

Synchronic Bayesian updating and the generalized Sleeping Beauty problem

TERRY HORGAN

The two principal answers defended in the literature on the Sleeping Beauty problem (Elga 2000) are $1/2$ and $1/3$. Roger White (2006) poses a generalized version of the problem. ‘When the main arguments for the answer $1/3$ are extended to the generalized case,’ he maintains, ‘they have an unacceptable consequence, whereas extending the halfer’s reasoning turns out rather nicely’ (114). Here I will argue that although my favoured reasoning for the $1/3$ answer (Horgan 2004, in press) does have the consequence in question, this consequence is quite acceptable given my treatment of the original problem. Any apparent unacceptability the consequence initially might seem to possess, I claim, results from a failure to appreciate the nature of newly available information in the Sleeping Beauty problem – in both the original version and White’s generalized version. Once this (easily overlooked) new information and its evidential relevance are taken into account – as is done in my treatment of the original problem – the allegedly unacceptable consequence, in the case of the generalized problem, can be seen to be unobjectionable.

Some thirders (e.g. Elga) disagree with me about the presence of new information in the Sleeping Beauty problem; they deny that the correct answer in the original problem turns on the acquisition of new information. I will argue that White’s generalized Sleeping Beauty problem poses a genuine difficulty for *those* thirders, even though my own version of thirdism handles the problem straightforwardly.

1. The original Sleeping Beauty problem, White’s generalized version, and White’s challenge

The original Sleeping Beauty problem goes as follows. On Sunday Sleeping Beauty learns that she will be put into dreamless sleep for the next two days in a sleep laboratory. She will be awakened briefly on Monday by the experimenters, and then returned to dreamless sleep. If a fair coin that is to be tossed prior to Monday evening lands Heads, then she will sleep through until Wednesday and will awaken by herself. If the coin comes up Tails, then on Monday evening her memory of the Monday awakening will be erased, and she will be briefly awakened again on Tuesday morning by the experimenters. When she is awakened on Monday, with no memory

of a prior awakening, what probability should she assign to the proposition that the coin lands Heads? (One version has the coin toss occurring before the Monday awakening; another has it occurring afterwards. Most who have discussed the problem, including both White and myself, maintain that it doesn't matter.)

White poses the following generalized version of the problem, involving a random waking device with an adjustable randomizer:

A random waking device has an adjustable chance $c \in (0,1]$ of waking Sleeping Beauty when activated on an occasion. In those circumstances in the original story where Beauty was awakened, we now suppose only that this waking device is activated. When $c = 1$, we have the original Sleeping Beauty problem. But if $c < 1$, the case is significantly different. For in this case Beauty cannot be sure in advance that she will be awakened at all during the experiment. When she does wake up she clearly gains some relevant information. (116)

White's challenge to thirder has two parts. First, he claims that their arguments for the $1/3$ answer in the original problem also apply directly to the generalized version, regardless of the setting of the parameter c . (He defends this claim in the body of his paper vis-à-vis Elga's argument and an argument by Dorr (2002) and Arntzenius (2003), and in note 2 vis-à-vis my argument.) Second, he reasons as follows about the Dorr-Arntzenius arguments (and about mine, as note 2 makes explicit):

So according to the Elga and Arntzenius-Dorr [and Horgan] arguments, then, the introduction of the variable c has no effect on the answer to the problem. But this, I submit, cannot be right. As we have noted, if $c < 1$ then when Beauty wakes up she clearly does gain some information, namely

W: Beauty is awake at least once during the experiment.

And this is clearly relevant to whether

H: The coin landed heads.

For the likelihood of W is greater given $\sim H$ than given H. Any answer must take into account the impact of this information on Beauty's credence. For the difference between the likelihoods $P_-(W|H)$ and $P_-(W|\sim H)$ increases as c decreases (where P_- is Beauty's rational credence function prior to waking). The *degree* to which Beauty has a better chance of being awakened given two opportunities rather than one depends on how small c is. *So whatever else we might say about Beauty's rational credence in H when she wakes up, it should vary to some degree with the value of c.* This is the result that the

thirder, insofar as he follows the Elga and Arntzenius-Dorr [or Horgan] arguments, cannot accommodate. (117–18, my italicization of the penultimate sentence)

I will call the inference embodied in the italicized sentence the *H-variability inference*.

2. *Synchronic Bayesian updating and the original Sleeping Beauty problem*

Let me focus first on the original Sleeping Beauty problem, before addressing the generalized version and White's challenge. We thirders agree with one another that in the original problem, for Beauty the epistemic probability of H changes from 1/2 on Sunday to 1/3 when she is awakened on Monday. But we disagree about why this is so. Elga maintains that the epistemic probability changes even though Beauty gains no new relevant information upon being awakened. I maintain, on the contrary, that she does gain new relevant information – and that the change in epistemic probability of H results from what I will here call *synchronic Bayesian updating* on this information.¹

Let me briefly rehearse the reasoning I recommend. When Beauty is awakened on Monday, she thereby acquires a piece of information she did not possess on Sunday – information she expresses with the indexical statement *I am awakened today by the experimenters*. This counts as relevant new information because one possibility about today that is consistent with her Sunday information – but not with her current total information – is *H and today is Tuesday*. (Were this how things actually are today, then of course she would currently be in dreamless asleep, and would not be pondering the problem. Nonetheless, it is a possibility about today that is consistent with what she knew on Sunday.) She considers the following four possibilities (where H and T are *HEADS* and *TAILS* respectively):

Today_{H,Mon}: H and today is Monday.

Today_{H,Tues}: H and today is Tuesday.

Today_{T,Mon}: T and today is Monday.

Today_{T,Tues}: T and today is Tuesday.

¹ Weintraub (2004) explicitly agrees with me that Beauty acquires new relevant information upon being awakened by the experimenters, and about the nature of this information. Although Dorr (2002) and Arntzenius (2003) are not explicit on the matter, I believe that my own approach captures the underlying spirit of both of theirs. Bostrum (in press) claims, as I do, that Beauty acquires new information and that it is indexical; but his approach is otherwise very different.

She first assigns *preliminary probabilities* to these four possibilities – probabilities they possess relative to a certain *proper subset* of her total current information. The proper subset in question excludes her information that she has been awakened today by the experimenters.² (But it does not exclude her information that today is either Monday or Tuesday.) These preliminary probabilities, expressed by ‘P_’, are:

$$P_-(\text{Today}_{H,\text{Mon}}) = 1/4$$

$$P_-(\text{Today}_{H,\text{Tues}}) = 1/4$$

$$P_-(\text{Today}_{T,\text{Mon}}) = 1/4$$

$$P_-(\text{Today}_{T,\text{Tues}}) = 1/4$$

She also assigns the following preliminary *conditional* probabilities to these four possibilities, relative to the same proper subset of her total current information (and letting ‘A_{Today}’ symbolize ‘I am awakened today by the experimenters’):

$$P_-(\text{Today}_{H,\text{Mon}}|A_{\text{Today}}) = 1/3$$

$$P_-(\text{Today}_{H,\text{Tues}}|A_{\text{Today}}) = 0$$

$$P_-(\text{Today}_{T,\text{Mon}}|A_{\text{Today}}) = 1/3$$

$$P_-(\text{Today}_{T,\text{Tues}}|A_{\text{Today}}) = 1/3$$

She now does Bayesian updating by means of these preliminary conditional probabilities, thereby taking into account her additional current information that A_{Today}. The resulting updated probabilities are:

$$P(\text{Today}_{H,\text{Mon}}) = 1/3$$

$$P(\text{Today}_{H,\text{Tues}}) = 0$$

$$P(\text{Today}_{T,\text{Mon}}) = 1/3$$

$$P(\text{Today}_{T,\text{Tues}}) = 1/3$$

And so,

$$P(H) = P(\text{Today}_{H,\text{Mon}}) = 1/3$$

$$P(T) = [P(\text{Today}_{T,\text{Mon}}) + P(\text{Today}_{T,\text{Tues}})] = 2/3$$

The Bayesian updating she employs is *synchronic*, because the four possibilities to which she assigns preliminary probability all have arisen post-awakening – by virtue of the fact that upon being awakened, she no longer knows which day it is (although she does know that today is either Monday or Tuesday). But it is Bayesian updating nonetheless, even though it does not conform to the familiar special case in which the preliminary

² It also excludes her information that she is awake *right now*, since if right now were Tuesday and the coin had landed heads, then she would now be in dreamless sleep.

probabilities are *temporally prior* probabilities.³ Accordingly, I will call my position *Bayesian thirdism*.⁴

3. Synchronic Bayesian updating and the generalized Sleeping Beauty problem

Let me now explain how my recommended form of reasoning extends to White's generalization of the Sleeping Beauty problem. With the variable parameter c in the picture, Beauty's preliminary probabilities are still the same as before:

$$\begin{aligned} P_{-}(\text{Today}_{H,\text{Mon}}) &= P_{-}(\text{Today}_{H,\text{Tues}}) = P_{-}(\text{Today}_{T,\text{Mon}}) \\ &= P_{-}(\text{Today}_{T,\text{Tues}}) = 1/4 \end{aligned}$$

Her preliminary probabilities for the pertinent conjunctions involving 'A_{Today}', again bracketing the information about having been awakened today, are these:

$$\begin{aligned} P_{-}(\text{Today}_{H,\text{Mon}} \ \&\ A_{\text{Today}}) &= 1/4c \\ P_{-}(\text{Today}_{H,\text{Tues}} \ \&\ A_{\text{Today}}) &= 0 \\ P_{-}(\text{Today}_{T,\text{Mon}} \ \&\ A_{\text{Today}}) &= 1/4c \\ P_{-}(\text{Today}_{T,\text{Tues}} \ \&\ A_{\text{Today}}) &= 1/4c \end{aligned}$$

Thus, concerning the possibility that she is awakened today by the experimenters, her preliminary probability is this:

$$\begin{aligned} P_{-}(A_{\text{Today}}) &= P_{-}(\text{Today}_{H,\text{Mon}} \ \&\ A_{\text{Today}}) + P_{-}(\text{Today}_{T,\text{Mon}} \ \&\ A_{\text{Today}}) + \\ &P_{-}(\text{Today}_{T,\text{Tues}} \ \&\ A_{\text{Today}}) = 3/4c \end{aligned}$$

So, plugging into the definition of conditional probability, her preliminary conditional probabilities (with the awakening-information still bracketed) are:

$$\begin{aligned} P_{-}(\text{Today}_{H,\text{Mon}} | A_{\text{Today}}) &= P_{-}(\text{Today}_{H,\text{Mon}} \ \&\ A_{\text{Today}}) / P_{-}(A_{\text{Today}}) \\ &= (1/4c) / (3/4c) = 1/3 \end{aligned}$$

³ It bears emphasis that, even in instances of the familiar special case, one's Bayesian updating still employs preliminary probabilities that accrue to various possibilities relative to a proper subset of one's total *current* information. In the special case, this proper subset happens to coincide with a body of information that was *previously* one's *total* pertinent information.

⁴ This label is intended to highlight the role of Bayesian updating in my treatment of the problem. Needless to say, one can advocate Bayesian thirdism without embracing the package of subjectivist views about the nature of epistemic probability, and about constraints on rational belief-formation, commonly called 'Bayesianism'. I myself am no fan of Bayesianism, although I am a Bayesian thirder.

$$\begin{aligned} P_{-}(\text{Today}_{H,\text{Tues}}|A_{\text{Today}}) &= P_{-}(\text{Today}_{H,\text{Tues}} \ \& \ A_{\text{Today}})/P_{-}(A_{\text{Today}}) \\ &= 0/(3/4c) = 0 \end{aligned}$$

$$\begin{aligned} P_{-}(\text{Today}_{T,\text{Mon}}|A_{\text{Today}}) &= P_{-}(\text{Today}_{T,\text{Mon}} \ \& \ A_{\text{Today}})/P_{-}(A_{\text{Today}}) \\ &= (1/4c)/(3/4c) = 1/3 \end{aligned}$$

$$\begin{aligned} P_{-}(\text{Today}_{T,\text{Tues}}|A_{\text{Today}}) &= P_{-}(\text{Today}_{T,\text{Tues}} \ \& \ A_{\text{Today}})/P_{-}(A_{\text{Today}}) \\ &= (1/4c)/(3/4c) = 1/3 \end{aligned}$$

She now does Bayesian updating by means of these preliminary conditional probabilities, thereby taking into account her additional current information that A_{Today} . The resulting updated probabilities, just as in the original problem, are:

$$\begin{aligned} P(\text{Today}_{H,\text{Mon}}) &= 1/3 \\ P(\text{Today}_{H,\text{Tues}}) &= 0 \\ P(\text{Today}_{T,\text{Mon}}) &= 1/3 \\ P(\text{Today}_{T,\text{Tues}}) &= 1/3 \end{aligned}$$

So again, just as in the original problem,

$$P(H) = P(\text{Today}_{H,\text{Mon}}) = 1/3$$

$$P(T) = [P(\text{Today}_{T,\text{Mon}}) + P(\text{Today}_{T,\text{Tues}})] = 2/3$$

Thus the original reasoning, employing synchronic Bayesian updating, extends straightforwardly to the generalized version of the problem. Beauty's newly acquired information that she has been awakened *today* by the experimenters has the same effect in the general case as in the special case where $c = 1$: this information changes the epistemic probability of H, which was 1/2 on Sunday, to 1/3.

4. Bayesian thirdism and the H-variability inference: diagnosing the fallacy

Against the background of the preceding discussion, let us now consider White's H-variability inference. The possibility he labels W can be usefully reformulated as follows, without prejudice to his argument (and in a way that conforms with the first-person indexical language employed in A_{Today}):

A_1 : I am awakened by the experimenters at least once during Monday and Tuesday.

Thus reformulated, the premiss of the inference is this:

PREMISS: If $c < 1$ then $P_{-}(A_1|\sim H) > P_{-}(A_1|H)$; furthermore, 'the degree to which Beauty has a better chance of being

awakened given two opportunities rather than one depends on how small c is' (118).

On my account, this premiss is certainly true. For, letting 'A_{Mon}' and 'A_{Tues}' respectively symbolize 'I am awakened by the experimenters on Monday' and 'I am awakened by the experimenters on Tuesday',

$$\begin{aligned} P_-(A_1|H) &= P_-((A_{\text{Mon}} \vee A_{\text{Tues}})|H) \\ &= P_-(A_{\text{Mon}}|H) + P_-(A_{\text{Tues}}|H) - P_-(A_{\text{Mon}} \& A_{\text{Tues}}|H) \\ &= c + 0 - 0 \\ &= c \end{aligned}$$

$$\begin{aligned} P_-(A_1|\sim H) &= P(A_1|T) \\ &= P_-((A_{\text{Mon}} \vee A_{\text{Tues}})|T) \\ &= P_-(A_{\text{Mon}}|T) + P_-(A_{\text{Tues}}|T) - P_-(A_{\text{Mon}} \& A_{\text{Tues}}|T) \\ &= c + c - (c \times c) \\ &= 2c - c^2 \end{aligned}$$

Thus, when $c = 1$,

$$P_-(A_1|H) = 1$$

$$P_-(A_1|\sim H) = (2 - 1) = 1,$$

while, as c approaches 0, the ratio of the quantity $(2c - c^2)$ to the quantity c grows increasingly greater.

White reasons as follows, on the basis of *PREMISS*: 'So, whatever else we might say about Beauty's rational credence in H when she wakes up, it should vary to some degree with the value of c ' (118). (This is what I above labelled the H-variability inference.) But, on my account, this inference is mistaken. For, as argued already, the right way to assign epistemic probabilities to H and T is by means of synchronic Bayesian updating (using the information that A_{Today}) on the following preliminary conditional probabilities:

$$P_-(\text{Today}_{H,\text{Mon}}|A_{\text{Today}}) = P_-(\text{Today}_{T,\text{Mon}}|A_{\text{Today}}) = P_-(\text{Today}_{T,\text{Tues}}|A_{\text{Today}}) = 1/3$$

$$P_-(\text{Today}_{H,\text{Tues}}|A_{\text{Today}}) = 0$$

And the result of such updating is the same, *regardless of the value of the parameter c* :

$$P(H) = P(\text{Today}_{H,\text{Mon}}) = 1/3$$

$$P(T) = [P(\text{Today}_{T,\text{Mon}}) + P(\text{Today}_{T,\text{Tues}})] = 2/3$$

What makes the H-invariability inference seem initially plausible, even though it is actually fallacious? Well, suppose that one believes – as many who have written about the Sleeping Beauty problem *do* believe – that, in the original version of the problem, Beauty obtains no new relevant information upon being awakened on Monday by the experimenters. Then one will also be strongly inclined to believe, concerning the generalized Sleeping Beauty problem, that the strongest new relevant information Beauty receives on Monday is the information A_1 . And if one believes *that*, then White’s H-variability inference becomes extremely plausible. After all, White is right to claim that ‘the *degree* to which Beauty has a better chance of being awakened given two opportunities rather than one depends on how small c is’ (118). So, if indeed the strongest new relevant information Beauty receives is A_1 , then it becomes *very* hard to see how the information that A_1 could fail to have a differentially stronger effect on the epistemic probability of H, depending on how small c is.

Moreover, again on the assumption that A_1 is the strongest new relevant information Beauty receives upon being awakened, the natural-looking way to take account of this information is to do *diachronic* Bayesian updating, by means of the preliminary probability $P_{-}(H|A_1)$. (The updating is diachronic because A_1 is a state of affairs expressible without any temporal-indexical term like ‘today’, and thus $P_{-}(H|A_1)$ has the same numerical value on Sunday that it has when Beauty is awakened on Monday.) By Bayes’s theorem, as White shows,

$$\text{As } c \rightarrow 1, P_{-}(H|A_1) \rightarrow 1/2$$

$$\text{As } c \rightarrow 0, P_{-}(H|A_1) \rightarrow 1/3$$

Thus, diachronic Bayesian updating via $P_{-}(H|A_1)$ gives these results:

$$\text{As } c \rightarrow 1, P(H) \rightarrow 1/2 \text{ (the halfer’s answer to the original problem)}$$

$$\text{As } c \rightarrow 0, P(H) \rightarrow 1/3 \text{ (the thirder’s answer to the original problem)}$$

So White’s H-variability inference, and also his use of diachronic Bayesian updating via the preliminary conditional probability $P_{-}(H|A_1)$, are both very plausible and natural, *given the assumption that the strongest newly acquired relevant information, in the generalized Sleeping Beauty Problem, is A_1 .*⁵

But on my account this assumption is *false* – as is the assumption’s consequence that in the original problem *no* new relevant information is

⁵ The presence of this assumption in White’s reasoning surfaces explicitly in the following passage: ‘Halfers are suspicious of any shift in credence that is not in response to new relevant information. So in the generalized case they insist that Beauty should simply update her credence in the standard way *by conditionalizing on her strongest new information, namely W* ’ (118, my italics – where W = ‘Beauty is awake at least once during the experiment’).

acquired. Upon being awakened by the experimenters, Beauty actually acquires not merely the information that she is awakened *at least once* by the experimenters, but also the stronger (more specific, essentially indexical) information that she is awakened *today* by the experimenters. (The latter piece of information entails the former under the conditions of the problem, but not conversely.) Thus, the fallacy in the H-variability inference, and in diachronic Bayesian updating via $P_{-}(H|A_1)$, is that they *ignore relevant new information*. The correctly updated epistemic probability for H is obtained by conditionalizing not on the partial new relevant information A_1 , but rather on the strongest new relevant information – viz., A_{Today} . And, regardless of the value of the parameter c , $P(H) = P_{-}(H|A_{\text{Today}}) = 1/3$.

5. Bayesian thirdism v. non-Bayesian thirdism

As I said, it is very commonly assumed in the literature on the original Sleeping Beauty problem that Beauty acquires no new relevant information upon being awakened by the experimenters. Some thirders hold this view – notably Elga, who is explicit about it. These non-Bayesian thirders cannot and do not argue, as I do, that the basis for the change in $P(H)$ from $1/2$ on Sunday to $1/3$ on Monday is Bayesian updating on newly acquired information. Nor can they reply to White, as I do, that White’s H-variability inference in the generalized Sleeping Beauty problem – and likewise his appeal to diachronic Bayesian updating via $P_{-}(H|A_1)$ – are fallacious by virtue of ignoring relevant new information.

On the assumption that Beauty gains no new relevant information in the original problem, and on the correlative assumption that the strongest new relevant information she gains in the generalized problem is A_1 , White’s H-variability inference does look very plausible – as does diachronic updating by means of $P(H|A_1)$. The generalized Sleeping Beauty problem thus poses a serious challenge to non-Bayesian thirders, even though it makes no trouble for Bayesian thirdism. So, unless and until the non-Bayesian thirders provide a plausible treatment of their own of the generalized problem, and a plausible account of their own of how White’s reasoning goes wrong, the capacity of Bayesian thirdism to smoothly handle the generalized problem provides new dialectical support for my own contention, over against the non-Bayesian thirders, that the right form of thirdism is Bayesian. The correct way to reason, in both the original and the generalized versions of the problem, is to invoke synchronic Bayesian updating on newly acquired indexical information.⁶

⁶ My thanks to Nathan Ballantyne, David Chalmers, Adam Elga, Justin Fisher, Michaela Mueller, Joel Pust, Mark Timmons and Roger White for helpful comments and discussion.

The University of Arizona
Tucson, AZ 85721-0027, USA
thorgan@email.arizona.edu

References

- Arntzenius, F. 2003. Some problems for conditionalization and reflection. *Journal of Philosophy* 100: 356–70.
- Bostrum, N. In press. Sleeping Beauty and self-location: a hybrid model. *Synthese*.
- Dorr, C. 2002. Sleeping Beauty: in defence of Elga. *Analysis* 62: 292–96.
- Elga, A. 2000. Self-locating belief and the Sleeping Beauty problem. *Analysis* 60: 143–47.
- Horgan, T. 2004. Sleeping Beauty awakened: new odds at the dawn of the new day. *Analysis* 64: 10–21.
- Horgan, T. In press. Synchronic Bayesian updating and the Sleeping Beauty problem: reply to Pust. *Synthese*.
- Weintraub, R. 2004. Sleeping Beauty: a simple solution. *Analysis* 64: 8–10.
- White, R. 2006. The generalized Sleeping Beauty problem: a challenge for thirders. *Analysis* 66: 114–19.