# The Halfers are right, but many were anyway wrong: Sleeping Beauty Problem

Minseong Kim

**Abstract.** In this paper, I will examine the representative halfer and thirder solutions to the Sleeping Beauty Problem. Two solutions give different answers for the probability of today being Monday that the sleeping beauty should rationally assign. Then by examining the definition of events, it is concluded that the representative thirder solution is wrong and the halfers are right, but that the representative halfer solution also contains wrong logical arguments.

## 1. Introduction: Sleeping Beauty Problem

The description of Sleeping Beauty Problem appears in Elga (2000), and is given the following:
"Some researchers are going to put you to sleep. During the two days that your sleep will last (Monday, Tuesday), they will briefly wake you up either once (only on Monday) or twice (both Monday and Tuesday), depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will ask you, the sleeping beauty, on what probability you will assign to the outcome of the coin toss turning out to be the Heads. Then they will put you, the sleeping beauty, to back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?"

## 2. Introduction: The Thirder - Adam Elga

Elga (2000)'s argument can be summarized as the following:
From now on, let $P(head) = 1/2$ be the unconditional probability that a fair coin toss will produce head. Therefore, $P(tail) = 1/2$.
Given that the result of the coin toss is tail, the probability that the sleeping beauty wakes up on Monday and the probability that the sleeping beauty wakes up on Tuesday should not be different. Therefore,

$$P(Monday|tail) = P(Tuesday|tail)$$

As such (by Bayes' rule),

$$P(Monday \cap tail) = P(Tuesday \cap tail)$$

As the result of coin toss only affects what happens on Tuesday (awake, not awake),

$$P(tail|Monday) = P(head|Monday)$$

$$P(Monday \cap tail) = P(Monday \cap head)$$

Thus,

$$P(Monday \cap head) = P(Monday \cap tail) = P(Tuesday \cap tail)$$

As $P(Monday \cap head) + P(Monday \cap tail) + P(Tuesday \cap tail) = 1$,

$$P(Monday \cap head) = \frac{1}{3}$$

Therefore, when first awakened, the sleeping beauty should assign 1/3 to the probability that the outcome of the coin toss is Heads.
The followings are not directly in Elga (2000), but they will aid our discussions:

$$P(Monday) = P(Monday \cap head) + P(Monday \cap tail) = \frac{2}{3}$$

$$P(Tuesday) = \frac{1}{3}$$

Let the unconditional probability that the sleeping beauty is only awakened on Monday be $P(EM)$ ($EM$ represents experiment on Monday only) from now on. $P(EM) = P(head)$ tautologically. $P(ET) = P(tail)$, where $ET$ represents experiment on Tuesday also - meaning that the sleeping beauty is awakened also on Tuesday.
Solving the equations of the following:

$$P(Monday) = P(Monday|EM)P(EM) + P(Monday|ET)P(ET)$$

$$P(Tuesday) = P(Tuesday|EM)P(EM) + P(Tuesday|ET)P(ET)$$

$$P(Monday|EM) + P(Tuesday|EM) = 1$$

$$P(Monday|ET) + P(Tuesday|ET) = 1$$

where $P(Tuesday|EM) = 0$ by the settings of the experiment produces

$$P(Monday|EM) = 1$$

$$P(Tuesday|EM) = 0$$

$$P(Monday|ET) = \frac{1}{3}$$

$$P(Tuesday|ET) = \frac{2}{3}$$

## 3. Introduction: The Halfer - David Lewis

Lewis (2001)'s argument can be summarized as the following:
Because no new information has been presented to the sleeping beauty, when awakened, the sleeping beauty should assign $1/2$ to the probability that the outcome of the coin toss is Heads. Because $P(head) = 1/2 = P(head \cap Monday)$, $P(tail) = 1/2 = P(Monday \cap tail) + P(Tuesday \cap tail)$ and $P(Monday \cap tail) = 1/4$. Therefore,

$$P(head|Monday) = \frac{1/2}{(1/2 + 1/4)} = \frac{2}{3}$$

$$P(tail|Monday) = \frac{1}{3}$$

The following is not included in Lewis (2001), but deriving them will aid our discussions:

$$P(head|Monday) = \frac{P(head \cap Monday)}{P(Monday)} = \frac{2}{3}$$

$$\frac{1}{2} = \frac{2}{3}P(Monday)$$

$$P(Monday) = \frac{3}{4}$$

$$P(Tuesday) = \frac{1}{4}$$

Also,

$$P(Monday) = P(Monday|EM)P(EM) + P(Monday|ET)P(ET)$$

$$P(Tuesday) = P(Tuesday|EM)P(EM) + P(Tuesday|ET)P(ET)$$

$$P(Monday|EM) + P(Tuesday|EM) = 1$$

$$P(Monday|ET) + P(Tuesday|ET) = 1$$

produces

$$P(Monday|EM) = 1$$

$$P(Tuesday|EM) = 0$$

$$P(Monday|ET) = \frac{1}{2}$$

$$P(Tuesday|ET) = \frac{1}{2}$$

## 4. But is the event being defined correctly?

It is noticeable that two representative approaches produce different answers for $P(Monday)$ and $P(Tuesday)$. By examining these answers, we may actually see which argument is right. I will argue that both approaches are wrong, but in the end halfers are right. Let me slightly change the question, but this change should not really affect the experiment. Instead of not just waking up the sleeping beauty, the sleeping beauty is told before the experiment is carried out that if the coin toss turns out to be head, the sleeping beauty will be awakened on Monday and be killed. If the coin toss turns out to be tail, then the sleeping beauty will be brought back to sleep with the drug and that be awakened on Tuesday. On Tuesday, the sleeping beauty will just be interviewed and not be killed in any case.
Now let us think about the events. Because the number of events we are considering is considered finite, events can safely be used.
Now $P(tail) = P(survive)$ and $P(head) = P(dead\ on\ Monday)$, unconditionally speaking. But we have two different possible sub-events for the event *survive*: $survive \cap Monday = survived\ on\ Monday$ and $survive \cap Tuesday = survived\ on\ Tuesday$. Should we consider these two events separately?
The answer, I believe is no. This is because these two events are not actually different. If I survive on Monday (represented by conditional $p$), then I will survive on Tuesday (represented by conditional $q$). If I am seen as surviving on Tuesday ($q$), I should have survived on Monday ($p$). $p \to q$ and $q \to p$, therefore $p \leftrightarrow q$. Invoking Kolmogorov probability theory and representing $p$ and $q$ as sets, $p$ and $q$ have to be the same set. Therefore, these two sub-events are actually tautological to each other. It is wrong to consider them as separate two sub-events that form the event *survive*. The only event that exists is *survive*, not $survived\ on\ Monday$ and $survived\ on\ Tuesday$.

This analogy explains where the thirder representative approach went wrong: they thought that $Monday \cap tail$ and $Tuesday \cap tail$ are actually separate sub-events that form the event $tail$. By applying the analogy above, it is shown that these sub-events are not actually separate sub-events, so while $P(Monday|tail) = P(Tuesday|tail)$ is true, this does not mean that $P(Monday \cap head) + P(Monday \cap tail) + P(Tuesday \cap tail) = 1$. What actually happened is double-counting. It should have been just $P(Monday \cap head) + P(tail) = 1$, where $P(tail) = P(Tuesday \cap tail) = P(Monday \cap tail)$, because these three events/sub-events are actually referring to the same event.

We can now see that the case is back to basic coin toss issue, and $P(head) = 1/2$ will be what the sleeping beauty responds when asked for the probability of the coin toss outcome.

But the representative halfer approach also has the same "event" clarification issue. Would there be any consequence that is actually wrong? For this, one only needs to look at $P(Monday)$ and $P(Tuesday)$.

For the representative halfer approach, $P(Monday) = 3/4$, while $P(Tuesday) = 1/4$. I will argue that for these questions, the thirder approach is right, that $P(Monday) = 2/3$ and $P(Tuesday) = 1/3$.

Let us go back to the analogy above. For Monday, the sleeping beauty either survives or dies, that is $P(Monday|Monday) = 1 = P(Monday \cap survive|Monday) + P(Monday \cap dead|Monday)$ and $P(Monday \cap survive|Monday) = 1/2$, $P(Monday \cap dead|Monday) = 1/2$. But if the event $Monday \cap survive$ happens, then $Tuesday \cap survive$ tautologically follows. As thus, the ratio between $P(Monday)$ and $P(Tuesday)$ is $P(Monday) : P(Tuesday) = 2 : 1$. As $P(Monday) + P(Tuesday) = 1$, $P(Monday) = 2/3$ and $P(Tuesday) = 1/3$.

## 5. Conclusion

This short paper examined both the halfer and thirder solution for the Sleeping Beauty problem. It is concluded that both solutions went astray because they applied incorrect understanding of events in probability theory. And then the paper examines what probability should be assigned for the questions like $P(Monday)$, $P(Tuesday)$.

## References

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

2 D. Lewis (2001). *Sleeping Beauty: Reply to Elga*, Analysis, **61(3)**, p. 171–176.