Eliciting Sleeping Beauty’s Credences*

Jessi Cisewski, Joseph B. Kadane, Mark J. Schervish, Teddy Seidenfeld, and Rafael Stern

Carnegie Mellon University

*Acknowledgment: We thanks the participants of the UC-Irvine Philosophy Department colloquium series, and the participants of the seminar on the Philosophy of Probability at Columbia University for their helpful suggestions in preparing this paper.
Abstract

We review the debate between Thirderers and Halfers regarding the Sleeping Beauty problem. We use de Finetti’s [1974] theory of coherent previsions to highlight two themes that are central to the debate:

Theme 1: Depending upon the details of the betting game, *credences* and *fair prices* need not be the same.

Theme 2: When using the Law of Total Probability to reason from a set of conditional probabilities given by an exhaustive set of cases, it is essential, also, that the cases are pair-wise exclusive.

We show that in the canonical version of the Sleeping Beauty problem, these themes underscore the Halfers’ solution.
1. Introduction and outline.

The Sleeping Beauty problem is an unusual decision problem with several exceptional features that, as a matter of history, have led to considerable controversy over its solution. In order to analyze the controversial aspects of this puzzle, first we review the canonical Sleeping Beauty problem and its principal, rival solutions.

On Sunday, Sleeping Beauty learns that she will participate in the following Experiment. A fair coin is to be flipped on Monday night. She will be awakened on Monday morning when she will be asked these two questions:

What is your degree of belief that the coin flip Monday night lands Heads?

What are your betting odds on the same event?

Provided the coin flip lands Tails, and only if it land Tails, she will be awaked a second time, on Tuesday morning, and asked again the same two questions, in the same way as on Monday. Thus, the event of the second awakening on Tuesday is correlated with the coin flip landing Tails Monday night: the one occurs if and only if the other does.

Each time that she is asked for her betting odds that the coin flip Monday lands Heads, Sleeping Beauty is required to make a fresh betting contract. However, when she retires for the night on Monday Sleeping Beauty loses all her memories of that day. Thus, if she is awakened on Tuesday (because the coin flip lands Tails Monday), she will have totally forgotten Monday’s events. So, whenever she is awake during the Experiment, on Monday morning and (if the coin lands Tails) also on Tuesday morning, she cannot distinguish one day from the other. Sleeping Beauty does not know if she is wagering on Heads with the
first or with the second contract. However, she is required to make a fresh contract whenever she is asked to bet. Last, she anticipates all this happening from her perspective on Sunday, and remembers the rules of the Experiment whenever she is awake during the Experiment.

What are Sleeping Beauty’s answers to the two questions posed to her during the Experiment? This problem was first introduced by Piccione and Rubinstein (1997, Example 5), a variant of their Absentminded Driver paradox. Elga (2000) used the name Sleeping Beauty when discussing the same problem. A rather large literature has grown up around it. Already, in Spring 1999, the newsgroup rec.puzzles reported several thousand threads discussing the Absentminded Driver paradox, Wedd (2006).

The problem is a puzzle and the literature is large because there is a continuing controversy how Sleeping Beauty should answer the two questions posed when awake during the Experiment. Next, we summarize two conclusions that we find dominate the controversy: Thirders versus Halfers.

Some Thirders argue that, when awake during the Experiment, Sleeping Beauty’s fair betting odds ought to be 1/3 for Heads and this value also reveals her credence, or degree of belief in the proposition that the coin flip Monday night lands Heads. We label this the Thirders’ argument #1 – reasoning about credences from fair betting odds. That argument is as follows:
If we consider a large number $n$ of probabilistically independent repetitions of the Experiment, with probabilistically independent flips of the same fair coin, on about half of the trials, on about $n/2$ trials, the fair coin lands Heads and on about $n/2$ trials it lands Tails. When the coin lands Tails she is asked separately on both Monday and on Tuesday to contract a bet on Heads. So, on about $n/3$ of all the occasions when Sleeping Beauty is awake and asked to bet with a fresh contract, the outcome is Heads. So, her fair betting odds on Heads ought to be 1:2, i.e., Sleeping Beauty should give a fair betting rate of 1/3 on Heads whenever asked during the Experiment. If these fair odds also elicit her credences about the coin flip, as is typical with ordinary cases of fair betting, then when awake during the Experiment her degree of belief that the coin lands Heads also should be 1/3.

Some Thirders argue differently, using a perspective of “centered” possibilities to argue directly about her rational credence in Heads. See, e.g., Elga (2000). These Thirders’ conclude that, while awake during the Experiment, Sleeping Beauty’s credence should be 1/3 for Heads. This analysis addresses solely the second of the two questions posed to Sleeping Beauty each day, namely, “What is your degree of belief the coin lands Heads?” The second Thirders’ argument does not address the question “How will you bet on Heads?”

We label the following reasoning the Thirders’ argument #2.

During the Experiment, while awake Sleeping Beauty recognizes these three “centered” possibilities as exhaustive:

A  It is now Monday and the fair coin will land Heads.
It is now Monday and the fair coin will land Tails.

It is now Tuesday and the fair coin landed Tails.

Let E be the event that Sleeping Beauty is awake according to the rules during the Experiment. Let $P_E(\cdot | \cdot)$ denote her rational conditional credence function while awake during the Experiment. Then, her conditional probabilities should satisfy these two conditions:

(i) $P_E(\text{Heads} | \text{It is now Monday}) = P_E(\text{A} | \text{A or B}) = 1/2$

(ii) $P_E(\text{It is now Monday} | \text{Tails}) = P_E(\text{B} | \text{B or C}) = 1/2$.

Assume that during the Experiment, whenever she is awake, the pair, \{\text{It is now Monday, It is now Tuesday}\} partitions her space of “centered” possibilities. This assumption requires that Sleeping Beauty’s space of “centered” possibilities uses “Now is Monday” and “Now is Tuesday” as both jointly exhaustive and mutually exclusive events.

With this assumption, by the Law of Total Probability:

$$P_E(\text{Heads}) = P_E(\text{Heads} | \text{Monday})P_E(\text{Monday}) + P_E(\text{Heads} | \text{Tuesday})P_E(\text{Tuesday})$$

$$= (1/2)(2/3) + 0(1/3) = 1/3.$$\(^1\)

We note that if in the Thirders’ argument #2, premise (ii), Elga’s premise from (2000), is generalized so that $P_E(\text{It is now Monday} | \text{Tails}) = x$ with $x < 1$, then the Law of Total Probability as applied in this argument, yields the conclusion

$$P_E(\text{Heads}) = (1/2)(2x/[x+1]) = x/[x+1] < 1/2.$$

Thus, the Thirders’ argument #2 stands in opposition to the Halfers’ conclusion that, while awake during the Experiment, Sleeping Beauty’s credence for Heads is 1/2.
Contrary to the Thirders’ position, a basic Halfer’s argument about Sleeping Beauty’s rational degrees of belief is this:

On Sunday, Sleeping Beauty’s credence is 1/2 that the coin lands Heads on Monday.

Let $P_S$ denote her rational credence from Sunday’s perspective. Then

$$P_S(\text{Heads}) = 1/2.$$ 

With $E$ the event that she is awake during the Experiment and aware of that fact, according to the rules during the Experiment,

$$P_S(E) = 1.$$ 

So, $P_S(\text{Heads} \mid E) = 1/2$, since conditioning on a sure-event leaves probabilities unchanged. But when Sleeping Beauty is awake during the Experiment, that fact (i.e., that the Experiment is running) is all that she learns has happened since going to sleep on Sunday. That is, event $E$ represents the totality of her new evidence between retiring Sunday and being awakened during the Experiment. Hence, using conditional probability as her rule for updating upon awakening – we model Sleeping Beauty as a canonical Bayesian, one who uses Bayes’ rule to update her degrees of belief – then,

$$P_E(\text{Heads}) = P_E(\text{Heads} \mid E) = P_S(\text{Heads} \mid E) = 1/2.$$ 

\[2\] Our analysis here shows that Sleeping Beauty may apply Bayesian conditionalization to update her coherent opinions from Sunday with respect to the evidence that she acquires when she is awake during the Experiment. This is possible even though she is required to suffer the memory loss of Monday’s events and, therefore, understands that she does not know whether it is Monday or Tuesday when awake during the Experiment. Thus, we dispute Pust’s (2012, fn. 3, 296) account of what the position in Schervish, Seidenfeld, and
We have recreated a familiar debate between Thirders and Halfers regarding what should be Sleeping Beauty’s credence in Heads whenever she is awake during the Experiment. In this paper we focus on two aspects of this debate.

**Theme 1:** Eliciting probabilities with fair prices for gambles depends upon the details of the gambles being priced.

Specifically, we establish in Section 2 of this paper that, because Sleeping Beauty is required to make a second contract for a bet on Heads placed on Tuesday, which occasion arises only if the coin lands Tails, and because of her mandated forgetfulness, her *fair price* for betting on Heads and her *credence* in Heads are not the same quantities. Though her fair price for this bet is in a 1-1 relation with her credence in Heads, nonetheless the two are different quantities. As we establish, when awake during the Experiment her *fair price* for a bet on Heads is $1/3$ *if and only if* her *credence* in Heads is $1/2$. (We note that both Bradley and Leitgeb (2006) and Yamada (*no date*) also distinguish between what might be Sleeping Beauty’s *credence* and her *fair betting odds* for Heads.)

**Theme 2:** The *Law of Total Probability* fixes unconditional probabilities from a collection of conditional probabilities, thought of as constraints, provided that the conditional probabilities are taken over a (finite) *partition* of events.

Kadane (2004) entails about a rational agent’s ability to apply conditionalization in case of an anticipated memory loss of the kind that Sleeping Beauty faces during the Experiment.

As we show in Section 3, it is not sufficient that the three events \(\{A, B, C\}\) used in the Thirders’ argument #2 are collectively exhaustive. We show that the indexical ‘now’ creates an exhaustive collection of three events \(\{A, B, C\}\) that do not form a partition from Sleeping Beauty’s perspective during the time of the Experiment! When dealing with a collection of exhaustive but not necessarily disjoint conditions, it is *The Law of Too Much Probability* that applies, rather than the *Law of Total Probability*. Using the appropriate *Law* and the constraints from the Thirders’ argument #2, we show that Sleeping Beauty’s rational credence in Heads is 1/2.

2. Fair bets and credences in the Sleeping Beauty Problem.

2.1 Coherence. We begin with a summary of de Finetti’s (1974) theory of coherence for previsions and conditional previsions – which has betting on events as a special case.

Let \(\Omega = \{\omega_1, \ldots\}\) be a set of possible states (constituents) that partition what for the rational agent, the *Bookie*, is the sure-event, which we also identify with \(\Omega\). (In Section 3 we discuss how this partition is identified.) Let \(\chi = \{X: \Omega \rightarrow \mathbb{R}; \ i = 1, \ldots\}\) be a collection of real-valued variables over \(\Omega\).

*Coherence as fair-prices:* Consider a 2-person, 0-sum sequential prevision game. Player #1, the *Bookie*, moves first and sets a fair (buy/sell) price \(P(X)\) for each \(X \in \chi\). Player #2, the *Gambler*, then acts on the *Bookie*’s offers. The *Gambler* may make finitely many (non-trivial) contracts at the *Bookie*’s announced prices. That is, for finitely many \(X\), the *Gambler* fixes a non-zero real number, \(\beta_X\) that defines a contract, as follows.
In state $\omega$, a contract has an outcome to the Bookie (and negative outcome to Gambler) of

$$\beta_X[X(\omega) - P(X)].$$

When $\beta_X > 0$ Bookie is buying the variable $X$ from Gambler at price $P(X)$.

When $\beta_X < 0$ Bookie is selling the variable $X$ to Gambler at price $P(X)$.

The Bookie’s net outcome in state $\omega$ is the sum of the payoffs from the finitely many non-zero contracts:

$$\sum_{X \in \chi} \beta_X[X(\omega) - P(X)].$$

**Definition:** The Bookie’s previsions are incoherent if there is a finite combination of acceptable gambles which, with respect to the partition $\Omega$, yields a uniformly negative net-payoff to the Bookie.

Otherwise the Bookie’s previsions are coherent.

Let $B$ be an event, identified with its indicator function, $B(\omega)$.

$$B(\omega) = 1 \text{ if event } B \text{ obtains in state } \omega.$$  

$$B(\omega) = 0 \text{ if } B \text{ does not obtain in state } \omega.$$  

When the variables in $\chi$ are indicator functions, we have the familiar set up of betting on an event, where:

- the $\beta$-term fixes the stake of the bet and whether the Bookie is buying or selling the variable,

and the prevision, $P(B)$, is the Bookie’s fair betting rate,
where \( \beta_\omega [B(\omega) - P(B)] \) is the gain/loss in state \( \omega \) to the Bookie from this winner-take-all bet.

De Finetti’s Coherence Theorem (1974, chapter 3) establishes that a set of previsions \( \{P(X_i)\} \) are coherent if and only if they coincide with the expected values of these variables, \( \mathcal{E}_P(X_i) \), where the expectations are derived from some (finitely additive) probability \( P \).

When the variables are indicator functions, then the coherent prevision satisfy:

\[
P(B_i) = P(B_i),
\]

where \( P(B_i) \) is the probability \( P(B_i) \) of the event \( B_i \), which equals the \( P \)-expected value of the indicator function \( B_i(\omega) \). That is, \( \mathcal{E}_P(B_i) = P(B_i) \).

### 2.2 Elicitation of coherent credences:

Since the Bookie does not know in advance whether the Gambler will choose \( \beta > 0 \) or \( \beta < 0 \), in order to make each contract fair, i.e. with expected value 0, the Bookie uses the probability \( P(\cdot) \) that is her/his personal degree of belief, her/his credence function, to fix previsions.

De Finetti’s Prevision game extends to conditional previsions using called-off contracts.

Let \( B \neq \emptyset \) be an event that is not the impossible event. The conditional prevision \( P(X \mid B) \) is the Bookie’s fair price in contracts of form:

\[
[\beta_{X,B} B(\omega)[X(\omega) - P(X \mid B)]].
\]

In the special case of previsions and conditional previsions for events, coherence then requires that previsions satisfy the equation

\[
P(AB) = P(A \mid B)P(B).
\]
Note well that, even with conditional previsions, the game is synchronic. There are no learning or dynamic features to (de Finetti’s) coherence criterion.

2.3 Sleeping Beauty’s fair previsions. Next we apply de Finetti’s theory to determine Sleeping Beauty’s fair previsions while awake during the Experiment, where she plays the role of the Bookie. We begin by defining a class of $SB$-variables that covers random quantities that she might be asked to price while in the unusual circumstances of the Sleeping Beauty problem.

Let MH, MT, and TuT be the following three events:

<table>
<thead>
<tr>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH</td>
<td>$SB$ is awake on Monday and the fair coin land Heads.</td>
</tr>
<tr>
<td>MT</td>
<td>$SB$ is awake on Monday and the fair coin lands Tails.</td>
</tr>
<tr>
<td>TuT</td>
<td>$SB$ is awake on Tuesday and the fair coin lands Tails.</td>
</tr>
</tbody>
</table>

**Definition** A general Sleeping Beauty variable $X_{a,b,c}$ is one that

- pays $a$ if event MH occurs
- pays $b$ if event MT occurs,
- and pays $c$ if event TuT occurs.

That the events MT and TuT are not exclusive is important in understanding how Sleeping Beauty’s fair price for betting on an event and her credence for that same event may be two different quantities. The event MT occurs if and only if the event TuT occurs. They are not disjoint events. That is, within the space of possibilities for the Sleeping Beauty
Experiment, there is no history of the world in which one of these two occurs without the other occurring as well.

While awake during the Experiment, let $P_{a,b,c}$ be *Sleeping Beauty’s* announced price (*prevision*) for the gamble that pays $X_{a,b,c}$. Then, the net cash flow in state $\omega$ from a contract on a generalized SB-variable can be written:

$$
\beta[\text{Heads}(\omega)(a - P_{a,b,c}) + \text{Tails}(\omega)2(b + c - P_{a,b,c})]
$$

We see this as follows. Recall that *Sleeping Beauty* has to make a fresh contract each time she is awake. When the fair coin lands Heads, then *Sleeping Beauty* makes only one contract, on Monday, with net proceeds from the contract equal to $\beta(a - P_{a,b,c})$. When the fair coin lands Tails, *Sleeping Beauty* makes two contracts during the Experiment, one on Monday and another on Tuesday. Each contract has net proceeds of $\beta(b + c - P_{a,b,c})$. Then, when the fair coin lands Tails, the net proceeds from both contracts equals $2\beta(b + c - P_{a,b,c})$.

The term $\beta$ is chosen by the *Gambler* after *Sleeping Beauty* announces $P_{a,b,c}$. During the Experiment, *Sleeping Beauty* is required to give the same *fair price* $P_{a,b,c}$ for gambling on $X_{a,b,c}$ whenever she is awake, as she cannot tell the days apart.\(^4\) We require that the *Gambler* must choose the same $\beta$-term each time *Sleeping Beauty* makes her offer, since we are not

\(^4\) In the canonical *Sleeping Beauty* problem, which is the topic of our analysis in this paper, *Sleeping Beauty* has the same total evidence whenever she is awake during the Experiment.
allowing the *Gambler* to take advantage of knowing, e.g., that the elicitation is on Tuesday, which entails that the event TuT occurs.

Let

\[ y = P_E(\text{Heads}). \]

That is, *y* is Sleeping Beauty’s credence, when awake during the Experiment, that the coin lands Heads. Then, from Sleeping Beauty’s perspective when awake during the Experiment, her expected cash flow from a generalized *SB*-gamble on \( X_{a,b,c} \) whose announced price is \( P_{a,b,c} \) is:

\[
\beta[y(a - P_{a,b,c}) + (1-y)2(b + c - P_{a,b,c})].
\]

The *fair price* that makes this expectation equal 0, satisfies

\[
P_{a,b,c} = \frac{[y(a - 2(b + c)) + 2(b + c)]}{(2-y)} \tag{2}
\]

The *standard SB-bet* on Heads is the special case in which \( a = 1 \) and \( b = c = 0 \). Thus, Sleeping Beauty’s prevision – her announced *fair price* during the Experiment – is the value for \( P_{1,0,0} \) where

\[
P_{1,0,0} = \frac{y}{(2-y)} = P_E(\text{Heads}) / [2 - P_E(\text{Heads})]. \tag{3}
\]

This is a continuous, 1-1 function that relates her *fair price* for an *SB*-bet \( P_{1,0,0} \) with her *credence* for Heads.

\[
y = 2P_{1,0,0}/(1+P_{1,0,0}) \tag{4}
\]

If, by the Thirders’ argument #1, Sleeping Beauty should announce \( P_{1,0,0} = 1/3 \) then

\[
y = 2P_{1,0,0}/(1+P_{1,0,0}) = 1/2.
\]

So the Thirders’ argument #1 for pricing the *SB*-bet on Heads at 1/3 leads to the Halfers’ conclusion that during the Experiment her credence in Heads is 1/2! Conversely, if during the Experiment her credence in Heads is 1/2, she ought to announce a price of 1/3.
Sleeping Beauty can express this reasoning also from her _ex ante_ perspective on Sunday. From that point of view, gambling on \( X_{a,b,c} \) during the Experiment is merely a conditional gamble from Sunday’s perspective – conditioned on the event \( E \) that she is announcing her prices when awake during the Experiment. That is, from Sunday’s perspective, she is pricing the called-off gamble:

\[
E(\omega)[\text{Heads}(\omega)(a - P^*_{a,b,c|E}) + \text{Tails}(\omega)2(b + c - P^*_{a,b,c|E})].
\]

But since Sleeping Beauty is a canonical Bayesian, during the Experiment she applies Bayes’ rule to update her _ex ante_ credences from Sunday by conditioning on her total evidence, \( E \), that she is awake during the Experiment according to the Experimental protocol. Thus, her fair price for the gamble on \( X_{a,b,c} \) during the Experiment, \( P_{a,b,c} \), equals her _ex ante_ fair price \( P^*_{a,b,c|E} \) for the called-off gamble on \( X_{a,b,c} \) given the event \( E \).

Examination of equation (2), which gives Sleeping Beauty’s fair price for a generalized _SB_-gamble while awake during the Experiment, shows that her prevision for the variable \( X_{abc} \) depends at most on her one degree of belief: \( y = P_E(\text{Heads}) \). To express the same point in different terms, there is no generalized _SB_-variable whose prevision depends upon Sleeping Beauty’s credences for the “centered” possibilities, \{ _Now is Monday, Now is Tuesday_ \} that comprise the Thirders’ argument #2. For example, the generalized _SB_-gamble on the event “ _SB is awake on Monday_ ” (\( a = b = 1, c = 0 \)), has a fair price \( P_{1,1,0} = (-y+2)/(2-y) = 1 \). This makes intuitive sense since, under the conditions of the Experiment, it is certain that she is awake on Monday. The coherent _fair price_ for the sure event is 1.
3. **Reasoning by Cases and the Law of Total Probability.**

A fact that we explore in this section, in connection the Thirders’ argument #2, is that, from her perspective within the Experiment, though the three events \{MH, MT, TuT\} are exhaustive for Sleeping Beauty, they do not form a partition of her space of serious possibilities. We begin by addressing how this situation distinguishes deductive from inductive *Reasoning by Cases*.

A familiar rule of deductive inference, *Reasoning by Cases*, is formulated as follows. Let $\Sigma \models \xi$ mean that sentence $\xi$ is deductively entailed by the set of sentences $\Sigma$. Then from the $n+1$ many assumptions that:

\[
\Sigma \models (\chi_1 \text{ or } \ldots \text{ or } \chi_n)
\]

\[
\Sigma \cup \{\chi_i\} \models \xi \quad (\text{for } i = 1, \ldots, n)
\]

it follows by elementary logic that $\Sigma \models \xi$.

Consider an inductive version, which we call *Probabilistic Reasoning by Cases*, where conditional probability plays the role of the entailment relation. Let $P_\Sigma(\cdot | \cdot)$ be a rational agent’s conditional credence function, and $P_\Sigma(\cdot)$ the same agent’s unconditional credence function, relative to background knowledge $\Sigma$. Let $\{C_1, \ldots, C_n\}$ be an exhaustive list of events: relative to $P_\Sigma$ it is certain that at least one occurs. Hence, $P_\Sigma(\bigcup_{i=1}^n C_i) = 1$. By *Probabilistic Reasoning by Cases* we mean the following rule.

Let $E$ be an event, $c_1$ and $c_2$ constants, such that for each $i = 1, \ldots, n$,

\[c_1 \leq P_\Sigma(E | C_i) \leq c_2.\]
Then, \( c_1 \leq P_{\Sigma}(E) \leq c_2 \).

When the set \( \{C_1, \ldots, C_n\} \) forms a partition, \emph{Probabilistic Reasoning by Cases} is valid and follows from the \emph{Law of Total Probability}:

\[
P_{\Sigma}(E) = \sum_i P_{\Sigma}(E | C_i) P_{\Sigma}(C_i).
\]

\emph{Probabilistic Reasoning by Cases} then is an instance of what de Finetti (1974) calls \emph{conglomerability}.\(^5\) However, when the finite set \( \{C_1, \ldots, C_n\} \) is exhaustive, but does not form a partition, \emph{Probabilistic Reasoning by Cases} is invalid even for countably additive probabilities. This is in contrast with deductive \emph{Reasoning by Cases}, where there is no corresponding requirement that the cases, \( \{\chi_1, \ldots, \chi_n\} \), form a partition relative to \( \Sigma \). See de Finetti (1972, 104). We illustrate with the following example.

Example: (For convenience, we suppress the subscript \( \Sigma \).) Consider a large, fair lottery with \( 2n \) many tickets, \( n > 1 \). Let \( T_i \) be the event that ticket \#\( i \) wins. Then \( \{T_i : i = 1, \ldots, 2n\} \) constitutes a partition. Let \( E \) be the event that the winning ticket has a number between 1 and \( n \), inclusive. So, from the assumption that the lottery is fair, \( P(E) = 1/2 \). Let \( C_i \) be the event \( E \cup \{T_{n+i}\} \), for \( i = 1, \ldots, n \). The collection of events, \( \{C_1, \ldots, C_n\} \) is exhaustive, i.e. the event \( C = \bigcup_i C_i \) is certain to occur. However, in violation of \emph{Probabilistic Reasoning by Cases}, though

\[
P(E | C_i) = n/(n+1) > 1/2 \quad \text{for each } i = 1, \ldots, n,
\]

\(^5\) For \emph{Reasoning by Cases} to be valid when \( \{C_1, \ldots, C_n, \ldots\} \) is a countably infinite partition, \( P \) is required to be countably additive. See, Schervish, Seidenfeld, and Kadane, (1984).
nonetheless, \[ P(E) = \frac{1}{2}. \]

With conditions that are exhaustive but not exclusive, valid probabilistic reasoning follows what we call the *The Law of Too Much Probability*. We express that *Law* for the elementary case of two exhaustive, but not mutually exclusive conditions, \( \{A', B'\} \). Let \( A', B' \) be events with \( A' \cup B' = \Omega \) and \( A' \cap B' = C \neq \emptyset \), and let \( D \) be an event.

*The Law of Too Much Probability:*

\[
P(D) = P(D|A')P(A') + P(D|B')P(B') - P(D|C)P(C).
\]

When the conditions \( \{A', B'\} \) are exhaustive and mutually exclusive, then \( C = \emptyset \) and the *Law of Too Much Probability* reduces to the *Law of Total Probability*.

A correct version of the Thirders argument #2 applies the *Law of Too Much Probability*, not the *Law of Total Probability*, as we argue next. If the coin lands Tails, then from Sleeping Beauty’s perspective while awake during the Experiment, the indexical ‘*now*’ refers both to Monday and to Tuesday. She has no resources, no events or other generalized SB-random variables that she can use for distinguishing the two instances of ‘*now*’. From her memory-restricted perspective \{Sleeping Beauty is *(now)* awake on Monday, Sleeping Beauty is *(now)* awake on Tuesday\} forms an exhaustive set, but not a partition. Next we apply the *Law of Too Much Probability* to this case.

Let \( A' \) and \( B' \) be the two (exhaustive) “centered” events:

- \( A' \) is the event \{*Now*, awake on Monday\}.
- \( B' \) is the event \{*Now*, awake on Tuesday\}, with \( P_{E}(B') = z \).
Let C and D be the two (mutually exclusive) “non-centered” events:

C is the event \{\text{Tails}\}, with \(P_E(C) = (1-y)\)

and D is the event \{\text{Heads}\}, with \(P_E(D) = y\).

The Thirders’ argument #2 posits that Sleeping Beauty’s conditional credence in Heads given that it is “now” Monday is 1/2, i.e. \(P_E(D|A') = 1/2\). Because Sleeping Beauty cannot use the ordinary indexical ‘now’ to distinguish Monday from Tuesday, and since she is surely awake on Monday, \(P_E(A') = 1\). Moreover, though the pair \{A', B'\} is exhaustive, they do not form a partition. \(A' \cap B' = C \neq \emptyset\). Then, by the Law of Too Much Probability

\[
P_E(D) = P_E(D|A')P_E(A') + P_E(D|B')P_E(B') - P_E(D|C)P_E(C).
\]

\[
= (1/2)1 + 0z - 0(1-y) = 1/2.
\]

Thus, the Law of Too Much Probability applied with the conditions from the Thirders’ argument #2 yields the Halfers’ conclusion that, during the course of the Experiment Sleeping Beauty’s credence in Heads is 1/2.

What is perhaps counterintuitive in this analysis is that, within her predicament of the Experiment, Sleeping Beauty cannot partition the space of possibilities using the exhaustive set of events \{I’m awake on Monday, I’m awake on Tuesday\}. This is further explained using de Finetti’s account of constituents (1974, 43-45, 106-110).

Constituents form the elements of the partition \(\Omega = \{\omega_i, \ldots \}\), the set of states needed for the theory of coherence. Those are the events that form the privileged partition used to
determine whether or not the Bookie’s previsions \( \{P(X_i)\} \) for the variables \( \{X_i\} \) are subject to a sure-loss when the Gambler chooses a strategy, the (finite) set of non-zero terms \( \{\beta_i\} \). How are constituents determined in de Finetti’s theory of coherence? The answer is to consider all the Boolean combinations events defined in terms of those random variables for which the Bookie assigns previsions. For instance, assume the Bookie assesses previsions \( P(X) \) and \( P(Y) \) for the two (simple) random variables \( X \) and \( Y \) having (respectively) sample spaces \( \{x_1, \ldots, x_m\} \) and \( \{y_1, \ldots, y_n\} \). Then the set of constituents formed from these two variables is the partition generated by considering all the \( mn \)-many Boolean combinations of the form \( (X = x_i) \& (Y = y_j) \) with \( 1 \leq i \leq m \) and \( 1 \leq j \leq n \).

Within the class of generalized SB-variables, defined in Section 2, are included three that serve as the indicator functions for these three events, which are jointly exhaustive:

- MH  Sleeping Beauty is awake on Monday and the fair coin land Heads.
- MT  Sleeping Beauty is awake on Monday and the fair coin lands Tails.
- TuT Sleeping Beauty is awake on Tuesday and the fair coin lands Tails.

Specifically, included among the generalized SB-variables are these three:

\[
X_{1,0,0} = 1 \text{ if and only if the event MH obtains; } X_{1,0,0} = 0, \text{ otherwise.}
\]

\[
X_{0,1,0} = 1 \text{ if and only if the event MT obtains; } X_{0,1,0} = 0, \text{ otherwise.}
\]

---

6 It is an important technical point within de Finetti’s theory that coherence is assessed relative to the privileged partition of constituents and not relative to arbitrary partitions. That is, because each merely finitely additive probability fails conglomerability in some denumerably infinite partition, sure-loss cannot be assessed in arbitrary (infinite) partitions. See, Seidenfeld and Schervish (1983) for further discussion of this point.
and $X_{0,0,1} = 1$ if and only if the event TuT obtains; $X_{0,0,1} = 0$, otherwise.

However, since $X_{0,1,0} = 1$ if and only if $X_{0,0,1} = 1$, the constituents formed from the class of generalized SB-variables coincides with whatever partition is created using the collection of gambles formed with these three variables. That is, the constituents are defined by considering the linear span (the set of linear combinations) of the class of generalized SB-variables and noting that the resulting partition is given by the pair \{Heads, Tails\}.\(^7\)

In equation (2), we observed that from within Sleeping Beauty’s perspective during the Experiment, each SB-variable has a prevision that depends solely on her credence over the elements of the binary partition \{Heads, Tails\}. Specifically, according to (2) her fair price for the SB-variable $X_{abc}$ satisfies $P_{a,b,c} = \frac{[y(a - 2(b + c)) + 2(b + c)]}{2 - y}$ where $y = P_{E}(\text{Heads})$.

So, while she is awake during the Experiment, the constituents which Sleeping Beauty has access to are given by the partition, \{Heads, Tails\}. She does not have access to

---

\(^7\) That the linear span is sufficient for determining the constituents is not generally the case for an arbitrary class $\chi$ of variables that the Bookie assesses with previsions. For an illustration when the Boolean combination of events defined by variables in $\chi$ and the linear span of the variables in $\chi$ generate different spaces, see Example 2, p. 525, of Seidenfeld, Schervish, and Kadane (2012).
{It is now Monday}, {It is now Tuesday} as a pair of constituents, since these two events are not exclusive for her in her predicament. If the coin lands Tails she experiences them both as ‘now’, though of course she has no knowledge of that fact.

There is an additional consideration that may help to explain the difficulty with the Thirders’ argument #2. As before, denote the three relevant “centered” events:

A  It is now Monday and the fair coin lands Heads.
B  It is now Monday and the fair coin lands Tails.
C  It is now Tuesday and the fair coin lands Tails.

The pivotal assumptions for the Thirders’ argument #2 are these two (as noted in fn. 1):

Assumption 1 \{A, B, C\} form a partition of the sure-event.

and  Assumption 2 \(P(B \mid \text{Tails}) = x\) with \(x < 1\),

from which it follows by the Law of Total Probability that \(P(\text{Heads}) < 1/2\).

But, as we have just seen, from her perspective while awake during the Experiment, Assumption 1 is false for Sleeping Beauty. During the Experiment, Assumption 2 is false for each of the others, the observers in the Experiment, who – unlike Sleeping Beauty – are able to tell whether ‘now’ refers to Monday or exclusively whether ‘now’ refers to Tuesday. There is no coherent point of view from within the Sleeping Beauty puzzle that supports both assumptions needed for applying the Law of Total Probability as in the Thirders’ argument #2. Neither Sleeping Beauty’s perspective, nor the perspective of the other characters in the story provides a model for the conjunction of Assumptions 1 and 2.

It is our purpose in this paper to emphasize two themes that we think help to address the counter-intuitive aspects of the Sleeping Beauty puzzle.

**Theme 1:** When relating *fair prices* for indicator functions to a rational agent’s *credences*, check that the elicitation process is straightforward.

It might be that, though you can identify the rational agent’s credences from her/his fair prices, the two are not identical. This is illustrated by Sleeping Beauty’s coherent price for the bet on Heads. The Thirder’s conclusion #1 for pricing the standard SB-bet on Heads, her *fair price* is 1/3, if and only if her *credence* in Heads is 1/2, in accord with the Halfer’s conclusion.

**Theme 2:** When applying probabilistic *Reasoning by Cases* through the *Law of Total Probability*, check that the cases are supported by reasoning in the partition of *constituents*.

If the cases are exhaustive but not mutually exclusive, you may have to reason by the *Law of Too Much Probability* and not the *Law of Total Probability*. We illustrate this by correcting the reasoning in the Thirders’ argument #2. Just as with the Thirder’s argument #1, correct reasoning leads to the Halfer’s conclusion that Sleeping Beauty’s credence in Heads ought to be 1/2.

We emphasize that Sleeping Beauty’s inability to create *constituents* for her space of possibilities corresponding to the familiar referents of the ordinary temporal indexical ‘now’ or ‘today’ is very different from an ordinary “senior moment” that you might experience
when you merely lose track of the time. Suppose you are teaching a class that ends promptly on the hour. There is no classroom clock to observe and you are concerned that you haven’t left enough time to finish the lecture as planned. You wonder to yourself, “I am only at page 5 of my notes for this class, and there are still three pages of notes left to give. I need another 10 minutes to finish the lecture as planned. Are there fewer than 10 minutes left now, or more time than that?”

You are not in Sleeping Beauty’s predicament! You can define the variable corresponding to “time that I’ve taken to present the first five pages of my notes.” This denotes some standard time, with a sample space of possible times corresponding to ‘now’ about which you are uncertain. This variable generates constituents that partition time in a familiar way.

Sleeping Beauty has no recourse to such a random variable. She has only the class of $SB$-variables to assess. She cannot identify a variable that plays the role in the ordinary case of “The time that I’ve taken to finish five pages of lecture notes” for distinguishing Monday from Tuesday. When the coin lands Tails, Sleeping Beauty’s total evidence is the same on Monday as on Tuesday. When the coin lands Tails, whatever experiences she has on Monday, she has the same experiences, in the same relation to one another, also on Tuesday.

The class of $SB$-variables does not provide Sleeping Beauty with a counterpart to this familiar variable (“time consumed completing the first 5 pages of the lecture”). When the coin lands Tails, the variable $X_{abc}$ has the outcome $X_{abc}(\text{Tails}) = b+c$, which does not permit her to distinguish a different conceivable outcome on Monday than on Tuesday.
Specifically, consider the two variables $X_{010}$ and $X_{001}$. The first has outcome 1 when Sleeping Beauty is awake on Monday and the coin lands Tails. The second has outcome 1 when Sleeping Beauty is awake on Tuesday and the coin lands Tails. Superficially, these appear to distinguish for her being awake on Monday when the coin lands Tails from being awake on Tuesday when the coin lands Tails. But that is an illusion. The two variables $X_{010}$ and $X_{001}$ have identical outcomes in each constituent generated by the class of $SB$-variables. When the coin land Heads, each has the outcome 0, and when it lands Tails, each has the outcome 1. Sleeping Beauty cannot conceive of an experience where these two variables have different outcomes, since when the coin lands Tails, all she can conceive on Monday is identical with what she can conceive on Tuesday.

Theme 2 reminds the decision maker that her/his space of serious possibilities, the set of constituents, is intimately linked to the class of random variables that she/he can assess with previsions. In that sense, the canonical Sleeping Beauty problem is an extreme and unusual case of forgetting. Fortunately, when we do become unsure of the current time (or the current location) we rarely, if ever become a Sleeping Beauty.
References


Yamada, M. (no date) Laying sleeping beauty to rest. philpapers.org/archive/YAMLSB