewise, the argument goes through just as well if heads and tails are replaced by two different hypotheses we have other reasons for having $\frac{1}{2}/\frac{1}{2}$ credences in.

In the sleeping beauty case it's uncontentious that the Principal Principle applies on Sunday, and thus that you should have $\frac{1}{2}/\frac{1}{2}$ credences in heads and tails. Some of the sleeping beauty literature has focused on whether the Principal Principle should also apply after you wake up on Monday.[13] The question is whether you get admissible evidence when you wake up on Monday. If so, the thought goes, then the Principal Principle should still apply, and your credences in heads and tails should remain $\frac{1}{2}/\frac{1}{2}$.

It follows from Elga's argument that upon waking our credences in heads and tails should be $\frac{1}{3}/\frac{2}{3}$. So if Elga's argument is sound, you do get inadmissible evidence when you wake up on Monday. But I think debating admissibility and the Principal Principle is the wrong way to approach the problem. First, there is no agreement as to what counts as admissible evidence.[14] This makes it hard to make progress in a debate over whether someone's evidence is admissible. Second, focusing on the issue of whether the Principal Principle applies on Monday gets us relatively little. As we just saw, the argument goes through just as well if heads and tails are replaced by two different hypotheses we have other reasons for having $\frac{1}{2}/\frac{1}{2}$ credences in. Concluding one thing or another about the Principal Principle doesn't tell us what to say in these other cases. Finally, suppose we decide that we don't receive inadmissible evidence upon waking, and therefore that Elga's argument is incorrect. We still need to decide what part of Elga's argument to reject, since the argument entails the $\frac{1}{3}/\frac{2}{3}$ result without making any assumptions about the admissibility of your evidence on Monday. The argument only requires that the Principal Principle hold on Sunday, before you go to sleep. Given this, I think it's better to assess the merits of Elga's argument and then see what implications this has regarding admissibility than to use admissibility to assess the merits of Elga's argument.

If one accepts Elga's argument, then belief changes that increase the number of doxastic alternatives at a world will generally increase one's credence in that world relative to worlds without such an increase. Likewise, one's credence in a proposition which multiplies doxastic alternatives will generally increase relative to propositions that don't multiply alternatives. One can see why this should be so for the proponent of Elga's response: to endorse Elga's response is to think that one's credence in tails should increase relative to one's credence in heads when the number of alternatives given tails increases (and the number of alternatives given heads does not).

However, accepting Elga's argument leads to counterintuitive consequences. Consider the following case, which I owe to Tim Maudlin:

[13] See Lewis (2001) and Dorr (2002).

[14] Though see Hall (2004) and Meacham (2005) for proposals regarding admissibility.

> *The Many Brains Argument:* Consider the hypothesis that you're a brain in a vat. I take it that this is epistemically possible and (perhaps) nomologically possible. Your current credence in this possibility, however, is presumably very low. Now 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 proposition be $0 < p < 1$, and your credence that there will be no multiplication of doxastic alternatives be $1-p$. If you accept Elga's argument then your credence in this hypothesis should be constantly increasing and will converge to 1. Thus, if you hold such a position you should come to believe (if not yet, then in a little while) that these brains in vats are being created.

It follows from Elga's Indifference Principle that your credences should be spread evenly among the doxastic alternatives at a world. So as you become certain that these brains in vats are being created, you should become certain that you're a brain in a vat.

The many brains argument assumed that brain in a vat duplication is the only proposition in which you have a non-zero credence that multiplies doxastic alternatives. Now suppose that you also have a small credence in the proposition that you're in a world where duplicates of you are constantly being created on distant but qualitatively identical planets. Then you'll come to believe (if not yet, then in a little while) that these brains in the vats are being created *or* that these duplicates of you are being created. Likewise, you'll come to believe that you are a brain in a vat *or* a duplicate on a distant planet. By a similar process, you can generalize the result of the many brains argument to any number of propositions that multiply alternatives.

In general, if you accept Elga's argument then you will come to believe that you're in a world where you have many doxastic alternatives. These are strange worlds. So if we accept Elga's argument, we'll come to believe (if not yet, then in a little while) that we live in a strange world. This is an unwelcome consequence.

6 Lewis' Response to Sleeping Beauty

In his criticism of Elga's account of sleeping beauty, Lewis (2001) claims that you do not receive inadmissible evidence when you wake up on Monday. Thus the Principal Principle should still apply on Monday, and your credence in heads/tails should remain $\frac{1}{2}/\frac{1}{2}$. I've said above why I think this is the wrong way to approach the problem. And as we saw, even if Lewis is right there remains the task of deciding what's wrong with Elga's argument, since the argument only requires that the Principal Principle apply on Sunday. So how would Lewis address Elga's argument? To reject the argument, Lewis needs to reject one of the four premises the argument employs. With Elga, Lewis accepts that the Principal Principle entails that our credences in heads and tails on Sunday should be $\frac{1}{2}/\frac{1}{2}$. Furthermore, Lewis endorses Elga's Indifference Principle and (centered) conditionalization. So Lewis must reject Elga's Continuity Principle.

In Lewis' (2001) account of the sleeping beauty case, he proposes that after waking up our credence in heads/tails should be $\frac{1}{2}/\frac{1}{2}$, the latter split evenly between Monday and Tuesday. If we then learn it's Monday, he proposes that our credence in heads/tails should become $\frac{2}{3}/\frac{1}{3}$. Lewis' proposal follows from five principles:

1. Centered Conditionalization
2. The Principal Principle
3. Elga's Indifference Principle
4. A Continuity Principle
5. The No-Increase Principle

The first three premises are familiar. The fourth premise is another Continuity Principle. Although Lewis must reject Elga's Continuity Principle, we can use it to characterize a Continuity Principle that will suit Lewis' purposes. Elga's Continuity Principle requires that any old and new alternative that satisfy the following conditions be continuous: (i) both are centered at the same world and individual, (ii) the new alternative is centered at a later time than the old alternative, and (iii) there's no other new alternative satisfying (i) and (ii) that's centered at an earlier time. Lewis' Continuity Principle requires that any pair of alternatives that satisfies these conditions be continuous *iff* the number of alternatives at that world has not increased.

Lewis' Continuity Principle needs to deny that pairs of alternatives that satisfy these conditions are continuous when the number of alternatives at that world increases. This leaves us with the question of what constraints, if any, should be imposed on your credences at worlds where the number of alternatives increases. Lewis' position seems to be that in cases where you don't get evidence about the world—where you don't gain or lose doxastic worlds—increases in the number of alternatives at a world should leave your credence in that world unchanged. I'll call this the No-Increase Principle.

Given these five principles, Lewis' proposal follows. As before, the Principal Principle and centered conditionalization entail that hp(H∧SUN) = hp(T∧SUN). When you wake up on Monday your Sunday alternatives are replaced by Monday alternatives at the heads worlds and by Monday and Tuesday alternatives at the tails worlds. By the Continuity Principle your Sunday alternatives at your heads worlds are continuous with your Monday alternatives at your heads worlds, and on centered conditionalization this entails that hp(H∧SUN) = hp(H∧MON).[15] By the No-Increase Principle the increase in alternatives at your tails worlds leaves your credence in tails unchanged, so your credence in tails after waking up on Monday is the same as your credence in tails on Sunday. Given centered conditionalization, this entails that hp(T∧SUN) = hp(T∧(MON∨TUE)) = hp(T∧MON) + hp(T∧TUE). Elga's Indifference Principle and centered conditionalization entail that hp(T∧MON) = hp(T∧TUE). Taken together, these equalities entail hp(H∧MON) = hp(H∧SUN) = hp(T∧SUN) = hp(T∧MON) + hp(T∧TUE) = 2·hp(T∧MON) = 2·hp(T∧TUE). When you wake up your doxastic possibilities are H∧MON, T∧MON and T∧TUE, so on centered conditionalization your credences after waking up on Monday are cr(H∧MON) = $\frac{1}{2}$ and cr(T∧MON) = cr(T∧TUE) = $\frac{1}{4}$.

Now what if you're woken up at 9 am and told at 9:01 am that it's Monday? After learning it's Monday you will have one alternative at each world, and by the Continuity

[15]We can actually derive one's credences after waking on Monday without using the Continuity Principle. But (a) the Continuity Principle *is* required for the derivation of what your credences should be if you then learn it's Monday, and (b) using it in this derivation as well makes things a bit clearer.

Principle these alternatives will be continuous with your Monday 9 am alternatives, i.e., $\text{hp}(\text{H}\wedge\text{MON}(9{:}00)) = \text{hp}(\text{H}\wedge\text{MON}(9{:}01))$ and $\text{hp}(\text{T}\wedge\text{MON}(9{:}00)) = \text{hp}(\text{T}\wedge\text{MON}(9{:}01))$. We know from above that $\text{hp}(\text{H}\wedge\text{MON}(9{:}00)) = 2{\cdot}\text{hp}(\text{T}\wedge\text{MON}(9{:}00))$, so it follows that $\text{hp}(\text{H}\wedge\text{MON}(9{:}01)) = 2{\cdot}\text{hp}(\text{T}\wedge\text{MON}(9{:}01))$. So on centered conditionalization your credences after learning it's Monday are $\text{cr}(\text{H}\wedge\text{MON}(9{:}01)) = \frac{2}{3}$ and $\text{cr}(\text{T}\wedge\text{MON}(9{:}01)) = \frac{1}{3}$.

Elga's account ran into problems because it entailed that belief changes that multiply alternatives at a world generally increase one's credence in that world. Lewis avoids this result by adopting a different Continuity Principle and the No-Increase Principle. But while on Lewis' account belief changes that multiply alternatives at a world don't increase one's credence in that world, belief changes that decrease the number of alternatives at a world generally do decrease one's credence in that world. And this leads to counterintuitive consequences for his account as well. Consider the following case:

> *The Sadistic Scientists Argument:* Consider the hypothesis that you're in a world where every second some scientists will create n brains in vats in situations subjectively identical to your own. A half second after the brains are created, the scientists will destroy them. Let your credence in this proposition be $0 < p < 1$, and your credence that there will be no creation or destruction of doxastic alternatives be $1 - p$. When the brains are created your credence that you are in such a world will remain the same (No-Increase Principle), and this credence will be evenly split between your $n + 1$ alternatives (Indifference Principle). As a half second passes and these brains are destroyed, your credence that you are in such a world will decrease by the appropriate amount (Continuity Principle and centered conditionalization). So as each second passes, your credence that you are in such a world will decrease and converge to 0. Thus, if you hold Lewis' position you should come to believe (if not yet, then in a little while) that these brains in vats are not being created.

The sadistic scientists argument assumed that brain in vat destruction is the only proposition you have a non-zero credence in that diminishes alternatives. Now suppose that you also had a small credence in the proposition that duplicates of you on distant but qualitatively identical worlds were being created and destroyed. Then you'd come to believe (if not yet, then in a little while) that neither of these propositions was true. The result generalizes to any number of propositions that diminish alternatives. In general, if you accept Lewis' argument then you'll come to believe that you're not in a world where continual doxastic elimination is taking place.

I take this result to be counterintuitive. If the result as stated does not move you, imagine a case in which you are living in a world where brain-in-a-vat creation technology is cheap and easily accessible. An enemy of yours who would enjoy destroying brains in vats in your subjective state tells you that at midnight she'll spend an hour creating n such brains, and at 1 am she'll spend an hour destroying them. This enemy has the resources to carry out this threat, and reliably carries out the threats she makes. If n is big enough, and you uphold the Lewis' account, then though you're now almost certain that she will carry out her threat, when you wake up tomorrow morning you'll be almost certain that she didn't. Indeed, if n is big enough, you could even go with her and watch as she creates the brains and destroys them; if you watch for long enough you won't believe your eyes!

7 The Varied Brains Argument

In the last two sections I've argued that Elga's and Lewis' positions lead to counterintuitive consequences. Now let's turn a critical eye toward my account.

Consider a case like sleeping beauty, but with the following twist. If the coin toss comes up tails, they'll put you in a black room on Monday and a white room on Tuesday. If the coin toss comes up heads, they'll flip another coin to determine whether to put you in a black or white room on Monday.

What should your credences be in this case on the three accounts we've looked at? On all three accounts your credences in heads and tails on Sunday will be the same as in the sleeping beauty case. Likewise, on all three accounts your credences in heads and tails after waking up on Monday before you open your eyes will be the same as in the sleeping beauty case. What about your credences in heads and tails after you open your eyes and see a black room? On Elga's and Lewis' accounts your credences will be the same: half of the heads worlds are eliminated and half of the tails alternatives are eliminated, and after renormalizing you get the same credences in heads and tails as before. Not so for compartmentalized conditionalization. When you eliminate half of the heads worlds you distribute this credence among the other worlds, but when you eliminate half of the tails alternatives you give that credence to the other alternative at the same world. So when you open your eyes and see a black room, your credence in tails worlds will increase, and your credence in heads/tails will become $\frac{1}{3}/\frac{2}{3}$.

This raises a natural worry for my account. I offered the many brains argument as a criticism of Elga's $\frac{1}{3}/\frac{2}{3}$ response to the sleeping beauty case. In the black and white room version of sleeping beauty compartmentalized conditionalization also ends up assigning $\frac{1}{3}/\frac{2}{3}$ credences to heads and tails. Is there an argument analogous to the many brains argument against compartmentalized conditionalization?

Yes and no. Let's look at how such an argument might go. The many brains argument itself won't work because on compartmentalized conditionalization multiplying alternatives at a world doesn't increase the likelihood of that world. As long as our doxastic worlds remains the same, our credences in worlds will remain the same. To get an argument analogous to the black and white room case, we need an argument where the normal worlds are eliminated but the alternative multiplying worlds are not. So consider the following case:

> *The Varied Brains Argument:* Assume that your doxastic worlds are such that they can be divided into two kinds of worlds, normal worlds and strange worlds. Throughout your doxastic worlds there are n subjectively distinguishable experiences that you might experience in the next second. Assume that you have some normal doxastic world compatible with each experience, and you have no subjective duplicates at your normal doxastic worlds. Assume that at each of your strange doxastic worlds there are scientists that will create n brains in vats a second from now, each brain compatible with one of your possible experiences. Now, at the end of a second you'll have some experience, say that of eating chocolate ice cream. This will eliminate the many normal worlds in which you don't have the experience of eating chocolate ice cream. On the other hand, at all of your

> strange worlds there's a brain in a vat which has the experience of eating chocolate ice cream, so no strange worlds will be eliminated. By compartmentalized conditionalization, your credence in your strange doxastic worlds should increase relative to your credence in your normal doxastic worlds.

We can extend this case by replacing 'second' with longer units of time, and as the unit of time grows larger, the number n of distinguishable experiences you might experience grows larger as well. By making the unit of time arbitrarily large, we can get a case in which, on compartmentalized conditionalization, your credence in your strange doxastic worlds grows arbitrarily large.

How bad is this?

One might question whether this result is counterintuitive. This is an interesting, if murky, question. But it is worth looking at how things stand if we decide that the result is counterintuitive.

In the varied brains case, your credence in your strange worlds increases relative to your credence in your normal worlds because of the artificial way in which these doxastic worlds have been selected: all the strange worlds under consideration are ones that will end up matching whatever you experience, whereas many of your normal worlds won't match what you experience. If we restricted the normal worlds to those compatible with eating chocolate ice cream, your credence in your strange worlds would not increase relative to your credence in your normal worlds. Likewise, if we placed no restrictions on which strange worlds were allowed, then the experience of eating chocolate ice cream would eliminate lots of strange worlds as well as lots of normal worlds. Whether your credence in strange worlds increases relative to your credence in normal worlds depends on which strange and normal worlds are your doxastic worlds—which worlds our priors and evidence lead us to believe could be ours. And it's reasonable to think that if you have doxastic worlds like ours, your credence in strange worlds will not gain on your credence in normal worlds.

Skeptical results can be roughly divided into two kinds. First, there are results which entail that people like us in situations like ours should be lead to skepticism. Second, there are results which entail skeptical consequences for people in outlandish situations, but which have little bearing on people like us. I take it that the first kind of result is worse than the second. Our general sentiment is that our intuitions in outlandish situations are less reliable—and thus easier to discard—than our intuitions in situations we're familiar with. Likewise, it's easier to bite the bullet with counterintuitive cases that have little impact on our everyday lives.

The varied brains argument is a result of the second kind; it entails that people with certain idiosyncratic doxastic set-ups will come to believe something counterintuitive. The many brains argument, on the other hand, is a result of the first kind; it entails that people like us should come to believe that we live in a strange world. So the skeptical arguments considered weigh more heavily against Elga's account than they do against the account I favor.

What about the sadistic scientists argument? This too is a result of the second kind. While people like us will become more and more sure we're not in a 'diminishing' world, this will have little effect on overall belief distribution since our credences in such

possibilities are so small. Only people whose initial credence in these strange worlds are high will be lead to counterintuitive results. So the skeptical arguments, considered in isolation, don't leave us with a reason to favor the account I advocate over Lewis' account. It is other considerations, such as the *prima facie* plausibility of the view, the implications with regards to reflection and continuity, etc., that will decide between the two views.

8 Why Compartmentalized Conditionalization?

We can sum up the intuitive difference between the three accounts with the following case:

> *The Up-and-Down Case:* Suppose you learn that you'll be part of the following experiment. Some scientists will flip a fair coin tonight. If it comes up tails, then every day at noon the scientists will create n brains in vats in states subjectively identical to yours, and at midnight will destroy $\frac{n}{2}$ of them. If it comes up heads, no brains will be created or destroyed.

If you endorse Elga's account then your credence that the coin came up tails will converge to 1, regardless of your evidence (knowledge of objective chances, etc.) to the contrary. If you endorse Lewis' account then your credence that the coin came up heads will converge to 1, regardless of your evidence (knowledge of objective chances, etc.) to the contrary. If you endorse my account, then your credences in heads and tails will remain $\frac{1}{2}/\frac{1}{2}$.

In this paper I've offered two reasons to adopt the third option. First, the dynamics of my account has a substantial advantage over the dynamics of Elga's and Lewis' with regards to accommodating the continuity of our beliefs. Accounts that adopt centered conditionalization need to invoke special purpose Continuity Principles in order to accommodate the diachronic coordination of our credences, and need to work out when pairs of alternatives are continuous. Accounts that adopt compartmentalized conditionalization get the diachronic coordination of our credences for free, and don't need to work out when alternatives are continuous. Second, I've shown that while all three accounts arguably suffer from counterintuitive consequences, the consequences faced by my account are better than those faced by Elga's account, and no worse than those faced by Lewis' account.

These aren't the only considerations relevant to the assessment of these three accounts. There are further questions about betting arguments, reflection principles, and the like.[16] But if what I've said is right, these two considerations provide compelling reasons in favor of my account.[17]

[16] See footnote 3 for some references to literature on these topics, and their bearing on my account.

[17] I'd like to thank Frank Arntzenius, Maya Eddon, Adam Elga, Hilary Greaves, John Hawthorne, David Manley, Tim Maudlin, Adam Sennet, Ted Sider and Jonathon Weisberg for valuable comments and discussion. In particular, I owe much to Tim Maudlin, who inspired my interest in these issues, and David Manley, for raising the black and white room case. Finally, I owe a special thanks to John Hawthorne, Maya Eddon, and Frank Arntzenius for comments on a number of drafts of this paper.

References

Arntzenius, F. (2002) "Reflections on Sleeping Beauty", *Analysis*, pp. 53-61

Arntzenius, F. (2003) "Self-locating Beliefs, Reflection, Conditionalization and Dutch Books", *Journal of Philosophy*, pp. 356-70

Dorr, C. (2002) "Sleeping Beauty: in defense of Elga", *Analysis*, pp. 292-6

Elga, A. (2000) "Self-locating belief and the Sleeping Beauty problem", *Analysis*, pp. 143-7

Elga, A. (2004) "Defeating Dr. Evil with self-locating belief",*Philosophy and Phenomenological Research*, forthcoming

Hall, N. (2004) "Two Mistakes About Credence and Chance", *Australasian Journal of Philosophy* 82, pp. 93-111

Halpern, J. (2004) "Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems", *Proceedings of the Twentieth Conference on Uncertainty in AI*, pp. 226-34

Hitchcock, C. (2004) "Beauty and the Bets", *Synthese*, pp. 405-20

Howson, C. and Urbach, P. (1993) *Scientific Reasoning: The Bayesian Approach, 2nd Ed.*

Lewis, D. (1979) "Attitudes *De Dicto* and *De Se*" in *The Philosophical Review*, pp. 513-43

Lewis, D. (1980) "A Subjectivist's Guide to Objective Chance" in *Studies in Inductive Logic and Probability, Vol. 2*

Lewis, D. (2001) "Sleeping Beauty: reply to Elga", *Analysis*, pp. 171-6

Meacham, C. (2005) "Three Proposals for a Theory of Chance", *Philosophical Perspectives*, forthcoming