Sleeping Beauty Remains Undecided

Marc A. Burock

ABSTRACT

The Sleeping Beauty problem remains controversial with disagreement between so-called Halfers and Thirders, although the Thirders appear to be leading these days. I analyze three popular arguments for the Thirder position, including the long-run frequency argument, Egla's 'symmetry' argument, and new-information arguments, and find problems with each. The long-run frequency argument is almost unequivocally thought to strongly support Thirders, but in formalizing the argument for an arbitrary number of repetitions, I show that the expected proportion of Heads-Awakenings for a single-trial experiment is unambiguously 1/2. My criticisms of Elga's symmetry argument and the new-information arguments point to subtle misalignments between the narrative/causal description of thought-experiments and the mathematical probability expressions and theory we use to describe these narratives. I end with distinguishing two varieties of possibility—a dynamic forward type and static historical type—that help clarify the Sleeping Beauty problem, nullify the main criticism against Lewis's Halfer argument, and have applicability to probability theory in general.

---

1. Introduction

The so-called Sleeping Beauty Problem is a simple thought experiment involving sleep and memory erasure, made infamous by Adam Elga ([2000]) and David Lewis ([2001]). It goes like this. A researcher is going to put you to sleep on Sunday and then toss a fair coin. If Heads then the researcher will wake you up on Monday, ask you a question, then erase your memory and put you back to sleep until Wednesday. If Tails, the researcher will wake you up on Monday and Tuesday, both times asking you the same question, then erasing your memory, and putting you back to sleep. You will wake on Wednesday for the completion of the experiment with no memory of what happened. Here is the question asked during each awakening: what is your degree of belief (credence) that the coin toss was Heads?

We can imagine this experiment taking place in the actual world, and anyone with an introductory understanding of probability can take a crack at it. Yet starting with seemingly straight-forward premises, an analysis of the problem leads to two (or more) conflicting solutions—a degree of belief in Heads of 1/3 (so-called Thirders) or 1/2 (Halfers). Try for yourself a solution if you have not yet taken a side in the dispute. Despite an impressive literature on Sleeping Beauty by academic scholars and hobbyists, the controversy continues.

Thirders appear to be ahead these days based upon some comments in the literature. In a recent article connecting Everettian quantum mechanics and the SB (Sleeping Beauty) problem, Wilson says that Everettian's 'need not be saddled with the unpopular Halfer conclusion' (Wilson [2018], p. 574). Winkler, in providing a friendly overview of the problem, history, and proposed solutions to Sleeping Beauty concludes that 'the evidence (to me) is that the Thirder position has emerged as the dominant view' (Winkler [2017], p. 581). Luna claims that 'The strong law of large numbers and considerations concerning additional information strongly suggest that Beauty upon awakening has probability 1/3 to be in a Heads-awakening' (Luna [2020], p. 1069).

In what follows I will undermine three popular arguments supporting the Thirder position. I will first analyze the so-called long-run frequency argument for 1/3 from a new perspective and will surprisingly argue that for a single-run of the experiment it favors Halfers. Next, I will look at Elga's original symmetry argument for the 1/3 position and show that only by neglecting relevant admissible information does this argument go through. Then I will show that several SB arguments that attempt to argue that knowledge of being awake counts as new, relevant information lean on so-called analogous examples that are disanalogous to SB in a particular way that invalidates the arguments. I follow with distinguishing two temporal varieties of possibility, that when disentangled, help to clarify the Sleeping

Beauty problem, and have further ramifications for probability theory in general. I end with presenting a neat argument for the Halfer position.

2. Expected proportions of Heads-awakenings for n-trial experiments

One of the strongest arguments supporting the Thirder solution is the so-called long-run frequency argument which goes something like this: if the sleeping beauty experiment is repeated a large number of times, then in the long run the approximate proportion of Heads-awakenings, relative to all awakenings, will approach 1/3. Since awakenings are indistinguishable to SB, the probability of any particular Heads-awakening should be 1/3. Elga made this compelling argument right out of the gate and it continues to persuade many of the 1/3 solution. When we align our credence of an event with the proportion of that event's occurrence over many repetitions, the Thirder solution seems to follow.

There is something correct about this argument on the surface, but the reasoning leaves open the question whether the long-run frequency of Heads-awakenings applies to situations where SB only undergoes one trial of the experiment (or two, or n times in general). Perhaps your credence of being in a Heads-awakening should converge to 1/3 for an infinitely repeated experiment, yet for the single trial experiment your credence ought to be something other than 1/3. During the following I will derive a general expression of this convergence for an arbitrary number of repetitions. Bolstrom ([2007], p. 72) alludes to this 'N-fold' generalization without deriving an exact formula, but our analyses of repetition are similar in spirit.

This derivation will require two sorts of repetition—within and between experiment repetition. In my use of the word experiment, a single SB experiment may include multiple SB trials. So, three tosses of a fair coin with three sets of awakenings constitutes a single experiment, rather than three separate experiments. A three-trial experiment results in three to six separate awakenings, depending upon the coin flips. The sample proportion of Heads-awakening can be calculated for a single three-trial experiment, or for any arbitrary n-trial experiment. This allows us to calculate the expected (or long-run frequency) proportion of Heads-awakenings for an arbitrary number of SB trials per experiment.

To perform these calculations, it helps to think in terms of random variables rather than events. Given an experiment and the corresponding set of possible outcomes (the sample space), a random variable associates a particular number with each outcome of the experiment. For SB experiments, the two obvious candidate random variables are the number of Heads and Tails awakenings, N_H and N_T, respectively. Next associate probabilities with each possible outcome. When the total experiment

involves only one trial (one coin flip), P(N_H= x) = { 1/2 if x=0 ; and 1/2 if x=1 }, where x represents the number of Heads-awakenings. SB will have either 0 or 1 Heads-awakenings during one run of a one-trial experiment, where the probability of each outcome in N_H follows from the toss of a fair coin. Similarly, P(N_T=y) = {1/2 if y=0 ; and 1/2 if y=2 }, where y is the number of Tails-awakenings for a single experiment, 0 or 2 per trial.

From these two random variables, construct a new random variable representing the proportion of Heads-awakenings for a single run of the experiment: P_H=N_H/(N_H+N_T). This new random variable is a function of our two base random variables, and we can use it to calculate the expected proportion (in the statistical sense) of Heads-awakenings for a one-trial experiment. To do so, compose the joint probability function $p(N_H,N_T)$ of N_H and N_T, as per figure 1A, and then apply the definition of expected value to our random variable P_H, giving $\mathbf{E}[P_H]$=1/2. [1] The steps are as follows:

N_H = Number of Heads Awakening per experiment
N_T = Number of Tails Awakenings per experiment

$$P_{H(1)} = \frac{N_H}{N_H+N_T}$$

$$\mathbf{E}\left[P_{H(1)}\right] = \mathbf{E}\left(\frac{N_H}{N_H+N_T}\right)$$

$$= \sum_{N_H,N_T}\left(\frac{N_H}{N_H+N_T}\right)p(N_H,N_T)$$

$$= \left(\frac{1}{1+0}\right)\frac{1}{2} + \left(\frac{0}{0+2}\right)\frac{1}{2} = \frac{1}{2}$$

This result may seem unexpected at first – how can the expected proportion of Heads-awakenings be 1/2 for one run of the SB experiment? But when thinking in proportions, half of the time the proportion of Heads-awakening will be 1 (all Heads-awakenings), and half the time it will be 0 (all Tails-awakenings). The mean or expected proportion is thus 1/2 for a single trial. This is like calculating the expectation of the sample proportion of a binomial random sample with n=1 trials and probably of success equal to 1/2. You may be tempted to say this is just counting experimental trials that resulted Heads relative to the total number of experiment trials, but this is not correct because we are not counting experimental trials at all, we are directly counting awakening events for a single trial and calculating the proportion of Heads-awakenings.

---

[1] The statistically wrong way to calculate the expected proportion of Heads-awakenings would be to apply the expectation operator to the numerator and denominator separately: $\frac{\mathbf{E}(\mathrm{N_H})}{\mathbf{E}(N_H+N_T)} = \frac{\mathbf{E}(N_H)}{\mathbf{E}(N_H)+\mathbf{E}(N_T)} = \frac{\mathbf{E}(N_H)}{\mathbf{E}(N_H)+2\mathbf{E}(N_H)} = 1/3.$ This invalid procedure, which is independent of the number of trials, gives the Thirder result.

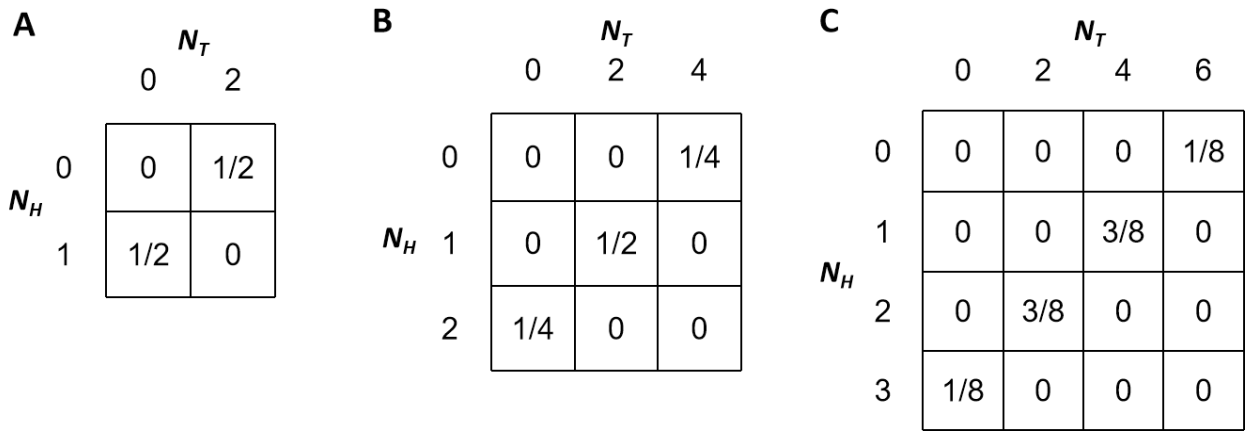

Figure 1. Joint probability distribution $p(N_H,N_T)$ for (A) one-trial experiment. (B) two-trial. (C) three-trial.

It's interesting to consider the 1/3 solution to SB for a one-trial experiment when thinking in terms of proportions of Heads-awakenings. Thirders should also say that the expected proportion of Heads-awakenings, for 1 trial experiments, ought to be 1/3 to align with their credence or probability of a Heads-awakening being 1/3. But the proportion of Heads-awakenings, for 1 trial, will either be 1 or 0 and nothing in between, with an explicit expectation of 1/2. The fact that SB either has two Tails-awakenings versus one Heads-awakening per trial does not have any bearing on the proportion of Tails-awakenings for a single trial—the proportion is either 0 or 1 for a single trial. A Thirder is forced to argue that the expected proportion of Heads-awakenings *for a single trial* ought to be different than her credence in a Heads-awakening—a credence that many Thirders support by invoking the long-run frequency argument which is an argument about proportions of Heads-awakenings.

It becomes more obvious that this calculation reflects something other than trial counting when a complete SB experiment encompasses more than one trial per experiment. A multiple trial SB experiment is one where the coin is tossed multiple times, and SB is awakened one or two times, depending upon an outcome of Heads or Tails for each coin toss. SB's memory is still erased after every awakening. A two-trial SB experiment then has four potential outcomes for the two-coin tosses (HH, HT, TH, TT), which results in two Heads-awakenings for HH, one Heads-awakening and two Tails-awakenings for HT or TH, and four Tails-awakenings for TT. Construct random variables for the number of Heads and Tails-awakenings similar to the one-trial experiment, with probabilities assigned according to Figure 1B. The expected proportion of Heads-awakenings for the two-trial experiment $\mathbf{E}[P_{H(2)}]$ is then:

$$\mathbf{E}\left[P_{H(2)}\right] = \left(\frac{2}{2+0}\right)\frac{1}{4} + \left(\frac{1}{1+2}\right)\frac{1}{2} + \left(\frac{0}{0+4}\right)\frac{1}{4} = \frac{5}{12}$$

The expected proportion of Heads-awakenings for the two-trial experiment is less than 1/2, but still not equal to 1/3. The possible proportions of Heads-awakenings for a two-trial experiment are 1, 1/3, and 0; with probabilities of 1/4, 1/2, and 1/4, respectively. A Thirder will generally be uncomfortable with this two-trial result, just as he was uncomfortable with the one-trial result. The Thirder's credence in a Heads-awakening for the two-trial experiment will continue to be 1/3, especially when appealing to the long-run frequency argument—he has little reason to differentiate a two-trial experiment from the one-trial experiment. Yet we do have good reason to differentiate the experiments—the expected proportions of Heads-awakenings differ between the two experiments, and this knowledge should be considered and explained by the Thirder, regardless of the justification for 1/3.

Although we use the expectation operator to calculate the result 5/12, we could also repeat the two-trial experiment many times, and calculate the sample mean of the proportion of Heads-awakenings for each experiment to arrive at the same result. A sequence of proportions of Heads-awakenings for the two-trial experiment might look something like { 1/3, 1, 1/3, 0, 0 ,1/3, 1/3, 1, 1, 1/3,…}, and for large enough repetitions, the average would converge to 5/12.

The three-trial SB experiment results in the following expected proportion of Heads-awakenings, where the possible outcomes for coin tosses are (HHH, HHT, HTH, THH, TTH, THT, HTT, TTT), and probability assignments for N_H and N_T in figure 1C:

$$\mathbf{E}\left[P_{H(3)}\right] = \left(\frac{3}{3+0}\right)\frac{1}{8} + \left(\frac{2}{2+2}\right)\frac{3}{8} + \left(\frac{1}{1+4}\right)\frac{3}{8} + \left(\frac{0}{0+6}\right)\frac{1}{8} = \frac{31}{80}$$

This expected proportion is less than 5/12, and it appears that the expected proportion of Heads-awakenings decreases as the number of trials per experiment increases. We can also begin to appreciate a pattern in these calculations and induce the general form for an arbitrary number of trials per experiment. The terms in parentheses are the possible proportions of Heads-awakenings for the complete experiment. The numerator is the number of Heads-awakenings, and the denominator is the total number of awakenings. For the case of n=3 trials, the numerator of each proportion goes incrementally from 3 to 0 in the four proportions, and the denominator goes from 3 to 6. The pattern is similar for *n*=2. The probabilities for each proportion correspond to the binomial probability distribution Binomial(*n*, 1/2), like those found in sequences of independent fair coin tosses. Putting these pieces together:

$$\mathbf{E}\left[P_{H(n)}\right] = \frac{1}{2^n}\sum_{k=0}^{n}\frac{n-k}{n+k}\binom{n}{k}$$

Which is the general form for the expected proportion of Heads-awakenings in the SB experiments with an arbitrary number or trials (coin flips) per experiment. The formula works for the examples above. SB's credence—if it tracks expected proportions—of being in a Heads-awakening is dependent upon how many repetitions she undergoes. What happens as the number of trials per experiment goes to infinity? Computational analysis suggests that:

$$\lim_{\mathrm{n}\to\infty}\mathbf{E}\left[P_{H(n)}\right] = \frac{1}{3}$$

Readers more proficient at limits may solve the limit directly, although plotting out values for large *n* strongly suggests the 1/3 result. The expected proportion of Heads-awakenings approaches 1/3 as the number of trials per SB experiment approaches infinity. This is the result Thirders are after when appealing to the long-run frequency argument, but it is only valid when the experiment includes an infinite number of trials per experiment. Further, Halfers should generally agree with this result, and there is some unification between the two camps in the limit. It would be odd, however, for a Thirder to agree that this limit and the analysis leading up to the limit are valid, yet deny that the analysis, when applied to a finite number of trials per experiment, is somehow invalid. Everything leading up to this limit result only considered awakening events and proportions of awakening events using canonical statistical procedures.

Although multiple authors have argued that the long-run proportion of Heads-awakenings converges to 1/3 in the infinite limit, no one has produced a correct formal expression for that limit before. The expression is complicated because the denominator—which represents the total number of awakenings over repetitions—is a random quantity and not a constant, and because the numerator and denominator are dependent quantities. If the experiment is repeated 100 times, for instance, the total number of awakenings will vary between 100 and 200, and will be dependent upon the number of Heads-awakenings. We cannot compute the relative frequency of Heads-awakenings by simply dividing the number of Heads-awakenings by a constant representing the number of repetitions.[2]

[2] Luna([2020], p. 1073) does imply the long-run frequency result of 1/3 can be calculated by dividing the number of Heads-awakenings by a constant representing the number of trials. The specific calculation is not demonstrated.

Perhaps the expected proportion of Heads-awakenings has nothing to do with one's credence in a Heads-awakening, but that position presents problems for the long-run frequency argument for 1/3, which is an argument about proportions in a limit sense. Further, it is not clear that the infinite long-run proportion of Heads-awakenings applies to an experiment of just one trial. Our result formally demonstrates that it does not and turns the long-run frequency argument on its head. When repetition is formally analyzed and considered for an arbitrary number or trials per experiment, the long-run frequency argument supports the 1/2 position for SB. That is, we get the long-run frequency result of 1/3 *through* the 1/2 result of a single trial experiment by appealing to a uniform way of analyzing proportions of Heads-awakenings with arbitrary trials per experiment, thus explicitly connecting the single trial result to the infinite trial result. The naive long-run frequency argument only applies to the infinite case. Minimally, this analysis demonstrates that the long-run frequency argument is ambivalent about 1/2 or 1/3 for SB experiments involving a single trial.

3. Asymmetry of Elga's symmetry argument

When you wake up during the course of the SB experiment, you reason that you are in one of three possible waking situations:

H1 HEADS and it is Monday.
T1 TAILS and it is Monday.
T2 TAILS and it is Tuesday.

Elga derives P(*H1*)=1/3 by arguing for two separate equalities, and then transitively stitches those equalities together. The first equality I will call the Tails Equality and is it relatively uncontroversial and I accept it. Briefly, if the coin lands Tails, then the waking on Monday is indistinguishable from the waking on Tuesday, so SB's credences P(*T1*) and P(*T2*) ought to be equal. Good enough. Our concern for this section is rather the Monday Equality which concludes P(*H1*)=P(*T1*). Proving the Monday Equality requires us to consider two distinct methods of performing the Sleeping Beauty experiment:

M1: first tossing the coin Sunday night and then waking you up either once or twice depending on the outcome; or

M2: first waking you up once on Monday, and then tossing the coin Monday night to determine whether to wake you up a second time.

Elga says that upon awakening the subject's credence in Heads ought to be the same regardless of whether **M1** or **M2** is used to do the experiment, and proceeds to reason assuming that the researchers use **M2** to justify the Monday equivalence. Lewis agrees that the experimental method ought not change the result, yet the method of the experiment does influence the process of our reasoning, and Elga explicitly makes use of **M2** to make the argument, so the informational content of experiments differ. It would be helpful to clarify these differences and similarities.

First the similarities. Both ways of performing the experiment produce the same statistical distribution of awakening events with respect to the day of the week and side of the coin, when viewed from the third-person or uncentered perspective. An outside observer running the SB experiment will tabulate the same long-run frequency of Tails-awakening and Heads-awakenings for both ways of performing the experiment. And the distribution of Monday awakenings and Tuesday awakening will be the same (in the long run), and also for the distribution of Monday Tails-awakenings relative to Tuesday Tails-awakenings. No matter how we carve up the events—whether by the side of the coin, awakening-events, or by day of the week, the observed distributions will be identical in the long-run for both ways of performing the experiment. The two methods are *outcome* equivalent.

The differences between the two experiments have nothing to do with the observed statistical outcomes of events, which is why Elga and Lewis and most subsequent authors have felt that SB's credence ought to be the same regardless of the protocol. Yet although her credence ought to be the same, her process of reasoning and information available need not be identical, rather, as Elga makes explicit, SB need consider **M2** to complete her argument for the 1/3 position. But how does the alternative arrangement **M2** pull in new information useful for solving the sleeping beauty problem given it produces statistically equivalent outcomes?

Timing of the coin flip differentiates **M1** and **M2**. The coin is flipped on Sunday night under **M1** but on Monday night for **M2**. This adds another piece of uncertainty to SB's uncertainty about the side of the coin and the day of the week upon awakening—specifically, uncertainty as to whether the coin was already flipped or not when she first awakens. While this uncertainty does not exist under **M1**, since the coin is flipped Sunday night and you know this upon awakening; under **M2** you may awaken on Monday with the coin still to be flipped, or Tuesday with the coin flip already decided. When you first awaken under **M2**, you do not know if the coin flip has occurred or not, but if told it is Monday, then you also know that the coin flip has yet to occur. Elga uses this crucial piece to then argue that SB's credence in Heads and Monday is equal to her credence in Tails and Monday, but this equivalence is actually an

equivalence of conditional credences given that the coin has yet to be flipped, in addition to conditioning on Monday.

Assuming **M2**, Elga says that being told it is Monday amounts to you having the conditional credence P(*H1*|*H1* or *T1*), and that this credence ought equal 1/2 because the chance of a fair coin yet to be tossed is 1/2. From P(*H1*|*H1* or *T1*)=1/2 it follows that P(*H1*)=P(*T1*). I think this leg of the argument is unintentionally misleading. The reason Elga can say that P(*H1*|*H1* or *T1*)=P(*T1*|*H1* or *T1*)=1/2 is only indirectly, and not necessarily, related to knowledge that SB is told it is Monday upon awakening. Rather, the Principal Principle and knowledge that *a fair coin has yet to be flipped* drives the intuition for P(*H1*|*H1* or *T1*)=1/2, but Elga's derivation ignores the information and uncertainty of the coin toss itself. If *C* is the event 'the coin was already flipped', and ¬*C* 'the coin was not flipped yet', then this leg of the argument should start with the credence P(*H1*|(*H1* or *T1*) and ¬*C*)=1/2. In words, your credence in being in a Heads-awakening, given that it is Monday *and* the coin was not yet flipped is 1/2. But now the Monday equivalence P(*H1*)=P(*T1*) does not follow.

We need event *C* to distinguish the methods under analysis. The expression P(*H1*|(*H1* or *T1*) and ¬*C*) models our complete narrative information for **M2** after being told it is Monday, just as P(*H1*|(*H1* or *T1*) and *C*) does so for **M1**. Being pulled out of the SB experiment under **M2** and told that it is Monday provides immediate information about *H1* precisely because learning it is Monday tells you that a fair coin has not yet been flipped. Under **M1**, being told it is Monday does not tell you anything directly about *H1* because the coin flip already occurred and past events that have been decided are no longer chancy—which is why the alternative arrangement was imagined. But the only thing Elga's Monday equivalence argument tells us under **M2** is that P(*H1*|¬*C*)=P(*T1*|¬*C*)=1/2, and that does not help solve the problem. This result cannot be attached to the unconditional Tails equivalence argument that P(*T1*)=P(*T2*) to give the Thirder result[3].

To save the Monday Equality, we need to argue that conditioning on ¬*C* is unnecessary and can be dropped without loss of generality, but without this explicit conditional, then we don't know if the credence expression applies to the case of **M1** or **M2**. Yes, the information is in the narrative of **M2**, but this narrative knowledge should also align with and be included in the particular credence expression as well. For instance, I may in narrative say that SB is awakened and told it is Monday and ask about your

[3] We can attempt to make a conditional Tails equivalence argument, such as *P*(*T1*|¬*C*)=*P*(*T2*|¬*C*), in order to line up the credences for the Monday Equality, but there is an immediate problem: *P*(*T2*|¬*C*)=0. Your credence in being in a Tails-awakening on Tuesday given the coin was not tossed is zero because the conditional implies it's Monday. The Tails equivalence argument appears to be invalid under **M2**, yet the Monday equivalence requires **M2** to even begin.

credence in Heads for that situation, but I wouldn't claim that writing the expression P(*H1*), without the conditional, adequately reflects the credence I am asking about, even though conditionalizing on Monday is implied in the narrative. An author cannot simply drop a relevant variable from a credence expression that will be subject to further manipulation using the probability calculus—but this is precisely how Elga derives the Monday Equality.

Further, knowledge if the coin flip has been decided dominates knowledge about Monday for the following reason: if you did not know which method the experiment was performed and you were told, upon awakening, that it was Monday, you could not directly reason that P(*H1*|Monday)=1/2; yet being told that the coin was not yet flipped justifies P(*H1*|¬*C*)=1/2 immediately. Similarly, coin flip knowledge also allows you to know which way the experiment is being performed if you were not told in advance, whereas being told it is Monday does not differentiate the experiments. Knowledge of ¬*C* is quite informative, which makes ¬*C*'s absence from these expressions even more glaring.

The two methods for performing the experiment produce identical long-run statistical outcomes, but reasoning about the evidence, when considering knowledge of whether the coin was flipped or not, does not permit Elga's original argument to go through. Elga's Thirder argument requires ignoring evidence about the coin flip, but this seems to be an impressive oversight given ongoing discussions about the importance of subtle indexical knowledge in the SB problem. Knowledge that the coin was flipped and decided seems to be as important as knowledge that 'Today is a waking day' for instance.

When including knowledge about the coin toss, it is also easy to explain the double-Halfer solution and to clarify Lewis's odd claim that P(*H1*|Monday)=2/3. P(*H1*|Monday and ¬*C*)=1/2, but P(*H1*|Monday and *C*)=2/3 and P(*H1*)=1/2. Wilson ([2014], p. 588) acknowledges that 'Halfers must say there is a difference between the toss-on-Sunday and toss-on-Monday cases' but incorrectly concludes that when 'Beauty knows the coin is tossed Sunday night, Halfing is the correct solution; but that in a case in which Beauty knows the coin is tossed Monday night, thirding is the correct solution.' I say that Halfing appears to be the correct solution regardless of which way SB performs the experiment, it just so happens that **M2** cannot be used to justify the 1/2 or the 1/3 solution because the causal efficacy of the fair coin yet to be tossed dominates probabilistic reasoning, independent of other conditional evidence. This leg of Elga's argument cannot be coupled to the leg involving the symmetry between Monday and Tuesday Tails-awakenings when the coin was known to be tossed Sunday night. Luna ([2020], p. 1087) also concludes (for very different reasons) that 'it is very dubious that a result based on the alternative arrangement can be transferred to the original one.' The above analysis formalizes Luna's instinct.

4. One-sided new information

Let *W* be the event 'I am awake now'. Many Halfers simply do not grasp how Thirders can claim that SB obtains new relevant information given *W*—relevant in the sense that it would compel SB to update her so-called prior or preliminary probabilities for each awakening event (Lewis [2001]; Pust [2008], [2014]; Bradley [2003], [2010]; White [2006]). I count myself one of them. Although *W* seems like an informative statement, this information has never surprised me or reduced my uncertainty in the information theoretic sense. Nonetheless, the thoughtful work of Dorr ([2002]), Horgan ([2004]), Karlander and Spectre ([2010]), Arntzenius ([2003]), Weintraub ([2004]) and others, contain reasonable arguments that SB does receive new relevant information upon awakening.

The literature on conditionalization is deep, and my narrow focus here is upon the authors who attempt to convince us that SB receives new relevant information upon awakening by constructing supporting thought experiments, that, although not identical to the SB experiment, claim to model a similar sort of 'waking day' evidence in a more transparent and immediately obvious mode. These analogous experiments (as I will call them) show us how waking-day type evidence can be used to conditionalize on a prior probability distribution to ultimately produce a 1/3 credence for an outcome in the analogous experiment. The authors then argue that it would be unreasonable to believe that these analogous experiments differ enough from the SB experiment to justify anything other than 1/3 for SB as well. I argue that these analogous experiments, in sharing a common feature not contained in the SB experiment, hold the key to understanding why Halfers should reject several new-information arguments based upon analogy, generalizing the initial insight of Bradley ([2003]) to other cases.

Let's first look at Karlander and Spectre's analogous experiment that they call 'BELL', which is well thought-out and explained. This analogous SB experiment also includes memory erasure and a fair coin toss but differs in that you are guaranteed to be awake on both days and adds a bell ring. You are put asleep Sunday night and a fair coin is tossed. If the coin lands Heads a bell rings at 6:00 pm on Monday but not on Tuesday (although you are awake), and if Tails the bells rings at 6:00 pm on both days. Your memory will be erased Monday night before sleeping, so you will not know what day it is during the experiment.

BELL attempts to convince us that holding prior probabilities in the SB experiment is justified, and that updating these priors after becoming awake leads us directly to the Thirder solution. Analysis of BELL goes like this: prior to going to sleep Sunday night, your prior probability of awakening in a bell-ringing day is 3/4 (because 3 of the 4 awakenings are associated with bells) and 1/4 of awakening in a not-bell ringing day. Let us suppose you start the experiment and are awakened. At 6:00 pm you hear

a bell, then your conditional probability in being in a Tails-awakening given hearing the bell is P(Tails|Bell)=P(Tails and Bell)/P(Bell)=2/3. I completely agree with Karlander and Specter's analysis for BELL, including the priors and conditionalization, but it is not analogous to Sleeping Beauty for the following reason: you can conditionalize on hearing a bell *and* not hearing a bell similarly in BELL , but not in SB for awakenings. In fact, you will conditionalize on not hearing a bell in about 1/4 of all awakenings if BELL was repeated many times. The complete set of conditional probabilities—the updated probabilities you ought have after 6 pm – are P(Tails|Bell)=2/3, P(Heads|Bell)=1/3, P(Tails|¬Bell)=0, and P(Heads|¬Bell)=1. These are all live possibilities and credences that you should have in BELL if updating according Karlander and Specter. In contrast, SB will never conditionalize on not-waking, not once.

The live possibility of conditioning on an event and its negation is the crucial difference between these analogous experiments meant to push us to consider awakening as evidence in the SB experiment. While becoming awake provides some sort of information, SB will never condition on ¬*W*, making *W* a performative tautology during the SB experiment–it is a true statement, but its negation will never be satisfied in the original sleeping beauty experiment, and it is the possibility of this negation which must be satisfied to complete any argument for conditionalizing on the full probability space. The analogous SB experiments allow for complete conditionalization which I suspect makes them more compelling, but disanalogous to the original.

It is one thing to condition on the negation in symbolic terms when manipulating probabilities, and another to imagine, in the actual world, to condition on that negated evidence. When contemplating *W* we imagine SB or ourselves becoming awake during the course of an actual experiment in the world and can condition on that evidence in some sense. The event *W* lives in symbols on the page as well as a part of what it means to receive actual evidence through some mechanism in the world. In contrast, the negated event ¬*W* never becomes a piece of information or evidence that I will condition upon while participating in the SB experiment. I claim that to justify the full conditional space, you must be able to condition on each of the conditional pieces of evidence in the sample space via a similar mechanism or process. While SB is not awake on Tuesday after a Heads toss, that is a third-person observation, and not a piece of information available to SB at any point under repeated experiments. Nor does being not awake, in-itself in the original SB experiment, trigger any mechanism that could be thought as updating credences.

Here is another way to put it. Assume that state *A* represents your prior or unconditional credence space on Sunday prior to being put asleep. Assume that while asleep, your unconditional credence

space *A* persists in some way, unchanged. Then, if it is Monday you will awaken, note the new evidence *W*, and can condition on the evidence, producing a change in your credence space to state *B* for the SB experimental outcomes. However, if it is Tuesday and Heads, your updated credence space state *B* will have been erased and set back to *A* (you have no memory of awakening Monday), you will remain asleep the entire day Tuesday, and you will remain in the unconditional credence space *A*. There is no reason for your credence space to change. You will never note that ¬*W* and condition on that information to conclude that your credence in Heads is 1, but that situation is precisely what is required to justify the complete conditional space. If a piece of information can be used to update credences, then the negative of that information must also be possible (at some point, under multiple repetitions) in order to justify moving to the conditional space in total. Again, while we can manipulate probabilistic symbols equally for waking or not-waking, the actual events waking and not-waking in the world need to also mechanistically move us from the unconditional to the conditional space.

I am not restricting my conception of evidence reception to that human consciousness when I suggest these things, rather, whatever the mechanism of evidence reception is, that mechanism must be similarly receptive to the statement and it's negation. The negation must initiate some sort of causal downstream process to be relevant in the conditional space, just as the affirmative must do so, in some way, to differentiate itself from the unconditional prior space. In the original SB experiment, this causal change does not occur when you sleep through Tuesday – and it's not so much about the sleeping or consciousness as it is about a lack of change. When the coin lands Heads, a causal event that correlates with updating your credence to P(Heads|¬*W*)=1 on Tuesday simply does not exist. Further, during the experiment, you are also asleep on Monday night, and this state of sleeping is no different than the state of sleeping on Tuesday. What then is SB's credence on Monday night for P(Heads|¬*W*), for instance?

Experiment BELL contains the absence of an event: the lack of hearing a bell sound at 6:00 PM. While this absence may appear analogous to being asleep in the SB experiment, it is not. In BELL, the time 6:00 PM functions as a trigger or decision point determined by prior information in the experimental set-up. Although the bell sound is not present before or after 6:00 PM on Tuesday when the coin lands Heads, the lack of a bell after 6:00 PM triggers an update in knowledge, specifically P(Heads|¬Bell after 6:00)=1. In contrast, when you sleep during the original SB experiment (on Sunday night, Monday night, and Tuesday day), there is no trigger or event that correlates with an update in knowledge. This sort of trigger need not rely upon human consciousness—it could be envisioned by a mechanical device with a timer and sensor, that depending upon the presence or absence of a bell

sound after 6:00 PM, initiates differing causal events. There is no such trigger for the sleeping periods in the original SB experiment.

Weintraub ([2004)] puts forth an analogous SB experiment meant to show how becoming awake can be seen as new relevant information, but this analogous experiment harbors a similar disanalogy as BELL. In this variant, you are put asleep, a fair coin is tossed, and upon awakening are shown three light flashes in sequence. Your memory is erased after each flash. If the coin lands Heads, one of the three flashes will be red and the two others green. If the coin lands Tails, one will be green and two will be red. Weintraub says you should obviously believe P(Heads|Red)=1/3 , and I again agree with the analysis here, but it is not analogous to SB. You can condition—while in the middle of this experiment—on seeing a green flash as well as a red flash, with the complete set of conditional probabilities being P(Heads|Red)=1/3, P(Heads|Green)=2/3, P(Tails|Red)=2/3, and P(Tails|Green)=1/3. SB can do nothing of the sort when colors are taken to be analogous to awakening or not.

Dorr ([2002]) imagines an analogous SB experiment where you will definitely awaken on both Monday and Tuesday, with drug-induced amnesia on Tuesday, but the type of amnesia drug you are given depends upon a fair coin toss on Monday night. If Tails, you will receive the amnesia-inducing drug from the original SB experiment. If Heads, you will receive a weaker amnesia drug that only delays (by minutes) the onset of memories from the previous day, with the consequence that you will remember being in an experiment on Monday and conclude it is Tuesday (with certainty) and that the outcome was Heads. Since you will also know, with certainty, when you did not receive the weaker drug, you can conditionalize on this 'negated' event as well. Knowing you took the weaker drug is equivalent to knowing it is a Heads-awakening and Tuesday (*H2*). Upon awakening, you can conclude that P(*H1*|¬*H2*)=P(*T1*|¬*H2*)=P(*T2*|¬*H2*)=1/3 or P(*H1*|*H2*)=P(*T1*|*H2*)=P(*T2*|*H2*)=0, depending upon your evidence. This sort of complete conditionalization does not occur in the original SB experiment. Bradley ([2003], p. 268) first made a similar observation when commenting on this variant, saying 'There is no possibility that Beauty will find out for certain that it fell Heads. So she cannot update on the lack of such information.' I agree.

In an analogous experiment by Arntzenius ([2003]), everything in the original SB experiment occurs, except SB is also assumed to be a frequent dreamer, and in fact if SB is not awake at 9 AM, she dreams that she is woken up at 9 AM anyway, and that she cannot (initially) distinguish dreaming from actually waking in reality. There are then four prior probabilities in this experiment: Monday & Tails & Awake, Monday & Heads & Awake, Tuesday & Tails & Awake, and Tuesday & Heads & Dreaming. Arntzenius adds that SB can know for certain she is awake by pinching herself—if she feels pain after pinching, she

knows she is awake, but if dreaming the pinch does not hurt and she knows she is not awake. SB can conditionalize both on being awake or not, and thus this analogous experiment harbors the same disanalogy to the original SB experiment as the others.

The lesson here is that symbolic conditioning on events using probabilistic expressions ought to track the causal events that generate those updates. During the original SB experiment, this alignment exists for P(Heads|*W*), where awakening may be viewed as a causal event capable of updating a credence. In contrast, for P(Heads|¬*W*), continuing to be asleep is not associated with a new causal event capable of providing evidence or new information *to the participant of the experiment*.[4] The expression P(Heads|¬*W*) is then either undefined, or made irrelevant by setting P(¬*W*)=0. If P(Heads|¬*W*) is undefined, then the entire joint probability distribution between the side of the coin and waking-state becomes undefined or at least problematic. Analogous experiments by Thirders attempt to evade this dilemma by positing fully aligned conditioning. I happen to agree that 1/3 is the solution for most of these variant cases with full conditioning, but I think 1/2 is correct for the original SB case.

A simple modification of the SB experiment creates further challenges for those who attempt to assign preliminary probabilities to SB-style events and then condition on awakening. Suppose that if the coin lands Heads, SB will still be awoken on Monday, but instead of only sleeping through Tuesday, SB will sleep for seven days prior to being awoken and told the experiment is over. If the coin lands Tails nothing changes; SB will be awoken on Monday and Tuesday and told the experiment is over on Wednesday. Those who believe assigning preliminary probabilities to days when SB will be asleep (during the experiment) should do so for each of the days SB will be asleep if the coin lands Heads. If you assign equal probabilities to each of the possible days of the experiment (3 waking days and 7 sleeping days), then your preliminary probability of Heads will be 4/5—which is difficult to justify given the coin is fair.

The other alternative is to preserve the preliminary fairness of the coin by preliminarily assigning 1/4 to each Tail's awakening, and 1/16 to each Head's day (1 waking and 7 sleeping). Upon awakening if you condition on being in a waking day then your updated probability for Heads will be 1/9. Your credence in Heads upon awakening will be highly dependent upon the number of sleeping days. Taken to the extreme, as the number of sleeping days given Heads increases, your credence in Heads upon awakening gets closer to zero, even though you know you will wake after a fair coin toss on one of three

[4] Watching SB sleep through Tuesday and Heads is informative to a third-person observer of the experiment, but not to SB. Gao ([2018]) believes carefully separating first and third-person perspectives solves the SB problem.

days. Halfers in general will not have a difficulty with this modification of the SB experiment since they do not assign any preliminary credence to sleeping days.

If you believe both that assigning preliminary probabilities to sleeping days and that subsequent conditionalization of being in a waking-day are justified, then changing the number of sleeping days in the experiment results in unintuitive credences regardless of how preliminary credences are assigned. This consequence has not been explored by those who advocate that new, relevant information is obtained from awakening during the SB experiment.

5. Two Varieties of Temporal Possibility

Luna makes a pertinent observation that the SB experiment involves two distinct types of possibility, and I suspect this feature of the experiment fuels much of the confusion and difficulty in analyzing the problem. In Luna's language:

> Note that the outcomes of Beauty's experiment are not ordinary A-worlds, that is, A -worlds representing the branching future of the world resulting from a physical random experiment (e.g. what side the coin came up), but self-locating alternatives ensuing from an epistemic random experiment (i.e. Beauty's asking what day it is). The only physical random experiment branching the future into ordinary A-worlds in our story is the coin toss. (Luna [2020], p. 1080).

An 'A-world' for Luna is a possible future of the actual world in the sense of Aristotelian potency; the potential for a physical object like a coin to land on one side or the other, as opposed to a metaphysical possibility. This is the sort of future possibility associated with objective chance in the Lewisian sense, or the probabilistic propensity for a physical system to dynamically enter one future state versus another. We typical think of these possibilities as developing causally through time. The physical events cause the world to branch down a particular path, where each world branch can be associated with an objective probability for those who believe in such things. As opposed to this future-oriented branching variant of possibility, within the SB experiment, SB also experiences a (primarily) epistemic situation regarding when along a particular branch she finds herself upon waking. Awakening on either of two possible days, Monday or Tuesday, is not associated with a physical event or potency that branches the world one way or the other. Waking up does not dynamically cause the day to become Monday or Tuesday, rather, the day has been set already in-advance—SB simply does not know what day it is.

The common way to draw this distinction in philosophical SB discussions is using the language of centered vs uncentered worlds, yet the proximate distinguishing feature of this possibility involves

temporality—future versus the past and present. Let's call these two concepts of possibility *dynamic* possibility and *static* possibility. Dynamic possibility is future-oriented and evolves in time, while static possibility refers to a present or past unknown state of the world. This distinction follows from the common metaphysical assumption that the future is open in some sense, whereas the present and past are fixed. A fair coin about to be tossed can land Heads or Tails, where Heads and Tails are dynamic possibilities of the toss. After the coin is tossed and the side decided, if the outcome is unknown to you then the possibility for the coin to be Heads or Tails is a static possibility. The side has been fixed, and the possibility exists epistemically alone. Wilson ([2014]) draws a distinction along similar lines with his concept of effective chanciness – which aligns somewhat with our notion of dynamic possibility.

Static possibility is a concept of possibility that includes centered possibility as a subset. Asking 'what day is today?', which is typically understood to involve centered possibility, is also a static possibility of the day, just as 'what side did the coin land?' presumes a static possibility of a past event. Carroll and Seben ([2018]), when discussing Everett's many-worlds approach to quantum mechanics and the quantum Sleeping Beauty variant, similarly make the distinction between these two sorts of uncertainty. A quantum event or measurement may cause the wavefunction to dynamically branch into multiple world branches associated with specific probabilities—so-called between-branch possibility or uncertainty—in addition to the uncertainty of events that occur within a specific branch, which aligns with the concept of static possibility.

The distinction between dynamic and static possibility hopefully makes a difference, and before I attempt to demonstrate this, I suggest the following principle that explains why we have previously overlooked the distinction:

Unconditional Temporal Continuity. Let H be an event that is decided at time t. H_D is the dynamic possibility of H at time t- prior to t, and H_S is the static possible state corresponding to H at time $t+$ later than t. If between t- and $t+$ you acquire no new admissible information regarding the outcome of H other than the fact that H was decided at t, then your unconditional credence $P_{t-}(H_D)$ should be equal to your unconditional credence $P_{t+}(H_S)$.

Suppose someone flips a fair coin at time t. Prior to the flip, the outcome is an objective chancy event in the Lewisian sense and your credence in Heads should be 1/2. If at a later time $t+$ you know the coin was flipped, but you do not know the outcome of the flip, your credence in Heads for that toss should still remain 1/2, even though there is no chance for the outcome to be any other way than it is.

The result seems obvious and the principle perhaps unneeded, but things become more complex when conditional evidence is explicitly considered in your credence.

Suppose there is a light that is red, it is now on, and you are about to flip a fair coin. If the coin lands Heads the light will change from red to green with 50% chance, but if the coin lands Tails the light will stay red. Your conditional credence of the dynamic possibility $P(H_D|Red)$ should be equal to 1/2—it is a fair coin yet to be tossed, and the color of the light ought not change that. The situation for $P(H_S|Red)$ differs. If the light is red after the toss, but I don't know the result, then my conditional credence in Heads after the toss is $P(H_S|Red)=1/3$. Therefore $P(H_D|Red)\neq P(H_S|Red)$ and the temporal continuity principle does not apply in this conditional setting. Knowledge that an event occurred can be admissible evidence when updating a conditional credence and does not require a mysterious violation of the Principal Principle. More, it would be an error to assume that $P(H_D|Red)=P(H_S|Red)$ and perhaps lead to a contradiction if H_D and H_S were not properly distinguished.

It's not hard to imagine a sleeping beauty type experiment somewhat similar to this color flip. Take the situation above and suppose that you awaken either before or after the coin flip—you don't know which one—and that the light is red when you awaken. Now, what is your credence that the coin will land or has landed heads given that the light is now red? I used the tensed verbs to illustrate the two sorts of possibility contained in the uncertainty of the situation. It may be more common to say something like 'what is your credence in a Heads world given the light is red?', ignoring the temporal directedness of your possibility, but this lumping of possibility entangles two situations that are not equivalent.

The dependency structure between the side of the coin and the color of the light changes after the coin is flipped. Because this dependency structure changes, within probability theory, we cannot collapse H_D and H_S into a single possibility called Heads, for this new single possibility would not appropriately reflect our knowledge of the dependencies at each time point. The relationship between the possibility Heads and the light color is temporally dependent. To unify H_D and H_S into a single entity for analysis, at best we could create another event or variable corresponding to if the coin was already flipped or not and construct a multi-dimensional probability distribution with random variables for the coin's side, the color, and flipping status. The point here is that any casual reference to Heads as a possible outcome of the coin in the simple experiment above may conflate two separate possibilities that are not equivalent. We need keep H_D and H_S separate. Again, Unconditional Temporal Continuity allows us to say that $P(H_D)=P(H_S)$, but that principle does not entail that H_D and H_S are the same entity and does not account for acquired dependencies that may develop after the event is decided.

I am not claiming this example is analogous to SB for the purposes of justifying an answer to the SB problem, rather, the analogous portion I care about involves the distinction between dynamic and static possibility, and how that distinction, if ignored, may lead to confusion. The SB experiment, when using **M2**, just does create dynamic and static possibilities for the Tails outcome. On Monday, the possibility for the coin to land Tails is a chancy, dynamic possibility, but on Tuesday, whether the coin landed Tails is a static possibility. If you buy the preceding argument, or at least have become a little suspicious, then you would now be hesitant to talk casually about the credence P(Tails|Monday), for instance. While P(T_D|Monday) can be well-defined as equal to 1/2, P(T_S|Monday) is poorly defined because T_S is only a possibility on Tuesday. Wilson, although aware of the distinction on one level, ignores the ramifications when making this argument for the sleeping beauty problem:

> Using a standard notion for subjective credence: Cr(Heads|Monday) = 1/2, by the Principal Principle. Cr(Heads|Tuesday) = 0, since Beauty is awakened on Tuesday only if the coin lands Tails. So if Cr(Tuesday) > 0, then Cr(Heads) < 1/2. This is already enough to establish the falsity of the Halfer conclusion. (Wilson [2014], p. 586)

But the possibility Heads in the expression Cr(Heads|Monday) is dynamic, while in Cr(Heads|Tuesday) the possibility Heads is static—on Tuesday the coin flip was already decided. We cannot assume these conditional expressions are commeasurable because the dependence structures between the side of the coin and the day of week are not necessarily equivalent. Put another way, Heads dynamic and Heads static are distinct entities and cannot be combined using the total law of probability without further justification. Specifically, Wilson's Halfer falsification argument conflates three notions of Heads when he assumes:

Cr(Heads)=Cr(Heads|Monday)Cr(Monday)+Cr(Heads|Tuesday)Cr(Tuesday)

Heads in the first term Cr(Heads) implies an atemporal concept of Heads. The second credence makes use of the dynamic possibility of Heads, and the third credence is the static possibility of Heads. Harking back to the section on Elga's Thirder argument, we could also note Wilson ignores relevant temporal information about the coin toss in these credence expressions, and including the coin toss events *C* and ¬*C* in these expressions would also invalidate using the total law of probability. More work needs to be done to falsify the Halfer conclusion.

The distinction between dynamic and static possibility is another way to clarify Lewis' claim that SB should believe that P(Heads|Monday)=2/3 when she is awoken on Monday prior to the coin flip. By distinguishing dynamic and static possibility, it is easy to sort out, so long as we also distinguish the two

methods of doing the SB experiment. If the experiment is done according to **M1** (where the coin is flipped Sunday night), then when SB wakes on Monday the coin flip has already been decided—she simply does not know the outcome. The possibility in question is not a dynamic possibility of the future, but a static possibility of the past involving epistemic uncertainty and not chance. Therefore, it is not problematic for SB to hold a credence P(H_S|Monday)=2/3 that differs from 1/2. If the experiment was done according to **M2**, however, and SB wakes on Monday and knows the coin has not yet been flipped, then she ought to believe P(H_D|Monday)=1/2. **M2** requires us to consider both dynamic and static possibilities, and potential dependency changes after the toss that greatly complicate analysis. **M1** is far less problematic given all of the possibility is static, after the toss.

I have argued that mixing dynamic and static possibilities can lead to confusion, while Luna has suggested that mixing centered and uncentered sample spaces is the problem. Although I am sympathetic to Luna's warning about mixing centered and uncentered sample spaces, this mixing is distinct from the problem of conflating dynamic and static possibilities. The following is a direct argument for the Halfer position that mixes centered and uncentered evidence without any clear conflict.

The Mirror Argument. Imagine that you participate in the SB experiment and find yourself within an awakening. You contemplate these possibilities and credences:

H_D= the fair coin will land Heads (dynamic, uncentered, unconditional)
H_S= the fair coin landed on Heads (static, uncentered, unconditional)
A_H = the awakening I am in is a Heads-awakening (static, centered, unconditional)

Then:

(**R1**) P(H_D)=P(H_S)=1/2, by Unconditional Temporal Continuity

(**R2**) P(H_S|A_H)=1, or in words, your credence the coin landed Heads given you are in a Heads-awakening is one.

(**R3**) P(A_H|H_S)=1, your credence that you are in a Heads-awakening given the coin landed Heads is one.

(**R4**) ∴ P(A_H)=P(H_S) which follows from (**R2)**, (**R3)**, and the definition of conditional probability. Further, if you accept (**R1**), then P(A_H)=P(H_S)=1/2 .

The Mirror Argument concludes that your *credence* of being in a Heads-awakening and the coin landing Heads are equivalent. When coupled to (**R1)** it gives P(A_H)=1/2. If you deny (**R1**) but assume

that P(A_H)=1/3 as do Thirders, then the Mirror Argument requires that your uncentered credence in Heads is 1/3 as well. In other words, it requires you to believe that a fair coin has a chance of 1/3 landing Heads! Those who attempt to deflate the Sleeping Beauty Problem by claiming your credences for H_S and A_H can differ (Groisman [2008]; Luna [2020]) must counter this simple argument. Luna attempts to defend differing credences but fails because Luna's argument is grounded in the belief that repetition arguments unequivocally conclude P(A_H)=1/3. I have shown that is not the case.

6. Conclusion

The Thirder position for Sleeping Beauty seems to have more support in the literature, yet several of the most popular arguments for 1/3 are not as sturdy as they appear. The long-run frequency argument, which has almost unequivocally been taken as strongly supporting 1/3, breaks down when we look at repetition with more detail and attempt to make some connection to the single-trial experiment. The statistical question "What is the expected proportion of Heads-Awakenings for a single-trial SB experiment?" has an unambiguous answer of 1/2, just as the similar question for an infinite-trial SB experiment has a clear answer of 1/3. Whether you believe credences should track expected proportions is another matter, but the naïve long-run frequency argument assumes that you should. This work demonstrates how the results of a long-run sequence may not carry over to the single case.

My criticisms of Elga's symmetry argument and the new-information arguments point to subtle misalignments between the narrative description of a thought-experiment in the world and the mathematical probability symbols and theory we use to describe these narratives. Elga's symmetry argument ignored information about the coin toss when translating the narratives into probabilistic expressions, permitting a conclusion that is only possible because a symbol was implied yet left off the page, but we need that missing symbol to correctly encode our relevant information about the experiment. For the new-information arguments, I drew the distinction between updating probabilistic expressions with conditionalization 'on the page' versus the actual causal mechanisms by which this updating must take place 'in the world' and suggest that they must be aligned to be valid. This alignment appears to be implicitly acknowledged by Thirders whose analogous SB experiments demonstrate proper alignment when updating, in contrast to the original SB experiment. Understanding this general alignment suggests an avenue for future research.

Perhaps the most controversial aspect of this work will be my distinction between dynamic and static possibility in probabilistic expressions, but the distinction follows from a widely shared and longstanding metaphysical difference between future and past events. We need this distinction in

probability theory to account for acquired dependences that may arise after an event is decided but the result is unknown. Knowledge that an event has been decided—without knowing the result—is admissible information that can alter a *conditional* credence. We can incorporate that information as a separate conditional variable in our probability expressions or use the language of dynamic and static possibilities. This is another example of delicately aligning our probabilistic symbols with the complete narrative of a thought experiment. The temporal uncertainty of the Sleeping Beauty problem makes this alignment tricky.

REFERENCES

Arntzenius, F. [2003]: 'Reflections on Sleeping Beauty', *Analysis*, **62**, pp. 53–62.

Bostrom, N. [2007]: 'Sleeping Beauty and Self-location: A Hybrid Model', *Synthese*, **157**, pp. 59–78.

Bradley, D. [2003]: 'Sleeping Beauty: A note on Dorr's argument for 1/3', *Analysis*, **63**, pp. 266–8.

Bradley, D. [2011a]: 'Self-location is no Problem for Conditionalization', *Synthese*, **182**, pp. 393–411.

Bradley, D. [2011b]: 'Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty', *British Journal for the Philosophy of Science*, **62**, pp. 323–42.

Carroll, S. and Seben, C. [2018]: 'Self-locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics', *British Journal for the Philosophy of Science*, **69**, pp. 25–74.

Dorr, C. [2002]: 'Sleeping Beauty: In defence of Elga', *Analysis*, **62**, pp. 292–6.

Elga, A. [2000]: 'Self-locating Belief and the Sleeping Beauty Problem', *Analysis*, **60**, pp. 143–7.

Gao, X. [2018]: 'Perspective Reasoning and the Solution to the Sleeping Beauty Problem', [Preprint] *URL: http://philsci-archive.pitt.edu/id/eprint/15355* (accessed 2022-08-01).

Groisman, B. [2008]: 'The End of Sleeping Beauty's Nightmare', *British Journal for the Philosophy of Science*, **59**, pp. 409–16.

Horgan, T. [2004]: 'Sleeping Beauty awakened: New odds at the dawn of the new day', *Analysis*, **64**, pp. 10–21.

Horgan, T. [2007]: ‘Synchronic Bayesian updating and the generalized Sleeping Beauty problem’, *Analysis*, **67**, pp 50-9.

Karlander, K. and Spectre, L. [2010]: ‘Sleeping Beauty meets Monday’, *Synthese* **174**, pp. 397–412.

Lewis, D. [2001]: ‘Sleeping Beauty: Reply to Elga’, *Analysis*, **61**, pp. 171–6.

Lewis, D. K. [1980]: ‘A Subjectivist’s Guide to Objective Chance’, in R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Volume II, Berkeley: University of California Press, pp. 263–93.

Liao, S. [2012]: ‘What are centered worlds?’, *The Philosophical Quarterly*, **62**, pp. 294–316.

Luna, L. [2020]: ‘Sleeping Beauty: Exploring a Neglected Solution’, *British Journal for the Philosophy of Science,* **71**, pp. 1069–92.

Pust, J. [2008]: ‘Horgan on Sleeping Beauty’, *Synthese,* **160**, pp. 97–101.

Pust, J. [2012]: ‘Conditionalization and Essentially Indexical Credence’, *Journal of Philosophy*, **109**, pp. 295–315.

Pust, J. [2014]: ‘Beauty and Generalized Conditionalization: Reply to Horgan and Mahtani’, *Erkenntnis*, **79**, pp. 687-700.

Titelbaum, M. G. [2012]: ‘An Embarrassment for Double-Halfers’, *Thought*, **1**, pp. 146–51.

Weintraub, R. [2004]: ‘Sleeping Beauty: A simple solution’, *Analysis*, **64**, pp. 8–10.

White, R. [2006]: ‘The generalized Sleeping Beauty problem: A challenge for thirders’, *Analysis*, **66**, pp. 114–9.

Wilson, A. [2014]: ‘Everettian Confirmation and Sleeping Beauty’, *British Journal for the Philosophy of Science*, **65**, pp. 573-98.

Winkler, R. [2017]: ‘The Sleeping Beauty Controversy’, *The American Mathematical Monthly*, **124**, pp. 579-87.