Ten Reasons to Care About the Sleeping Beauty Problem

Michael G. Titelbaum*
University of Wisconsin-Madison

Abstract

The Sleeping Beauty Problem attracts so much attention because it connects to a wide variety of unresolved issues in formal epistemology, decision theory, and the philosophy of science. The problem raises unanswered questions concerning relative frequencies, objective chances, the relation between self-locating and non–self-locating information, the relation between self–location and updating, Dutch Books, accuracy arguments, memory loss, indifference principles, the existence of multiple universes, and many–worlds interpretations of quantum mechanics. After stating the problem, this article surveys its connections to all of these areas.

Adam Elga (2000) introduced the Sleeping Beauty Problem to the philosophical literature[1] as follows:

> Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking.

> When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?

Since Elga's article, a virtual flood of arguments concerning this problem has appeared – a quick online search reveals over 40 articles about Sleeping Beauty in just the major philosophy journals over the last decade. So besides wondering what the answer to the problem is, someone new to Sleeping Beauty might wonder why so many people care. Philosophers love a good puzzle, and part of the attraction of Sleeping Beauty is surely that no one has solved it to everyone's satisfaction (or even to most people's satisfaction). But that's true of a number of puzzles in philosophy. I will suggest that the Sleeping Beauty Problem has attracted so much attention because it highlights a wide variety of unresolved issues in formal epistemology, decision theory, and the philosophy of science. The Sleeping Beauty Problem raises unanswered (and sometimes previously unasked) questions concerning: relative frequencies, objective chances, the relation between self-locating and non–self-locating information, the relation between self-location and updating, Dutch Books, accuracy arguments, memory loss, indifference principles, the existence of multiple universes, and many–worlds interpretations of quantum mechanics.

A few stipulations before we begin: (1) I will follow the literature in referring to the agent in the story as 'Beauty';[2] (2) I will refer to the time before the researchers put Beauty to sleep as Sunday night, and assume that before putting her to sleep, they fill her in on all the details of the experiment (including any modifications to the experiment we make later on); (3) I will assume that however the coin comes out Beauty first awakens on Monday morning, and if

the coin lands tails, her second awakening occurs on Tuesday morning; (4) I will embrace the Subjective Bayesian positions that agents' degrees of belief (or 'credences') in propositions can be represented on a scale from 0 (certainty of falsehood) to 1 (certainty of truth) and that *rational* agents' degrees of belief obey Kolmogorov's (1950) probability axioms;[3] and (5) I will adopt the unanimous opinion among Sleeping Beauty writers that because the coin is fair, on Sunday night Beauty should have a 1/2 credence in heads.

1. Sleeping Beauty and Relative Frequencies

Elga thinks that when Beauty awakens on Monday morning, her credence in heads should be 1/3. His initial argument invokes relative frequencies. If you expect that in the long run, a particular experimental setup would yield a particular result fraction *f* of the time, it seems you should have credence *f* that the result will occur if you run the experiment just once. If we ran the Sleeping Beauty experiment 1000 times, say, we would expect the coin to come up heads roughly 500 times. Each time the coin came up heads, Beauty would be awakened once, so over the 1000 trials, we would expect her to have 500 awakenings on which the coin had come up heads. We would also expect the coin to come up tails roughly 500 times, each of which would generate two awakenings. So we would expect 500 heads awakenings and 1000 tails awakenings from the 1000 trials. Out of the 1500 total expected awakenings, 500 of those would be heads awakenings. Since Beauty expects that in the long run, one third of her awakenings would be heads awakenings, when she awakens Monday morning (unsure whether it's Monday morning), she should be 1/3 confident that she is in the middle of a heads awakening. In other words, she should be 1/3 confident that the coin came up heads.

Principles requiring credences to match long run relative frequencies face a notorious difficulty: the Reference Class Problem.[4] When I refuse to loan my teenage son my car, he might protest that only a small fraction of overall drivers get in serious accidents. Yet my credence that harm will befall him is probably set using a narrower reference class: data about what fraction of *teenage boys* get into serious accidents. Different reference classes often yield different frequency values for the same result, and it can be difficult to know which frequency one should rely on in setting credences. In Sleeping Beauty, why is the relevant statistic the fraction of *awakenings* that involve the coin's having come up heads? Notice that if we ran the experiment 1000 times, we would expect 1/2 of the *experimental runs* to yield a heads result. So when Beauty finds herself awake in the middle of an experimental run, why shouldn't she be 1/2 confident that it's a run in which the coin comes up heads – and therefore 1/2 confident in heads?

Our principles for setting credences in line with relative frequencies don't seem settled enough to choose between these two reference classes. So relative frequencies don't seem to settle the Sleeping Beauty Problem.[5]

2. Sleeping Beauty and Objective Chances

There's something strange about the 1/3 answer to Sleeping Beauty. The coin is fair; it has a 1/2 objective chance of landing heads; Beauty is certain of this. When Beauty awakens on Monday morning, she seems to know nothing about the outcome of the coin flip other than this objective chance. So shouldn't her credence in heads be 1/2?

Elga responds to the 'halfer' position on Sleeping Beauty – that Beauty should be equally confident of heads and tails on Monday morning – by pointing out an unwelcome consequence. Notice that the outcome of the coin flip doesn't affect anything in the experimental protocol until *after* Beauty is put back to sleep on Monday. So Elga suggests making two changes to the experiment: First, the coin flip is moved from Sunday night to

Monday night. Because it doesn't change the protocol, this shouldn't have any effect on Beauty's rational Monday-morning credences. Second, Elga imagines that after waking Beauty up on Monday and letting her sit around a while, the experimenters tell Beauty that it's Monday before putting her back to sleep (and erasing her memory of everything that occurred on the Monday awakening). Let's call the time on Monday when Beauty is certain that it's Monday 'Monday evening'.

Historically, Subjective Bayesians have embraced a norm for updating credences called Conditionalization.[6] Elga notes that if Beauty updates her credences between Monday morning and Monday evening by conditionalizing on her newly gained information that it's Monday, her confidence in heads will increase. Some mathematics is involved, but the basic idea is that when Beauty awakens on Monday morning, she entertains three possibilities: It's Monday, and the coin comes up heads; it's Monday, and the coin comes up tails; or it's Tuesday, and the coin comes up tails.[7] (If the coin comes up heads, Beauty doesn't awaken on Tuesday.) Learning that it's Monday eliminates the third possibility, and eliminating a tails possibility should increase Beauty's confidence in heads.

So it looks like the halfer must maintain that upon learning it's Monday, Beauty should be more than 1/2 confident in heads. That's the unwelcome consequence. After all, it's Monday evening, the coin hasn't been flipped yet (because of Elga's first modification), the coin is fair, and Beauty is certain of all this. So how can her rational credence in heads be other than 1/2?

David Lewis (2001) recognized this unwelcome consequence of his own halfer view. He also was famously committed to the Principal Principle (Lewis, 1980), which requires an agent to assign credences equal to chance values of which she is certain unless she has what Lewis called 'inadmissible' evidence. An obvious example of inadmissible evidence is information causally downstream from the outcome of the chance process in question. For example, if you watch a fair die roll come up 3, you're allowed to change your credence in that outcome while remaining certain that its objective chance was 1/6. But Beauty doesn't seem to have any information causally downstream from the coin flip on Monday evening. So Lewis cast about for a 'novel and surprising' (2001, p. 175) kind of inadmissible evidence that would authorize a greater-than-1/2 Monday-evening credence in heads.[8]

Uncomfortable with Lewis's maneuvers, many halfers nowadays take a different route – they deny that Beauty should generate her Monday-evening credences by conditionalizing her Monday-morning credences on the information that it's Monday. (As we'll discuss in Section 4, Conditionalization has independent problems with Sleeping Beauty, so this move isn't as *ad hoc* as it might sound.[9]) Denying the conditionalization move opens up a 'double-halfer' position, on which Beauty maintains a 1/2 credence in heads throughout her Monday awakening.

So there are three major positions on the Sleeping Beauty Problem, summarized in this table (where the numerical values represent Beauty's credence in heads at the given time):[10]

	Thirder	Lewisian halfer	Double-halfer
Sunday night	1/2	1/2	1/2
Monday morning	1/3	1/2	1/2
Monday evening	1/2	> 1/2	1/2

The double-halfer position seems to have the advantage that it allows Beauty to keep her credences equal to objective chances throughout the experiment. But that triumph is

short-lived. Suppose that besides moving the fateful coin flip to Monday night (as Elga suggests), the experimenters also tell Beauty that they're going to perform a second fair coin flip on Tuesday night. The Tuesday flip is guaranteed to have no consequences for Beauty's experience; the experimenters just enjoy flipping their coin. When Beauty awakens Monday morning (uncertain whether it's Monday or Tuesday), she can assign a credence to the proposition that *today's* flip will come up heads. After a bit of calculation, it turns out that on *any* halfer position – Lewisian or double, with or without Conditionalization – Beauty's credence that today's flip will come up heads must be greater than $1/2$.[11] This, despite the fact that Beauty is certain today's flip is fair and has not yet occurred!

It looks like no matter what one's position, at some point Beauty's credences must diverge from the objective chances on some coin-flip proposition. At some point in the Sleeping Beauty Problem, Beauty comes to possess a novel and surprising kind of inadmissible evidence. Since our understanding of inadmissible evidence is inadequate to determine exactly when that point is, reasoning from objective chances doesn't seem to solve the Sleeping Beauty Problem.

3. Sleeping Beauty, Self-Location, and Relevance

The Principal Principle is a *synchronic* principle; it relates an agent's credences at a given time to her credences at the *same* time about objective chances. But there's also a *diachronic* argument for halfing. Everyone agrees that on Sunday night, Beauty should be $1/2$ confident in heads. Between Sunday night and Monday morning, all she learns is 'Today is Monday or Tuesday.' Information about the passage of time doesn't seem relevant to facts about coin flips, so Beauty's credence in heads on Monday morning shouldn't *change* from what it was on Sunday night – namely $1/2$.

This was Lewis's main argument for halfing; its crucial premise is that 'Today is Monday or Tuesday' is irrelevant to credences about the outcome of the coin flip. One might argue for that irrelevance by distinguishing *uncentered* information about what the world is like (such as 'Columbus reaches the New World in 1492') from *centered* information about one's identity or spatio-temporal location in the world (such as 'I'm Columbus'). Beauty learns only centered information between Sunday night and Monday morning, while propositions about the flip outcome are uncentered.

That means halfing can be defended with the Relevance-Limiting Thesis: An agent who learns only centered information between two times should not change her credences about uncentered matters.[12] We can recast this thesis using the distinction between centered and uncentered *worlds*. An uncentered world is a standard possible world, while a centered world selects within such an uncentered world a particular time and individual. In those terms, the Relevance-Limiting Thesis says that if what you learn between two times does not *remove* any uncentered worlds from consideration, it shouldn't change your credence *distribution* over uncentered worlds either. The thesis can be motivated by considering cases in which an agent is certain in advance exactly what is going to happen to her over a particular period of time. As events unfold, the agent's centered credences (tracking where she is in the course of her experience) definitely change, but these predictable alterations don't authorize any change in her opinions about what the world is generally like.[13]

Unfortunately, the Relevance-Limiting Thesis is false and can be shown to be false by examples independent of the Sleeping Beauty Problem. Imagine you're one of ten people standing in a circle in a room. Some scientists will flip a fair coin: If it comes up heads, they'll put nine black balls in a bag and one white; if it comes up tails, they'll put in one black and nine white. The bag will then be passed around the room, and each person will draw one

ball without replacement. Everyone in the room is informed of the experimental protocol, but no one is allowed to see the outcome of the coin flip or the ball anyone else has drawn.

You draw your ball and see that it's black. This should increase your confidence in heads from 1/2 to 9/10. But now suppose you have no non–indexical way of picking yourself out from among the ten people. You all look the same (so you can't uniquely describe yourself by appearance), the room is cylindrical (so you can't pick yourself out by absolute position), etc. Then, the only way to describe your new evidence is 'I picked a black ball.'[14] But that's centered evidence, and it has increased your rational credence in the uncentered proposition that the coin came up heads. So the Relevance-Limiting Thesis is false.[15] There is no general relation between centered and uncentered information (or self–locating and non–self-locating propositions[16]) as simple as the Relevance-Limiting Thesis maintains.

4. Sleeping Beauty, Self-Location, and Updating

Shouldn't we be able to resolve the relevance of self-locating beliefs using extant Bayesian principles? As I've already mentioned, most Subjective Bayesians endorse a precise norm – updating by Conditionalization – for changing credences over time. Shouldn't we be able to take Beauty's Sunday-night credence distribution (on which everyone agrees), conditionalize it on her new Monday-morning evidence 'Today is Monday or Tuesday', and resolve any questions about her Monday-morning credence in heads?

The trouble with this proposal is that Conditionalization systematically fails for cases involving self-locating degrees of belief. So does the Reflection Principle, another popular Bayesian norm involving degrees of belief at distinct times.[17] Combining Kolmogorov's probability rules with some simple math, one can show that Conditionalization requires an agent certain of any proposition to remain certain of that proposition into perpetuity. But while Beauty is certain on Sunday night that it's Sunday night, she shouldn't be certain of that on Monday morning. Similarly, Reflection requires an agent who is certain that she'll be certain of some proposition in the future to be certain of that proposition now. On Sunday night, Beauty is certain that she will on Monday morning be certain that it's either Monday or Tuesday, but that doesn't mean Beauty should be certain that it's Monday or Tuesday on Sunday night.

These sorts of problems have motivated Subjective Bayesians to invent new updating norms that yield the same results as Conditionalization when only uncentered information is involved, but avoid absurdities in cases involving centered propositions.[18] The invention of such norms has become a cottage industry; it would take us too far afield to investigate them here.[19] For our purposes, the relevant result is that there are updating schemes that say Beauty should assign 1/2 to heads on Monday morning and updating schemes that say she should assign 1/3. The literature on updating self-locating credences has not settled on a univocal answer to the Sleeping Beauty Problem.[20]

That literature has, however, highlighted an important feature of the problem. Kierland and Monton (2005, p. 391) note that while most authors describe the Sleeping Beauty Problem as a case of experience duplication, Beauty's awakenings need not be subjectively indistinguishable. For example, she might wear red pajamas on Monday and blue pajamas on Tuesday, or *vice versa*. As long as the color/day correlation is determined by a chance process unobserved by Beauty and probabilistically independent of the fateful coin flip, seeing that she is wearing a particular color when she awakens does not tip Beauty off as to which day it is and so does not change her rationally required credence in heads.

If Beauty's awakenings are subjectively distinguishable, then when she awakens, her new evidence is not strictly centered/self-locating. For example, she might learn 'Beauty gets to awaken on the red-pajama day,' something she wasn't certain of on Sunday night. This is

important because updating schemes that treat centered and uncentered evidence very differently (especially those that endorse the Relevance-Limiting Thesis) may recommend a different credence in heads for Beauty in colored pajamas than they did when her awakenings were indistinguishable. The early halfer schemes of Halpern (2005) and Meacham (2008), for example, recommended that Beauty be 1/2 confident in heads on Monday morning if her awakenings are indistinguishable but only 1/3 confident in heads if pajamas are involved.[21] This result is now generally agreed to be highly implausible since the pajama-selection apparatus was carefully designed to be independent of the coin flip.

5. Sleeping Beauty and Dutch Books

Subjective Bayesians don't endorse Kolmogorov's axioms and updating by Conditionalization for no reason – they have *arguments* that these principles represent genuine norms of rationality. Perhaps some of those arguments could select among the formal schemes for updating self-locating credences mentioned in the previous section, or among rival answers to the Sleeping Beauty Problem?

Among the most popular arguments for Bayesian norms are Dutch Book and Dutch Strategy arguments. A Dutch Book is a set of bets a bookie could place with an agent at a given time such that the agent's credences judge each individual bet as fair, yet the entire package of bets guarantees the agent a loss. That a Dutch Book can be made against an agent is taken to show that the agent's credences are irrational; various theorems show that an agent whose credences violate Kolmogorov's axioms is subject to Dutch Book.[22]

A Dutch Strategy is a set of bets placed against an agent over time, such that if the agent adopts a particular policy for updating her credences, she can predict that she will view each of the bets as fair when placed, yet over the entire course of bets, she is guaranteed to lose funds. For example, David Lewis constructed a Dutch Strategy that will guarantee a sure loss to any agent who plans to update her credences in a manner inconsistent with Conditionalization.[23]

A Dutch Strategy can be assembled against thirders in the Sleeping Beauty Problem. Hitchcock (2004) proposes that on Sunday night, the bookie sells Beauty a bet that costs $15 and pays $30 if the coin comes up heads. Since Beauty is 1/2 confident in heads on Sunday night, she will accept this bet as fair. The bookie also tells Beauty on Sunday night that when she awakens Monday morning, he will sell her a bet for $20 that pays $30 if the coin comes up tails. If Beauty plans on being a thirder, she is certain that she will accept this bet as fair on Monday morning. (Notice that the bookie places this bet *only once* – on Monday morning – however the coin flip comes out.) Yet now Beauty is guaranteed to shell out a total of $35 for two bets which together will pay her $30, no matter the flip outcome. Planning to be a thirder exposes Beauty to a sure loss of $5.

The trouble is that a Dutch Strategy can also be assembled against halfers. Hitchcock suggests that the bookie start with a Sunday-night bet that costs $15 and pays $30 on tails. Then, *every* time Beauty awakens during the experiment (once if heads, twice if tails), the bookie sells her a bet that costs $10 but pays $20 on heads. If Beauty plans on Sunday night to be a halfer whenever she awakens, she knows she will accept these bets as fair. Totaling up is a bit more complicated because the number of bets depends on the outcome of the coin flip, but Beauty is once more guaranteed to lose $5 no matter how the flip comes out.

Which of these Dutch Strategies is the real deal? We want to know what rules a Dutch Strategy must follow to truly be revelatory of an underlying rational inconsistency. On the one hand, it seems suspicious that in order for a bookie to implement the Strategy against thirders, he would have to place a bet on Monday that he didn't place on Tuesday. This would require the bookie to know something when offering the bet (which day it is) that

Beauty doesn't know when she accepts it. On the other hand, it seems suspicious that the anti–halfing Strategy correlates the number of bets placed with the truth of the proposition being wagered upon.

The jury is still out on which of these Dutch Strategies – if either – reveals a genuine rational inconsistency and so can be used to decide the Sleeping Beauty Problem.[24] Our general theory of Dutch Strategies has not developed to the point of sorting these matters out. Yet I should note one interesting suggestion: It's possible that which Dutch Strategy counts as probative depends on one's views elsewhere in decision theory. Briggs (2010) argues that causal decision theorists should embrace the Dutch Strategy against halfing (and become thirders), while evidential decision theorists should embrace the strategy against thirders (and so be halfers).[25] If that's true, then one's opinions on causal vs. evidential decision theory may affect one's opinions on the Sleeping Beauty Problem – or *vice versa*!

6. Sleeping Beauty and Accuracy

Another type of argument for Bayesian principles starts with the idea that one's credences should be as *accurate* as possible. We can define an accuracy measure for individual credence values (such that accuracy improves when you have high credence in truths and low credence in falsehoods) and then calculate the accuracy of entire credence *distributions* (either by averaging the accuracy of each credence or by totaling over all the credences assigned). Joyce (1998) showed that if we choose our accuracy measure from a historically significant class (the 'proper scoring rules'), an agent whose credence distribution violates the Kolmogorov axioms can be confronted by a distribution that is more accurate than hers *in every possible world*. Greaves and Wallace (2006) later showed that if an agent is selecting an updating policy for responding to future courses of evidence, she can minimize her expected inaccuracy by selecting Conditionalization.[26]

One might think that aiming to minimize *average* inaccuracy and aiming to minimize *total* inaccuracy would always favor the same approach. But Kierland and Monton (2005) show that while thirding minimizes Beauty's expected total inaccuracy over the course of the Sleeping Beauty Problem, halfing minimizes her expected average inaccuracy. The trick is that the flip outcome affects how many times Beauty awakens, and so correlates with the total number of credence assignments she makes over the course of the experiment. Since Beauty awakens twice as many times if the coin comes up tails, being close to the truth when tails comes up will be twice as important to her *total* inaccuracy as being accurate when the coin comes up heads. Totaling encourages Beauty to be more confident in tails.[27]

Kierland and Monton think that there's little to recommend average over total inaccuracy in this case – or *vice versa*. Deciding between them parallels the Dutch Book question of whether we should allow the bookie to bet with Beauty once or twice when the coin comes up tails. So perhaps it is unsurprising that, after revising Kierland and Monton's technique for calculating expected inaccuracies, Briggs (2010) reaches the same conclusion that she reached with Dutch Books: Causal decision theorists should be thirders, while evidential decision theorists should be (double-) halfers. In any case, inaccuracy-minimization considerations alone are inconclusive for the Sleeping Beauty Problem.[28]

7. Sleeping Beauty and Memory Loss

As I noted in Section 4, the Sleeping Beauty Problem is difficult to solve with standard Bayesian updating norms like Conditionalization because Beauty's crucial piece of new evidence on Monday morning is a centered proposition. Yet the Sleeping Beauty Problem also features another notorious difficulty for Conditionalization: memory loss. Clearly, if

an agent can forget earlier certainties (or have those certainties forcibly erased from her memory) without violating the requirements of rationality, there are going to be cases in which her credences update rationally but fail to conform to the certainty-retaining Conditionalization rule.[29] Even worse, Arntzenius (2003) argues that just the *threat* that an agent's memory has been altered can make it rational for her to update in ways that violate Conditionalization. Sleeping Beauty certainly faces that threat when she awakens on Monday morning. As a result, Arntzenius (2002, p. 61) went so far as to declare, 'self-locating learning plays no relevant role in the Sleeping Beauty case. The real issue is how one deals with known, unavoidable, cognitive malfunction.'

This claim seems a bit overblown[30] because we can construct Beauty-like scenarios that involve no memory loss or the threat thereof. For example, instead of awakening Beauty twice if the coin lands tails, the experimenters could awaken her only once but create a duplicate of her that they awaken at the same time. Instead of being uncertain what day it is when she awakens, Beauty would be uncertain whether she was Beauty or the doppelgänger.[31] Given the structural similarities between this Doppelgänger Beauty Problem and the original Sleeping Beauty Problem, many authors analyze both problems the same way (and recommend the same Monday-morning credence in heads in each case). Yet Beauty never forgets anything (or suspects she may have) in the doppelgänger case.

Yet not everyone takes Doppelgänger Beauty to be perfectly analogous to the original problem. While they refuse to choose between expected average and expected total inaccuracy for Sleeping Beauty, Kierland and Monton think that reasoning about the doppelgänger case should definitely be governed by averaging. Expected average inaccuracy helps Beauty estimate how far off the mark a particular updating policy will leave *her*; expected total inaccuracy adds into that calculation inaccuracies of individuals (namely the doppelgänger) who aren't Beauty. Why should Beauty care about the inaccuracy of other people's opinions? This is the accuracy-theoretic version of a more general point Briggs (2010) and Meacham (2010) make about updating and doppelgängers. Philosophers think we have some grasp on why my current opinions should line up with opinions held by my former selves. But I owe no such allegiance to opinions held by *others* in the past. To the extent, then, that Beauty when she awakens on Monday morning entertains the possibility that she is someone *other* than Beauty, why should her opinions about the coin flip line up in any tidy way with Beauty's opinions on Sunday night? Updating-based arguments may not transfer so tidily between Sleeping Beauty and the doppelgänger case.

8. Sleeping Beauty and Indifference Principles

Duplication brings us to a feature of the Sleeping Beauty Problem I haven't discussed yet. The main controversy in Sleeping Beauty is how confident Beauty should be on Monday morning that the coin comes up heads. But once Beauty has decided how much credence to allocate various possible coin flip outcomes, she has a further decision to make: how much credence to allocate various possibilities about which day it is. If the flip came up heads, it's Monday (because she's awake), but if the flip came up tails, it could be Monday or Tuesday. So conditional on the supposition of tails, how confident should Beauty be in Monday versus Tuesday?

Despite their differences on the main question in Sleeping Beauty, Elga (2000) and Lewis (2001) agreed on the application of what Elga called 'a highly restricted principle of indifference' to set this conditional value at 1/2. In other words, Elga and Lewis both thought that however much credence Beauty assigns on Monday morning to tails, she should divide that credence equally between the possibility that the coin comes up tails and it's Monday and the possibility that the coin comes up tails and it's Tuesday. Later, in his

(2004), Elga generalized this indifference principle to maintain that whenever an agent considers an uncentered possible world with multiple centers subjectively indistinguishable from her current experience, she should divide her credence in that uncentered world equally among those centers.

The Bayesian literature has generally given indifference principles a (well-deserved) bad name.[32] But most indifference principles direct agents to divide their credences equally among uncentered worlds. Elga's indifference principle escapes many traditional objections because it divides credence only among centers within the *same* uncentered world. Still, Elga's principle is subject to a number of objections. Weatherson (2005) worries about Elga's internalist conception of experience, the intransitivity of indistinguishability, Elga's failure to distinguish 'uncertain' from 'risky' propositions, and the applicability of Elga's principle to cases involving infinitely many possibilities.[33] But perhaps the most troubling consequence of Elga's indifference principle is raised by Elga himself. Suppose you receive a letter making you certain that yesterday, someone made one thousand duplicates of the individual you (up until you read the letter) thought of as yourself – each of whom then had identical subjective experiences. According to Elga's indifference principle, after reading the letter, you should be highly confident (1000/1001) that you are one of the copies as opposed to the original. This seems somewhat disturbing.[34]

Elga presents his indifference principle as a new constraint on rational credences, independent of the Kolmogorov axioms, Conditionalization, and the like.[35] But particular applications of that principle (such as its application to Sleeping Beauty) can be derived from well-established Bayesian ways of thinking. For example, if we adopt the expected relative frequencies approach of Section 1, we can argue that in the long run, Beauty should expect 1/2 of her tails awakenings to be Tuesday awakenings. So, on any given awakening, she should be 1/2 confident that it's Tuesday conditional on tails. This is exactly the result one gets from Elga's indifference principle. It is an open question which positions on the Sleeping Beauty Problem – and which types of arguments for those positions – are committed to Elga's principle in its full generality.

9. Sleeping Beauty and the Multiverse

Elga's indifference principle distributes agents' credences among subjectively indistinguishable centers within a given uncentered world. On its own, Elga's principle has no implications for agents' credence distributions across uncentered worlds. But Elga incorporates that principle into an argument for thirding, and some authors see thirding as an invitation to serious skepticism about one's identity.

The original Sleeping Beauty Problem and Doppelgänger Beauty share the following structure: On Sunday night, Beauty entertains two possible uncentered worlds, a world in which the coin comes up heads and a world in which it comes up tails. The tails world contains two centers with a particular experience – the experience of awakening in the experimenters' room.[36] The heads world contains only one center with such an experience. On Sunday night, Beauty is equally confident in each of these uncentered worlds, but the thirder thinks that when Beauty awakens and finds herself at one of the centers in question, she should shift confidence towards the tails world. In other words, finding herself at one of the centers should shift Beauty's credence towards the uncentered world in which more centers like that exist.

This conclusion looks a lot like the Self-Indication Assumption (Bostrom 2002a): When an agent considers her own existence as evidence, she should take it to favor hypotheses implying the existence of many agents over hypotheses implying the existence of fewer. The Self-Indication Assumption was originally proposed to deflect the Doomsday Argument, which suggests that we should dramatically increase our confidence that we are

among the last humans who will ever live. Self-Indication also intersects with a hot issue in physics: whether there exist universes other than our own. If the Self-Indication Assumption is correct, our very existence is evidence for hypotheses positing many (populated) universes over hypotheses that posit just one.[37]

But Self-Indication has problems of its own. For example, the Self-Indication argument for the multiverse in the previous paragraph seems too easy – supporting a hypothesis as ontologically extravagant as the existence of other universes should require more empirical investigation. Similar problems arise if thirding means that we should generally be more confident of hypotheses implying the existence of many agents with experiences like our own than hypotheses implying fewer. Now you need not actually to receive the letter described in Section 8 to start worrying about duplicates – you need only to imagine it. As long as you merely *entertain* the possibility that someone out there is making copies of you, you will assign that possibility a nonzero credence, and our current reasoning will require that credence to be high.[38] If we add in Elga's indifference principle, however confident you are that the copies exist most of that confidence will be allotted to your being one of the copies. As a number of authors have noted, replacing the doppelgängers in this scenario with brains in vats mimicking your subjective experience offers a quick road to skepticism.[39]

10. Sleeping Beauty and Quantum Mechanics

All of this might seem far-fetched, were philosophers of physics not currently engaged in a debate between two empirically robust theories that differ in how many agents they take the world to contain. On the classical (Copenhagen) interpretation of quantum mechanics, whenever a quantum experiment is run a genuinely indeterministic event occurs, determining how that experiment will be measured to have turned out. For example, when a physicist runs a Stern-Gerlach experiment, it is genuinely chancy whether the electron in the experiment will be measured as 'spin-up' or 'spin-down'.[40] Given aspects of the experimental setup, the 'Born rule' from quantum mechanics provides specific numbers for those chances – the chance of spin-up might be, say, 70% – and by the Principal Principle, the physicist should have corresponding credences that particular outcomes will be measured. The rival many-worlds interpretation (Everett 1957) reads the quantum situation very differently. Roughly, this interpretation holds that when the physicist runs her quantum experiment, the universe splits in two, with each resulting universe containing a physicist who observes a different outcome.[41]

A number of problems arise when we try to understand credences under the many-worlds interpretation. Most importantly, the 'incoherence problem'[42] for many-worlds asks how we should make sense of non-extreme credences about experimental outcomes before the experiments are performed. When our physicist has set up her Stern-Gerlach experiment and is about to run it, her best theory of quantum mechanics tells her that there is a 70% chance of measuring spin-up. But if the many-worlds interpretation is correct (and she is certain of that), the physicist should be 100% confident that each possible outcome will be measured (by *someone*). So what does the 70% mean?

The incoherence problem is a general problem for the many-worlds interpretation. But supposing we can solve it, there remain questions about how to generate the physicist's non-extreme credences at various stages using extant Bayesian theory. For instance, take the time immediately after the experiment has been run but before anyone has observed its outcome. At that time, there are two agents in two universes. Each of those agents is about to measure a different outcome, but each should be 70% confident that she herself is the one about to measure spin-up.

Now consider the physicist's credence before the experiment is performed that she herself will measure spin-up. Depending on your theory of personal identity, the physicist should either be certain that she is identical with the individual who will observe spin-up or be certain that she is not identical with that individual. Either way, if the agents who exist after the experiment generate their credences by conditionalizing the physicist's earlier distribution, each of them will either remain certain that she is the spin-up agent or remain certain that she is not. Neither option allows these agents to land where they should be, at 70%.

Matters are not improved by applying Elga's indifference principle from Section 8. After the experiment is performed but before its result is observed, the agents in the two universes have subjectively identical experiences.[43] So Elga's principle recommends that each of them be 50% confident that she will observe spin-up. Unfortunately, there are many real-world Stern-Gerlach setups for which 50% is not the chance allocated to spin-up by our best quantum theories. Once again, we are unable to obtain the credence number (70%, or whatever) the actual physics recommends.

Finally, the many-worlds interpretation offers a direct challenge to thirders about Sleeping Beauty. Imagine the experimenters tell Beauty that they're only going to wake her up once (on Monday), that no coin will be flipped after she goes to sleep on Sunday night, but that they will perform a quantum experiment and reveal the outcome to her on Monday morning. Suppose further that Beauty is unsure whether the classical or many-worlds interpretation is the right way to understand quantum mechanics. Notice that if the classical interpretation is correct, there will be only one of her on Monday morning, while many-worlds posits one Monday-morning Beauty for each possible quantum outcome. Thus, if she is a thirder and if thirding requires being more confident of hypotheses featuring many individuals in one's subjective state, then when Beauty awakens Monday morning, she will become more confident that the many-worlds interpretation is true. And since there are quantum events happening around us all the time (and splitting us into multiples if Everett is correct), it seems that if we are thirders, we should constantly be increasing our confidence in many-worlds over the Copenhagen interpretation.[44]

11. Conclusion

The Sleeping Beauty Problem has connections to a variety of unsettled issues. The problem can be attacked from many directions, including some I haven't even mentioned here.[45] If any of those attacks should land, the solution to Sleeping Beauty would reverberate through those connections to disparate areas of philosophy. One simple question – how confident should Beauty be in heads when she awakens? – could provide answers for the theory of content, the measurement of accuracy, and the interpretation of quantum mechanics.[46]

Short Biography

Michael G. Titelbaum received his PhD from the University of California, Berkeley, in 2008 and is now an assistant professor of philosophy at the University of Wisconsin-Madison. He has been thinking about the Sleeping Beauty Problem for what feels like a very long time.

Notes

* Correspondence: Department of Philosophy, University of Wisconsin-Madison, 600 North Park Street, Madison, WI 53706, USA. Email: titelbaum@gmail.com.

[1] Elga notes that 'this problem appears as Example 5' of Piccione and Rubinstein (1997) (an article from the decision theory literature on Piccione and Rubinstein's 'Absent-Minded Driver Paradox'). Elga also reports that the 'Sleeping

Beauty' moniker was introduced by Robert Stalnaker, who 'first learned of examples of this kind in unpublished work by Arnold Zuboff' (Elga 2000, note 1).

[2] Grimms' fairy tales refer to the protagonist of the Sleeping Beauty story as either 'Briar-Rose' or 'the princess' (Grimm and Grimm 1972, pp. 237–241). In the Disney film, she is 'Aurora' and also 'Briar-Rose' (Teitelbaum 2004). To my knowledge, philosophers were the first to call her 'Beauty'.

[3] Kolmogorov's axioms require all probabilities to be non-negative, the probability of any tautology to be 1, and the probability of a disjunction to be the sum of the probabilities of its disjuncts (assuming the disjuncts are mutually exclusive). For more on Subjective Bayesianism, see Easwaran (2011a, 2011b).

[4] The problem traces back to Venn (1876); the name dates to Reichenbach (1949). (Thanks to Hájek (2007) for these pointers.)

[5] One might suggest that the awakenings reference class is somehow *narrower* (or more *specific*) than the experimental runs class and whenever an agent has precise long run expected frequencies for two reference classes, she should set her credences by the narrower one. This is essentially the position of Oscar Seminar (2008) in arguing for 1/3. But see the response from Pust (2011) and then the response to Pust from Thorn (2011).

[6] Conditionalization requires an agent's later credence in a proposition to be her earlier credence in that proposition conditional on everything she learns between the two times. For more discussion and a formal presentation, see the introductory section of Easwaran (2011a).

[7] Schwarz (2012) and Hawley (2013) both entertain the position that before Beauty even talks to the experimenters on Monday, she should already be certain that it's Monday. But this position is hard to swallow, as it requires Beauty to assign *no* credence to a possibility (that it's Tuesday and the coin came up tails) which doesn't seem to have been ruled out by her evidence at that time.

[8] For attempts to make sense of Lewis's proposed inadmissible evidence, see Bradley (2011b) and Titelbaum (2013, Ch. 9).

[9] There are actually two ways to avoid the consequence that Beauty's Monday-evening credence in heads must be greater than her Monday-morning credence: You can either deny that Conditionalization applies in this case or grant that Beauty should update by conditionalizing but claim that she should conditionalize on more information than just that it's Monday. The former approach is taken by Meacham (2008, 2010), Briggs (2010), Cozic (2011), and Pust (2012), while Bostrom (2007) takes the latter.

[10] The thirder value for Monday evening is calculated by taking the thirder's recommended Monday-morning credence distribution and conditionalizing on the information that it's Monday. Note that from this point forward, we will continue to work with a version of the Sleeping Beauty Problem that includes Elga's modifications.

[11] For the relevant calculations, see Titelbaum (2012). (To my knowledge, that is the first article to propose this Tuesday-flip scenario.) Again, I'm assuming that upon awakening Monday morning, Beauty has some positive credence that it's Tuesday (as explained in note 7 above).

[12] The Relevance-Limiting Thesis was first articulated in Titelbaum (2008). See Bradley (2011b), Briggs (2010), Titelbaum (2013, Chapter 10), and Draper (ms) for discussion.

[13] As Bostrom (2007) and Briggs (2010) point out, only the double-halfer can consistently maintain the Relevance-Limiting Thesis. The thirder violates this thesis when he has Beauty change her credence in heads between Sunday night and Monday morning. The Lewisian halfer violates it when he has Beauty change her heads credence between Monday morning and Monday evening. (Between those two times, Beauty has gained only the centered information that today is Monday.)

[14] Your new evidence does *not* include 'At least one person picks a black ball.' You were already certain of that once the experimental protocol was explained.

[15] This example appears in Titelbaum (ms; 2013, Chapter 10). It is similar to a counterexample to the Relevance-Limiting Thesis that appears in Bradley (2011b, Section 9), which Bradley in turn attributes to Matt Kotzen.

[16] Although Lewis (1979) is happy to use centered/uncentered terminology that he takes from Quine (1969), Lewis thinks of both centered and uncentered information as 'self-locating': Uncentered information locates you within the space of possible worlds, while centered information locates you within the particular world you occupy. Nowadays, it's more typical to describe only centered information as 'self-locating'; I will follow that practice.

[17] For the Reflection Principle, see van Fraassen (1995). Arntzenius (2003) raises the problems for both Conditionalization and Reflection that I'll be discussing here, and ties them to Sleeping Beauty. (See also Schervish et al. (2004).)

[18] For reasons independent of self-location concerns, some Subjective Bayesians prefer Jeffrey Conditionalization (Jeffrey 1983) to Conditionalization as an updating norm. Jeffrey Conditionalization doesn't require agents to retain certainties, so one might think that it could handle self-locating update without modification. Kim (2009), however, shows that Jeffrey Conditionalization has just as much trouble with self-locating credences as standard Conditionalization.

[19] Distinct formal systems for updating self-locating credences have been proposed by Halpern (2005), Meacham (2008, 2010), Titelbaum (2008, 2013), Stalnaker (2008), Kim (2009), Briggs (2010), Schulz (2010), Cozic (2011), Schwarz (2012), Moss (2012), and Santorio (ms). For an overview of all these formalisms that divides them into three general approaches, see Titelbaum (ms).

[20] In discussing the problems that self-location poses for Conditionalization and Reflection, I am using a Lewisian approach to propositions on which 'Today is Sunday' expresses the same (centered) proposition whenever it is uttered. Some authors reject this approach; Stalnaker (2008) has even argued that problems about self-locating *update* can be entirely solved by properly understanding self-locating *content*. If that's right, Sleeping Beauty is tied to much larger debates in the philosophy of mind and language about how to understand self-locating beliefs. (Stalnaker is a thirder.)

[21] See the discussion of 'Technicolor Beauty' in Titelbaum (2008) and Briggs (2010) and the discussion of the black-and-white-room case in Meacham (2008, 2010).

[22] See Easwaran (2011a, Section 1.1) for discussion and references.

[23] Teller (1973) first described this Dutch Strategy in print, but attributed it to Lewis.

[24] See Arntzenius (2002), Hitchcock (2004), Vineberg (ms), Bradley and Leitgeb (2006), Draper and Pust (2008), Lewis (2010), Briggs (2010), and Schwarz (ms). More broadly, Dutch Strategies for situations involving self-locating credences have not been investigated generally enough to select among the rival formalisms for self-locating update discussed in Section 4.

[25] See also Draper and Pust (2008), Shaw (2013), and Schwarz (ms). Roughly, the distinction between causal and evidential decision theorists is that the former use conditional probabilities tracking *causal* influences to calculate expected utilities, while the latter use conditional credences reflecting *any* evidential correlation between act and outcome. For more information, see Weirich (2012).

[26] For further developments, see Joyce (2009) and Leitgeb and Pettigrew (2010a, 2010b). We calculate *expected* inaccuracy by taking the inaccuracy of the credence distribution an agent's policy would yield in a given possible world, weighting it by her current credence that that world is actual, and then summing over worlds.

[27] Interestingly, Kierland and Monton (2005) note that both expected total inaccuracy and expected average inaccuracy are minimized if Beauty plans to assign 1/2 credence to heads on Monday evening. This constitutes an argument against Lewisian halfing but does not discriminate between thirding and double-halfing.

[28] See also Pettigrew (ms).

[29] See Talbott (1991).

[30] And, in fairness, Arntzenius seems to have backed off from it by his (2003).

[31] Duplication stories in this vicinity have been discussed by Arntzenius (2003), Elga (2004), Kierland and Monton (2005), Bostrom (2007), and Meacham (2008), among others. To my knowledge, the first to propose this specific doppelgänger variation of Sleeping Beauty were Kierland and Monton.

[32] See, for example, van Fraassen (1989).

[33] For a general discussion of Sleeping Beauty's interaction with principles of the infinite – and especially Countable Additivity – see Ross (2010).

[34] Quantum physicists note that if the universe is infinite in size, the chance is 1 that somewhere in this vastness, quantum fluctuations have caused a clump of particles to coalesce into a brain that is having the same subjective experiences as you *right now*. In fact, the chance is 1 that there are many such 'Boltzmann brains' sharing your current experiences. So if we ever discover that the universe is infinite, Elga's indifference principle will force each of us to become highly confident that he or she is in fact a randomly assembled brain floating out in space, doomed to survive for only a moment longer. (Thanks to Kenny Easwaran for introducing me to this objection against Elga.)

[35] Of course, Elga provides arguments for the new constraint, involving what seem to be reasonable verdicts about a string of cases intimately related to the duplication case.

[36] In referring to each of these as a single 'center', I'm ignoring the experience's temporal duration. For an investigation of that aspect, see Fisher (ms).

[37] See Bostrom (2002a) and a nice summary in Bradley (2012), which ties Doomsday and arguments for the multiverse to the Sleeping Beauty Problem and many of the other topics discussed here. For more on treating one's own existence as evidence in a Bayesian context, see Pust (2007), Kotzen (2012), and Manley (ms). (Thanks to an anonymous referee for pointing me to these sources.)

[38] Matters are made worse by the fact that Self-Indication and thirding seem to assign a larger share of one's credence to a hypothesis involving duplicates the more duplicates that hypothesis involves.

[39] Bostrom (2002b) considers the rival hypotheses that the universe is either infinite or finite. As I pointed out in note 34 above, an infinite universe is almost sure to contain many Boltzmann brains with experiences just like yours. Self-Indication/thirding looks like it will make you highly confident of the infinite-universe hypothesis containing such brains, and Elga's indifference principle will then make you confident that you're one of them!

[40] This example is adapted from Greaves (2004).

[41] The many-worlds interpretation of quantum mechanics is different from the multiverse hypothesis discussed in the previous section. The multiverse hypothesis typically holds that there are many universes, causally closed off from each other, and featuring different physical constants (or even different laws of physics). In many-worlds, the separate 'universes' are different terms in the equation of a single (superposed) quantum state; they share the same physical constants and physical laws and can even re-merge under the right conditions.

[42] See Lewis (2007) for discussion and references.

[43] While these two physicists occupy different 'universes' from the quantum-mechanical point of view, they still belong to the same uncentered possible world in the philosopher's sense – the sense needed to apply Elga's indifference principle.

[44] The quantitative concerns described in this section are discussed in Lewis (2007), Price (ms), Bradley (2011a, 2012), and Schwarz (2012). Titelbaum (2013, Chapter 11) tries to resolve them by first showing that careful application of devices like Kierland and Monton's colored pajamas (from Section 4) can generate many of the consequences of Elga's indifference principle without adopting that principle as an independent constraint. When these devices are combined with a thirder-friendly updating policy and applied to versions of the many-worlds interpretation that avoid the incoherence problem, agents wind up with credences that match the Born rule's recommendations and do not confirm many-worlds over the classical interpretation. (See also Wilson (ms).)

[45] For instance, the epistemic probabilities approach of Horgan (2004), Bovens' (2010) link between Sleeping Beauty and Judy Benjamin, and the symmetry arguments of Dorr (2002).

[46] Thanks to Hayley Clatterbuck, Jesse Steinberg, and Reuben Stern for feedback on earlier drafts; to Darren Bradley for discussion; and to a very helpful anonymous referee for *Philosophy Compass*.

Work Cited

Arntzenius, F. Reflections on Sleeping Beauty. *Analysis* 62 (2002): 53–62.

——. Some problems for conditionalization and reflection. *The Journal of Philosophy* 100 (2003): 356–370.

Bostrom, N. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge, 2002a.

——. Self-locating belief in big worlds: Cosmology's missing link to observation. *Journal of Philosophy* 99 (2002b): 607–623.

——. Sleeping Beauty and self-location: A hybrid model. *Synthese* 157 (2007): 59–78.

Bovens, L. Judy Benjamin is a Sleeping Beauty. *Analysis* (2010): 23–26.

Bradley, D. Confirmation in a branching world: The Everett interpretation and Sleeping Beauty. *British Journal for the Philosophy of Science* 62 (2011a): 323–342.

——. Self-location is no problem for conditionalization. *Synthese* 182 (2011b): 393–411.

——. Four problems about self-locating belief. *Philosophical Review* 121 (2012): 149–177.

——. and H. Leitgeb. When betting odds and credences come apart: More worries for Dutch Book arguments. *Analysis* 66 (2006): 119–127.

Briggs, R. Putting a Value on Beauty. *Oxford Studies in Epistemology*, Volume 3. Eds. T. S. Gendler and J. Hawthorne. Oxford: Oxford University Press, 2010. 3–34.

Cozic, M. Imaging and Sleeping Beauty: A case for double-halfers. *International Journal of Approximate Reasoning* 52 (2011): 137–143.

Dorr, C. Sleeping Beauty: In defence of Elga. *Analysis* 62 (2002): 292–296.

Draper, K. The evidential relevance of self-locating information. *Philosophical Studies* (ms); Forthcoming.

——. and J. Pust. Diachronic Dutch Books and Sleeping Beauty. *Synthese* 164 (2008): 281–287.

Easwaran, K. Bayesianism I: Introduction and arguments in favor. *Philosophy Compass* 6 (2011a): 312–320.

——. Bayesianism II: Criticisms and applications. *Philosophy Compass* 6 (2011b): 321–332.

Elga, A. Self-locating belief and the Sleeping Beauty Problem. *Analysis* 60 (2000): 143–7.

——. Defeating Dr. Evil with self-locating belief. *Philosophy and Phenomenological Research* 69 (2004): 383–396.

Everett, H. 'Relative state' formulation of quantum mechanics. *Reviews of Modern Physics* 29 (1957): 454–462.

Fisher, J. The Java problem: Another puzzle about self-locating belief. (ms). Unpublished manuscript.

Greaves, H. Understanding Deutsch's probability in a deterministic universe. *Studies in History and Philosophy of Modern Physics* 35 (2004): 423–456.

——. and D. Wallace. Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind* 115 (2006): 607–632.

Grimm, J. L. K. and W. K. Grimm. *The Complete Grimm's Fairy Tales*. New York: Pantheon Books. Translated by Margaret Hunt and James Stern, 1972.

Hájek, A. The reference class problem is your problem too. *Synthese* 156 (2007): 563–585.

Halpern, J. Y. Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. *Oxford Studies in Epistemology*, Volume 1. Eds. T. Gendler and J. Hawthorne. Oxford: Oxford University Press, 2005. 111–142.

Hawley, P. Inertia, optimism and Beauty. *Noûs* (2013): 85–103.

Hitchcock, C. R. Beauty and the bets. *Synthese* 139 (2004): 405–420.

Horgan, T. Sleeping Beauty awakened: New odds at the dawn of the new day. *Analysis* 64 (2004): 10–21.

Jeffrey, R. C. *The Logic of Decision* (2nd edn). Chicago: University of Chicago Press, 1983.

Joyce, J. M. A nonpragmatic vindication of probabilism. *Philosophy of Science* 65 (1998): 597–603.

——. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. *Degrees of Belief*. Eds. F. Huber and C. Shmidt-Petri. Springer, 2009. 263–300.

Kierland, B. and B. Monton. Minimizing inaccuracy for self-locating beliefs. *Philosophy and Phenomenological Research* 70 (2005): 384–395.

Kim, N. Sleeping Beauty and shifted Jeffrey conditionalization. *Synthese* 168 (2009): 295–312.

Kolmogorov, A. N. *Foundations of the Theory of Probability*. Ed. Nathan Morrison. New York: Chelsea Publishing Company, 1950.

Kotzen, M. Selection biases in likelihood arguments. *British Journal for the Philosophy of Science* (2012): 825–839.

Leitgeb, H. and R. Pettigrew. An objective justification of Bayesianism I: Measuring inaccuracy. *Philosophy of Science* 77 (2010a): 201–235.

——. An objective justification of Bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science* 77 (2010b): 236–272.

Lewis, D. Attitudes *de dicto* and *de se*. *The Philosophical Review* 88 (1979): 513–543.

——. A Subjectivist's Guide to Objective chance. *Studies in Inductive Logic and Probability*, Volume 2. R. C. Jeffrey. Berkeley: University of California Press, 1980. 263–294.

——. Sleeping Beauty: Reply to Elga. *Analysis* 61 (2001): 171–6.

Lewis, P. J. Uncertainty and probability for branching selves. *Studies in History and Philosophy of Modern Physics* 38 (2007): 1–14.

——. Credence and self-location. *Synthese* 175 (2010): 369–382.

Manley, D. Self-location, existence, and evidential uniqueness. (ms). Unpublished manuscript.

Meacham, C. J. G. Sleeping Beauty and the dynamics of de se beliefs. *Philosophical Studies* 138 (2008): 245–270.

——. Unravelling the tangled web: Continuity, internalism, non-uniqueness and self-locating beliefs. *Oxford Studies in Epistemology* 3 (2010): 86–125.

Moss, S. Updating as communication. *Philosophy and Phenomenological Research* (2012): 225–48.

Oscar Seminar, T. An objectivist argument for thirdism. *Analysis* (2008): 149–155.

Pettigrew, R. Self-locating beliefs and the goal of accuracy. (ms). Unpublished manuscript.

Piccione, M. and A. Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* 3 (1997): 3–24.

Price, H. Probability in the Everett world: Comments on Wallace and Greaves. (ms). Unpublished manuscript.

Pust, J. Cartesian knowledge and confirmation. *Journal of Philosophy* (2007): 269–289.

——. Sleeping Beauty and direct inference. *Analysis* 71 (2011): 290–293.

——. Conditionalization and essentially indexical credence. *Journal of Philosophy* (2012): 295–315.

Quine, W. V. Propositional Objects. *Ontological Relativity & Other Essays*. New York: Columbia University Press, 1969.

Reichenbach, H. *The Theory of Probability*. Berkeley: University of California Press, 1949.

Ross, J. Sleeping Beauty, countable additivity, and rational dilemmas. *Philosophical Review* 119 (2010): 411–447.

Santorio, P. Cognitive relocation. (ms). Unpublished manuscript.

Schervish, M. J., T. Seidenfeld, and J. Kadane. Stopping to reflect. *Journal of Philosophy* 101 (2004): 315–322.

Schulz, M. The dynamics of indexical belief. *Erkenntnis* 72 (2010): 337–351.

Schwarz, W. Changing minds in a changing world. *Philosophical Studies* 159 (2012): 219–239.

——. Lost memories and useless coins: Revisiting the absentminded driver. (ms). Unpublished manuscript.

Shaw, J. R. *De se* belief and rational choice. *Synthese* (2013): 491–508.

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

Talbott, W. J. Two principles of Bayesian epistemology. *Philosophical Studies* 62 (1991): 135–150.

Teitelbaum, M. *Sleeping Beauty*. (Disney Princess ed.). Golden Books Publishing Company, 2004.

Teller, P. Conditionalization and observation. *Synthese* 26 (1973): 218–258.

Thorn, P. D. Undercutting defeat via reference properties of differing arity: A reply to Pust. *Analysis* 71 (2011): 662–667.

Titelbaum, M. G. The relevance of self-locating beliefs. *Philsophical Review* (2008): 555–605.

——. An embarrassment for double-halfers. *Thought* (2012): 146–51.

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

——. Self-Locating Credences. *The Oxford Handbook of Probability and Philosophy*. A. Hájek and C. R. Hitchcock. Oxford: Oxford University Press, (ms). Forthcoming.

van Fraassen, B. C. Indifference: The Symmetries of Probability. *Laws and Symmetry*. Oxford: Oxford University Press, 1989. 293–317.

——. Belief and the problem of Ulysses and the Sirens. *Philosophical Studies* 77 (1995): 7–37.

Venn, J. *The Logic of Chance* (2nd edn). MacMillan and Co., Limited, 1876.

Vineberg, S. Beauty's cautionary tale. (ms). Unpublished manuscript.

Weatherson, B. Should we respond to evil with indifference? *Philosophy and Phenomenological Research* 70 (2005): 613–35.

Weirich, P. Causal Decision Theory. *The Stanford Encyclopedia of Philosophy*. Ed. E. N. Zalta (Winter 2012 ed.), 2012.

Wilson, A. Everettian confirmation and Sleeping Beauty. *British Journal for the Philosophy of Science*. (ms). Forthcoming.