Sleeping Beauty: theme and variations

J. Finkelstein

Sleeping Beauty theme Sleeping Beauty, before she goes to sleep on Sunday, knows that she is the subject of the following experiment: A fair coin will be tossed; if it lands showing heads, she will be awakened Monday morning. If the coin lands showing tails, she will be awakened Monday morning, will be put to sleep Monday evening with a drug which causes her to forget everything she experienced on Monday, and will be awakened a second time on Tuesday morning. Since on Tuesday morning she will have no recollection of Monday, Tuesday morning and Monday morning will be indistinguishable for her, and so (whichever way the coin lands) when she wakes she will not know whether it is Monday or Tuesday.

This story was published by Elga (2000) to pose the following question: when Sleeping Beauty wakes, what should be her credence that the coin landed showing heads? Elga, and a majority of the participants in the ensuing discussion ("thirders") claim that the answer is one-third (quickly, because if this experiment were run a large number of times, the expected fraction of awakenings with the coin showing heads would be one-third), but others, e.g. David Lewis (2001) ("halfers") argue that the answer is one-half (quickly, because when she wakes she will know exactly as much as she did on Sunday evening, when her credence for Heads was surely one-half). In this paper, I will describe a series of variations on this theme, constructed so that once we agree on what should be her credence in a given variation, that will help us to see what it should be in the variation following in the series, which will ultimately lead back to the Sleeping Beauty theme.

Variation #1 This variation is similar to the "simplified version" of the Sleeping Beauty story told by Peter J. Lewis (2007) in his discussion of the relationship between the Sleeping Beauty story and the many-worlds interpretation of quantum mechanics. There is no coin used in this variation. Sleeping Beauty is awakened on three mornings (in Peter J. Lewis's version it was two), and in the first and second evenings made to forget that she had been awakened, so that when she wakes she can not tell whether it is the first, second, or third morning. I take it to be non-controversial that when she wakes her credence for it being the first morning should be one-third.

In this variation it is not necessary to specify that the three mornings in which she is awakened must be on consecutive days of the week. It could be, for example, that she is awakened on a Monday morning, then made to sleep until the following Friday when she is awakened again, and then made to sleep until Wednesday of the following week and awakened for the third time. In fact, although I will always assume that she is fully aware of the experimental protocol, it would not matter if she did not know on which days of the week the awakenings were to occur.

Variation #2 As in variation #1, she is awakened on three mornings. In variation #2, a picture is placed, where she cannot see it, facing up on the first morning and facing down on the last two mornings. When she wakes, what should be her credence for the picture facing up?

She knows that the picture faces up if and only if it is the first morning, so her credence for Up should be the same as her credence for First morning, which as we have seen in variation #1, should be one-third.

Variation #3 A fair coin is tossed. If the coin lands showing heads, the experiment proceeds exactly as in variation #2: she is awakened on three mornings, with the picture facing up on the first morning and down on the second and third mornings. If the coin lands showing tails, she will still be awakened on three mornings, but the picture will be face down on the first and second mornings and up on the third morning. When she wakes, she does not know the result of the coin toss; what should be her credence for the picture facing up?

If she knew that the coin had shown heads, she would know that the experiment proceeds as in variation #2, so her credence for Up would be the same as in that variation, namely one-third. And a similar argument[1] shows that, if she knew the coin had shown tails, her credence for Up should again be one-third . Since her credence for Up should be one-third whether the coin showed heads or tails, then, without knowing the result of the coin toss, her credence for Up should be one-third.

Variation #4 This variation takes two weeks, and is summarized in the table below. A fair coin is tossed on Sunday evening. If the coin shows heads, she is awakened on the first Monday morning with the picture facing up, awakened for a second time on Monday of the second week with the picture facing down, and a third time on Tuesday of the second week with the picture facing down. If the coin shows tails, she is awakened on the first Monday with the picture facing down, again on the first Tuesday with the picture facing down, and for a third time on Monday of the second week with the picture facing up.

	first Monday	first Tuesday	second Monday	second Tuesday
heads	up	(not awake)	down	down
tails	down	down	up	(not awake)

When she wakes, what should be her credence for Up?

[1]In variation #1, her credence for the third morning should also be one-third. Instead of variation #2, I could have defined variation #2a, in which the picture was face down for the first two mornings and face up for the third. Her credence for Up in variation #2a should be the same as her credence for the third morning, namely one-third. If the coin has shown tails, variation #3 proceeds as in variation #2a.

Variation #4 differs from variation #3 only in that the days on which she will be awakened are specified, in a way which depends on the result of the coin toss. But as far as Sleeping Beauty's experience is concerned, variation #4 is just a special case of variation #3, and so her credence for Up should be the same as in variation #3, namely one-third.[2]

Variation #5 The same as variation #4, except that after she is awakened, she is told whether it is the first or the second week (but she is not told the day of the week). Suppose that she is told that it is the first week; what should be her credence for Up?

Let *P(first week)* represent her credence, after she wakes but before she is told which week it is, for it being the first week; analogously define *P(second week)*, with

$$P(\text{first week}) + P(\text{second week}) = 1. \quad (1)$$

Let $P_{\text{before}}(Up)$ represent her credence for Up after she wakes but before she knows which week it is; then

$$P_{\text{before}}(Up) = P(\text{first week})\,P(Up \mid \text{first week}) + P(\text{second week})P(Up \mid \text{second week}). \quad (2)$$

Since the coin was fair, the symmetry of the protocol table (given in variation #4) implies that[3]

$$P(Up \mid \text{second week}) = P(Up \mid \text{first week}) \quad (3)$$

These three equations imply

$$P(Up \mid \text{first week}) = P_{\text{before}}(Up). \quad (4)$$

We see from variation #4 that $P_{\text{before}}(Up)$ = one-third. Thus after she wakes but before she is told which week it is, Sleeping Beauty thinks that if it is the first week, the probability of the picture facing up is one-third. Then when she is told that it *is* the first week, her credence for Up is one-third.

---

[2]Vaidman and Saunders (2001) and Bostrom (2007) also discuss versions of the Sleeping Beauty story which take two weeks. In those two versions a coin is tossed twice, once for each week, while in variation #4 there is only one coin toss. Vaidman and Saunders' "experiment iii", which takes 5218 weeks, also has some similarity with variation #4. Vaidman and Saunders rely on approximations which are valid when the number of weeks is large to relate their experiment iii to Elga's Sleeping Beauty story , but no such approximation will be needed here (which is fortunate, because two, the number of weeks in variation #4, is not a very large number!).

[3] Thirders will want to say that, given it is the first week, her credence for Heads, and thus her credence for Up, should be one-third, and that, given it is the second week, her credence for Tails, and thus her credence for Up, should be one-third. Halfers will want to say that each credence should be one-half. Either way, they should all agree with the equality (3).

Variation #6 Same procedure as in variation #5, but with a different question. After she is told that it is the first week, what should be her credence that the tossed coin showed heads?

During the first week, the picture is face up if and only if the tossed coin showed heads. Therefore once she knows that it is the first week, her credence for Heads should be the same as her credence for Up, which from variation #5 is one-third. This brings us

Back to theme The first week of variation #6 *is* the Sleeping Beauty theme, the story told by Elga. True, that story does not make any mention of a picture facing one way or the other, but since in variation #6 Sleeping Beauty never sees the picture at all, its presence should not affect her credence for Heads. Once she is told that it is the first week, Sleeping Beauty in variation #6 knows exactly as much as does Sleeping Beauty in the theme, whose credence for Heads should therefore be one-third.

So it seems that the thirders are correct. To deny this conclusion, to assert for example that her credence in variation #3 should be one-third but her credence in variation #4 should be one-half, would seem to require not just separate calculations for her credence in those two variations, but also an explanation of how her credence in those two variations could be different.

J. Finkelstein
Department of Physics and Astronomy
San Jose State University
San Jose, CA 95192-0106 U.S.A.
JLFinkelstein@lbl.gov

References

Bostrom, N. 2007. Sleeping beauty and self-location: A hybrid model. Synthese 157: 59-78.
Elga, A. 2000. Self-locating beliefs and the Sleeping Beauty problem. Analysis 60: 143-147.
Lewis, D. 2001. Sleeping Beauty: reply to Elga. Analysis 61: 171-176.
Lewis, P. J. 2007. Quantum Sleeping Beauty. Analysis 67: 59-65.
Vaidman, L. and Saunders, S. 2001. On Sleeping Beauty controversy.
http://philsci-archive.pitt.edu/archive/00000324