Synthese (2007) 157:59–78
DOI 10.1007/s11229-006-9010-7

ORIGINAL PAPER

Sleeping beauty and self-location: A hybrid model

Nick Bostrom

Received: 8 April 2005 / Accepted: 5 April 2006 / Published online: 1 June 2006

Abstract The Sleeping Beauty problem is test stone for theories about self-locating belief, i.e. theories about how we should reason when data or theories contain indexical information. Opinion on this problem is split between two camps, those who defend the "1/2 view" and those who advocate the "1/3 view". I argue that both these positions are mistaken. Instead, I propose a new "hybrid" model, which avoids the faults of the standard views while retaining their attractive properties. This model *appears* to violate Bayesian conditionalization, but I argue that this is not the case. By paying close attention to the details of conditionalization in contexts where indexical information is relevant, we discover that the hybrid model is in fact consistent with Bayesian kinematics. If the proposed model is correct, there are important lessons for the study of self-location, observation selection theory, and anthropic reasoning.

Keywords Self-locating belief · Probability indexical information · Epistemology sleeping beauty problem · Anthropic principle

Introduction

> *Sleeping Beauty*
>
> On Sunday, Beauty is put to sleep. She is awakened once on Monday, and put to sleep again after being administered a memory-erasing drug that causes her to forget her awakening. A fair coin is tossed. If and only if the coin falls tails, Beauty is awakened again on Tuesday. She knows all this. When she awakes on Monday, what should her credence be that the coin will fall heads?

N. Bostrom (✉)
Faculty of Philosophy, University of Oxford,
Suite 7, Littlegate House,
16/17 St Ebbes st,
Oxford Ox11PT, Uk
e-mail: nick@nickbostrom.com

The Sleeping Beauty problem is a variation of some very similar problems of "imperfect recall" that have been discussed for some time in the game theoretic literature.[1] It was named by Robert Stalnaker, who had learnt about similar cases from Arnold Zuboff. The problem was brought to the attention of the philosophical community through an exchange between Adam Elga, who argued for the answer 1/3 (hereafter the "1/3 view"), and David Lewis, who defended the 1/2 view. The last few years have seen a burst of publications advocating one or the other of the two competing doctrines. To date, neither side seems to have gained a decisive advantage.

Sleeping Beauty is an example of a problem involving self-locating beliefs, i.e., beliefs that an agent, or a temporal part of an agent, might have about its own location. An agent-part that knew exactly which possible world is actual can still be ignorant about its own location in that world. That can happen if the world contains two or more agent-parts whose evidential states are subjectively indistinguishable. These agent-parts would then be unable to determine with certainty their own spatiotemporal location. (Even if Beauty knew that the outcome of the coin toss would be tails, she could not know whether it was currently Monday or Tuesday.) Another way of expressing this is by saying that the agent-parts would be ignorant about which *centered possible world* they are in even though they know which possible world they are in. Yet another formulation is that agent-parts, or "observer-moments", possess all non-indexical information about the world but lack some indexical information. Here, we shall use these expressions interchangeably.

The Sleeping Beauty problem is but one piece of the larger puzzle of how to relate indexical to non-indexical information in our reasoning. If one studies the problem in isolation from this wider context, one risks coming up with answers and principles that do not fit with the other parts of the puzzle. I will first argue that both the 1/3 view and the 1/2 view suffer from such a misfit. That done, I will propose a new "hybrid" model which incorporates aspects of both the 1/3- and the 1/2-view but is identical to neither. The hybrid view overcomes the problems associated with the purebred answers. It also suggests an explanation for why the 1/3- and the 1/2-view have both managed to appeal to their fan bases: they each encapsulate a part of the truth.

The 1/3 view

The 1/3 view is that upon awakening, Beauty should assign credence 1/3 to HEADS. Elga's argument for this view is as follows.[2]

When Beauty wakes up, she knows that she is in one of three situations:

H1 HEADS and it is Monday,
T1 TAILS and it is Monday,
T2 TAILS and it is Tuesday.

[1] Robert Stalnaker gave the problem its name, after having of examples of a similar kind in unpublished work by Arnold Zuboff. Closely related problems have also been discussed in the game theory literature; see volume 20 of the journal *Games and Economic Behavior* (1997).

[2] Elga (2000). For some other defenses of the 1/3 view (see Dorr 2002; Monton 2002; Weintraub 2004).

Given that the Monday and the Tuesday awakening would be evidentially indistinguishable, we have

$$P(\mathrm{T1}) = P(\mathrm{T2}) \quad \text{[By an indifference principle].}$$

If Beauty were to learn that it is Monday, her credence in HEADS should be 1/2 because this is then just the credence that a coin that is about to be tossed, and is known to be fair, will fall heads. Hence,

$$P(\mathrm{H1}|\mathrm{H1} \vee \mathrm{T1}) = 1/2 \quad \text{[By appeal to intuition].}$$

Since $P(\mathrm{H1} \mid \mathrm{H1} \vee \mathrm{T1})$ can be rewritten as $P(\mathrm{H1})/[P(\mathrm{H1}) + P(\mathrm{T1})]$, we thus have

$$P(\mathrm{H1}) = P(\mathrm{T1}) = P(\mathrm{T2}).$$

These credences sum to 1, so it follows that $P(\mathrm{H1}) = 1/3$.

The argument replies on an appeal to intuition in the middle step. One could try to support this step by invoking the *Principal Principle*. This principle says, roughly, that one's credences should accord with one's estimates of objective chances. As formulated by David Lewis, the Principal Principle came with a proviso: one's credences are constrained by one's beliefs about objective chances according to the principle *only if* one does not have relevant "inadmissible" information.[3] Lewis's reason for introducing the proviso was to bracket off cases involving oracles, time travel, and the like, where one might get information about the future that is not mediated by information about current chances. It is possible to maintain, however, that some problems involving self-location also include inadmissible information and should hence be exempted from the Principal Principle's domain of applicability. In light of the positive argument *against* the 1/3 view, which I will present below, we do in fact have strong grounds for regarding Beauty's indexical information as inadmissible.

Given Lewis's own "best-system" analysis of chance, it would be unsurprising to find that indexical information can be sometimes inadmissible. According to the best-system analysis, chances are a kind of concise partial summary of patterns of local, non-modal, occurrent facts. Since chances, defined in this way, do not even purport to summarize indexical information, there is no reason to suppose that all relevant information about future events is always implied by knowledge of the chances—not if we have reason for thinking that indexical information might also be relevant. (This is, in fact, the stance that Lewis adopted in his critique of Elga's argument.[4])

Since our intuitions about what information is admissible in this type of case are no more secure than our direct intuitions about what credence Beauty should assign to HEADS, the Principal Principle fails to settle the Sleeping Beauty controversy. What counts as admissible information in the Sleeping Beauty problem must emerge from its solution—which needs to be independently justified—rather than assumed at the outset.

Another way in which one may attempt to support the middle step in Elga's argument is by invoking Bas van Fraassen's *Reflection Principle*.[5] (The Reflection Principle, in its simplest form, postulates that your credence at a time t is constrained by your credence at a later time t' according to $P_t(X|P_{t'}(X)=x)=x$.) Here we encounter the same problem again: the applicability of the principle to Sleeping Beauty is at least as

[3] (see Lewis 1980, 1994). A precursor to the Principal Principle was formulated by Mellor (1971).

[4] (see Lewis 2001). A similar point was made, independently, in Bostrom (2001).

[5] (see van Fraassen 1984).

problematic as the 1/3 view itself. The Sleeping Beauty problem postulates a breakdown of rationality. Beauty faces the possibility of drug-induced amnesia. Even if we ask about Beauty's credence on Monday, before the drug has been administered, she will not at that point know whether or not her memory has already been tampered with. It is independently known, from other cases, that the Reflection Principle should not be followed where forgetting takes place or is suspected. For example, if I knew that 1 year from now I will have credence 1/2 in the proposition that it rained today, that plainly does not imply that I should assign the same credence now—when I still vividly remember spending the day reading in the garden. The Reflection Principle, therefore, is of no more avail to the 1/3 view than is the Principal Principle.

We are left with the unsupported appeal to intuition in the step of the argument that assumes that $P(\mathrm{H1} \mid \mathrm{H1} \vee \mathrm{T1}) = 1/2$. How are we to evaluate this intuition? One way is to examine what else we are led to accept if we adopt the 1/3 view. If the consequences are unacceptable, we should revise our intuition. Let us, therefore, consider some variations of the original Sleeping Beauty to explore the wider ramifications of the 1/3 view.

If we set $P(\mathrm{H1}) = P(\mathrm{T1}) = P(\mathrm{T2})$, it follows that Beauty should, upon awakening, assign $P(\mathrm{HEADS}) = 1/3$ and $P(\mathrm{TAILS}) = 2/3$. The only relevant difference between HEADS and TAILS is that there would be more awakenings of Beauty on the latter hypothesis. Since the structure of the situation does not depend on the particular numbers involved, we can test our intuitions by considering a more extreme version of the problem.

Extreme Sleeping Beauty

This is like the original problem, except that here, if the coin falls tails, Beauty will be awakened *on a million* subsequent days. As before, she will be given an amnesia drug each time she is put to sleep that makes her forget any previous awakenings. When she awakes on Monday, what should be her credence in HEADS?

By reasoning exactly parallel to that which Elga used to support the 1/3 view in the original version, we obtain

$$P(\mathrm{H1}) = P(\mathrm{T1}) = P(\mathrm{T2}) = P(\mathrm{T3}) = \cdots = P(\mathrm{T1,000,001}).$$

That is, upon awakening, Beauty should assign $P(\mathrm{HEADS}) = 1/1{,}000{,}002$. The result can be generalized: following this line of reasoning, Beauty takes her observation "I am awake now" as evidence in favor of hypotheses that imply that there are many such awakenings of her. The degree of support is proportional to the number of awakenings postulated by the hypotheses. This consequence in Extreme Sleeping Beauty is counterintuitive. It seems like a rather excessive confidence in the proposition that a fair coin, yet to be tossed, will fall tails.

We can bring out even worse implications if we consider cases in which more than one agent is involved. Suppose that a possible world contains several agent-parts $\{a_i\}_{i \in R}$ that are subjectively indistinguishable, that is, these agent-parts are in such similar evidential situations that individual agent-parts cannot tell which one they are in. Elga has argued for an indifference principle stating that one should assign the same credence to each centered proposition of the form "My current agent-part is a_i", for $i \in R$, and this is supposed to hold whether or not the agent-parts in $\{a_i\}_{i \in R}$ all belong

to the same agent or to different agents.[6] This highly restricted indifference principle follows as a special case from a somewhat stronger indifference principle that appears to be needed to make sense of many seemingly legitimate scientific inferences.[7] The principle is thus well supported. But if we use it together with the reasoning in the 1/3 view, we obtain implausible results.

Beauty and Doppelganger

This is like the original Sleeping Beauty problem, except here Beauty is never woken up after being put to sleep on Monday. Instead, if the coin falls tails, another person is created and awoken on Tuesday. This new person will spend her Tuesday wakening in a state that is subjectively indistinguishable from Beauty's Monday state (she will have the same apparent memories and have experiences that feel just the same as Beauty's). When Beauty awakes on Monday, what should be her credence in HEADS?

By Elga's weak indifference principle, Beauty should have the same conditional credence, given TAILS, in her being Beauty and Doppelganger. And by the same reasoning that the 1/3 view relies on in the original version of the problem, Beauty should have the same conditional credence, given MONDAY, in being Beauty as in being Doppelganger. From this we can derive, just as before, that the awakened Beauty should assign an equal credence in three centered propositions:

$$\begin{aligned} &P(\text{“I am Beauty and HEADS”}) \\ &\quad = P(\text{“I am Beauty and TAILS”}) \\ &\quad = P(\text{“I am Doppelganger and TAILS”}). \end{aligned}$$

Since the credence given to these three centered propositions sum to 1, it follows that upon awakening, Beauty should hold $P(\text{HEADS}) = 1/3$.

Now consider the extreme version of Sleeping Beauty and Her Doppelganger, where on tails there will be a million different doppelgangers, each having an awakening on some subsequent day that will be subjectively indistinguishable from Beauty's Monday awakening. It is easy to show, by the same steps as before, that Beauty should upon awaking have credence $P(\text{HEADS}) = 1/1{,}000{,}002$.

The argument does not depend on whether the coin is tossed before the experiment starts or only after Beauty has been put back to sleep on Monday. It is, in fact, irrelevant what the objective chance of HEADS is when Beauty is making her assessment.[8] The relevant factor is Beauty's prior credence in HEADS (relative to her non-indexical background information). For example, Beauty's posterior credence in HEADS would thus be unaffected if the existence of additional awakenings by her doppelgangers were made to depend, not on the outcome of a coin toss, but instead on whether the trillionth digit in the decimal expansion of π is even—provided only that Beauty's prior credence in this proposition is 1/2. We can therefore generalize the foregoing thought experiment:

[6] (see Elga 2004).

[7] (see Bostrom 2002a, b).

[8] If the coin were tossed before she woke up, the chance of HEADS would not be 1/2, but either 1 or 0.

Beauty and Doppelganger (generalized)

Let φ be a (non-indexical) proposition, to which Beauty assigns a prior credence of 1/2. Beauty is never woken up again after being put to sleep on Monday. If φ is true then there will be a total of $N>0$ awakenings of doppelgangers in states that are subjectively indistinguishable from Beauty's Monday awakening.

As before, we get $P(\text{HEADS})=1/(N+2)$. Let us consider what this recommendation amounts too. Each (awake) agent-part, merely by taking into account the indexical fact that "I am currently an agent-part of this kind", should give a greater credence to hypotheses in proportion as they imply that there is a greater number of subjectively indistinguishable agent-parts of that kind. At this point, those familiar with the literature on observation selection theory may notice an disturbing similarity between the reasoning behind this position and the so-called "Self-Indication Assumption". That assumption states that each observer should regard her own existence as evidence supporting hypotheses that imply the existence of a greater total population of observers in the world, the degree of support being proportional to the implied (expected) number of observers.

The Self-Indication Assumption was originally introduced in discussions about the Doomsday argument, as an attempt to neutralize that argument. It turns out, however, that the assumption has implications of its own that are perhaps even more counterintuitive than those of the Doomsday argument. It seems that the following thought experiment, in particular, gives us fairly strong grounds for rejecting the Self-Indication Assumption.

Presumptuous Philosopher

It is the year 2100 and physicists have narrowed down the search for a theory of everything to only two remaining plausible candidate theories, T_1 and T_2 (using considerations from super-duper symmetry). According to T_1 the world is very, very big but finite and there are a total of a trillion trillion observers in the cosmos. According to T_2, the world is very, very, *very* big but finite and there are a trillion trillion trillion observers. The super-duper symmetry considerations are indifferent between these two theories. Physicists are preparing a simple experiment that will falsify one of the theories. Enter the presumptuous philosopher: "Hey guys, it is completely unnecessary for you to do the experiment, because I can already show to you that T_2 is about a trillion times more likely to be true than T_1!" (Whereupon the presumptuous philosopher explains the Self-Indication Assumption.)[9]

By modifying this example slightly, we can substitute the reasoning embodied in the 1/3 view for that of the presumptuous philosopher. To do this, suppose that the two theories that the physicists have come up with differ not only in regard to how many observers there are but also in regard to how many agent-parts there are that are subjectively indistinguishable from your own current one. Elga's 1/3 view can now take the place of the Self-indication Assumption.

It is worth noting that the situation described in this modified version of Presumptuous Philosopher is by no means a farfetched possibility. Contemporary cosmologists face essentially that predicament. They are trying to determine whether the universe

[9] (see Bostrom 2002a, b; 2003)

is finite or infinite. Given the standard Big Bang model and the assumption that spacetime is singly connected, the universe is infinite if and only if it is either open or flat. Whether it is open or flat, or closed, depends on whether the cosmic energy density, Ω, exceeds a certain threshold value. Current measurements indicate that the actual density is very close to the critical value, $\Omega \approx 1$. It is an important open empirical question whether the actual value is above, below, or exactly at the critical level. Measurements are being conducted to obtain a better estimate of the cosmic energy density. If the universe is infinite then with probability one there are an infinite number of agent-moments in states subjectively indistinguishable to your current one.[10] Therefore, if Elga's 1/3 view is correct, we could conclude that we already have "infinitely strong" evidence that the universe is infinite. The consequence that it would be a waste of money to carry out the planned experiments because we can predict the outcome from our armchair (with probability 1), is extremely implausible. Somebody wishing to toe this line should be willing to bet at practically *any* odds on the outcome of these future experiments.[11]

We have seen that the original argument given for the 1/3 view is inconclusive, that neither the Principal Principle nor the Reflection Principle could be successfully invoked to buttress it, and that when we unfold the reasoning embedded in the 1/3 view, we find that highly counterintuitive consequences follow. We have good reason to reject the 1/3 view.[12]

We have not yet considered another argument in favor of the 1/3 view, one that is based on long-run frequency or betting considerations. We will discuss this argument in a later section. But first, let us turn our gaze to the 1/2 view. We shall argue that this view, too, should be rejected.

The 1/2 view

According to the 1/2 view presented by David Lewis, Beauty should upon awakening have credence 1/2 in HEADS, and her conditional credence in HEADS given MONDAY should be 2/3.

$$P(\mathrm{H1}) = 1/2,$$
$$P(\mathrm{H1}|\mathrm{H1} \vee \mathrm{T1}) = 2/3.$$

Suppose that Beauty is informed that it is Monday, and let P_+ be her new credence function after she has obtained this information. Lewis claims that P_+ should be

[10] (see Bostrom 2002b).

[11] Even if during the next 200 years we obtained overwhelming empirical evidence that the universe is finite, we should, on this view, continue to assign credence 1 to the universe being infinite.

[12] Kierland and Monton, in a very recent paper (Kierland and Monton 2005), argue that the case where persons are duplicated (as in Doppelganger) should be treated in a fundamentally different way from cases where awakenings of the same person are duplicated (as in the original Sleeping Beauty). They would therefore likely resist the implications that we have seen follow from Elga's position; and they could do this by rejecting the interpersonal version of the Self-Indication Assumption. I will not discuss this option in detail in this paper, but it is worth noting that the Extreme Sleeping Beauty thought experiment is unaffected by this issue. Furthermore, one can imagine possible situations in which the number of awakenings that an individual should except to experience is strongly correlated with the truth of some physical theory. One may therefore be able to derive counterintuitive implications from this view similar to the ones described above, albeit perhaps only in less empirically realistic situations.

obtained by conditionalizing P on MONDAY. Thus,

$$P_+(\text{HEADS}) = P(\text{HEADS}|\text{MONDAY}) = 2/3.$$

Lewis's argument for this claim is simple: before the experiment, Beauty should assign credence 1/2 to the proposition that a fair coin to be tossed in the future will fall heads. She already knows that she will be awakened. Therefore, when she awakes, she obtains no new relevant information; so her credence in HEADS should remain 1/2.

This argument starts to look peculiar when we compare it to Lewis's explanation of why Beauty *should* increase her credence in HEADS upon being informed that it is Monday:

> Now when Beauty is told during her Monday awakening that it's Monday, ... she is getting evidence–centered evidence–about the future: namely that she is not now in it. That's new evidence: before she was told that it is Monday, she did not yet have it... This new evidence is relevant to HEADS, since it raises her credence in it by 1/6 (i.e. from 1/2 to 2/3).[13]

On this reasoning, it would seem, one could similarly argue that when Beauty awakes on Monday (but before she is informed that it is Monday) she likewise gets relevant evidence–centered evidence–about the future: namely that she *is* now in it. Since it makes no difference whether the coin is tossed before the experiment begins or on Monday evening (a point of agreement between Lewis and Elga), let us suppose the case where the coin is tossed just before Beauty awakens on Monday. If being in "the future" means being in the period after the coin has been tossed, Beauty now has new relevant information about her current location relative to this period (namely, that she is in it now). Lewis is thus committed to the view that one's beliefs about a chance event such as a coin toss *can* be affected by obtaining evidence that is purely about one's own current location. Yet he offers no argument for why only centered evidence that it is Monday, but not centered evidence that one is currently in the "experimental phase" (i.e. that it is either Monday or Tuesday, rather than, say, the preceding Sunday) can be relevant to HEADS. Absent such an argument, his claim that Beauty upon awakening should assign credence 1/2 to HEADS is a completely unsupported assumption, one which those who disagree with the 1/2 view should feel free to reject. Opponents of the 1/2 view can simply insist that Beauty *does* get centered relevant evidence when she finds herself awake in the experimental (Monday or Tuesday) phase. Lewis's argument for the 1/2 view therefore fails.

If we unpack the implications of accepting the 1/2 view, we find that it has implications no less counterintuitive than those of the 1/3 view. Let us begin by considering again the amplified version of the Sleeping Beauty problem.

> *Extreme Sleeping Beauty*
>
> This is like the original problem, except that here, if the coin falls tails, Beauty will be awakened *on a million* subsequent days. As before, she will be given an amnesia drug each time she is put to sleep that makes her forget any previous awakenings. When she awakes on Monday, what should be her credence in HEADS?

The adherent of the 1/2 view will maintain that Beauty, upon awakening, should retain her credence of 1/2 in HEADS, but also that, upon being informed that it is Monday,

[13] (Lewis 2001, p175).

she should become extremely confident in HEADS:

$$P_+(\text{HEADS}) = 1,000,001/1,000,002.$$

This consequence is itself quite implausible. It is, after all, rather gutsy to have credence 0.999999% in the proposition that an unobserved fair coin will fall heads.

We can extract an even more counterintuitive consequence by modifying the example slightly. Instead of using a single coin toss, with a prior probability of heads equal to 1/2, we could stipulate a sequence of 10 independent tosses of the same coin. The prior probability that all of these tosses will come up heads is 2^{-10}, which is less than one in a thousand ($\approx$0.00098%). Suppose that unless the coin comes up heads all ten times, Beauty will not be awakened again after the Monday awakening. If, however, the tossing does yield ten heads, then Beauty will be awakened on a million subsequent days. We can then ask what odds Beauty could reasonably accept if offered to bet on such a sequence of coin tosses.

Beauty the High Roller

> Beauty is awakened on Monday and after having been awake for an hour she is offered a bet. She is told that a fair coin will be tossed ten times. If it lands heads all ten times then Beauty wins $1,000. If it lands tails at least once, then Beauty loses $100,000. But there is a twist: If Beauty wins, the experiment ends at that point. If Beauty loses, she will be put to sleep, given an amnesia drug that causes her to forget her awakening, and then awoken again the next day; and this procedure will be repeated for a total of one million days. (On each of these subsequent awakenings, Beauty will spend an hour in a state of ignorance about what day it is before she is put to sleep. No bet is offered after the initial Monday awakening.)

Beauty awakes on Monday and prudently decides to reject the bet that she is offered. But just as she is about to declare her decision, David Lewis's ghost appears in a puff of smoke. The ghost explains the 1/2 view reasoning and argues that Beauty's credence in the proposition that all ten tosses will come up heads should be very close to unity. In fact, the ghost calculates that, even taking into account the low prior probability of this proposition, Beauty should nevertheless assign it a posterior credence of 99.8% after taking into account that she has just learnt that her current awakening is the initial Monday awakening.[14] The expected value of the gamble to Beauty is therefore positive:

$$\text{EV} \approx 0.998 \times \$1,000 + 0.002 \times (-\$100,000) = \$998 - \$200 = \$798.$$

So according to the ghost's reckoning, Beauty ought to take the bet. But surely it would be crazy for Beauty to follow the ghost's advice.[15] Hence we should reject the 1/2 view.

[14] Let H^* be the proposition that the coin falls heads all ten times. Let M be the centered proposition stating that Beauty's is currently awake on the first Monday. We then have

$$P(H* | M) = \frac{P(M|H*)P(H*)}{P(M|H*)P(H*) + P(M|\neg H*)P(\neg H*)}.$$

With $P(H*) = 2^{-10}$, $P(\neg H*) = 1 - P(H*)$, $P(M|H*) = 1$, and $P(M|\neg H*) = 2/1,000,002$, it easy to check the ghost's calculation.

[15] We assume that Beauty is risk-neutral and that her utility function is linear in money. If she has a diminishing marginal utility of money, or is risk-averse, we can simply adjust the stakes or the number

A hybrid model?

If the 1/3- and the 1/2-view both have unacceptable consequences, how can we build a better model for reasoning with indexical information?

Consider again the 1/3 view. The problem with that view was that it led to a bias in favor of hypotheses entailing that there are many subjectively indistinguishable duplicates of one's current agent-part. When all the hypotheses under consideration agree on the number of such duplicates, the bias does not manifest itself; but it causes trouble in cases like Sleeping Beauty. The obvious way to correct this "many-duplicates" bias is to divide one's credence in being any one particular duplicate with the total number of duplicates.

Consider the following two possible worlds, in which the only agent-parts are those in the Sleeping Beauty experiment. (We shall assume throughout that all "agent-parts" are of equal duration.)

w1:	h1		[The "heads" world],
w2:	t1	t2	[The "tails" world].

As before, H1, T1, and T2 are the centered propositions expressing that one is currently h1, t1, and t2, respectively; HEADS is the proposition that the actual world is w1; and TAILS the proposition that the actual world is w2. The proposal for removing the many-duplicate bias is that we set

$$\begin{aligned} &\mathrm{P}(\mathrm{H1}) = 1/2, \\ &\mathrm{P}(\mathrm{T1}) = \mathrm{P}(\mathrm{T2}) = 1/4. \end{aligned}$$

Of course, we still have

$$\begin{aligned} &\mathrm{P}(\mathrm{H1}|\mathrm{HEADS}) = 1, \\ &\mathrm{P}(\mathrm{T1}|\mathrm{TAILS}) = \mathrm{P}(\mathrm{T2}|\mathrm{TAILS}) = 1/2. \end{aligned}$$

It follows that $P(\mathrm{HEADS}) = P(\mathrm{TAILS}) = 1/2$. Thus, there is no general tendency, upon finding oneself awakened in the experiment, to favor hypotheses implying that there are many such awakenings. This means that we are immune from objections of the "Presumptuous Philosopher"-type. Our new de-biasing postulate implies that the *conditional probability* of HEADS given that it is MONDAY is greater than 50%:

$$P(\mathrm{HEADS}|\mathrm{H1} \vee \mathrm{T1}) = 2/3, \quad \textit{Constraint } P.$$

Now, if the credence function P_+ that Beauty should have upon learning that it is MONDAY were obtained by conditionalizing P on $(\mathrm{H1} \vee \mathrm{T1})$, then we would fall into the trap illustrated in Beauty the High Roller. To avoid falling into this trap, it *seems* as though we require that

$$P_+(\mathrm{HEADS}|\mathrm{H1} \vee \mathrm{T1}) = 1/2, \quad \textit{Constraint } P_+.$$

"Presumptuous Philosopher" and "Beauty the High Roller" form a Scylla and a Charybdis, which we must avoid, and yet it looks like the only way to satisfy the constraints from these thought experiments involves violating Bayesian

(footnote 15 continued)
of awakenings that would occur so that the calculation still favors her taking the gamble, without affecting the basic point of the thought experiment.

conditionalization.[16] I believe, however, that this sacrifice is not necessary. The matter is somewhat subtle.

Suppose that Beauty will be told after awakening on Monday that it is Monday. The situation is then different from the one described above. It must instead be represented as follows:

w1: h1 h1m [The "heads" world],

w2: t1 t1m t2 [The "tails" world].

The situation we are confronting involves *five* possible agent-parts, not three. The added terms, "h1m" and "t1m", denote the agent-parts of Beauty that know that it is Monday (in the heads and the tails world, respectively). Let us retain the P unchanged (i.e. the credence function for Beauty at the times when she is unaware that it is Monday). Now consider more carefully Constraint P_+, which *seemed* like it was a constraint on the P_+ (i.e. the credence function that Beauty has when she knows that it is Monday). Constraint P_+ contains the expression "$\mathrm{H1} \vee \mathrm{T1}$". But this expression does *not* describe what Beauty knows after she has learnt that it is Monday, for at that point she should set:

$$P_+(\mathrm{H1}) = 0, \qquad P_+(\mathrm{T1}) = 0.$$

This is because at that point Beauty knows that her current agent-part is either h1m or t1m. The information she has just obtained is therefore *not* $(\mathrm{H1} \vee \mathrm{T1})$, but rather $(\mathrm{H1M} \vee \mathrm{T1M})$, where H1M is the centered proposition expressed by "My current agent-part is h1m" and T1M is the centered proposition expressed by "My current agent-part is t1m". The correct formulation of Constraint P_+ is therefore as follows:

$$P_+(\mathrm{HEADS}|\mathrm{H1M} \vee \mathrm{T1M}) = 1/2, \quad \textit{Constraint } P_+ \text{ [corrected]}.$$

This correction eliminates the conflict with Constraint P and allows us to avoid violating Bayesian conditionalization. The corrected Constraint P_+ is precisely what we need to save Beauty from ruin in the High Roller thought experiment.

One may still wonder what conditional credence Beauty should assign, before being informed about it being Monday, to HEADS given that she is currently an agent-part that knows that it is Monday:

$$P(\mathrm{HEADS}|\mathrm{H1M} \vee \mathrm{T1M}) = ?$$

However, there is no need to assign a value to this expression. Note that

$$P(\mathrm{HEADS}|\mathrm{H1M} \vee \mathrm{T1M}) = P(\mathrm{HEADS} \,\&\, [\mathrm{H1M} \vee \mathrm{T1M}])/P(\mathrm{H1M} \vee \mathrm{T1M}).$$

Since $P(\mathrm{H1M} \vee \mathrm{T1M}) = 0$, this expression is undefined. And so it should be.[17]

Let us take a step back and consider the point more generally. Whenever an agent receives some evidence E, we could distinguish the earlier agent-part, α^-, that lacked this evidence, and the later agent-part, α^+, which has come to possess it. According to

[16] It has been argued that we should indeed violate conditionalization in the Sleeping Beauty problem (Kierland and Monton 2005). Kierland and Monton argue for the 1/3 answer on grounds which they claim do not lead to the counterintuitive result in the Beauty and Doppelganger. Their position thus diverges significantly from Lewis's 1/2 view.

[17] The model used here presupposes that agent-parts know what their evidence is. This simplifying assumption may be inappropriate in certain cases, but we shall not here discuss how such cases should be modeled.

the reasoning just described, we cannot automatically conclude that the conditional probability $P(X|E$ & "I am currently α^-"), conditionalized on E, yields the correct posterior credence that α^+ should assign to X. Only if

$$P(X|E \,\&\, \text{"I am currently } \alpha^-\text{"}) = P^+(X|E \,\&\, \text{"I am currently } \alpha^+\text{"}),$$

can the kinematics be represented in the simplified form $P^+(X) = P(X|E)$. This standard representation is thus elliptic as it omits some changes in indexical information.

In ordinary cases, such changes in indexical information are irrelevant to the hypotheses being considered and can hence be safely ignored. The standard elliptic representation of Bayesian conditionalization can then be used without danger. In certain special cases, however, such delicate changes in indexical information can be relevant, and it is then crucial to recognize and make explicit the hidden intermediary step. Sleeping Beauty, on the model proposed here, turns out to be just such a special case.

To recapitulate, I have argued that a Bayesian can coherently accept both Constraint P and the corrected Constraint P_+, even though superficially this seems to violate Bayesian conditionalization. The reason why we not only *can* but *should* accept both these constraints was given earlier: to avoid the counterintuitive consequences that follow if either of these constraints is violated, as shown by the "Presumptuous Philosopher" and the "Beauty the High Roller" thought experiments.

The long-run frequency argument

One other important argument for the 1/3 view needs to be examined as it might be thought to pose a problem for the hybrid model. The discussion of this argument will also serve to further elucidate how the proposed model works.

A proponent of the 1/3 view could argue that Sleeping Beauty awakened ought to have credence 1/3 in HEADS because if the experiment were repeated many times, then approximately 1/3 of all her awakenings would be heads-awakenings (and 2/3 would be tails-awakenings). In the infinite limit, this ratio would, with probability 1, be approached arbitrarily closely. This argument could be buttressed by introducing betting considerations. In the infinite limit, Beauty would *have* to assign credence 1/3 to HEADS, else she would be guaranteed a loss if she put her money where her mouth is.

For a betting argument to have any bite, the hypothesized bookie must have the same information as Beauty. If a bet were only offered on the Monday awakening in each run of the experiment, and if both the bookie and Beauty knew this, then Beauty could infer from the fact that she was offered a bet that it was Monday. According to the hybrid model, she should then assign credence 1/2 to the proposition that the coin will fall heads in that trial. This will match the long-run frequency of bets that she will win, so in this case betting considerations pose no problem.

The betting argument therefore requires that Beauty be offered a bet each time she is awakened. Then she cannot infer what day it is from the fact that she is being offered a bet. In this case, in the long run, Beauty would be expected to lose 2/3 of her bets if she consistently bet on heads. How does this square with the prescription of the hybrid model that Beauty, upon awakening (but before learning which day it is) assigns credence 1/2 to heads?

One possible response to this argument is to deny that betting considerations provide a valid guide to credence assignment in the present case. Since there would be a different number of bets placed depending on how the coin fell, one might regard the test as unfair.[18] In support of this response one may note that the *presumptuous philosopher* would also be vindicated if we assumed that an agent-part's credence should be determined by the betting-odds at which the expected net gains and losses of the collective of all his duplicate agent-parts would be zero. Since there would be a trillion times more duplicates of the agent-part if theory T_2 is true then if T_1 is true, each agent-part would have to assign a trillion times greater odds to T_2 than to T_1 in order for the expected value of all the bets made by the collective of agent-parts to be zero. And yet we argued that it seems wrong for an agent-part to assign a trillion times greater credence to T_2 than to T_1.

However, this response does not address the case of the repeated Sleeping Beauty problem. For in this case, in the infinite limit, there is no uncertainty about the total proportion of awakenings in tails- and heads-runs of the experiment. Beauty knows that (with probability one) there will actually be two times as many awakenings in tails-trials as in heads-trials. In this case, therefore, betting considerations unambiguously suggest that Beauty upon awakening should assign the 2/3 credence to tails. Here one could not justify a divergence of credence assignment from betting odds by saying that there would be a different number of bets placed depending on which of the hypotheses under consideration is true, because the total number of bets placed is not (significantly) variable when Beauty is put through a large number of repetitions of the experiment.

There is, consequently, strong reason for recommending that Beauty assign credence 1/3 to heads when she knows that the experiment will be repeated very many times. This, however, is not an objection to the hybrid model proposed above. The hybrid model, as we shall now see, implies the very same credence assignment as the betting considerations suggest. Betting considerations, far from being an embarrassment to the hybrid model, actually agree with its implications and support it.

Up until this section, our discussion has focused on (variations of) the single-shot Sleeping Beauty problem, where there are no repetitions of the experiment. This is the simplest case: the world contains no other relevant agent-parts than those existing within a single implementation of the Sleeping Beauty experiment. Let us now apply the hybrid model to the situation that arises if the experiment is repeated many times. But first, as an intermediary step, consider the following case.

Three Thousand Weeks (non-random)

> Beauty lives for three thousand weeks. On odd-numbered weeks she is awakened once, on Mondays. On even-numbered weeks she is awakened twice, on Mondays and Tuesdays. After each awakening she is given an amnesia drug that causes her to forget her previous awakenings. Beauty knows all this.

The hybrid model that I propose implies that in this case, Beauty should have credence 1/3 in the centered proposition "My current awakening is taking place in an odd-numbered week" (or "ODD" for short). This is because Beauty, when she wakes up, knows that ODD is true for one third of all the agent-parts that are in the same subjective evidential state as her current agent-part. (The credence assignment

[18] (see also Arntzenius 2002).

follows from the very weak indifference principle which Lewis, Elga, and I all accept.) Crucially, these agent-parts are all *actual* agent-parts, as opposed merely possible ones. We thus have

$$P(\text{ODD}) = 1/3.$$

Further, it is easy to show that

$$P(\text{ODD}|\text{MONDAY}) = 1/2.$$

If we suppose that every Monday, just before being put to sleep, Beauty is told that it is Monday, we also have

$$P_+(\text{ODD}|\text{MONDAY}) = 1/2.$$

Note that, for the reasons explained earlier, "MONDAY" denotes a different centered proposition in each of these two conditional credence expressions. (In the first expression, "MONDAY" refers to a proposition that is centered on an agent-part that does not know that it is Monday; in the second expression, specifying Constraint P_+, "MONDAY" refers to a proposition centered on an agent-part that *does* know that it is Monday.) In the present case, however, the conditional credences work out the same. The hybrid view therefore coincides with the 1/3 view in this example.

The key difference between the original Sleeping Beauty problem and 3,000 weeks and is that in the latter case—but not in the former—there are twice as many *actual* awakenings of one type as of the other. This means that in 3,000 weeks, the prior credence in ODD, before Beauty learns that it is Monday, is unaffected by the correction we made to eliminate the bias in favor of hypotheses entailing the existence of more duplicates. (Such a bias would result from applying the indifference principle to a class of agent-parts that included merely possible as well as actual agent-parts.) In 3,000 weeks, ODD is true for one-third of the agent-parts that are ignorant about whether it is Monday, and for one half of the agent-parts who know that it is Monday; correspondingly, the credence in ODD is 1/3 for the first type of agent-part and 1/2 for the second type.[19]

Let us now apply this analysis to a more straightforwardly repeated version of the original Sleeping Beauty problem:

The N-fold Sleeping Beauty Problem

This is like the original Sleeping Beauty problem repeated N times on consecutive weeks. Beauty knows that the experiment is repeated N times, but she is unable to determine which run of the experiment she is currently in.

For $N = 1$, this reduces to the original Sleeping Beauty problem, and Beauty's credence in HEADS should be 1/2, both before and after learning that it is Monday. If N is some large number, such as $N = 3,000$, then the case approximates 3,000 weeks, and Beauty's credence in HEADS should be approximately 1/3 before learning that it is Monday, and 1/2 after being told that it is Monday. ("HEADS" here stands for "My current awakening is in one of the trials where the coin fell heads".) The larger N

[19] We say that ODD "is true for" an agent-part if the centered proposition which that agent-part would express by saying "ODD" is true. When writing down a symbol like "ODD", we need to be careful about whether we take this to refer to a specific centered proposition or to a function that yields a centered proposition when given an agent-part as an argument. In the text, the context should make it clear what is intended in each case.

is, the more exact will the approximation be. The credences in the 3,000-fold Sleeping Beauty Problem are not *exactly* equal to those in 3,000 weeks because the total number of awakenings is not strictly fixed. There is, however, a very high chance that there will be roughly 3,000 tails-awakenings and 1,500 heads-awakenings in the 3,000-fold Sleeping Beauty Problem, so it closely approximates 3,000 weeks.

Illustration: The hybrid model to the $N = 2$ case

It may be instructive to calculate the exact credences for the $N = 2$ case. There are four possible outcomes of the coin tosses: heads-heads, heads–tails, tails–heads, and tails–tails. We can represent these four possibilities along with the possible agent-parts they would realize as follows:

Week 1				Week 2	
w1:	h1			h2	
w2:	h3			t1	t2
w3:	t3	t4		h4	
w4:	t5	t6		t7	t8

Each of these four possibilities has an equal chance of occurring ($p = 1/4$). Since each of these agent-parts are in the same evidential situation, Beauty's conditional credence, given one of the four possibilities, is divided equally between the agent-parts that that possibility would realize. Hence, her unconditional credence in being any particular possible agent-part is obtained by multiplying this conditional credence with her prior credence in the possibility in question (i.e. 1/4). Thus, we get the following assignment of credence to the centered propositions that she is currently a particular agent-part:

Week 1				Week 2	
w1:	**1/8**			**1/8**	
w2:	**1/12**			1/12	1/12
w3:	1/12	1/12		**1/12**	
w4:	1/16	1/16		1/16	1/16

We obtain P(HEADS) by summing the credences of the centered propositions that imply HEADS (indicated with boldface):

$$P(\text{HEADS}) = 1/8 + 1/8 + 1/12 + 1/12 = 5/12.$$

Since P(HEADS | MONDAY)= P(HEADS & MONDAY)/P(MONDAY), we likewise get

$$P(\text{HEADS}|\text{MONDAY}) = (5/12)/(17/24) = 10/17.$$

This, however, is not the credence that Beauty should assign to HEADS if she were told that it is Monday. For the same reasons as noted above in the discussion of the original (one fold) Sleeping Beauty problem, the relevant quantity is instead P_+(HEADS | MONDAY). To determine this quantity, we again represent four possibilities, but these now include agent-parts that know that it is Monday (these are the agent-parts in the middle columns, whose names end with the letter 'm'):

	Week 1				Week 2		
w1:	h1	**h1m**		|	h2	**h2m**	
w2:	h3	**h3m**		|	t1	t1m	t2
w3:	t3	t2m	t4	|	h4	**h4m**	
w4:	t5	t3m	t6	|	t7	t4m	t8

Since the number of agent-moments that know that it is Monday is the same in all four possibilities (i.e., two in each case), each of these agent-parts (who are in the same evidential situation) should assign the same credence to being a particular one of these agent-parts, namely $(1/4)(1/2) = 1/8$, and they should assign zero credence to being some other agent-part. Thus:

	Week 1				Week 2		
w1:	0	**1/8**		|	0	**1/8**	
w2:	0	**1/8**		|	0	1/8	0
w3:	0	1/8	0	|	0	**1/8**	
w4:	0	1/8	0	|	0	1/8	0

To obtain $P_+(\text{HEADS} \mid \text{MONDAY})$, we sum the credences of the centered propositions that imply both HEADS and MONDAY (indicated in boldface), and divide this by the sum of the credences that imply MONDAY:

$$P_+(\text{HEADS}|\text{MONDAY}) = (1/8 + 1/8 + 1/8 + 1/8)/1 = 1/2.$$

The hybrid model thus implies that when Beauty learns that it is Monday, she should have credence 1/2 in HEADS. This is so both in the original one-shot version of the Sleeping Beauty problem and in the repeated ("N-fold") versions where $N \geq 1$.

Discussion

We have argued that the standard arguments for the standard positions on the Sleeping Beauty problem, the 1/2 view and the 1/3 view, are, if not directly question-begging then at least inconclusive in that they rely on eminently deniable premises. To evaluate the standard positions, therefore, we need to seek for further constraints. We presented two such constraints in the form of two thought experiments. The Presumptuous Philosopher thought experiment, in a version adapted for application to the Sleeping Beauty case, strongly suggests that the 1/3 view is wrong. The Beauty the High Roller thought experiment strongly suggests that the 1/2 view is wrong. On these grounds, we concluded that *both* the standard models for reasoning about self-location are unacceptable.

In the second, constructive part of the paper we proposed a new model. This model seeks to combine the most attractive features of the 1/3- and the 1/2-view, so we termed it the hybrid model. It implies that Beauty should not take the fact that she is currently awake as evidence that there are large numbers of awakenings. But it also implies that when Beauty discovers that it is currently Monday, she should not take this as evidence against the hypothesis that there will be many more awakenings in the future.

If the hybrid model is correct, it might explain the fact that both the 1/3- and the 1/2-views have some intuitive appeal. According to the hybrid model, *both* these views

get something right. The 1/3 view is right that Beauty's *posterior* credence in HEADS after being informed that it is Monday should be one-half. The 1/2 view is right that Beauty's *prior* credence in HEADS, after awakening but before learning that it is Monday, should be one-half.

The 1/3 view is also right that in the version of the Sleeping Beauty where the experiment is repeated a large number of times, Beauty should (in the infinite limit), upon awakening, assign a prior credence of 1/3 to the centered proposition that the coin fell heads in that particular trial. The hybrid view distinguishes between actual and merely possible agent-parts. In the N-fold Sleeping Beauty problem, for $N \gg 1$, it is (almost certainly) the case that approximately one-third of all actual agent-parts of Beauty are in trials in which the coin fell heads, and the total number of awakenings is (with high probability) approximately determined in advance. By contrast, in the one fold version, it is *not* the case that one-third of all actual agent-parts of Beauty are in a heads-trial. There, either all are, or none. Moreover, in the one fold version, the total number of awakenings is strongly correlated with which hypothesis, HEADS or TAILS, is true. The hybrid model corrects for the bias in favor of many awakenings that is inherent in the 1/3 view. (In cases where N is small but larger than 1, the hybrid model gives a prior credence that is intermediate between the that of the 1/3 view and the 1/2 view, thus avoiding any sharp discontinuity. In general, for $N \geq 1$, we have $1/3 \leq P(\text{HEADS}) \leq 1/2$.)

The main concern about the hybrid model is that it appears to violate Bayesian conditionalization. I argued, however, that this violation is merely apparent. If we pay close attention to the changing indexical information available to different agent-segments, we find that the model does not violate Bayesian conditionalization. A lesson here is that while indexical evidence is irrelevant and can be ignored in most ordinary cases of Bayesian updating, there are special cases–Sleeping Beauty included–where such evidence is relevant. In these special cases, certain implicit assumptions in the common way of applying Bayesian conditionalization are false.

In closing, I will address one challenge that could be directed at the hybrid model.[20] If Beauty follows this model and agrees to betting odds matching her credence function, she can be Dutch-booked.

The Beauty and the Bookie

This is like the original one-shot version but with an added Bookie, who is put to sleep at the same time as Beauty and given the same amnesia drug. (We put the Bookie through this procedure to make sure that he does not have any relevant information that Beauty lacks.) Upon awakening, on both Monday and Tuesday, before either knows what day it is, the Bookie offers Beauty the following bet:

Beauty gets \$10 if HEADS and MONDAY.
Beauty pays \$20 if TAILS and MONDAY.
(If TUESDAY, then no money changes hands.)

[20] I'm grateful here to one anonymous referee. A similar Dutch-book argument has recently been advanced in Hitchcock (2004).

On Monday, after both the Bookie and Beauty have been informed that it is Monday, the Bookie offers Beauty a further bet:

Beauty gets \$15 if TAILS.

Beauty pays \$15 if HEADS.

If Beauty accepts these bets, she will emerge \$5 poorer.

Since Beauty is able to anticipate the result of accepting all the bets, it is clear that she should not do so.

Following the hybrid model, Beauty should have no objection to accepting the second Monday bet. The hybrid model implies that $P_+(\text{HEADS}|\ \text{MONDAY}) = P_+(\text{TAILS} \mid \text{MONDAY}) = 1/2$. Being offered a single straightforward bet on HEADS at even odds, knowing that it is Monday, she has no reason to refuse it.

It is the other set of bets that she should reject. The hybrid model implies that Beauty, before learning that it is Monday, assigns $P(\text{HEADS} \mid \text{MONDAY}) = 2/3$. This appears to justify her accepting the bookie's first offer. But here the situation is more complicated. Since neither party knows whether it is Monday, the Bookie cannot offer this bet only on Monday. He must offer it on both awakenings. This means that the total number of bets will vary depending on how the coin falls: if heads, the first type of bet is offered only once; but if tails, it is offered twice. Moreover, we may assume that Beauty will either accept it on both occasions or reject it on both occasions, as she has no effective way of telling which occasion she is currently encountering.[21] So Beauty knows that she would be accepting two bets if TAILS and one bet if HEADS.

Now, we already know from other examples that when the number of bets depends on whether the proposition betted on is true, then the fair betting odds can diverge from the correct credence assignment. For instance, suppose you assign credence 9/10 to the proposition that the trillionth digit in the decimal expansion of π is some number other than 7. A man from the city wants to bet against you: he says he has a gut feeling that the digit is number 7, and he offers you even odds—a dollar for a dollar. Seems fine, but there is a catch: if the digit is number 7, then you will have to repeat exactly the same bet with him one hundred times; otherwise there will just be one bet. If this proviso is specified in the contract, the *real* bet that is being offered you is one where you get \$1 if the digit is not 7 and you lose \$100 if it is 7. That you should reject *this* bet is quite unproblematic and does not in any way undermine your original assessment that the probability of the trillionth digit being 7 is 1/10.

A similar situation can arise in a more subtle way. We can construct a scenario where, even though no "catch" is explicitly part of the contract, you nevertheless know that you will be put in a position where you will end up betting a hundred times if you are wrong but only one time if you are right. This could happen, e.g. if there is a machine that will determine the correct answer and then, on the basis of what this answer is, will decide whether to repeatedly administer an amnesia drug to you that makes you forget whether you have already betted. The machine could do this in such a way that you end up making a larger number of bets if you are wrong. If you believe that you are facing a situation of this kind, you should take corrective action to limit the distortive effects of the memory erasure on your decision-making. In particular, you may decide to reject bets that seem fair to you and that may have been perfectly acceptable in the absence of the forced irrationality constraint.

[21] If Beauty could opt for a mixed strategy, she could decide to accept the bet at a given occasion with a certain probability. This would complicate the argument but would not affect the conclusion.

Let us return to the case of Beauty and the Bookie. Beauty knows that she faces the risk of having her memory erased and thus of becoming irrational. (Memory erasure entails a form of irrationality.) For reasons such as those described above, Beauty may therefore reject the bookie's first set of bets as a form of damage control to minimize the impact of the failures of rationality from which she knows she is at risk. If the deviation of her optimal betting odds from her credence assignment can be justified on these grounds, then she can use the hybrid model and still avoid being Dutch booked.

It is interesting that in Beauty and the Bookie, Beauty's betting odds should deviate from her credence assignment even though the bet that might be placed on Tuesday would not result in any money changing hands. In a sense, the bet that Beauty and the bookie would agree to on Tuesday is void. Nevertheless, it is essential that this bet is included in the example. The bookie is unable to pursue the policy of only offering bets on Monday since he does not know which day it is when he wakes up. If we changed the example so that the bookie knew that is was Monday immediately upon awakening, then Beauty and the bookie would no longer have the same relevant information, and the Dutch book argument would fail. If instead we changed the example so that Beauty as well as the bookie knew that it was Monday immediately upon awakening, then Beauty's credence in HEADS & MONDAY would be 1/2 throughout Monday, so again she would avoid a Dutch book.[22]

In conclusion, the hybrid model combines the comely aspects of the 1/2 view and the 1/3 view while avoiding their faults. The main concern with the hybrid model is that it may appear to violate Bayesian conditionalization. I have presented (tentative) arguments suggesting that the violation is merely apparent. At any rate, one might hope that having a third contender for how Beauty should reason will help stimulate new ideas in the study of self-location.[23]

References

Arntzenius, F. (2002). Reflections on Sleeping Beauty. *Analysis, 62(1)*, 53–62.
Bostrom, N. (2001). The Doomsday argument, Adam & Eve, UN++, and Quantum Joe. *Synthese, 127(3)*, 359–387.
Bostrom, N. (2002a). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge: New York.
Bostrom, N. (2002b). Self-locating belief in big worlds: cosmology's missing link to observation. *Journal of Philosophy, 99(12)*, 607–623.
Bostrom, N. (2003). The mysteries of self-locating belief and anthropic reasoning. *Harvard Review of Philosophy, 11*, 59–74.
Dorr, C. (2002). Sleeping beauty: in defense of Elga. *Analysis, 62(4)*, 292–296.
Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis, 60(2)*, 143–147.
Elga, A. (2004). Defeating Dr. Evil with self-locating belief. *Philosophy and Phenomenological Research, 69(2)*.
Hitchcock, C. (2004). Beauty and the bets. *Synthese, 139*, 405–420.

[22] If Beauty would know on Monday that it is Monday, then she would also be able to infer on Tuesday—from the fact that she does not know then that it is Monday—that it is Tuesday. So she would always know what day it is. (We assume that Beauty always know the general setup of the experiment she is in.)

[23] For comments and discussions, I am grateful to Adam Elga, Bradley Monton, Brian Kierland, Simon Saunders, and anonymous referees.

Kierland, B., & Monton, B. (2005). Minimizing inaccuracy for self-locating belief. *Philosophy and Phenomenological Research, 70(2)*, 384–395.
Lewis, D. (1980). A subjectivist guide to objective chance. In R.C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*. University of California Press: Berkeley, p. 2.
Lewis, D. (1994). Humean supervenience debugged. *Mind, 103(412)*, 473–490.
Lewis, D. (2001). Sleeping beauty: reply to Elga. *Analysis, 61(271)*, 171–176.
Mellor, H. (1971). *The Matter of Chance*. Cambridge University Press: Cambridge.
Monton, B. (2002). Sleeping beauty and the forgetful Bayesian. *Analysis, 62(1)*, 47–53.
van Fraassen, B. (1984). Belief and the will. *Journal of Philosophy, 81*, 235–256.
Weintraub, R. (2004). Sleeping beauty: a simple solution. *Analysis, 64(1)*, 8–10.