

A mathematical analysis of the Sleeping Beauty problem

by

Jeffrey S. Rosenthal*

(August, 2008.)

1. Introduction.

The *Sleeping Beauty problem* (Elga, 2000; see also Piccione and Rubinstein, 1997) is a philosophical dilemma related to conditional probability. It may be succinctly described as follows. Sleeping Beauty is put to sleep, and a fair coin (say, a nickel) is tossed. If the nickel shows heads, then Beauty is interviewed on Monday only, while if the nickel shows tails, Beauty is interviewed on both Monday and Tuesday (and given an amnesia-inducing drug between the two interviews, so she does not remember the first interview during the second). In each interview, without access to any additional information (such as the result of the coin toss, or the existence of any previous interviews, or the day of the week), Beauty is briefly woken and is asked to assess the probability that the nickel showed heads. The question is, what probability should she assign to this?

One possible answer (e.g. Lewis, 2000; Arntzenius, 2002; Bostrom, 2007; Pust, 2008) is $1/2$, since the coin was fair originally (i.e., she surely would have assigned probability $1/2$ *before* being put to sleep), and Beauty has not really gained any new information since then (because she knew she would be interviewed at least once in any case). Another possible answer (e.g. Elga, 2000; Dorr, 2001; Monton, 2002; Weintraub, 2004; Horgan, 2004; Neal, 2007) is $1/3$, since in the long run Beauty will be interviewed twice as often if the nickel shows tails than if it shows heads, so if she bets \$2 on tails versus \$1 on heads then she will break even in the long run. Which (if either) answer is correct?

Mathematically speaking, it seems that we are being asked to compute the conditional probability that the nickel showed heads, conditional on the fact that Beauty is currently being interviewed. Generally speaking, conditional probabilities are well understood and should be unambiguously analysable by straightforward mathematics, using the formula

*Department of Statistics, University of Toronto, 100 St. George Street, Room 6018, Toronto, Ontario, Canada M5S 3G3. Email: jeff@math.toronto.edu. Web: <http://probability.ca/jeff/>

$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$. However, in this case such analysis is somewhat problematic, the difficulty being that a precise mathematical interpretation of “conditional on currently being interviewed” seems to be unclear. So, how can we analyse this problem mathematically?

This paper attempts to reconsider the problem in such a way that precise mathematical reasoning can be applied to prove that the answer $1/3$ is correct. It is hoped that this mathematical approach avoids the philosophical ambiguities inherent in some of the previous arguments.

2. A Subproblem: the Sleeping Peon.

Consider the following simple subproblem. We find a Peon and put him to sleep, and then flip a fair coin, say a nickel. If the nickel shows tails, we wake Peon and interview him (just once), asking him to assess the probability that the nickel showed heads. If the nickel shows heads, then we flip a second fair coin, say a dime. If the dime shows tails, we similarly wake the Peon and interview him once. If not (i.e., if the nickel and dime both show heads), then we do not bother to wake or interview Peon at all. (In summary, we interview Peon once if either the nickel or the dime or both show tails, otherwise we interview him zero times. In particular, the overall probability that Peon is interviewed is equal to $3/4$.) Under these circumstances, what probability should Peon assign, upon being interviewed, to the event that the nickel showed heads?

For this subproblem, the solution seems clear. Indeed, let *Interviewed* be the event that “Peon was interviewed (at all)”, and let *NickelHeads* be the event “the nickel showed heads”, and similarly *DimeHeads*, etc. Then Peon is being asked to assess the probability of *NickelHeads*, conditional on knowing only that the event *Interviewed* occurred. (Indeed, since this subproblem involves no amnesia or multiple interviews, *all* that Peon learns is whether or not he is interviewed at all, i.e. whether or not the event *Interviewed* occurs, so it is mathematically clear that *Interviewed* is the event Peon should condition on.)

That is, Peon is being asked to compute the conditional probability $P(\textit{NickelHeads} | \textit{Interviewed})$. He would do so as follows:

$$\begin{aligned} P(\textit{NickelHeads} | \textit{Interviewed}) &= \frac{P(\textit{NickelHeads} \text{ and } \textit{Interviewed})}{P(\textit{Interviewed})} \\ &= \frac{P(\textit{NickelHeads} \text{ and } \textit{DimeTails})}{P(\textit{NickelTails} \text{ or } \textit{DimeTails} \text{ or both})} = \frac{1/4}{3/4} = 1/3. \end{aligned}$$

Thus, for this simple subproblem, the correct probability that Peon should assign (during the interview) to the event that the nickel showed heads is equal to $1/3$. I consider this answer to be correct and clear and unambiguous, following directly from straightforward mathematical laws of conditional probability. I shall now argue that the original Sleeping Beauty problem can essentially be reduced to this simple Peon subproblem.

3. The Original Problem Revisited.

To make use of the above Peon subproblem in analysing the original Sleeping Beauty problem, we add one additional element. We assume that in addition to the previous elements (the nickel, Sleeping Beauty herself, the amnesia-introducing drug, etc.), we also have at our disposal another fair coin, say a dime. We make use of the dime as follows. If the nickel showed tails, then the dime is simply placed so that it shows heads during Beauty's Monday interview, and then repositioned so that it shows tails during Beauty's Tuesday interview. If instead the nickel showed heads (so Beauty will only be interviewed once), then the dime is instead simply flipped once in the usual fashion at the beginning of the experiment, and allowed to show its actual flipped result (either heads or tails, with probability $1/2$ each) throughout the experiment (in particular, during the one interview that will take place, on Monday). Furthermore, we assume that Beauty is *not* allowed to see the dime at all (and might not even know of its existence).

Thus, the dime does not in any way affect or control or interfere with any aspect of the original problem (including the nickel, the interviews, Beauty's knowledge and amnesia, etc.). However, we shall see that the dime does permit a precise mathematical analysis of the problem.

We now reason as follows. Call an interview a "heads-interview" if it takes place while the dime shows heads. Then if the nickel showed tails, then there will certainly be precisely one heads-interview. If the nickel showed heads, then there will be either one or zero heads-interviews, with probability $1/2$ each. So, the number of heads-interviews behaves just like the total number of interviews in the Peon subproblem.

Now, if Beauty were *told* just before her interview that the dime shows heads, then she would learn that a heads-interview did indeed occur. This would then put her in precisely the same situation as that of the Peon in the subproblem above. Then, just like the Peon, she would assign the probability $1/3$ that the nickel showed heads. In summary, *if Beauty were told that the dime showed heads*, then the correct answer to the problem would be $1/3$.

In mathematical terms, we can write this conclusion as $P(\text{NickelHeads} | \text{DimeHeads}) = 1/3$. That is, if Beauty is informed that the dime shows heads during her interview, then conditional on this information, she should assign the probability $1/3$ that the nickel also shows heads.

Similarly, if Beauty is informed (just before her interview) that the dime shows tails, then by identical reasoning, the answer would again be $1/3$. In summary, the answer would be $1/3$ if Beauty could see the dime, regardless of whether the dime was currently showing heads or tails. We can write this formally as

$$P(\text{NickelHeads} | \text{DimeHeads}) = P(\text{NickelHeads} | \text{DimeTails}) = 1/3.$$

In the actual problem, we assume that Beauty *cannot* see the dime. However, we now argue that, as far as probabilities for the nickel are concerned, that fact is irrelevant, and Beauty should still assign the probability $1/3$ even if she does not know what the dime shows. To see this, write $P(\text{NickelHeads})$ for the overall probability that Beauty should assign to the event that the nickel showed heads upon being interviewed (but now without knowing about the dime). Then it follows by the Law of Total Probability that

$$\begin{aligned} P(\text{NickelHeads}) &= P(\text{DimeHeads}) P(\text{NickelHeads} | \text{DimeHeads}) \\ &\quad + P(\text{DimeTails}) P(\text{NickelHeads} | \text{DimeTails}) \\ &= P(\text{DimeHeads}) (1/3) + P(\text{DimeTails}) (1/3) = 1/3. \end{aligned}$$

(We are using the fact that $P(\text{DimeHeads}) + P(\text{DimeTails}) = 1$, since during any specific interview the dime must show either heads or tails but not both.)

Thus, the answer for this version of the problem seems unambiguously and mathematically to be $1/3$. And, since the mere existence of the dime (which Beauty cannot see and has no knowledge of) cannot change Beauty's probabilities, I submit that this argument shows unambiguously that the answer to the original Sleeping Beauty problem is also $1/3$.

4. Some Related Issues.

While the above completes my main argument, I now consider a few other related issues.

4.1. A slight variant: randomised Sleeping Beauty.

Consider a very slight variant of the original Sleeping Beauty problem. As before, if the nickel is tails we will interview Beauty twice, once on Monday and once on Tuesday (with amnesia). And, as before, if the nickel is heads we will interview Beauty just once. The only modification is that if the nickel is heads, then rather than necessarily interviewing Beauty on Monday, we will first flip another fair coin (say, a dime), and then conduct our (one) interview on Monday if the dime is heads, or on Tuesday if the dime is tails. (We assume, as usual, that Sleeping Beauty cannot tell what day it is.)

For this variant, if Beauty were *told* that her interview was taking place on Monday, then this would reduce precisely to the Peon subproblem above. That is, as far as Monday interviews go, if the nickel showed tails then she would certainly have precisely one, while if the nickel showed heads then she would have one only with probability $1/2$ (i.e., only if the dime showed tails), otherwise zero. Furthermore, the fact that the interview is actually taking place on Monday tells her that she did indeed have one Monday interview. Thus, Beauty is in precisely the same situation as the Peon in the above subproblem. So, just as in the subproblem, the correct answer for the probability that the nickel showed heads would be $1/3$.

Similarly, if Beauty were told that her interview was taking place on Tuesday, the answer would again be $1/3$. (The reasoning is identical to the above, except that the roles of “heads” and “tails” for the dime are interchanged.) In summary, the answer would be $1/3$ if she knew which day it was, regardless of whether that day were Monday or Tuesday.

Now, in the actual problem, Beauty is *not* told which day it is. However, by the Law of Total Probability just as before, it follows that since she would have assigned probability $1/3$ upon being told either that it is Monday or that it is Tuesday, she should still assign probability $1/3$ even if she does not know which day it is. Thus, the answer for this variant of the problem again seems unambiguously and mathematically to be $1/3$.

Now, it seems clear that this variant is probabilistically equivalent to the original Sleeping Beauty problem, since in the original problem it is not relevant whether the one interview (if the nickel shows heads) takes place on Monday or Tuesday. So, this provides another (similar) argument for why the answer to the original problem is $1/3$.

4.2. Yet another variant: sleeping twins.

Consider the following variant of the Sleeping Beauty problem. Suppose there are two

twins, Beauty1 and Beauty2. We put them both to sleep (in separate, soundproof rooms), and flip a fair nickel. If the nickel shows tails, we wake and interview each of them (separately). If the nickel shows heads, we flip a dime. If the dime shows tails we interview Beauty1 only, while if the dime shows heads we interview Beauty2 only. What probability should each of them assign, upon being interviewed, to the event that the nickel showed Heads?

It is clear that in this variant, the situation for Beauty1 is precisely the same as that of the Peon in the above subproblem. Hence, as in that subproblem, Beauty1 should assign probability $1/3$ to the nickel showing heads. Similarly, Beauty2 should also assign probability $1/3$.

On the other hand, if we regard Beauty1 and Beauty2 as a “unit”, then together they behave (probabilistically speaking) just like Sleeping Beauty in the original problem. Indeed, the total number of times that Beauty1 and Beauty2 will be interviewed is two if the nickel is tails, and one if the nickel is heads. So, since each of Beauty1 and Beauty2 should assign the probability $1/3$, this suggests that Sleeping Beauty in the original problem should also assign probability $1/3$. Indeed, this argument is very similar to, and perhaps more intuitive than, the argument given in Section 3 above. However, it is not completely definitive, due to the possible confusion over conditioning on the same person being interviewed twice (in the original problem), versus two different people each being interviewed once (in this variant).

4.3. A simple argument why $1/2$ must be wrong.

Another mathematical insight into the original Sleeping Beauty problem can be gained by conditioning on the day of the interview, i.e. by considering how the probabilities would change if Beauty knew which day it was. Recall that, in the original problem, Beauty is interviewed on both Monday and Tuesday if the nickel showed tails, but is interviewed on Monday alone if the nickel showed heads.

Suppose first that Beauty is informed that her interview is taking place on Monday. Then, since precisely one interview would be conducted on Monday regardless of whether the nickel showed heads or tails, she should at that point assign equal probabilities to the nickel showing heads or tails. In other words, we must have $P(\text{NickelHeads} \mid \text{Monday}) = 1/2$, where *Monday* is the event that “the interview is taking place on Monday”.

On the other hand, suppose Beauty is informed that her interview is taking place on Tuesday. Then, since it is *impossible* to have an interview on Tuesday if the nickel shows

heads, she should at that point assign probability zero to the nickel showing heads. That is, we must have $P(\text{NickelHeads} | \text{Tuesday}) = 0$.

It then follows, again by the Law of Total Probability, that

$$\begin{aligned} P(\text{NickelHeads}) &= P(\text{Monday}) P(\text{NickelHeads} | \text{Monday}) + P(\text{Tuesday}) P(\text{NickelHeads} | \text{Tuesday}) \\ &= P(\text{Monday}) (1/2) + P(\text{Tuesday}) (0) = P(\text{Monday}) / 2. \end{aligned}$$

Now, it is not clear what value Beauty should assign to $P(\text{Monday})$, the probability (without any additional knowledge) that her interview is in fact taking place on Monday. Is it $1/2$, since she could be interviewed on either day? Or $2/3$, since two of the three possible interview situations (heads-Monday, tails-Monday, tails-Tuesday) involve Monday? Or $3/4$, since the probabilities of those three possible interview situations are respectively $1/2$, $1/4$, and $1/4$, and $1/2 + 1/4 = 3/4$?

In any case, since *sometimes* interviews will take place on Tuesday, we must have $P(\text{Tuesday}) > 0$, whence $P(\text{Monday}) = 1 - P(\text{Tuesday}) < 1$, whence $P(\text{NickelHeads}) < 1/2$. Hence, this argument allows us to see directly that the answer $1/2$ cannot be correct.

(Of course, once we agree that $P(\text{NickelHeads}) = 1/3$ is the correct answer to the original problem, then working backwards we can conclude that $P(\text{Monday}) = 2/3$.)

4.4. Generalisation to other numbers and probabilities.

Of course, once we accept the above reasoning, then it can also be applied to various generalisations of the original problem.

For example, if Beauty will instead be interviewed n times (with amnesia each time) if the nickel showed tails, but just once if the nickel showed heads, then it follows (by replacing the dime by an n -sided die) that the answer becomes $1/(n + 1)$. The original problem corresponds to $n = 2$.

Or, if the nickel actually was not a fair coin but instead had a priori probability q of coming up heads (and $1 - q$ of coming up tails), then the answer would become $(q/2)/(q/2 + (1 - q)) = q/(2 - q)$. The original problem corresponds to $q = 1/2$.

If we combine both of the above modifications, then the answer would become $q/(n - (n - 1)q)$. The original problem then corresponds to the values $n = 2$ and $q = 1/2$.

Many other similar variations can be similarly solved.

Acknowledgements: I am very grateful to Gary Malinas and Calvin Normore for discussing these issues with me.

References:

- Arntzenius, F. (January 2002), Reflections on Sleeping Beauty. *Analysis* **62**, 53–62.
- Bostrom, N. (July 2007), Sleeping Beauty and self-location: a hybrid model. *Synthese* **157**, 59–78.
- Dorr, C. (October 2002), Sleeping Beauty: in defence of Elga. *Analysis* **62**, 292–296.
- Elga, A. (April 2000), Self-locating belief and the Sleeping Beauty problem. *Analysis* **60**, 143–147.
- Horgan, T. (January 2004), Sleeping Beauty awakened: New odds at the dawn of the new day. *Analysis* **64**, 10–21.
- Lewis, D. (July 2001), Sleeping Beauty: reply to Elga. *Analysis* **61**, 171–176.
- Monton, B. (January 2002), Sleeping Beauty and the forgetful Bayesian. *Analysis* **62**, 47–53.
- Neal, R. M. (2006), Puzzles of anthropic reasoning resolved using full non-indexical conditioning. Technical Report No. 0607, Dept. of Statistics, University of Toronto.
- Piccione, M. and Rubinstein, A. (July 1997), On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* **20(1)**, 3–24.
- Pust, J. (January 2008), Horgan on Sleeping Beauty. *Synthese* **160**, 97–101
- Weintraub, R. (January 2004), Sleeping Beauty: a simple solution. *Analysis* **64**, 8–10.