The Sleeping Beauty Problem: How Analytic Philosophy Generates Unsolvable Puzzles

Joseph Wayne Smith, Saxon J. Smith, and N. Stocks

PRINCE FLORIMOND FINDS THE SLEEPING BEAUTY

(Wikipedia, 2023)

1. Introduction: Sleeping Beauty and Probabilities

The Sleeping Beauty problem was formulated in the mid-1980s (Zuboff, 1990; Elga, 2000), and predictably enough, it has been intensely debated by Analytic philosophers, and also by some mathematicians, with no clear resolution. The problem arises from an imaginary experiment that somehow gets ethical approval. Sleeping Beauty, for some reason best known to her, volunteers for a bizarre experiment on a Sunday. On that day she will be injected with a drug that will put her to sleep. Either once or twice during the experiment she will be awakened by the administrator, by another drug. Once awake, she will be interviewed, be told that it is Monday, then given yet another drug that causes amnesia, so that she will not remember awaking or anything that occurred during the interview. On this basis, Sleeping Beauty will not be able to determine what day it is, or whether she has been awakened at any time. Her mind, in this respect, is a complete blank.

The experimenters will toss a fair coin to determine the following:

(SB1) If the coin turns up heads, Sleeping Beauty will be awakened and only interviewed on Monday.

(SB2) If the coin comes up tails, Sleeping Beauty will be awakened and interviewed on Monday and Tuesday, and given the amnesia causing drug.

Whatever happens, Sleeping Beauty will be awakened on Wednesday, and the experiment concludes with no interview, but probably some peer reviewed psychology paper will be churned out. In any case, after the experiment concludes, Sleeping Beauty is asked: what is your credence that the coin landed heads?

2. The Problem of Solutions

The first main response, which is known as the "thirder" position, asserts that the probability of heads is 1/3, a position taken by the majority of philosophers addressing this problem (Elga, 2000; Dorr, 2002; Monton, 2002; Arntzenius, 2003; Hitchcock, 2004; Horgan, 2004). When Sleeping Beauty awakes, she knows that she is in one of these situations:

Heads/Monday: The coin landed heads and it is Monday.

Tails/Monday: The coin landed tails and it is now Monday.

Tails/Tuesday: The coin landed tails and it is now Tuesday.

Now, by the principle of indifference, her credence that it is Monday is equal to her credence that it is Tuesday, because there is no fact of the matter to distinguish these events. Hence:

(SB3) Pr (Monday|Tails) = Pr (Tuesday|Tails).

Therefore:

(SB4) Pr (Tails & Tuesday) = Pr (Tails & Monday).

On the other hand, if Sleeping Beauty believed that it was Monday and given that Pr (Tails|Monday) = Pr (Heads|Monday), then:

(SB5) Pr (Tails & Tuesday) = Pr (Tails & Monday) = Pr (Heads & Monday).

Since the three components of the equation (SB5) exhaust the possibilities, the probability of their sum is 1, and by the principle of indifference, the probability of each component is 1/3 (White, 2006: pp. 115-116). In short, for every "heads-waking, there are two tails-waking" (Weintraub, 2004: p. 8), so the probability is 1/3.

Contrary to all this, David Lewis argued for the "halfer" position (Lewis, 2001), that Sleeping Beauty's credence that the coin landed heads is ½. The reason is simply that, given before the experiment was conducted the probability of heads was ½, and since, as the problem is set up, she has no information about what happened while she was asleep, then she would still have a credence that the probability of heads is ½. As Bischoff summarizes it:

> After all, one could even flip the coin before sending Sleeping Beauty to sleep. By the experiment's design, she does not have any extra clues to the situation, so logically she should state the probability as ½. (Bischoff, 2023)

Before discussing the state of play with these purported solutions to the Sleeping Beauty problem, it is instructive to compare and contrast this problem with another probability teaser, the Monty Hall problem, where there is a similar structure and set of intuitions about probability which lead many people, including some mathematicians, astray. In this case, there is a game show (similar to *Let's Make a Deal*, originally hosted by Monty Hall), where there are three doors. Behind one door is a worthwhile prize, while the other two doors have no prize (or, some prefer a goat, but we know people who think goats are worthwhile prizes). The host knows what is behind the doors, and when you choose a door, he/she will open one door which he/she knows has no prize behind it; a different door from the one which is chosen is always opened. You are then invited to either stay with your original choice or switch. Is it advantageous to switch (Selvin, 1975)?

While many people including some mathematicians have initially said that there is now an equal-probability, ½ chance with the two doors, because one door has a worthwhile prize while the other has no prize, the received answer is to switch, which has a 2/3 probability of winning the prize, while staying with the original choice has only a 1/3 chance of winning. The 50-50 chance position would be true if the show host opened the door at random, which seems to be implicitly assumed, but falsely, by many people (Rosenthal, 2009: p. 36). However, this is not so, since the host will open a door that is dependent upon the choice originally made. Hence, the assumption of independence, while intuitively seductive, is false (Falk, 1992; Herbranson & Schroeder, 2010; Granberg, 2014). Initially you have a 1/3 chance of choosing the door with the prize. But when the host opens the door with no prize there is a 2/3 chance that the prize is behind the door that is other than the original choice, because the 1/3 probability "shifts" to that door; because there is a 2/3 chance that the prize was behind the two doors, and when one is eliminated, the probability still remains as 2/3.

Thus, a problem such as the Monty Hall probability teaser, is one in which intuition about a probability fails, but when the problem is given a mathematical analysis, the problem resolves itself easily enough (Rosenthal, 2009). Yet this is simply not so with the Sleeping Beauty problem. There does not seem to be any implicit intuition leading to mistaken probability, but rather, there are two seemingly compelling arguments for two different probabilities (Titelbaum, 2013).

3. Both Positions are Objectionable

The arguments for both the halfer and thirder positions have been subjected to a sustained critique in the literature. The situation is interesting, since supporters of each position advance arguments against the other side, which seem to hold even if the independent arguments refuted their position. This is like two Wild West gunfighters, in a fast draw showdown, each drawing and shooting the other.

Only a small sample of the literature can be mentioned here, but the reader can set out on an internet journey via Google Scholar to verify this claim for her/himself. Thus, for Ross, there is a major problem for thirders with a situation of inconsistency with the principle of countable additivity (Ross, 2010). And White argues that there is a problem for thirders based on a generalized Sleeping Beauty problem that halfers do not face (White, 2006).

Some think that the Sleeping Beauty problem might be tackled using the many worlds interpretation of quantum mechanics. Peter J. Lewis argues for a halfer position (Lewis, 2007), while Groisman, Hallakoun and Vaidman go for a thirder position (Groisman, Hallakoun & Vaidman, 2013). Who is right, and in what world? Well, it is "many worlds," after all.

Others, such as Garisto, see neither position as inherently right or wrong:

> [I]t is a matter of how we define "self"—we do not give an answer about which camp is "right" because they are each right given a reasonable set of assumptions. (Garisto, 2020)

Without spelling out Garisto's theory in detail, the upshot is that Sleeping Beauty assigns to heads 1/3 if "Beauty sees herself as being in all three observer moments and "1/2" if she sees herself as living in an H world or a T world." Still, we do not know what "world" Sleeping Beauty sees herself in, and perhaps at the end of the day, she does not know herself.

Bostrom argues that both the ½ and 1/3 views are wrong, because both positions have question begging assumptions (Bostrom, 2007). He considers the "extreme Sleeping Beauty" which we think is certain not to get ethics approval. Nevertheless, in this thought experiment, it is like the original problem only if the coin falls tails, Sleeping Beauty is awakened on one million subsequent days, and given the amnesia drug and put back to sleep. When awakened on Monday, what is her credence in heads? There, Pr (Heads) = 1/1,000,002, with the degree of support being proportional to the number of awakenings, which is counter-intuitive (Bostrom, 2007: p. 63). Bostrom writes:

> It implies that Beauty should not take the fact that she is currently awake as evidence that there are large numbers of awakenings. But it also implies that when Beauty discovers that it is currently Monday, she should not take this as evidence against the hypothesis that there will be many more awakenings in the future. (Bostrom, 2007: p. 74)

Bostrom also outlines another thought experiment, the Presumptuous Philosopher, which he believes shows that the 1/3 position is incorrect; we will not sketch it here. But according to a hybrid model, which Bostrom believes solves the problem, in an N-fold Sleeping Beauty problem, for N ≥ 1:

(SB6) Pr (HEADS) ≥ 1/3 & ½ ≥ Pr (HEADS) (Bostrom, 2007, 75).

So, what then is the probability of HEADS in the one fold version, which we are interested in? Bostrom says:

> [I]n the one fold version, it is *not* the case that one-third of all actual agent-parts of Beauty are in a heads-trial. There, either all are, or none. Moreover, in the one fold version, the total number of awakenings is strongly correlated with which hypothesis, HEADS or TAILS, is true. (Bostrom, 2007: p. 75)

Apparently then, there does not seem to be one probability for the one fold problem; we searched the pages and did not find it, and in accordance with (SB6) it could be any number n ε R greater than or equal to 1/3, but less than or equal to ½. We do not see how that solves the *original* Sleeping Beauty problem, where what is sought is a single probability from Sleeping Beauty.

This problem of the solutions also rules out a paraconsistent response, where it could be argued that if there were utterly compelling arguments for *both* the 1/3, and ½ positions, then both are correct. In the context of the logic-semantical paradoxes, Graham Priest has said:

> Here we have a set of arguments that appear to be sound, and yet end in contradiction. *Prima facie*, then, they establish that some contradictions are true. Some of the arguments are two and a half thousand years old. Yet despite intense attempts to say what is wrong with them in a number of logical epochs, including our own, there are no adequate solutions. […] trying to solve them is simply barking up the wrong tree: we should just accept them [the logico-semantical paradoxes] at face value, as showing that certain contradictions are true. (Priest, 2006: p. 83)

However, even the paraconsistency response fails, since in the case of the Sleeping Beauty problem, it is not the case that there are compelling arguments for both the 1/3 and ½ positions, but instead the case that both these positions fail because of sustained critical arguments in the literature.

4. Conclusion

The inevitable conclusion is that there is no existing satisfactory solution to the Sleeping Beauty problem. Moreover, given the level of intense scrutiny devoted to the problem, and the critical arguments against all proposed solutions, one is also justified in concluding that no solution will likely to be forthcoming. Therefore, this is an example—and we will argue elsewhere that there are many—of Analytic philosophy generating unsolvable problems for itself. If Sleeping Beauty, via the amnesia drug, does not retain any memories of awakenings, it might seem to people who are not Analytic philosophers, that she will not have any idea whatsoever about the probabilities in the problem, and thus will have no answer at all to the Analytic philosopher's question. That dismal result would not be published in a highly-ranked Analytic philosophy journal, but with the world the closest it has been to nuclear war since the Cuban missile crisis (Anonymous, 2023), the fact that Sleeping Beauty does not know what day it is, should not be a major concern for humankind at large. Nevertheless, the fact that Analytic philosophers have wasted so much time and energy on such absurd Scholastic insolubilia—indeed, may we be permitted to say, on such pseudo-philosophical bullshit (Frankfurt, 1988)?—merely for the self-interested purposes of fattening the lists of their publications and forwarding their professional academic careers, should be a major concern for anyone concerned about real philosophy.

REFERENCES

(Anonymous, 2023). Anonymous. "Nuclear War is Imminent—North Korea." *Global Village Space*. 16 August. Available online at URL = <https://www.globalvillagespace.com/nuclear-war-is-imminent-north-korea/>.

(Arntzenius, 2003). Arntzenius, F. "Some Problems for Conditionalization and Reflection." *Journal of Philosophy* 100: 356-370.

(Bischoff, 2023). Bischoff, M. "Why the 'Sleeping Beauty Problem' is Keeping Mathematicians Awake." *Scientific American*. 4 May. Available online at URL = <https://www.scientificamerican.com/article/why-the-sleeping-beauty-problem-is-keeping-mathematicians-awake/>.

(Bostrom, 2007). Bostrom, N. "Sleeping Beauty and Self-Location: A Hybrid Model." *Synthese* 157: 59-78.

(Dorr, 2002). Dorr, C. "Sleeping Beauty: In Defence of Elga." *Analysis* 62: 292-296.

(Elga, 2000). Elga, A. "Self-Locating Belief and the Sleeping Beauty Problem." *Analysis* 60, 2: 143-147.

(Falk, 1992). Falk, R. "A Closer Look at the Probabilities of the Notorious Three Prisoners." *Cognition* 43, 3: 197-223.

(Frankfurt, 1988). Frankfurt, H. "On Bullshit." In H. Frankfurt, *The Importance of What We Care About*. Cambridge: Cambridge Univ. Press. Pp. 117-133.

(Garisto, 2020). Garisto, R. "How to Select Observers." *Physical Review Research* 2: 033464.

(Granberg, 2014). Granberg, D. *The Monty Hall Dilemma: A Cognitive Illusion Par Excellence*. Lumad/CreateSpace.

(Groisman, Hallakoun & Vaidman, 2013). Groisman, B. Hallakoun, N. & Vaidman, L., "The Measure of Existence of a Quantum World and the Sleeping Beauty Problem." *Analysis* 73, 4: 695-706.

(Herbranson & Schroeder, 2010). Herbranson, W. T. & Schroeder, J., "Are Birds Smarter than Mathematicians? Pigeons (*Columba livia*) Perform Optimally on a Version of the Monty Hall Dilemma." *Journal of Comparative Psychology* 124, 1: 1-13.

(Hitchcock, 2004). Hitchcock, C. “Beauty and the Bets.” *Synthese* 139: 405-420.

(Horgan, 2004). Horgan, T. “Sleeping Beauty Awakens: New Odds at the Dawn of the New Day.” *Analysis* 64: 10-21.

(Lewis, 2001). Lewis, D. “Sleeping Beauty: Reply to Elga.” *Analysis* 61, 3: 171-176.

(Lewis, 2007). Lewis, P. J., “Quantum Sleeping Beauty.” *Analysis* 67, 11:, 59-65.

(Monton, 2002). Monton, B. “Sleeping Beauty and the Forgetful Bayesian.” *Analysis* 62: 47-53.

(Priest, 2006). Priest, G. *Doubt Truth to be a Liar*. Oxford: Oxford Univ. Press.

(Rosenthal, 2009). Rosenthal, J.S. “A Mathematical Analysis of the Sleeping Beauty Problem.” *Mathematical Intelligencer* 31, 3: 32-37.

(Ross, 2010). Ross, J. “Sleeping Beauty, Countable Additivity and Rational Dilemmas.” *Philosophical Review* 119, 4: 411-447.

(Selvin, 1975). Selvin, S. “On the Monty Hall Problem (Letter to the Editor).” *American Statistician* 29, 3: 134.

(Titelbaum, 2013). Titelbaum, M. “Ten Reasons to Care about the Sleeping Beauty Problem.” *Philosophy Compass* 8, 1: 1003-1017.

(Weintraub, 2004). Weintraub, R. “Sleeping Beauty: A Simple Solution.” *Analysis* 64, 1: 8-10.

(White, 2006). White, R. “The Generalized Sleeping Beauty Problem: A Challenge for Thirders.” *Analysis* 66, 2: 114-119.

(Wikipedia, 2023). Wikipedia. “Sleeping Beauty.” Available online at URL = <https://en.wikipedia.org/wiki/Sleeping_Beauty>.

(Zuboff, 1990). Zuboff, A. “One Self: The Logic of Experience.” *Inquiry* 33, 1: 39-68.