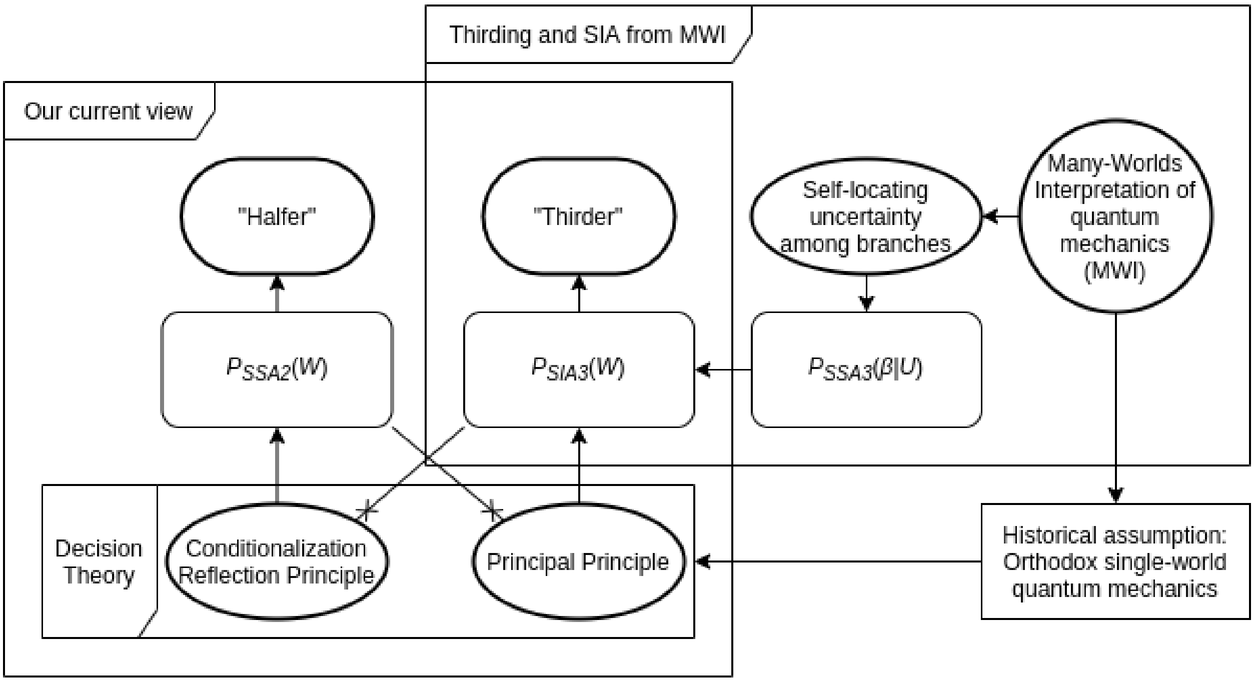
CHALMERS
UNIVERSITY OF TECHNOLOGY



Everettian Illusion of Probability and its Implications for Doomsday and Sleeping Beauty

Master's thesis

JOPPE WIDSTAM

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020
www.chalmers.se

MASTER'S THESIS 2020

Everettian Illusion of Probability and its Implications for Doomsday and Sleeping Beauty

JOPPE WIDSTAM

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Bayesian Statistics group
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Everettian Illusion of Probability and its Implications for Doomsday and Sleeping Beauty
JOPPE WIDSTAM

Supervisor & examiner: Olle Häggström, Department of Mathematical Sciences

Master's Thesis 2020
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Bayesian Statistics group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Our current view of incompatible conclusions in the Sleeping Beauty problem and the Doomsday Argument, stemming from the historical assumption of a single-world reality. If the Many-Worlds interpretation of quantum mechanics is correct, the illusion of probability based on self-locating uncertainty among world-branches is the source of the conflicting intuitions in our current view.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Everettian Illusion of Probability and its Implications for Doomsday and Sleeping Beauty
JOPPE WIDSTAM
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

An attempt at a satisfying compromise for the resolution to the Sleeping Beauty problem is presented. It is satisfying in virtue of explaining why the solution has eluded us for so long. The proposed reason for that the solution has eluded us is that the correct analysis of the Sleeping Beauty problem hinges on the consequences of the Many-Worlds interpretation of quantum mechanics being correct, consequences that have not been considered in the literature on the Sleeping Beauty problem.

The premise that the Many-Worlds interpretation is correct makes the argument speculative. If the speculation is correct though, this account avoids problems of automatic confirmation of hypotheses postulating a greater number of observers (e.g., the Presumptuous Philosopher problem) and avoids problems of credence about future chance events differing from objective chance (e.g., if the coin is tossed on Monday evening in the Sleeping Beauty problem), two problems that usually can't both be avoided.

The cost of accepting this account is that a distinction is made between credence in events with objective chance and credence from ignorance about deterministic facts. Beauty is a Thirder if both possibilities that she is uncertain between correspond to actually existent world-branches (i.e., resulting from a chance event) and a Halfer if they don't (i.e., ignorance about which possibility is actually existent). This distinction was proposed in this context by Wilson (2014) and is criticized in Bradley (2015). This account challenges the criticism by justifying the distinction from considerations about the physical nature of reality.

Keywords: Bayesian reasoning, Decision-theoretic rationality, Sleeping Beauty problem, Doomsday Argument, Everett, Many-Worlds interpretation.

Acknowledgements

To Professor Olle Häggström, my supervisor. For the conversations: They never failed to spark insight. For the opportunity: Because this is the thesis I was dreaming of writing. Truly, thank you!

I would also like to take this opportunity to express my genuine gratitude, and great debt, to the Swedish Board of Student Finance (CSN) and for the privilege of experiencing education being available freely and widely.

Joppe Widstam, Gothenburg, April 2020

Contents

List of Figures

List of Tables

1. Introduction

Imagine yourself in the position of Joe: He might not seem like much, but in fact he is in the peculiar situation of inhabiting one of the incredibly rare possible universes that allows for life to develop. This is not very strange to Joe, it is actually necessarily true for Joe that he inhabits such a universe. While the probability is very low, *a priori*, that a specific universe is able to harbor life, that probability is 1 *conditional* on that Joe exists in that universe. The observations of the universe that Joe, an observer, makes are biased, the sample will not be representative of the population being studied. This is known as a *selection effect*. Selection effects whose source are the circumstances of the observer's own existence are called *observation selection effects* or *anthropic bias*.

Again, imagine yourself in the position of Joe. There is a sense in which Joe could think of himself as a randomly sampled observer. Should Joe consider universes with more observers more likely than universes with fewer observers to contain himself, the one doing the sampling? Or is it still just whether a universe is observer-allowing or not that Joe should consider for the probability that he inhabits a particular universe? Should Joe consider himself a random sample from the *possible observers in all possible universes* or as a random sample from the *actual observers in the randomly selected actual universe*? These two options correspond to opposing positions, each with their virtues and problems, and their proponents and opponents, in two different but related concrete instances in formal epistemology. The two instances – the Doomsday Argument and the thought experiment known as the Sleeping Beauty problem – and how they are related is introduced in Section 2 and formalized as probability distributions in Section 3.

Section 4 proposes a resolution to the disagreement. The proposal presented is that the illusion of probability in the Everettian Many-Worlds interpretation of quantum mechanics entails that both positions apply, but in different circumstances. When credence is based in ignorance of deterministic facts the observer should consider herself a random sample from the *actual observers in the randomly selected actual universe* and when credence is based in the illusion of probability from quantum branching the observer should consider herself as a random sample from the *possible observers in all possible universes*. This proposal draws from the similar account of Wilson (2014) in which the role that in my account is filled by the illusion of probability from quantum branching instead is filled by the objective chance of the outcome of a chance event. Bradley (2015) criticizes this account, saying that applying the two positions in different circumstances is insufficiently motivated. This criticism, other approaches and ethical implications are discussed in Section 5.

The thesis concludes with a discussion on how resolving the Sleeping Beauty problem and the Doomsday Argument is a contributing reason for accepting the Many-Worlds interpretation.

2. Background

2.1 The Doomsday Argument

In the complete duration of the existence of the human race, some number of humans have come before you and some number of humans will come after you. What can probabilistically be said about the number of future humans given the number of past humans? First, ponder the following exercise in probability theory:

> There are two opaque urns, urn A and urn B. Urn A contains 100 balls numbered 1 to 100 and urn B contains one ball numbered 1. You draw one ball from one of the urns not knowing which urn is which. You look at the ball and see that it is numbered 1. What is the probability that the ball came from urn B?

Apply Bayes' Theorem with the prior probability distribution $P(A) = P(B) = 1/2$:

$$\begin{aligned} P(B|n=1) &= \frac{P(B)P(n=1|B)}{P(A)P(n=1|A)+P(B)P(n=1|B)} \\ &= \frac{1/2 \cdot 1}{1/2 \cdot 1/100 + 1/2 \cdot 1} \\ &= \frac{100}{101} \end{aligned} \tag{2.1}$$

The probability that the ball came from urn B is a little bit more than 99 percent. The posterior probability reflects that the probability of drawing ball $n = 1$ when drawing from urn B is much higher than when drawing from urn A. Now, ponder the following thought experiment, courtesy of Bostrom (2001):

> *The Incubator*
> *Stage (a)*: The world consists of a dungeon with one hundred cells. The outside of each cell has a unique number painted on it (which can't be seen from the inside); the numbers being the integers between 1 and 100. The world also contains a mechanism which we can term the *incubator*. The incubator first creates one observer in cell #1. It then activates a randomization mechanism; let's say it flips a fair coin. If the coin falls tails, the incubator does nothing more. If the coin falls heads, the incubator creates one observer in each of the cells [#2, #3, #4, ... #100]. Apart from this, the world is empty. It is now a time well after the coin has been tossed and any resulting observers have been created. Everyone

> knows all the above.
> *Stage (b)*: A little later, you have just stepped out of your cell and discovered that it is #1. (Bostrom, 2001, p. 6)

The Incubator is very similar to the urn example. In the urn example, one ball is sampled from one of two different possible distributions. In the Incubator, you are the sample and you could have been sampled from one of two different distributions. Given that the sample has a specific attribute ($n = 1$) the question is: What are the posterior probabilities for the distribution of these distributions? If Equation 2.1 also holds for the Incubator it suggests that the posterior probability distribution is $P(Heads) = 1/101$ and $P(Tails) = 100/101$, i.e., you should be more than 99% certain that the fair coin showed Tails and that you were sampled from the distribution where no other observers than you were created.

Translating the Incubator into our own cosmological situation we have a Doomsday Argument. Assume two equally probable hypotheses on the longevity of the human species:

h_1 = A total of 10^{12} humans will have ever lived
h_2 = A total of 10^{15} humans will have ever lived

Every human has a number in line from the first to the last human born. Call that number that human's *birth rank*. Conditionalizing on your birth rank produces a Bayesian probability shift on the probability distribution for the longevity of the human species, redistributing the ignorant prior probabilities towards smaller number of humans. Obtaining the evidence that your birth rank is $n = 10^{11}$ and conditionalizing on this evidence (with the prior probability distribution $P(h_1) = P(h_2) = 1/2$) we obtain the posterior probability distribution on the hypotheses.

$$
\begin{aligned}
P(h_1|n = 10^{11}) &= \frac{P(h_1)P(n = 10^{11}|h_1)}{P(h_1)P(n = 10^{11}|h_1) + P(h_2)P(n = 10^{11}|h_2)} \\
&= \frac{1/2 \cdot 10^{-12}}{1/2 \cdot 10^{-12} + 1/2 \cdot 10^{-15}} \\
&= \frac{1000}{1001} \\
P(h_2|n = 10^{11}) &= \frac{P(h_2)P(n = 10^{11}|h_2)}{P(h_1)P(n = 10^{11}|h_1) + P(h_2)P(n = 10^{11}|h_2)} \\
&= \frac{1/2 \cdot 10^{-15}}{1/2 \cdot 10^{-12} + 1/2 \cdot 10^{-15}} \\
&= \frac{1}{1001}
\end{aligned}
\tag{2.2}
$$

The probabilities have been redistributed from the prior probability distribution $P(h_1) = P(h_2) = 1/2$ into a posterior probability distribution where smaller number of humans are more probable: $P(h_1) = 1000/1001$ and $P(h_2) = 1/1001$. Doom is expected sooner.

2.2 The Sleeping Beauty Problem

Beauty is participating in an experiment. The setup of the experiment is as follows. On Sunday, Beauty will go to sleep and the experimenters will flip a fair coin. If the coin comes up Heads they will wake Beauty on Monday. She will then sleep through Tuesday. If the coin comes up Tails, they will wake her up on Monday and Tuesday. Between Monday and Tuesday they administer a memory-erasure drug that will make her forget that she has been awake on Monday. (See Figure 2.1.)

Beauty is well-versed in Bayesian rationality and knows the setup of the experiment. When Beauty wakes up, what should be her credence, her probability estimate, for that the coin came up Heads? Is it 1/2? Is it 1/3? Proponents of the former conclusion, e.g., Lewis (2001), are called *Halfers* and proponents of the latter, e.g., Elga (2000), are called *Thirders*.

When waking up, Beauty has two different kinds of uncertainty. She is uncertain about the outcome of the coin toss, i.e., what the world is like, and she is uncertain about the current day, i.e., her temporal location in the world. In possible world semantics, the two kinds of uncertainty is about which *uncentered* and *centered* possible world is actual. An uncentered possible world is what the world could be like and a centered possible world is an uncentered possible world indexed with a specific individual within that world at a specific time. The Halfer and Thirder positions can be seen as different ways of dealing with the interplay of centered and uncentered uncertainty.

	Heads	Tails
Monday	Awake	Awake
Tuesday	zzz	Awake

Figure 2.1: The structure of the Sleeping Beauty problem. The columns are uncentered worlds representing the different outcomes of the coin toss. The memory-erasure drug administered between Monday and Tuesday in the Tails-world makes Beauty uncertain about her location among the three centered worlds she could have awakened in.

2.2.1 Halfers

When waking up, Beauty obtains the centered evidence that she is in one of the three centered possible worlds. Between going to sleep on Sunday and waking she obtains no new uncentered evidence relevant to whether the coin came up Heads or Tails.

The Halfer-conclusion follows from two premises. The first premise is that change in credence is only produced by new relevant evidence. Violating this premise violates the *principle of conditionalization* (Lewis, 1999; Teller, 1973), one of the cornerstones of Bayesian reasoning. The other premise is that the centered evidence that she is in one of the three centered possible worlds is not relevant evidence for whether the coin came up Heads or Tails. Violating this premise violates another cherished principle in Bayesian reasoning: Van Fraassen's (1984) Reflection Principle. The principle states that if you know that you in the future will assign a certain credence to an uncentered proposition you should already now assign that credence. If the evidence was relevant, the Reflection Principle would mandate a credence in Heads of 1/3 already on Sunday.

From these premises it follows that her credence when she is awakened during the experiment, P_{awake}, remains unchanged from her credence on Sunday, $P_{objective}$, which matches the objective chance for the outcome of the coin toss.

$$\begin{aligned} P_{awake}(Heads) = P_{objective}(Heads) = 1/2 \\ P_{awake}(Tails) = P_{objective}(Tails) = 1/2 \end{aligned} \tag{2.3}$$

If Beauty knows that the coin came up Heads then she is certain it is Monday.

$$P_{awake}(Monday|Heads) = 1 \tag{2.4}$$

Further, if Beauty knows that the coin came up Tails, the only way she can differentiate between the propositions “Today is Monday” and “Today is Tuesday” is by the names of the days. The Principle of Indifference is a principle applicable in situations like this. The principle prescribes assigning equal credence to propositions differentiable only by their label. In this case both Halfers and Thirders are motivated by this principle to say that Beauty should assign equal credences to the two propositions.

$$P_{awake}(Monday|Tails) = P_{awake}(Tuesday|Tails) = 1/2 \quad \text{(Principle of Indifference)}$$

Beauty’s credence in Heads and Tails is the same as the objective chance and according to the Principle of Indifference her centered credence is distributed as shown in Figure 2.2.

	Heads	Tails
Monday	1/2	1/4
Tuesday	zzz	1/4

Figure 2.2: Beauty’s credence for each awakening according to Halfers

2.2.2 Thirders

The problem does not change significantly if the coin is tossed on Monday evening instead of Sunday. In either case, Beauty will wake up on Monday and then with probability 1/2 she will wake up on Tuesday. Her credences about which awakening she is currently experiencing are also unaffected by the timing of the coin toss.

If Beauty were to learn that the current day is Monday, and she knows that the experimenters will toss the coin on Monday evening then she knows that the coin toss takes place in her future. Beauty’s credences in the centered propositions “Heads and today is Monday” and “Tails and today is Monday” are then her credences about the outcome of a fair coin about to be tossed. The probability that a coin will land Heads (Tails) is the objective chance of Heads (Tails). There is a quite reasonable principle in Bayesian reasoning, called the *Principal Principle*, that states that credence should match objective chance, i.e., real physical randomness (Lewis, 1980). Beauty's credences are then $P_{awake}(Heads|Monday) = P_{objective}(Heads)$ and $P_{awake}(Tails|Monday) = P_{objective}(Tails)$.

The coin is fair so the objective chance for the outcome of the coin toss is $P_{objective}(Heads) = P_{objective}(Tails) = 1/2$. From there it is just a matter of reasoning backward through Bayes' Theorem to find out what credences she ought to have had at the time before she learned that it is actually Monday. Doing this we find $P_{awake}(Heads)$ and, analogously, $P_{awake}(Tails)$ to be:

$$\begin{aligned}
P_{awake}(Heads) &= \frac{\frac{P_{objective}(Heads)}{P_{awake}(Monday|Heads)}}{\frac{P_{objective}(Heads)}{P_{awake}(Monday|Heads)} + \frac{P_{objective}(Tails)}{P_{awake}(Monday|Tails)}} \\
&= \frac{\frac{1/2}{1}}{\frac{1/2}{1} + \frac{1/2}{1/2}} \\
&= 1/3 \\
P_{awake}(Tails) &= \frac{\frac{P_{objective}(Tails)}{P_{awake}(Monday|Tails)}}{\frac{P_{objective}(Heads)}{P_{awake}(Monday|Heads)} + \frac{P_{objective}(Tails)}{P_{awake}(Monday|Tails)}} \\
&= \frac{\frac{1/2}{1/2}}{\frac{1/2}{1} + \frac{1/2}{1/2}} \\
&= 2/3
\end{aligned} \tag{2.5}$$

Distributing the credence $P_{awake}(Tails) = 2/3$ equally across Monday and Tuesday according to the Principle of Indifference we obtain the credences shown in Figure 2.3.

	Heads	Tails
Monday	1/3	1/3
Tuesday	zzz	1/3

Figure 2.3: Beauty's credences for each awakening according to Thirders

2.2.3 The Disagreement

The two positions are argued by from reasonable premises but have incompatible conclusions (Table 2.1).

Table 2.1: Conditionalization, the Reflection Principle, and the Principal Principle and their respective conclusions in the Sleeping Beauty problem (SBP)

Conditionalization Reflection Principle	1/2
Principal Principle	1/3

By maintaining that Beauty's credences should not change between Sunday and her waking up, as not to violate the principle of conditionalization or the Reflection Principle, Halfers are committed to assigning credences $P_{awake}(Heads|Monday) = 2/3$ and $P_{awake}(Tails|Monday) = 1/3$. This is in direct contradiction of the Thirders' premise, motivated by the Principal Principle, that $P_{awake}(Heads|Monday) = 1/2$ and $P_{awake}(Tails|Monday) = 1/2$.

The conclusion Thirders draw, that $P_{awake}(Heads) = 1/3$ and $P_{awake}(Tails) = 2/3$, entails that between Sunday and her waking up during the experiment there is a change in credence. Either this change in credence was produced without obtaining any new relevant evidence (violating conditionalization) or it was relevant evidence that Beauty is in one of the three centered possible worlds (violating the Reflection Principle), thus contradicting at least one of the Halfers' premises.

Many authors consider their attempts at solutions to the Sleeping Beauty problem successful. Several of these arguments attempt to derive the correct conclusion to the Sleeping Beauty problem from non-straightforward applications of decision-theoretic principles, i.e., conditionalization, Reflection, and the Principal Principle. Attempts at justifying non-straightforward applications of these principles include arguing why it is acceptable to violate at least one of the principles in the Sleeping Beauty context (Elga, 2000; Monton, 2002; Arntzenius, 2003; Dieks, 2007) and suggesting plausible modifications or provisos for at least one of the principles (Lewis, 2001; Arntzenius, 2002; Bradley, 2011b).

2.3 The Similarity Between the Sleeping Beauty Problem and the Doomsday Argument

Let us consider the Doomsday Argument with two equally probable uncentered possible worlds, $W_{N=1}$ and $W_{N=2}$ with $N = 1$ and $N = 2$ total number of humans. An observer can have birth rank $n = 1$ or $n = 2$. The structure of the Doomsday Argument is then similar to the structure of the Sleeping Beauty problem (Figure 2.4), which has two hypotheses on the outcome of the coin toss, Heads and Tails, associated with $N = 1$ and $N = 2$ awakenings respectively (Dieks, 2007). Beauty can wake up on either Monday (Day #1) or Tuesday (Day #2).

	$W_{N=1}$	$W_{N=2}$
$n = 1$	Exists	Exists
$n = 2$		Exists

	W_{Heads}	W_{Tails}
Day #1	Awake	Awake
Day #2	zzz	Awake

Figure 2.4: Analogy between the Doomsday Argument and the Sleeping Beauty problem.

To complete the analogy we need to grant certain similarities between Beauty's situation and our own cosmological situation. The centered worlds in the Sleeping Beauty problem and the observers in our cosmological situation should in the relevant ways be treated in the same way. A centered world is an uncentered world indexed by a specific observer within that world at a specific time or duration of time.

A centered world can then be indexed with a specific observer for that observer's lifetime. The centered worlds in the Sleeping Beauty problem involve Beauty at different observer-moments (Bostrom, 2002a) corresponding to the different awakenings. Because of the peculiar situation Beauty is in, these observer-moments are not psychologically continuous with each other. Beauty could, just like us in our cosmological situation, have been any of these observer-moments and is uncertain which observer-moment she is currently experiencing. To be further convinced by the similarity, consider the structurally similar Beauty and Doppelganger thought experiment:

> *Beauty and Doppelganger*
> This is like the original Sleeping Beauty problem, except here Beauty is never woken up after being put to sleep on Monday. Instead, if the coin falls tails, another person is created and awoken on Tuesday. This new person will spend her Tuesday wakening in a state that is subjectively indistinguishable from Beauty's Monday state (she will have the same apparent memories and have experiences that feel just the same as Beauty's). When Beauty awakes on Monday, what should be her credence in HEADS? (Bostrom, 2007, p. 63)

Taking another small step, consider a scenario where no Beauty existed on Sunday but one or two Beautys are created depending on the outcome of the coin toss. Any Beauty that is created is informed about the situation and asked about her credence that the coin came up Heads. This scenario is very similar to the Incubator thought experiment used to motivate the Doomsday Argument in Section 2.1.

In the Sleeping Beauty problem there is disagreement on what are the correct credences for an observer to have about the uncentered worlds. Halfers think that Beauty's credences should match the objective chances of the uncentered worlds. Thirders think instead that Beauty's credences when she knows it is Monday should match the objective chances. In the Doomsday Argument we assumed, like the Halfers, that an observer's credences match the objective chances of the uncentered worlds, but this might not be the case. Section 2.4 will introduce an alternative assumption.

2.4 SSA and SIA

Imagine these two scenarios:

> *Cloakroom*
> You are getting your coat from the cloakroom. The room is completely dark and you can't see where it ends. You find your coat half a meter into the room.

> *Lost-Property Warehouse*
> You are looking for your coat in a big lost-property warehouse. The warehouse has lines and lines of aisles with different lengths. You search through these one by one. In one aisle you find your coat half a meter in. (Wikipedia contributors, 2018)

The cloakroom scenario is akin to the reasoning employed in the Doomsday Argument. How far inside you find your coat strongly influences your credence on how deep the cloakroom is. The implication in the lost-property warehouse scenario is that since long aisles contain more lost property than short aisles your coat is more likely to be stored in a long aisle *a priori*. How far into the aisle you find your coat influences your credence in the length of the aisle but is offset by the prior probability favoring longer aisles. The same considerations can be employed in our cosmological situation where possible universes with more observers might be *a priori* more probable than possible universes with few observers thus nullifying the Doomsday Argument (Table 2.2). Bostrom (2002a) formulated these two assumptions as follows:

> *The Self-Sampling Assumption* (SSA)
> All other things equal, an observer should reason as if they are randomly selected from the set of all *actually existent* observers (past, present and future) in their reference class.
>
> *The Self-Indication Assumption* (SIA)
> All other things equal, an observer should reason as if they are randomly selected from the set of all *possible* observers.

Table 2.2: The Self-Sampling Assumption (SSA) and the Self-Indication Assumption (SIA) and their respective conclusions in the Doomsday Argument (DA)

SSA	Doom
SIA	No doom

In the analogy between the urn example (Equation 2.1) and our cosmological situation in the Doomsday Argument from Section 2.1, the set of balls in the urn that is selected to be drawn from is analogous to the set of observers in the world that is selected to exist. In this analogy an observer is randomly selected from the set of all actually existent observers, like in the Self-Sampling Assumption (SSA). If SSA is the correct assumption, this analogy is correct for our cosmological situation.

In the lost-property warehouse metaphor for the Self-Indication Assumption (SIA), any place in any aisle is equally likely to contain your coat. Drawing from this metaphor, we can modify the urn example to one where any ball is equally likely to be drawn, one where the ball is instead randomly drawn from the union of the sets of balls in each urn. In the analogy between the modified urn example and our cosmological situation, the union of the sets of balls in each of the urns is analogous to the union of the sets of observers in each of the possible worlds. In this analogy an observer is randomly selected from the set of all possible observers, like in SIA. If SIA is the correct assumption, this is the analogy that is correct for our cosmological situation.

3. Probability of a Possible World

The two positions in the Doomsday Argument correspond to the two positions in the Sleeping Beauty problem. In this section each pair of positions is formalized as a probability measure on possible worlds. Table 3.1 summarizes the section.

3.1 Halfing and SSA as $P_{SSA2}(W)$

We have seen that with SSA, an observer's credences in uncentered possible worlds should be proportional to the objective probabilities of those worlds. Observers reasoning according to SSA have this in common with Beautys that are Halfers. Define $P_{SSA2}(W)$, with the index referring to SSA and 1/2, as:

$$\begin{aligned} P_{SSA2}(W_{N=k}) &= P_{objective}(W_{N=k}) \sum_{n=1}^{k} P(n|W_{N=k}) \\ &= P_{objective}(W_{N=k}) \sum_{n=1}^{k} 1/k \\ &= P_{objective}(W_{N=k}) \end{aligned} \tag{$P_{SSA2}(W)$}$$

In Equation 3.1 $P_{SSA2}(W)$ is used with the two equally probable worlds $W_{N=1}$ and $W_{N=2}$ to get the conditional probabilities of the worlds given $n = 1$. The equation can be interpreted as describing the Doomsday Argument or Beauty's credence after obtaining and conditionalizing on the evidence that it is Monday.

$$\begin{aligned} P_{SSA2}(W_{N=1}|n=1) &= \frac{P_{SSA2}(W_{N=1})P(n=1|W_{N=1})}{P_{SSA2}(W_{N=1})P(n=1|W_{N=1}) + P_{SSA2}(W_{N=2})P(n=1|W_{N=2})} \\ &= \frac{P_{objective}(W_{N=1})P(n=1|W_{N=1})}{P_{objective}(W_{N=1})P(n=1|W_{N=1}) + P_{objective}(W_{N=2})P(n=1|W_{N=2})} \\ &= \frac{1/2 \cdot 1}{1/2 \cdot 1 + 1/2 \cdot 1/2} = \frac{2}{3} \\ P_{SSA2}(W_{N=2}|n=1) &= \frac{P_{SSA2}(W_{N=2})P(n=1|W_{N=2})}{P_{SSA2}(W_{N=1})P(n=1|W_{N=1}) + P_{SSA2}(W_{N=2})P(n=1|W_{N=2})} \\ &= \frac{P_{objective}(W_{N=2})P(n=1|W_{N=2})}{P_{objective}(W_{N=1})P(n=1|W_{N=1}) + P_{objective}(W_{N=2})P(n=1|W_{N=2})} \\ &= \frac{1/2 \cdot 1/2}{1/2 \cdot 1 + 1/2 \cdot 1/2} = \frac{1}{3} \end{aligned} \tag{3.1}$$

3.1.1 P_{SSA2} and Objective Chance of Future Events

A problem for the Halfer-position, i.e., applying $P_{SSA2}(W)$ in the Sleeping Beauty problem, is that after learning that it is Monday, Beauty's credence in the coin toss that determines if she is awakened differs from the objective chance of the outcome of the coin toss. Her rational credence differs from objective chance even if the coin toss happens in her future, for example on Monday evening, or even if she tosses the coin herself.

3.2 Thirding and SIA as $P_{SIA3}(W)$

With SIA, the probability of sampling any specific possible observer is proportional to the objective probability of the world that observer inhabits. The probability of a randomly selected observer (selected according to SIA) inhabiting a specific uncentered world is proportional to the just mentioned objective probability of that uncentered world and to the number of observers inhabiting that world.

Recall the conditional probability of an uncentered world given your birth rank in the Doomsday Argument, given again in Equation 3.1. The term $P_{objective}(W)$ in this equation needs to be modified for SIA. The inverse of the term $P(n = 1|W)$ is exactly the proportionality constant for adjusting the objective probability, $P_{objective}(W)$, of the world W to be proportional to the number of observers inhabiting that world.

$$\frac{1}{P(n = 1|W_{N=k})} = k \tag{3.2}$$

In the general case, define $P_{SIA3}(W)$, with the index referring to SIA and 1/3, as:

$$\begin{aligned} P_{SIA3}(W_{N=k}) &= \frac{P_{objective}(W_{N=k})\frac{1}{P(n=1|W_{N=k})}}{\sum\limits_{W} P_{objective}(W_{N=h})\frac{1}{P(n=1|W_{N=h})}} \\ &= \frac{P_{objective}(W_{N=k}) \cdot k}{\sum\limits_{W} P_{objective}(W_{N=h}) \cdot h} \end{aligned} \tag{$P_{SIA3}(W)$}$$

Ross (2010) defines the structurally identical generalization for the Thirder-position in the Sleeping Beauty problem as follows:

> *Generalized Thirder Principle*
> In any Sleeping Beauty problem, defined by a partition S, upon first awakening, Beauty's credence in any given hypothesis in S should be proportional to the product of its objective chance and the number of times Beauty awakens if this hypothesis is true. (Ross, 2010, p. 414)

In Equation 3.3 we apply $P_{SIA3}(W)$ to get the conditional probabilities of the two possible worlds, $W_{N=1}$ and $W_{N=2}$, given that $n = 1$. The result is that these conditional probabilities equal the objective probabilities of the worlds. Equation

3.3 can have the interpretation that Doomsday is averted since the objective probabilities are restored after obtaining and conditionalizing one's birth rank. Another interpretation is that since $n = 1$ in the Doomsday Argument is analogous to "Today is Monday" in the Sleeping Beauty problem, the result corresponds to that if Beauty knows it is Monday then her credences in the uncentered worlds equal their objective chances. That Beauty's credences when she knows it is Monday should the equal the objective chances is the premise for the Thirder-conclusion, the argument from the Principal Principle. See Equation 2.5 for how the Thirder-conclusion follows from this premise.

$$\begin{aligned}
P_{SIA3}(W_{N=1}|n=1) &= \frac{P_{SIA3}(W_{N=1})P(n=1|W_{N=1})}{P_{SIA3}(W_{N=1})P(n=1|W_{N=1}) + P_{SIA3}(W_{N=2})P(n=1|W_{N=2})} \\
&= \frac{\frac{\frac{P(n=1|W_{N=1})}{P(n=1|W_{N=1})}P_{objective}(W_{N=1})}{\sum\limits_{W} P_{objective}(W_{N=h})\cdot h}}{\frac{\frac{P(n=1|W_{N=1})}{P(n=1|W_{N=1})}P_{objective}(W_{N=1})}{\sum\limits_{W} P_{objective}(W_{N=h})\cdot h} + \frac{\frac{P(n=1|W_{N=2})}{P(n=1|W_{N=2})}P_{objective}(W_{N=2})}{\sum\limits_{W} P_{objective}(W_{N=h})\cdot h}} \\
&= \frac{P_{objective}(W_{N=1})}{P_{objective}(W_{N=1}) + P_{objective}(W_{N=2})} \\
&= \frac{P_{objective}(W_{N=1})}{P(\Omega)} \\
&= P_{objective}(W_{N=1}) \\
P_{SIA3}(W_{N=2}|n=1) &= \frac{P_{SIA3}(W_{N=2})P(n=1|W_{N=2})}{P_{SIA3}(W_{N=1})P(n=1|W_{N=1}) + P_{SIA3}(W_{N=2})P(n=1|W_{N=2})} \\
&= \frac{\frac{\frac{P(n=1|W_{N=2})}{P(n=1|W_{N=2})}P_{objective}(W_{N=2})}{\sum\limits_{W} P_{objective}(W_{N=h})\cdot h}}{\frac{\frac{P(n=1|W_{N=1})}{P(n=1|W_{N=1})}P_{objective}(W_{N=1})}{\sum\limits_{W} P_{objective}(W_{N=h})\cdot h} + \frac{\frac{P(n=1|W_{N=2})}{P(n=1|W_{N=2})}P_{objective}(W_{N=2})}{\sum\limits_{W} P_{objective}(W_{N=h})\cdot h}} \\
&= \frac{P_{objective}(W_{N=2})}{P_{objective}(W_{N=1}) + P_{objective}(W_{N=2})} \\
&= \frac{P_{objective}(W_{N=2})}{P(\Omega)} \\
&= P_{objective}(W_{N=2})
\end{aligned} \tag{3.3}$$

3.2.1 P_{SIA3} and the Presumptuous Philosopher

Applying $P_{SIA3}(W)$ entails that hypotheses that postulate more observers are favored over hypotheses postulating fewer observers. The consequences of this are highly implausible, as highlighted by the following thought-experiment:

> *The Presumptuous Philosopher*
> It is the year 2100 and physicists have narrowed down the search for a theory of everything to only two remaining plausible candidate theories:

> T_1 and T_2 (using considerations from super-duper symmetry). According to T_1, the world is very, very big but finite and, there are a total of a trillion trillion observers in the cosmos. According to T_2 the world is very, very, *very* big but finite and there are a trillion trillion trillion observers. The super-duper symmetry considerations are indifferent between these two theories. Physicists are preparing a simple experiment that will falsify one of the theories. Enter the presumptuous philosopher: "Hey guys, it is completely unnecessary for you to do the experiment, because I can already show you that T_2 is about a trillion times more likely to be true than T_1!" (Bostrom, 2007, p. 64)

Table 3.1: The premises formalized and their conclusions and implausible consequences in the Sleeping Beauty problem and the Doomsday Argument.

		Sleeping Beauty problem	
Conditionalization Reflection Principle	$P_{SSA2}(W)$	1/2	Credences for future outcomes differ from their chances
Principal Principle	$P_{SIA3}(W)$	1/3	Presumptuous Philosopher

		Doomsday Argument	
SSA	$P_{SSA2}(W)$	Doom	Credences for future outcomes differ from their chances
SIA	$P_{SIA3}(W)$	No doom	Presumptuous Philosopher

4. Best of Both Worlds

4.1 Non-Chancy Sleeping Beauty

Wilson (2014) suggests that the Sleeping Beauty problem should be treated differently depending on if the uncentered uncertainty is expectation based on the objective chance of a chance event or is quantified ignorance about a deterministic fact. Following this suggestion, Beauty's credence should be 1/3 in Heads in the ordinary Sleeping Beauty problem and 1/2 in Odd in the Non-chancy Sleeping Beauty problem, outlined here:

> *Non-chancy Sleeping Beauty*
> On Sunday night, Beauty has credence 1/2 that an even number of stars will be visible in total on Monday night. She will be awakened on Monday if there is an odd number, and on both Monday and Tuesday (again with her memories from Monday erased) if there is an even number. Beauty knows all this. The puzzle is to say what credence Beauty should have on Monday in the proposition that the number of stars is Even. (Bradley, 2015, p. 8)

Wilson argues that the difference in treatment between the chancy and non-chancy versions of the Sleeping Beauty problem is motivated by the fact that the argument from the Principal Principle for the Thirder-position fails in non-chancy situations. The Principal Principle restricts credence to conform to objective chance (real physical randomness), and there is no objective chance in the quantification of one's ignorance in non-chancy situations. This amounts to adopting $P_{SIA3}(W)$ when W corresponds to outcomes of chance events and adopting $P_{SSA2}(W)$ when W corresponds to hypotheses that do not correspond to chance events. Bradley (2015) criticizes this account on two counts:

> First, most thirder arguments do still apply to Non-chancy Sleeping Beauty. Wilson would have to hold that all of these fail, but Elga's succeeds. Second, Non-chancy Sleeping Beauty seems to show that, intuitively, chancy cases in which the principal principle can be applied should get the same verdict as non-chancy cases in which the principal principle cannot be applied. Wilson has to explain away this intuition. (Bradley, 2015, p. 9)

The first count is countered in Section 5.2. The second count is about the motivation for applying $P_{SSA2}(W)$ and $P_{SIA3}(W)$ in different circumstances instead of applying $P_{SSA2}(W)$ in all of them. We will see in Section 4.4 that even if SSA is assumed, the Many-Worlds Interpretation of quantum mechanics (MWI) entails that, in at least a limited sense, the correct probability function in the Sleeping Beauty problem and in our cosmological situation is $P_{SIA3}(W)$. It is correct only in a limited sense because we will assume that the role of uncentered possible worlds in $P_{SIA3}(W)$ is filled by the actually existent parallel world-branches proposed by MWI, an insight previously explored in Jebari (2014). When the possible worlds are not the actually existent world-branches $P_{SSA2}(W)$ applies instead.

4.2 The Illusion of Probability in Many-Worlds

An appealing view of reality is that it is causally deterministic but practically infeasible to perfectly predict and that probability expresses uncertainty in the predictions. However one of the lessons from the most popular interpretations of quantum mechanics is that our reality is actually fundamentally probabilistic. Even when knowing everything about a system and the dynamics of its time evolution it is still not possible, even in theory, to predict the future state of the system. At best, what can be known are the probabilities, known as the Born probabilities, of the possible future states of the system.

The Many-Worlds Interpretation of quantum mechanics turns the lesson of the previous paragraph on its head. In the Many-Worlds Interpretation, reality is deterministic but still compatible with our observations of indeterminism. This is because the greater reality is that all possible future states of a system will exist and each will contain different future versions of the observer observing the system, each version only observing one possible future state of the system, one possible world-branch. The time evolution of this greater reality, the multiverse, is deterministic.

In orthodox interpretations of quantum mechanics only one possible future becomes the actual future and for each possible future there is some objective probability (the Born probabilities) that it becomes the actual. In the Many-Worlds interpretation there is an illusion of probability for the observer (Vaidman, 2012, 2018). In the future each version of the observer has the impression that their past self evolved through a single path in time to their current self, but that their past self could have evolved through another path and at the time were ignorant about which path would obtain. This impression of ignorance in the past about the present extrapolates into an impression of ignorance in the present about the future.

Lacking objective probability of the future for an observer, the required ignorance must be found somewhere else. A proposed solution depends on the apparent self-locating uncertainty of an observer in the period between the time a world-branch and the observer it contains has branched into different world-branches and the time when the observer observes which world-branch they inhabit (Vaidman, 1998). It turns out that the only rational credence assignment to each world-branch is the same value as the Born probability of the corresponding outcome in the orthodox interpretations of quantum mechanics (McQueen and Vaidman, 2019). In the Many-Worlds interpretation, and its illusion of probability, this prescription of credence is dubbed the *Born-Vaidman rule* (Vaidman, 2018; Tappenden, 2011) and the value is called the *measure of existence* of a world-branch (Vaidman, 1998).

4.3 Self-Location Among World-Branches

The Sleeping Beauty problem and the Doomsday Argument involve uncertainties about what the world is like and where or when in the world one is located. In the Many-Worlds interpretation and its illusion of probability both of these kinds of uncertainty are self-locating uncertainty (Groisman et al., 2013; Sebens and Carroll, 2018). Unlike in single-world theories where some possible observers exist and others don't, there is no qualitative difference between the possible observers in MWI. The uncentered question of what the world is like, when the worlds are world-branches of the universal wave-function, is actually the centered question of which world-branch one is in.

Using the notation from Sebens and Carroll (2018), the wave-function $|\Psi(t_i)\rangle$, describing the multiverse at time t_i, can be separated into the wave-function of the branch $|\beta_i\rangle$ containing the observer O_i at time t_i and the wave-function of the remainder of the multiverse $|\beta_i^{\perp}\rangle$.

$$|\Psi(t_i)\rangle = b_i|\beta_i\rangle + a_i|\beta_i^{\perp}\rangle \tag{4.1}$$

Each possible observer inhabits a world-branch (some of them the same), each with some measure of existence, but the observer is uncertain which of these observers she is. The measure of existence of branch β_i is the modulus square of b_i.

$$w(O_i) = |b_i|^2 \tag{4.2}$$

The credence an observer in this state of self-locating uncertainty should assign to being each of the possible observers is proportional to the measure of existence of the world-branches the respective possible observers inhabit. Equation 4.3 gives the probability of being observer O_i. The denominator is a normalizing factor and $O(U)$ is the set of all observers in the multiverse.

$$P(O_i|U) = \frac{w(O_i)}{\sum\limits_{O_j \in O(U)} w(O_j)} \tag{4.3}$$

When, as in Wilson (2013), the possibility space of possible worlds is restricted to the actually existent world-branches in the multiverse, U, the probability distribution on world-branches is given by the probability function $P_{SSA3}(\beta|U)$ below.

$$\begin{aligned} P_{SSA3}(\beta_{t(N=k)}|U) &= \sum_{O_i \in O(\beta_{t(N=k)})} P(O_i|U) = \frac{\sum\limits_{O_i \in O(\beta_{t(N=k)})} w(O_i)}{\sum\limits_{O_j \in O(U)} w(O_j)} \\ &= \frac{\sum\limits_{O_i \in O(\beta_{t(N=k)})} w(O_i)}{\sum\limits_{\beta \in U} \sum\limits_{O_j \in O(\beta_{t(N=k)})} w(O_j)} \\ &= \frac{|b_{t(N=k)}|^2 \cdot k}{\sum\limits_{\beta \in U} |b_{t(N=k)}|^2 \cdot h} \end{aligned} \tag{$P_{SSA3}(\beta|U)$}$$

The probability distribution on world-branches can be constructed by summing the measure of existence of the parts of the observers contained in the world-branch of interest, $O(\beta)$, over the sum of the measure of existence of all observers, $O(U)$. The sum in the denominator of $P_{SSA3}(\beta|U)$ can be separated into the non-overlapping world-branches that constitute U at the time of their last observer, $t(N)$. The measure of existence of an observer is divided into different of these non-overlapping world-branches. Different observers in the same world-branch have an equal measure of existence contained in it, equaling the measure of existence of the world-branch, $|b_{t(N)}|^2$. Since the terms in the summation are equal, the summation can be replaced with a multiplication by the number of observers in the world-branch.

This form of $P_{SSA3}(\beta|U)$ has the same form as $P_{SIA3}(W)$. Jebari (2014) refers to $P_{SSA3}(\beta|U)$ as "physical SIA" and $P_{SIA3}(W)$ as "metaphysical SIA", highlighting that the possibility spaces for these two probability distributions are the physically existent possible worlds (world-branches) and the metaphysically possible worlds, respectively. The simplification that all possibilities are restricted to be actually existent means that SSA and SIA agree on the set of observers that an observer should reason that she is randomly selected from, and that, given that W is a world-branch, the probability distribution on world-branches resulting from assuming SSA is described by $P_{SIA3}(W)$:

$$
\begin{aligned}
P_{SSA3}(\beta_{t(N=k)}|U) &= \frac{|b_{t(N=k)}|^2 \cdot k}{\sum\limits_{\beta \in U} |b_{t(N=h)}|^2 \cdot h} \\
&= \frac{P_{objective}(W_{N=k}) \cdot k}{\sum\limits_{W} P_{objective}(W_{N=h}) \cdot h} \\
&= P_{SIA3}(W_{N=k})
\end{aligned}
\tag{4.4}
$$

The index of $P_{SSA3}(\beta|U)$ refers to that, in the Many-Worlds setting, even if SSA is assumed instead of SIA, this probability function assigns 1/3 to Heads in the Sleeping Beauty problem.

4.4 Non-Branchy Sleeping Beauty

The account offered in this thesis is in the same vein as the account of Wilson (2014) from Section 4.1 but the difference in treatment between chancy and non-chancy situations is in the account in this thesis instead between branchy and non-branchy situations. Under the assumption that MWI is correct the outcomes of the coin toss are realized in different branches (assuming that the objective chance of the coin toss is quantum mechanical) and what was called the objective chance of a chance event in orthodox single-world interpretations is in the Many-Worlds setting instead the credence prescribed by the Born-Vaidman rule based on the observer's self-locating uncertainty among branches. In the Many-Worlds setting, let us call the ordinary chancy version of the Sleeping Beauty problem the *branchy Sleeping Beauty problem* and the non-chancy version the *non-branchy Sleeping Beauty problem*.

The branchy Sleeping Beauty problem results in two branches and all three possible observer-moments are existent. That they are existent means that there is no qualitative difference between the possible observer-moments and thus not only SIA but also SSA prescribe the credence assignment 1/3 to Heads, by $P_{SSA3}(\beta|U)$. Having the uncentered uncertainty about the outcome of the coin toss replaced by the centered uncertainty about which branch one is on has implications for van Fraassen's Reflection Principle, one of the premises for the Halfer-position described in Section 2.2.1. Since there is no uncentered proposition to assign credence to in the branchy Sleeping Beauty problem the principle does not apply and thus $P_{SIA3}(W)$ is not contradicted by it in the branchy Sleeping Beauty problem.

The non-branchy Sleeping Beauty problem does not result in two branches. Only one branch containing either one or two observer-moments of Beauty is actually existent and Beauty starts out with a credence of 1/2 for each of the two possibilities. In this case there is a qualitative difference between the possible observer-moments of Beauty, some of them exist and some do not. Beauty is uncertain which observer she is but this uncertainty is not self-locating uncertainty among existent possible observers, so in this case SSA prescribes the credence assignment 1/2 to each of the two possibilities, by $P_{SSA2}(U)$ below.

$$\begin{aligned} P_{SSA2}(U_{N=k}) &= P(U_{N=k}) \sum_{O_i \in O(U_{N=k})} P(O_i|U_{N=k}) \\ &= P(U_{N=k}) \frac{\sum_{O_i \in O(U_{N=k})} w(O_i)}{\sum_{O_j \in O(U_{N=k})} w(O_j)} \\ &= P(U_{N=k}) \end{aligned} \tag{$P_{SSA2}(U)$}$$

If MWI is correct there are two good reasons for accepting that $P_{SSA2}(W)$ and $P_{SIA3}(W)$ should each be applied in different circumstances over accepting that one or the other applies across all these circumstances.

The first reason is that it explains the source of the dividing disagreement that the Sleeping Beauty problem and the Doomsday Argument has given rise to. Equation 4.4 gives the reason for why $P_{SIA3}(W)$ is correct when credence is based on objective chance, since objective chance has its basis in self-locating uncertainty among world-branches as described by $P_{SSA3}(\beta|U)$. For $P_{SIA3}(W)$ to also be correct when credence is based on ignorance of a deterministic fact a separate reason needs to be given, since this credence does not have a basis in self-locating uncertainty among world-branches. The fact that $P_{SSA2}(U)$ is the correct prescription of credence in the latter and $P_{SSA3}(\beta|U)$ in the former is the source of the conflicting intuitions about the correct conclusion to the Sleeping Beauty problem and the Doomsday Argument (see Figure 4.1 for a schematic overview).

The second reason for accepting that $P_{SSA2}(W)$ and $P_{SIA3}(W)$ apply in different circumstances is that MWI entails avoiding the most implausible consequences of unrestrictedly applying any of them across these circumstances. The next two sections describe how these implausible consequences are avoided.

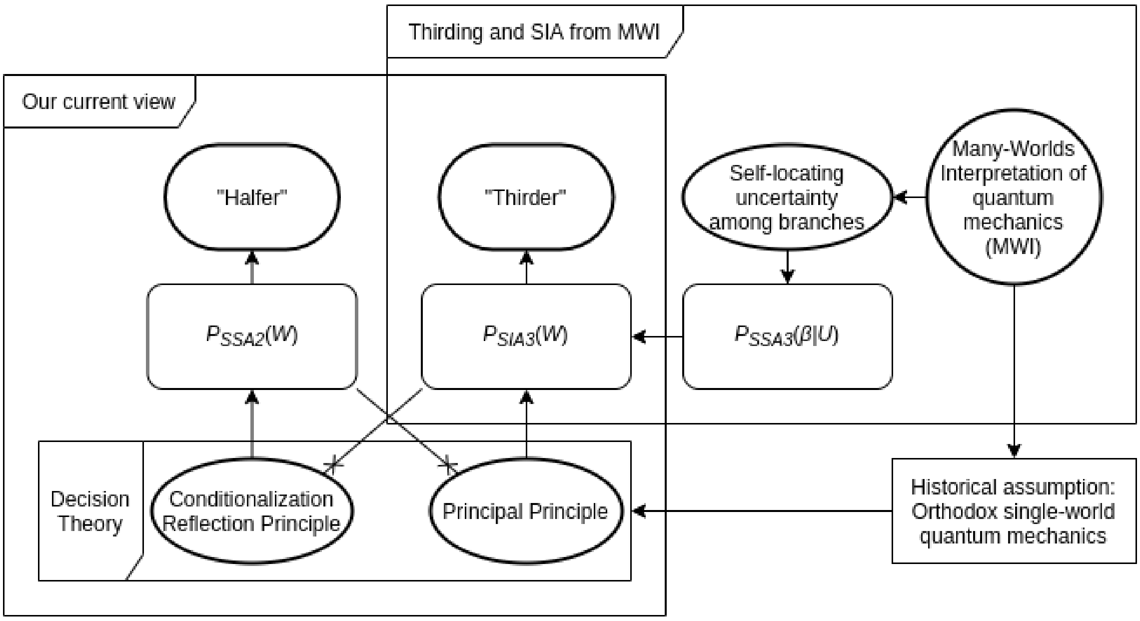


Figure 4.1: Schematic of MWI entailing $P_{SIA3}(W)$ and our current view of incompatible conclusions in SBP and DA, stemming from the historical assumption of a single-world reality. If MWI is correct, the illusion of probability based on self-locating uncertainty among world-branches is the source of the conflicting intuitions in our current view.

4.5 No Chance, all is in Order

In the non-branchy Sleeping Beauty problem, Beauty's credences are prescribed by $P_{SSA2}(U)$ and as in Equation 3.1 her credences after having learned that it is Monday are 1/3 for Odd and 2/3 for Even. These are the credences that lead to the supposedly problematic conclusions in Section 3.1.1 of credences differing from the objective chances of outcomes of future chance events. However, these credences are all in order. In the non-branchy version the future event is not a chance event so it does not result in branching. In this case it is fine for Beauty to have credences that differ from those she had on Sunday after having learned that the current awakening takes place on a Monday. Learning that it is Monday is relevant evidence about how many observer-moments of Sleeping Beauty that exists and how many observer-moments of Beauty that will exist is relevant to the probability of the future non-branchy event.

4.6 The Philosopher's Presumptuous Worldview

Recall the Presumptuous Philosopher (described in Section 3.2.1) and her presumptuous suggestion that a hypothesis on the "theory of everything" receives confirmation just by postulating more observers than competing hypotheses. Assuming that MWI is true and that SSA applies in general, i.e., that $P_{SSA3}(\beta|U)$ applies for classical worlds in the actually existent universe and $P_{SSA2}(U)$ applies to the probability of which universe is the actually existent, then the presumptuous suggestion is disarmed in one of two ways.

If the presumptuous suggestion is that classical worlds with many observers receive automatic confirmation then the suggestion is correct. In the Many-Worlds interpretation physically real observers exist in different classical worlds and worlds with more observers are more likely to contain a particular observer. Under this interpretation the suggestion is not very presumptuous and it shouldn't really claim to be a suggestion about the theory of everything.

If the suggestion is interpreted as being about theories of everything, about which possible universe is actually existent, then the suggestion is false. When evaluating the probability of hypotheses that don't correspond to different world-branches in the multiverse then $P_{SSA2}(U)$ applies. The presumptuous conclusion does not arise because $P_{SSA2}(U)$ does not assign higher credence to hypotheses that postulate more observers.

5. Discussion

5.1 Principal Principle: The Right Route to 1/3

Recall the criticism of the account of Wilson (2014) from Section 4.1:

> First, most thirder arguments do still apply to Non-chancy Sleeping Beauty. Wilson would have to hold that all of these fail, but Elga's [argument from the principal principle] succeeds. Second, Non-chancy Sleeping Beauty seems to show that, intuitively, chancy cases in which the principal principle can be applied should get the same verdict as non-chancy cases in which the principal principle cannot be applied. Wilson has to explain away this intuition. (Bradley, 2015, p. 9)

Wilson (2014) argues that $P_{SIA3}(W)$ is correct when the Principal Principle can be applied but that $P_{SSA2}(W)$ is correct otherwise. Instead of welcoming this small concession to $P_{SSA2}(W)$, his preferred position, Bradley (2015) first criticizes the account on the count that most arguments for $P_{SIA3}(W)$ apply equally well to chancy as to non-chancy versions of the Sleeping Beauty problem. Second, he defends $P_{SSA2}(W)$ by requiring that the intuitive appeal of applying $P_{SSA2}(W)$ in all the circumstances to be explained away.

If the Many-Worlds interpretation is correct, the second concern was countered in Section 4.4 by explaining away the intuitive appeal of $P_{SSA2}(W)$ in branching situations by showing that MWI entails the Thirder-position in these situations by $P_{SSA3}(\beta|U)$. The following section counters the first concern by outlining the most prominent arguments for $P_{SIA3}(W)$, other than the argument from the Principal Principle, and argues for how they are to be resisted. Under the account of the non-chancy and non-branchy Sleeping Beauty in Section 4.1 and Section 4.4 the conclusion of these arguments is correct in the ordinary, chancy or branchy, Sleeping Beauty problem but the arguments overreach as they reach the incorrect conclusion in the non-chancy and the non-branchy versions.

5.2 Overreaching Arguments for Thirding

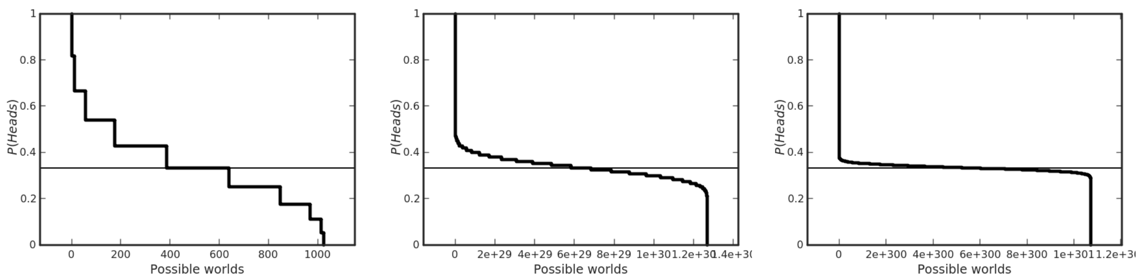
5.2.1 The Long-Run Frequency Argument

One view of probability equates the probability of an event with its relative frequency in a long run of trials. One of the most prominent arguments for $P_{SIA3}(W)$ in the

Sleeping Beauty problem, present already in Elga's (2000) article that popularized the problem, is based on repeating the Sleeping Beauty experiment a large number of times. In this repeated version, Beauty doesn't know in which iteration of the experiment her current awakening takes place and each time she is awakened she is asked what her credence is of the coin having come up Heads. Her credence for the repeated version should be 1/3 for Heads as this is the expected frequency of Heads-awakenings in the long run. The long-run frequency argument in the Sleeping Beauty problem hinges on the assumption that the long-run frequency of Heads- and Tails-awakenings determines Beauty's credence for the one-time case. Since the long-run frequency of Heads is 1/3 and $P_{SIA3}(W)$ prescribes the credence 1/3 in Heads in the one-time case, this assumption entails that $P_{SIA3}(W)$ is the correct probability distribution of worlds.

The long-run frequency definition of probability works well in ordinary cases involving chance because there is only one probability of an event that when repeated yield the long-run frequency observed. This is not true for the Sleeping Beauty problem. We will see that no matter if Beauty's credence in a Heads-awakening in the one-time case is 1/3 or 1/2, the long-run frequency of Heads-awakenings is 1/3.

The generalizations of the Halfer-position and the Thirder-position in the form of $P_{SSA2}(W)$ and $P_{SIA3}(W)$ give the probability of a possible world. In the one-time Sleeping Beauty problem each possible world corresponded to an outcome of the coin toss. In the repeated Sleeping Beauty problem a possible world is a sequence of outcomes of coin tosses. Figure 5.1 shows the probability of a Heads-awakening for all possible sequences of outcomes in three sequences of different lengths.



(a) 10 repetitions **(b)** 100 repetitions **(c)** 1000 repetitions

Figure 5.1: The possible worlds, corresponding to different sequences of outcomes of coin tosses, for 10, 100, and 1000 repetitions and the ratio of Heads-awakenings for each world.

As the length of the sequence approaches the infinite limit, a ratio 1/3 Heads-awakenings is, with probability 1, approached arbitrarily closely (Bostrom, 2007, p. 70). No matter if $P_{SSA2}(W)$ or $P_{SIA3}(W)$ is assumed, the long-run frequency of Heads-awakenings is 1/3. This is the intuitive result of repeating the Sleeping Beauty experiment. Since both $P_{SSA2}(W)$ and $P_{SIA3}(W)$ is compatible with the result, the argument fails to distinguish between $P_{SSA2}(W)$ and $P_{SIA3}(W)$, thus the claim that $P_{SIA3}(W)$ would be established by the argument is false.

Wilson (2014) argues that the long-run frequency argument fails to establish $P_{SIA3}(W)$ for the non-chancy Sleeping Beauty problem because having the experiment repeated with a deterministic event will have the same result every time. This might be so, but if the reasoning above is correct then the long-run frequency argument fails to establish $P_{SIA3}(W)$ under any circumstance, leaving the argument from the Principal Principle to establish it under the circumstances it is applicable.

5.2.2 Dorr's Thirder Argument

	Heads	Tails
Monday	H_1	T_1
Tuesday	H_2	T_2

Figure 5.2: Dorr's Thirder argument. Four symmetrical situations before learning and conditionalizing on it not being Tuesday in the Heads-world.

There is symmetry between the four situations in Figure 5.2. If Beauty could have woken up in any of these situations and, because of a memory-erasure drug, would be uncertain of which, her credence would be 1/4 for each of the situations being her current situation. If she would learn and conditionalize on the fact of whether her current awakening is H_2 or not, then, if it is not, her credence would be 1/3 for each of the remaining situations being her current situation.

Dorr (2002) attempts to show that there is no relevant difference between this scenario and Beauty's situation in the original Sleeping Beauty problem. His argument involves a variant of the Sleeping Beauty problem:

> Again, Sleeping Beauty knows for certain on Sunday that she is to be the subject of an experiment. This time, the experimenters will definitely wake her both on Monday and on Tuesday, administering an amnesia-inducing drug between the two awakenings. However, they have two amnesia-inducing drugs, and they will decide which one to administer by tossing a fair coin on Monday night. If the outcome of the toss is Tails, they will administer the amnesia-inducing drug that was used in the original version of the experiment. If the outcome is Heads, they will administer a much weaker amnesia-inducing drug, which merely *delays* the onset of memories from the previous day, rather than destroying them entirely. If Beauty receives this weaker drug, the first minute of her awakening on Tuesday will be just as it would have been if she had received the stronger drug, but after that the memories of Monday's awakening will come flooding back. She will then realize that it is Tuesday, and that the outcome of the toss must have been Heads. (Dorr, 2002, p. 293)

After the minute has passed and Beauty has not received a flood of memories her credence will be 1/3 in Heads as in the scenario described previously. Not receiving a flood of memories is evidence that can be conditionalized on. Dorr's line of argument is that the conclusion of the argument doesn't change if the delay until the memories come back is decreased and Dorr finds it absurd that a very short delay compared to the memories already being present when waking up would be the difference between Beauty's credence in Heads being 1/3 and it being 1/2.

Bradley (2003), in my view, successfully locates the disanalogy between the two cases. In Dorr's variant, if Beauty learns that it is H_2 (by receiving the flood of memories of her previous day) her credence in Heads increases to 1. In the original Sleeping Beauty problem, Beauty has no possibility of access to any uncentered evidence like this and is thus unable to conditionalize on the lack of such evidence. Bradley concludes that the conclusion in Dorr's variant case has no bearing on the original case.

5.2.3 Technicolor Beauty

Titelbaum (2008) argues that if Beauty is able to distinguish between the different days she will end up with credence 1/3 in Heads. He poses the variant Technicolor Beauty to attempt to demonstrate this.

> *Technicolor Beauty*
> Everything is exactly as in the original Sleeping Beauty Problem, with one addition: Beauty has a friend on the experimental team, and before she falls asleep Sunday night he agrees to do her a favor. While the other experimenters flip their fateful coin, Beauty's friend will go into another room and roll a fair die. (The outcome of the die roll is independent of the outcome of the coin flip.) If the die roll comes out odd, Beauty's friend will place a piece of red paper where Beauty is sure to see it when she awakens Monday morning, then replace it Tuesday morning with a blue paper she is sure to see if she awakens on Tuesday. If the die roll comes out even, the process will be the same, but Beauty will see the blue paper on Monday and the red paper if she awakens on Tuesday.
>
> Certain that her friend will carry out these instructions, Beauty falls asleep Sunday night. Some time later she finds herself awake, uncertain whether it is Monday or Tuesday, but staring at a colored piece of paper. What does ideal rationality require at that moment of Beauty's degree of belief that the coin came up heads? (Titelbaum, 2008, p. 591)

Titelbaum's argument is that gaining the uncentered evidence "There is an awakening on which the red paper is observed" (and correspondingly for the blue paper if the paper Beauty observes is blue) confirms Tails as this evidence is more likely if Beauty wakes both days. Starting out with credences according to $P_{SSA2}(W)$ and conditionalizing on this evidence makes Beauty end up with a credence of 1/3 in Heads.

Bradley (2011a, 2012) argues (successfully, in my view) that being able to distinguish between different days make no difference because the evidence that Beauty's should conditionalize on for it to have bearing on uncentered propositions, her strongest evidence, is the centered proposition "The awakening on which a red paper is observed is *today*". This proposition has the probability 1/2 for $P_{SSA2}(W)$ no matter if the coin toss came up Heads or Tails and thus observing the color of the paper does not confirm Tails over Heads. Figure 5.3 illustrates Bradley's argument.

	World 1	World 2	World 3	World 4
	Heads		Tails	
Monday	Blue	Red	Blue	Red
Tuesday	zzz	zzz	Red	Blue
$P_{SSA2}(W)$ (unnormalized) after waking				
Monday	1	1	1/2	1/2
Tuesday	zzz	zzz	1/2	1/2
Normalized	1/2		1/2	
$P_{SSA2}(W)$ (unnormalized) after observing Red				
Monday	0	1	0	1/2
Tuesday	zzz	zzz	1/2	0
Normalized	1/2		1/2	
$P_{SIA3}(W)$ (unnormalized) after waking				
Monday	1	1	1	1
Tuesday	zzz	zzz	1	1
Normalized	1/3		2/3	
$P_{SIA3}(W)$ (unnormalized) after observing Red				
Monday	0	1	0	1
Tuesday	zzz	zzz	1	0
Normalized	1/3		2/3	

Figure 5.3: The Technicolor Beauty variant of the Sleeping Beauty problem with four uncentered worlds corresponding to Heads and Tails and if Monday was a Red paper day or a Blue paper day. (Unnormalized) occurrence of the events that Beauty wakes up and that a Red paper is observed. Observing information that is irrelevant to the probability of the outcome of the coin toss eliminates centered worlds with equal parts of the total occurrence for both Heads and Tails and thus does not result in a change in credence in Heads and Tails.

5.2.4 Dutch Book Arguments

The subjectivist approach to probability uses betting odds to determine an agent's rational credences (Ramsey, 1926; De Finetti, 1937). Rational credences correspond to those betting odds that make the agent immune to *Dutch books*, series of bets that results in a sure loss for the agent whatever the outcome of the events the bets are on. The possibility that a Dutch book can be made is guaranteed for agents that have credences that violate the axioms of probability and an agent whose credences follow the axioms of probability is guaranteed to be immune to Dutch books.

However, there are several considerations that make Dutch book arguments less than straight-forward in the Sleeping Beauty problem (Bradley and Leitgeb, 2006; Armstrong, 2011; Briggs, 2010; Ross, 2010). One consideration is whether Beauty's winnings and losses are shared between her different temporal parts. Depending on how the problem is interpreted in relation to this consideration the acceptable betting odds are differently defined (Armstrong, 2011).

Another consideration is that whether Beauty is an evidential decision theorist or a causal decision theorist determines whether she needs to adopt $P_{SSA2}(W)$ or $P_{SIA3}(W)$ in order to avoid a Dutch book (Briggs, 2010). The illusion of probability from the Many-Worlds interpretation could potentially shine some light on why this is so. This could be a starting point for further research on the role that uncertainty with and without basis in self-location among world-branches plays in decision theory.

5.3 Comparison to Other Approaches

There have been several approaches to comparing the treatment of probability in the Many-Worlds interpretation and in the Sleeping Beauty problem, owing to the branching structure of experience in common to the two domains.

Section 3.2.1 addressed the issue that $P_{SIA3}(W)$ entails automatic confirmation of theories postulating many observers. The issue is debated in Bradley (2011a, 2015) and Wilson (2014) in which the Many-Worlds interpretation has the role of a hypothesis postulating a large number of observers. Unlike the account in this thesis, that debate does not consider the consequences of the unfamiliar foundation for probability that the Many-Worlds interpretation entails[1].

Another debate is between proponents of the account of (illusion of) probability in MWI (Vaidman, 2012; Groisman et al., 2013; Sebens and Carroll, 2018) from Section 4.3 on one side and Lewis (2007) on the other. Lewis (2007) compares the pre-branching uncertainty about which outcome an observer will experience in a (quantum) chance event in MWI to the structurally similar simplified Sleeping Beauty problem:

> Sleeping Beauty is woken up on Sunday, and told the following: "We will wake you for an hour on Monday, and for an hour on Tuesday, and on Monday night we will administer a drug that will cause you to forget the Monday waking." (Lewis, 2007, p. 1)

Beauty has no pre-branching uncertainty about which day she will experience in this simplified Sleeping Beauty problem, she knows she is awakened on both days with probability 1.

Lewis (2007) argues from the premise that there is no good reason for these two cases to be treated differently and treats the latter kind of branching as another branching resulting from a (quantum) chance event in MWI. This maneuver entails the Halfer-position in the Many-Worlds setting since each of the Tuesday-awakenings are half as likely as the Monday-awakening. Vaidman (2012), Groisman et al. (2013), and Sebens and Carroll (2018) disagree with the premise of Lewis (2007) and argue

[1]Although, Wilson (2013) previously expressed this view: "If [MWI] is correct, then what we ordinarily think of as objective probability is – and has always been – constituted by [the measure of existence of world-branches]. [...] they are a familiar phenomenon under a new and unfamiliar mode of presentation." In fact, the main conclusion of this thesis is entailed by this view combined with the view of Wilson (2014) that the Thirder-position requires that the possibilities under consideration results from a chance process.

(successfully, in my view) that the different treatment between the two cases is motivated by the fact that only the quantum branching involves a change in measure of existence of the observer. They arrive at the Thirder-position by considering the self-locating uncertainty inherent to the unfamiliar foundations of probability in MWI (see Section 4.3).

Groisman et al. (2013) cites Greaves and Myrvold (2010) to support an assertion that this result applies to the Sleeping Beauty problem in single-world theories of quantum mechanics as well. However, the framework outlined in Greaves and Myrvold (2010) aims to show compatibility in the opposite direction, that all the epistemic and decision-theoretic work that chance does in orthodox interpretations of quantum mechanics can be done by the illusion of probability in MWI. That results from analyses that assume MWI are applicable to decision theory in general is too strong of a claim.

5.4 Ethical Implications: Longtermism

While the Doomsday Argument influences our estimation of the likelihood of human extinction scenarios, the research field of Existential Risks, which concerns itself about the long-term survival of humanity (Baum et al., 2019), is, with the notable exception of the early writings of Bostrom (2002a,b), largely silent on it. Another exception is Häggström (2016). After a critical discussion on the Doomsday Argument, he concludes:

> [T]here is a long list of concrete risks, beginning with the most familiar example of nuclear Holocaust, and ending who-knows-where. For each of these risks, there is (in most cases plenty of) scope for discussions both about how large the risks are and what we can do to mitigate them. In contrast, the Doomsday Argument concerns only how large the total risk is (summed across the entire list of concrete risks) and offers no useful idea about how to mitigate the risk. (Häggström, 2016, p. 181)

The fact of whether the Doomsday Argument is sound or not is unhelpful for evaluating individual domains of existential risks; however, getting the conclusion right is important for prioritizing between policy decisions that affect different horizons of the longevity of our civilization. For example, if one accepts the Doomsday Argument and takes it to imply that us colonizing the galaxy is extremely unlikely, then resources that could go towards realizing this (in my view) very beneficial outcome would be put towards more near-term, less beneficial, causes.

It is sometimes said that whether the Many-Worlds interpretation or some other interpretation of quantum mechanics is correct makes no or very little practical difference. However, if the account in this thesis is correct then the Many-Worlds interpretation entails that a human's birth rank does not provide her any evidence that is probabilistically restricting the horizon of the longevity of the human race.

6. Conclusion

There is a feeling of dissonance between the common intuition that the Sleeping Beauty problem can be solved within decision theory and the fact that it seemingly hasn't been solved to satisfaction. The principles of decision theory have proven very useful and coherent so it is expected that applying them straight-forwardly would also solve this problem. However, no consensus on the correct conclusion has yet been reached. In this thesis, the underpinnings of decision theory is questioned and the conclusion is that the illusion of probability that arises in the Everettian Many-Worlds interpretation of quantum mechanics, based on self-locating uncertainty between branches, straight-forwardly resolves the controversial disagreement in the Sleeping Beauty problem and the Doomsday Argument.

In orthodox single-world theories, what are considered uncentered propositions (propositions about what the world is like) are in the Many-Worlds interpretation sometimes uncentered propositions about what the universe (multiverse) is like and sometimes centered propositions about which observer in which world(-branch) one is. The Many-World interpretation entails that the reasoning employed for SSA and the Halfer-position is applicable for uncentered propositions about the multiverse and that the reasoning employed for SIA and the Thirder-position is applicable for centered proposition about which world-branch one is in. Restricting the two opposing positions to the different circumstances happens to avoid the implausible consequences of each position.

Kent (2015), in his criticism of the account of Sebens and Carroll (2018), states that even if prerequisites important for the Many-Worlds interpretation are granted, postulating more than one real world is still unmotivated:

> Moreover, it is worth underlining again here that, if we were able to find reasonably natural postulates that respected physical symmetries and defined an objective branching structure for the universal wave function, it would be superfluous to postulate many independent real worlds. It would be simpler and more natural to postulate that precisely one of the branches is randomly chosen (using the Born weight distribution) and realised in nature. (Kent, 2015, p. 4)

The conclusion in this thesis provides a counter-example. The account of Sebens and Carroll (2018), Vaidman (2012, 2018), and Groisman et al. (2013) entails the virtue of resolving the controversial Sleeping Beauty problem and Doomsday Argument precisely because it is not the case that "one of the branches is randomly chosen and realised in nature" but that all branches are real and the nature of the observers contained in them are qualitatively indistinguishable. The Many-Worlds interpretation entails a piece of internal consistency that is left inconsistent in competing single-world interpretations, a piece of consistency that has proved very elusive. The fact that an account of reality based on the illusion of probability happens to satisfyingly resolve this dividing disagreement makes, at least to me, such an account more satisfying than an account of probability as fundamental.

Bibliography

Armstrong, S. (2011). Anthropic decision theory. *arXiv preprint arXiv:1110.6437*.

Arntzenius, F. (2002). Reflections on sleeping beauty. *Analysis*, 62(1):53–62.

Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy*, 100(7):356–370.

Baum, S. D., Armstrong, S., Ekenstedt, T., Häggström, O., Hanson, R., Kuhlemann, K., Maas, M. M., Miller, J. D., Salmela, M., Sandberg, A., et al. (2019). Long-term trajectories of human civilization. *Foresight*, 21(1):53–83.

Bostrom, N. (2001). The Doomsday Argument Adam & Eve, UN++, and Quantum Joe. *Synthese*, 127(3):359–387.

Bostrom, N. (2002a). *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge.

Bostrom, N. (2002b). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and technology*, 9.

Bostrom, N. (2007). Sleeping Beauty and self-location: A hybrid model. *Synthese*, 157(1):59–78.

Bradley, D. (2003). Sleeping Beauty: a note on Dorr's argument for 1/3. *Analysis*, 63(3):266–268.

Bradley, D. (2011a). Confirmation in a branching world: The Everett interpretation and Sleeping Beauty. *The British journal for the Philosophy of Science*, 62(2):323–342.

Bradley, D. (2011b). Self-location is no problem for conditionalization. *Synthese*, 182(3):393–411.

Bradley, D. (2012). Four problems about self-locating belief. *The Philosophical Review*, 121(2):149–177.

Bradley, D. (2015). Everettian confirmation and Sleeping Beauty: Reply to Wilson. *British Journal for the Philosophy of Science*, 66(3):683–693.

Bradley, D. and Leitgeb, H. (2006). When betting odds and credences come apart: More worries for dutch book arguments. *Analysis*, 66(2):119–127.

Briggs, R. (2010). Putting a value on Beauty. *Oxford studies in epistemology*, 3:3–34.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68.

Dieks, D. (2007). Reasoning about the future: Doom and Beauty. *Synthese*, 156(3):427–439.

Dorr, C. (2002). Sleeping Beauty: in defence of Elga. *Analysis*, 62(4):292–296.

Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60(2):143–147.

Greaves, H. and Myrvold, W. (2010). Everett and evidence. *Many worlds*, pages 264–304.

Groisman, B., Hallakoun, N., and Vaidman, L. (2013). The measure of existence of a quantum world and the Sleeping Beauty problem. *Analysis*, 73(4):695–706.

Häggström, O. (2016). *Here be dragons: Science, technology and the future of humanity*. Oxford University Press.

Jebari, K. (2014). The sleeping beauty problem and the self-indication assumption (SIA). Unpublished Manuscript.

Kent, A. (2015). Does it make sense to speak of self-locating uncertainty in the universal wave function? remarks on Sebens and Carroll. *Foundations of Physics*, 45(2):211–217.

Lewis, D. (1980). A subjectivist's guide to objective chance. In *Ifs*, pages 267–297. Springer.

Lewis, D. (1999). *Papers in metaphysics and epistemology*, volume 2. Cambridge University Press.

Lewis, D. (2001). Sleeping beauty: reply to Elga. *Analysis*, 61(3):171–176.

Lewis, P. J. (2007). Quantum sleeping beauty. *Analysis*, 67(1):59–65.

McQueen, K. J. and Vaidman, L. (2019). In defence of the self-location uncertainty account of probability in the many-worlds interpretation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 66:14–23.

Monton, B. (2002). Sleeping Beauty and the forgetful Bayesian. *Analysis*, 62(1):47–53.

Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and other Logical Essays*, chapter 7, pages 156–198. McMaster University Archive for the History of Economic Thought.

Ross, J. (2010). Sleeping Beauty, countable additivity, and rational dilemmas. *Philosophical Review*, 119(4):411–447.

Sebens, C. T. and Carroll, S. M. (2018). Self-locating uncertainty and the origin of probability in Everettian quantum mechanics. *The British Journal for the Philosophy of Science*, 69(1):25–74.

Tappenden, P. (2011). Evidence and uncertainty in Everett's multiverse. *British Journal for the Philosophy of Science*, 62(1):99–123.

Teller, P. (1973). Conditionalization and observation. *Synthese*, 26(2):218–258.

Titelbaum, M. G. (2008). The relevance of self-locating beliefs. *Philosophical Review*, 117(4):555–606.

Vaidman, L. (1998). On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory. *International studies in the Philosophy of science*, 12(3):245–261.

Vaidman, L. (2012). Probability in the many-worlds interpretation of quantum mechanics. In *Probability in physics*, pages 299–311. Springer.

Vaidman, L. (2018). Many-worlds interpretation of quantum mechanics. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition.

Van Fraassen, C. (1984). Belief and the will. *The Journal of Philosophy*, 81(5):235–256.

Wikipedia contributors (2018). Self-indication assumption doomsday argument rebuttal — Wikipedia, the free encyclopedia. [Online; accessed 24-November-2019].

Wilson, A. (2013). Objective probability in Everettian quantum mechanics. *The British Journal for the Philosophy of Science*, 64(4):709–737.

Wilson, A. (2014). Everettian confirmation and Sleeping Beauty. *The British Journal for the Philosophy of Science*, 65(3):573–598.

DEPARTMENT OF MATHEMATICAL
SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY