# Sleeping Beauty's Credences

Jessi Cisewski, Joseph B. Kadane, Mark J. Schervish, Teddy Seidenfeld, and Rafael Stern*

The Sleeping Beauty problem has spawned a debate between "thirders" and "halfers" who draw conflicting conclusions about Sleeping Beauty's credence that a coin lands heads. Our analysis is based on a probability model for what Sleeping Beauty knows at each time during the experiment. We show that conflicting conclusions result from different modeling assumptions that each group makes. Our analysis uses a standard "Bayesian" account of rational belief with conditioning. No special handling is used for self-locating beliefs or centered propositions. We also explore what fair prices Sleeping Beauty computes for gambles that she might be offered during the experiment.

**1. Introduction and Outline.** The Sleeping Beauty puzzle is an unusual decision problem with several exceptional features that, as a matter of history, have led to considerable controversy over its solution. Many of the solutions rely on attempts to reconcile credences expressed as probabilities with centered propositions and possible-world semantics (see, e.g., Halpern 2005; Meacham 2008; Titelbaum 2008; Cozic 2011). In contrast, we take an approach based on probability modeling, which includes conditioning for updating credences. In order to analyze the controversial aspects of this puzzle, and to identify its exceptional features, first we review the basic Sleeping Beauty problem and its principal, rival solutions.

*1.1. The Basic Sleeping Beauty Problem.* On Sunday, Sleeping Beauty learns that she will participate in the following experiment. Sunday night Sleeping Beauty will be put into a controlled sleep. A fair coin is to be flipped at some time before Tuesday morning, but its result is not revealed to Sleep-

Received January 2015; revised December 2015.

*To contact the authors, please write to: Teddy Seidenfeld, Carnegie Mellon University, Department of Philosophy, Baker Hall 161, 5000 Forbes Ave., Pittsburgh, PA 15213; e-mail: teddy@stat.cmu.edu.

Philosophy of Science, 83 (July 2016) pp. 324–347. 0031-8248/2016/8303-0002$10.00

ing Beauty until Wednesday. She will be awakened on Monday morning for a brief period when she will be asked the question "What is your *degree of belief* (or *credence*) for the event that the coin lands heads?"

Then, she will be returned to her induced state of sleep. At this point, by the design of the experiment, Sleeping Beauty loses all of her memories of Monday. If the coin flip lands tails, and only if it land tails, she will be awakened briefly for a second time during the experiment, on Tuesday morning, and again asked the same question, in the same way as on Monday. Then she returns to her state of sleep until she awakes normally, Wednesday, after the experiment is over.

*1.2. What Is Sleeping Beauty's Credence in Heads during the Experiment?* This problem was first introduced by Piccione and Rubinstein (1997, example 5), a variant of their Absentminded Driver paradox. Elga (2000) used the name Sleeping Beauty when discussing the same problem. A rather large literature has grown up around it. Already, in spring 1999, the newsgroup rec.puzzles reported several thousand threads discussing the Absentminded Driver paradox (Wedd 2006).

The problem is a puzzle as evidenced by the large literature and continuing controversy over how Sleeping Beauty should answer the two questions posed when awakened during the experiment. Next, we summarize the original arguments and conclusions from two rival factions, *thirders* and *halfers*, that dominate the controversy. Alternative arguments for both sides have developed over time, but it is not our goal to refute all arguments. Rather we offer a unified probability model that shows how both rival conclusions derive from different model assumptions. The summaries that we give should help the reader to better understand how our approach differs from the reasoning that others have used.

The following text, which we call "the halfers' argument" summarizes the original halfers' reasoning about Sleeping Beauty's rational degrees of belief.

> On Sunday, Sleeping Beauty's credence is 1/2 that the coin lands heads, since the coin is stipulated to be fair. Let P(·) denote her rational credence from Sunday's perspective. Then P(heads) = 1/2.
>
> With E the event that during the experiment she is awake and aware of that fact, according to the rules during the experiment, P(E) = 1.
>
> So, P(heads|E) = 1/2, since conditioning on a sure-event leaves probabilities unchanged. But when Sleeping Beauty is awake during the experiment, that fact (i.e., the proposition E that the experiment is running and she is awake) is all that she learns has happened since going to sleep on Sunday. That is, event E represents the totality of her new evidence between retiring Sunday and being awakened during the experiment. We

> model Sleeping Beauty as a canonical Bayesian, one who uses Bayes's rule to update her degrees of belief when augmenting what she knows with new evidence. Define $P_E(\cdot) = P(\cdot|E)$. Then $P_E(\text{heads}) = 1/2$.[1]

The halfers' argument (as presented above) relies on "that she is awake during the experiment" being the totality of Sleeping Beauty's knowledge (beyond what she knew on Sunday) when she tries to assess her credence in how the coin will land. If she awakes with a crick in her neck or a slight case of indigestion, is it necessary or even plausible that such experiences make no difference in her assessment of credences?[2] It might be difficult to argue how such experiences are related to the flip of a coin, but might they shed some light on how many times she is awakened during the experiment? To answer these questions and others, section 2 presents a general probability model of what Sleeping Beauty experiences during the experiment.

The original "thirders' argument" (see Elga 2000 for an example) uses centered possibilities to argue that Sleeping Beauty's credence in heads while awake during the experiment should be 1/3. We summarize that argument next.

> During the experiment, while awake, Sleeping Beauty recognizes these three centered possibilities as exhaustive:
>
> A. It is now Monday and the fair coin will land heads.
> B. It is now Monday and the fair coin will land tails.
> C. It is now Tuesday and the fair coin landed tails.
>
> Let E be the event that Sleeping Beauty is awake according to the rules during the experiment. Let $P_E(\cdot|\cdot)$ denote her rational conditional credence

[1] The analysis in this article shows that Sleeping Beauty may apply Bayesian conditionalization to update her coherent opinions from Sunday with respect to the evidence that she acquires when she is awake during the experiment. This is possible even though she is required to suffer the memory loss of Monday's events and, therefore, understands that she does not know whether it is Monday or Tuesday when awake during the experiment. Thus, we dispute Pust's (2012, 296 n. 3) account of what the position in Schervish, Seidenfeld, and Kadane (2004) entails about a rational agent's ability to apply conditionalization in case of an anticipated memory loss of the kind that Sleeping Beauty faces during the experiment.

[2] Schwarz (2015, 3023–24) explicitly mentions this sort of knowledge that Sleeping Beauty might obtain while awake. Instead of treating it as evidence on which to condition, he introduces a principle called doxastic conservatism into the analysis. Hawley (2013) argues that Sleeping Beauty learns no relevant information and uses an inertia principle to justify the halfers' argument.

function while awake during the experiment. Then her conditional probabilities should satisfy these two conditions:

i) $P_E$(heads|it is now Monday) = $P_E$(A|A or B) = 1/2.
ii) $P_E$(it is now Monday|tails) = $P_E$(B|B or C) = 1/2.

Assume that during the experiment, whenever she is awake, the pair {it is now Monday} and {it is now Tuesday} partition her space of centered possibilities. This assumption requires that Sleeping Beauty's space of centered possibilities uses "it is now Monday" and "it is now Tuesday" as both jointly exhaustive and mutually exclusive events. With this assumption, by the Law of Total Probability (see theorem 3 below): $P_E$(heads) = $P_E$(heads|it is now Monday)$P_E$(it is now Monday) + $P_E$(heads|it is now Tuesday)$P_E$(it is now Tuesday) = (1/2)(2/3) + 0(1/3) = 1/3.

In the descriptions of the above arguments, we have stated assumptions explicitly without commenting on their plausibility or compatibility. The arguments are distilled from numerous earlier publications.[3] In the remainder of the article we present a probability model for Sleeping Beauty's acquisition of knowledge during the experiment. Special cases of the model lead to the conclusions drawn from the halfers' argument and the thirders' argument, as well as a number of alternative conclusions. The same principles used in the model allow the modeling of more familiar types of forgetting (as in sec. 5).

## 2. A Day in the Life of Sleeping Beauty

*2.1. What Sleeping Beauty Knows.* On Sunday, and during each experimental awakening, Sleeping Beauty knows that she will (or did) awake on Monday, regardless of how the coin lands. She also knows that she will (or did) awake on Tuesday if and only if the coin lands tails. Beyond these simple assertions, there is little agreement among the discussants of the problem about what she knows or believes.

We begin by assuming that, on Sunday, Sleeping Beauty has a joint probability model for the flip of the coin and what she might know or experience while awake during the experiment. She is also welcome to have a probability distribution over other things that she definitely will not know or experience while awake during the experiment, but we will not make use of such

[3] For illustrations from the literature, the thirders' reasoning is presented in Elga (2000), Dorr (2002), Weintraub (2004), Titelbaum (2008), and Rosenthal (2009). The halfers' conclusion is argued for in Lewis (2001), Cozic (2011), and Hawley (2013), as well as in Elga (2000), although with reasons different from what we presented here.

additional distributions.[4] The only thing that she definitely will not know or experience, but for which she needs a distribution, is the coin flip. We put no restrictions on her probability model except that the space of possible values for "what she knows or experiences while awake during the experiment" is a countable set $\mathcal{X}$. The elements of $\mathcal{X}$ can be sequences of possible sensory inputs that Sleeping Beauty might experience or any other sorts of objects that she cares to use for describing her knowledge at a time when she is awake during the experiment.

Like many other authors, we find the need to refer to Sleeping Beauty's credences at more than one time. Because forgetting is such an important aspect of the problem, we do not want to make heavy-handed assumptions about how she handles updating her credences from one time to the next when forgetting might intervene. We do assume, as do virtually all writers on the subject, that at no point while awake during the experiment does Sleeping Beauty ever forget what she knew on Sunday. This is why we use Sunday as the time at which she formulates her probability model. Whenever she wants to update her credences during the experiment, she takes stock of what she knows, namely, some uncentered element $x$ of $\mathcal{X}$, which includes what she knew on Sunday and updates using the Sunday probability model.

Suppose that, while awake during the experiment, Sleeping Beauty contemplates updating her credences at a time that she labels $t$. The label $t$ might refer to the same time when she contemplates the update (i.e., "now"), or it might refer to some time in the future. She may not know what day or what clock-time corresponds to $t$, and she may not even remember whether $t$ is the first time that she will or did update her credences. She uses $t$ as a label in case she needs to refer to multiple assessments of credences during the experiment. Define $X_{Mt}$ to be a random object that takes values in $\mathcal{X}$ and whose realization is what Sleeping Beauty knows at time $t$ while awake on Monday. Similarly, let $X_{Ut}$ be another random object taking values in $\mathcal{X}$ and whose realization is Sleeping Beauty's knowledge at time $t$ if and while awake on Tuesday.

Based on her Sunday probability model, let $f_{Mt}(\cdot|H)$ be her conditional probability function for $X_{Mt}$ given heads. That is, for each $x \in \mathcal{X}$, $f_{Mt}(x|H)$ is her conditional probability on Sunday that $X_{Mt} = x$ given that the coin lands heads.[5] Also, let $g_t(\cdot, \cdot|T)$ be her conditional joint probability function for

[4] In app. B, we extend Sleeping Beauty's probability model to cover other random variables about which she might be uncertain and about which she might be asked to gamble.

[5] The model that we present allows dependence between the coin flip and what Sleeping Beauty knows. This might appear to violate the spirit of the original Sleeping Beauty problem. The added generality is needed in order to model the example of Rosenthal (2009), which we describe in sec. 3.1.

$(X_{Mt}, X_{Ut})$ given tails. That is, for all $a, b \in \mathcal{X}$, $g_t(a, b|T)$ is her conditional probability on Sunday that $X_{Mt} = a$ and $X_{Ut} = b$ given that the coin lands tails. Our main concern is in determining how Sleeping Beauty updates her credences by conditioning between Sunday and the time at which she announces her credence in heads to the experimenters.

Sleeping Beauty's conditional probability functions for $X_{Mt}$ and $X_{Ut}$ given that the coin lands tails are respectively

$$f_{Mt}(x|T) = \sum_{b\in\mathcal{X}} g_t(x, b|T),$$

and

$$f_{Ut}(x|T) = \sum_{a\in\mathcal{X}} g_t(a, x|T).$$

Let $C_{xMt}$ stand for the event that $X_{Mt} = x$, and let $C_{xUt}$ stand for the event that $X_{Ut} = x$. If Sleeping Beauty's knowledge while awake at time $t$ is $x$, then she is observing the event

$$C_{xt} = C_{xMt} \cup C_{xUt}.$$

If both $X_{Mt}$ and $X_{Ut}$ contain $x$, then when she knows $x$ she will not know whether she has observed $X_{Mt} = x$ or $X_{Ut} = x$. She knows "now" only that she has observed $C_{xt}$.

As $x$ is the totality of Sleeping Beauty's accumulated knowledge at time $t$, she must update her credences at time $t$ by conditioning on $C_{xt}$. If $P(C_{xt}) > 0$, her updated credence in an arbitrary event $K$ is

$$P(K|C_{xt}) = \frac{P(C_{xt} \cap K)}{P(C_{xt})},$$

where $P(\cdot)$ stands for her Sunday probability. The denominator of this expression is

$$P(C_{xt}) = 0.5 f_{Mt}(x|H) + 0.5 f_{Mt}(x|T) + 0.5 f_{Ut}(x|T) - 0.5 g_t(x, x|T). \quad (1)$$

If, for example, $K = H$, the event that the coin lands heads, then

$$P(C_{xt} \cap H) = 0.5 f_{Mt}(x|H). \quad (2)$$

It follows that her conditional credence at time $t$ in the coin landing heads given that her knowledge is $x$ is

$$P(H|C_{xt}) = \frac{f_{Mt}(x|H)}{f_{Mt}(x|H) + f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x, x|T)} \ . \quad (3)$$

The remainder of this section is devoted to characterizing those cases in which (3) coincides with the conclusions drawn in the halfers' and thirders' arguments. We can express the conclusions to those two arguments in terms of (3):

> **Halfers' conclusion:** Sleeping Beauty's credence in heads given what she knows at time *t* is 1/2, no matter what she knows. That is, (3) equals 1/2 for all $x$ such that $P(C_{xt}) > 0$.

If the halfers' conclusion holds at a time *t*, we say that *Sleeping Beauty is a halfer at time t*. Since Sleeping Beauty's probability of heads is 1/2 on Sunday, the halfers' conclusion is equivalent to the coin flip being independent of what she learns according to her probability distribution $P(\cdot)$.

> **Thirders' conclusion:** Sleeping Beauty's credence in heads given what she knows at time *t* is 1/3, no matter what she knows. That is, (3) equals 1/3 for all $x$ such that $P(C_{xt}) > 0$.

If the thirders' conclusion holds at a time *t*, we say that *Sleeping Beauty is a thirder at time t*. Both the halfers' and the thirders' conclusions seem very strong. In sections 2.2 and 2.3, we show that each of these conclusions is equivalent to a strong assumption about what Sleeping Beauty can learn while awake during the experiment.

### *2.2. The Halfers' Argument.*

The explicit assumption made in the halfers' argument in section 1.2 is that the totality of experience that Sleeping Beauty has when she assesses her credences is that she is awake during the experiment, an event to which she had assigned probability 1 on Sunday. To express this in the language of the model of section 2.1, there must be a single element $x_0 \in \mathcal{X}$ (representing what she knows while awake during the experiment at time *t*) in such a way that $f_{Mt}(x_0|H) = g_t(x_0, x_0|T) = 1$. Expressed in these terms, the assumption appears rather strong and possibly implausible. However, a weaker and slightly more plausible assumption also implies the halfers' conclusion. (We provide the proofs to the theorems and corollaries in app. A.)

> Theorem 1 (*halfers' assumption for time t*): A necessary and sufficient condition for (3) to equal one-half for all $x$ with $P(C_{xt}) > 0$ is $f_{Mt}(x|H) = g_t(x, x|T)$ for all $x$ such that $P(C_{xt}) > 0$.

For the remainder of the article, we refer to the necessary and sufficient condition in theorem 1 as "the halfers' assumption for time *t*." We justify this name as follows. The conditions of theorem 1 are necessary and sufficient for drawing the halfers' conclusion at time $t$. If, on Sunday, Sleeping Beauty wishes to draw the halfers' conclusion at time $t$, she implicitly or explicitly makes the halfers' assumption for time $t$.

There is a useful corollary to theorem 1, which highlights the conflict between the halfers' assumption and what we call the thirders' assumption in section 2.3.

Corollary 1. The halfers' assumption implies that $\sum_x g_t(x, x|T) = 1$, for all $x \in \mathcal{X}$. In words, if the coin lands tails, Sleeping Beauty believes that what she knows on Tuesday at time $t$ must be identical to what she knows on Monday at time $t$.

According to corollary 1, the halfers' assumption entails that, with probability 1, everything that Sleeping Beauty knows on Monday at time $t$ (including every ache, pain, and bodily function) will be known again on Tuesday if the coin lands tails. We leave it to the readers to decide whether the halfers' assumption is what was intended when the Sleeping Beauty problem was posed.[6]

*2.3. The Thirders' Argument.* The explicit assumption made in the thirders' argument is that, while awake during the experiment, {it is now Monday} and {it is now Tuesday} partition Sleeping Beauty's sure event. Assessing the plausibility of this assumption as well as its compatibility with the halfers' assumption in theorem 1 requires understanding the centered propositions "it is now Monday" and "it is now Tuesday." But, we can assess the thirders' conclusion directly using only probability theory. In particular, we ask the simpler question: "What assumption is equivalent to (3) being equal to 1/3 for all $x$?" The answer is contained in the next theorem.

Theorem 2 (*thirders' assumption for time t*): A necessary and sufficient condition for (3) to equal one-third for all $x \in \mathcal{X}$ with $P(C_{xt}) > 0$ is (i) $\sum_x g_t(x, x|T) = 0$, and (ii) $f_{Mt}(x|H) = 0.5[f_{Mt}(x|T) + f_{Ut}(x|T)]$, for all $x$ such that $P(C_{xt}) > 0$.

In words, (i) if the coin lands tails, Sleeping Beauty believes that what she knows at time $t$ on Tuesday must be different from what she knows at time $t$ on Monday, and (ii) the conditional distribution of what she knows at time $t$ on Monday given heads is the average of the conditional distribution of what she knows at time $t$ on Monday given tails and the conditional distribution of what she knows at time $t$ on Tuesday given tails.

[6] As if in anticipation of the Sleeping Beauty problem, Alpern (1988) introduced agents in multiagent games who have limited memory and who reach the same information set at multiple times during the game without knowing how often they have done so.

For the remainder of the article, we refer to the necessary and sufficient conditions in theorem 2 as “the thirders’ assumption for time *t*.” The conditions of theorem 2 are necessary and sufficient for drawing the thirders’ conclusion. Hence, if Sleeping Beauty wishes to draw the thirders’ conclusion at time *t*, she implicitly or explicitly makes the thirders’ assumption for time *t*. The clearest incompatibility between the halfers’ and thirders’ assumptions is as follows. The halfers’ assumption requires that whatever Sleeping Beauty knows on Monday she must also know on Tuesday if the coin lands tails, while the thirders’ assumption requires that what she knows on Monday and Tuesday must be different if the coin lands tails.

It is comforting to see that both halfers and thirders can reach their desired conclusions without violating any of the mathematical theory of probability, so long as they each carefully state the assumptions that they are making. If neither the halfers’ assumption nor the thirders’ assumption holds for some time *t*, then Sleeping Beauty’s credence in heads at time *t* could vary with the *x* that she knows and could even take values outside of the interval [1/3, 1/2], depending on the specific version of the model for her knowledge. Although some of those versions are interesting, pursuing them all would divert us from the main points of this article. In appendix C, we illustrate one version that we find interesting primarily for its having been ignored in so much of the Sleeping Beauty literature. We show that, for every *q* between 1/3 and 1/2, there are distributions for what Sleeping Beauty might learn with the property that (3) equals *q* for all *x*. In other words, halfers and thirders should not have a monopoly on the controversy. They are merely the extremes of a continuum of *q*-ers for all $1/3 \leq q \leq 1/2$.

**3. Examples of the Thirders’ Assumption.** Very few authors explicitly entertain assumptions anything like the thirders’ assumption. Notable exceptions are Meacham (2008), Titelbaum (2008), and Rosenthal (2009), which we consider next. These papers all have one thing in common: they introduce possible information that Sleeping Beauty might learn during the experiment that has the property that the conditional probability of heads given every possible value of this information is 1/3. But, they all insist on concluding that she should then assign probability 1/3 to heads even if she does not learn the information. They want to draw the conclusion that would follow from conditioning without doing the conditioning. In section 4, we explain why this is not justified within the theory of probability.

*3.1. Rosenthal’s Dime.* Rosenthal (2009) introduced a variation on the Sleeping Beauty problem in which she (or somebody else) contemplates another coin flip, whose result is a special case of our $X_{Mt}$ or $X_{Ut}$ information that she might observe. First, Rosenthal refers to the coin that is flipped in the original Sleeping Beauty problem as *nickel*, and the two possible values of the flip are called *NickelHeads* and *NickelTails*. The new coin is called *dime*. Precisely how dime is used is more complicated than how nickel is used. In particular, there is dependence between dime and nickel. Specifically, we quote from Rosenthal (2009, 33):

> If the nickel showed tails, then the dime is simply placed so that it shows heads during Beauty's Monday interview, and then repositioned so that it shows tails during Beauty's Tuesday interview. If instead the nickel showed heads (so Beauty will only be interviewed once), then the dime is instead simply flipped once in the usual fashion at the beginning of the Experiment, and is allowed to show its actual flipped result (either heads or tails, with probability 1/2 each) during the one interview that will take place on Monday. Furthermore, we assume that Beauty is not allowed to see the dime at all, and might not even know of its existence.

To express the use of dime in terms of the model in section 2.1, let $t$ be the time at which dime becomes observable. Define $X_{Mt} = 1$ or $X_{Ut} = 1$ if dime shows heads on the day corresponding to the subscript, and let $X_{Mt} = 0$ or $X_{Ut} = 0$ if dime shows tails.[7]

It follows that $g_t(1,1|T) = g_t(0,0|T) = g_t(0,1|T) = 0$, $g_t(1,0|T) = 1$, $f_{Mt}(1|H) = f_{Mt}(0|H) = 0.5$, and $f_{Ut}(x|T) = 1 - x$, for $x = 0, 1$. These numbers satisfy the thirders' assumption for time $t$; hence, (3) equals 1/3 for both $x = 0$ and $x = 1$. As the end of the above quote makes clear, Rosenthal assumes that Sleeping Beauty does not observe the result of the dime. In section 4, we explain why Sleeping Beauty needs to observe an event at time $t$ that satisfies the thirders' assumption for time $t$ (such as the result of the dime) in order to change her credence in heads from one-half to one-third.

*3.2. Titelbaum's Technicolor Beauty.* Titelbaum (2008, 591–92) introduces a variant of the Sleeping Beauty problem in which she is offered knowledge of a specific sort:[8]

> Everything is exactly as in the original Sleeping Beauty Problem, with one addition: Beauty has a friend on the experimental team, and before she falls asleep Sunday night he agrees to do her a favor. While the other experimenters flip their fateful coin, Beauty's friend will go into another room and roll a fair die. (The outcome of the die roll is independent of the outcome of the coin flip.) If the die roll comes out odd, Beauty's friend will

[7] For ease of notation, we indicate only the new evidence Sleeping Beauty acquires during the experiment, without repeating all that she recalls from Sunday.

[8] Meacham (2008, 263) gives a similar example in which "see a red paper" is replaced by "wake up in a black room," and "see a blue paper" is replaced by "wake up in a white room."

> place a piece of red paper where Beauty is sure to see it when she awakens Monday morning, then replace it Tuesday morning with a blue paper she is sure to see if she awakens on Tuesday. If the die roll comes out even, the process will be the same, but Beauty will see the blue paper on Monday and the red paper if she awakens on Tuesday.
>
> Certain that her friend will carry out these instructions, Beauty falls asleep Sunday night. Some time later she finds herself awake, uncertain whether it is Monday or Tuesday, but staring at a colored piece of paper. What does ideal rationality require at that moment of Beauty's degree of belief that the coin came up heads?

To express Technicolor Beauty in terms of our model from section 2.1, let $t$ be the time at which she observes the colored paper, and let $\mathcal{X} = \{R, B\}$. Then $f_{Mt}(x|H) = f_{Mt}(x|T) = f_{Ut}(x|T) = 1/2$ for all $x \in \mathcal{X}$, and $g_t(R, B|T) = g_t(B, R|T) = 1/2$. It follows from theorem 2 that Sleeping Beauty is a thirder at time $t$. Like Rosenthal, Titelbaum wants to be able to claim that Sleeping Beauty's credence in heads should be 1/3 even if she does not see the colored paper. We explain why we disagree in section 4.

However, Titelbaum makes a very insightful comment about what happens if Sleeping Beauty does get to see the color of the paper (2008, 592): "However, the addition of the colored papers has given Beauty a uniquely denoting context-insensitive expression for 'today.' On Monday morning, Beauty is certain that 'the red paper day' uniquely picks out the denotation of 'today.'" In section 5, we see how the idea expressed in this quote helps to distinguish the Sleeping Beauty problem from more familiar cases of forgetting.

**4. The Law of Total Probability: When It Applies and When It Does Not.** Assume the thirders' assumption as stated in theorem 2. If Sleeping Beauty knows (on Sunday) that she is going to learn something that causes her to assign probability 1/3 to heads, why does she not assign probability 1/3 before she learns that something?[9] The probabilistic intuition behind this question is the following well-known theorem.

Theorem 3 (*Law of Total Probability*): Let $B_1, B_2, \dots, B_n$ be events that satisfy $P(\cup_{i=1}^{n} B_i) = 1$ and $P(B_i \cap B_j) = 0$ for $i \neq j$. Then, for every event $A$, $P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j)$.

[9] This question can be reexpressed as, "Does Sleeping Beauty violate the Reflection Principle?" See van Fraassen (1995) for a statement of the Reflection Principle. Elga (2000, sec. 3) discusses the Reflection Principle in the context of Sleeping Beauty, seemingly without being aware of the distinction between the Law of Total Probability and theorem 4. For more discussion of the Reflection Principle, see Schervish et al. (2004).

The proof of theorem 3 is straightforward and omitted.[10] In the special case in which $P(A|B_j) = c$, for all $j$, $P(A) = c$ follows from the Law of Total Probability. If each $B_j$ is the event that Sleeping Beauty learns one of the things that will cause her to change her credence in heads from 1/2 to 1/3, why does theorem 3 not tell her that her credence in heads should be 1/3 before observing one of the $B_j$ events? The reason is one of the subtle features of the Sleeping Beauty problem that challenges the intuition. In the examples from section 3, and under the thirders' assumption in general, the $C_{xt}$ events do not satisfy the assumptions of theorem 3. Their intersections have positive probability.[11]

*Example 1*: If $\sum_x g_t(x,x) < 1$, there must exist $x, y \in \mathcal{X}$ with $x \neq y$, such that $g_t(x,y) > 0$. It follows that $P(C_{xt} \cap C_{yt}) \geq (1/2)g_t(x,y) > 0$.

There is a theorem that applies when $P(\cup_{i=1}^{n} B_i) = 1$, but at least one of the intersections of the sets has positive probability. The proof of theorem 4 is virtually identical to the proof of the formula for the union of a finite number of events and is not given here.[12]

Theorem 4: Let $B_1, B_2, \ldots, B_n$ be events that satisfy $P(\cup_{i=1}^{n} B_i) = 1$. Then, for every event $A$,

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j) - \sum_{j \neq k} P(A|B_j \cap B_k)P(B_j \cap B_k) + \ldots + (-1)^{n+1} P(A|\cap_{j=1}^{n} B_j)P(\cap_{j=1}^{n} B_j).$$

We illustrate theorem 4 with Technicolor Beauty. Titelbaum (2008, 596) explicitly considers a label $s$ that Technicolor Beauty assigns to a time after she awakens but before she sees the colored paper. What does probability theory say is her updated credence at time $s$? It depends, of course, on what she knows at time $s$. For example, suppose that what she knows at time $s$ satisfies the halfers' assumption, but later (at time $t$) she will see the colored paper. Let $R$ stand for the event that she sees the red paper at time $t$, and let $B$ stand for the event that she sees the blue paper at time $t$. Also, let $x_s$ (with

[10] The Law of Total Probability has a conditional version as well. If $C$ is a further event such that for each $j$, $B_j \cap C \neq \phi$, then $P(A|C) = \sum_{j=1}^{n} P(A|B_j \cap C)P(B_j|C)$.

[11] While probability theory without a partition is coherent, the Law of Total Probability is unavailable in these cases.

[12] See, e.g., DeGroot and Schervish (2012, sec. 1.10). Theorem 4 also has a conditional version given a further event $C$. Just make every probability in the statement of theorem 4 conditional on the intersection of $C$ with those events on which it is already conditional.

$P(C_{x_s s}) > 0$) stand for what she knows at time $s$. For simplicity, assume that she also knows that (i) there will be no forgetting between times $s$ and $t$, (ii) all she will know at time $t$ is $C_{x_s s}$ and the color of the paper, and (iii) $C_{x_s s}$ is independent of $R$ and $B$ according to $P(\cdot)$.[13] Then, at time $s$, what she knows is that she has observed the event $C_{x_s s}$, such that $P(H|C_{x_s s}) = 1/2$. Because of conditions i and ii she also knows that at time $t$ she will have observed either $C_{x_t t} = C_{x_s s} \cap R$ or $C_{x_t t} = C_{x_s s} \cap B$.

Let $P_s(\cdot) = P(\cdot|C_{x_s s})$ stand for her probability distribution at time $s$, after she knows $x_s$. Then the question we need to answer is, "What is the nature of $P_s(\cdot)$?" First, we know that $P_s(H) = 1/2$, because her probability satisfies the halfers' assumption for time $s$. Second, we know how she will condition on possible knowledge at time $t$ using $P_s(\cdot)$. For example, she can compute $P_s(H|R)$ using $P_s(R)$ and $P_s(H \cap R)$. By assumption iii, we can compute these using the distributions in section 3.2. In particular,

$$P_s(R) = P_s(H)P_s(R|H) + P_s(T)P_s(R|T) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1 = \frac{3}{4},$$

because she will observe $R$ with probability 1/2 given heads but with probability 1 given tails. Similarly, $P_s(H \cap R) = 1/4$. Hence, $P_s(H|R) = 1/3$, as expected. Similarly, $P_s(B) = 3/4$, and $P_s(H|B) = 1/3$. Also, $B \cap R = T$, so $P_s(B \cap R) = 1/2$. Theorem 4 applies to conclude that

$$\begin{aligned} P_s(H) &= P_s(H|R)P_s(R) + P_s(H|B)P_s(B) - P_s(H|R \cap B)P_s(R \cap B) \\ &= \frac{1}{3} \times \frac{3}{4} + \frac{1}{3} \times \frac{3}{4} - 0 \times \frac{1}{2} = \frac{1}{2}. \end{aligned}$$

Finally, we notice that $P_s(\cdot|R) = P(\cdot|C_{x_s s} \cap R)$, and similarly for $B$. If $x_t$ is what she knows at time $t$, then $C_{x_t t} = C_{x_s s} \cap R$ or $C_{x_t t} = C_{x_s s} \cap B$ depending on whether she sees the red paper or the blue paper at time $t$. Hence, if she sees the red paper, $P(H|C_{x_t t}) = P_s(H|R) = P(H|C_{x_s s} \cap R) = 1/3$, and similarly if she sees the blue paper.

The import of the calculations immediately above is that it is consistent with probability theory and conditioning for Sleeping Beauty to have credence 1/2 in heads at a time $s$ while awake in the experiment and to know that she is about to observe data such that her credence will drop to 1/3, regardless of the particulars of those data, so long as the probabilities for those data do not satisfy the Law of Total Probability. The same argument

[13] Assumption i could be weakened at the expense of requiring a model for what Sleeping Beauty remembers and forgets as the day advances. Assumption ii could be weakened at the expense of notational clutter to represent intervening knowledge acquisition. Assumption iii is needed so that the description of the Technicolor Beauty problem means the same thing at time $s$ as it does on Sunday.

applies to Meacham's white and black rooms. A similar argument applies to Rosenthal's dime, but it involves different numbers because of the dependence between dime and nickel.

The preceding results may seem counterintuitive, but there is some intuition to support them. Under the thirders' assumption, for every $x$, the probability of learning $x$ given tails is twice as large as the probability of learning $x$ given heads because Sleeping Beauty has two opportunities to learn $x$ given tails compared to only one opportunity given heads. If she starts with equal probabilities for heads and tails, then learning $x$ will make the probability of tails twice as high as the probability of heads. So, Sleeping Beauty's probability of heads drops to 1/3 after she learns $x$ but not before. Even when Sleeping Beauty will acquire a unique context-insensitive expression for the day, such as a red paper or a white room or Rosenthal's dime, she is not entitled to use the Law of Total Probability with the events $\{C_{xt} : x \in \mathcal{X}\}$ because their intersections have positive probability. Rather, if she wants to compute the probability of heads before observing a $C_{xt}$, she must apply theorem 4. After she learns the unique context-insensitive expression for the day, she can condition on having learned it and update her credences accordingly.

**5. The Absentminded Professor.** As a more common example of forgetting, imagine an absentminded professor who loses track of the time while delivering a lecture. There is no clock in the room, but he knows that students will start to leave if he goes on more than a few minutes past the scheduled end of class. He asks himself, "What time is it?" He does not have a clock, so he looks at his notes and sees that he has just finished 10 out of 20 pages of notes that he had prepared for the lecture. Of course, he generally prepares more pages than needed because it is more difficult to recover from running out of material than it is to pick up where he left off last time. So, he is pretty sure that he is more than halfway through the scheduled lecture period but is uncertain about precisely how much time is left.

The absentminded professor has access to what he knows, which includes the fact that he has completed 10 pages of the lecture that he has been delivering along with any other experiences he remembers that help to distinguish this lecture from any other similar lecture that he may have given in the past. What matters to him is the probability distribution of the amount of time that would elapse from the start of the lecture until he finished 10 pages of the notes. This will allow him to compute the probability distribution of the time remaining in the lecture along with the probability distribution of "what time it is."

We should stress that the absentminded professor's forgetting of the time is very different from Sleeping Beauty's forgetting. For one thing, the absentminded professor does not believe that there is positive probability that he is reliving an earlier experience that he has forgotten. Nor does he believe that there is positive probability that he will have to relive the current experience after being forced to forget it. In Titelbaum's language (quoted earlier) the absentminded professor has "a uniquely denoting context-insensitive expression" for the current time. At time $t$ while awake, Sleeping Beauty has a uniquely denoting context-insensitive expression for the day if and only if she has learned an $x$ such that $g_t(x, x|T) = 0$. Otherwise there is positive probability that she is either reliving an earlier experience or will relive her current experience after forced forgetting.

## 6. Discussion.

The halfers' argument (in sec. 1.2) relies on a strong assumption about what Sleeping Beauty knows while awake during the experiment. We noted earlier that the halfers' conclusion at time $t$ is equivalent to the coin flip being independent of $C_{xt}$ for all $x$ according to Sleeping Beauty's Sunday distribution. Intuition would suggest that

$$f_{Mt}(\cdot|H) = f_{Mt}(\cdot|T) = f_{Ut}(\cdot|T) \tag{4}$$

expresses independence between the coin flip and what Sleeping Beauty knows at time $t$. But intuition often fails in the Sleeping Beauty problem. For example, the event that Sleeping Beauty observes is $C_{xt}$, which takes into account the fact that she gets two opportunities to observe $x$ if the coin lands tails but only one opportunity if the coin lands heads. In order for $C_{xt}$ to be independent of the coin flip, the second opportunity to observe $x$ cannot change the probability of observing $x$. This fact is what makes the halfers' assumption so strong. From the experimenter's viewpoint, (4) expresses independence between the coin flip and what Sleeping Beauty knows but not from Sleeping Beauty's viewpoint.

Rather than appeal to additional principles in order to accommodate possible-world semantics, we have identified necessary and sufficient conditions for both halfers and thirders to be able to achieve their desired conclusions within the realm of probability theory. Not surprisingly, the conditions needed by the two groups are incompatible with each other. But at least we now understand from whence the differences arise.

It is difficult to determine which assumption (if either) is more sensible or more compatible with the original intention of the Sleeping Beauty problem. Even Elga (2000, 145) fails to acknowledge that Sleeping Beauty's knowledge while awake during the experiment might account for her change in credence from one-half to one-third: "This belief change is unusual. It is not the result of your receiving new information—you were already certain that you would be awakened on Monday." Sleeping Beauty's belief change may appear "unusual," but whether she changes her belief will follow from standard probability calculus once she is explicit about how she models her "new information."

We believe that the main contribution of this article is in making explicit the assumptions that are necessary and sufficient for drawing either of the competing conclusions. We have accomplished this by using a probability model for Sleeping Beauty's knowledge while awake during the experiment. The model (in sec. 2.1) is the most general model possible if one assumes that what she knows lies in a discrete space. Under this general model, at each time *t* during an awakening, she is a thirder at time *t* if and only if she satisfies the thirders' assumption for time *t*, and she is a halfer at time *t* if and only if she satisfies the halfers' assumption for time *t*. As a side point, we show in appendix C that halfers and thirders do not have the whole show to themselves. There are *q*-ers for every *q* from 1/3 to 1/2. They just have not been as prolific in contributing to the literature.

We did not take the approach of creating variations on the Sleeping Beauty problem in order to support our reasoning. We did analyze a few of the existing variations (see secs. 3 and 4) to illustrate the wide applicability of our modeling approach. Other variations are also amenable to our analysis, but these would require more complicated models because they modify the assumptions of the problem in more fundamental ways. For example, White (2006) offers one of several possible changes to the assumptions about how/when Sleeping Beauty awakens. There are also several possible ways to change the assumptions about how/when she forgets. We illustrate how probability theory can shed light on the controversy surrounding the most elementary versions of the problem.

A secondary contribution is that we have compared and contrasted the type of forgetting that plagues people in everyday life to the contrived situation in which Sleeping Beauty finds herself (see sec. 5). A third contribution is that we have clarified both the relationship and the differences between Sleeping Beauty's credence in heads and her fair price for betting on heads in the experimental setting (see app. B).

## Appendix A

### Proofs of Theorems

*Proofs of Theorem 1 and Corollary 1.* First, we prove sufficiency of the condition. The condition stated in the theorem and the formulas for $f_{Mt}(x|T)$ and $f_{Ut}(x|T)$ imply that

$$f_{Mt}(x|T) = f_{Ut}(x|T) = g_t(x, x|T) = f_{Mt}(x|H), \tag{A1}$$

for all $x$ such that $P(C_{xt}) > 0$. Inserting (A1) into (3) yields 1/2 for all $x$ such that $P(C_{xt}) > 0$.

Next, we prove necessity of the condition. Notice that (3) equal to 1/2 for all $x$ such that $P(C_{xt}) > 0$ implies that

$$f_{Mt}(x|H) = f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x,x|T), \tag{A2}$$

for all $x$ such that $P(C_{xt}) > 0$. The sum of the left-hand side of (A2) is 1, while the sum of the right-hand side is $2 - \sum_x g_t(x,x|T)$. Hence,

$$\sum_x g_t(x,x|T) = 1, \tag{A3}$$

which proves corollary 1. It follows from (A3) that $g_t(x,y|T) = g_t(y,x|T) = 0$, for all $y \neq x$. It follows from the formulas for $f_{Mt}(x|T)$ and $f_{Ut}(x|T)$ and (A2)–(A3) that $f_{Mt}(x|H) = g_t(x,x|T)$. Finally, note that the verbal description in corollary 1 is equivalent to (A3).

*Proof of Theorem 2.* For sufficiency, note that i implies that $g_t(x,x|T) = 0$ for all $x$. Substitute this and ii into (3) and the result is 1/3. For necessity, notice that (3) equal to one-third implies

$$2f_{Mt}(x|H) = f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x,x|T). \tag{A4}$$

A necessary condition for (A4) to hold for all $x$ with $P(C_{xt}) > 0$ is that the sums over all $x$ with $P(C_{xt}) > 0$ of the two sides of (A4) be equal. This implies $\sum_x g_t(x,x|T) = 0$, which is i. As before, i implies $g_t(x,x|T) = 0$ for all $x$. Substituting this into (A4) implies ii. The verbal descriptions of i and ii are clearly the same as their formulas.

## Appendix B

### Gambling during the Experiment

*A Thirders' Gambling Argument.* There is a second thirders' argument that is designed to answer the question, "What fair price should Sleeping Beauty offer for a bet on whether the coin lands heads?" Some thirders reason that this fair price is also her credence, or degree of belief in the proposition that the coin lands heads. The second argument is as follows:

> If we consider a large number $n$ of probabilistically independent repetitions of the experiment, with probabilistically independent flips of the same fair coin, on about $n/2$ trials, the fair coin lands heads and on about $n/2$ trials it

> lands tails. When the coin lands tails Sleeping Beauty is asked separately on both Monday and on Tuesday to contract a bet on heads. So, on about *n*/3 of all the occasions when Sleeping Beauty is awake and asked to bet with a fresh contract, the outcome is heads. So, her fair betting odds on heads ought to be 1:2; that is, Sleeping Beauty should give a fair betting rate of 1/3 on heads whenever asked during the experiment. If these fair odds also elicit her credences about the coin flip, as is typical with ordinary cases of fair betting, then when awake during the experiment her degree of belief that the coin lands heads also should be 1/3.

We agree with the thirders that Sleeping Beauty's fair price for betting on heads is 1/3 and not 1/2. But, as has been noted by others (Bradley and Leitgeb 2006; Briggs 2010; Yamada 2011) because there is a negative correlation between Sleeping Beauty's betting opportunities and the outcome heads, the special circumstances of the Sleeping Beauty problem provide grounds for distinguishing between what might be Sleeping Beauty's credence and her fair price for betting on heads. In deriving her fair price for betting on heads, equation (B3) below, we note that the relationship between what she knows at the time of the bet and her fair price is not the same as the relationship between what she knows and her credence.

We discuss these points in detail in the remainder of this appendix, where we apply de Finetti's (1972, 1974) theory of fair gambles to show that one may elicit Sleeping Beauty's credence for heads from her fair price for betting on heads, although these are not the same quantities. Assume that some time is being considered and all random variables and distributions are indexed by that time.

*General Gambles.* When awake during the experiment, Sleeping Beauty can be offered a gamble on any random variable about which she is uncertain. In the basic Sleeping Beauty problem, the indicator $H$ of the event that the coin lands heads is commonly used as the only example of such a random variable. We will first extend the model of section 2.1 to include random variables that remain unobserved while Sleeping Beauty is awake during the experiment and then show how she should set her fair prices for betting on such random variables. A general gamble on a single random variable $Y$ can be expressed as

$$\beta B(Y-p), \tag{B1}$$

where $p$ is a price specified by a *bookie* (Sleeping Beauty in this case), $B$ is (the indicator function for) an event such that the gamble is called off if $B$ fails to occur, and $\beta$ is a real number chosen by a *gambler* (the experimenter in this case). The value in (B1) is the amount the gambler receives (and the bookie pays) when the bet is settled. In order for the gamble in (B1) to be fair (to the bookie), the bookie's expected value of (B1) must be 0.[14]

Sleeping Beauty has the opportunity (requirement?) to gamble each time that she is awake during the experiment. On Monday she will not realize that she is gambling for the first time, and if the coin lands tails, she can gamble again on Tuesday, but she will not realize that she is gambling a second time. In the basic Sleeping Beauty problem, she is asked to gamble on the same random variable $Y = H$ on both days, with the Tuesday gamble called off if $T$ fails to occur. In principle, it is not necessary that the same random variable be the object of the gamble both days. What is required is that (i) the Monday and Tuesday random variables (and events, if any, for calling off the gambles) are known on Sunday, (ii) Sleeping Beauty's announced fair price and the experimenter's $\beta$ can depend only on what Sleeping Beauty knows at the time of each gamble, (iii) the mappings from Sleeping Beauty's knowledge to the price $p$ and the coefficient $\beta$ must be known on Sunday, and (iv) the Monday and Tuesday random variables and the function $\beta$ are all bounded. The third condition is to prevent possible cheating by the experimenter who might have "inside" information. The fourth condition avoids mathematical contortions that are required for unbounded gambles. The combined effect of the gambles to which she is subject is

$$\sum_{x \in \mathcal{X}} \beta(x)\{B_M C_{xMt}[Y_M - p(x)] + B_U C_{xUt}[Y_U - p(x)]\}, \qquad \text{(B2)}$$

where the sum is over all $x$ that Sleeping Beauty might know right before being asked to give her fair price, $Y_M$ and $Y_U$ are the bounded random variables on which the Monday and Tuesday gambles respectively are based, and $B_M$ and $B_U$ are events such that the Monday and Tuesday gambles are respectively called off if the corresponding event fails to occur. It is required that $T \subseteq B_U$ because the Tuesday gamble is called off if tails fails to occur. Sleeping Beauty avoids sure loss if and only if the expected value of (B2) is 0 for all bounded $\beta(\cdot)$ functions. The expected value of (B2) is 0 for all $\beta(\cdot)$ functions if and only if, for each $x$ with $P(C_{xt}) > 0$, the conditional expected value of the part of (B2) between the $\{\cdots\}$ symbols is 0. The resulting conditional mean in question is

$$\frac{\mathrm{E}(B_M C_{xMt} Y_M) - p(x)\mathrm{E}(B_M C_{xMt}) + \mathrm{E}(B_U C_{xUt} Y_U) - p(x)\mathrm{E}(B_U C_{xUt})}{P(C_{xt})},$$

[14] Theorems B.139 and B.141 of Schervish (1995) show that a bounded sum of a countable collection of gambles of the form of (B1) avoids sure loss if and only if there exists a probability $Q(\cdot)$ such that $Q(BY) = pQ(B)$ for each gamble. If $Q(B) > 0$, this is equivalent to $p = Q(Y|B)$. Gambles such as (B1) are designed to elicit the conditional mean of $Y$ given $B$.

which is 0 for all $x$ with $P(C_{xt}) > 0$ if and only if

$$p(x) = \frac{\mathrm{E}(B_M C_{xMt} Y_M) + \mathrm{E}(B_U C_{xUt} Y_U)}{\mathrm{E}(B_M C_{xMt}) + \mathrm{E}(B_U C_{xUt})}.$$

It is interesting to compare the fair price $p(x)$ to a conditional mean given $C_{xt}$. Consider the special case in which $Y_M = Y_U = Y$, $B_M = \Omega$ (the sure event), and $B_U = T$; we find that

$$p(x) = \mathrm{E}(Y|C_{xt})\frac{f_{Mt}(x|H) + f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x,x|T)}{f_{Mt}(x|H) + f_{Mt}(x|T) + f_{Ut}(x|T)},$$

which equals $\mathrm{E}(Y|C_{xt})$ if and only if $g_t(x,x|T) = 0$. In words, the gambles that make up (B2) elicit Sleeping Beauty's conditional means given $C_{xt}$ if and only if she satisfies part i of the thirders' assumption. In particular, if she knows an $x$ to which she assigns positive probability of knowing on both days, her fair price will necessarily be lower than her conditional mean. We examine the implications of this for the gamble on heads in the next section.

*If the Coin Is Fair.* The special case of most immediate interest is that of the original Sleeping Beauty problem in which the coin is fair, $B_M = \Omega$ (the sure event), $B_U = T$, and $Y_M = Y_U = H$ (the indicator of heads). In that case,

$$p(x) = \frac{0.5f_{Mt}(x|H)}{0.5f_{Mt}(x|H) + 0.5f_{Mt}(x|T) + 0.5f_{Ut}(x|T)}. \tag{B3}$$

Under both the halfers' and the thirders' assumptions, (B3) equals 1/3 for all $x$. So halfers and thirders agree that Sleeping Beauty should offer 1/3 as a fair price for a gamble on heads that gets executed on Monday and then again on Tuesday if the coin lands tails. But they do not agree on her credence in the event that the coin lands heads.[15] The mathematical derivation of (B3) proves that halfers and thirders agree on the fair price, but there is some intuition about why this happens in spite of the differing credences. Once a thirder observes $C_{xt}$, she knows that she is subject to one and only one of the two gambles in (B2) that correspond to $x$, namely, either $\beta(x)\{C_{xMt}[H - p(x)]\}$ or $\beta(x)\{C_{xUt}[H - p(x)]\}$ but not both. Unfortunately, she does not know which. If the coin lands tails, she will also be subject to one and only one of a different pair of gambles corresponding to a different $x'$. Because she has a uniquely denoting context-insensitive expression for the day, she can apply the Law of Total Probability conditional on $C_{xt}$ using the partition $C_{xMt}$ and

[15] If Sleeping Beauty's model for what she knows satisfies (4), then (B3) equals 1/3 for all $x$ even if she is neither a halfer nor a thirder.

$C_{xUt}$. The weighted average of the two conditional fair prices given these two events will be $P(H|C_x) = 1/3$. The same thing happens on both days, with different $x$ values, if the coin lands tails. A halfer, however, knows that if the coin lands tails she will know the same $x$ on both days, and she has no partition of $C_{xt}$ available for use with the Law of Total Probability. Hence, she is subject to both gambles $\beta(x)\{C_{xMt}[H - p(x)]\}$ and $\beta(x)\{C_{xUt}[H - p(x)]\}$ and must choose $p(x)$ to make the expected value of the sum equal to 0. The result will not be her credence but rather the formula for $p(x)$ in (B3). In effect, the thirder adjusts her credence on the basis of the potential two opportunities to observe $x$ and is then able to use her credence as a fair betting price. The halfer makes no adjustment in her credence because she observes nothing that can change her credence. But she adjusts the fair betting price because the two gambles to which she is subject do not elicit her credence.

*If the Coin Is Unfair*. In order to better understand the relationship between the Sunday credence in heads, the credence in heads during the experiment, and the fair price for a bet on heads, instead suppose that Sleeping Beauty believes on Sunday that the probability of heads is $z \in (0, 1)$. It is straightforward to see that (1) and (3) would change to

$$P(C_{xt}) = zf_{Mt}(x|H) + (1 - z)[f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x, x|T)]. \tag{1$'$}$$

$$P(H|C_{xt}) = \frac{zf_{Mt}(x|H)}{zf_{Mt}(x|H) + (1 - z)[f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x, x|T)]} \cdot \tag{3$'$}$$

To generalize the halfers' argument, Sleeping Beauty's credence in heads remains unchanged when she awakes during the experiment, so it is $z$. We state the following modification of theorem 1 without proof because the proof is almost the same as the proof of theorem 1.

Proposition 1. Assume that Sleeping Beauty's probability of heads on Sunday is $z$. A necessary and sufficient condition for (3′) to equal $z$ for all $x$ with $P(C_{xt}) > 0$ is $f_{Mt}(x|H) = g_t(x, x|T)$ for all $x$ such that $P(C_{xt}) > 0$.

To generalize the thirders' argument is slightly more complicated. It still seems intuitive that (in the same notation as in sec. 1.2)

i) $P_E$(heads|it is now Monday) = $P_E$(A|A or B) = $z$.
ii) $P_E$(it is now Monday|tails) = $P_E$(B|B or C) = 1/2.

Assuming (as in the original thirders' argument) that A, B, and C form a partition, we compute

$$\mathrm{P_E}(\mathrm{A}) = \frac{z}{2-z},$$

$$\mathrm{P_E}(\mathrm{B}) = \frac{1-z}{2-z},$$

and

$$\mathrm{P_E}(\mathrm{C}) = \frac{1-z}{2-z}.$$

Pretending as if the Law of Total Probability applied, one would compute

$$\begin{aligned}\mathrm{P_E}(\text{heads}) &= \mathrm{P_E}(\text{heads}|\text{Monday})\mathrm{P_E}(\text{Monday}) \\ &\quad + \mathrm{P_E}(\text{heads}|\text{Tuesday})\mathrm{P_E}(\text{Tuesday}) \\ &= z\left(\frac{1}{2-z}\right) + 0\left(\frac{1-z}{2-z}\right) = \frac{z}{2-z}.\end{aligned}$$

That is, thirders replace 1/3 by $z/(2-z)$ if the coin is not fair. We now state a modification of theorem 2 when the coin is not fair.

> Proposition 2. Assume that Sleeping Beauty's probability of heads on Sunday is $z$. A necessary and sufficient condition for (3′) to equal $z/(2-z)$ for all $x$ with $P(C_{xt}) > 0$ is (i) $\sum_x g_t(x,x) = 0$ and (ii) $f_{Mt}(x|H) = 0.5[f_{Mt}(x|T) + f_{Ut}(x|T)]$, for all $x$ such that $P(C_{xt}) > 0$.

The proof of proposition 2 is almost the same as the proof of theorem 2. For the gamble in the basic Sleeping Beauty problem, (B3) changes to

$$p(x) = \frac{zf_{Mt}(x|H)}{zf_{Mt}(x|H) + (1-z)f_{Mt}(x|T) + (1-z)f_{Ut}(x|T)}. \tag{B3′}$$

With an unfair coin, both halfers and thirders agree that $p(x)$ equals $z/(2-z)$ for all $x$ such that $P(C_{xt}) > 0$. If Sleeping Beauty is either a halfer or a thirder (and the experimenter knows this), the experimenter can recover her Sunday probability of heads from her fair price by the formula

$$z = \frac{2p(x)}{p(x)+1}.$$

## Appendix C

### Thirders, Halfers, and Everyone in Between

Setting equation (3) equal to $q$ for all $x$ such that $P(C_{xt}) > 0$ is equivalent to

$$f_{Mt}(x|H) = \frac{q}{1-q}[f_{Mt}(x|T) + f_{Ut}(x|T) - g_t(x,x|T)], \tag{C1}$$

for all $x$ such that $P(C_{xt}) > 0$. The left-hand side of (C1) adds to 1, while the right-hand side adds to $q/(1-q)$ times a number bounded between 1 and 2. It follows easily that $q$ must lie between 1/2 and 1/3 (including the endpoints). Theorems 1 and 2 give necessary and sufficient conditions to achieve the endpoints. The values interior to the interval are slightly more complicated to achieve. In a special class of cases, we can determine when (C1) holds. Assume that (4) holds, and call the common function $f_t(\cdot)$. Then (C1) becomes

$$\frac{3q-1}{q} f_t(x) = g_t(x,x|T), \tag{C2}$$

for all $x$ such that $P(C_{xt}) > 0$. Examples of (C2) are easy to construct. For instance, start with Technicolor Beauty but assume that the friend might not be able to change the color of the paper on Tuesday if the coin lands tails.[16] To be specific, suppose that the color of the paper remains the same on Tuesday as it was on Monday with probability $r$ and changes as in the description of Technicolor Beauty with probability $1-r$. We assume that Sleeping Beauty knows all of this so that $r = 0$ is Technicolor Beauty. Now $f_t(R) = f_t(B) = 1/2$, while $g_t(R,R) = g_t(B,B) = r/2$ and $g_t(R,B) = g_t(B,R) = (1-r)/2$. These satisfy (C2) with $r = (3q-1)/q$, so that $q = 1/(3-r)$. As $r$ runs from 0 to 1, $q$ runs from 1/3 to 1/2. After seeing the colored paper, Sleeping Beauty could be anything from a thirder to a halfer depending on what she believes about how the colored paper is revealed to her. Perhaps there is a new or old principle that could tell Sleeping Beauty what to believe in this example, but probability theory does a pretty good job on its own.

## REFERENCES

Alpern, S. 1988. "Games with Repeated Decisions." *SIAM Journal on Control and Optimization* 26 (2): 468–77.

Bradley, D., and H. Leitgeb. 2006. "When Betting Odds and Credences Come Apart: More Worries for Dutch Book Arguments." *Analysis* 66:119–27.

Briggs, R. 2010. "Putting a Value on Beauty." In *Oxford Studies in Epistemology*, vol. 3, ed. T. Gendler and J. Hawthorne, 1–342. Oxford: Oxford University Press.

Cozic, M. 2011. "Imaging and Sleeping Beauty: A Case for Double-Halfers." *International Journal of Approximate Reasoning* 52:137–43.

de Finetti, B. 1972. *Probability, Induction, and Statistics*. London: Wiley.

[16] At the expense of cluttering the notation, we could assume that Technicolor Beauty has already assessed her credences at an earlier time $s$, as we did in sec. 4. We could then make the same three assumptions we made there about how $s$ and $t$ relate. The simplified presentation here leads to the same conclusion with less complicated notation.

———. 1974. *Theory of Probability.* Vol. 1. New York: Wiley.

De Groot, M. H., and M. J. Schervish. 2012. *Probability and Statistics*. 4th ed. Boston: Addison-Wesley.

Dorr, C. 2002. "Sleeping Beauty: In Defence of Elga." *Analysis* 62:292–96.

Elga, A. 2000. "Self-Locating Belief and the Sleeping Beauty Problem." *Analysis* 60:143–47.

Halpern, J. Y. 2005. "Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems." In *Oxford Studies in Epistemology*, vol. 1, ed. T. Gendler and J. Hawthorne, 111–42. Oxford: Oxford University Press.

Hawley, P. 2013. "Inertia, Optimism, and Beauty." *Nôus* 47 (1): 85–103.

Lewis, D. 2001. "Sleeping Beauty: A Reply to Elga." *Analysis* 61:171–76.

Meacham, C. J. G. 2008. "Sleeping Beauty and the Dynamics of De Se Beliefs." *Philosophical Studies* 138:245–69.

Piccione, M., and A. Rubinstein. 1997. "On the Interpretation of Decision Problems with Imperfect Recall." *Games and Economic Behavior* 20:3–27.

Pust, J. 2012. "Conditionalization and Essentially Indexical Credence." *Journal of Philosophy* 109:295–315.

Rosenthal, J. S. 2009. "A Mathematical Analysis of the Sleeping Beauty Problem." *Mathematical Intelligencer* 31:32–37.

Schervish, M. J. 1995. *Theory of Statistics*. New York: Springer.

Schervish, M. J., T. Seidenfeld, and J. B. Kadane. 2004. "Stopping to Reflect." *Journal of Philosophy* 101:315–22.

Schwarz, W. 2015. "Lost Memories and Useless Coins: Revisiting the Absentminded Driver." *Synthese* 192:3011–36.

Titelbaum, M. 2008. "The Relevance of Self-Locating Beliefs." *Philosophical Review* 117:555–605.

van Fraassen, B. C. 1995. "Belief and the Problem of Ulysses and the Sirens." *Philosophical Studies* 77:7–37.

Wedd, N. 2006. "Some Sleeping Beauty Postings." http://www.maproom.co.uk/sb.html.

Weintraub, R. 2004. "Sleeping Beauty: A Simple Solution." *Analysis* 64:8–10.

White, R. 2006. "The Generalized Sleeping Beauty Problem: A Challenge for Thirders." *Analysis* 66 (2): 114–19.

Yamada, M. 2011. "Laying Sleeping Beauty to Rest." Philpapers. http://philpapers.org/archive/YAMLSB.