Sleeping Beauty: Awakenings, Chance, Secrets, and Video

Nathan Salmon

Abstract A new philosophical analysis is provided of the notorious Sleeping Beauty Problem. It is argued that the correct solution is one-third, but not in the way previous philosophers have typically meant this. A modified version of the Problem demonstrates that neither self-locating information nor amnesia is relevant to the core Problem, which is simply to evaluate the conditional chance of heads given an undated Monday-or-Tuesday awakening. Previous commentators have failed to appreciate the significance of the information that Beauty gains upon waking, and which is relevant to the conditional chance of heads: *de re* acquaintance with the awakening itself and the non-locating knowledge that it is an experimental awakening. David Lewis and company are committed to several unjustifiable and unacceptable probability assessments. Previous commentators have in effect confused the information that Beauty undergoes this particular experimental awakening for the information that she undergoes some experimental awakening or other. Lewis in particular thereby illegitimately tips the scales both in favor of heads and in favor of Monday. The Sleeping Beauty Problem is equivalent to a ball-in-urn word problem in elementary probability theory.

Keywords Sleeping beauty problem · Propositional attitudes · *De re* belief · Epistemology · Probability assessment

N. Salmon (✉)
Department of Philosophy, University of California, Santa Barbara, Santa Barbara, CA, USA
e-mail: nsalmon@ucsb.edu

In unpublished work, (Zuboff, 1990) introduced an ingenious problem now known as the Sleeping Beauty Problem. Twenty years later, (Elga, 2000) published what was to become the classic Problem. A vigorous controversy followed. During roughly the same period, philosophers of language undertook extensive investigations into both the semantics and the pragmatics of ascriptions of propositional attitudes, with special attention to belief of singular propositions and the concomitant notion of *de re* belief. Belief of singular propositions raises thorny difficulties involving the pragmatics of belief attributions. Although both lines of investigation have a focus on the phenomenon of belief, for the most part, the two lines have gone their separate ways. This is unfortunate. Significant developments in the theory of *de re* belief and concerning the semantics and pragmatics of belief attributions have a direct bearing on the solution to the Problem.

The Sleeping Beauty Problem scenario goes as follows: Princess Aurora, the experimental subject and an infallible expert on probability theory, is informed of the setup. She will sleep Sunday night. A fair coin, having a 50/50 chance (even odds) of landing heads, will be tossed before Tuesday. (Here and throughout, the word "chance" is used in the ordinary sense that it customarily has as occurring in phrases like "game of chance," "chances are," "off chance," and "50/50 chance.") Aurora will be briefly awakened on Monday at precisely time t_1 and interviewed. A drug will then be administered that has the effects of erasing her memory of the awakening and of her interview and of putting her back to sleep. If the coin lands heads, she will be allowed to sleep undisturbed through Tuesday, and the experiment will be concluded. If the coin lands tails, she will be briefly awakened on Tuesday at t_2 and interviewed again. She will then be put back to sleep. She will wake on Wednesday and the experiment will be concluded. Knowing all this, upon waking on Monday how does Aurora assess the probability of heads?

Some philosophers, most notably (Lewis, 2001), argue that the answer is one-half.[1] As many others have argued, most notably Elga, the correct solution is in fact one-third. Some, e.g., (Bostrom, 2007) at p. 62, object that an assessment of one-third reflects excessive expectation of tails—especially if the coin is not yet tossed, and especially since it was a foregone conclusion that Aurora would be awakened on Monday regardless. This objection is understandable but wrongheaded.

Whereas many writers have argued that the correct solution is one-third, my position crucially differs from theirs in at least two significant respects. First, the orthodox Sleeping Beauty Problem is formulated in terms of Aurora's "credence" that the coin lands heads, i.e., her "degree of belief" of heads.[2] Previous advocates of

---

[1] See notes 2 and 5 below.

[2] Numerous philosophers distinguish (following Lewis, roughly following Carnap) two kinds of probability: subjective *credence*, or degree of belief; and objective *chance*, which is supposed to be closely related to relative frequency in the long run. Credence is seen as epistemic, chance as physical or otherwise metaphysical, and indeed as *"propensity*, understood as making sense in the single case" (in contrast to relative frequency, which requires a plurality of cases). See (Lewis, 1980), pp. 263–293, at p. 263. Propensity theorists see the chance of an outcome as something like

the one-third solution (so-called *thirders*) judge that upon being awakened, Aurora's degree of belief that the coin lands heads decreases from one-half belief to one-third. In response to the question "To what degree does Aurora believe on Monday that the coin lands heads?", my own answer is less than 50%, less than 33%, less even than 1%. I maintain that *on waking Aurora has no positive belief or confidence whatsoever* that the coin lands heads. Assessing the probability of heads as one-third is a way of *doubting* heads, which is not the same thing as believing heads a little bit. One may argue—and argument is required—that upon being awakened, Aurora comes partially to believe that the coin lands tails. But throughout the entire experiment, Aurora never forms even so much as a partial belief that the coin lands heads. In the ordinary sense of 'believe', at no time does Aurora come to believe heads at all, not even partly.

Second, under Lewis's influence, philosophers who have previously advocated the one-third solution have typically taken for granted that the objective chance that the coin lands heads remains constant throughout. I agree that in a fair coin toss, the default chance of heads is one-half. In repeated fair tosses, the coin will therefore land heads around one-half of the time. Knowing this, on Sunday, Aurora has no belief one way or the other on the question of heads versus tails. Upon the Monday awakening, she comes to doubt heads. On my view, it is not merely Aurora's beliefs that change. What decreases from one-half to one-third is not how much Aurora believes in heads, but the objective chance of heads given all that Aurora knows by then, in the relevant sense of 'chance'. I contend that alleged phenomenon of partial belief is in fact a colossal distraction, driven largely by the excessive influence of Lewis's eccentric theories of probability and belief. *The central issue is the probability of heads—the objective chance of heads, in the ordinary sense—given all that Aurora knows upon being awakened on Monday.* That probability is, surprisingly, one-third. Aurora is certain of it. Her beliefs and doubts—her "credences"—concerning what day it is and the outcome of the coin toss appropriately follow suit.[3]

---

its degree of unpreventability. They maintain that before a fair coin is tossed, the chance that it will land heads is one-half, quite irrespective of what anyone knows or believes. The notion of conditional chance is thereby rendered problematic, since a phenomenon's propensity is typically unaffected by one's information concerning it. Propensity theorists do best (albeit not well) to understand Chance($p \mid q$) as p's dispositional propensity were q's propensity to be 100%, i.e., what p's propensity would become if q were presently unpreventable.

It is far from clear that belief is a matter of degree. I have argued that degrees of confidence are not epistemic but doxastic; whereas chance, although objective, is epistemic. Cf. (Salmon, 2019), pp. 644–662. See especially pp. 644n1, 659. The epistemic nature of chance, once it is recognized, ultimately renders metaphysical propensity an idle conceptual wheel.

[3] As mentioned, propensity theorists about chance use the word 'chance'. for an alleged physical or otherwise metaphysical and non-epistemic property of external phenomena, including coin tosses and the like. However, they also use the word in the ordinary sense familiar from coin tosses, dice, dreidels, and other so-called games of chance. In effect, propensity theorists use the word ambiguously (although they presumably deny doing so). The propensity theorist's assessment that the chance of heads is one-half both before and upon the Monday awakening conflicts with the real chances, in the ordinary sense used here. See notes 2 and 9. Arguably, Aurora's degree of belief of heads should be negative—exactly −33⅓%, according to the rule that $Belief_x(p) = 2C_x(p)-1$,

These two disagreements are not entirely verbal, but to the extent that they are, I do not think that the other side can claim to have the support of common usage. Many will regard my departures from previous philosophers' investigations as excessively radical. Radical though the differences may be, I regard my claims as little more than philosophical common sense, both about probability and about belief. Indeed, to insist that in the ordinary sense of 'believe' (or of *'croire'*, *'glauben'*, *'credere'*, etc.) Aurora partially believes that the coin lands heads would be to betray a serious misunderstanding. It is not to my purpose to engage with the sizeable literature on the Sleeping Beauty Problem. Rather, my principal objective is to provide a partly novel defense of the one-third solution to an unorthodox conceptualization of the Problem, one that I take to stick closer to the facts from the perspective of the philosophy of cognition. Where others speak of one's credence that the coin lands heads, I shall speak instead of one's assessment of the probability of heads. Accordingly, I adopt a temporary policy of replacing the formal epistemologist's jargon "a's credence toward p is n," where $0 \leq n \leq 1$ (e.g., 1/2, or 1/3), with the non-philosopher's vernacular "a assesses the probability/chance of p to be n." (Readers who strongly resist my views—evidently the vast majority of those who have thought about the problem—are invited reciprocally to replace my wording here with their preferred phraseology, although the two are decidedly not synonymous.)[4]

Here, we supplement the classic Problem setup with the following: A video recording is taken of Aurora's Monday awakening. If Aurora is awakened again on Tuesday, her Tuesday awakening will be streamed live on a video monitor. The video recording or live-stream video would all look and sound exactly the same irrespective of the coin toss. At some time t_0 on Sunday, Prince Phillip, another infallible expert on probability theory, is informed of the experimental setup. At t_1 on Monday, Phillip correctly assesses the probabilities of various propositions about the experiment, taking into account all that he knows. At t_2 on Tuesday, the video recording of Aurora's Monday awakening is shown to Phillip, but he is informed only that it is either a live stream in real time of an awakening happening as he watches, or else a recording of the Monday awakening. (Phillip understands that even if Aurora is presently waking at t_2, the video could nevertheless be a recording of yesterday's awakening.) Phillip also instantaneously correctly assesses the probabilities of various propositions, taking into account all he then knows. (Phillip is not administered any drug. He is not awakened from sleep. He is not shown any other video of Aurora waking.)

Let h be the proposition that the coin lands heads. Tails is then effectively equivalent to ~h. Since the coin is fair, prior to the experimental setup $\Pr(h) = \Pr$

---

where C_x is x's probability-assessment function (or what may be tantamount to the same thing, x's subjective confidence expressed as a probability).

[4]Whereas the orthodox problem concerns Aurora's degree of belief in heads, some writers have presented the problem instead in terms of Aurora's assessment of the probability of heads. See for example the (The Oscar Seminar, 2008), pp. 149–155, at p. 151.

$(\sim h) = 1/2$. Let "a" be an individual constant (name) for the witnessed experimental awakening, which is the occurrent, token event of Aurora's awakening at t_1. The constant "a" should be thought of on the model of the demonstrative phrase "this awakening". (It could instead be a name whose reference is fixed by the definite description "the awakening I am presently witnessing". However, it cannot be an abbreviation for "the Monday experimental awakening".) Let 'F' be a predicate for the property of being an experimental awakening (a Monday or a Tuesday awakening, one or the other). Let us define $1F(\alpha) =_{df} \forall y(Fy \leftrightarrow y = \alpha)$, i.e., α and nothing else is (will have been) F; and $2F(\alpha) =_{df} \exists z[z \neq \alpha \& \forall y(Fy \leftrightarrow y = z \lor y = \alpha)]$, i.e., α is one of exactly 2F's. The experimental setup includes that $[h \leftrightarrow \exists x 1F(x)] \& [\sim h \leftrightarrow \exists x 2F(x)]$.

Expressed in this notation, Phillip makes the following assessments:

Assessment	Time	Proposition	Probability
1	t_0	h	1/2
2	t_0	$\sim h$	1/2

Assessment	Time	Proposition	Probability
1′	t_1	h	1/2
2′	t_1	$\sim h$	1/2
3′	t_1	$h \leftrightarrow \exists x 1F(x)$	1
4′	t_1	$\sim h \leftrightarrow \exists x 2F(x)$	1

Therefore,

Assessment	Time	Proposition	Probability
5′	t_1	$\exists x 1F(x)$	1/2
6′	t_1	$\exists x 2F(x)$	1/2

Assessment	Time	Proposition	Probability
3″	t_2	$h \leftrightarrow 1F(a)$	1
4″	t_2	$\sim h \leftrightarrow 2F(a)$	1
5″	t_2	$1F(a)$	1/3
6″	t_2	$2F(a)$	2/3

Therefore,

Assessment	Time	Proposition	Probability
1″	t_2	h	1/3
2″	t_2	$\sim h$	2/3

Although controversial, assessments 5″and 6″ are conclusively established *a priori* through a simple thought experiment establishing relative frequency in the long run in hypothetical repeated trials, what is called an "hypothetical frequency."

To illustrate, suppose the same procedure is performed 1000 times during the same four-day period. Then on average, in 500 of those trials, give or take, the coin lands heads and there are around 500 experimental awakenings, all on Monday. In the remaining trials, where the coin lands tails, there are around 500 experimental awakenings on Monday and 500 more on Tuesday. Roughly 500 awakenings among the trials are the only awakening in their trial, compared with 1000 that are each paired with another awakening. The *de re* hypothesis concerning each experimental awakening that it was the only one in its trial is correct of about 500, incorrect of about 1000, approaching a relative frequency of 500/1500. The hypothetical frequency of sole experimental awakenings is thus one-third. It follows that, given an undated experimental awakening, as Phillip is, the objective chance that it is Aurora's only experimental awakening is only one-third. Therefore, given an undated experimental awakening, the objective chance that it is a heads awakening is one-third. Probability expert Phillip runs this argument to arrive at assessments 5″ and 6″.

In stipulating only one awakening in case of heads and two awakenings in case of tails, the experimental setup biases the coin toss, halving the conditional chance of heads *vis-à-vis* tails given an undated experimental awakening. Given an undated experimental awakening, it is twice as likely to be one of a pair of experiment awakenings as to be a sole experimental awakening.

To repeat: This result is obtained *a priori*. No actual trials have been performed, or need be (if they can be). Indeed, even if by some fluke repeated trials yielded a different outcome, it would make no difference to the results obtained by thought experiment. The hypothetical long-run relative frequency is not merely one consideration favoring one-third as the answer. *Rather, the thought experiment constitutes a proof that the real probability of heads given Aurora's state of knowledge and ignorance is one-third.*[5]

---

[5] Recall that I am taking the Problem to concern the conditional probability of heads given Aurora's information upon waking, not Aurora's "credence" in heads. On my view, Aurora has no positive degree of belief in heads on either Sunday or Monday (and therefore she does not have one-third of full belief, whatever that is).

See note 3. The members of the Oscar Seminar (*op. cit.*) draw a distinction between two alleged kinds of "objectivist" (non-Bayesian) probabilities—"definite" and "indefinite"—and they take indefinite probability to be the more basic notion epistemically. They postulate a defeasible inference rule ("direct inference") that sanctions identifying the definite probability that F*a* with the indefinite probability that F*x*, for an unspecified individual *x*, given that *x* satisfies a restrictive condition, that G*x* (the "reference" condition), which *a* is known to satisfy. I do not endorse the conceptual apparatus of the Oscar Seminar and their critics. On my view, so-called indefinite probability is basically a species of conditional *de re* probability. The indefinite probability that F*x* given that G*x* might even be defined in terms of so-called definite probability: the number *n* (if such exists) such that for all *x*, *n* = the conditional definite probability that F*x* given no *a posteriori* information involving *x* save that G*x* (and the *a priori* consequences thereof).

More significantly, I conceive of the Problem and its proper solution very differently from the Oscar Seminar and their critics. The Seminar's proposed solution has Aurora defeasibly infer the definite probability that the relevant coin lands heads from the indefinite probability of a coin landing heads given a particular complex restrictive reference condition on coin tosses. I conceive

A cautionary word: One must not confuse (as a number of theorists have) the above thought experiment, which is attributed to Phillip and is carefully devised so as to address the Problem, for an alternative hypothetical-frequency scenario, which is neither. Phillip's thought experiment must be sharply distinguished (as Phillip does) from an immaterial and irrelevant one that instead invokes 1000 probability experts playing the role of Phillip. In about one-half of the trials of this alternative scenario, the subject playing Phillip's role will witness at t_2 the only experimental awakening, and in the remaining trials, the subject will witness one of two awakenings. This alternative relative-frequency scenario addresses the following question: Given the experimental setup, what is the antecedent chance that the event Phillip will witness at t_2 is the only experimental awakening rather than one of two? The answer: 50–50. *But that is not the issue at hand*. The Problem concerns instead the subject's considered response, given all that she/he knows about the situation, to a considerably more difficult question:

> What is the *de re* probability about *this* [the currently witnessed awakening], that *it* is the only experimental awakening rather than one of two?

Confusion between these two questions has likely contributed to the on-going controversy. The reasoning that Phillip employs to arrive at his t_2 assessments does not invoke hypothetical trials involving himself (or his counterparts) and his observations. It instead invokes, and properly so, trials involving the awakenings themselves. It establishes the correctness of Phillip's t_2 assessments of particular *de re* probabilities about the awakening he is witnessing at that moment, not antecedent probabilities about his witnessing of it.

Phillip's assessment 1′ does not make use of his knowledge that Aurora was to be awakened at t_1 on Monday. He knows at t_1 that there is an awakening, but only by

---

of the Problem more simply, as that of determining the conditional *de re* probability about the witnessed awakening that it is a sole experimental awakening (the only alternative being that it is one of a pair), given that it is an experimental awakening. On the solution proposed here, Aurora does not defeasibly infer the definite probability of heads from an indefinite probability given some complex (and hence potentially problematic) restrictive reference condition on coin tosses. Instead she calculates the (definite) probability of heads given her newly acquired information that a is an experimental awakening. She performs this calculation knowing that this value is also the *de re* probability, for any value of 'x', that x is a sole experimental awakening *given only that x satisfies the simple reference condition of being an experimental awakening*.

Elga (*op. cit.*) and others argue for the one-third solution on significantly more complex grounds. See note 10 below. Cf. (Rosenthal, 2009), pp. 32–37. In his original presentation, Elga pointed out in passing (*op. cit.*, p. 143) that the long-run relative frequency of heads in repeated trials is one-third, but he presented this as merely one consideration among others. On the contrary, that ground alone constitutes decisive proof, which Aurora can carry out, concerning the actual probability of heads. (Titelbaum 2013), at p. 1004, argues that the relative-frequency thought experiment does not settle the issue. His counter consideration has Aurora disregard the relevant information that she gains upon being awakened. However, the Problem specifically concerns the probability of heads *given all that Aurora knows upon waking*. (I here ignore the exceedingly implausible prospect that Aurora's assessment of the probability of heads deviates from the conditional probability given all she knows.)

knowing that there is a Monday awakening irrespective of the coin toss. The chance of heads given his total information at t_1 remains one-half. To update his assessment, Phillip must await further information, specifically knowledge of the occurrence of a particular awakening but with no further information concerning whether it is the Monday awakening or a Tuesday awakening, or whether it is the only experimental awakening or one of two. The chance of heads is correctly assessed by Phillip at t_1 as one-half and at t_2 instead as one-third. The revision is based merely on the *a posteriori* information that Phillip gains through watching the video, i.e., through his witnessing that awakening, together with his antecedent knowledge of the experimental setup and his *a priori* reasoning from 3″-6″.

Aurora's epistemic situation upon being awakened at t_1 is the same as Phillip's epistemic situation at t_2 in all respects relevant to the Problem. Aurora's t_1 assessments are therefore all the same as Phillip's t_2 assessments. In particular, Aurora is as able as anyone is (and far more inclined than non-thirders are) to conduct the same thought experiment that Phillip conducts at t_2. Consequently, upon waking at t_1 Aurora runs through the same relative-frequency proof, establishing assessment 1″. On the other hand, the differences between Aurora's and Phillip's situations filter out some irrelevancies. Part of the original Problem's charm is that it assigns to a single person roles I have divided and delegated: Aurora's role on Monday as beautiful awakened experimental subject; and Phillip's role on Tuesday as expert probability assessor in the relevant epistemic situation. One significant point that emerges is that amnesia is irrelevant to the core Problem and its solution. At t_2, unlike Aurora, Phillip remembers the day before. So-called self-locating (or token-reflexive or *de se* or "centered") information is likewise inessential. Unlike Aurora, at t_2 Phillip knows that it is Tuesday. Also unlike Aurora, he cannot truthfully say "I am being awakened now." What matters is the assessor's knowledge and ignorance concerning the witnessed awakening, irrespective of whether it be from a first-person or from a third-person perspective. Like Aurora, Phillip knows that it is a Monday or Tuesday awakening, but not whether it is a Monday awakening or a Tuesday awakening, or whether it is the only experimental awakening or one of two. This is critical to their epistemic situations being relevantly the same.[6]

Jeffrey Rosenthal says, "Mathematically speaking, it seems that we are being asked to compute the conditional probability that the [coin] showed heads, conditional on the fact that Beauty is currently being interviewed" (*op. cit.*, first page). This description evidently conflicts with Phillip's assessment 1′. Rather, the issue is the probability of heads on the condition that the undated witnessed event is an experimental awakening—with no information that it is a Monday awakening, that it is a Tuesday awakening, that it is the only experimental awakening, or that it is one of two. In short, the Sleeping Beauty Problem is to evaluate $\Pr(h \mid Fa)$.

---

[6] As David Braun points out, it is arguable that upon waking Aurora already knows that it is Monday simply by knowing what she expresses by "It is today." This sort of issue should be bypassed in discussion of the Problem by taking "Monday" and "Tuesday" to be abbreviations for "the second day of the experiment" and "the third day of the experiment," respectively.

On Sunday, Aurora correctly assesses the objective chance of heads as one-half. She also correctly predicts that when she awakens on Monday, the chance, given all she will then know, dips to one-third. How can it be that the chance of heads decreases just because Aurora is awakened on Monday—especially if the coin is not yet tossed, and especially since it was a foregone conclusion that she would be awakened on Monday regardless?

Key to the Sleeping Beauty controversy is the fact that chance (in the ordinary sense used here), although objective, varies with one's information. The chance of a prospect is an objective epistemic ratio, in effect a numeric comparison of the epistemically (in contrast to metaphysically) possible scenarios that include the prospect in question to the antecedently understood full range of epistemically possible scenarios. In the most straightforward type of case—a finite equiprobable space representing a fair coin, a pair of fair dice, or the like—the chance of p is the classical Laplacian ratio of the number of epistemically "equipossible" (equally weighted), mutually exclusive, scenarios in which p obtains to the number of the jointly exhaustive totality of relevant, epistemically equipossible, mutually exclusive scenarios. That the probability of heads given, or on the condition, that a is an experimental awakening is one-third is not to be confused with the conditional claim that $Pr(h) = 1/3$ if Fa. In Phillip's epistemic situation at t_1, that conditional is false (unbeknownst to him) yet $Pr(h \mid Fa) = 1/3$. The mere fact that Fa, of itself, in no way alters the fact that $Pr(h) = 1/2$. The influence of the information that Fa on the objective chance of heads is not causal. It is a non-causal, numeric relationship, but that relationship is very real and the information thereby affects the Laplacian ratio for heads. On the condition that this witnessed event is an experimental awakening—given the undated experimental awakening *as additional information*—the chance of heads decreases from one-half to one-third.

Theorists who favor the answer of one-half, as well as some who favor the correct answer, have failed to appreciate the nature of the information Aurora gains upon waking. Elga says that Aurora's revised probability assessment at t_1 is not a result of her receiving new information, since she was already certain she would be awakened on Monday (*op. cit.*, p. 145). Lewis says that when Aurora is awakened on Monday "the only evidence she gains is ... that she is presently undergoing either the Monday awakening or the Tuesday awakening" (*op. cit.*, p. 173), but he adds that this new evidence is irrelevant to h *vs.* $\sim h$ (p. 174, which contains a misprint). Lewis also says that "if she is awakened on Tuesday the memory erasure on Monday will make sure that her total evidence at the Tuesday awakening is exactly the same as at the Monday awakening" (p. 171). On the contrary, *Aurora gains crucial information on Monday that she lacks on Sunday*, and also lacks on Tuesday if she is awakened then.

Others argue that the knowledge that Aurora gains upon being awakened that justifies her updated assessment of the probability of heads as one-third is of "locating" information (or conditionally locating information), expressible only by utilizing an indexical expression like "I," "today," or "now" (or by using a name whose reference is fixed ultimately by means of an indexical). See (Horgan, 2004) and (Weintraub, 2004), respectively, at pp. 19–20 and p. 9. See also (Moss, 2012);

(Stalnaker, 2008), at pp. 65 and 78, and (Stalnaker, 2011; and (Titelbaum, 2013a, 2013b), Sect. 10.2, pp. 238–243.[7] Cf. (Salmon, 2019), p. 658. I agree that in some cases the probability of a proposition is relative to a proposition guise, a way of taking that proposition. But Phillip's epistemic situation shows that Aurora gains information that justifies downgrading her assessment of the probability of heads and that is not "essentially indexical" or token-reflexive, or even locating information in any straightforward way.

Both Aurora and Phillip know at t_1 that Aurora is being awakened. But Aurora is already able to update the prior probability assessment of heads to one-third, whereas Phillip cannot yet do so. This is precisely because Aurora gains additional information at t_1 that Phillip does not have until t_2: the information that Fa. Contrary to both Elga and Lewis, Aurora gains the information not merely that she is undergoing some experimental awakening or other, but furthermore that she is undergoing *this particular* experimental awakening. At different times, she and Phillip gain knowledge by acquaintance of a itself—empirical knowledge that alters the probabilities. As with Phillip's t_2 assessments, Aurora's t_1 assessments are based on the *a posteriori* acquaintance with a itself, together with her antecedent knowledge and a priori reasoning.[8]

At t_1, Aurora knows only that it is Monday or Tuesday; she does not know which. For all Aurora knows—and for all Phillip knows at t_2—the coin landed tails and a is a Tuesday awakening. The information that a is the Monday awakening is very relevant to h vs. ~h. Given it, the chance of heads is restored to one-half. The misinformation that a is a Tuesday awakening is also very relevant. Given it, the chance of heads collapses to zero. The information that Fa, though meager, is also relevant. In the experimental setup, two mutually exclusive epistemic possibilities are exhaustive and equiprobable:

p_1 : h; $\exists \times$ [1F(x) & x occurs on Monday]
or

p_2: ~h; $\exists x$[2F(x) & x occurs on Monday & $\exists y$(Fy & $x \neq y$) occurs on Tuesday].

At t_1, Aurora learns that Fa. Given that Fa, p_1 becomes an epistemic possibility concerning a. Because there are two F's in it, p_2 instead bifurcates into two mutually exclusive equiprobable epistemic possibilities concerning a. Given the new information, three mutually exclusive epistemic possibilities are exhaustive:

p_1' : h; 1F(a); a occurs on Monday
or

p_2': $\sim h$; 2F(a); a occurs on Monday; $\imath y$(Fy & $a \neq y$) occurs on Tuesday

or else

p_2'': $\sim h$; 2F(a); a occurs on Tuesday; $\imath y$(Fy & $a \neq y$) occurs on Monday.

What are their separate probabilities? Clearly, Pr(p_2' | Fa) = Pr(p_2'' | Fa). Those who hold that Pr(h | Fa) = 1/2 are compelled to judge that Pr(p_1' | Fa) = 1/2, whereas Pr(p_2' | Fa) = 1/4. These assessments in turn lead to further, more immediately unacceptable assessments. Having declared that upon waking on Monday, Aurora learns nothing relevant to whether the coin lands heads, Lewis reasons that if she is now informed it is Monday ($p_1' \vee p_2'$), she will revise her assessment of heads up from one-half to two-thirds, twice the probability of tails.[9] The reasoning is insufficiently attentive to the epistemic situation. The information that it is Monday rather than Tuesday does increase the likelihood of heads. But there is a Monday experimental awakening regardless of how the coin lands. Given that a in particular is a Monday awakening, the odds that it is a heads awakening are even; the long-run relative frequency of heads among Monday awakenings in repeated trials is one-half. Informed that it is Monday, Aurora will indeed raise her heads assessment—not from one-half but, correctly, *to* one-half. Closely related, Lewis and company are committed to Aurora judging it twice as probable upon waking that it is Monday and the coin lands heads as that it is Monday and the coin lands tails. This tipping of the scales in favor of heads is likewise quite unjustifiable. Even more dramatically, Lewis, et al. are committed to Aurora assessing the probability upon being awakened that it is Monday as fully 75%, *three times as probable as that it is Tuesday*. Such a lopsided favoring of Monday over Tuesday is completely unjustifiable. The probability upon waking that it is Monday is intuitively, and in truth, simply the long-run relative frequency of Monday awakenings: two-thirds. The erroneous assessments all result from attributing too much probability upon waking to heads, not enough to tails. The facts dictate that Pr(p_1' | Fa) = Pr(p_2' | Fa). It follows that p_1', p_2', and p_2'' are all three equiprobable, and therefore that Pr(p_1'| Fa) = Pr(h | Fa) = 1/3.[10]

The objective chance of heads given all that Aurora knows at t_1, and the hypothetical relative frequency of heads awakenings in repeated trials in the long run, are the same: one-third. Lewis says ("Subjectivist's Guide," p. 263) that chance

---

[9]*Op.* cit., pp. 174–175. Lewis's position is that upon being informed it is Monday, Aurora believes the chance of heads is one-half yet her "credence" increases to two-thirds. See notes 2 and 3 above. He purports to reconcile the conflict by claiming that on being informed it is Monday, Aurora gains "inadmissible" information about the future—that now is not in it—and the knowledge that it will never be now justifies two-thirds confidence of heads despite allegedly knowing that the chance of heads is only one-half. Lewis says he finds this "fairly convincing, independently of wishing to follow where my argument leads." I find Lewis's rationale entirely devoid of plausibility and his position incoherent.

[10]This proof, though arrived at independently, is related to Elga's primary argument for one-third, *op.cit.*, at pp. 144–145. See note 5 above.

plays much the same role in his thought that long-run relative frequency played in Carnap's thought. Despite this, Lewis's interconnected, idiosyncratic (albeit influential) views concerning belief, chance, credence, evidence, knowledge, probability, propositions, and more dictate that Aurora's judgment on Monday is one-half. The correct solution to the Sleeping Beauty Problem constitutes a disproof of a broad segment of Lewis's philosophical system. (See notes 2 and 9.)

In effect, those who answer one-half confuse the *de re* proposition that Fa for the general proposition that $\exists x$Fx. The latter is a trivial logical consequence of the setup proposition that $[h \leftrightarrow \exists x1\text{F}(x)]$ & $[\sim h \leftrightarrow \exists x2\text{F}(x)]$. Since $\text{Pr}(\exists x\text{F}x) = 1$, $\text{Pr}(h \mid \exists x\text{F}x) = \text{Pr}(h) = 1/2$. But the problem concerns $\text{Pr}(h \mid \text{F}a)$, which is not the same as $\text{Pr}(h \mid \exists x\text{F}x)$. That Aurora gains more information on Monday than merely that she is undergoing some experimental awakening or other should come as no surprise to those who embrace the apparatus of direct reference, Twin Earth, and singular propositions. That eye-opening apparatus beautifully interrupts dogmatic slumber and removes much of the maleficent mystery.

The Problem is not as much a philosophical problem as it is a vexed word problem in elementary probability theory.[11] It is equivalent to the following variant: On Monday, a ball is placed in an empty urn. A fair coin is tossed. If the coin lands heads, no other balls are placed in the urn. If it lands tails, on Tuesday, a second ball is placed in the urn. On Wednesday, a ball is drawn from the urn, at random if there are two. (In fact, the original Monday ball is drawn, but that is both unknown and irrelevant.) Absent knowledge of the result of the coin toss, what is the probability that the drawn ball was the only ball in the urn? We reinterpret our earlier notation as follows: Let *'a'* denote the drawn ball; let 'F' be a term for being a ball in the urn on Tuesday. The predicates '1F' and '2F' are defined in terms of 'F' exactly as before. The new setup includes the conjunction '$[h \leftrightarrow \exists x1\text{F}(x)]$ & $[\sim h \leftrightarrow \exists x2\text{F}(x)]$' with 'F' reinterpreted. So interpreted, what is $\text{Pr}(h \mid \text{F}a)$?

See again the *cautionary word* above. One significant benefit of the balls-in-urn rendering of the Problem is that it sets the irrelevant off from the relevant. The balls represent the Sleeping Beauty awakenings; the drawing of a ball represents a witnessing of an awakening. On Wednesday, before any ball drawing, it is equally likely that there is only one ball in the urn as that there are two. Unquestionably, $\text{Pr}(h \mid \exists x\text{F}x) = 1/2$. It is given that the probability is 50% that the Wednesday witnessing is of an unaccompanied ball. But the relevant question is *de re*, specifically concerning the drawn ball itself, not the drawing of it. There are twice as many balls in the urn in case of tails as there are in case of heads. Regarding the drawn ball, it is equally likely: that it was all alone in the urn; that it was the Monday ball and had a companion; and that it was the Tuesday ball and had a companion. It is therefore half as likely, given and concerning the drawn ball, that *it* was alone in the urn as that

---

[11] Cf. Rosenthal, *op. cit.*

it had company. The probability that the drawn ball was alone in the urn is precisely one-third—the analog of Phillip's assessment 5″. It follows that $\Pr(h \mid Fa) = 1/3$.[12]

References

Bostrom, N. (2007). Sleeping beauty and self-location: A hybrid model. *Synthese, 157*(1), 59–78. https://doi.org/10.1007/sl1229-006-9010-7

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis, 60*(2), 143–147. https://doi.org/10.1093/analys/60.2.143

Horgan, T. (2004). Sleeping beauty: New odds at the Dawn of the new day. *Analysis, 64*(1), 10–21.

Lewis, D. (1980). A Subjectivist's guide to objective chance. In R. Jeffrey (Ed.), *Studies in inductive logic and probability* (Vol. II, pp. 263–293). University of California Press.

Lewis, D. (2001). Sleeping beauty: Reply to Elga. *Analysis, 61*(3), 171–176. https://doi.org/10.1093/analys/61.3.171

Moss, S. (2012). Updating as Communication. *Philosophy and Phenomenological Research, 85*(2), 225–248.

Rosenthal, J. S. (2009). A mathematical analysis of the sleeping beauty problem. *The Mathematical Intelligencer, 31*(3), 32–37., *CiteSeerX* 10.1.1.151.2326. https://doi.org/10.1007/s00283-009-9060-z

Salmon, N. (2019). Impossible odds. *Philosophy and Phenomenological Research, 99*(3), 644–662. https://doi.org/10.1111/phpr.12517

Stalnaker, R. (2008). *Our knowledge of the internal world*. Oxford University Press.

Stalnaker, R. (2011). Response to Weatherson. *Philosophical Studies, 155*, 471–475. https://doi.org/10.1007/s11098-010-9614-0

The Oscar Seminar. (2008). An objectivist argument for Thirdism. *Analysis, 68*(2), 149–155.

Titelbaum, M. G. (2013b). *Quitting certainties: A Bayesian framework modeling degrees of belief.* Oxford University Press.

Titelbaum, M. G. (2013a). Ten reasons to care about the sleeping beauty problem. *Philosophy Compass, 8*(11), 1003–1017. https://doi.org/10.1111/phc3.12080

Weintraub, R. (2004). Sleeping beauty: A simple solution. *Analysis, 64*(1), 8–10.

Zuboff, A. (1990). One self: the logic of experience. *Inquiry: an interdisciplinary Journal of Philosophy, 33*(1), 39–68. https://doi.org/10.1080/00201749008602210

---

[12] As with the relative-frequency thought experiment involving Aurora, a repeated-trial thought experiment involving balls in urns needs to be correctly interpreted, and must not be confused with a thought experiment that addresses the wrong question. The problem question is addressed by taking the ratio of balls placed alone in the urn to the totality of balls placed. In 1000 trials there will be approximately a total of 1500 ball placings. The final tally of unaccompanied balls will be roughly 500. This confirms that the probability is one-third concerning the drawn ball that it is unaccompanied.