Erkenntnis
https://doi.org/10.1007/s10670-019-00212-4

ORIGINAL RESEARCH

Bayesian Beauty

Silvia Milano[1]

Received: 2 May 2019 / Accepted: 23 December 2019
© The Author(s) 2020

Abstract
The Sleeping Beauty problem has attracted considerable attention in the literature as a paradigmatic example of how self-locating uncertainty creates problems for the Bayesian principles of Conditionalization and Reflection. Furthermore, it is also thought to raise serious issues for diachronic Dutch Book arguments. I show that, contrary to what is commonly accepted, it is possible to represent the Sleeping Beauty problem within a standard Bayesian framework. Once the problem is correctly represented, the 'thirder' solution satisfies standard rationality principles, vindicating why it is not vulnerable to diachronic Dutch Book arguments. Moreover, the diachronic Dutch Books against the 'halfer' solutions fail to undermine the standard arguments for Conditionalization. The main upshot that emerges from my discussion is that the disagreement between different solutions does not challenge the applicability of Bayesian reasoning to centered settings, nor the commitment to Conditionalization, but is instead an instance of the familiar problem of choosing the priors.

1 Introduction

Adam Elga (2000) introduced to the philosophical literature what has come to be known as the *Sleeping Beauty problem*:

> 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

I am very grateful to Christian List, Anna Mahtani, Richard Bradley, Aron Vallinder, Bryan Roberts, and two anonymous referees for their comments on earlier drafts of this paper, as well as the audiences at the LSE Choice Group, ELP Konstanz, Edinburgh Graduate Epistemology Conference, CHPS at CU Boulder, the MCMP Logic Colloquium, and the participants in the Formal Epistemology and Decision Theory Workshop at Bristol.

✉ Silvia Milano
silvia.milano@oii.ox.ac.uk

[1] University of Oxford, 1 St. Giles', Oxford OX1 3JS, UK

Published online: 13 February 2020

> are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads? (Elga 2000, p. 143).

Since then, the Sleeping Beauty problem has attracted considerable attention in the literature as a paradigmatic example of how self-locating uncertainty creates problems for the principles of Conditionalization and Reflection. The current consensus is that self-locating uncertainty puts pressure on the standard Bayesian framework (Titelbaum 2016, 2013; Spohn 2017; Schwarz 2011). While there is no general agreement on a solution, those that have been proposed all raise some tension between different Bayesian commitments. The 'thirder' solutions (see Sect. 3.1 below) appear to violate standard Conditionalization and Reflection, and yet are not vulnerable to diachronic Dutch Books (see e.g. Bradley and Leitgeb 2006; Bovens and Rabinowicz 2011). This is at odds with a well known result (Teller (1973), who attributes the idea to D. Lewis). See also Lewis (2010), Skyrms (2009), which establishes that an agent can avoid being vulnerable to diachronic Dutch Books only by planning to update via Conditionalization. On the other hand, it is often accepted that 'halfers' should not bet at the odds that reflect their beliefs in cases like the Sleeping Beauty (Bradley and Leitgeb 2006; Briggs 2009). But this puts pressure on the idea that an agent's credences can generally be interpreted (or operationalised) as the betting odds that the same agent would consider fair, undermining a standard argument for probabilism (de Finetti 1937).

The current state of the literature indicates that the issues raised by the Sleeping Beauty problem constitute a serious threat to the Bayesian approach to subjective probability. The puzzle is compounded by the fact that self-locating beliefs are pervasive, and they are key to modelling ordinary evidence, which normally comes in the form of indexical observations (*e.g.* 'it is raining now', 'this coin toss landed Tails').

Given this background, this paper makes two contributions. Firstly, I give a Bayesian representation of the Sleeping Beauty problem (Sect. 2), and show that the different numerical solutions discussed in the literature can all be derived within this framework, depending on the choice of priors (Sect. 3). Secondly, I show that, despite the appearances, Sleeping Beauty does not call into question the Bayesian commitment to Conditionalization. This is done in two steps. I first defend the thirder solution and address the objection that it violates the principles of Conditionalization and Reflection (Sect. 4). I then show that Diachronic Dutch book arguments against the halfer solution do not provide a counterexample to the results by Lewis and Skyrms (Sect. 5).

Before proceeding further, I should also be clear on the limitations of this paper. I do not aim to give a comprehensive review of the vast literature on the Sleeping Beauty problem. Moreover, I will only be interested in solutions that aim to preserve the validity of standard Bayesian commitments, and specifically the principle of Conditionalization. I will not consider the merits of the (numerous) other existing revisionary approaches that propose alternatives to Conditionalization here,[1] as well as those of solutions that do not rely on a Bayesian framework (such as, for instance, the objectivist one discussed in *Oscar Seminar* 2008).

Instead, my goal is simply to show that it is possible to represent the Sleeping Beauty problem within a Bayesian framework, and that this representation is consistent

[1] For an overview of some of these, see e.g. Titelbaum (2016).

with the Bayesian commitments to Conditionalization and Reflection. Given that the Bayesian framework is widely regarded as a powerful framework for reasoning under uncertainty, and that there are independent reasons for accepting Conditionalization and Reflection, the results I present shift the burden of proof to those that want to argue for a departure from these principles.

2 The Problem

Elga's description of the Sleeping Beauty problem grants the following natural assumptions:

1. The experiment lasts two full days, from the moment Beauty is put to sleep at the end of day 0, to the moment when she is woken up and dismissed at the beginning of day 3.
2. There are two possible outcomes to the experiment: either the coin toss comes up Heads, and Beauty is woken up on day 1, but left to sleep on day 2; or the coin toss comes up Tails, and Beauty is woken up both on day 1 and on day 2. Each outcome has a prior probability that is equal to $\frac{1}{2}$.
3. When she wakes up during the experiment, Beauty does not know which day it is.

From Beauty's standpoint, the task is to determine the probability of Heads, after she wakes up during the experiment.

On each day, Beauty could be in either of two states: she is either awake or she is asleep. Representing an awakening by w and a sleep-through by s, we know from the outset that day 1 involves an awakening, while day 2 may involve either an awakening or a sleep-through, depending on the result of the coin toss. A first characterisation of the state space for the whole experiment is therefore:

$$\Omega = \{ws, ww\}$$

Let $H = \{ws\}$ be the event that the coin toss comes up Heads and $T = \{ww\}$ be the event that the coin toss comes up Tails. By assumption 2, the prior probability of H is the same as the prior probability of T, that is, $P(H) = P(T) = \frac{1}{2}$. In the context of the experiment undergone by Sleeping Beauty, the probabilities of the events H and T are given as priors, as they are fixed by the experimental setup.

During the course of the experiment, Beauty is allowed to make some observations that potentially provide her with side information about the outcome of the experiment. Each observation consists in waking up on a given day, and noting that 'Beauty wakes up today', where 'today' picks out a day $i \in \{1, 2\}$. (Recall that, by assumption 3, Beauty does not know which day it is when she wakes up, and that by assumption 1 each outcome spans over two days. So, Beauty believes that 'today' could be either day 1 or day 2.) How does this observation affect the probability of H? In order to answer this question, we need to represent the observation that 'Beauty wakes up today' (i.e., 'Beauty wakes up on day $i \in \{1, 2\}$') as an event within the same state space as the event H. Let us call this event W. A quick glance at Ω reveals that without some further elaborations, that space is not sufficiently rich to express the event W.

This is because, if the outcome of the experiment is *ws* (if, that is, the result of the coin toss is Heads), it is indeterminate whether Beauty wakes up on day i. To see this, consider how the outcome *ws* consists of an awakening followed by a sleep-through, so for $i = 1$, Beauty wakes up, but for $i = 2$, Beauty does not wake up.

The difficulty with modelling the event W that Beauty wakes up on day i ('today') is due to the fact—expressed by assumption 3—that Beauty does not know which day it is when she wakes up. To represent her uncertainty, we need to refine the outcome space, taking into account that the experiment spans over two days (assumption 1), and that it could be either day 1 or day 2, since the experiment lasts two days regardless of the result of the coin toss.[2] The resulting refined state space is:

$$\Omega' = \{ws1, ws2, ww1, ww2\}$$

Here, the elements of the state space are indexical states, centred on different times.[3] Relative to Ω' we can express the event W as the set of states $\{ws1, ww1, ww2\}$—these are all the states in which Beauty is awake during the experiment. W is false at $ws2$, which corresponds to the state where the result of the coin toss is Heads, it is day 2, and Beauty is sleeping. The event corresponding to a Heads result of the coin toss is $H = \{ws1, ws2\}$, while a Tails result corresponds to $T = \{ww1, ww2\}$. The same prior constraints set by the experimental setup should apply to the refined state space Ω', as they did to the more coarse-grained version of the state space Ω. In particular, by assumption 2, the probabilities assigned to H and T relative to Ω' should be equal.[4] Table 1 summarises the refined state space and the probabilities associated to each outcome, subject to the constraint that $P(H) = P(T) = \frac{1}{2}$.

The parameters α and β in Table 1 both take values in the $[0, 1]$ interval and represent the conditional probability that it is day 1, given that the result of the coin toss is either Heads or Tails; more precisely, $P(D_1|H) = \alpha$ and $P(D_1|T) = \beta$. Expressing the

[2] In a sense, speaking about refining the state space may seem suspicious: after all, $ws1$ and $ws2$ *both* happen (sequentially) if the result of the coin toss is Heads. So, from an atemporal point of view, they are not mutually exclusive. However, here we are not interested in the atemporal viewpoint, but in the temporally located viewpoint that Beauty occupies at the time that she considers the problem. From this temporally located perspective, $ws1$ and $ws2$ are indeed mutually exclusive.

[3] The indexical states within the state space Ω' can be interpreted as *centred worlds*. Using centred worlds to capture the content of self-locating propositions is a standard move in the philosophical literature Lewis (1979). However, the present model is also compatible with other ways of interpreting indexical content, such as for instance as Fregean propositions (see, e.g. Magidor 2015). The limited space of this paper does not allow me to detail this discussion, but I develop this point further elsewhere.

[4] A possible objection that could be moved against this step is the following: assumption 2 constrains the probabilities of the *uncentred* events Heads and Tails, which are perfectly captured in the state space Ω. But once we move to the *centred* state space Ω', it is unclear why the same constraint should apply. In other words, assumption 2 constrains prior uncentred probabilities, but not necessarily the centred probability of the events once we move to the (centred) space Ω'. I do not find this objection convincing, for at least two reasons. First, distinguishing between 'uncentred' and 'centred' reasoning about events such as H and T would make it unclear what are the agent's true prior probabilities, when they could differ. Second, it is not at all clear why we should interpret assumption 2 as applying preferentially to events defined in Ω. Reading it as applying to the events in Ω' is at least as natural, and supports my argument that Ω' is the right state space to capture the relevant events in the Sleeping Beauty case. In light of these, we should preserve correspondence between the prior probabilities assigned to uncentred events relative to both state spaces, as breaking it would have higher costs and is not warranted by anything in the problem description.

Table 1 The Sleeping Beauty problem

Outcomes	(Heads)		(Tails)	
	$ws1$	$ws2$	$ww1$	$ww2$
Probabilities	$\frac{1}{2}\alpha$	$\frac{1}{2}(1-\alpha)$	$\frac{1}{2}\beta$	$\frac{1}{2}(1-\beta)$

probabilities assigned to the elements of the state space in terms of these two parameters is not necessary, as we could instead just focus on the probability that is assigned to each element of the state space in isolation. However, introducing the parameters α and β is convenient simply because it preserves the equal probability ratio between the hypotheses H and T, making sure that their respective prior probabilities remain fixed however we distribute probabilities to the events within this partition.

The conditional probabilities expressed by α and β are not fixed by the experimental setup, so the description of the problem leaves us free, in principle, to set them however seems best. One might worry, however, that the state $ws2$ could never be experienced by Sleeping Beauty, and so should not be part of the space of possibilities. I will address this worry in the next subsection, but note for now that as long as we accept that all the states in Ω' represent logical possibilities, the description of the experimental setup does not give explicit information regarding how Beauty should apportion these probabilities, and therefore we should, at this stage of representing the problem, leave open how she sets the values of both α and β. A discussion of which are the correct or the most plausible values for these parameters is left until later (see Sect. 3 below).

We are now in a position to formally state the problem (which we originally formulated as: *What probability should Beauty assign to Heads, given that she wakes up today?*) in terms of computing the posterior probability of H, given that W is observed, or $P(H|W)$. By Bayes' theorem, we immediately have:

$$P(H|W) = \frac{P(W|H)P(H)}{P(W)}$$

Since H and T partition the outcome space Ω', by the law of total probability we have that:

$$P(W) = P(W|H)P(H) + P(W|T)P(T)$$

Moreover, we know that $P(H) = P(T) = \frac{1}{2}$, $P(W|T) = 1$ (since Beauty wakes up every day if T) and $P(W|H) = \alpha$ (since the probability that Beauty wakes up, given that the coin toss comes up Heads, is equal to the probability that it is day 1 given H). The previous equation simplifies to:

$$P(W) = \frac{1+\alpha}{2} \tag{1}$$

The solution to the Sleeping Beauty problem is therefore given by the equation:

$$P(H|W) = \frac{\alpha}{1+\alpha} \tag{2}$$

Answering Sleeping Beauty's original question therefore depends solely on the value that we assign to parameter α.[5]

2.1 Further Questions

The original Sleeping Beauty problem involves computing the value of H given that W is observed. But once this is done, there are many questions we can still ask. For example, what credence should Beauty assign to the coin toss having come up Heads, if after waking up she were informed that it is day 1? Or, similarly, what would her credence in Heads be if the experimenters told her, after waking her up, that today is the last time she wakes up during the experiment?

Another advantage of the refined state space Ω' is it allows us to model these further questions. The first question (*What is the probability of Heads, if today is day 1?*) can be answered by computing the posterior probability of H, given $D_1 \cap W$ (which coincides with D_1):

$$P(H|D_1) = \frac{\alpha}{\alpha + \beta} \tag{3}$$

Let $L = \{ws1, ww2\}$ be the event that Sleeping Beauty wakes up for the last time. To answer the second question (*what is the probability of* H, *if today is the last time you wake up?*), we just need to compute the probability of H, given the information that L is the case, which is:

$$P(H|L) = \frac{\alpha}{1 + \alpha - \beta} \tag{4}$$

Another question that will be interesting to consider might be posed to Beauty before the experiment actually begins: *Suppose that it is either day 1 or day 2. What is the probability that you wake up?* In order to answer this question, Beauty should effectively state what is the prior probability of W, on the supposition that it is either the first or the second day (that is, given $D_1 \cup D_2$). Since $D_1 \cup D_2 = \Omega'$, the conditional probability of W given $D_1 \cup D_2$ is equal to its unconditional probability:

$$P(W|D_1 \cup D_2) = P(W) \tag{5}$$

Things are a bit different if the question specifies which particular day is to be supposed, *e.g. Suppose it is day 1 (day 2). What is the probability that you are awake?* In this case, since we know that $P(D_1) = \frac{\alpha+\beta}{2}$ and $P(D_2) = \frac{2-\alpha-\beta}{2}$, by a simple calculation we have:

[5] Baratgin and Walliser (2010) propose a solution to the Sleeping Beauty problem that is formally similar to Eq. 2. A difference between their approach and the one presented here is that they assume that the probabilities of $\{ww1\}$ and $\{ww1\}$ are both equal to $\frac{1}{4}$ by the Restricted Principle of Indifference (an assumption that, as I argue, we should dispense with). Because they do not consider parameter β, their approach is also not able to rationalise the double halfer solution (Hawley 2013), which is possible within my approach (see note 12 in Sect. 3.2 below).

$$P(W|D_1) = 1 \qquad P(W|D_2) = \frac{1-\beta}{2-\alpha-\beta} \tag{6}$$

Equation 6 states an interesting result. The probability that Beauty wakes up, given that it is day 1, is equal to 1—just as we would expect, since the experimental setup specifies that Beauty always wakes up on day 1. However, the probability that Beauty wakes up on day 2 does not necessarily equal $\frac{1}{2}$, as this depends on what values α and β take.

2.2 Prior Perspectives

Before we move on, I want to consider two worries that might arise in relation to my construction of the state space Ω' to represent the Sleeping Beauty problem. One worry is that one of the elements of Ω', namely $ws2$, is not compatible with Beauty's evidence at any point in time. Since Beauty is asleep in $ws2$, she could never consciously experience being in that state. In other words, we could say that $ws2$ is a 'blind' state: Beauty could never learn that this is her current state, and any evidence that she might have while she is awake automatically rules out $ws2$. But, then, maybe we should not consider $ws2$ to be a genuine epistemic possibility for Beauty, and it should not be included in the state space Ω' at all. Another, perhaps related, worry is that the state space Ω' is only available to Beauty when she wakes up, but not before, as none of the states in Ω' are compatible with Beauty's evidence before being put to sleep, when she is certain about the current time. This is a serious worry because, if Ω' is not always available to Beauty, then it would be unclear in what sense we can speak of her having prior probabilities relative to Ω', upon which she may conditionalize after learning new evidence during the course of the experiment.

Let us start by addressing the first worry, that $ws2$ may not be a genuine epistemic possibility for Beauty, since she could never learn that she is in that state. First of all, there are some *prima facie* objections to the claim that we could not assign positive probability to 'blind' states that are logically possible. Consider the following case. Suppose you are preparing to leave for a journey, and buying travel insurance brings you to mind the possibility that your flight will crash. Given what you believe about the chances of an accident, you assign a positive probability to the possibility of your own death. This, however, is not a proposition that you are in a position to ever learn. So, in some cases at least, it seems plausible to assign a positive probability to 'blind' states.

Further to this, whether we should include 'blind' states such as $ws2$ in the state space and whether they should be assigned a positive probability are two issues that can be kept separate. In the construction of Ω', and in what follows, I assume that the state space contains all the logically possible states, including those that (like $ws2$) may never represent live epistemic possibilities for Beauty. This still leaves open the option of assigning zero probability to these states. Doing this in the Sleeping Beauty case would support Lewis's solution to the problem, as we will see in Sect. 3.2 below.

The second objection—that the state space Ω' is not available to Beauty before being put to sleep—raises a more general worry for my approach. In one sense, this worry reflects a common intuition: namely, that the *experience* of being in one of the

states in Ω' is only available to Beauty when the state occurs, and it is not available to her at earlier (or later) times. In other words, Beauty could not make the observation that 'today is day 1' before being put to sleep. This observation is not compatible with her circumstances before the experiment starts, when she is certain that the current time is not day 1. But the idea that observations may not be available at all points in time should not be particularly surprising. Suppose that you are considering whether to pack a warm sweater for an upcoming trip: you may look up the weather forecast at your destination, and decide to pack it on the basis of the probability that you come to assign to the proposition that it will be chilly in Vienna at the time of your visit. In other words, you make a decision now, based on the expectation (formed on the basis of the weather forecast) that you will experience a chilly day when you travel to Vienna. The observation 'it is chilly today in Vienna' is not available to you now. But what is relevant for you now is the expectation of experiencing this state in the future. This can be modelled by your assigning a conditional probability to the proposition 'it is chilly', conditional on it being the day when you visit Vienna. If this conditional probability is sufficiently high, based on your current estimates, then you will pack a warm sweater. Similarly for Sleeping Beauty, the observation 'I am awake and today is day 1' would only be available to her on the first day of the experiment—she could not experience these circumstances at any other times. But the expectation of making this experience *is* available (and relevant) to her at other times, specifically before being put to sleep, and it is equal to the conditional probability that she assigns to being awake, given day 1, or $P(W|D_1)$.

A natural objection to what I just said will be that Beauty is certain that it is not day 1 on (what we will call) day 0, before being put to sleep. So, the conditional probability of W, given D_1, is simply undefined, since on day 0 $P(D_1) = 0$. Of course, I do not want to deny that, on day 0, Beauty should be certain that it is not day 1.[6] But notice that, by the same reasoning, it would also seem that we have an argument for the claim that, in the sweater case, the probability that it is chilly in Vienna when I am there is undefined. The way in which we should solve the issue, I propose, is to allow Beauty's prior probabilities to be defined as a conditional probability structure, assigning probabilities to each logically possible state, conditional on different times, as is implicit in my construction of the sample space Ω'. This proposal, versions of which have been discussed in the literature on self-locating beliefs (see Moss 2012; Stalnaker 2008; Meacham 2016; Schwarz 2017; Wenmackers 2017),[7] has the advantage of allowing

[6] To be precise, to model Beauty's beliefs on day 0 we should expand the state space to include a time point relative to day 0. Consider the resulting expanded state space $\Omega'' = \{ws0, ws1, ws2, ww0, ww1, ww2\}$: since $P(D_0|D_0) = 1$, and since D_0 and D_1 are mutually exclusive events, $P(D_1|D_0)$ has to be equal to 0. The expanded state space Ω'' can be used to more accurately model Beauty's credences before being put to sleep, but since the propositions that we are interested in are those relative to her states during the experiment (that is, conditional on it being either day 1 or day 2), there is no loss in using Ω' to solve the equations in Sects. 2–2.1.

[7] See also Horgan (2004) and Neal (2006) for related discussions. The 'generalised Conditionalization' principle defended by Horgan differs from standard Conditionalization by its reliance on what Horgan calls 'preliminary probabilities', which are distinct from prior probabilities. Horgan defends the thirder position, but as I show the halfer position is also consistent with Conditionalization.

for a unified and straightforward treatment of both cases like 'sweater' and Sleeping Beauty.[8]

Having clarified the nature of Beauty's priors, in the next section I turn to the task of deriving a numerical solution within the framework I have presented.

3 Solution

As we've seen in the previous section, the solution to the Sleeping Beauty problem comes down to computing the value of $P(H|W)$ in Eq. 2. To do this, we need to specify what is the value of α, which is not explicitly fixed by the description of the experimental setup. Is there a correct or most plausible assignment of value to α? Moreover, does the value assigned to α match the value assigned to β, or do they differ? (Although the value of β does not matter to the solution of the original Sleeping Beauty problem, it affects the solution to the further questions I described in the previous section.)

There are, I think, two possible ways to go from here. One possibility would be to say that we simply cannot assign any value to α and β, since these are left unspecified by the description of the experimental setup. This would leave us unable to compute the posterior probability of Heads given that Beauty wakes up.[9] Although theoretically coherent, this solution is not very attractive.

A second way to go is to proceed and assign a value to α (and to β, if we have any interest in answering the further questions that Sleeping Beauty might consider, described in Sect. 2.1). If we take this route, it remains to be decided which values are the most appropriate. The most natural assignment—in my view—is $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$. The rationale for this assignment is the following. Since Beauty is uncertain about what day it is, we should conceptualise the day i that she observes as if it had been selected through some sort of randomising mechanism. We don't need to be very specific about the nature of the sampling mechanism that we imagine.[10] The point is simply that, in order to represent Beauty's uncertainty about what day it is, the way in which we model the problem must respect the intuition that both day 1 and day 2 are possible. The parameter α (respectively, β) represents the prior probability that a day randomly selected through this hypothetical mechanism is day 1, given that the

[8] This proposal generalises the one defended by Moss (2012) (who, in turn, expands the analysis of the Sleeping Beauty problem presented in Stalnaker 2008), by giving up the assumption of reducibility of *de se* to *de dicto* beliefs. Due to space constraints, I do not address here some further concerns related to the choice of priors discussed by Meacham (2016), though I pursue these further elsewhere. In any case, these issues do not affect the main point of this paper, that is providing a Bayesian formulation of the Sleeping Beauty problem. This issue is separate from that of the choice of priors.

[9] An additional possibility, that I will not explore here (but that would constitute a possible extension of the representation that I give in this paper), would be to introduce imprecise probabilities and allow a set or an interval of possible values for the posterior (I thank an anonymous referee for drawing my attention to this possibility). See Bovens and Ferreira (2010) for a related discussion, though they do not apply it specifically to the Sleeping Beauty case.

[10] In the terminology of Bovens and Ferreira (2010), this would corresponds to different *protocols* through which Beauty could learn the relevant pieces of information. I resist using this terminology here, as the parameters α and β do not, in themselves, constitute protocols, but they could be seen as consistent with different choices of protocols.

result of the coin toss is Heads (respectively, Tails). In other words, $\alpha = P(D_1|H)$ and $\beta = P(D_1|T)$.

Since the experiment is expected to run over two days, regardless of the result of the coin toss, it makes sense to assume a uniform prior distribution over the two days in both scenarios (whether the coin toss comes up Heads or Tails). On this view, when considering what values to set for α and β, Beauty employs the Principle of Indifference (PI, for short): since she has no reason to believe, assuming that the result of the coin toss is Heads (respectively, Tails) and before anything else is learned, that any one of the two experimental days is more likely to be selected than the other, she should assign them an equal probability, and therefore set $\alpha = \frac{1}{2}$ (respectively, $\beta = \frac{1}{2}$).

PI is a notoriously controversial principle, and I will not attempt to defend it here, although a consideration that might favour it in this particular instance is the fact that points in time are linearly ordered, so each day should intuitively have the same weight. But it is important to note that my application of PI is conceptually distinct from a related line of reasoning that is often applied to the Sleeping Beauty problem, which is known in the literature as the *Restricted* Principle of Indifference (RPI, for short). According to RPI, any two events that take place within the same history (such as the awakenings on day 1 and day 2, in the event of the coin toss coming up Tails) and which would be *subjectively indistinguishable* for an agent experiencing them given some evidence E, should receive equal *posterior* probability after learning that E. This principle places a substantive restriction on the *posterior* probabilities that an agent may come to have upon a learning event, and has come under independent criticism (see Marcoci 2018, Weatherson 2005). In contrast, PI—as I have employed it—gives us a way to set Beauty's *prior* probabilities, which I take to be more in line with the general commitments of Bayesian reasoning (I also consider alternative ways to fix the prior parameters in Sect. 3.2).

3.1 A Numerical Answer

To represent Beauty's uncertainty about which day it is when she wakes up within a probabilistic framework (on the assumption that Beauty reasons as a Bayesian agent) we have refined the simple state space Ω, and then we employed the Principle of Indifference to generate Beauty's prior probabilities for the events that we defined relative to the refined state space Ω'.

If we now plug in the chosen values for α and β to Eqs. 1 and 2 from Sect. 3, we are finally able to compute the desired probabilities:

$$P(W) = \frac{1 + \frac{1}{2}}{2} = \frac{3}{4} \tag{7}$$

$$P(H|W) = \frac{\frac{1}{2}}{1 + \frac{1}{2}} = \frac{1}{3} \tag{8}$$

And moreover, answering the further questions in Sect. 2.1:

$$P(H|D_1) = \frac{1}{2} \tag{9}$$

$$P(H|L) = \frac{1}{2} \tag{10}$$

$$P(W|D_1) = 1 \qquad P(W|D_2) = \frac{1}{2} \tag{11}$$

None of the above answers seems particularly surprising. If Beauty is told that it is day 1, she intuitively is in the same situation as someone who doesn't (yet) know the result of a fair coin toss, and this explains why we have a strong intuition that she should assign a probability of $\frac{1}{2}$ to Heads. Similarly, if Beauty learns that this is the last time she wakes up, but does not know if it is day 1 or day 2, she knows that it is Heads if and only if it is day 1, and Tails if and only if it is day 2. This means that $P(H|L) = P(D_1|L)$, and it is very plausible that both should equal $\frac{1}{2}$. Finally, the probabilities in Eq. 11 are simply in line with the description of the experimental setup.

3.2 Tweaking the Parameters

Although the motivations I gave to support it in Sect. 3 are different, my numerical solution to the Sleeping Beauty problem agrees with the one put forward by Elga (2000),[11] and which is known in the literature as the 'thirder' solution. However, Elga's original paper did not settle the answer to the Sleeping Beauty problem, as attested by a literature growing around it to this day. In this section, I review another prominent alternative solution to the Sleeping Beauty problem, Lewisian halfing. I show how this solution too can be derived within the framework that I gave in Sect. 1, and then critically examine a possible rationale for it.

An early reply to Elga by David Lewis (2001) advocated a different solution to the original problem, which has come to be known as 'halfing'. According to Lewis, Beauty's credence in Heads should not change between the time before she is put to sleep and when she wakes up on day 1, but should stay equal to $\frac{1}{2}$. In other words, for Lewis, both $P(H) = \frac{1}{2}$ *and* $P(H|W) = \frac{1}{2}$. The rationale given by Lewis to defend this 'halfer' solution is that, upon waking up, Beauty does not learn anything new. She was aware all along that she would wake up at least once during the experiment, and therefore an awakening does not give her additional clues about the outcome of the coin toss. Given this consideration, halfers argue that Beauty should not change her credence in Heads upon waking up. Since she knows the coin to be fair, she should maintain a credence of $\frac{1}{2}$ in Heads.

[11] Dorr (2002) also gives an argument in favour of Elga's solution based on a variation of the problem where Sleeping Beauty is awakened every day with her memory temporarily erased, but remembers the Monday awakening after a little while on Tuesday if Heads. Due to space constraints, I do not elaborate on this variation here, but note that Dorr's argument is different from the one presented here, as different solutions cannot be rationalised in his approach.

Lewis's solution, like Elga's, can be derived within the refined state space Ω' that I have given in Sect. 1 (see Table 1). This is because both Lewis and Elga agree that we need to model Beauty's uncertainty regarding what day it is, since this information is relevant to the probability of Heads. They also agree that, if the coin lands Tails, Beauty is equally likely to wake up on day 1 as she is on day 2, that is $\beta = \frac{1}{2}$. However, in order to get Lewis's solution, the conditional probability of day 1 given Heads should be set equal to 1, that is, $\alpha = 1$.[12] When we set the parameters in this way, the prior probability of the event W is:

$$P(W) = \frac{1+1}{2} = 1 \tag{12}$$

And the resulting numerical solution to the Sleeping Beauty problem is therefore:

$$P(H|W) = \frac{1}{1+1} = \frac{1}{2} \tag{13}$$

The halfer solution is known to generate some counter intuitive answers when it comes to the further questions I formulated in Sect. 2.1. These can all be easily derived in the formal model I have given in Sect. 1. One problem for the Lewisian halfers is that the probability that Beauty assigns to Heads appears to increase if, upon awakening, she is informed that it is day 1, as (by Eq. 3):

$$P(H|D_1) = \frac{1}{1+\frac{1}{2}} = \frac{2}{3}$$

This result is clearly puzzling, since Beauty's awakening on day 1 happens independently of the result of the coin toss. Lewis himself acknowledged the puzzling nature of this result, arguing that it constitutes an interesting case of getting evidence 'about the future' (Lewis 2001, p. 175).

In spite of the puzzling result it generates, Lewisian halfing remains a relatively popular solution. This is because it makes a basic appeal to an intuition that is shared by many people, regarding what is the content of Beauty's evidence upon waking up. The idea behind Lewisian halfing is the following: if the coin toss comes up Heads, Beauty can only wake up on day 1. Moreover, Beauty only observes a day if she gets to wake up on that day. Therefore, if Heads, day 1 is experienced with certainty. There is no possible scenario in which Beauty gets to observe day 2, if Heads, and so the prior probability assigned to day 2 given Heads should be 0.

Although it appears intuitive, this motivation for the halfer solution may rest on a misunderstanding of what is the observable event W in the Sleeping Beauty problem. As I explained in Sect. 1, W contains all the outcomes of the refined state space Ω' in which Beauty is awake. When she is put to sleep, Beauty considers it possible that

[12] Another popular solution, that I will not review here, is 'double halfing' (see e.g. Meacham 2008; Cozic 2011). This solution, too, can be derived within my framework by setting $\alpha = 1$ and $\beta = 1$, but has the drawback of violating the third assumption from Sect. 2 (though see Hawley (2013) for an argument in favour of this option). It can also be subjected to a similar DDB argument as the one for the Lewisian halfer that I discuss in Sect. 5 below.

she will not wake up on every day during the experiment. This is because she knows that it is possible that she will sleep through day 2, namely if the result of the coin toss is Heads. So, even though she knows that she will not be consciously aware of it if and when it happens, *day 2 given Heads* is a live possibility at the outset, to which she intuitively should assign a positive prior probability. If the halfer solution were correct, however, then Beauty would be certain that the prior probability of W is equal to 1, since (plugging in $\alpha = 1$ in Eq. 1) $P(W) = \frac{1+1}{2} = 1$. Moreover, puzzlingly, she would also be certain to wake up, conditional on it being day 2, as we can see by solving Eq. 6: $P(W|D_2) = \frac{1-\frac{1}{2}}{2-1-\frac{1}{2}} = 1$. In other words, according to Lewisian halfing, Beauty would be certain that she wakes up on every day during the experiment—even though this clearly is contrary to the description of the experimental setup, which specifies that the prior probability that she wakes up on day 2 is equal to the probability that the coin toss comes up Tails—which, the coin being fair, is in turn equal to $\frac{1}{2}$.

Another reason to believe that the halfer intuition may rest on a misunderstanding is that halfers often justify the value of $\beta = \frac{1}{2}$ on the basis of the same Restricted Principle of Indifference advocated by Elga (but which is not necessary to derive the thirder solution, as I have shown). This means that (at least in the case where the coin toss comes up Tails) Lewisian halfers allow for the possibility that we should think of the day Beauty observes as if it were randomly selected from the set of possible days within a Tails run. But why should this same reasoning not apply to the Heads run, as well? After all, the indifference should reflect the ignorance of which day it is according to Beauty's priors, and not be taken as a way to set her posterior probabilities.

4 Matters of Principle

In this section, I examine how the proposed representation and solution tally with two principles, Conditionalization and Reflection, which concern the relationship between an agent's credences at different times. The fact that the thirder solution seems to violate Conditionalization and Reflection could be taken as indication that these principles are not necessary conditions for diachronic rationality. However, I show that the thirder solution derived within my framework upholds both Conditionalization and Reflection, when appropriately construed.

4.1 Conditionalization

The Sleeping Beauty problem is generally taken to present a challenge to the principle of Conditionalization.[13] Conditionalization is the way in which Bayesian reasoners are expected to update their credences over time, upon learning new pieces of information. It works like this: suppose that at time t_1, you learn a new piece of evidence E (and nothing else). For any event A, the probability that you assign to A at t_1 after learning that E should be equal to the conditional probability you used to assign to $(A|E)$ at the

[13] See Titelbaum (2016), p. 667: 'The current consensus in the self-locating credence literature is that obtaining a general updating scheme for degrees of belief in both centered and uncentred propositions requires us to alter (or at least supplement) conditionalization in some way.'

time t_0 just before learning E. More formally, denoting by P_{t_0} and P_{t_1} your credences at t_0 and t_1, respectively, Conditionalization places the following constraint on how your credences should change between t_0 and t_1, when the only thing that you learn in the interval between these two times is E:

Definition 1 *(Conditionalization).* $P_{t_1}(A) = P_{t_1}(A|E) = P_{t_0}(A|E)$

The question now is: Does Beauty update her credences via Conditionalization upon waking up on day 1? It is often argued that if Beauty is a thirder, then the way her credence in Heads is updated when she wakes up on day 1 is not compatible with Conditionalization. Elga himself makes this point:

> Before being put to sleep, your credence in H was 1/2. […] when you are awakened on [day 1], that credence ought to change to 1/3. 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 [day 1]. (Elga 2000, p. 146)

The upshot, for Elga, is that Conditionalization does not always apply. In cases where an agent receives only centered evidence, his or her credences may change in ways that conflict with Conditionalization.

In light of the analysis I have offered in Sect. 2, we can see how Elga's argument here cannot be right. To say that the change in Beauty's credence in Heads 'is not a result of [her] receiving new information' implies that Beauty is certain that she will receive evidence W, or—more precisely—it implies that the prior probability $P(W)$ equals 1. But, as we have seen, relative to the assignment of values to α and β consistent with the thirder solution, this is not true, because for $\alpha = \frac{1}{2}$ we have that $P(W) = \frac{3}{4} \neq 1$. In other words, if she is a thirder, Beauty is *not* certain that she always learns W. Moreover, as I argued in Sect. 3, learning W *is* relevant to the probability of H.

The last sentence from Elga's quote indicates where the problem lies. When Elga says that Beauty does not receive new information, that is because she is certain of waking up on day 1. This explains why, intuitively, on day 0 she is certain that she will receive evidence W at least once in the future—namely, on day 1. Conditional on her being awake and it being day 1, Beauty's credence in Heads should indeed remain unchanged (as I also argued in Sect. 3), since $P(H)$ is independent of $P(W \cap D_1)$—that is, $P(H|W \cap D_1) = \frac{1}{2} = P(H)$. However, upon waking up, Beauty does *not* learn that $W \cap D_1$. Instead, her evidence is just W, and since $P(H)$ is not independent of $P(W)$, this is relevant information upon which she should update her credences via Conditionalization. My solution allows this, and thus vindicates Conditionalization.

Lewis's halfer solution—contrary to Elga's—does not entail a violation of Conditionalization. Lewis simply starts from the assumption that the evidence W is irrelevant to H, and as we have seen this can be achieved within the representation I have given by setting $\alpha = 1$ and $\beta = \frac{1}{2}$. Given this setting, the prior probability $P(W) = 1$, and so Beauty is indeed certain that she will receive evidence W, which then gives us $P(H|W) = \frac{1}{2} = P(H)$, without any violations of Conditionalization.

Based on this discussion, we can now see that the key difference between the halfer and thirder solutions is the characterisation of the event W. For halfers, W is

certain, and so learning W does not affect the probability of Heads. For thirders, on the contrary, W is not always certain, and therefore learning it affects the probability of Heads, via Conditionalization. Given these results, we can see that once the problem is correctly represented, the solution to the Sleeping Beauty problem does not challenge the validity of Conditionalization as a principle for updating one's credences in the face of newly acquired evidence.

4.2 Reflection

Another rationality principle that appears to be violated in the Sleeping Beauty case is van Fraassen's Reflection principle (Van Fraassen 1984). Suppose that you are a rational Bayesian agent, that you always plan to update your credences via Conditionalization, and you do not expect to suffer any cognitive mishap that would lose you some of your previous evidence. Then let, as before, P_{t_i} denote your credences at a time t_i, and P_{t_j} denote your credences at some later time $t_{j>i}$. If you know, at t_i, that your later credence $P_{t_j}(A)$ in some event A will be equal to some real number $0 \leq p \leq 1$, then, intuitively, your credence at $P_{t_0}(A)$ should match that same value. That is, stated somewhat more formally (see Schervish et al. 2004):

Definition 2 *(Reflection).* $P_{t_i}(A|P_{t_j}(A) = p) = p$.

Clearly, you do not typically know what probability you will assign to an uncertain event in the future. This is because you do not generally know in advance which possible pieces of evidence you will learn in the future, and so you do not know what posterior probability you will assign to A by the time t_j. However, if you were certain that you will receive a particular piece of evidence E (and nothing more) by t_j, which would lead you to update your credence in A (via Conditionalization) to $P_{t_j}(A) = P_{t_i}(A|E) = p$, then it seems reasonable to suppose that you should *already* have the same credence $P_{t_i}(A) = p$ at the earlier time t_i. This is indeed confirmed by the probability calculus: to be certain, at t_i, that you will receive evidence E just means that $P_{t_i}(E) = 1$, and so naturally $P_{t_i}(A) = P_{t_i}(A|E) = p$.

Despite this natural reading, the principle of Reflection has come under considerable critical scrutiny (Mahtani 2016). The Sleeping Beauty problem, in particular, provides one instance where the principle of Reflection appears to be violated. If Beauty is a thirder, and assigns a probability of $\frac{1}{3}$ to Heads upon waking up on day 1, it seems that her prior credences on day 0, before the experiment begins, violate Reflection. This is because she knows, at t_0 = day 0, that she will receive the evidence W at t_1 = day 1. By Reflection, then, it seems that her earlier credence in H at t_0 should be $P_{t_0}(H|P_{t_1}(H) = \frac{1}{3}) = \frac{1}{3}$. Beauty's credence in Heads on day 0, however, is not equal to $\frac{1}{3}$ but to $\frac{1}{2}$, in accordance with what she knows about the experimental setup, which explicitly sets the prior $P(H) = \frac{1}{2}$. So, it seems that either the initial probability of Heads is not $\frac{1}{2}$, or Beauty's credences do not satisfy Reflection. Both alternatives seem very bad: the former flatly contradicts the setup of the problem, while the latter is inconsistent with the probability calculus, under the assumption that Beauty is a rational agent who updates her credences via Conditionalization. What can possibly have gone wrong?

The puzzle, I think, derives from the rather informal statement of Reflection, which has led us to a subtle mis-interpretation. To find the solution to the puzzle, we need only look more closely into the conditions under which Beauty expects to learn W. This is because it is well-known (see Mahtani 2016; Briggs 2009; Schervish et al. 2004) that a rational agent's credences at t_0 should not reflect her credences at a successive time t_1 when the agent at t_0 should not 'trust' her later self at t_1 because either:

1. the agent expects to suffer memory loss between t_0 and t_1, losing some relevant evidence; or
2. it is the case that both (a) at t_1, the agent does not know that it is t_1;[14] and (b) learning that t_1 has arrived would give her new evidence, which would change her credences at t_1.

We can easily verify that conditions 2.a and 2.b are both satisfied in the Sleeping Beauty case, hence why her credence at t_0 is not required by Reflection to match her expected credence at t_1. Given what she knows about the experimental setup, Beauty expects to receive evidence W on day 1, since she is certain that the experimenters wake her up on day 1 irrespective of the coin toss. This consideration is reflected in the prior probability $P(W|D_1) = 1$, as can be easily verified (see Sect. 3). So, when we say that Beauty is certain to learn W (and, as a consequence, to update the probability of Heads to $P(H|W) = \frac{1}{3}$), what we really mean is that Beauty is certain to experience an awakening *on day 1*. But, clearly, t_1 is not a stopping time for Beauty, since upon waking up she does not know what day it is. If at t_1 she was in a position to conditionalize on learning $W \cap D_1$, then indeed Reflection would be satisfied, as expected: $P_{t_0}(H|W \cap D_1) = P_{t_1}(H|W \cap D_1) = \frac{1}{2}$. However, Beauty does not learn $W \cap D_1$ at t_1, but only W. This explains why her credence at t_0 does not reflect her credence at t_1: that is not because she is irrational, or violates Conditionalization in the way she updates her credences between these two times, but because at t_0 she can only be certain that she learns W *given* D_1, but the latter event is not part of her evidence at t_1. It would be incorrect to say that $P_{t_0}(W) = 1$, since $P_{t_0}(W) = \frac{3}{4}$. Therefore, Beauty is not *certain* of W at the earlier time, and she can't reflect on it.

Notice that the same argument for why Reflection does not hold in the Sleeping Beauty case also applies to the halfer solution, since it is also true under this solution that day 1 is not a stopping time for Beauty and, moreover, learning that it is day 1 would lead the halfer to increase her credence in H from $\frac{1}{2}$ to $\frac{2}{3}$. This indicates that the fact that the halfer does not change her credence in H between t_0 and t_1 is not due to an application Reflection.

5 Bets and Odds

As I showed in the previous sections, both the thirder and the halfer solutions can be represented within a Bayesian framework, in a way that is compatible with the principles of Conditionalization and Reflection. A well known result by Lewis, reported by Teller (1973[15] and generalised by Skyrms 2009) shows that a Bayesian agent can

[14] That is, in Schervish et al. (2004)'s terms, t_1 is not a *stopping time*.

[15] See also Lewis (2010).

avoid falling victim to a *Diachronic Dutch Book* (DDB, for short) only if she plans to update her credences via Conditionalization. A DDB, in this context, is a series of bets offered to the agent before and after she learns a given piece of evidence, such that each individual bet is fair, but when taken in combination they guarantee her a sure loss. For illustration, imagine that Betty is a Bayesian who plans to update her credences via Conditionalization. Before a fair die is rolled, we can assume that she might accept a bet X that pays £5 if the die shows a 3, and loses £1 otherwise. The expected value of this bet for Betty now is 0, as she expects to lose £1 with a probability of $\frac{5}{6}$, and win £5 with a probability of $\frac{1}{6}$. She also currently estimates that the probability that she wins the bet, conditional on the die showing an odd number, is equal to $\frac{1}{3}$, which is higher than her current unconditional probability of winning. So, Betty would also be prepared now to accept a conditional bet Y that pays £9 if the die shows a 3, and loses £3 otherwise, all conditional on the die showing an odd number (that is, the bet is void if the die shows an even number, but gives 1:2 odds on 3, conditional on an odd number). Suppose that later, the die is rolled and Betty receives the information that it shows an odd number. At that point, given that she plans to update her credences via Conditionalization, she will be prepared to accept an unconditional bet Z on the die showing a 3, at the same odds as the conditional bet Y. The expected value of Z, later, is the same for Betty as the expected value of Y now—so, if the expected value of Y is non-negative, the same must be true for Z.

Regardless of whether she is a halfer or a thirder, on the solution I have given in Sect. 3 Beauty does not violate Conditionalization. So, by Lewis's result, she should not be Dutch bookable, just as Betty would not be in the example I just gave. Indeed, this solves one side of the puzzle that I identified in Sect. 1: since, on the solution that I have proposed, thirders do not violate Conditionalization (see Sect. 4.1), this vindicates the fact that no genuine DDB against thirders has been found in the literature.[16]

What is, however, surprising is that the halfer solution should be vulnerable to a DDB—completing the puzzle that I identified in Sect. 1. As I have shown in Sect. 3.2, the halfer solution can be derived within my framework with a choice of prior probabilities that assigns a value of 1 to the parameter α. Moreover, since under this prior probability assignment Beauty is certain to learn the evidence W, her credences do not change upon waking up during the experiment, and she does not violate Conditionalization. So, since they satisfy Conditionalization, we would not expect halfers to be vulnerable to DDBs. But consider the following set of bets: On day 0, the bookie offers Beauty a bet on H at even odds, that pays £15 if the result of the coin toss is Tails, and loses £15 if Heads. Since Beauty assigns equal probability to Heads versus Tails, she can accept the bet. Then, every time that Beauty wakes up during the experiment, the bookie offers her another bet, also at even odds, but which pays £10 if Heads, and loses £10 if Tails. Again, assuming that Beauty is a halfer (and so her probability for Heads does not change upon waking up) she can accept these bets as well. But accepting all the bets that she is offered amounts to buying a Dutch Book: no matter

[16] See Hitchcock (2004), Bradley and Leitgeb (2006), Bovens and Rabinowicz (2011).

what the result of the coin toss, by the end of the experiment she is certain to lose £5.[17] Notice that the bookie does not need to have any more information than Beauty does on each awakening to be able to offer her these bets—in fact, we can imagine that the bookie undergoes the same series of awakenings as Beauty. How can we explain this? Is the existence of DDBs against the halfer proof that this solution is incoherent, or does it provide a counterexample to Lewis's result, undermining a standard argument for Conditionalization?

Given that the halfer solution can be derived within the framework I put forward in Sect. 2, we know that this solution is probabilistically coherent, so we can dismiss the first worry. Meanwhile, the second worry merits closer consideration. A crucial feature of the DDB against halfers is that it requires the possibility of an additional bet, that is placed on day 2 only if the result of the coin toss is Tails. What this means is that, in the event of Tails, the book purchased by Beauty contains three bets, which, as we have seen, taken together have a negative value. But the DDB considered in Lewis/Teller (1973) only takes into account an agent's probability assignments at *two* distinct points in time, before and after learning a new piece of evidence E. In other words, let $P_{-}(A)$ be the probability that the agent assigns to an event A at t_0, and $P_{+}(A)$ be the probability that the agent assigns to A at t_i, after learning a piece of evidence E. Lewis shows that, unless $P_{+}(A) = P_{-}(A|E)$, there is a DDB that can be formulated against the agent by offering her some bets at t_0 and at t_i. But even if the agent updates via Conditionalization between t_0 and t_i, Lewis's result does not say anything about bets that the agent could take at some additional time t_j, $j \neq i$. So, technically, the DDB against the halfer does not constitute a counterexample to Lewis's Dutch Book argument for Conditionalization. Is this enough to get standard Conditionalization off the hook?

Maybe not quite. While this vindicates Lewis's and Skyrms's results, the problem remains that planning to update via Conditionalization is not sufficient to ensure that the agent does not commit to a series of bets that guarantee a sure loss. A way out of the impasse, discussed by Briggs (2010), may be to argue that halfers should adopt evidential decision theory. But this option has several drawbacks, among others the fact that it would generate a DDB against the thirder solution (also discussed in Briggs 2010), and that it would give unstable recommendations to the halfer under slight variants of the problem (Conitzer 2015). Perhaps, then, one should bite the bullet and try to argue that only pairwise DDBs are significant means of eliciting an agent's epistemic attitudes, while allowing DDBs to vary in length introduces strategic considerations that can influence Beauty's betting odds, as she needs to coordinate the choice she makes on the bet she is offered with the choices she would make on bets that might be offered to her at other times. Ultimately, the most plausible lesson to draw from the DDB against the halfer may be that satisfying Conditionalization is a necessary, but not sufficient, condition to ensure invulnerability from DDBs of varying length.

[17] If Heads, Beauty loses £15 on the first bet on day 0, and wins £10 on the second bet she takes on day 1, amounting to an overall loss of £5. If Tails, Beauty wins £15 with the bet taken on day 0, but then accepts the second bet both on day 1 and on day 2, each time losing £10, again amounting to an overall loss of £5.

6 Conclusion

The Sleeping Beauty problem has generated a great deal of controversy, as all the main attempts to solve it in the literature appear to violate some or other rationality constraint (Titelbaum 2016). This creates a puzzling state of affairs, as thirder solutions are usually thought to violate the principles of Conditionalization and Reflection, while halfer solutions seem vulnerable to a diachronic Dutch Book.

I have shown that it is possible to model a range of possible solutions to the Sleeping Beauty problem in a Bayesian framework. All the solutions that can be derived within this framework, including the thirder solution that I have defended, respect Conditionalization, thus explaining why they are not subject to Lewis-style DDBs. Moreover, DDBs against the halfer solution do not provide a counterexample to the claim that Conditionalization is a necessary condition for diachronic rationality.

I take the main lessons that can be drawn from my discussion to be the following:

1. Bayesian reasoning can be naturally applied to self-locating uncertainty. The locus of disagreement between different solutions to the Sleeping Beauty problem concerns only the choice of priors.
2. In order to avoid puzzling conclusions, we should be careful to model what is the prior probability of receiving different pieces of evidence.
3. The case of Sleeping Beauty does not present a counterexample to the principles of Conditionalization and Reflection. It does, however, raise interesting questions with respect to whether Conditionalization is sufficient to ensure invulnerability to certain types of DDBs.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Baratgin, J., & Walliser, B. (2010). Sleeping Beauty and the absent-minded driver. *Theory and Decision*, *69*(3), 489–496.

Bovens, L., & Ferreira, J. L. (2010). Monty Hall drives a wedge between Judy Benjamin and the Sleeping Beauty: A reply to Bovens. *Analysis*, *7*(3), 473–481.

Bovens, L., & Rabinowicz, W. (2011). Bets on hats: On Dutch Books against groups, degrees of belief as betting rates, and group-reflection. *Episteme*, *8*(3), 281–300.

Bradley, D., & Leitgeb, H. (2006). When betting odds and credences come apart: More worries for Dutch Book arguments. *Analysis*, *66*(2), 119–127.

Briggs, R. (2009). Distorted reflection. *The Philosophical Review*, *118*(1), 59–85.

Briggs, R. (2010). Putting a value on Beauty. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 3, pp. 3–34). Oxford: Oxford University Press.

Conitzer, V. (2015). A Dutch Book against sleeping beauties who are evidential decision theorists. *Synthese*, *192*(9), 2887–2899.

Cozic, M. (2011). Imaging and Sleeping Beauty: A case for double-halfers. *International Journal of Approximate Reasoning*, *52*(2), 137–143.
de Finetti, B. (1937). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg & H. E. K. Smokler (Eds.), *Studies in subjective probability*. Huntington, NY: Robert E. Kreiger Publishing Co.
Dorr, C. (2002). Sleeping Beauty: In defence of Elga. *Analysis*, *62*(276), 292–296.
Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, *60*(2), 143–147.
Hawley, P. (2013). Inertia, optimism and Beauty. *Noûs*, *47*(1), 85–103.
Hitchcock, C. (2004). Beauty and the bets. *Synthese*, *139*, 405–420.
Horgan, T. (2004). Sleeping Beauty awakened: New odds at the dawn of the new day. *Analysis*, *64*, 10–20.
Lewis, D. K. (1979). Attitudes *de dicto and de se*. *The Philosophical Review*, *88*(4), 513–543.
Lewis, D. K. (2001). Sleeping Beauty: Reply to Elga. *Analysis*, *61*(3), 171–176.
Lewis, D. K. (2010). Why conditionalize? In A. Eagle (Ed.), *Philosophy of probability: Contemporary readings* (pp. 403–407). Abingdon: Routledge.
Magidor, O. (2015). The myth of the *De Se*. *Philosophical Perspectives*, *29*(1), 249–283.
Mahtani, A. (2016). Deference, respect and intesionality. *Philosophical Studies*, *174*, 163–183.
Marcoci, A. (2018). Monty Hall saves Dr. Evil: On Elga's *restricted principle of indifference. Erkenntnis.* https://doi.org/10.1007/s10670-018-0018-4.
Meacham, C. J. G. (2016). Ur-priors, conditionalization, and ur-prior conditionalization. *Ergo: An Open Access Journal of Philosophy*. http://doi.org/10.3998/ergo.12405314.0003.017.
Meacham, C. J. G. (2008). Sleeping Beauty and the dynamics of de se beliefs. *Philosophical Studies*, *138*(2), 245–269.
Moss, S. (2012). Updating as communication. *Philosophy and Phenomenological Research*, *85*(2), 225–248.
Neal, R. M. (2006). Puzzles of anthropic reasoning resolved using full non-indexical conditioning. Manuscript. arXiv:math/0608592.
Oscar Seminar (2008). An objectivist argument for thirdism. *Analysis*, *68*(2), 149–155.
Schervish, M. J., Seidenfeld, T., & Kadane, J. B. (2004). Stopping to reflect. *Journal of Philosophy*, *101*(6), 315–322.
Schwarz, W. (2017). Diachronic norms for self-locating beliefs. *Ergo: An Open Access Journal of Philosophy*. https://doi.org/10.3998/ergo.12405314.0004.025.
Schwarz, W. (2011). Changing minds in a changing world. *Philosophical Studies*, *159*(2), 219–239.
Skyrms, B. (2009). Diachronic coherence and radical probabilism. In M. C. Galavotti (Ed.), *Bruno de Finetti, radical probabilist, volume 8 texts in philosophy* (pp. 85–96). London: College Publications, King's College London.
Spohn, W. (2017). The epistemology and auto-epistemology of temporal self-location and forgetfulness. *Ergo: An Open Access Journal of Philosophy*. https://doi.org/10.3998/ergo.12405314.0004.013.
Stalnaker, R. C. (2008). *Our knowledge of the internal world*. Oxford: Oxford University Press.
Teller, P. (1973). Conditionalization and observation. *Synthese*, *26*, 218–258.
Titelbaum, M. G. (2013). Ten reasons to care about the Sleeping Beauty problem. *Philosophy Compass*, *8*(11), 1003–1017.
Titelbaum, M. G. (2016). Self-locating credences. In A. Hájek & C. Hitchcock (Eds.), *The Oxford handbook of probability and philosophy*. Oxford: Oxford University Press.
Van Fraassen, B. C. (1984). Belief and the will. *The Journal of Philosophy*, *81*(5), 235–256.
Weatherson, B. (2005). Should we respond to Evil with Indifference? *Philosophy and Phenomenological Research*, *70*(3), 613–635.
Wenmackers, S. (2017). The Snow White problem. *Synthese*, *196*, 4137–4153. https://doi.org/10.1007/s11229-017-1647-x.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.