The Sleeping Beauty Problem Is Just Another Fairy Tale

Percy Forest

E-mail: nosleepingbeautyproblem@yahoo.com

ABSTRACT : The Sleeping Beauty Problem is shown to be misconceived and therefore incoherent. A model of the experiment is presented along with a properly conceived treatment of betting scenarios. The views of 'thirders', 'halfers' and 'double halfers' are examined when this model comports with the experience of Sleeping Beauty. Bayesian inference and the proper role of credence are also discussed.

This paper is concerned *only* with the Sleeping Beauty Problem as defined below. It is *not* concerned with such ideas as possible world semantics, self locating beliefs, de dicto and de se beliefs, centred worlds, doxastic worlds, centred propositions, or what forms of conditionalization are appropriate with respect to them; no opinion is offered about these ideas and none should be inferred.

1. The Problem

Since it's inception the problem has been presented as having two components: a description of the problem and a requirement that any solution must respect.

Component one -

Adam Elga [1] described the problem:
'Some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you 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 coin toss is Heads?'

A second question is usually considered:
When you are awakened and told the day, to what degree ought you believe that the outcome of the coin toss is Heads?

Component two -

It is required that a solution assign probabilities, which sum to one, to the mutually exclusive 'predicaments' or 'centred epistemic possibilities' in which Sleeping Beauty might find herself upon awakening: (Mon, heads), (Mon, tails), and (Tue, tails). Thirders and halfers insist that her credence be consistent with the three premises:
P(Mon,tails) = P(Tue, tails), P(heads) = P(Mon, heads), and
P(tails) = P(Mon,tails) + P(Tue,tails).

Following Walters [2], the experiment may be modelled by a Markov chain:

```
                                Awakened and questioned
                                     |
                                     |    Experiment ends
                                     |       |
                 (Mon, heads) ---------------(Tue, heads)----------------------->
                  /                  1                         1
   Drugged    /   1/2
        |       /
(Sun)--------                                 Drugged
          |   \                                  |
       toss   \   1/2                            |
                \               1                |                     1
                 (Mon,tails)----------------->(Tue, tails)--------------------->
                              \                         /        |
                               \                       /        Experiment ends
                                \                     /
                              Awakened and questioned
```

The initial state is (Sun). The transition probability from (Sun) to the 'epistemic possibilities' (Mon, heads) and (Mon,tails) is 1/2 , and the transition probability from (Mon, tails) to (Tue, tails) is 1. It is generally agreed that it makes no difference if the coin is tossed before or after the Monday awakening.

It is also agreed that Sleeping Beauty remembers the experimental protocol and upon awakening is ignorant only of the day.

2. Three Solutions

1) Thirders

Elga, perhaps the first thirder, reasons as follows:

If, upon awakening, you were to learn that the toss was tails, then you would have no reason to prefer either (Mon, tails) or (Tue, tails) over the other, and therefore
P(Mon, tails) = P((Mon,tails) | (Mon, tails) or (Tue, tails)) and
P(Tue, tails) = P(Tue, tails) | (Mon, tails) or (Tue, tails)).
It follows that P(Mon, tails) = P(Tue, tails).

Upon learning that it is Monday, Elgas' Sleeping Beauty updates her probabilities as one would in a shell game when a randomly chosen shell is seen not to contain the pea.
Thus the updated probability P*(Mon, heads) satisfies
P*(Mon, heads) = P(Mon, heads)/(P(Mon,heads) + P(Mon, tails)) = 1/2.
P*(heads) = 1/2 = P*(tails).

If, upon awakening, you were to learn that it is Monday, and given that the coin toss was fair, your credence that the coin landed heads should be 1/2, and this should be the same as P((Mon, heads) | (Mon, heads) or (Tue, tails)). It follows that
P*(Mon, heads) = P*(Tue, tails) = 1/2.

Since the probabilities of the three 'epistemic possibilities' must sum to one he concludes that P(Mon, heads) = P(Mon, tails) = P(Tue, tails) = 1/3, and that P(heads) = 1/3 and P(tails) = 2/3.

Despite mentioning that his argument does not depend on betting considerations, he does not seem averse to such arguments, stating - 'Imagine the experiment repeated many times. Then in the long run, about 1/3 of the wakings would be Heads-wakings — wakings that happen on trials in which the coin lands Heads. So on any particular waking, you should have credence 1/3 that that waking is a Heads-waking, and hence have credence 1/3 in the coin’s landing Heads on that trial. This consideration remains in force in the present circumstance, in which the experiment is performed just once.'

2) Halfers

David Lewis, perhaps the first halfer, in his reply to Elga [3], reasons as follows:

If, upon awakening, you are not told the day, then you should assign the same probabilities to the coin toss as you would have on Sunday since only new evidence can produce a change in credence. Therefore P(heads) = P(tails) = 1/2 and
P(Mon,heads) = 1/2 = P(Mon, tails) + P(Tue, tails),
P(Mon,tails) = 1/4 = P(Tue,tails).
Upon learning that it is Monday, probabilties are updated as in a shell game to
P*(heads) = 2/3, P*(tails) = 1/3.

In summary, Elga and Lewis both accept the following three premises:

1) P(Mon,tails) = P(Tue, tails)
2) P(heads) = P(Mon, heads)
3) P(tails) = P(Mon,tails) + P(Tue,tails).

They differ in how a probability is to be assigned to (Mon, heads) and what probabilities are, as a consequence, to be assigned to the other epistemic possibilities.

3) Double halfers

Double halfers maintain that at all awakenings, even when told that it is Monday, Sleeping Beauty should always say that P(heads) = 1/2 = P(tails).
Double halfers include Cozic [4], Bostrom [5], Walters [6], Yamada [7], and Mariolis [8]. No two of them arrive at this conclusion in the same way and it is unfortunate that they are all given the same label.

All three camps agree that when told it is Tuesday the updated probability P*(tails) = 1 and when told the toss was heads P*(Mon) = 1.

3. Beauty and the Bookmaker

Before considering wagers, certain quantities representing objective properties of the experiment should be derived.

Let $P_{obj}(H)$ and $P_{obj}(T)$ be, respectively, the objective probabilties of heads and tails.
Let (#AaH) and (#AaT) be, respectively, the number of awakenings after a toss of heads, and after a toss of tails, i.e. - 1 or 2.
The expected, or average, number of awakenings per trial is:
$E(A) = 3/2 = P_{obj}(H)$ (#AaH) + $P_{obj}(T)$ (#AaT), where $P_{obj}(H) = 1/2 = P_{obj}(T)$.
The expected number of awakenings after a toss of heads per trial is:
$P_{obj}(H)$ (#AaH) = 1/2.
The expected number of awakenings after a toss of tails per trial is:
$P_{obj}(T)$ (#AaT) = 1.

Beginning with Elga, thirders have made much about the importance of the ratio
[$P_{obj}(H)$ (#AaH)] / E(A)] = 1/3.

Once again, here is Elga:

'Imagine the experiment repeated many times. Then in the long run, about 1/3 of the wakings would be Heads-wakings — wakings that happen on trials in which the coin lands Heads. So on any particular waking, you should have credence 1/3 that that waking is a Heads-waking, and hence have credence 1/3 in the coin’s landing Heads on that trial.”

So what is this ratio of averages?

Since each awakening corresponds to an 'epistemic possibility' we may view E(A) as E(E.P.) = the expected number of 'epistemic possibilities' encountered per trial.
Clearly, E(A) = E(E.P.).
We may also view the number of (Mon, heads) after a toss of heads, (#(Mon, heads)aH), as equal to (#AaH).
Since 2 = 1+1 we may view [the number of (Mon, tails) after a toss of tails plus the number of (Tue,tails) after a toss of tails] as equal to (#AaT).

E(A) = E(E.P.)
= $P_{obj}(H)$(#(Mon,heads)aH) + $P_{obj}(T)$(#(Mon, tails)aT) + $P_{obj}(T)$(#Tue,tails)aT)
= 3/2

The expected number of 'epistemic possibilities' encountered after N trials is simply N(E(E.P.)).

The expected number of (Mon,heads), after a toss of heads, encountered after N trials is N($P_{obj}(H)$ (#(Mon,heads)aH)), and similarly for the other two 'epistemic possibilities'.

If a record of the 'epistemic possibilities' encountered during N trials were to be kept, say by throwing color-coded balls into an urn, then the following conclusions hold:

a) As N increases E(E.P.) becomes an increasingly accurate estimate of the number of balls in the urn.

b) As N increases N(P_{obj}(H) (#(Mon,heads)aH)) becomes an increasingly accurate estimate of the number of balls in the urn representing the 'epistemic possibility' (Mon, heads), and similarly for the the other two 'epistemic possibilities'.

c) As N increases the ratio of averages ([P_{obj}(H) (#(Mon,heads)aH)] / E(E.P.)) becomes an increasingly accurate estimate of the proportion of balls in the urn which represent (Mon, heads), and similarly for the other two 'epistemic possibilities'.
[Exercise: let (#(M,H) / #B) be the ratio of the number of (M,H) balls in the urn to the total number of balls in the urn - find a function of this ratio which serves as an increasingly accurate estimate of P_{obj}(H) as N increases.]

d) And of course the actual number of tosses after N trials is N, and the expected number of heads after N trials is N(P_{obj}(H)).

Since all camps seem to think that repeated trials of the experiment or betting considerations based upon such trials are relevant to their positions a general description of betting scenarios is presented.

Before the trials begin Sleeping Beauty may assign any probabilities she wishes to the 'epistemic possibilities' and from these determine the probabilities, representing her credence, of the coin toss. Thirders and halfers must do so in a way that is consistent with the three premises given above. Since she is presumed to be unaware of the day and the result of the toss she must have the same credence at each 'epistemic possibility'.

At each 'epistemic possibility' she must guess the outcome of the toss according to the credence she has assigned by means of a random variable whose output represents her guess. Thus three independent, identically distributed random variables are required.

She is willing to bet a fixed amount at each awakening on her guess being correct.

The Bookmaker is informed of her credence and decides on the betting odds he is willing to offer.

They must both agree to any wager, and both will accept a fair bet; a fair bet being one which neither can be expected to profit from in the long run.

Each trial proceeds as in the diagram given previously.

Though equivalent to the proposed experiment, this set-up is much simpler. It requires no drugs, no trials, and no simulations. All that is needed is pencil and paper. Sleeping Beauty can even stay awake.

After each awakening, a correct guess will result in winning a bet, and an incorrect guess will result in a losing bet.

Expressing Sleeping Beauty's credence as $P_{SB}(H)$ and $P_{SB}(T)$, a few moments thought will show that the expected number of wins per trial, E(#W) is

$E(\#W) = P_{obj}(H)\ P_{SB}(H) + P_{obj}(T)\ P_{SB}(T) + P_{obj}(T)\ P_{SB}(T)$

and that the expected number of losses per trial E(#L) is

$E(\#L) = P_{obj}(H)\ P_{SB}(T) + P_{obj}(T)\ P_{SB}(H) + P_{obj}(T)\ P_{SB}(H)$.

[Exercise: show that if Sleeping Beauty was unaware of the objective probabilty of heads, as well as being unaware of the day and toss, that she could estimate $P_{obj}(H)$ if at some point in the trials she were told how many bets she had won and how many bets she had lost over N trials. Show that the estimate $(2 – (1/N)(\#W + \#L))$ becomes increasingly accurate as N increases.]

The Bookmaker computes the odds of a fair bet by taking the ratio $(E(\#L) / E(\#W))$ and, since Sleeping Beauty always bets the same fixed amount, setting the denominator to one. So we have $(E(\#L) / E(\#W)) = B/1 = B$, where B is the amount bet by the Bookmaker given that Sleeping Beauty always bets 1.

If both Sleeping Beauty and the Bookmaker are intelligent the only wager acceptable to both will be a fair bet.

Sleeping Beauty knows *all of this* before the experiment and after each awakening.

There is really nothing more to be said. And yet...

The Sleeping Beauty problem has been discussed extensively on the Internet. One can only speculate about how these disputes might appear in journals [9].

Brad DeLong, economist and thirder, here [10] attacks all those who disagree in no uncertain terms, in particular setting his sights on the physicist and halfer Lubos Motl.

After restating only the first component of the problem he writes:

'"Thirders" say that he should answer "1/3" when awakened, and "1/2" after being told it is Monday. If you run the experiment over and over again, 1/3 of the correct answers to the first question are "heads", and 1/2 of the correct answers to the second question are heads. You make money if you bet on heads as the answer to the first question at better than 2-1 odds, and if you bet on heads as the answer to the second questions at better than 1-1 odds. Sensible probabilities should converge in the long run to sample averages, and should guide you to betting decisions that do not lose money.

"Double-Halfers" say that he should answer "the odds are 1/2 that the coin came up heads and it is Monday, and 1/2 that it is heads and I will or have answered this question twice" when awakened, and "1/2" after being told it is Monday--for the definition of a fair coin is that its probability of coming heads is 1/2. Never mind that betting on heads as your answer to the first question will lose you money, and that only 1/3 of the occasions on which the first question is asked will the right answer be "heads".

"Lewis-Halfers" say he should answer "1/2" when awakened, and "2/3" after being told it is Monday. Why? Because they do not understand how probabilities work, or how to properly apply the Rule of the Reverend Thomas Bayes.'

The argument is remarkably fatuous.

Of course 'You make money if you bet on heads as the answer to the first question at better than 2-1 odds, and if you bet on heads as the answer to the second questions at better than 1-1 odds.'

But so will anyone! Delong's Sleeping Beauty is crowing because she has found a stupid Bookmaker willing to accept a wager he can only lose.

The sample averages that guide DeLong are not the expected number of heads over the number of tosses, but ($[P_{obj}(H)$ (#(Mon,heads)aH)] / E(E.P.)), i.e. - the expected proportion of (M,H) *awakenings* in the set of all *awakenings* as N increases. *These are different things* and, as the industrious reader will have worked out, are directly related to each other.

And of course double halfers will lose if they accept an even money bet on heads when not told the day.

But so will anyone! Everyone could be a winner if only DeLong would give up the name of his stupid Bookmaker!

In addition, the credence of a thirder Sleeping Beauty will result in fair betting odds of 4/5. On each awakening, the Bookmaker will bet 4/5 of the fixed amount bet by Sleeping Beauty.

The credence of a halfer or double halfer Sleeping Beauty when not told the day will result in fair betting odds of 1/1. On each awakening, the Bookmaker will bet the same fixed amount as Sleeping Beauty.

Funnily enough, when $P_{obj}(H) = 1/2 = P_{obj}(T)$, *whatever* credence is applied after each awakening, the credence applied to the Monday awakenings may be updated *in any manner whatsoever* without affecting the fair betting odds of the trial.
[Exercise: show this]

DeLong never does explain 'how to properly apply the Rule of the Reverend Thomas Bayes'.

Responding to DeLong, Motl [11] quickly makes three incompatible arguments against the thirder position:

First:

'The most popular wrong answer is that the probability of "heads" drops to 1/3 when she's woken up because heads-Monday, tails-Monday, and tails-Tuesday are three indistinguishable (by her), and therefore equally likely, arrangements of the coin state and the day.

Also, if the experiment is repeated many times, 1/3 of the awakenings simply will be "heads" ("heads-Monday") ones which, the "thirders" believe, implies that the probability is equal to this fraction.

However, both of these two arguments in favor of the answer 1/3 are wrong.

The argument that the probability should be 1/3 because "there are three equally good arrangements and 100% must be equally divided between them" is wrong because if there are three options, it simply doesn't mean that they are equally likely. One would need something like the ergodic theorem (clearly not applicable here because different coin_state-time combinations can't be compared and thermalized); or some Z3 or S3 symmetry to establish that the three probabilities are equal.

However, no such thermalization and no such symmetry exists here which is why there is absolutely no reason to expect that the three probabilities are equally likely. And indeed, a different, correct argument may be employed to show that they are not equal.'

Second:

'Sampling bias

The other argument, based on the "counting of awakenings" over many weeks, is wrong because the ratio cannot be interpreted as the probability due to the sampling bias.'

Third:

'Another popular wording for the basic frequentist argument of the "thirders" is the following:

The answer "the probability of heads is 50%" cannot be correct because this probability would imply that one wins 2 times the price of the lottery ticket and if the sleeping beauty made a 50% bet on "heads" every time she is woken up, she would be losing money because only 1/3 of the awakenings would be "heads" so she would only win on 1/3 of her lottery tickets but she would need to win 1/2 of the tickets for her investment to return.

Why isn't this argument right? Because the bookmaker isn't honest, and that's why the numbers calculated from this experimental setup cannot be interpreted as probabilities.'

But – we have seen that:

First: as N increases, the proportion of each 'epistemic possibility' in the set of awakenings approaches 1/3.

Second: each 'epistemic possibility' must be counted to determine the average number of awakenings; there is no sampling bias, selection bias, or double counting –
2 apples + 1 orange = 3 fruits; the proportion of (Mon, heads) in the set of awakenings is not the same as the probability of heads.

Third: the 'dishonesty' of the Bookmaker is not in counting the tails awakenings twice, but in suggesting an even money bet; in fact, Sleeping Beauty has stupidly accepted a wager she must always lose.

4. Bayesian Inference

In an earlier post [12]. Motl offered the Bayesian analysis alluded to above, first 'deriving' the probability 1/3 for each 'epistemic possibility' and then 'correcting' the result (in this post the roles of Monday and Tuesday are reversed w.r.t. to the presentation given above; the analysis is given in full because it contains most of the misconceptions and mistakes which crop up in this debate):

'The correct answer is obviously P=1/2. Both results of the coin toss are equally likely, P=1/2, and she knows that this is how the coin works. Both of these hypotheses predict that she would be woken up. When she learns that she woke up, she only knows that she would have woken up at least once. But both hypotheses, "tails" and "heads", guarantee that this observation would take place. So the observation gives her no information whatsoever. Bayes' theorem guarantees that her posterior probabilities therefore remain the same after the observation, P=1/2 both for "heads" and "tails".

The people who want to defend Ptails=1/3 like to imagine that one throws some marbles or fruits into a jar whenever she is woken up. See e.g. "another experiment" by Cristi Stoica. After a long time, the number of heads-fruits in the jar will exceed the number of tails-fruits by a factor of two which they interpret by saying that the probability that she wakes up after "heads" is twice as high.

They completely miss the obvious problem, the sampling bias. The fruits are more likely to be thrown to the jar if they're heads-fruits. She knows about this sampling bias as well. It's a bias, a mistake, and it has to be removed if you want to determine the actual probabilities of the underlying phenomenon – the coin, in this case. She is asked about the probability of a statement about the coin, not about fruits, so if she finds fruits helpful to answer a question about the coin, she must do it correctly. She demonstrably knows about this sampling bias as well – there are two heads-fruits added to the jar in the case of the "heads coin toss" – so she has to take it into account. It means that she must divide the number of heads-fruits by two, and then interpret the ratios of fruits as the ratio of probabilities. When she does it right, of course that she will determine that the probability is still Ptails=1/2.

Instead of fooling herself by the "same-size fruit" for differently likely events, she – if she wants to avoid later calculations – could directly throw half a pound of brown butter into the jar whenever she wakes up after "heads", and one pound of yellow butter if she wakes up after "tails". Then, after a long time, the relative amount of the butter of different colors could be directly interpreted as the probability ratio, and be sure that "tails" and "heads" would be equally likely once again, therefore P=1/2.

All the "thirders", defenders of P=1/3, seem to be very sloppy so no one has actually presented a clear argument why the answer should be P=1/3. Let me try to fill this gap.

A Bayesian derivation of P=1/3

There are 7 days in the week and 2 possible results of the coin toss. It means that we have 14 competing hypotheses Hcd describing both the "state of the coin" c "today's day in the week" d. The hypotheses are Monday-tails, Monday-heads, Tuesday-tails, Tuesday-heads, and so on.

By the Z7 symmetry between the days and Z2 symmetry between the sides of the coin, and by the a priori independence of these two quantities, we may argue that the probability of each combination is Pcd=1/14, OK? Let me write the full table of prior probabilities for you.'

Probabilities	Heads	Tails
Monday	1/14	1/14
Tuesday	1/14	1/14
Wednesday	1/14	1/14
Thursday	1/14	1/14
Friday	1/14	1/14
Saturday	1/14	1/14
Sunday	1/14	1/14 '

Now, when she wakes up, only 3 possibilities out of the 14 survive while the remaining 11 combinations are eliminated, right? So the table after this "collapse" – after a step of the Bayesian inference – seems to be

Probabilities	Heads	Tails
Monday	1/14	1/14
Tuesday	1/14	0
Wednesday	0	0
Thursday	0	0
Friday	0	0
Saturday	0	0
Sunday	0	0

These are the conditional probabilities P(wakeup|Hcd).

The denominator in Bayes' formula only means that we have to uniformly renormalize all the probabilities of the 14 hypotheses to guarantee that they sum up to one. That means that we must adjust the probabilities above to

Probabilities	Heads	Tails
Monday	1/3	1/3
Tuesday	1/3	0
Wednesday	0	0
Thursday	0	0
Friday	0	0
Saturday	0	0
Sunday	0	0

The table shows the conditional probabilities P(Hcd|wakeup).

At any rate, the three surviving arrangements, "heads-Monday", "heads-Tuesday", and "tails-Monday", seem to be equally likely. We may sum them up which means that after she learns that she was just woken up, the probability of Monday is 2/3, the probability of Tuesday is 1/3. The probability of heads is 2/3, the probability of tails is 1/3.

It sounds fairly good, doesn't it?

Maybe but what's more important is that it is wrong. Let's look what Bayes' formula actually quantifies the posterior probability of the hypothesis as

$P_t=P_{PWI}(H_{cd}|\text{wakeup}) = P(\text{wakeup}|H_{cd})P(H_{cd}) / P(\text{wakeup})$

where $c\in\{\text{heads,tails}\}$ is the state of the coin and d is one of the seven days of the week (the proposition is "today is d"). The denominator is just a normalization factor not dependent on the hypotheses that is calculated so that the sum of the hypotheses remains equal to one after we perform one step of Bayesian inference.

An important detail is that I added the subscript t=PWI which means that these are subjective probabilities evaluated right after a "particular wakeup incident" (PWI). You know, subjective probabilities depend on time – that's why we have the "inference". It is a bit dangerous to parameterize them by some "subjective time" but it is nevertheless possible to do so, if we are careful.

The prior probabilities Hcd are all equal to 1/14 in our case. I think that this is a claim that the "thirders" would endorse because the democracy between these 14 options is really the main "experience" that impresses them into thinking that the surviving 3 options are equally likely, too.

OK, the last piece of the story that matters are the probabilities is

Pt = PWI(wakeup|Hcd)

which is the probability of the "wakeup" predicted by one of the 14 "hypotheses" (coin-day arrangements) Hcd. I've said that it's obvious what the "thirders" assume about the value of Hcd: it's equal to 1 for the three surviving options and 0 for the remaining eleven options. That's why they believe that the probabilities of the three surviving options remain the same.

I am keeping the t=PWI as the subscript to indicate "when" the probabilities are evaluated.

However, we must avoid sloppiness if we want to calculate these "predicted" conditional probabilities – and therefore also the final result of the "sleeping beauty problem" – correctly. I have deliberately been sloppy so far because I wanted to emulate a thirder (after his IQ was increased by 10 points or so; I am not a good enough actor to play truly dense people).

What does the evidence in the conditional probabilities, "wakeup", really says if we are careful and not sloppy? The actual evidence that she can extract from being woken up tells her that

I was woken up today, i.e. on an unknown day. It isn't guaranteed that the current time t=PWI is equal to a particular day d.

When she is woken up, she doesn't really learn any particular unambiguous information about the "day" that is today. So her opinion about "the day today" doesn't necessarily have to agree with a particular value of the index d of a hypothesis we evaluate.

You might think that I am just being talkative or picky and that the "thirders" realize that and take that correctly into account. However, they don't. This clarification makes a lot of difference. It's all the difference you need to correct the wrong result P=1/3 and replace it by the right result P=1/2.

How does it work? Let me write the full table of the conditional probabilities – the predictions for the wakeup – because the tables look elegant given the ease with which I learned how to press CTRL/C and CTRL/V. :-)

Pt=PWI(wakeup|Hcd)	Heads	Tails
Monday	1/2	1
Tuesday	1/2	0
Wednesday	0	0
Thursday	0	0
Friday	0	0
Saturday	0	0
Sunday	0	0

These are the correct probabilities. After you normalize them to switch the order of the things in the conditional probability, the table becomes

Pt=PWI(Hcd|wakeup)	Heads	Tails
Monday	1/4	1/2
Tuesday	1/4	0
Wednesday	0	0
Thursday	0	0
Friday	0	0
Saturday	0	0
Sunday	0	0

The probability of "tails" is P=1/2 while the two wakeup days share the remaining P=1/2 reserved for the "heads".

OK, let's return to the previous table with the entries 1/2,1/2,1 and justify it. The hypothesis "tails Monday" predicts the "wakeup on Monday" with the probability 100%. It's really the uncontroversial part. The part where the "thirders" would be doing their mistake if they were able to present a Bayesian derivation at all is in the values

Pt=PWI(wakeup|Hheads,Monday) = 1/2
Pt=PWI(wakeup|Hheads,Tuesday)=1/2

Their wrong assumption leading to the value P=1/3 for the sleeping beauty problem is that both of these probabilities are equal to one. Why are these predicted probabilities equal to one-half and not one?

It's because

the hypothesis "heads Monday" (just like "heads Tuesday") predicts that there is a wakeup both on Monday and Tuesday – these two wakeups are logically connected, not mutually exclusive – so even if we assume that "today is Monday", there is a 50% probability that t=PWI is Monday and 50% that t=PWI is Tuesday.

Both "Monday heads" and "Tuesday heads" hypotheses share the possibility that she wakes up and t=PWI= Monday or Tuesday, so these predicted conditional probabilities are 50%.

Nearly everything in this 'derivation' that makes any sense at all is wrong.
So let's do it right.

Nothing is lost if we consider only two days, Monday and Tuesday, and replace the Z_7 symmetry by another Z_2 symmetry. We return to the convention of two awakenings after tails. This yields:

Probabilities	Tails	Heads
Monday	1/4	1/4
Tuesday	1/4	1/4

We now have four mutually exclusive'epistemic possibilities' and two sets of probabilities. The probabilities for the toss and the day are independent, and the probabilities of the 'epistemic possibilities' are products of these. In addition, on three of the possibilities Sleeping Beauty is awakened and questioned, and on the fourth, (Tue, heads), she is not. Call this the two probability model.

Hence:

P (A&Q) = 3/4   P(not(A&Q)) = 1/4

Where P(A&Q) = P((Mon, heads) or (Tue, tails) or (Tue, tails)) and
P(not(A&Q)) = P(Tue, heads).

Sleeping Beauty, upon being awakened and questioned, but before stating her credence, can make the following Bayesian inference:

P((Mon, heads) | (A&Q)) = (P((A&Q | (Mon, heads)) P(Mon, heads) / P(A&Q)) = 1/3
= P*(Mon, heads)
and similarly for the the other possibilities -
P((Mon, tails) | (A&Q)) = 1/3 = P*(Mon, tails),
P((Tue, tails) | (A&Q)) = 1/3 = P*(Tue, tails),
P((Tue, heads) | (A&Q)) = 0 = P*(Tue, heads).

Her credence that the coin landed Heads is therefore P*(heads) = 1/3.

When told that it is Monday she can make another inference:

With (Mon) as ((Mon,heads) or (Mon, tails))

P*((Mon, heads) | (Mon)) = (P*((Mon) | (Mon, heads)) P*(Mon,heads)) / P*(Mon))

= (1 (1/3) / 2/3) = 1/2

= P**(Mon,heads).

Where, *in this case,* P*((Mon,heads) or (Mon, tails)) = P*(Mon, heads) + P*(Mon, tails).

Motl, to combat supposed sampling bias, initially advocated the 'butter strategy' , which here, were records of trials being kept, would amount to re-labelling the (Tue, heads) balls as (Mon, heads) and placing them in the urn. The proportions of the possibilities in the urn would then approach those he considers to be correct, and no weighting of the tails possibilities would be necessary. Ironically, this would be, not sampling bias, but sampling fraud.

Later he claims that the first inference needs to be corrected because (Mon,tails) and (Tue, tails) are *not* mutually exclusive and decides that the conditional probabilities P((A&Q) | (Mon,heads)), P((A&Q) | (Mon,tails)), and P((A&Q) | (Tue, tails)), must be weighted to achieve the desired result. He doesn't say whether or not the 'epistemic possibilities' are mutually exclusive after being weighted.

The argument is inconsistent with what has gone before, utterly ad hoc, and far from lucid.

In another post [13] he gives an understandable, but mistaken, rationale for the weightings, based on a post by Walters [2].

Walters, considering the original experiment and using the Markov chain shown above, derived P((Mon, tails) or (Tue, tails)). Rather than reading the probabilties off the diagram and using the elementary addition rule, like this -

P(Mon, tails) + P(Tue,tails) - P(Mon, tails) P((Tue, tails) | (Mon, tails))
= 1/2 + 1/2 - (1/2) 1
= 1/2,

Walters took P(Mon, tails) to be the probability of arriving at the node after one step, and P(Tue,tails) to be the probability of arriving at the next node after two steps. He wrote the correct answer as

P((Mon, tails) or (Tue, tails)) = 1/2 = 1/2(P((Mon, tails) or (Tue, tails)).

It would seem that Motl took this expression as warrant for assigning weights to the two possibilities, which in the objective experiment both have probability of 1/2.

Walters did nothing of the sort and in fact did not really address the second component of the Sleeping Beauty problem at all.

Motl has simply accepted the second component of the problem. Putting aside a spurious analogy to statistical mechanics, what argument remains is similar to that given by Lewis. Motl's argument is *sheer Flapdoodle.*

But this does not vindicate thirders. Some awkward problems remain. (see section 6)

5. But credence is ***subjective...***

Both thirders and halfers seek to justify the assignment of probabilities to the 'epistemic possibilities' and their resultant assignment of probabilities to the outcomes of the coin toss by argueing that of course everyone knows the *objective* probabilities of all these things but Sleeping Beauty is in a unique situation, namely that of being unaware of the day whn she awakens. Her credence must therefore reflect and be dependent upon this fact and she must comply completely with second component.

Motl, in the comment thread of [11], states:

'the probability that the coin landed tails is - at any objectively defined time, Monday, Tuesday, anything - always 1/2. Everyone agrees with that, I think. The only problems begin when one evaluates the probabilities at moments that are defined subjectively - "after her awakening" That's where the difference between halfers and thirders emerges, and you haven't started to solve the problem yet at all.',

and later,

'the "moment after she was woken up" is an ambiguous term that may mean one of the 1 or 2 (or, in total, 3) possible awakenings, and the decision "which of them is occurring now" is a matter of psychological credence.

The sleeping beauty problem makes no sense whatsoever with purely "objective" notions of probability and time. That's why all meaningful formulations of the problem talk about the Bayesian probability or credence or confidence, and if they only talk about "probability", they mean "Bayesian probability" - a refinement of the definition of probability that depends on subjective knowledge.

All the complexity or controversy of the problem is derived from the question what is the right way to probabilistically answer the question "what time is it now" based on some knowledge of the spacetime into which we were thrown.'.

And here DeLong [10] takes the double-halfer Walters to task:

' Double-Halfer Sleeping Beauty gives those answers because her goal is not to give the right answer to the question or win money by betting but rather to be right about the state of the world. In the Monday world, because it is pre-coin flip, the answer she should give if she wants to be right about the state of the world to P(HEADS) is 1/2. In the Tuesday world, because she is not awakened if the coin is heads, the answer she should give if she wants to be right to P(Heads) is 0. Double-Halfer Sleeping Beauty is right about the probabilities all the time on Monday, and wrong on Tuesday. Thirder Sleeping Beauty is wrong about the probabilities all the time: on Monday they are not 1/3 but 1/2, and on Tuesday they are not 1/3 but 0). But Double-Halfer Sleeping Beauty loses if she bets on her probabities.

By contrast, the Thirder gives coherent and consistent answers: precisely because the day is uncertain and the coin is correlated with the day, the probability the coin was or will be heads is 1/3; the probability that it is Tuesday (and tails) is 1/3; and the probability that it is Monday-Tails is 1/3.

The problem with Walters--and here I suspect the problem is that he did not read Elga's paper sufficiently caerfully--is that Sleeping Beauty is not asked "what is the probability": she is not asked about a parameter describing the world. In fact, the word "probability" does not appear in Elga's initial paper. The words "belief" and "credence" do. The phrase "decision theory" does. She is asked not to nail the value of a parameter describing the world, but rather to form rational beliefs in order to guide her actions.'

The fatuity of betting considerations has already been remarked upon. When combined with such willful ignorance and massive condescension the mix is quite distasteful.

He continues:

'Double-Halfers get confused because they want to reserve "probability" for cases in which you are unsure which possible world you are in. They do not think "probability" applies to cases in which you are unsure where you are in one particular possible world. You can perhaps restrict the meaning of "probability" to use it in such a way. But you cannot reserve "belief" or "credence" in that way.

Whenever you are awakened, the probability that the coin falls heads and you are being awakened once is one-half; and the probability that the coin falls tails, that you are being awakened twice, and that this is one of those two awakenings is one-half. That is why the Double-Halfers are confused. But are they definitively wrong? Not quite.

Are they right? No. For Elga asks:

> When you are awakened...

that is, at that specific moment...

> ...to what degree ought you believe...

that is, not what are the objective probabilities of various possible worlds that might have happened, but what does it make sense for you to subjectively think are the odds...

> ...that the outcome of the coin toss is Heads?

And the answer to this question is crystal clear.

When you are awakened you do not know the date. It might be Monday, in which case Pm(HEADS)=1/2. It might be Tuesday, in which case Pt(HEADS)=0. Since experimental subjects are awakened twice as often on Monday as on Tuesday Tuesday, it is twice as likely that you are in a Monday as in a Tuesday Tuesday. So even though it is a fair coin, for Sleeping Beauty P(HEADS) = (2/3)Pm(HEADS) + (1/3)Pt(HEADS) = 1/3.'

The mind boggles. Again he mistakes a ratio of averages for a probability. What can be said in response? Plenty.

6. The Phantom Update

The second component of the problem is:

It is required that a solution assign probabilities, which sum to one, to the mutually exclusive 'predicaments' or 'centred epistemic possibilities' in which Sleeping Beauty might find herself upon awakening: (Mon, heads), (Mon, tails), and (Tue, tails). Thirders and halfers insist that her credence be consistent with the three premises:
P(Mon,tails) = P(Tue, tails), P(heads) = P(Mon, heads), and
P(tails) = P(Mon,tails) + P(Tue,tails).

Thirders and halfers seem to accept a model in which Sleeping Beauty goes to sleep on Sunday and proceeds directly to one of three (out of four) 'epistemic possibilities' where she is awakened and questioned.

```
                  0
            |-------------(Tue,heads)
            |     X
            |-------------(Mon, heads)
Sun---------|     Y
            |-------------(Mon, tails)
            |     Y
            |-------------(Tue, tails)
```

Where 0, X, andY are the transition probabilities.

These possibilities are mutually exclusive. They are determined by the probability of the toss and the newly discovered probability of the day. They determine a joint distribution of the probabilities of the coin toss and of the day. The probabilities of the toss and the day can be read off as marginals of this joint distribution.

What's more, these probabilities result from an update made upon awakening. The priors are known to her – P(H) = 1/2 and. given P(H), P(Mon), though having appeared as if by magic, must be equal to 1/2 since the updated probabilities for (Mon, tails and (Tue, tails) are equal. This also accounts for the 'missing' possibility (Tue, heads). The probability of each of the three 'epistemic possibilities' must equal 1/3.

In short – *The requirements of the second component demand that Sleeping Beauty adopt a model of the experiment that is identical to the two probability model, given above, that correctly models Bayesian inference. In accepting all these requirements both thirders and halfers commit themselves, albeit subjectively, to a model of the experiment that is inconsistent with the original experiment. In fact, the requirements literally* ***define*** *the two probability model.*

Elga, seemingly unaware of this, finds a rationale for the update [1]:

'Let *H* be the proposition that the outcome of the coin toss is Heads. Before being put to sleep, your credence in *H* was 1/2. I've just argued that when you are awakened on Monday, that credence ought to change to 1/3. This belief change is unusual. It is not the result of your receiving new information — you were already certain that you would be awakened on Monday.3 (We may even suppose that you knew at the start of the experiment exactly what sensory experiences you would have upon being awakened on Monday.) Neither is this belief change the result of your suffering any cognitive mishaps during the intervening time—recall that the forgetting drug isn't administered until well *after* you are first awakened. So what justifies it?

The answer is that you have gone from a situation in which you count your own temporal location as *irrelevant* to the truth of *H*, to one in which you count your own temporal location as *relevant* to the truth of *H*.',

and in closing:

'It is no surprise that the manner in which an agent counts her own temporal location as relevant to the truth of some proposition can change over time. What is surprising — and this is the first lesson — is that this sort of change can happen to a perfectly rational agent during a period in which that agent neither receives new information nor suffers a cognitive mishap.

At the start of the experiment, you had credence 1/2 in *H*. But you were also certain that upon being awakened on Monday you would have credence 1/3 in *H*—even though you were certain that you would receive no new information and suffer no cognitive mishaps during the intervening time. Thus the Sleeping Beauty example provides a new variety of counterexample to Bas Van Fraassen's 'Reflection Principle' (1984:244, 1995:19), even an extremely qualified version of which entails the following:
Any agent who is certain that she will tomorrow have credence x in proposition

R (though she will neither receive new information nor suffer any cognitive mishaps in the intervening time) ought *now* to have credence x in R.

David Lewis once asked ‘what happens to decision theory if we [replace the space of possible worlds by the space of centered possible worlds]?’ and answered ‘Not much.’ (Lewis 1983:149) A second lesson of the Sleeping Beauty problem is that something *does* happen. Namely: at least one new question arises about how a rational agent ought to update her beliefs over time.'

Elga's rationale conflicts with the model defined by his assumptions.

Lewis, ironically, is quite unaware of the 'predicament' in which he finds himself. Correctly believing that only new relevant evidence produces a change in credence, he answers the first question, when not knowing the day, i.e. P(H) = 1/2.
But, having assigned probabilities to the 'epistemic possibilities', he is unable to apply this principle when told that it is Monday and must say, incorrectly, that P*(H) = 2/3.

Lewis justifies this, in a manner similar to Elga, by saying that when told it is Monday Sleeping Beauty learns that it is not Tuesday and that an update is necessary. But because he has incorrectly, within the two probability model, assigned probabilities to the 'epistemic possibilities', his update is also incorrect.

Lewis' assignment of probabilities conflicts with the model defined by his assumptions.

If a halfer such as Motl, from knowledge or belief about the experiment, can reject the assignment of probabilities made necessary by accepting the second component then why can't he simply reject the second component altogether?

And why should Sleeping Beauty accept the second component?

Sleeping Beauty wakes up *fully believing* that she is a participant in the experiment which has been described to her. She *fully believes* that the coin was fair and that she will be awakened twice after a toss of tails. Being unaware of the day is entirely consistent with her beliefs and could even be taken as confirmation of them.

She *fully believes* that betting considerations are irrelevant and that the long term sample averages can be derived from what she *fully believes* is the value of $P_{obj}(H)$. Everything that she might learn about the results of the trials is consistent with what she *fully believes.*

If, upon awakening, it were for the first time to be suggested to Sleeping Beauty that she simply consider the second component of the problem before giving her answer would any of her beliefs, which determine her credence, change as a result?

Of course not! Why would they?

She knows that the consequences of complying with the second component and assigning probabilities, which sum to one, to the 'epistemic possibilities' are inconsistent with what she *fully believes.*

She *fully believes* that these consequences are absurd, and that they become a million times more absurd in Bostrom's version of the Extreme Sleeping Beauty [5].

She knows that it is not possible to assign objective or subjective probabilities to Monday and Tuesday in a manner consistent with what she *fully believes*.

She knows and *fully believes* that if given information about the trials, sample averages, and bets that she could tell the difference between the original experiment and the two probability experiment.

[Exercise: show this]

Those who think otherwise have some questions to answer -

For objective Bayesians:

> *What does Sleeping Beauty know? And when does she knowit?*
>
> *What does she forget? And when?*
>
> *What does she learn that makes the second component seem reasonable?*

For subjective Bayesians:

> *What does Sleeping Beauty believe? And when does she believe it?*
>
> *When the second component is suggested why do her beliefs change?*

Until they do there is no reason to take their views seriously. Since they can't, there will never be a reason

A rather Hobbesian explanation in terms of 'epistemic possibilities' can be found here [14]

Conclusion

The two components of the Sleeping Beauty problem are contradictory and cannot both be accepted without inconsistency.
If the first component is taken alone then the questions posed can be sensibly answered. Without new relevant information Sleeping Beauty's credence remains unchanged when not told the day and when told that it is Monday.
When told that it is Tuesday or that the toss was heads, only logical inference is possible.

ACKNOWLEDGEMENTS: The author would like to thank RAF IV for insight into thirders and halfers, and for correcting his sums. Thanks also to Lubos Motl, his blog The Reference Frame, and TRF commenter Tony. This remedial text was produced with support from the Catherine Foundation, which has been aiding victims of philosophical abuse for over four decades.

References

[1] Elga, A. (April 2000), Self-locating belief and the Sleeping Beauty problem. Analysis 60(2): 143 - 147

[2] Walters, R. F. C., http://rfcwalters.blogspot.it/2014/08/the-sleeping-beauty-problem-how.html

[3] Lewis, D. (July 2001), Sleeping Beauty: reply to Elga. Analysis, 61(3): 171 - 176

[4] Cozic, Mikal, Imaging and Sleeping beauty A Case for Double-Halfers, International Journal of Approximate Reasoning 52 (2):137-143 (2011

[5] Bostrom, N. (July 2007), Sleeping Beauty and self-location: a hybrid model. Synthese 157, 59 - 78

[6] Walters, R. F. C., http://rfcwalters.blogspot.it/2014/08/the-sleeping-beauty-problem-how-some.html

[7] Yamada, Masahiro, http://philpapers.org/rec/YAMLSB

[8] Mariolis, Iannis, Revealing the Beauty behinf the Sleeping beauty Problem, http://arxiv.org/abs/1409.3803

[9] http://www.mcsweeneys.net/articles/bar-fight-insults-as-academic-papers

[10] DeLong, Bradford http://www.bradford-delong.com/2015/03/sleeping-beauty-again-thirders-correct-double-halfers-confused-halfers-wrong-festival-of-fools-blogging.html

[11] Motl, Lubos http://motls.blogspot.com/2015/03/sleeping-beauty-and-beast-named-brad.html

[12] Motl, Lubos http://motls.blogspot.com/2014/08/sleeping-beauty-in-guantanamo-bay.html

[13] Motl, Lubos http://motls.blogspot.com/2014/08/sleeping-beauty-thirders-rudimentary.html

[14] http://ace.mu.nu/Windows-Live-Writer/34740215485e_1105A/too_true_8_2.jpg