Preliminary: Comments Invited Mar. 2018

The Sleeping Beauty Problem:

A Change in Credence?

Everything should be made as

simple as possible, but not simpler.

Attributed to Albert Einstein

Abstract. Adam Elga (2000) put forward an experiment: Sleeping Beauty (SB) is put to sleep. She will be awakened once if a coin toss is Heads and twice if Tails. A drug erases her memory of each awakening. When awakened, what should she believe the chances are the toss was *H*?

Two answers dominate the literature: 1/2 (the halfers) and 1/3 (thirders).

This paper makes two main points:

a) Both halfers and thirders write that SB believes 1/2 before going to sleep on Sunday. This is incorrect. If SB is a thirder on Monday, she will be a thirder on Sunday.

b) Conditional credences are used in Elga's paper to support the thirder case, but they are unnecessary. The case may be made much more simply.

The Sleeping Beauty (SB) controversy has been with us for several decades. Its main features are reviewed by Peter Winkler, who concludes (2017: 586): "I am under no illusions that controversy about its [the SB problem's] solution will ever entirely disappear."

The SB problem has inspired about 100 articles in academic journals. To simplify, I will concentrate on two: Elga's seminal article (2000) that inspired much of the controversy, and Winkler's review (2017).

1. The Experiment

Elga's article (2000: 143) posed the following experiment:

> Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads; once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking.
>
> When you are first awakened, to what degree ought you to believe that the outcome of the coin toss is Heads?

The experiment is explained on Sunday. SB may be in three possible positions when awakened: With heads (*H)*, there is a single awakening H_1, on Monday. If tails (*T)*, the first of two awakenings, T_1, occurs on Monday; the second, T_2, on Tuesday.

Throughout, SB remembers the details explained on Sunday, even though her memory of each wakening is erased.

To avoid confusion, note that *T* stands for tails, not Tuesday, with subscripts indicating the day. Note also that Elga states that the coin is fair, with an equal probability of *H* or *T*:

$$P(H) = P(T) = 1/2 \qquad (1)$$

If it's not fair, anything goes.

"Halfers" argue that, when awakened, SB believes there is a 50% chance the coin has[1] turned up *H*; "thirders" maintain that she believes there is only one chance in three.

This paper argues that several problems arise in the literature:

a) Does SB change her credence in *H* when she awakens?

This question creates fundamental difficulties for both sides; both maintain that SB has a credence of 1/2 before going to sleep on Sunday. But this is not so. If she is a thirder upon awakening, she will be a thirder on Sunday.

[1]Elga (2000: 144) considers two possibilities: the coin is flipped before SB is awakened on Monday, or after she goes back to sleep on Monday. He notes that it doesn't make any difference, and has the flip occurring after she goes back to sleep on Monday. To avoid grammatical tangles, I have the flip occur before the Monday awakening.

b) Elga's seminal paper uses conditional credences to support the thirder case, but they are unnecessary. The case may be made more simply.

c) The puzzle must be looked on two ways; it is important not to confuse them if the flip is *T.* The *whole experiment* must be considered, where both T_1 AND T_2 occur. But, we must also consider a *single awakening*. Each time SB is asked about the chances of *H*, she is at a single awakening. With a *T* flip, she can be in only one position, T_1 OR T_2.

Sections 2 and 3 below explain thirder and halfer positions. Section 4 presents Elga's thirder case. Section 5 explains my key point: if SB is a thirder upon awakening, she already has enough information to be a thirder before going to sleep on Sunday. There is no need for her to change her credence. Section 6 considers what happens if SB bets on the outcome of the flip.

2. The Thirder Case, Simply Put

A fair coin lands *T* half the time, and SB is quizzed at T_1 and T_2. Thus, *during a complete experiment*:

$$P(H_1) = P(T_1) = P(T_2) = 1/2 \qquad (2)$$

Each of the three outcomes is equally probable, at 50%, with *T* being "observed" twice as often as *H*.

When asked where she is during any single awakening, SB can be only in one of the three positions: H_1, T_1, or T_2. She cannot stick to Eq. 2. She must reduce the sum of the probabilities to one. (The probabilities in Eq. 2 add up to 150% because the three positions are not mutually exclusive; T_2 occurs each time that T_1 does. In a complete experiment, 1.5 awakenings occur on average.) The three positions are equally probable. Thus, she sees the probabilities *in a single awakening*:

$$P(H_1) = P(T_1) = P(T_2) = 1/3 \qquad (3)$$

Note once more that the issue is not whether the coin is fair. It is. The issue is one of reporting. There is a *bias in*

reporting, with SB responding to the question twice if *T*, but only once if *H*. The chances are 50-50 that a fair coin will come up *T*. But two thirds of the time that she is asked, the coin will have come up *T*. That is the basis for the thirder's case.

3. Where is Sleeping Beauty? A Halfer Case (and Rebuttal)

We can be led toward the halfer conclusion when starting, not with the whole experiment (as in Section 2), but with the question, "in what single position is SB when awakened, H_1, T_1, *or* T_2?"

There is a 50% chance the flip will be *H,* and H_1 will be observed. There is a 50% chance the flip will be *T*, and T_1 or T_2 will be observed. That is:

$$P(T) = P(T_1) + P(T_2) = 1/2 \qquad (4)$$

With the probability of T_1 or T_2 presumably being the same:

$$P(T_1) = P(T_2) = 1/2 \times 1/2 = 1/4 \qquad (5)$$

Hence :

$$P(H_1) + P(T_1) + P(T_2) = 1/2 + 1/4 + 1/4 = 1 \qquad (6)^2$$

SB is a halfer.

When the flip is *T*, we should keep in mind the distinction between what happens during a single SB awakening, at T_1 *or* T_2, and during a complete experiment, with SB awakening twice, at T_1 *and* T_2. In deriving Eq. 6, we have asked, "where is SB?" She can only be in one position; there is a single *T* awakening (during which she is asked her belief in the probability of an *H* flip). In a complete experiment, with two *T* awakenings, the *T* probabilities in Eq. 6 must be doubled. This leads us back to Eq. 2 and hence to Eq. 3. SB is a thirder.

Eq. 6 holds only if SB is awakened just once during the complete experiment, at one of the three possible positions. A coin is flipped; 50% of the time it comes up *H*, with SB

[2]In considering SB's credence that it is Tuesday, Winkler 2017, 584 notes that "the halfer answer is 1/4 [my Eq. 6], while the thirders claim 1/3 [my Eq. 3]. So what?" So, quite a bit. One (1/3) is a cornerstone of the the thirder case; the other (1/4) is part of the halfer case.

awakened at H_1; the other 50% of the time it comes up *T*, and SB is awakened at T_1 or T_2. Thereupon the experiment ends with the halfer declaring victory. But this contradicts Elga's experiment, where she is awakened twice with a flip of *T*.

This point may be expressed somewhat differently, defending Eq. 6 with a very peculiar experiment. SB is awakened twice with a *T* flip, but quizzed only once, at T_1 or T_2. During the other *T* awakening, she is put directly back to sleep without questioning. In this case, there are only two times at which she can be quizzed, once when the flip is *H* and once when *T*, with a 50% chance of each. Thus, she would be a halfer when asked. But this strange experiment seems pointless. If the flip is *T*, why awaken her twice but ask her only once? And it is not Elga's experiment, as seen most clearly when he considers repeated experiments (2000: 143).

Observe that, when we start with the whole experiment, in Section 2, we are drawn toward the thirder conclusion.

When we start by looking at where she is during a single awakening in this section, we are drawn toward the incorrect halfer conclusion.

This section has dealt with only part of the halfer case. The core of the case is considered in Section 5.

4. Where is Sleeping Beauty? Elga's Approach

In the paper that kicked off the debate, Elga makes a complicated thirder case, to show that she can change her credence from 1/2 on Sunday to 1/3 when awakening. He depends on two conditions.

Condition C1 (quoted from Elga 2000: 144)

> If (upon first awakening) you were to learn that the toss outcome is Tails, that would amount to you learning that you are either in T_1 or T_2. . . . Your credence that you are in T_1, after learning that the

> toss outcome is tails, ought to be the same as the conditional credence $P(T_1|T_1 \text{ or } T_2)$, and likewise for T_2.

Hence:

$$P(T_1) = P(T_2) \tag{7}$$

C1 is problematic. If SB learns that the toss is *T*, wouldn't she say, "Stop right there. You've let the cat out of the bag. I can answer your question. $P(T) = 1$; therefore $P(H)$ = zero. At least for this flip." Isn't C1 giving her new information? The issue of new information is considered further in fn. 3 and Section 5.

Things become more complex when we move to Condition C2 (quoted from Elga 2000: 145):

> If (upon wakening) you were to learn that it is Monday . . . your credence that the coin will land Heads . . . ought to be the same as the conditional

credence $P(H_1|H_1 \text{ or } T_1)$. So $P(H_1|H_1 \text{ or } T_1) = 1/2$, and hence

$$P(H_1) = P(T_1) \tag{8}$$

Combining Eq. 7 and 8, we get:

$$P(H_1) = P(T_1) = P(T_2) \tag{9}$$

When SB awakens, she is in one of these three equally probable positions. The probabilities must sum to one. Thus:

$$P(H_1) = 1/3 \tag{10}$$

Thus saith Elga.

We don't have to decide if Elga's logic, including conditional credences, is correct.[3] His complex derivation of key Eq. 9 is unnecessary. The simple statement of the thirder case in Section 2 suffices. Eq. 3 includes Eq. 9 and 10.

---

[3]Above, we questioned whether Elga violated the assumption of no new information with C1. The same question arises regarding C2, when Elga (2000: 145) writes that the belief change from 1/2 to 1/3 "is not the result of your receiving any new information — you [SB] were already certain that you would be awakened on Monday." Yes, but you weren't certain that, upon awakening, it would be Monday. Every time the experiment is run, SB wakes up on Monday. But not every time she wakes up is on Monday.

Conditional credences throw sand in our eyes. Very fine sand, to be sure, but sand nonetheless.

5. New Information? A Change in Credence?

The standard (but not universal) assumption in the SB controversy is that she is not provided new information after she goes to sleep on Sunday evening. Both sides agree that SB's credence is 1/2 on Sunday.

But this is not so; both sides are wrong. What is the core of the thirder case? That when SB is awakened, there are two chances in three the flip has been *T*, as reflected in Eq. 3. But on Sunday, SB *already knew that this would happen*. If a thirder when awakened, she should be a thirder on Sunday. *She needs no new information.*

This destroys the core of the halfer case. Winkler (2017: 580, ital. in original) writes that the halfer argues that

> before SB is put to sleep on Sunday, her credence that the fair coin will come up Heads is inarguably [!] 1/2. She knows she will be awakened, so when she inevitably is, *she has no new information*, and therefore, her credence in Heads cannot have changed.

But her mind doesn't have to change; her credence is already 1/3 on Sunday.[4]

Likewise, there are problems on the thirder side. Specifically, much of the thirder literature considers how she could change her credence without new information. These complicated arguments are unnecessary once we recognize she has enough information to be a thirder on Sunday. There is no need for a mysterious change in credence.

---

[4]As mentioned above (fn. 1), Elga has the coin flipped after she goes back to sleep on Monday. In describing the halfer position of Lewis (2001), Winkler (2017: 583) writes, "How can SB's credence in Heads be other than 1/2 if she knows the coin hasn't been flipped yet?" The answer: on Sunday, she knows that whenever it is flipped, there is a 1/2 chance of *T*, with two awakenings, and only one chance in three that the flip will have been *H* when she is awakened. Furthermore, she doesn't know whether the coin has been flipped or not. Her memory is erased. She doesn't know which of the three awakenings she is in. And Elga is correct in arguing that it doesn't matter if the coin is flipped before or after the Monday awakening.

Elga (2000: 146) notes that his argument — without any new information, SB changes her credence in $P(H)$ from 1/2 on Sunday to 1/3 upon being awakened on Monday — provides a counterexample to Bas Van Fraassen's Reflection Principle (1995: 19), which, simply put, says that if you get no new information, your credence should be the same now as it was yesterday. But SB has enough information to be a thirder on Sunday; the SB problem is consistent with the Reflection Principle.

When he summarizes the controversy, Winkler writes (2017: 581) that

> Philosophers are after much bigger game than the Sleeping Beauty problem itself. . . . How should [a] credence be updated with new information or the passage of time? Sleeping Beauty is a demanding test for any theory that addresses these questions, incorporating loss of consciousness, loss of memory, and absence of time indication.

But philosophers need not hack their way through the theoretical underbrush in search of prey; there is no need for SB to change her credence.

When he summarizes the controversy, Winkler writes (2017: 581) that

> Philosophers are after much bigger game than the Sleeping Beauty problem itself. . . . How should [a] credence be updated with new information or the passage of time? Sleeping Beauty is a demanding test for any theory that addresses these questions, incorporating loss of consciousness, loss of memory, and absence of time indication.

But philosophers need not hack their way through the theoretical underbrush in search of prey; there is no need for SB to change her credence.

Although "no new information" is the standard assumption, a number of writers explicitly depart from it. For example, White (2006) introduces a waking device that kicks in

once SB wakens and is about to go back to sleep again. Arntzenius (2003) has SB influenced by a vivid dream. Kim (2015: 1220) considers the case where SB wakes up with a large electronic calendar on the wall, telling her it's Monday — that is, in effect, Condition C2.

I do not pursue the "new information" issue further. Following Einstein's admonition, I have tried to keep the argument as simple as possible,[5] most notably by showing that Elga's conditional credences are unnecessary.

6. Betting

Another approach to the the SB problem, summarized by Winkler (2017: 580, ital. in original) is to ask her,

> upon each wakening, if she's willing to have $3 deducted from her bank account if the coin landed Heads, provided

[5]The extensive SB literature introduces a multitude of complications that I do not wish to address. Most notable, perhaps, are Lewis (2007), Peterson (2011), and Groisman, Hallakoun, and Vaidman (2013), who bring out the heavy artillery, introducing quantum theory into the controversy.

that $2 is *added* to her account if the coin landed Tails. . . .

As a thirder, SB should accept the bet. Her expectation is

$$1/3(-\$3) + 2/3(+\$2) \qquad (11)$$

which is greater than 0. She gains — assuming she is right in choosing the thirder position. But suppose she is a halfer. Her expected payoff (EP) is:

$$EP = 1/2(-\$3) + 1/2(+\$2) \qquad (12)$$

which is less than 0. She rejects the bet.

Clearly, this has gotten us nowhere. Each side uses its probabilities to reach the conclusion with which it started; a quick trip around a logical circle. But who is correct?

The halfer might immediately object that the thirder has unintentionally cheated, introducing a loaded coin that ends up *H* only one third of the time in expected payoff (my 11, from Winkler 580). This objection is considered in the puzzle below.

The thirder position should be based based, not on a loaded coin, but on a double observation if the coin lands *T*;

that is, SB wins \$2 at both T_1 and T_2. To make it clearer what is driving the thirder conclusion — a double winning rather than a loaded coin — we may restate SB's expected payoff, using a fair coin with a 50-50 chance of landing *H*:

EP = 1/2(-\$3) + 1/2(+\$2 +\$2) (13)

This seems like a strange bet: she gets a double payoff if she is correct in betting on *T*. (It is hard to imagine any other bet with such a double payoff for winning.) But the SB problem is indeed strange, to engage in understatement.

Suppose SB wants to bet. If the flip was *T* and she wakes a second time, on Tuesday, the coin must *not* be flipped a second time to see if she won. She would have an expected payoff of -\$0.50 on the second flip. The initial flip in the experiment must be used to determine whether she wins or loses throughout the experiment, upon each awakening. If she is only permitted to make one bet if the flip is *T* — at T_1 *or* T_2 — she gets only a single payoff (of -\$0.50), as seen in Eq. 12 (which is consistent with the peculiar experiment

discussed after Eq. 6). She should reject the bet. But this contradicts the "upon each wakening" clause in Winkler's statement of betting. Thus, betting as stated by Winkler supports the thirder case, provided that we stick to a single coin flip per experiment and a double payoff if the flip is *T*.

Observe, however, the discrepancy between Expected Payoff (11) — $0.33 — and the $0.50 of EP (13). The difference might seem trivial, but as noted in fn. 2, a small number can reflect a fundamental difference. The discrepancy leaves a puzzle.[6] Expected Payoff (13) leads toward the correct basis for the thirder conclusion; that is, two payoffs when the toss is *T*. Expected Payoff (11) suggests the thirders may be using an unfair coin.

---

[6]It is unclear where the puzzle comes from. In his summary of the controversy, Winkler (2017: 580) gives Expected Payoff (11) if SB is a thirder (suggesting a loaded coin?) But he then writes that "she ends up $4 ahead if the coin landed *T* and only $3 behind if it landed *H*," suggesting a double payoff for *T*, as in Expected Payoff (13).

For a more complicated analysis of betting, which uses the Dutch Book argument to support the thirder case, see Hitchcock (2004).

7. Concluding Comments and Summary

This paper comes down on the thirder side; there is only one chance in three that the flip has been *H* when she is awakened.

To reinforce this conclusion, consider an extreme example, with SB awakened 99 times with a flip of *T*. If SB is asked each time, the thirder becomes a one percenter. With a fair coin, a one percent chance of *H* might seem preposterous. But it is so, if we look not at the probability that a coin flip turns up *H* (50%), but rather, the probability the coin has turned up *H* when she is awakened. In 99 of 100 times, it is *T*.

This paper shows that if SB is a thirder when awakened, she is a thirder on Sunday. She has the same information on Sunday as when awakened: she knows that there is only one chance in three that, when she is awakened, the flip has been *H*. There is no need for a revision in credences. As a result:

a) The core halfer case collapses.

b) Much of the thirder discussion is overly complicated. It struggles with a non-existent problem, SB's mysterious change in credence.

The basic issue is really quite simple; the basic thirder argument was made in Section 2. But this is slippery stuff. It is easy to get things mixed up — as my earlier drafts illustrate to my embarrassment. I look at journalists with a combination of awe and terror: their first drafts are often published.

REFERENCES

Arntzenius, F. 2003. Some problems for conditionalization and reflection. Journal of Philosophy 199: 356–70.

Elga, A. 2000. Self-locating belief and the Sleeping Beauty problem. Analysis 60: 143–47.

Groisman, B., N. Hallakoun, and L. Vaidman 2013. The measure of existence of a quantum world and the Sleeping Beauty Problem. Analysis 73: 695-706.

Hitchcock, C. R. 2004. Beauty and the bets. Synthese 139: 405-20.

Kim, N. 2015. Titelbaum's theory of *De Se* updating and two versions of Sleeping Beauty. Erkenn 80: 1217-36.

Lewis, D. 2001. Sleeping Beauty: Reply to Elga. Analysis 61: 171-76.

Lewis, P. J. 2007. Quantum Sleeping Beauty. Analysis 67: 59-65.

Peterson, D. 2011. Qeauty and the books: a response to Lewis's quantum sleeping beauty problem. Synthese 181: 367-74.

van Fraassen, B. C. 1995. Belief and the problem of Ulysses and the sirens. Philosophical Studies 77: 7-37.

White, R. 2006. The generalized Sleeping Beauty Problem: a challenge for thirders. Analysis 66: 114–19.

Winkler, P. 2017. The Sleeping Beauty controversy. American Mathematical Monthly 124: 579-87.