

Alexander Janiak

12/09/2022

### The Sleeping Beauty Problem: The Perspective Argument

The Sleeping Beauty Problem is a thought experiment in the field of philosophy that has sparked much debate and discussion throughout several disciplines. At its core, the experiment raises fundamental questions about the nature of knowledge and belief, as well as ways in which probabilities are affected by our knowledge and lack thereof. In this essay, I will divide the problem into different perspectives and put forth a potential solution by arguing that certain perspectives are better “fit” than others.

The Sleeping Beauty Problem can be described as follows: Sleeping Beauty (“Beauty”) participates in an experiment in which the researchers plan to put her to sleep for two days starting on Sunday. After she is put to sleep, the researchers flip a coin and, depending on the outcome, wake her up briefly either once or twice. If the coin lands heads, they wake her up just on Monday. If the coin lands tails, they wake her up on both Monday and Tuesday. After each waking, they give her a memory-erasure drug that causes her to forget whether she had awoken before. When she is first woken up, to what degree of belief – or *credence* – should she have that the outcome of the coin toss is heads?

Within this essay, the credence that a proposition is true is represented numerically on a scale from 0 (certainty of falsehood) to 1 (certainty of truth). Furthermore, I assume that the coin the researchers use is fair (equal chance of landing heads or tails) and thus, accept the generally undisputed opinion that Beauty’s credence for the coin landing Heads on Sunday should be  $\frac{1}{2}$  (Titelbaum).

So what should Beauty believe? There are 2 main camps that people fall into: the “halfers” and the “thirders”. The “halfers” believe that Beauty’s credence should be  $\frac{1}{2}$  when she awakens. This belief rises from the idea that she knows that the probability of a coin landing heads is  $\frac{1}{2}$  and she learns nothing new about the flip when she wakes up. Specifically, the “halfers” employ the *Reflection Principle* – which states that if you know that you will have some credence later, then you should have that credence now – and the *Principal Principle* – which states that credences should reflect real-life probabilities (Peterson).

On the other hand, the “thirders” believe that Beauty’s credence should be  $\frac{1}{3}$ . One way to understand this is through the *Indifference Principle* – which states that if there are  $n$  different possibilities, we should be indifferent towards each of them and assign a credence of  $\frac{1}{n}$  towards each of them occurring (Peterson). Applying this principle to Beauty, there are 3 different possible states she could be in when she wakes up: the coin landed heads and it is Monday, the coin landed tails and it is Monday, and the coin landed tails and it is Tuesday. Thus she should assign her credence to be  $\frac{1}{3}$  because there are three possibilities she could be in and the coin landing heads is only one of them.

This discrepancy between the “halfers” and “thirders” has led to a lot of discussion regarding the consequences of each approach and potential reasons for their acceptance – but no concrete solution has been accepted. In the following argument, I propose that the “halfers” and “thirders” are just different perspectives that yield different credences for Beauty. In other words, Beauty’s situation can be viewed through different lenses that yield different potential credences. Furthermore, I use this approach to show that Beauty should switch to the “thirder” perspective when she first awakens because she learns of a more “fine-grained” perspective and thus ought to have a credence of  $\frac{1}{3}$  when she wakes up.

To understand how different perspectives on Beauty's situation could coexist while still resulting in different credences, imagine a scenario very similar to the Sleeping Beauty experiment, except this time, the experiment is led by quantitative researchers. When Beauty is woken up, she is asked to bet on whether the coin was flipped heads. This means that if she correctly predicts which face the coin lands on, she wins the bet. Beauty's goal is to correctly bet on which face the coin will land – attempting to win the bet. Which face should Beauty bet on when she first awakens?

In this situation, it would make sense for Beauty to align her credence that the coin will land on heads with the odds of winning the bet if she were to bet on heads. If her credence that the coin will land heads is  $\frac{1}{3}$ , then she should almost always bet on tails to maximize her chance of winning the bet. In fact, this betting example is logically equivalent to the original Sleeping Beauty Problem. The difference is that it provides a unique framework to think about the problem and a different lens for other agents – like an observer – to determine their credences.

First, let's evaluate the “halfer” perspective in terms of the betting scenario. Before Beauty goes to sleep on Sunday, it's easy to see that she should have a credence of  $\frac{1}{2}$ . The coin has two possible scenarios per experiment: either it lands heads or tails. Before she goes to sleep, she knows it is Sunday and that if she bets now, she has a  $\frac{1}{2}$  chance of winning. This perspective can be described as an observer's perspective because if there was an observer in the experiment – separate from the researchers and not present for the awakenings or the coin flip – they should bet on heads half of the time. To put it differently, the observer's perspective is part of the “halfer” perspective and views the experiment as a sort of “black box”. Beauty is put to sleep on Sunday, the coin is flipped during the experiment, and then she wakes up after it has concluded.

The key is that the observer does not experience the experiment. The “thirder” perspective, however, uses the lens of Beauty— who has awakened. Imagine that Beauty awakens and decides to bet on heads. Out of the 3 possible scenarios where she can be awoken, she loses the bet in  $\frac{2}{3}$  of the scenarios. This becomes more obvious when the number of times Beauty is woken up increases. For example, imagine instead that when the coin lands tails, Beauty is woken up 99 times instead of 2, and there are 100 possible states for her to be in. One state where she is awoken only once when the coin lands heads, and 99 states where she is awoken when the coin lands tails. Thus, when she first wakes up, she can be in any of the 100 states, and when she bets on heads, she loses the bet in  $\frac{99}{100}$  of those states. The main point here is that this perspective comes from Beauty’s knowledge that she has awoken. She knows she is within one of these states, but unsure of which one, and should therefore rely on the Indifference Principle.

When framed using an observer's perspective and Beauty’s perspective in the betting scenario, It can be seen that Beauty’s perspective includes something that the observer’s perspective does not. It includes Beauty’s experience of waking up, which ought to change her perspective from one that considers the experiment as a whole to one that focuses on the separate states within the trial. Why might this be?

Consider a similar use of the Indifference Principle by “halfers”. If the betting odds are viewed using enumerated trials, then Beauty should bet on heads half the time. This is because if the experiment is run 1000 times, it should be expected that 500 of the trials have a coin flip resulting in heads. Such a perspective allows for Beauty to have a credence of  $\frac{1}{2}$  while still allowing for the use of the Indifference Principle (Titelbaum). The fundamental difference between the Indifference Principle in this perspective and the “thirder” perspective is the reference class. In statistics, a reference class is a group of objects or events used as a basis for

calculating probabilities. In the lens of a “halfer”, Beauty’s reference class is enumerated trials, but while in the “thirder” scenario, her reference class are the different states within the trials. The reason one should favor the “thirder” scenario is because it utilized a more “fine-grained” approach.

To see why “fine-grained” approaches are more favorable, consider “rolling above a 4” and “rolling a 4 or below” on a six-sided fair die. One reference class can be found by being indifferent to these options and assigning a credence of  $\frac{1}{2}$  to rolling a 5 or a 6 on a fair die: either you roll above a 4 or you do not. It's obvious that this is not correct because there is more “fine-grained” information to the die – there are actually 6 different possible roll options. Two of these options are contained within “rolling above a 4”. Thus your credence for “rolling above a 4” should be  $\frac{1}{3}$  (Peterson).

Therefore, because the “thirder” perspective is more “fine-grained” than the “halfer” perspective using the Indifference Principle, Beauty should favor the “thirder” perspective. In general, Beauty should always switch from the observer (“halfer”) perspective to Beauty’s (“thirder”) perspective whenever she wakes up because she is able to reframe the experiment from the trial to the different states within it. She experiences these fine-grained possibilities and uncovers the “black-box” that is overlooked by the “halfers” – which use the experiment as a whole for their reference class.

A direct attack to this argument is that it seems to violate the Reflective Principle: Beauty has a credence of  $\frac{1}{2}$  towards the coin landing heads on Sunday, but then switches her credence to  $\frac{1}{3}$  when she wakes up – without gaining any new additional information. Within this argument, however, Beauty gains new spatial-temporal information that she has awoken and it is no longer Sunday. She discovers she must be within one of the potential states. In other words, she gains

the perspective of being within one of the states (Beauty's perspective) and as a result changes her credence to reflect that. Thus, due to Beauty gaining new information, she does not violate the Reflective Principle.

Given that the argument has Beauty update her credences due to new spatial-temporal information, another potential response is to say this argument fails the *Relevance-Limiting Thesis*. The Relevance-Limiting Thesis states that an agent who learns only centered information between two points in time should not change their credences regarding uncentered matters. Centered information is information regarding one's identity and spatio-temporal location in the world. Uncentered information, on the other hand, is information regarding one's environment independent of oneself (Titelbaum). In the Sleeping Beauty Problem, the flip of the coin is uncentered information while Beauty's knowledge that she has awoken is centered. Thus, Beauty can't possibly change her credence regarding the coin flip just from knowledge that she has awoken and is within one of the states. While problematic, the Relevance-Limiting Thesis has been shown to be false in cases other than the Sleeping Beauty Problem and thus the relation between uncentered and centered information remains undiagnosed (Titelbaum).

Another rather interesting counter could be to argue against a potential – somewhat disturbing – consequence of the Indifference Principle. Suppose you receive a letter saying that yesterday, someone had made 1000 copies of the individual you thought of as yourself. According to the Indifference Principle, you should have a credence of  $\frac{1000}{1001}$  regarding being one of the copies. Although it seems unsettling – given it is possible to make a copy of oneself with the same memories – this would seem correct (Titelbaum). Similar to the Sleeping Beauty Problem, after you receive the letter, you gain the knowledge that you are within one of the different possible states. There are 1000 states where you read the letter and are a copy, and only

1 state where you are not a copy. Because a copy's memories are the same as the original individual, it would be impossible, using centered information, to determine whether you are a copy. Thus, you should switch your perspective to a "fine-grained" approach that considers all the possible states and thus gives you a high confidence that you are a copy. The argument that this seems disturbing is an appeal to emotion that is contingent less on the Indifference Principle and more on the ability to create a copy of oneself.

The Sleeping Beauty Problem is a thought-provoking hypothetical experiment that has created widespread debate and questioned the foundations of knowledge and belief. Many arguments have been made to try to discredit the "halfers" or "thirders" but have proved unfruitful because both the "halfers" and "thirders" build their arguments on fundamental principles of probability. One would have to discredit those principles to prove the different arguments incorrect. Instead, a separate argument can be made that the "halfer" and "thirder" are just different perspectives of looking at the same situation which yield different credences for Beauty. The "halfer" perspective looks through the lens of an observer and views the experiment as a whole – disregarding Beauty's experience and just focusing on the coin flip. The "thirder" perspective, on the other hand, uses the lens of Beauty and considers the various states that she might take on within the experiment. The fundamental difference between these two perspectives is that they view the problem through different reference classes. Furthermore, because more "fine-grained" classes should be favored, Beauty ought to be a "thirder" when she first wakes up. This argument also uses Beauty's new knowledge that she is within one of the states, to allow her to change her credences without violating the Reflective Principle. It does however, fail the Relevance-Limiting Thesis because this new knowledge is uncentered and, according to the thesis, should not influence her credence on the centered information regarding the coin flip. The

Relevance-Limiting Thesis, however, has been proven false and thus fails to break the argument.

Therefore, this argument in favor of the “thirders” holds true.



Works Cited

Peterson, Daniel. "The Sleeping Beauty Problem." *1000-Word Philosophy: An Introductory*

*Anthology*, 24 June 2022,

<https://1000wordphilosophy.com/2019/06/15/the-sleeping-beauty-problem/>.

Titelbaum, Michael. *Ten Reasons to Care About The Sleeping Beauty Problem*.

<https://onlinelibrary.wiley.com/doi/abs/10.1111/phc3.12080>.