
The Sleeping Beauty Problem

Pallas-Athena Cain

Abstract. The Sleeping Beauty Problem is a puzzle that has been stumping philosophers and mathematicians for decades. The complexities lie in the fact the problem itself covers topics like decision theory, epistemology, self vs nonself-locating information, indifference principles and many more ([5][Titelbaum], 1003). In this article, I will give a brief introduction to the problem itself, why it is so complex, and the various arguments mathematicians and philosophers alike have tried to use to come to a conclusion on this difficult problem.

1. THE SCENARIO

The most prevalent version of the problem goes like this: Imagine Sleeping Beauty consented to be a part of a sleep experiment. In this experiment, the scientists tell Sleeping Beauty that they are going to put her to sleep on Sunday and flip a coin in a fair coin toss. If the coin lands heads then they wake her up Monday morning and ask her what is her credence that the coin landed on heads. However, if the coin lands on tails the scientists will wake her up Monday ask her the question, wipe away her memory of waking up, put her back to sleep, and wake her up again on Tuesday morning and ask her the same question. Whenever Sleeping Beauty wakes up, she will be asked the question about the chance that the coin landed on heads, however, she still does not know what day it is. Our job is to find what is the most logical response for Sleeping Beauty to respond.

One side makes the case that she will respond with a 1 in 2 chance of the coin landing heads while others make the case that she will respond with a 1 in 3 chance. Now, we can go into both lines of reasoning, the main points of contention, and why this problem has been up for debate for decades.

2. INTRODUCTORY ARGUMENTS

Argument One: $\frac{1}{2}$

The instinctual response to this problem for many people is 50%. It is a fair coin flip after all. Our sample space includes two different outcomes for the coin, heads and tails. Our event is when the coin lands on heads. Using the event space formula we can calculate the probability of the coin landing on heads.

$$\text{Sample Space } (\omega) = H, T$$

$$\text{Let Event } (E) = H$$

$$\frac{\# \text{ of items in } E \text{ subset}}{\# \text{ of items in } \omega} = \frac{1}{2}$$

We can also use this to find the probability that Sleeping Beauty is in either tails scenario, tails and waking up on Monday and tails and waking up on Tuesday. Since

there are two times she awakes when the coin lands on tails we can split the $\frac{1}{2}$ chance she is in a tails scenario between the two possible awakenings when a tail occurs leaving a $\frac{1}{4}$ th chance she is in either tails scenario ([4][Wireless Philosophy]).

This chart represents the distribution of Sleeping Beauty's credence that she is in any of the three awakenings based on the $\frac{1}{2}$ argument.

–	Monday	Tuesday
Heads	50%	Asleep
Tails	25%	25%

This is one-way people look at it however some move the focus onto how many awakenings she has and hence they come to a different result.

Argument Two: $\frac{1}{3}$

The second answer to this problem lies in using relative frequencies ([5][Titelbaum], 1004). There are three possible scenarios for Sleeping Beauty to wake up to ([1][Elga], 2). First, the coin landed heads and it is Monday (H1). Second, the coin landed on Tails and it is Monday before the memory wipe (T1). And last, the coin landed on tails and it is Tuesday (T2). For every 1 head awakening, there are 2 tails awakenings. This revelation shifts our equation from a sample space of 2 to a sample space of 3.

$$Sample\ Space(\omega) = H1, T1, T2$$

$$Let\ Event(E) = H1$$

$$\frac{\# of\ items\ in\ E\ subset}{\# of\ items\ in\ \omega} = \frac{1}{3}$$

Overall, the waking up scenario appears to have one heads scenario for every two tails scenario hence $\frac{1}{3}$ chance that the coin landed heads. We can also visualize this in a chart similar to the first argument.

–	Monday	Tuesday
Heads	33%	Asleep
Tails	33%	33%

3. EVIDENCE AND COUNTERARGUMENTS

Evidence 1: Epistemic Probability $\frac{1}{3}$

The main point of contention throughout this problem is basing probability around the available evidence, the epistemic probability ([2][Horgan], 13). When analyzing the epistemic probability of a problem we consider the subject's credence or rational degree of belief ([2][Horgan], 13).

The epistemic agent in the Sleeping Beauty problem is the loss of temporal location information or the loss of information about what day it is ([2][Horgan], 14). Additionally, we have to consider the possibility of a mishap occurring and Sleeping Beauty's memory of waking up the day previously being erased.

The $\frac{1}{3}$ argument can be strengthened by the existence of possible memory loss. Horgan states that the mishap alone brings to light the idea that the prior probability is not the same as the current probability ([2][Horgan], 17). If we say the coin toss occurred before Sleeping Beauty was put to sleep, she knew the coin toss was fair and so her answer was $\frac{1}{2}$. However, after she wakes up her answer would change to $\frac{1}{3}$ because the possible mishap increases the odds of tails occurring.

This argument relies on the fact that her answer should consider that all three situations look exactly the same from her perspective. Because of the memory mishap that occurs when the coin is flipped tails, Sleeping Beauty has no possible way to tell what day she is waking up on ([2][Horgan], 10). We can argue the fact that after Sleeping Beauty awakens in any of the instances she has lost her temporal location, or where she is at relative to the time outside of her being, and therefore can not decipher the difference between any of the three awakenings. From her perspective, they are equivalent awakening scenarios and she will treat it as such.

Counterargument One: Epistemic Probability $\frac{1}{3}$

The main issue with the $\frac{1}{3}$ viewpoint based on epistemic probability is that it assumes that Sleeping Beauty's answer should change based on when the coin itself was tossed and should change when she awakens. Like before if we say the coin toss occurred before Sleeping Beauty was put to sleep, her answer would be $\frac{1}{2}$. However, even if the coin was tossed after she fell asleep, sleeping beauty knows it was a fair coin toss. Her answer should not change based on when the coin itself was tossed because it has no impact on the experiment itself other than it has to be anytime before she is awakened ([3][Lewis], 173). With that in mind, we can also look into whether or not it is logical for her to change her answer when she wakes up.

It is understood that Sleeping Beauty gains no information between the moment she goes to sleep and the moment she wakes up. There was no change in the experiment itself. She only needs to answer her credence on the coin toss that occurred. Wherever she is temporally, asleep or awake, is irrelevant to the outcome of the coin toss itself ([1][Elga], 4). She could be in any of the awakenings and it would not change what she knows about the coin toss is fair.

Evidence Two: Long-run experiment

Another way to come to a result is to imagine this process was repeated 1000 times in a long-run experiment ([5][Titelbaum], 1004). Heads is expected to appear roughly 500 times and will correspond to 500 awakening scenarios. Then, the other 500 times tails will be expected. Because every time tails is the result generates two awakening scenarios, she will wake up 1000 times due to tails being flipped. This gives the proportion of 500 head awakenings to 1000 tails awakenings hence $\frac{1}{3}$ chance of the coin being flipped heads.

Counterargument Two: Awakening Logic

The logic of the $\frac{1}{3}$ side initially relies on the fact that there are exactly two tails awakenings. We can see the stability of this logic if we change the scenario slightly. Instead of waking up one time on Tuesday, we could wake her up 10000 times on Tuesday ([4][Wireless Philosophy]). We can use the same process of erasing her memory between awakenings. This means that there are now 10002 possible awakening sce-

narios. One the coin landed heads and it is Monday (H1). Second, the coin landed on Tails and it is Monday before the memory wipe (T1) and the 10000 times for the coin landed on tails and it being Tuesday (T2-10000). The sample space for the problem has now increased to 10002. This means that the probability of heads is now 1 in 10002. However, the coin toss itself was fair. The new probability is extremely biased toward the outcome being tails ([4][Wireless Philosophy]).

4. CONCLUSION

Whether you are a proponent of the $\frac{1}{2}$ or the $\frac{1}{3}$ approach, both have logical arguments and counterarguments which leads many to look for the source of the disagreement elsewhere. Some think the issue stems from the vagueness of the question. Such as “Are you asking sleeping beauty the chance that the coin came up heads or the chance she is going to be asked the question again?”. Some figure the issue is what the individual chooses to set as the focus for their answer. For instance they ask questions like “Is the focus on the timeline she is in or the day?”. The initial question itself is left extremely open to interpretation and just one shift in how you look at it can completely change your result. The verdict is still left undecided for the Sleeping Beauty Problem and will remain that way as long as there are so many variations in interpretation.

REFERENCES

1. Adam Elga. Self-locating belief and the sleeping beauty problem. *Analysis*, 60(2):143–147, 2000.
2. Terry Horgan. Sleeping beauty awakened: New odds at the dawn of the new day. *Analysis*, 64(1):10–21, 2004.
3. David Lewis. Sleeping beauty: Reply to elga. *Analysis*, 61(3):171–176, 2001.
4. Wireless Philosophy. Philosophy - epistemology: The sleeping beauty problem [hd].
5. Michael G. Titelbaum. Ten reasons to care about the sleeping beauty problem. *Philosophy Compass*, 8(11):1003–1017, 2013.