The generalized Sleeping Beauty problem: a challenge for thirders

ROGER WHITE

The two candidate answers to the Sleeping Beauty problem (Elga 2000) are 1/2 and 1/3, the proponents of which are known as halfers and thirders. By considering a generalization of the original puzzle, I pose a challenge to thirders: When the main arguments for the answer 1/3 are extended to the generalized case they have an unacceptable consequence, whereas extending the halfer's reasoning turns out rather nicely.

1. The original Sleeping Beauty problem

On Sunday Sleeping Beauty learns that she will be put to sleep for the next two days. If the fair coin that is to be tossed lands Heads, she will be awakened briefly on Monday. If it lands Tails, she will be awakened briefly on Monday, returned to sleep with her memory of that awakening erased, then awakened briefly again on Tuesday. When she awakens on Monday, what should Beauty's credence be that the coin landed Heads?

ANALYSIS 66.2, April 2006, pp. 114–19.

A natural first answer is 1/2. Since Beauty knew no more than that the coin was fair, her initial credence that the coin would land Heads should have been 1/2. Has she learnt anything new that should alter this judgment? She knew all along that she was to be awakened briefly during the experiment at some time. So it is no news to her when she finds herself awake at an unknown time. When awakened she may learn something that she would express as 'I am awake *now*.' But it is difficult at best to see what bearing this could have for her on whether the coin landed Heads. Hence surely her credence that the coin landed Heads should remain at 1/2.

Nevertheless, the majority of philosophers who have written on the puzzle have concluded that the correct answer is 1/3. Thirders include Arntzenius (2003), Dorr (2002), Elga (2000), Hitchcock (2004), Horgan (2004), Monton (2002) and Weintraub (2004). Lewis (2001) is the only explicit halfer that I know of in print, but Bradley (2003) challenges Dorr's argument for the 1/3 answer. There have been two main arguments for 1/3, which I will only briefly sketch here:

The Elga argument: When Beauty wakes up she knows that she is in one of the following 'predicaments':

H_{MON}: The coin landed Heads and it is now Monday.
T_{MON}: The coin landed Tails and it is now Monday.
T_{TUE}: The coin landed Tails and it is now Tuesday.

Let P be the rational credence function for Beauty when she wakes up on Monday. That her credence in the coin having landed Heads should be 1/3 follows from two lemmas:

(1) $P(T_{MON} | T_{MON} \text{ or } T_{TUE}) = P(T_{TUE} | T_{MON} \text{ or } T_{TUE})$

(2) $P(H_{MON} | H_{MON} \text{ or } T_{MON}) = 1/2$

Proof: (1) entails that $P(T_{MON}) = P(T_{TUE})$. (2) entails that $P(H_{MON}) = P(T_{MON})$. So $P(H_{MON}) = P(T_{MON}) = P(T_{TUE}) = 1/3$, since these predicaments are exhaustive and incompatible at a time. Beauty knows the coin landed Heads if and only if she is in H_{MON}, hence her credence that it did should be 1/3.

Argument for (1): Given that the coin has landed Tails, and hence that she is in either predicament T_{MON} or T_{TUE}, Beauty has no more reason to suppose that she is undergoing the first Tails waking rather than the second, or vice versa. Hence she should divide her credence equally.

Argument for (2): It should make no epistemic difference to Beauty if the coin is tossed before or after the first waking to determine whether she will be awakened again on Tuesday. Supposing then that it's the latter, if Beauty is informed that it is Monday and hence that she is in either H_{MON}

or T_{MON}, her credence that she is in H_{MON} should equal her credence that a fair coin which is yet to be tossed will land Heads, namely 1/2.

The Arntzenius-Dorr argument: Consider a variant case:

> *Modified Story*: Exactly as in the original story, except that if the coin lands Heads then Beauty is awakened again on Tuesday (her memory of the earlier waking erased). But after a brief pause she has an experience by which she can verify that the coin has landed Heads and it is Tuesday.[1]

Upon waking on Monday Beauty's credence should be divided evenly among the four predicaments: $\{H_{MON}, H_{TUE}, T_{MON}, T_{TUE}\}$. When she rules out H_{TUE} by failing to have the distinguishing experience, she learns nothing that should affect her distribution of credence among the remaining possibilities. Hence she should have credence of 1/3 in each. Since the total information that Beauty has to go on now – that she is in one of the three predicaments: $\{H_{MON}, T_{MON}, T_{TUE}\}$ – is the same in the original puzzle in which being awake in H_{TUE} is never an open possibility for her, the answer should be 1/3 in the original problem also.

2. *The generalized Sleeping Beauty problem*

The challenge that I have for thirders arises from the following generalization of the original puzzle setup. 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. For she has a greater chance of being awakened if the coin lands Tails, since she will in that case have two opportunities instead of one in which the device

[1] Much the same argument was arrived at independently by Arntzenius (2003) and Dorr (2002). In Arntzenius' version, in H_{TUE} Beauty is not strictly awake but enjoys a vivid dream whose only phenomenological difference from waking experience is that pinching herself doesn't hurt. In Dorr's version, if the coin lands Heads she is given only temporary amnesia after her Monday wakening, so that part way into H_{TUE} her memories flood back. Another difference of debatable relevance is that the version of the unmodified story that Dorr addresses has Beauty waking on Tuesday if the coin lands heads with her memories intact (in Elga's and Arntzenius' versions she remains asleep for two days.) Dorr appeals to sorites-style reasoning to support the equivalence of the two cases. My interpretation of the argument more closely follows Arntzenius' presentation.

might wake her. So even ardent halfers must agree that in this case Beauty's credence should shift toward the coin's having landed Tails.

But let's consider how Elga's argument should be extended to the case where $c < 1$, considering the two crucial lemmas in turn. (1) Once again, it appears that if Beauty were to awaken and learn that the coin landed Tails, she should divide her credence equally between T_{MON} and T_{TUE}. For she knows that if the coin lands Tails, the waking device is activated on Monday and again on Tuesday, with the same chance of her waking on each occasion. (2) Now we suppose that Beauty learns just that it is Monday. If we can suppose without crucially altering the case that the coin is yet to be tossed, it seems that her credence that it is Monday and the coin lands Heads, i.e. that she is in H_{MON}, should be 1/2. For she knows that whether she is depends on whether a fair coin that is yet to be tossed lands Heads. So we appear to have the required assumptions (1) and (2) to derive the answer 1/3. At any rate, we have *no less* reason to follow Elga's reasoning in the generalized case than we did in the original one. If we trust Elga's original argument we should conclude that Beauty's credence on Monday that the coin landed Heads should be 1/3, regardless of what value c takes.

We get the same result by extending the Arntzenius-Dorr argument. First we modify the case by supposing that if the coin lands Heads, then on Tuesday the waking device is activated again except that if awakened then, Beauty can soon discern that she is in predicament H_{TUE}. It appears that upon waking on Monday she should first distribute her credence equally among the four predicaments. The fact that she only has a chance of $c < 1$ of being awakened on any of these four occasions cannot affect the case. So once she has determined that she is not in H_{TUE}, her credence that she is in H_{MON} should shift to 1/3. Hence we arrive by the reasoning above that in the case which does not include H_{TUE} as a possibility for her to be awake in, her credence that she is in H_{MON} and hence that the coin landed heads should be 1/3.[2]

So according to the Elga and Arntzenius-Dorr arguments then, the introduction of 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.

[2] Horgan (2004) presents an argument he identifies as being in the same spirit as Dorr's but without the appeal to a modified case. He suggests that when Beauty wakes up she should assign *prior* probabilities of 1/4 to each of the 'statements' $\{H_{MON}, H_{TUE}, T_{MON}, T_{TUE}\}$. Her *current* probability is obtained by assigning zero to H_{MON}, and renormalizing to give 1/3 to the others. In so far as I understand the rationale being applied here, the conclusion will be unaffected by lowering the value of c. So Horgan's argument faces the same problem as Arntzenius' and Dorr's.

And this is clearly relevant to whether

H: The coin landed Heads.

For the likelihood of *W* is greater given ~*H* than given *H*. Any answer must take into account the impact of this information on Beauty's credence. But now the force of this impact must depend partly on the value of *c*. For the difference between the likelihoods P_(*W*|*H*) and P_(*W*|~*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 a result that the thirder, insofar as he follows the Elga and Arntzenius-Dorr arguments, cannot accommodate.[3]

From the sorry plight of the thirder, let's turn to the happier results of the halfer. 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*. Beauty's new credence in *H* should be

$$\begin{aligned} \mathrm{P}(H) &= \mathrm{P}\_(H|W) \\ &= \mathrm{P}\_(H)\,\mathrm{P}\_(W|H) \,/\, [\mathrm{P}\_(H)\,\mathrm{P}\_(W|H) + \mathrm{P}\_(\sim H)\,\mathrm{P}\_(W|\sim H)] \\ &= (1/2)c \,\Big/\, \left[(1/2)c + (1/2)\left(1-(1-c)^2\right)\right] \\ &= 1 \,/\, (3-c) \quad (\text{since } c > 0) \end{aligned}$$

Here we get an interesting result:

As $c \to 1$, $\mathrm{P}(H) \to 1/2$ (the halfer's answer to the original problem)
As $c \to 0$, $\mathrm{P}(H) \to 1/3$ (the thirder's answer to the original problem)

On the halfer's analysis, the 1/3 answer is correct only at the limit as the chance of being awakened on any occasion gets arbitrarily small.

[3] We can make the difficulty for the thirder more dramatic by considering a modified case. Suppose that $c = 0.1$, but whereas if the coin lands Heads, the waking device is activated only on Monday, if it lands Tails the device is activated once a day for twenty-five days. In this case if the coin lands Tails, Beauty's chance of being awakened at least once during the experiment is greater than 0.9, while on Heads it is only 0.1. This dramatic difference in likelihoods should surely make a difference to Beauty's credence when she wakes up. Yet according to the thirder's arguments, her credence should be no different from the credence in the case in which $c = 1$, where there is no difference in these likelihoods at all.

Without having diagnosed the exact error in the Elga and Arntzenius-Dorr arguments, the challenge I have raised should undermine their case for 1/3.[4]

New York University
New York, NY 10003-6688, USA
roger.white@nyu.edu

References

Arntzenius, F. 2003. Some problems for conditionalization and reflection. *Journal of Philosophy* 100: 356–70.
Bradley, D. 2003. Sleeping Beauty: a note on Dorr's argument for 1/3. *Analysis* 63: 266–67.
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.
Hitchcock, C. 2004. Beauty and the bets. *Synthese* 139: 405–20.
Horgan, T. 2004. Sleeping Beauty awakened: new odds at the dawn of the new day. *Analysis* 64: 10–21.
Lewis, D. 2001. Sleeping Beauty: reply to Elga. *Analysis* 61: 171–76.
Monton, B. 2002. Sleeping Beauty and the forgetful Bayesian. *Analysis* 62: 47–53.
Weintraub, R. 2004. Sleeping Beauty: a simple solution. *Analysis* 64: 8–10.

[4] Thanks to Cian Dorr, Adam Elga, Matt Kotzen and Sydney White for conversations about Sleeping Beauty.