How to Put Self-Locating Information in its Place

Paul Bartha
Department of Philosophy
University of British Columbia

How can self-locating propositions be integrated into normal patterns of belief revision? Puzzles such as *Sleeping Beauty* seem to show that such propositions lead to violation of ordinary principles for reasoning with subjective probability, such as Conditionalization and Reflection. I show that sophisticated forms of Conditionalization and Reflection are not only compatible with self-locating propositions, but also indispensable in understanding how they can function as evidence in *Sleeping Beauty* and similar cases.

1. Introduction. A proposition is *self-locating* if it pertains to my location (hereafter, temporal location) in the world. "Today is Monday" is *entirely self-locating*: it adds nothing to a complete description of the world beyond my current location. A proposition is *substantial* if it tells me something about the world. "Julius Caesar died in 44 B.C." is *entirely substantial*: it tells me nothing about my current location. Can a proposition be both self-locating and substantial? If so, how can learning such a proposition best be integrated into normal patterns of belief revision?

In recent years, several puzzles related to these questions have become familiar to philosophers, the most notorious being Elga's *Sleeping Beauty* problem. Elga (2000) believes in substantial self-locating propositions and thinks that they furnish counterexamples to venerable Bayesian principles for reasoning with subjective probabilities. Other philosophers, such as Lewis (2001), are reluctant to give up those principles and resist the idea that we can learn anything about the world from self-locating information.

This paper offers a framework for thinking about self-locating propositions and adjudicating some of the disputes. I concentrate on *Sleeping Beauty* and a few variants. I argue that we can accommodate substantial self-locating information with very little modification of traditional Bayesian principles. Rather than overturning those principles, puzzles such as *Sleeping Beauty* confirm their flexibility and viability.

2. The *Sleeping Beauty* problem.

Sleeping Beauty (A):

> Between Sunday and Wednesday, an experiment is conducted. Beauty goes into a deep, drug-induced sleep on Sunday night. A fair coin is tossed. If the result is *Heads*, Beauty is awakened only at noon on Monday. If the result is *Tails*, she is awakened at noon both on Monday and on Tuesday. After each awakening, the drug is administered and Beauty goes back to sleep. The drug causes limited amnesia. When Beauty awakens on Monday or Tuesday, she has no memory of anything after Sunday night. In particular, she has no recollection of previous awakenings (if any), and no idea whether it is Monday or Tuesday. On Wednesday, the experiment ends and Beauty goes home.

Figure 1 depicts the schedule of awakenings.

	Mon	*Tues*
Heads	Awake	Not awake
Tails	Awake	Awake

Fig. 1: Sleeping Beauty

The *Sleeping Beauty* problem is this: upon being awakened but not told which day it is, what should be Beauty's degree of belief (or subjective probability) that the result of the coin toss was *Heads*?

Everyone agrees that on Sunday, Beauty's subjective probability for *Heads* ought to be ½, since the coin is fair and Beauty should follow the 'Principal Principle' (Lewis 1980). Opinion is split about what should happen to Beauty's probability for *Heads* when she is awakened. 'Halfers' maintain that it stays ½. 'Thirders' think that it should drop to 1/3. By the end of the paper, I conclude that the correct answer is 'slightly above 1/3.'

3. The argument for ½. Lewis (2001) maintains that Beauty's subjective probability must remain unchanged at ½. Beauty learns that she is awake. Lewis agrees that this is new evidence. Beauty now knows that the disjunction (*Monday* and *Heads*) ∨ (*Monday* and *Tails*) ∨ (*Tuesday* and *Tails*) is true, where *Monday* = "Today is Monday," and similarly for *Tuesday*. That is self-locating or 'centred' evidence,[1] but Lewis thinks it is irrelevant to Beauty's beliefs about the coin toss.

Why does Lewis think it is irrelevant? Not just because it is self-locating. Lewis thinks self-locating evidence can be substantial: he believes that the 'centred' evidence *Monday* should raise Beauty's subjective probability for *Heads* above ½. His distinction between relevant and irrelevant self-locating evidence seems to be that the former rules out possibilities while the latter rules out nothing. Beauty's being awake is compatible with either toss result and with either day, so it cannot be relevant.

For a clearer case of a self-locating proposition that is substantial, consider a variation.

Complementary Sleeping Beauty (B):

> The set-up is as in version *A*, except that on a result of *Heads*, Beauty is awakened only on Tuesday. On a result of *Tails*, Beauty is not awakened at all.

Figure 2 illustrates the schedule of awakenings.

	Mon	*Tues*
Heads	Not awake	Awake
Tails	Not awake	Not awake

Fig. 2: Complementary Sleeping Beauty

Awakening definitely provides substantial self-locating information. It both tells Beauty that it is Tuesday and increases her probability for *Heads* to 1. Self-locating information can be substantial—the problem is to sort out when and how.

4. The symmetry argument. Hitchcock (2004) provides an argument for the 1/3 answer, based upon a theoretical model and two specific theses. The theoretical model consists of a description of the possible outcomes together with a probability function that reflects Beauty's degrees of belief. The two specific theses are *symmetry* and

[1] See section 4.

independence. Beauty's degrees of belief exhibit symmetry with respect to the two possible days (*Mon* or *Tues*) and probabilistic independence between the current date and the result of the coin toss.

I modify the details of Hitchcock's argument slightly. With Lewis and Elga, Hitchcock represents Beauty's situation using Quine's device of *centred worlds*: possible worlds with individuals-at-times. Since we are concerned with just one individual (Beauty), a centred world is an ordered pair $<w, t>$ where w is a possible world and t is a temporal location in w. If W is the set of all possible worlds and T is the set of all possible times, then the set of all centred worlds is just the Cartesian product $C = W \times T$,[2] representing possible worlds together with possible current locations for Beauty.

We restrict our attention to cases where W and T are finite. In *Sleeping Beauty*, the possible worlds can be represented as the two rows in **Fig. 1**: *Heads* and *Tails*. The possible times consist of the two columns, *Mon* and *Tues*. The set C of centred worlds consists of the four combinations represented as squares in **Fig. 1**. Each world corresponds to an elementary proposition $<O, D>$, where O is the outcome of the coin toss (*Heads* or *Tails*) and D is the current day (*Mon* or *Tues*).

Beauty's degrees of belief (on Monday, but not knowing that it is Monday) are represented by a probability function P defined over disjunctions of these elementary propositions. P is defined by attaching a probability to each elementary proposition, and that is where the two theses come into play.

Symmetry is the assumption that $P(Mon) = P(Tues) = ½$. Hitchcock suggests we may appeal to symmetry, "in the absence of any information favouring one day of the week over the others" (409). These probabilities do not take into account the information that Beauty has been awakened. They rest upon Beauty's appreciation of the experimental set-up. They represent beliefs that she ought to hold if she could be polled in her sleep, just before awakening.

Independence is the assumption that D and O are independent. There is no correlation between the current date and the result of the coin toss. It helps to think of independence as holding right up to the moment before Beauty is awakened. Hitchcock writes: "if she were given the information that it is Monday while sleeping, this should not affect her subjective probability that the coin landed heads."

From these assumptions, together with the unproblematic *P*(*Heads*) = *P*(*Tails*) = ½, it follows that each of the elementary propositions <*Heads*, *Mon*>, <*Tails*, *Mon*>, etc. has probability ¼. Given the schedule represented in **Fig. 1**, where *A* stands for an awakening, we have *P*(*A* / <*Heads*, *Mon*>) = *P*(*A* / <*Tails*, *Mon*>) = *P*(*A* / <*Tails*, *Tues*>) = 1, while *P*(*A* / <*Heads*, *Tues*>) = 0. It follows at once that *P*(*Heads* / *A*) = 1/3. Note also that *P*(*Heads* / ~*A*) = 1: if Beauty could somehow be made aware in her sleep that no awakening has taken place, she would conclude that the coin toss resulted in *Heads*.

5. Two objections. I believe that the preceding argument is decisive, with only a minor modification related to the symmetry assumption (see section 9). By contrast, Hitchcock raises two objections that lead him to conclude that the symmetry argument is only "suggestive".

The first is that the argument requires us to attribute "funny" subjective probabilities to Beauty. The probabilities *P*(*Mon*) = *P*(*Tues*) = ½ and *P*(*Heads* / ~*A*) = 1 are funny because, given our interpretation of *P* as Beauty's degree of belief, they correspond to degrees of belief that Beauty could never actually have. We want *P*(*Mon*) and *P*(*Tues*) to refer to degrees of belief that Beauty has just *before* she awakens, but at that time Beauty lacks the capacity to have degrees of belief. Also, *P*(*Heads* / ~*A*) requires conditionalization on a proposition that Beauty could never learn, namely, that she did not awaken. The concern, then, is that these subjective probabilities are meaningless.

Hitchcock's second (and more serious) objection is that we lack a convincing reason for thinking that conditionalization is the correct way to model how we assimilate self-locating information. Yet the symmetry argument depends upon conditionalization on a self-locating proposition. So a crucial step is unjustified.

Why should conditionalization on a newly learned self-locating proposition be a problem? Hitchcock points out that belief change due to 'learning' such propositions

---

[2] Simplifying assumption: the set of possible times is the same (or isomorphic) across worlds. This is fine for problems like *Sleeping Beauty*, where possible worlds do not vary in spatio-temporal structure.

may violate conditionalization. I currently believe with probability 1 the proposition *Mon*, 'today is Monday.' Upon awakening tomorrow, my probability for *Mon* drops from 1 to 0, while my probability for *Tues* jumps from 0 to 1. Such changes violate conditionalization. Thus, we cannot readily equate learning a self-locating proposition with updating our beliefs via conditionalization.

Of course, a similar point applies to most propositions upon which we conditionalize. Such propositions are typically assigned probability 1 as the result of direct observation (the coin lands *Heads*) or direct action (I accept such-and-such a bet), not conditionalization. Self-locating information is acquired in a direct manner, but that hardly precludes its use in conditionalization.

There are, however, some legitimate concerns. The first pertains to the fickle or non-cumulative nature of self-locating evidence. $P(Mon)$ rises to 1 and then drops to 0. That never happens with ordinary propositions upon which we conditionalize. If E is to count as evidence, we might expect that it should obey the following principle:

> *Stability of Evidence*. If $P_t(E) = 1$, then $P_{t'}(E) = 1$ for times t' later than t (barring cognitive malfunction).

A second, closely related concern is representing subsequent changes in terms of conditionalization. What happens to Beauty's degrees of belief on Wednesday, when she goes home? Here are four possible experimental protocols:

> *Protocol #1*: Beauty remembers everything. She must then assign probability 1 either to *Heads* or to *Tails* (depending on what happened on Tuesday).
>
> *Protocol* #2: Beauty remembers nothing. She reverts to probability ½ for *Heads*.
>
> *Protocol* #3: Beauty remembers just one randomly selected day, but it happens to be a day when she was awakened. The probability for *Heads* remains 1/3.
>
> *Protocol* #4 (the *default protocol*): Beauty remembers only Monday. Her subjective probability for *Heads* reverts to ½. (This also holds if Beauty remembers only the last day on which she was awake.)

Focus on protocol #4. If the initial move to probability 1/3 for *Heads* (on Monday) is the result of conditionalization, the return to ½ must be due to conditionalization on new information (received on Wednesday). But it is difficult to see how this can be.

Conditionalizing upon self-locating information, in short, appears to raise many questions. Nevertheless, I argue in section 7 that we can model Beauty's changing probabilities as successive cases of conditionalization.

6. The first objection. Hitchcock's "funny" subjective probabilities are a concern only because, in *Sleeping Beauty*, we cannot separate the *capacity* for having certain subjective probabilities from the *news* or information that leads to belief revision. Beauty's awakening does double duty. It both enables her to have degrees of belief and functions as a signal that leads to belief change. This is a distracting feature of the scenario rather than a serious problem.

Consider the situation of Beauty's younger sister, Cutie.

Sleeping Cutie (C):

> Cutie undergoes a similar experiment, but is awakened on both days at 9 a.m. If the result of the coin toss was *Heads*, a chime sounds at noon on Monday only. If the result was *Tails*, a chime sounds at noon on both Monday and Tuesday. Shortly after noon, Cutie goes back to sleep. Like Beauty, she has no memory of previous awakenings. Cutie's problem is to come up with a value for *P*(*Heads* / *Chime*).

Here is the picture:

	Mon	*Tues*
Heads	Chime	No Chime
Tails	Chime	Chime

Figure 3: Sleeping Cutie

The early awakening allows Cutie to have degrees of belief, represented by *P*. Hitchcock's first objection does not apply because *P*(*Mon*), *P*(*Tues*) and *P*(*Heads* / *No chime*) cease to be "funny". The first two correspond to degrees of belief that Cutie can and should have between 9 a.m. and 12 noon. The third is meaningful because (*pace* Hitchcock's second objection) Cutie can update by conditionalization on whether the chime sounds at noon. If it sounds, we have *P*(*Heads* / *Chime*) = 1/3.

Other than eliminating Hitchcock's first objection, *Sleeping Cutie* is not relevantly different from *Sleeping Beauty*. We can transform one into the other in three steps.

Step 1: Let Cutie sleep in.

> Let Cutie be awakened close to noon—indeed, precisely at noon, just as the chime sounds (or fails to sound).

At the limit of a noon awakening, the probabilities $P(Mon)$ and $P(Tues)$ cease to be subjective degrees of belief that Cutie can actually have. The continuity argument shows that this is unimportant. Cutie accepts $P(Mon) = P(Tues) = ½$ because had she been awakened one second earlier, she would have held those degrees of belief.

Subjective probability is not precisely the same as actual degree of belief. As van Fraassen (1984) points out, sometimes the conditional probability P(X / Y) "is not to be thought of as the probability we would accord X should we learn that Y". It may be that Y is not a proposition that we can ever come to believe. A similar comment applies to prior probabilities that we could never have. Even though Cutie sleeps in, P represents her understanding of the set-up, unchanged from Sunday night.

Step 2: Make the chime an alarm clock.

> Don't wake Cutie at all if the chime does not sound.

This step makes no difference to Cutie's subjective probabilities for the three cases out of four where she is awakened. The priors are unchanged. The policy about when to sound the chime is unchanged. Her being awake or not provides no information beyond that given by the chime.

Step 3: Dispense with the chime.

It doesn't matter what mechanism is used to awaken Cutie. All that matters is that she is awakened deliberately according to a prescribed schedule. So we can dispense with the chime, which makes Cutie's case identical to that of Beauty. Accordingly, there is no loss of generality if we focus on *Sleeping Cutie* when we turn to Hitchcock's second objection.

7. The second objection: conditionalization and product measures. Cutie conditionalizes her probability for *Heads* on whether or not the chime sounds. My main claim in this section is that it is perfectly in order to update our beliefs about *entirely substantial* propositions by conditionalizing on self-locating information. If A is an entirely substantial proposition (such as *Heads*) and E is newly acquired self-locating information (such as *Chime*), then it is appropriate to shift our degree of belief from $P(A)$ to $P(A / E)$.

In order to defend this claim, I first sharpen the symmetry argument by taking a closer look at the probability measure P on the set of centred possible worlds and the nature of conditionalization. As in section 4, let us represent the set of centred worlds for Cutie as $C = W \times T$, where $W = \{Heads, Tails\}$ and $T = \{Mon, Tues\}$. The natural choice for Cutie's subjective probability function is the product measure $P = P_W \times P_T$, where P_W and P_T are probability functions for W and T. If $\mathrm{A} \subseteq W$ and $\mathrm{B} \subseteq T$, then $P(A \times B) = P_W(A) \cdot P_T(B)$. P is then defined (additively) over finite unions of such rectangles.

The product measure comes with the built-in assumption of independence between entirely self-locating propositions and entirely substantial propositions. Independence is the correct way to formalize the idea that we can learn nothing from (entirely) self-locating information. Such information changes only P_T, not P_W, and can make no difference to our substantial beliefs. Let A be entirely substantial, i.e., a subset of W. Write $P(A)$ for $P(A \times T)$. Then

$$P(A) = P_W(A) \cdot P_T(T) = P_W(A),$$

so a change to P_T makes no difference to $P(A)$. Contrary to Lewis, learning that it is Monday can have no effect on Cutie's probability for *Heads*.[3]

Turn now to conditionalization with the product measure. Let A be an entirely substantial proposition, such as *Heads*. Let E be new evidence. Then conditionalization applied to the product measure gives:

[3] By the same token, losing information that is entirely self-locating makes no difference to the subjective probability for *Heads*. So even if, as Monton (2002) has argued, Beauty forgets which day it is between Sunday and Monday, that type of forgetting cannot explain her shift in probability for *Heads*.

(1) $$P(A/E) = \frac{P((A \times T) \cap E)}{P(E)}$$

This formula can be applied to any sort of evidence *E* at all. In particular, we can apply it to *Sleeping Cutie*. The evidence *E* she gets (on Monday) is *Chime*, which amounts to <*Heads, Mon*> ∨ <*Tails, Mon*> ∨ <*Tails, Tues*>. Here, *A* is the proposition *Heads* and *T* is {*Mon*, *Tues*}. Given that our other assumptions entail that each elementary proposition has probability ¼, we have

$$P(Heads/E) = \frac{P((Heads \times T) \cap E)}{P(E)}$$

$$= \frac{P < Heads, Mon >}{P < Heads, Mon > + P < Tails, Mon > + P < Tails, Tues >}$$

$$= \frac{(1/4)}{(3/4)}$$

= 1/3.

To conditionalize on *Chime* is simply to remove the top right box in **Fig. 3** and distribute the probability over the remaining L-shaped region.

Subsequent conditionalization on further evidence is no problem. We give Cutie the additional information *E′* that it is Monday. She conditionalizes again to obtain *P*(*Heads* / *E′* and *E*) = *P*(*Heads* / *E′*) = ½. The new information removes the bottom right box in **Fig. 3**.

The objections described in section 5, however, pertain to diachronic updating—changes in belief between Sunday and Monday, or between Monday and Wednesday. Instead of *P*, we should speak of P_t, the probability on day *t*. Cutie's subjective probabilities change during the day, but they also change from day to day. Can diachronic updating be represented by conditionalization on self-locating propositions?

Our first concern was that evidence should remain evidence, yet self-locating propositions appear to be ephemeral: they pass from impossible to certain and then back to impossible. In fact, barring memory loss, self-locating *evidence* is as permanent as any other sort of evidence! Granted, the self-locating proposition "today is Monday" is true on day *t* (Monday) and false on Wednesday. But the *evidence* Mon_t, that on day *t* it was

Monday, remains evidence on Wednesday (assuming the default protocol). Cutie might say: "that day I remember was Monday." Similarly, on Wednesday, Cutie still has the evidence $Chime_t$ that the chime sounded *on that remembered day*.

Remembered evidence E_t about time t remains evidence. At t (in fact, Monday), Cutie hears the chime and conditionalizes as above to obtain $P_t(Heads / Chime_t) = 1/3$. If t' is the next remembered day after t, then for any entirely substantial proposition A, I suggest

(2) $P_{t'}(A) = P_t(A / E_t)$

where E_t is the total information obtained on day t.

Acquiring centred evidence about the past time t (at later time t') is equivalent to acquiring the centred evidence at t. On Wednesday, Cutie learns Mon_t: the day she remembers was Monday. Conditionalization on this information is no more problematic than if she acquires that information on Monday. The reasoning leads to probability ½ for *Heads*. If she learns that the remembered day was Tuesday, she can conditionalize on $Tues_t$ to obtain probability 1 for *Tails* (since she already has the evidence $Chime_t$). And so on. Any possible change in Cutie's probability for *Heads* between Monday and Wednesday can be accounted for by conditionalization on new centred evidence about the past, provided she retains her memories of Monday.

This takes care of successive conditionalization, which was the second concern identified in section 5. Conditionalization on self-locating information is unproblematic, provided that the self-locating information is time-indexed and the object of conditionalization is an entirely substantial proposition.

8. Reflection. The 1/3 answer seems to violate standards of rationality. First consider the Sunday-Monday shift: Beauty foresees that her subjective probability for *Heads* will be 1/3 on Monday, but insists on assigning ½ on Sunday. Second, consider the Monday-Wednesday shift: Beauty foresees a change to ½ on Wednesday (assuming the default protocol), but insists on 1/3 on Monday. She twice refuses to take her foreseeable future beliefs as authoritative, despite acknowledging the rationality of the policy that leads to those beliefs.

The relevant 'standards of rationality' are van Fraassen's Reflection principles. General Reflection states:

> (GR) My current opinion about event E must lie in the range spanned by the possible opinions I may come to have about E at later time t, as far as my present opinion is concerned. (1995, 16)

For a person with sharp numerical subjective probabilities, General Reflection entails the Special Reflection Principle:

(SR1) $P(A / \mathrm{P}_t(A) = \mathrm{x}) = \mathrm{x}$,

where P_t stands for my subjective probability function at that later time t. More generally, if I am sure that I will have some precise probability for A at t,

(SR2) $P(A) = \sum_x x \cdot P(P_t(A) = x)$;

$P(A)$ is my expectation value for $P_t(A)$ (1995, 19).

The Sunday-Monday and Monday-Wednesday shifts clash with the Reflection principles. Yet in both cases, Beauty is right not to follow those principles because of her uncertainty about her location on Monday. As Hitchcock (2004) observes, an agent cannot be convicted of an irrational belief revision policy if the policy exhibits no internal inconsistency, but only a clash between the agent's internal beliefs and (inaccessible) outside information. From the outside, Beauty's degree of belief for *Heads* upon awakening is too low on Monday and too high on Tuesday. But that proves nothing about the rationality of her belief revision policy. Still, rather than rejecting Reflection, I propose a modest revision.

Let us say that our situation at t is *classical* if we are certain about our location: $P_t(T{=}t) = 1$. It is *non-classical* if there are times $t_1, \ldots, t_n$ such that $P_t(T{=}t_i) = p_i > 0$ for each t_i, $p_1 + \ldots + p_n = 1$, and t is one of $t_{1,\ldots,}t_n$. Call $\{t_1, \ldots, t_n\}$ the *orbit* of t. The values

p_i need not be equal, but the same values $p_1,\ldots,p_n$ must be assigned to $P_{ti}(T=t_1),\ldots,$ $P_{ti}(T=t_n)$ for each t_i. Otherwise, we could use our current subjective probability assignments to rule out some of the times in the orbit of t.

The two Reflection Principles should be modified when either the present or future time (or both) in the comparison is non-classical. Suppose first that at t_0 I foresee a non-classical period. Let t range over possible future times t_1, $t_2,\ldots,$ t_n in that period, and let $p_i > 0$ be the probability that I will assign at each such time to $t=t_i$. Restricting our attention to substantial propositions A, I suggest the following Reflection principles.

1. ***GRL (General Reflection with uncertain location)*:**

> For any time t, my current probability for A must lie in the range spanned by the probabilities I may have for A at any time in the orbit of t, as far as my present opinion is concerned.

The entire orbit (rather than any single day) is authoritative for the present. In normal cases, this reduces to the familiar principle of General Reflection. For *Sleeping Beauty*, GRL implies that Beauty's probability for *Heads* on Sunday must lie in the range spanned by 1/3 and 1 (since 1 = P(*Heads* / *not Awake*) is a potential subjective probability[4]), which it does.

2. SRL (Special Reflection with uncertain location):

> For any future time t, $P(A / p_{t1}(A)=x \,\&\, \ldots \,\&\, p_{tn}(A)=x) = x$, if $t_1,\ldots,$ t_n are the times I deem possible at t. [P is my current subjective probability.]
>
> More generally,
>
> $$P(A) = \sum_i p_i \sum_x x \cdot P(P_{t_i}(A) = x)\,,$$
>
> where p_i = $P_t(t=t_i)$, the prior probability that $t=t_i$. That is, we take the weighted sum of the expectation of A at each time that I deem possible.

The formula reduces to ordinary Special Reflection in normal cases. For *Sleeping Beauty*, we have $p_1 = p_2 = ½$ and

> P(*Heads*) = ½ (1/3) + ½ ([½ · 1/3 + ½ · 1]) = ½,

the correct value for Sunday. This takes care of the Sunday-Monday transition.

What of the Monday-Wednesday transition? Here, Beauty's uncertainty is about her present temporal location *t* rather than the future location *t**=Wednesday. She knows that on Wednesday, she will get centred information leading to a subjective probability of ½ for *Heads*, but that information might not be about *t* (it might be Tuesday)! Told that *t*=Monday, she will regard Wednesday as authoritative; told that *t*=Tuesday, she will ignore Wednesday and assign probability 0 to *Heads*. Special Reflection here takes the form of a linear combination of these two cases, weighted by her current (posterior) probabilities for the possible values of *t*:

$$P_t(A) = p_{Mon} \cdot P_{t*}(A) + p_{Tues} \cdot 0\,,$$

where *t* is the present moment and *t** is Wednesday. Since p_{Mon}=2/3 and $P_{t*}(A)$ = ½, we get $P_t(Heads)$ = (2/3)(1/2) = 1/3.

I do not offer a completely general version of Reflection. It is enough to see, first, that the Monday-Wednesday shift can be represented as an instance of Reflection, and second, that we only gain full insight into the rationality of that shift when we do so represent it. Reflection makes us smarter. Rather than leading us to reject Reflection, *Sleeping Beauty* should lead to increased respect for this guiding principle.

9. Symmetry revisited. Bostrom (2006) asks us to consider another case.

Extreme Sleeping Beauty (D):

> If the toss results in *Heads*, Beauty is awakened only on day one. If the toss results in *Tails*, Beauty is awakened on one million consecutive days! As usual, she has no memory of prior awakenings and no idea what day it is. What should be her probability for *Heads* when she awakens?

Employing the symmetry argument, we get $\frac{1}{1{,}000{,}001}$. That, says Bostrom, is highly implausible.

This objection signals a problem with the symmetry assumption. Both Hitchcock and Elga appeal to the Principle of Indifference in arguing that all days are equally probable. Hitchcock notes "the absence of any information favouring one day of the week over the others," while Elga notes that Monday and Tuesday are "subjectively

---

[4] Recall the discussion of subjective probability vs. degree of belief in section 7.

indistinguishable" (after a result of *Tails*). Yet we have reason to assign unequal prior probability to the days in *Extreme Sleeping Beauty*. Though subjectively identical, the days are not exactly alike simply because some are much closer to the beginning of the experiment. Before the experiment begins, Beauty knows that she may become ill, the scientists may get bored, or something else might put a halt to the proceedings. Her subjective probability that things will carry on for the full million days could be miniscule. Symmetry is not compulsory!

Suppose that Beauty thinks that there is a small (constant) chance, p, that the experiment will end each night, and $p_{n+1} = (1 - p)p_n$ is the relation between the prior probabilities of days n and $n+1$. Then $P(Heads\ /\ Awake)$ works out to approximately p for a sufficiently long experiment. If $p = 1/100$, then Beauty's degree of belief in *Heads* on being awakened is also around 1/100. This might not satisfy Bostrom but it provides a helpful reality check.

Even in the original *Sleeping Beauty* problem, we need not accept perfect symmetry between Monday and Tuesday. This observation is philosophically important even though the impact on Beauty's numerical probabilities is slight.[5] If $P(Mon)$ is slightly larger than $P(Tues)$, then $P(H\ /\ Awake)$ ends up slightly larger than 1/3. The significant point here is that *independence* (of day and coin toss result) does the real work in justifying Beauty's shift in probability for *Heads* away from ½. Symmetry, or lack of symmetry, simply drives that new subjective probability closer to either 1/3 or ½.

[5] One philosophical implication: Hitchcock notes that his DDB argument compels us to accept $P(Mon) = P(Tues) = ½$, which (if the present claims about symmetry are correct) counts against that argument.

10. Conclusion. A proper understanding of the role of self-locating propositions in belief revision is essential for a resolution of *Sleeping Beauty*. Such propositions are only problematic if they are the object of the belief being updated. Provided we put them in their place, as evidence used to update an entirely substantial belief, we can accommodate them using sophisticated forms of conditionalization and Reflection.

REFERENCES

Bostrom, Nicholas (2006), "Sleeping Beauty and Self-Location: A Hybrid Model", unpublished ms.

Elga, Adam (2000), "Self-Locating Belief and the Sleeping Beauty Problem", *Analysis* 60: 143-7.

Hitchcock, Christopher (2004), "Beauty and the Bets", *Synthese* 139: 405-20.

Lewis, David (1980), "A Subjectivist's Guide to Objective Chance", in Richard Jeffrey (ed.), *Studies in Inductive Logic and Probability*, *Vol. II.* Berkeley: University of California Press.

Lewis, D. (2001), "Sleeping Beauty: Reply to Elga", *Analysis* 61: 171-6.

Monton, Bradley (2002), "Sleeping Beauty and the forgetful Bayesian", *Analysis* 62: 47-53.

Quine, Willard Van Orman (1969), "Propositional Objects", in *Ontological Relativity and Other Essays*. New York: Columbia University Press.

van Fraassen, Bas (1984), "Belief and the Will", *The Journal of Philosophy* 81: 235-56.

van Fraassen, Bas (1995), "Belief and the Problem of Ulysses and the Sirens", *Philosophical Studies* 77: 7-37.