Sleeping Beauty on Monty Hall

Michel Janssen* and Sergio Pernice†

May 19, 2016

Abstract

We present a game show that we claim can serve as a proxy for the notorious Sleeping Beauty Problem. This problem has divided commentators into two camps, 'halfers' and 'thirders'. In our game show, the potential awakenings of Sleeping Beauty, during which she will be asked about the outcome of the coin toss that determined earlier how many times she is awakened and asked, are replaced by potential contestants, deciding whether to choose heads or tails in a bet they will get to place if chosen as contestants on the outcome of the coin toss that determined earlier how many of them are chosen as contestants. This game show bears out the basic intuition of the thirders. Our goal in this paper, however, is not to settle the dispute between halfers and thirders but to draw attention to our game-show proxy itself, which realizes a version of the Sleeping Beauty Problem without the ambiguities plaguing the original. In this spirit, we design similar game-show proxies for variations on the Sleeping Beauty Problem with stochastic experiments other than a coin toss. We do the same for a variation in which Sleeping Beauty must decide upon being awakened whether or not to switch doors in the famous Monty Hall Problem and have the number of awakenings during which she gets to make that decision depend on the door she picked before she was put to sleep.

1 The Three Stooges on Monty Hall

Consider the following puzzle. In a special edition of his famous game show, "Let's make a deal," Monty Hall calls the Three Stooges to the stage and has them collectively pick one of three doors, D_1, D_2 or D_3. Behind one are two checks for a

*Program in the History of Science, Technology, and Medicine, University of Minnesota; email: janss011@umn.edu

†Universidad del Cema, Buenos Aires; email: sp@cema.edu.ar

thousand dollars each, behind the other two is a goat. The Three Stooges pick D_2. Before the show goes to a commercial break, Monty tells the Three Stooges that either one or two of them will be called back after the break. If they were wrong, only one of them will return (but he doesn't tell them which one!); if they were right, the other two will. During the break, the Three Stooges are made to take a nap backstage. When the show resumes, they are sound asleep.

If the checks are *not* behind D_2, Monty wakes up Curly and brings him back to the stage, making sure he has no idea whether he is the first or the second one to be woken up and brought back. As usual, Monty opens either D_1 or D_3 (whichever one has a goat behind it) and offers Curly to switch from D_2 to the other door that remains unopened. Curly is given ample time to make up his mind. If the door he ends up choosing has the two checks behind it, he gets one of them. If not, he goes home empty-handed.

If the checks *are* behind D_2, Monty goes through this same routine twice, with Moe and Larry (not necessarily in that order), the only other difference being that he now has a choice whether to open D_1 or D_3. Monty once again makes sure that neither of them finds out whether they were woken up and brought back first or second (at least not until Monty opens the door with the checks behind it, one of which may be gone at that point).

What is a Stooge to do in this predicament? Should he switch? Should he stay with the door they originally picked? Does it make any difference whether he stays or switches?

2 Game-show proxies for (variations on) the Sleeping Beauty Problem

The puzzle in section 1 combines elements of two well-known puzzles challenging our intuitions about probability: the Monty Hall Problem and the Sleeping Beauty Problem.[1] The solution of the Monty Hall Problem is no longer controversial. A contestant in a normal episode of "Let's make a deal" (without the funny business of putting contestants to sleep and waking them up again) should always take Monty Hall up on his offer to switch. Since Monty Hall never opens the door with the prize behind it and thus has to know which door that is, we need to assume that he offers contestants to switch regardless of which door they initially picked but that (weak) assumption is routinely granted.

[1] There is a vast literature on both. Good places to start are the wikipedia entries for these two problems.

The intuition that the opening of a door and the offer to switch are just for dramatic effect and do not affect the contestant's chances of winning is simply wrong. The opening of one of the doors provides the contestant with important information. A simple and (judging by its ubiquity on the web) effective way to make the point is this. Initially, the contestant only has a 1/3 chance of picking the right door and a 2/3 chance of picking the wrong one. Suppose he (or she) switches. If he was right the first time, he will now be wrong. If he was wrong the first time, he will now be right. So he now has a 2/3 chance of being right and only a 1/3 chance of being wrong. In a slogan, he flips the odds by switching. A device often used to shore up one's intuitions in this case is to increase the number of doors. If there are n doors ($n \geq 3$), the contestant initially has a $1/n$ chance of picking the right one and a $(n-1)/n$ chance of picking the wrong one. After the contestant has made her (or his) initial choice, Monty Hall opens all but one of the remaining doors and offers her to switch. Once again, the contestant flips the initial odds by switching. By switching, she effectively guesses that the prize is not behind the door she initially picked but behind any one of the other $n-1$ doors, all but one of which Monty Hall has meanwhile opened for her.

Unlike the Monty Hall Problem, the Sleeping Beauty Problem remains controversial. The problem is essentially the following. Sleeping Beauty is told that a fair coin will be tossed after she's been put to sleep and that, when she is woken up, she will be asked what her degree of belief is that the coin came up heads. It depends on the outcome of that coin toss, however, how many times she is asked. If the coin comes up heads, she will only be woken up and asked once. If the coin comes up tails, she will be woken up and asked twice. The first time she's woken up after the coin comes up tails she is given some amnesia drug so that she won't remember the second time that she's been woken up before and asked the same question. Every precaution is taken to make sure that Sleeping Beauty, when she wakes up, cannot tell whether her current awakening is the one after the coin came up heads or one of the two after the coin came up tails. What should Sleeping Beauty's degree of belief be upon being awakened that the coin came up heads?

One knee-jerk response is that, no matter how often Sleeping Beauty is put to sleep, woken up and drugged, the probability that a fair coin comes up heads is and remains 1/2. Therefore, her answer every time she is asked should be 1/2. Another knee-jerk response is that, if the coin comes up tails, Sleeping Beauty is twice as likely to be asked than if the coin comes up heads. Therefore, her answer every time she is asked should be 1/3. Those who think the answer is 1/2 are known as halfers. Those who think the answer is 1/3 are known as thirders. The debate between halfers and thirders has long moved beyond this clash of intuitions upon first encountering the problem. In fact, it persists only because it is ambiguous exactly what Sleeping Beauty is being asked.

To get a better handle on the Sleeping Beauty Problem and inspired by the Monty Hall Problem, we design a game-show proxy for it, in which potential contestants take over the role of potential awakenings. Our analysis will show that both halfers and thirders are right, depending on how one interprets the question Sleeping Beauty is asked. That said, our game-show proxy will be much more congenial to thirders than to halfers as it implements the interpretation of the question they consider to be the interesting one. *Should the information that Sleeping Beauty is given ahead of time about what will happen depending on the outcome of one toss of a fair coin change her degree of belief upon being awakened that this particular coin toss resulted in heads?* Upon a little reflection, both halfers and thirders will agree that if *that* is the question, Sleeping Beauty would be just as mistaken to ignore this information as contestants on "Let's make a deal" would be to ignore the information Monty Hall is giving them by opening one of the three doors.

Just as Sleeping Beauty knows how a coin toss will decide how many of her potential awakenings will become actual awakenings, potential contestants in our game-show proxy know how a coin toss will decide how many of them become actual contestants. Both Sleeping Beauty and our potential contestants are asked to assess (one way or another) the probability that this coin toss resulted in heads when woken up or chosen as a contestant, respectively. Using Bayes' rule, they should update their degree of belief that the coin came up heads in light of the information about how potential awakenings/contestants become actual awakenings/contestants. In the spirit of a time-honored Bayesian tradition (based on the principle of that you put your money where your mouth is), we will cash out the degrees of belief of our game-show contestants in terms of betting behavior. If our contestants update according to Bayes' rule (and the same is true for Sleeping Beauty), their prior degrees of belief, determined by the symmetry of the coin, get replaced by posterior degrees of belief that are determined by the numbers of contestants (awakenings) chosen for the two possible outcomes of the coin toss—1 for heads, 2 for tails in the original problem but those numbers can be chosen arbitrarily.

Similar game-show proxies can be designed for variations on the Sleeping Beauty Problem in which the coin toss is replaced by a different stochastic experiment with an arbitrary number of equiprobable outcomes and arbitrary numbers of candidates (awakenings) for different outcomes. As in the case of the Monty Hall Problem, a good way to shore up one's intuitions in dealing with the Sleeping Beauty Problem, as we will see in sections 3–4, is to vary these parameters. In section 5, we illustrate this strategy for getting a handle on the Sleeping Beauty Problem by designing a game-show proxy for a variation on the problem that combines it with the Monty Hall Problem (for arbitrary numbers of doors and awakenings). The scenario with the Three Stooges in section 1 is our game-show proxy for this variation on the Sleeping Beauty Problem for three doors and three awakenings.

3 Pitching a new game show: "Ignore the (initial) odds!"

Given the central role of switching doors in Monty Hall's "Let's make a deal," his game show might as well have been called "Flip the odds." We will call the class of game shows introduced and analyzed in this section "Ignore the odds." In both cases, the odds in the title refer to the initial odds (better: probabilities, chances, degrees of belief), determined by symmetries of some physical object.

In "Ignore the odds," the host starts by selecting (not necessarily randomly) a pool of N candidates or potential contestants from the audience. These candidates are carefully sequestered before the host proceeds to perform a simple stochastic experiment with n equiprobable outcomes O_i ($i = 1, \ldots, n$): flipping coins, rolling fair dice, spinning a roulette wheel, drawing balls from a lottery ball machine, drawing cards from a deck, etc. Different stochastic experiments can be used in different versions or different episodes of the show. Depending on the outcome of the experiment, the host then calls up, one by one, one or more of the N candidates to become actual contestants on the show. Each contestant is given a certain amount of money, say a thousand dollars, to place a bet, at even odds,[2] on one of the n possible outcomes of the experiment the host just performed (that very experiment, not a repetition of it). If they are right, they go home with two thousand dollars; if they are wrong, they go home empty-handed.

If contestants were randomly selected from a pool of N candidates, their chances of winning the bet would be $1/n$ regardless of which outcome O_i they choose to put their money on. This probability is determined simply by symmetries of the physical object(s) used in the stochastic experiment. As long as he or she does not know anything about the selection procedure, a contestant's degree of belief that the outcome was O_i should be $1/n$ for all i. In Bayesian terms, these should be his or her prior degrees of belief or *priors* for short:

$$\Pr(O_i) = \frac{1}{n}. \tag{1}$$

[2]The restriction to bets at even odds can be relaxed but this would unnecessarily complicate the analysis.

In “Ignore the odds,” however, the selection procedure is anything but random and, more importantly, all candidates know that it is not. The procedure still gives all of them an equal chance of becoming a contestant,[3] but *how many* candidates become contestants depends on the outcome of the very stochastic experiment contestants will be betting on.

Figure 1: “Ignore the odds” with eight potential contestants and a lottery ball machine as the stochastic experiment (Drawing by Laurent Taudin).

Here is how it works in detail. At the beginning of the show the candidates are divided into groups assigned to different outcomes. The sizes of these groups will be different for different outcomes. The candidates are told the sizes of the groups for all outcomes but they are not told in which group they are. Candidates become contestants if and only if they happen to be in the group for the actual outcome of the stochastic experiment. The host will call up all members of this group, one by one, to place their bet.

[3]In section 5, however, we will consider a variation on the selection procedure used in sections 3 and 4, in which this is no longer true.

To make sure that a candidate, when chosen as a contestant, knows neither the outcome of the stochastic experiment nor whether any other candidates have already been called up as contestants, all candidates are put into individual cubicles after the host has explained to them how contestants will be chosen but before he carries out the stochastic experiment (see Fig. 1). When a cubicle's door is closed, the candidate inside cannot detect in any way what is going on outside of it. The candidates are given the sizes of the groups for all outcomes before they are sent to their cubicles.

Each cubicle is equipped with two lights, one on the inside, one on the outside, which will be switched on to let the candidate inside and the audience outside know that the candidate in that particular cubicle has been chosen as a contestant. The candidate-turned-contestant then leaves her (or his) cubicle, closes the door behind her (which turns off the lights), places the bet and finds out whether she lost or won. Unless she is the final contestant, she is only allowed to watch the rest of the show from a place where the contestants coming after her cannot see her (nor detect her presence in any other way). Once all members of the group of candidates chosen as contestants have placed their bets, the remaining candidates are allowed to leave their cubicles and the show is over. How can the candidates take advantage of the information they are given about the selection of contestants to improve their chances of winning their bets when chosen as a contestant?

Let N_i be the number of candidates in the group assigned to the outcome O_i. These N_i's can be any non-negative integers that add up to the total number of candidates N. Of course, the number of candidates will itself be a number chosen by the host or the producers of the show. Note that one or more N_i's are allowed to be zero. If $N_j = 0$ and O_j happens to be the result of the stochastic experiment, no candidates will become contestants. That is not a problem. The host can simply repeat the stochastic experiment until he gets an outcome O_k for which $N_k \neq 0$. We can think of the instances in which the stochastic experiment produces a result O_j for which $N_j = 0$ as runs of the game where the game is over before contestants are called to the stage.

The candidates can use Bayes' rule to update the priors in Eq. (1) to take into account the information they are given about how candidates become contestants and use their posterior degrees of belief (or *posteriors* for short) when deciding on which outcome to put their money if and when actually chosen as a contestant. They can do this updating as soon as they have been given the group sizes N_i, even though their beliefs do not really matter unless and until they become contestants.

We will first do this Bayesian updating for the simple case in which the stochastic experiment is a single coin toss. We then generalize the results to arbitrary stochastic experiments with n equiprobable outcomes.

For a single toss of a fair coin, we have

$$n = 2, \quad O_1 = H, \quad O_2 = T, \quad N_1 = N_H, \quad N_2 = N_T, \quad N = N_H + N_T. \tag{2}$$

H and T stand for 'Heads' and 'Tails', respectively. N_H and N_T are the number of candidates in the groups for heads and tails, respectively. These should add up to the total number of candidates, N. The prior degrees of belief for any candidate that the coin will come up heads or tails, respectively, are:

$$\Pr(H) = \Pr(T) = \frac{1}{2}. \tag{3}$$

After all candidates have been given the numbers N_H and N_T, they go into their cubicles and the doors are closed. The host then tosses the coin. If the coin comes up heads, the N_H candidates in the group for heads are chosen as contestants. If the coin comes up tails, the N_T candidates in the group for tails are. One by one, the candidates in the chosen group will see the light in their cubicle go on and get to place their bet.

We calculate the degrees of belief that an arbitrary candidate, sitting in her (or his) cubicle waiting for the light to go on, should have that the coin came up heads or tails, respectively, given everything she knows about how candidates are chosen to become contestants. In Bayesian terms, we need to calculate the posteriors $\Pr(H \mid A)$ and $\Pr(T \mid A)$, where

$$A = \text{I, a candidate, have been chosen as a contestant}$$

('A' stands for *actual* as opposed to *potential* contestant). Implicit in this definition of A is that the selection took place according to the rules of the show, including those specifying what all candidates will be told about the selection procedure.

According to Bayes' rule, these posteriors are given by:

$$\begin{aligned} \Pr(H \mid A) &= \frac{\Pr(A \mid H)\Pr(H)}{\Pr(A \mid H)\Pr(H) + \Pr(A \mid T)\Pr(T)}, \\ \Pr(T \mid A) &= \frac{\Pr(A \mid T)\Pr(H)}{\Pr(A \mid H)\Pr(H) + \Pr(A \mid T)\Pr(T)}. \end{aligned} \tag{4}$$

Since the priors $\Pr(H) = \Pr(T) = 1/2$ (see Eq. (1)), this simplifies to

$$\Pr(H \mid A) = \frac{\Pr(A \mid H)}{\Pr(A \mid H) + \Pr(A \mid T)}, \quad \Pr(T \mid A) = \frac{\Pr(A \mid T)}{\Pr(A \mid H) + \Pr(A \mid T)}. \tag{5}$$

The evaluation of the posteriors thus comes down to the evaluation of the *likelihoods* $\Pr(A \,|\, H)$ and $\Pr(A \,|\, T)$ on the right-hand side.

One might be tempted to set $\Pr(A \,|\, H) = \Pr(A \,|\, T) = 1/2$ on the argument that the probability of any candidate being chosen as a contestant is 1/2 regardless of the outcome of the coin toss. In that case, we would also have $\Pr(H \,|\, A) = \Pr(T \,|\, A) = 1/2$. We would be ignoring the information, however, that different numbers of candidates will become contestants depending on the outcome of the coin toss. That information also needs to be taken into account.

If the coin came up heads (H) and the candidate is chosen as a contestant (A), she knows that she must be one of the members of the group for heads. The likelihood $\Pr(A \,|\, H)$ is thus equal to the probability that she belongs to that group. Great precautions have been taken to make sure that, as far as she can tell, she could be any one of the N candidates. The probability that she (or any other candidate) belongs to the group for heads is therefore simply the number of candidates N_H in that group divided by the total number of candidates N. A similar argument can be given for $\Pr(A \,|\, T)$. Hence

$$\Pr(A \,|\, H) = \frac{N_H}{N}, \qquad \Pr(A \,|\, T) = \frac{N_T}{N}. \tag{6}$$

Inserting these expressions in Eq. (5), we arrive at

$$\Pr(H \,|\, A) = \frac{N_H}{N_H + N_T} = \frac{N_H}{N}, \quad \Pr(T \,|\, A) = \frac{N_T}{N_H + N_T} = \frac{N_T}{N}. \tag{7}$$

These posteriors replace the priors in Eq. (3). Since the posteriors in Eq. (7) only depend on N_H and N_T, candidates will already know as they are sitting in their cubicles waiting for the light to go on, whether they will choose heads or tails, should they be called up as contestants. If $N_H > N_T$, they choose heads. If $N_T > N_H$, they choose tails. If $N_H = N_T = N/2$, they get no information that would increase their chances of winning their bet. For $N_H = N$ (or, similarly, $N_T = N$), finally, the game becomes trivial: all contestants would be guaranteed to win. If the coin comes up heads, the host calls up all contestants, if it comes up tails, he just flips it again.

This is an extreme version of a general mechanism that helps us understand why the posteriors in Eq. (7) are what they are. Imagine that this episode of "Ignore the odds" (with the same coin and the same pool of potential contestants) is repeated X times (with X a very large integer). I now put myself in the shoes of an arbitrary potential contestant. I will be chosen as a contestant in about N_H/N of the roughly $X/2$ runs of the show in which the coin will come up heads and in about N_T/N of the roughly $X/2$ runs in which the coin will come up tails (if $N_H = 0$ or $N_T = 0$ the show will be over without any contestants being chosen in roughly $X/2$ runs and I will be chosen as a contestant in the roughly $X/2$ remaining runs). Overall, I will thus to be chosen as a contestant in about

$$\frac{N_H}{N}\frac{X}{2}+\frac{N_T}{N}\frac{X}{2} \tag{8}$$

runs. Since $N_H + N_T = N$, Eq. (8) confirms that, overall, I expect to be chosen as a contestant in about half of the X runs. However, Eq. (8) also shows that, if $N_T > N_H$, I expect to be a candidate in runs where the coin comes up tails more often than in runs in which the coin comes up heads. In other words, that I (or any other potential contestant) should use Eq. (7) and set the probabilities that a fair coin came up heads or tails equal to $N_H/N < \frac{1}{2}$ and $N_T/N > \frac{1}{2}$, respectively, is simply to take into account this *sampling bias*.

What Eq. (7) shows is that the procedure for selecting contestants from an initial pool of candidates has the effect of erasing the priors in Eq. (3), determined by the symmetry of the coin, and replacing them by posteriors determined by our choice of the sizes N_H and N_T of the two groups into which we divided the N candidates we started with. What Eq. (7) shows then is that we can basically change the priors to any posteriors we want. That is why we called this game show "Ignore the odds"—the initial odds that is.

These results can readily be extended to arbitrary stochastic experiments with n equiprobable outcomes O_i ($i = 1, \ldots, n$). The prior degrees of belief for any candidate that the experiment will result in O_i are

$$\Pr(O_i) = \frac{1}{n}. \tag{9}$$

For each outcome O_i there will be $N_i \geq 0$ potential candidates (the host will repeat the stochastic experiment until he gets an outcome O_j for which $N_j \neq 0$). Eq. (4) straightforwardly generalizes to:

$$\Pr(O_i \,|\, A) = \frac{\Pr(A \,|\, O_i)\Pr(O_i)}{\sum_{j=1}^{n}\Pr(A \,|\, O_j)\Pr(O_j)}. \tag{10}$$

Given Eq. (9), this reduces to (cf. Eq. (5)):

$$\Pr(O_i \,|\, A) = \frac{\Pr(A \,|\, O_i)}{\sum_{j=1}^{n}\Pr(A \,|\, O_j)}. \tag{11}$$

The candidate's knowledge about the selection procedure tells him or her that the likelihoods are given by (cf. Eq. (6) and the reasoning leading up to it)

$$\Pr(A \,|\, O_i) = \frac{N_i}{N}. \tag{12}$$

Inserting this expression for the likelihoods on the right-hand of Eq. (11), we see that Eq. (7) generalizes to:

$$\Pr(O_i \,|\, A) = \frac{N_i}{N}. \tag{13}$$

Once again, we see that the procedure for selecting contestants from an initial pool of candidates has the effect of erasing priors (see Eq. (9)), determined by symmetries of physical objects used in a stochastic experiment, and replacing them by posteriors (see Eq. (13)), determined by the choice of the numbers N_i, subject only to the requirement that they add up to N, the number of candidates we started out with.

Like Eq. (7), Eq. (13) can be seen as expressing a certain sampling bias. Imagine that this episode of "Ignore the odds" (with the same stochastic experiment and the same pool of potential contestants) is repeated X times (with X a very large integer). I put myself in the shoes of an arbitrary potential contestant in these X runs of the game. Suppose there are $m < n$ outcomes O_j for which $N_j = 0$. That means in that in roughly $m(X/n)$ runs of the game, no contestants are chosen. In the remaining $(n - m)(X/n)$ runs, I will be chosen as a contestant in about N_i/N of the roughly X/n runs in which the outcome of the stochastic experiment is O_i. Overall, I will thus be chosen in about

$$\sum_{i=1}^{n} \frac{N_i}{N}\frac{X}{n} \tag{14}$$

runs. Since $\sum_{i=1}^{n} N_i = N$, Eq. (14) confirms that I expect to be chosen as a contestant in about X/n runs overall. However, Eq. (14) also shows that, if $N_k > N_l > 0$, I expect to be a candidate in runs with outcome O_k more often than in runs with outcome O_l. In other words, that I (or any other potential contestant) should use Eq. (13) and choose a value other than $1/n$ for the probability that the stochastic experiment under consideration results in one of its n equiprobable outcomes is simply to take into account this sampling bias.

4 Game-show proxy for Sleeping Beauty Problem

Our proxy for the Sleeping Beauty Problem is a special case of the game show analyzed with malice aforethought in Eqs. (2)–(8) in Section 3. It is a version or episode of "Ignore the odds" in which the stochastic experiment is a single toss of a fair coin. We analyzed this game show for an arbitrary number of candidates N divided into groups for heads and tails, containing N_H and N_T candidates, respectively. To turn this into a proxy for the Sleeping Beauty Problem, we set $N = 3$ (the minimum number of candidates for which this version of "Ignore the odds" is non-trivial) and $N_H = 1$.

The parameters N_H, N_T and N introduced in Eq. (2) thus have the values:

$$N_H = 1, \quad N_T = 2, \quad N = 3. \tag{15}$$

Candidates (potential contestants) are the analogues of potential awakenings. The one candidate chosen to become an actual contestant if the coin comes up heads is the analogue of the one potential awakening of Sleeping Beauty that becomes an actual awakening if the coin comes up heads. The two contestants chosen after the coin comes up tails are the analogues of the two potential awakenings that become actual awakenings if the coin comes up tails. Just as there is probability of 1/2 for all three candidates in the game show to become contestants, there is a probability of 1/2 for all three potential awakenings of Sleeping Beauty to become actual awakenings. The drugging of Sleeping Beauty ensures that during subsequent awakenings she knows as little about preceding awakenings as any contestant coming out of his or her cubicle knows about other contestants in the game-show proxy for the Sleeping Beauty Problem.

It may seem that there is still an important difference between potential contestants and potential awakenings. It is easy to tell potential contestants apart: we could use their Social Security Numbers (SSNs), for instance. Every potential contestant will presumably know at all times what his or her SSN is. How can we tell Sleeping Beauty's potential awakenings apart? How can Sleeping Beauty herself tell her own potential awakenings apart? Fortunately, these questions have an easy answer. We can specify, for any potential awakening, at what time and on what date it will happen if it will happen at all (allowing ten minutes or so for each awakening). We then put a clock that displays both time and date in the room where Sleeping Beauty is awakened so that she can tell, every time she is woken up, which potential awakening has just become an actual awakening. Just as we had no need to refer to a potential candidate's SSN, we will have no need to refer to date and time of a potential awakening. The point is that, at least in principle, individuating potential awakenings is no more problematic than individuating potential contestants.[4]

[4] This way of providing potential awakenings with a time stamp should not be conflated with the time stamps used in the standard version of the Sleeping Beauty Problem. There Sleeping Beauty is told ahead of time that she will be woken up a second time on Tuesday if the coin comes up tails. In that case, Sleeping Beauty, when awakened, can obviously not be allowed to find out what day of the week it is (at least not until she has answered the question what her degree of belief is that the coin came up heads). In our version, she is told only that two of the three potential awakenings (labeled by the time and date they will happen) will become actual awakenings if the coin comes

Sleeping Beauty has no reason to doubt that the coin used to determine whether she will be woken up once or twice is fair. In other words, her prior degree of belief that the coin will come up heads is

$$\Pr(H) = \frac{1}{2}. \tag{16}$$

What is Sleeping Beauty's degree of belief, once she is woken up, that the particular toss of this coin that decided how many times she would be awakened resulted in heads? In Bayesian terms, was is her posterior $\Pr(H \,|\, A)$? Here A is defined as

$$A = \text{I, Sleeping Beauty, have been awakened,}$$

where 'A' stand for actual rather than potential awakening. Implicit in the definition of A is that which potential awakenings become actual awakenings is determined by the protocol spelled out in the Sleeping Beauty Problem, including the provisions about what Sleeping Beauty is to be told about this protocol.

With this reinterpretation of A and using the values of the parameters in Eq. (15), we can use the first half of Eq. (7) for Sleeping Beauty's posterior degree of belief that the coin came up heads:

$$\Pr(H \,|\, A) = \frac{N_H}{N} = \frac{1}{3}. \tag{17}$$

This follows directly from Eq. (6) for the likelihoods $\Pr(A \,|\, H)$ and $\Pr(A \,|\, T)$.

Since this is the crux of the matter, we transfer the reasoning leading up to Eq. (6) back to the Sleeping Beauty Problem. We do so for arbitrary values of N_H and N. If the coin came up heads (H) and she has been awakened (A), that awakening must be one of the N_H potential awakenings that become actual awakenings if the coin came up heads. The likelihood $\Pr(A \,|\, H)$ is thus equal to the probability that Sleeping Beauty's current awakening is one of those N_H potential awakenings. Great precautions have been taken to make sure that Sleeping Beauty, when she wakes up, has no way to tell whether her current awakening is one of the N_H potential awakenings after the coin came up heads or one of the N_T potential awakenings after the coin came up tails. The probability that her current awakening is one of former is therefore the number N_H of such potential awakenings divided by the number N of all potential awakenings.

---

up tails. She is not told which two. In that case, the date and the time she is awakened are as irrelevant to Sleeping Beauty's assessment of her degree of belief that the coin came up heads as the SSN of a potential contestant in "Ignore the odds" is to his or her calculation of the posteriors $\Pr(H \,|\, A)$ and $\Pr(T \,|\, A)$ (see Eqs. (4)–(7)).

In Section 2, we noted that the Monty Hall Problem becomes more intuitive if the number of doors is increased. The solution to the Sleeping Beauty Problem likewise becomes more intuitive if the number of potential awakenings or, in our game-proxy for the problem, the number of potential contestants is increased. Instead of the values in Eq. (15), we choose

$$N_H = 1, \quad N_T = 99, \quad N = 100. \tag{18}$$

In this case, Eq. (17) tells us that Sleeping Beauty's degree of belief that the coin came up heads plummets from its initial value of 1/2 to just 1/100.

Like Eq. (7) and Eq. (13), Eq. (17) can be seen as expressing a certain sampling bias. Imagine we repeat the experiment with Sleeping Beauty X times where X is an arbitrary large integer. In the roughly $X/2$ runs in which the coin comes up heads, the total number of awakenings of Sleeping Beauty will be about $N_H(X/2)$. In the roughly $X/2$ runs in which the coin will come up tails, the total number of awakenings will be about $N_T(X/2)$. In all runs combined, the total number of awakenings will thus be $N_H(X/2) + N_T(X/2) = N(X/2)$. If $N_T > N_H$, more of these awakenings occur during runs in which the coin comes up tails than during runs in which the coin comes up heads. That Sleeping Beauty uses Eq. (17) to set the probability that a fair coin came up heads equal to $N_H/N \neq \frac{1}{2}$ is simply to take into account this sampling bias.

Eq. (17) shows that, *as long as our game show can be used as a proxy for the Sleeping Beauty Problem*, the thirders are right and the halfers are wrong. There is a simple way, however, of reconciling the two positions. Once again imagine that the experiment with Sleeping Beauty is repeated many times. Ask Sleeping Beauty to answer the following pair of questions:

Question #1: What is the probability that in an arbitrarily chosen run the coin comes up heads? **Answer**: 1/2.

Question #2: What is the probability that an arbitrarily chosen awakening happens during a run in which the coin comes up heads? **Answer**: N_H/N (which works out to 1/3 for the values given in Eq. (15)).

One can debate—and commentators have!—whether question #1 or question #2 is the more natural way of interpreting the question in the original Sleeping Beauty Problem and whether the problem doesn't become trivial once you accept that halfers and thirders are just interpreting it differently. We will not get into those debates. We simply note that, although both questions are relevant to our game-show proxy (the answer to question #2 presupposes the answer to question #1), it is the second question that makes this game show and others like it intrinsically interesting.

5 Switch or stay? Sleeping Beauty on Monty Hall

Suppose Sleeping Beauty is a contestant on "Let's make a deal." She is asked to choose between doors D_1, D_2 and D_3. Behind one of those are two checks for a thousand dollars each; behind the other two is a goat. Sleeping Beauty chooses D_2. Monty Hall offers her the usual deal but with a twist. The twist is that he will only open one of the other doors and give her the opportunity to switch *after* she's been put under and woken up, possibly twice (in which case he will have to administer an amnesia drug the first time). Every time she is awakened she is offered the same deal and Monty Hall will open one of the doors (not necessarily the same one each time) and offer her to switch. If Sleeping Beauty was right the first time and the checks are behind D_2, she will be woken up twice. If she was wrong the first time and the checks are behind D_1 or D_3, she will only be woken up once. That she might have to be drugged once, Monty tells her, is well worth it, as she will get one of the checks each time she picks the right door and could thus conceivably walk away with two thousand dollars.

Sleeping Beauty takes the deal. When awakened, should she switch doors, should she stay with the one she originally picked or should she be indifferent between switching and staying? As with the original version of the Monty Hall Problem and Sleeping Beauty Problem, people will probably respond differently when first faced with this conundrum. Those who have not seen either problem before might want to say that the opening of one of the doors changes the probability that the checks are behind D_2 from 1/3 to 1/2 and that the putting to sleep, waking up and drugging of Sleeping Beauty does nothing to change that probability. Sleeping Beauty should thus be indifferent between staying and switching. Those familiar with the Monty Hall Problem but not with the Sleeping Beauty Problem will split into two camps. The knee-jerk 'halfer' response would be that the probabilities of the prize being behind one door and the goats behind the two others are what they are no matter how often Sleeping Beauty is put to sleep, woken up and drugged. The problem, therefore, is no different from the original Monty Hall Problem and Sleeping Beauty should switch. The knee-jerk 'thirder' response would be that, since Sleeping Beauty is woken up twice as often when she gets it right the first time than when she gets it wrong, her chance of being right the first time goes up to 2/3. Sleeping Beauty should thus stay with D_2.

Imagine an audience that is a mix of these three groups all eager to pipe up and tell Sleeping Beauty what to do. Unfortunately, the audience will have to remain absolutely silent during every awakening. Should Sleeping Beauty have to be woken up a second time, the audience already knows behind which door the checks are.

Unless they have it in for her, they would now all give her the same advice. During the second awakening, Monty therefore cannot allow any audience interference. That will be clear to Sleeping Beauty as well. Hence, Monty cannot allow any audience interference during the first awakening either. Sleeping Beauty should have no way of knowing whether this is her first or her second awakening.

The episode of "Let's make a deal" with the Three Stooges in section 1 is essentially our game-show proxy for this Sleeping Beauty on Monty Hall Problem. Rather than putting candidates or potential contestants (the Three Stooges in this case) to sleep, however, we put them in the kind of cubicles used in "Ignore the odds" (see section 3). We will call this new game-show proxy "Switch or stay?"

As is section 1, we assume the three candidates initially pick D_2. They are now divided in two groups. Two of them are assigned to D_2, one of them is assigned to the other two doors. This game then is different from the one we introduced and analyzed in section 3. There all candidates were assigned to one of the $n - m$ equiprobable outcomes O_i for which $N_i \neq 0$, which means that they all have the same chance $1/(n - m)$ of being chosen as a contestant regardless of the exact values of these non-zero N_i's. In this case, the three candidates are not assigned to equiprobable outcomes. As a result, their chance of being chosen as a contestant depends on the exact values of the non-zero N_i's and they *cannot just ignore* that the initial probability that the prize is behind D_1 or D_3 is twice as large as the initial probability that the prize is behind D_2. In other words, "Switch or stay?" is not simply a special case of "Ignore the odds." Its analysis, however, is completely analogous to the analysis of "Ignore the odds" in section 3.

In this case, we have two (non-equiprobable) outcomes

$$D_2 = \text{Checks behind 2nd door}, \quad \text{not-}D_2 = \text{Checks not behind 2nd door}. \tag{19}$$

The priors are

$$\Pr(D_2) = \frac{1}{3}, \quad \Pr(\text{not-}D_2) = \frac{2}{3}. \tag{20}$$

We have three candidates divided into two groups (one of two, one of one) assigned to these two outcomes:

$$N_{D_2} = 2, \quad N_{\text{not-}D_2} = 1, \quad N = N_{D_2} + N_{\text{not-}D_2} = 3. \tag{21}$$

The likelihoods are (cf. Eq. (6) and the reasoning leading up to it):

$$\Pr(A \,|\, D_2) = \frac{N_{D_2}}{N} = \frac{2}{3}, \quad \Pr(A \,|\, \text{not-}D_2) = \frac{N_{\text{not-}D_2}}{N} = \frac{1}{3} \tag{22}$$

The posteriors are given by Bayes' rule:[5]

$$\begin{aligned}
\Pr(D_2 \mid A) &= \frac{\Pr(A \mid D_2)\Pr(D_2)}{\Pr(A \mid D_2)\Pr(D_2) + \Pr(A \mid \text{not-}D_2)\Pr(\text{not-}D_2)}, \\
\Pr(\text{not-}D_2 \mid A) &= \frac{\Pr(A \mid \text{not-}D_2)\Pr(\text{not-}D_2)}{\Pr(A \mid D_2)\Pr(D_2) + \Pr(A \mid \text{not-}D_2)\Pr(\text{not-}D_2)}.
\end{aligned} \tag{23}$$

From Eqs. (20) and (22), we see that

$$\Pr(A \mid D_2)\Pr(D_2) = \Pr(A \mid \text{not-}D_2)\Pr(\text{not-}D_2) = \frac{2}{9}, \tag{24}$$

which means that the posteriors in Eq. (23) are

$$\Pr(D_j \mid A) = \Pr(\text{not-}D_j \mid A) = \frac{1}{2}. \tag{25}$$

In other words, our contestants (and, by analogy, Sleeping Beauty) should take the advice of those who tell her to simply ignore the putting to the sleep, the waking, the drugging and the opening of one of the three doors! The sampling bias introduced by our peculiar procedure of selecting contestants from an initial group of candidates wipes out the advantage of switching doors, which a contestant would have had under normal circumstances.

Of course, we can just as easily create a version of the Sleeping Beauty on Monty Hall Problem in which she increases her chances of picking the right door from 1/3 to 2/3 by staying. To do that, we replace the values of N_{D_2}, $N_{\text{not-}D_2}$ and N in Eq. (21) by

$$N_{D_2} = 4, \quad N_{\text{not-}D_2} = 1, \quad N = 5. \tag{26}$$

In this case, the Three Stooges in section 1 need to be replaced by, say, the Jackson Five. Inserting the numbers in Eq. (26) into Eq. (22) for the likelihoods, we find

$$\Pr(A \mid D_2) = \frac{N_{D_2}}{N} = \frac{4}{5}, \quad \Pr(A \mid \text{not-}D_2) = \frac{N_{\text{not-}D_2}}{N} = \frac{1}{5} \tag{27}$$

Inserting Eq. (20) for the priors and Eq. (27) for the likelihoods into Eq. (23) for the posteriors, we find

$$\Pr(D_2 \mid A) = \frac{2}{3}, \quad \Pr(\text{not-}D_2 \mid A) = \frac{1}{3}. \tag{28}$$

[5]Eq. (23) is the analogue of Eq. (4) and Eq. (10). However, since $\Pr(D_2) \neq \Pr(\text{not-}D_2)$ in this case, Eq. (23) does not reduce to the analogues of Eqs. (5) and (11), respectively.

In this case, Sleeping Beauty should clearly stay with D_2.

If we assign two or more awakenings to not-D_2, we can create a game-show proxy for the resulting version of the Sleeping Beauty on Monty Hall Problem with *three equiprobable* outcomes, just as in "Ignoring the odds" in section 3, rather than *two non-equiprobable* outcomes as we did above:

$$\begin{aligned} D_1 &= \text{Checks behind 1st door}, \\ D_2 &= \text{Checks behind 2nd door}, \\ D_3 &= \text{Checks behind 3rd door}, \end{aligned} \tag{29}$$

As in section 3, groups of size $N_i > 0$ ($i = 1, 2, 3$) will be assigned to each one of these three outcomes. When we use this format, candidates are told that they will become contestants if and only if they are in the group assigned to the door that has the prize behind it. In "Switch or stay?"—played by the rules analyzed in Eqs. (19)–(25)—candidates are told that they will become contestants if and only if they are either in a group assigned to one door (the one they picked) or in group assigned to *a pair of* doors (the two they did not pick).

Suppose, once again, that the candidates initially pick D_2. If we want contestants to be indifferent between switching and staying in this format of the game show, we need at least four candidates and choose the N_i's as follows:

$$N_1 = 1, \quad N_2 = 2, \quad N_3 = 1, \quad N = 4. \tag{30}$$

Eq. (13) from section 3 tells us that

$$\Pr(D_2 \,|\, A) = \frac{N_2}{N}, \quad \Pr(\text{not-}D_2 \,|\, A) = \Pr(D_1 \,|\, A) + \Pr(D_3 \,|\, A) = \frac{N_1}{N} + \frac{N_3}{N}. \tag{31}$$

Inserting the values in Eq. (30), we see that these posteriors are equal to 1/2 in this case. Once again, Sleeping Beauty should be indifferent between switching and staying.

All these results can easily be generalized from 3 to n doors (and $n-1$ checks!). We will do this only for the format in which we assign one candidate to *all* doors not picked by the candidates and for the special case that the right answer is that contestants (and thereby Sleeping Beauty) should be indifferent between switching and staying. In that case, we need $N = n$ candidates. Suppose they collectively pick the j^{th} door. We can now basically repeat the steps that got us from Eq. (19) to Eq. (25). The two (non-equiprobable) outcomes are:

$$D_j = \text{Checks behind } j^{\text{th}} \text{ door}, \quad \text{not-}D_j = \text{Checks not behind } j^{\text{th}} \text{ door}. \tag{32}$$

The priors are:

$$\Pr(D_j) = \frac{1}{n}, \quad \Pr(\text{not-}D_j) = \frac{n-1}{n}. \tag{33}$$

The sizes of the groups assigned to these two outcomes are:

$$N_{D_j} = n - 1, \quad N_{\text{not-}D_j} = 1. \tag{34}$$

The likelihoods are

$$\Pr(A \,|\, D_j) = \frac{N_{D_j}}{N} = \frac{n-1}{n}, \quad \Pr(A \,|\, \text{not-}D_j) = \frac{N_{\text{not-}D_j}}{N} = \frac{1}{n} \tag{35}$$

The posteriors are

$$\begin{aligned} \Pr(D_j \,|\, A) &= \frac{\Pr(A \,|\, D_j)\Pr(D_j)}{\Pr(A \,|\, D_j)\Pr(D_j) + \Pr(A \,|\, \text{not-}D_j)\Pr(\text{not-}D_j)}, \\ \Pr(\text{not-}D_j \,|\, A) &= \frac{\Pr(A \,|\, \text{not-}D_j)\Pr(\text{not-}D_j)}{\Pr(A \,|\, D_j)\Pr(D_j) + \Pr(A \,|\, \text{not-}D_j)\Pr(\text{not-}D_j)}. \end{aligned} \tag{36}$$

From Eqs. (33) and (35), we see that

$$\Pr(A \,|\, D_j)\Pr(D_j) = \Pr(A \,|\, \text{not-}D_j)\Pr(\text{not-}D_j) = \frac{n-1}{n^2}, \tag{37}$$

which means that

$$\Pr(D_j \,|\, A) = \Pr(\text{not-}D_j \,|\, A) = \frac{1}{2}. \tag{38}$$

In other words, contestants in this game (and Sleeping Beauty in the variation on the Sleeping Beauty Problem for which this game show is a proxy) should be indifferent between staying and switching.

The original Monty Hall Problem becomes more intuitive if we increase the number of doors (n). The original Sleeping Beauty Problem becomes more intuitive if we increase the number of potential awakenings (N). In this new problem that combines the Sleeping Beauty Problem with the Monty Hall Problem the number of doors is equal to the number of potential awakenings ($n = N$). Once again, the problem becomes more intuitive if we increase that number. The design of a game-show proxy for this new problem, moreover, nicely illustrates how variations on the Sleeping Beauty Problem can be turned into game shows that steer clear of the ambiguities plaguing the original problem.

Acknowledgments

We are grateful to Pradeep Mutalik, whose article, "Why Sleeping Beauty Is Lost in Time," in the on-line *Quanta Magazine* (March 31, 2016; Accessed April 17, 2006 at https://www.quantamagazine.org/20160331-why-sleeping-beauty-is-lost-in-time/) inspired our paper. We want to thank to John Norton for helpful comments on an earlier draft and Laurent Taudin for his lovely illustration of our piece. The work of one of us (MJ) was generously supported by Jürgen Renn's Department at the Max Planck Institute for the History of Science in Berlin.