SLEEPING BEAUTY: A NEW PROBLEM FOR HALFERS

by

MICHAEL NIELSEN

Under the direction of Andrea Scarantino

ABSTRACT

I argue against the halfer response to the Sleeping Beauty case by presenting a new problem for halfers. When the original Sleeping Beauty case is generalized, it follows from the halfer's key premise that Beauty must update her credence in a fair coin's landing heads in such a way that it becomes arbitrarily close to certainty. This result is clearly absurd. I go on to argue that the halfer's key premise must be rejected on pain of absurdity, leaving the halfer response to the original Sleeping Beauty case unsupported. I consider two ways that halfers might avoid the absurdity without giving up their key premise. Neither way succeeds. My argument lends support to the thirder response, and, in particular, to the idea that agents may be rationally compelled to update their beliefs despite not having learned any new evidence.

INDEX WORDS: Sleeping Beauty, Bayesian epistemology, conditionalization

SLEEPING BEAUTY: A NEW PROBLEM FOR HALFERS

by

MICHAEL NIELSEN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2014

Copyright by
Michael Nielsen
2014

SLEEPING BEAUTY: A NEW PROBLEM FOR HALFERS

by

MICHAEL NIELSEN

Committee Chair: Andrea Scarantino

Committee: Daniel Weiskopf

Neil Van Leeuwen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2014

Acknowledgments

Thanks first and foremost to my advisor, Andrea Scarantino. I count myself very fortunate to have had such a generous and supportive mentor. Thanks also to Dan Weiskopf, Neil van Leeuwen, and George Graham for their valuable comments on this thesis and other work, and for making my time at GSU so fruitful and stimulating. Finally, thanks to my family and my girlfriend, Tawni, for their love and support.

Table of Contents

Acknowledgments **iv**

1 Introduction **1**

2 The Original Sleeping Beauty Case **2**

3 The Thirder Response **4**

3.1 Formal Preliminaries 4

3.2 Elga's Argument 6

3.3 The Variant Sleeping Beauty Argument 6

3.4 The Dutch Book Argument 7

4 The Halfer Response **9**

5 A New Problem for Halfers: An Informal Sketch **10**

6 A New Problem for Halfers: A Formal Presentation **14**

6.1 Formalizing the Halfer Response to the Original Case 14

6.2 Formalizing the New Problem 16

7 Halfer Replies to the New Problem **17**

7.1 Explaining Away the Absurdity: Lewis' Appeal to the Principal Principle . . 17

7.2 Rejecting Conditionalization 21

8 Is the Generalized Sleeping Beauty Case a Problem for the Thirder Response? **24**

8.1 Finite Sleeping Beauty Cases 25

8.2 Infinite Sleeping Beauty Cases 27

9 Conclusion **31**

§1 Introduction

The Sleeping Beauty case, introduced by Elga [8], elicits two incompatible responses: the *halfer response* and the *thirder response*. The thirder response has received the most attention in the large body of literature on Sleeping Beauty, being endorsed by Elga [8], Dorr [6], Monton [25], Hitchcock [14], Weintraub [37], and Horgan [15, 16]. Although the halfer response has been less popular, arguments supporting it can be found in Lewis [22], White [38] , Bradley [2], and Pust [27]. The current status of the debate seems to be that there is no clear consensus on the correct response (Groisman, et al. [10]).

In this thesis, I will argue against the halfer response. After introducing the Sleeping Beauty case and the various arguments supporting the thirder response (§2–§3), I will focus on Lewis' [22] argument for the halfer response (§4). I'll argue that Lewis' key premise, the Halfer Premise, should be rejected by introducing a new problem for halfers (§5–§6): when the original Sleeping Beauty case is generalized, the *Halfer Premise* entails that Beauty must update her confidence in a fair coin's landing heads in such a way that it becomes arbitrarily close to certainty. This is clearly absurd. Thus, the Halfer Premise must be rejected on pain of absurdity. I will consider two halfer replies that don't involve rejecting the Halfer Premise (§7): the first is an elaboration of some cryptic remarks made by Lewis; the second is a rejection of Bayesian conditionalization. I argue that both replies are problematic for halfers. Finally, I will investigate whether the generalized Sleeping Beauty case raises a problem for the thirder response (§8). I argue that it does not. I conclude in §9.

If my argument succeeds, then the halfer response is left unsupported.[1] Not only that, but challenges to the thirder response that make use of the Halfer Premise, such as those found in White [38] and Bradley [2], are also undermined. Thus, although this thesis does

[1]The only other way of supporting the halfer response that I have seen comes from Pust [27]. Pust argues that Reichenbach's [28] principle of direct inference can be used to support the halfer response. The trouble with this way of supporting the halfer response, which Pust acknowledges, is that the principle of direct inference can also be used to support the thirder response, as shown by Seminar [30]. So, since the principle of direct inference does not support the halfer response uniquely, it is not suitable for resolving the Sleeping Beauty problem in favor of the halfer response.

not attempt to advance positive arguments for the thirder response, its conclusion lends thirders some support.

§2 The Original Sleeping Beauty Case

On Sunday, a group of experimenters informs Beauty that she will be participating in the following experiment (we can assume that Beauty fully believes everything that the experimenters tell her). Beauty will be put to sleep on Sunday night and woken up on Monday. At first, the experimenters will not tell Beauty what day it is, but after a few minutes they will tell her that it is Monday and then they will put her to sleep again. After she is put to sleep again on Monday, there are two ways that the experiment can continue: (a) Beauty remains asleep for the rest of the experiment, which lasts until Wednesday; or (b) Beauty is briefly woken up again on Tuesday with her memory of Monday erased (her last memory will be of Sunday night), and then put to sleep for the rest of the experiment, until Wednesday. If (b) occurs, then the Monday and Tuesday awakenings will be completely indistinguishable to Beauty; she will have the same total evidence available to her on both Monday and Tuesday as a result of the memory erasure. The experimenters will decide whether (a) or (b) occurs by flipping a fair coin on Sunday night after Beauty is put to sleep. If the coin lands heads (Heads), then (a) occurs; if it lands tails (Tails), (b) occurs. How confident should Beauty be in Heads

(i) on Sunday after she is apprised of the experimental setup,

(ii) upon first awakening on Monday, and

(iii) on Monday, after the experimenters tell her that it is Monday?

Halfer and thirder responses are typically distinguished by how they answer (ii). Halfers say that Beauty's level of confidence, or credence, in Heads, upon first awakening on Monday, should be 1/2; thirders say it should be 1/3.

Regarding (i), both halfers and thirders agree that Beauty's credence in Heads should be 1/2 on Sunday. This follows from a straightforward application of Lewis' [20] *Principal Principle* (PP). PP is a widely accepted constraint on rational credences which says that an agent's credence in the proposition that some chance event will occur (e.g. a fair coin's landing heads) should be equal to the expected chance of that event occurring. Since it's certain that the chance of a fair coin's landing heads is 1/2, Beauty's credence in Heads should be 1/2 on Sunday.

In answering (ii), halfers are motivated by the fact that Beauty doesn't learn anything that's relevant to Heads between Sunday night and her first awakening on Monday. Since upon first awakening on Monday, Beauty has no new evidence on which to update her Sunday credence in Heads, her credence in Heads should remain 1/2 upon first awakening. Thirders, on the other hand, are motivated by the fact that there are twice as many awakenings associated with a tails coin toss as a heads coin toss. The experiment guarantees that Beauty cannot distinguish between a Monday-heads awakening, a Monday-tails awakening, and Tuesday-tails awakening when she first wakes up on Monday, so she ought to distribute her credence evenly amongst the three awakenings. Only one of the three awakenings is associated with a heads coin toss (viz. Monday-heads), so Beauty's credence in Heads should be 1/3.

It is important to note that the thirder response has a rather surprising consequence: thirders claim that Beauty is rationally compelled to change her credence in Heads from 1/2 to 1/3 between Sunday night and Monday, even though she does not learn any new evidence between those two times. For recall that Beauty is certain of the experimental setup prior to being awakened on Monday—she knows that she will be awoken on Monday, and she knows that she won't be able to distinguish what day it is. A further consequence of the thirder response is that the Sleeping Beauty case provides a counterexample to the well-supported Bayesian norm of belief updating called conditionalization. This is because conditionalization says that an agent is rationally compelled to update her credences just in

case she learns new evidence.[2,3] Let us record these consequences of the thirder response as follows:

Thirder Consequence At t_1, an agent may be rationally compelled to update her credence in A, despite not having learned any new evidence about which she was previously uncertain between t_0 and t_1. Therefore, not all rational belief updating goes by way of conditionalization.

In the next section, I will present arguments that have been given in support of the thirder response.

§3 The Thirder Response

In the previous section, I noted that the thirder response is motivated by considering the frequency of awakenings associated with a heads coin toss. Since this consideration is at odds with the halfer's motivating thought—no belief change without new evidence—thirders have advanced a number of independent arguments to support their position. The aim of this section is to survey these various arguments.

3.1 Formal Preliminaries

The arguments that follow can be stated most clearly by taking a broadly Bayesian approach to rational belief. Let us start, then, by introducing the Bayesian formal apparatus.

[2]On the thirder response, the Sleeping Beauty case also provides a counterexample to van Fraassen's [34, 35] Reflection Principle, which entails that if an agent is certain that she will have a credence x in proposition H tomorrow, and if she will neither gain new information that is relevant to H nor suffer any cognitive mishaps between now and tomorrow, then her credence in H now ought to be x.

[3]Strictly speaking, conditionalization only applies to what Bradley [2] calls "eternal beliefs"—beliefs whose contents have the same truth-value over time—like Heads. Conditionalization does not apply to Bradley's "temporal beliefs," whose contents do not have the same truth-value over time. I will not be making use of this distinction because I only consider conditionalization on eternal beliefs. Also, we could make our informal characterization of conditionalization more precise as follows: conditionalization says that an agent is rationally compelled to update her credences just in case she learns new evidence about which she was previously uncertain. This qualification captures the fact that a Bayesian agent cannot update her beliefs on evidence E if her prior credence in E is 0 (see fn. 5). This qualification holds throughout, so I'll forgo mentioning it for the sake of felicitous prose.

Let P_- represent Beauty's belief state on Sunday night after being apprised of the details of the experiment. P_- is a probability function, or credence function, that maps propositions onto numbers in $[0, 1]$, where 1 represents full confidence, or certainty, in a proposition, and 0 represents certainty in a proposition's negation. Similarly, let P be Beauty's credence function on Monday before being told what day it is, and let P_+ be Beauty's credence function on Monday after being told that it is Monday. I will also sometimes refer to P_+ as Beauty's updated Monday credence.[4]

The Bayesian norm of belief updating is called *conditionalization*. Formally, conditionalization says that if a rational agent has credence function C at time t_0 and credence function C_+ at time t_1, and if the agent learns just E between t_0 and t_1, then the agent's credence in A at t_1 is equal to her conditional credence in A at t_0, given E:

Conditionalization $C_+(A) = C(A|E)$

We will adopt the standard definition of conditional probability which says that the probability of A given E is equal to the probability of A and E divided by the probability of E: $C(A|E) = C(A\&E)/C(E)$.[5]

Finally, let us introduce some useful shorthand for the propositions that Beauty is interested in. Let H_1 be the proposition that the coin landed heads and it is Monday; let T_1 be the proposition that the coin landed tails and it is Monday; and let T_2 be the proposition that the coin landed tails and it is Tuesday. Note that H_1 is equivalent to Heads because H_1 iff Heads. Similarly, $(T_1 \vee T_2)$ is equivalent to Tails because $(T_1 \vee T_2)$ iff Tails.

[4]This nomenclature follows Lewis [22]. The term "updated Monday credence" is my own, however.

[5]Note that conditionalization applies only if the conditional probability $C(A|E)$ is well defined, i.e. only if $C(E) > 0$. This condition holds in all of the cases that I consider here. Also note how the formal definition of conditionalization relates to the informal characterization given above (see fn. 3), according to which conditionalization says that rational agents update their credences just in case they learn evidence about which they were previously uncertain. The qualification that rational agents update on evidence "about which they were previously uncertain" captures the fact that $C(A|E)$ is undefined when $C(E) = 0$ and $C(A|E) = C_+(A)$ when $C(E) = 1$. Thus, according to conditionalization, rational updating occurs when $0 < C(E) < 1$, that is, when the agent's prior credence in E was uncertain (between 0 and 1).

3.2 Elga's Argument

Elga [8] argues for the thirder responsethat $P(H_1) = 1/3$from two premises. The first premise is that, upon first awakening, Beauty's credence in T_1 should be equal to her credence in T_2, on the supposition that the coin landed tails. In other words, if she were to learn that the coin landed tails, Beauty would have no reason to favor a Monday-tails awakening over a Tuesday-tails awakening, and would divide her credence equally between the two possibilities. This gives:

$$P(T_1|T_1 \vee T_2) = P(T_2|T_1 \vee T_2) \tag{E1}$$

The second premise is that Beauty's credence in Heads should be equal to 1/2, on the supposition that it is Monday. Elga argues that it should make no epistemic difference to Beauty whether the coin is tossed before or after her first awakening. Supposing that it's the latter, if Beauty were to learn that it's Monday, her credence in Heads should be equal to her credence in a future fair coin toss landing heads, or 1/2. This gives:

$$P(H_1|H_1 \vee T_1) = 1/2 \tag{E2}$$

Now, from (E1) it follows that $P(T_1) = P(T_2)$, and from (E2) it follows that $P(H_1) = P(T_1)$. So, $P(H_1) = P(T_1) = P(T_2)$. Since the three possibilities are exhaustive, they must sum to 1. Therefore, $P(H_1) = 1/3$. Upon first awakening on Monday, Beauty's credence in Heads should be equal to 1/3.

3.3 The Variant Sleeping Beauty Argument

The next argument we will look at has been presented independently by Dorr [6], Arntzenius [1], and Horgan [15]. It starts by considering a variant of the original Sleeping Beauty case.

Variant Case The experiment is identical to the original case, except for one difference. If the coin lands heads, Beauty will be woken up again on Tuesday and, after a few

moments, will be shown a flashing red light, which indicates that the coin landed heads and it's Tuesday.

Upon first awakening on Monday, Beauty's credence should be divided evenly between four possibilities: $\{H_1, T_1, H_2, T_2\}$, where H_2 is the proposition that the coin landed heads and it's Tuesday. When she doesn't see the flashing red light, Beauty can rule out H_2. But, in ruling out H_2, Beauty hasn't learned anything that supports one of the three remaining possibilities over the others. Hence, Beauty's credence should be evenly divided between $\{H_1, T_1, T_2\}$. Since Beauty's evidential situation is exactly the same in the variant case, after ruling H_2, as it is in the original case, where H_2 is never a possibility, Beauty's credences should be evenly divided in the original case as well. So, in the original case, $P(H_1) = 1/3$.

3.4 The Dutch Book Argument

The last thirder argument we will look at is Hitchcock's [14] Dutch Book argument. A Dutch Book is a book of bets, each of which an agent regards as fair, but that collectively guarantee a loss of funds. Thus, susceptibility to a Dutch Book is often taken to indicate a kind of evaluative incoherence. Hitchcock argues that if Beauty is a halfer, then she is susceptible to a Dutch Book, whereas if she's a thirder, she isn't.

A key assumption behind all Dutch Book arguments is that rational credences can be interpreted as behavioral dispositions to accept and reject bets. Suppose your credence in a fair coin landing heads is 1/2. Then, you should be willing to stake \$1 on a bet that pays at least \$2 if the coin lands heads.[6] In general, where c is your credence in A, you should be willing to stake \$S on a bet that pays at least $\$(S/c)$ if A.

On the assumption that Beauty's credences manifest themselves as dispositions to bet, Hitchcock shows that Beauty is susceptible to a Dutch Book if she is a halfer. On Sunday, Beauty will be willing to stake \$3 on a bet that pays \$6 if Tails because her credence in

[6]Dutch book arguments involve various idealizations. For example, it's assumed that agents only assign utilities to losses and gains of money, that utility assignments are linear, that agents are always disposed to accept fair bets, and so on.

Tails on Sunday is 1/2. Every time she is woken up, Beauty will be willing to stake \$2 on a bet that pays \$4 if Heads because she's a halferupon awakening, her credence in Heads is 1/2. If the coin lands tails, Beauty will accept two such bets because she will be awoken two times; if the coin lands heads she will accept one such bet because she will be awoken only one time. Thus, if the coin lands heads, Beauty will lose the first bet and win the second, netting $\$-1$; if the coin lands tails, Beauty will win the first bet and lose the second and third, netting $\$-1$. Beauty is guaranteed to lose \$1, even though she regards each bet as fair.

Hitchcock goes on to argue that Beauty is not susceptible to a Dutch Book if she's a thirder. On Sunday, Beauty will be willing to stake $\$X/2$ on a bet that pays $\$X$ if Heads because her credence in Heads on Sunday is 1/2. Every time she is woken up, Beauty will be willing to stake $\$Y/3$ on a bet that pays $\$Y$ if Heads because she's a thirder—upon awakening, her credence in Heads is 1/3. As above, if the coin lands tails she will accept two such bets, and if the coin lands heads she will accept one such bet. If the coin lands heads, Beauty will win both the first and second bets, netting $\$(X/2 + 2Y/3)$. If the coin lands tails, Beauty will lose all three bets, netting $\$-(X/2 + 2Y/3)$. The only way to guarantee that Beauty suffers a net loss is to choose X and Y such that both $\$(X/2 + 2Y/3)$ and $\$-(X/2 + 2Y/3)$ are negative. But this is impossible because one term is the negative of the other. So, if Beauty is a thirder, she isn't susceptible to a Dutch Book.

At this point we should pause to observe some general features of the thirder arguments mentioned above. Both Elga's Argument and the Variant Sleeping Beauty Argument provide positive reasons to accept the thirder response, but they don't show where the halfer response goes wrong. Only Hitchcock's Dutch Book Argument provides an argument against the halfer response, since it shows that halfers are susceptible to a Dutch Book. But many philosophers are unmoved by Dutch Book arguments. These philosophers (e.g. Joyce [19]) argue that Dutch Book arguments are not sufficiently epistemic—they don't show that an agent's credences are inaccurate or unsupported by her evidence. Rather, Dutch Book

arguments are merely prudential—they show what credences an agent ought to have in order to avoid bad outcomes, such as a guaranteed loss of money. If this is right, then the Dutch Book Argument does not provide sufficient reason to abandon the halfer response. The Dutch Book Argument merely shows that halfers have a prudential reason to adopt new credences, but it doesn't reveal what's epistemically bad about halfer credences. The argument in this thesis fills a hole in the current literature by showing that the halfer response prescribes credences that are patently absurd, and thus not supported by Beauty's evidence.

§4 The Halfer Response

The arguments surveyed above support the thirder response and entail the surprising Thirder Consequence that Beauty is rationally compelled to change her credence in Heads between Sunday and Monday, despite not having learned any new relevant evidence between those two times. Halfers take the denial of the Thirder Consequence as their starting point (Lewis [22]). It is simply unfathomable, on Lewis' view, that rationality can require belief change in the absence of new relevant evidence. The rejection of the Thirder Consequence gives rise to the following premise.

Halfer Premise Beauty is rationally compelled to update her credence in Heads just in case she learns new evidence that is relevant to Heads.

The Halfer Premise is consistent with conditionalization, the universal validity of which the halfer response is motivated to preserve (Bradley [2]).

Using the Halfer Premise, halfers argue as follows. On Sunday, before the experiment begins, Beauty is certain that the chance of a fair coin's landing heads is 1/2. So on Sunday, Beauty's credence in Heads ought to be 1/2. Both halfers and thirders agree on this. When Beauty wakes up on Monday, she does not learn any new evidence that is relevant to Heads. By the Halfer Premise, her credence in Heads should remain 1/2. Thus, the halfer's answer to (ii) is that, upon first awakening on Monday, Beauty's credence in Heads should be 1/2.

In the next section, I make a first pass at my new problem for halfers by considering how the halfer answers (iii)—i.e. how confident should Beauty be in Heads after the experimenters tell her that it is Monday?

§5 A New Problem for Halfers: An Informal Sketch

Following Lewis, let us extend the halfer response by answering (iii): how confident should Beauty be in Heads after she learns that it is Monday? Since the halfer thinks that rational belief change involves conditionalizing on new evidence, and since "It is Monday" is clearly new evidence, the halfer thinks that Beauty should update her credence in Heads by conditionalizing on "It is Monday".

In order to get an intuitive feel for the halfer response, we can think of conditionalization as a procedure that rational agents follow when they learn some evidence about which they were previously uncertain. In the first step of the procedure, the agent eliminates from her "space of possibilities", or *sample space*, all of the propositions that are inconsistent with her total evidence. Then, she increases her credences in the propositions that are consistent with her total evidence, such that disjoint and jointly exhaustive propositions sum to 1. With this in mind, here is how the halfer envisions Beauty updating her credence in Heads after learning that it is Monday:[7]

"As a halfer I'm committed to my credence in both Heads and Tails being 1/2 upon first awakening.[8] Suppose I were to learn that the coin landed tails. Then there would be two equally likely possibilities: either it's a Monday awakening or it's a Tuesday awakening. Let's call these possibilities *Monday-tails* and *Tuesday-tails*, respectively. Each of these possibilities occupies half of my 'tails sample space.' But my tails sample space is only half of my total sample space. So since Monday-tails and Tuesday-tails are equally likely, they each

[7]Of course, conditionalization does not require that rational agents actually reason in the manner suggested below. Beauty's reasoning is provided merely for illustrative purposes.

[8]This follows from: (i) Heads and Tails are disjoint and jointly exhaustive propositions, (ii) disjoint and jointly exhaustive propositions sum to 1 (by the axioms of the probability calculus), and (iii) according to the halfer response, Beauty's credence in Heads, upon first awakening on Monday, is 1/2.

must occupy one half of one half, or 1/4, of my total sample space. And since Heads occupies one half of my total sample space, and there is only one possible awakening associated with Heads, *Monday-heads*, it follows that Monday-heads occupies 1/2 of my total sample space. Now, I've learned that it's Monday so I can eliminate Tuesday-tails from my sample space, leaving me with Monday-heads and Monday-tails. Based on my previous reasoning, I can see that Monday-heads occupies twice as much of my sample space as Monday-tails, meaning my credence in Monday-heads should be twice my credence in Monday-tails. In order for my credence in Monday-heads to be twice my credence in Monday-tails, and in order for my credences in Monday-heads and Monday-tails to sum to 1, my credence in Monday-heads must now be 2/3, and my credence in Monday-tails must now be 1/3. Finally, since Monday-heads is logically equivalent to Heads, my credence in Heads must now be 2/3." I will subsequently refer to this reasoning as *intuitive halfer reasoning*.

Intuitive halfer reasoning provides the halfers answer to (iii): after Beauty learns that it is Monday, her credence in Heads should be 2/3. This is also Lewis' answer to (iii), as we will see in more detail in the next section. So far, this may not seem obviously problematic, but, when generalized, intuitive halfer reasoning leads to absurdity.

Consider the *generalized Sleeping Beauty case*. The generalized case is identical to the original Sleeping Beauty case except that, in the generalized problem, the experiment continues for n days after the first Monday awakening and there are n number of awakenings associated with a tails coin toss (in the original case $n = 2$). So, if the coin lands tails, Beauty is woken up n times and her memory is erased between each awakening—her last memory upon awakening will always be of Sunday night.

What credence should Beauty assign to Heads upon first awakening on Monday? As in the original problem, Beauty's credence in Heads on Sunday should be 1/2 (by PP), and she learns no new evidence that is relevant to Heads between Sunday and her first awakening on Monday. By the Halfer Premise then, Beauty's credence in Heads, upon first awakening on Monday, should be 1/2.

What credence should Beauty assign to Heads after the experimenters tell her that it is Monday? Here, the halfer will want to apply intuitive halfer reasoning. After learning that it is Monday, Beauty should eliminate all non-Monday awakenings from her sample space, and then increase her credence in Heads and Tails. The problem is that the amount by which Beauty's credences get increased depends on how many awakenings occur in the experiment (that is, it depends on the value of n). When there are only two awakenings ($n = 2$), intuitive halfer reasoning shows that Beauty's credence in Heads changes from 1/2 to 2/3 because her updated credence in Heads must be twice her updated credence in Tails, and the two credences must sum to 1. By the same reasoning, when there are n awakenings, Beauty's updated credence in Heads must be n times greater than her updated credence in Tails. And since the two credences must sum to 1, Beauty's updated credence in Heads must be $n/(n+1)$. (This will be proven formally in the next section.)

This is clearly an absurd result. Suppose that the number of awakenings n is very large, say, 999. Then, according to the halfer, after learning that it is Monday, Beauty's credence in Heads should be 999/1000. Beauty should be nearly *certain* that the coin landed heads, even though her evidence (that it's Monday) is compatible with Tails and does not favor Heads over Tails (since Beauty is always awoken on Monday whether Heads or Tails). In general, the halfer is committed to the following: in experiments with many awakenings, Beauty's credence in Heads should be *arbitrarily close to certainty* after she learns that it is Monday.

That this result is absurd can be illustrated by supposing that Beauty reasons as follows. "When I first woke up, because I'm committed to the Halfer Premise, it was rational for me to have a credence of 1/2 in Heads. But now that I've learned that it's Monday, I'm certain that the coin landed heads, even though the coin's landing tails is compatible with today being Monday!"

Here is another way of highlighting the absurdity. Plausibly, it's rational for Beauty's credence in Heads to correspond to the odds at which she would accept a bet on a fair coin's

landing heads. Suppose that the experiment has $300,000$ awakenings associated with a tails coin toss. Then, after learning that it is Monday, Beauty should be willing to pay $300,000 for a bet that will net her $1 if the coin lands heads. According to the halfer, after learning that it is Monday, it is rational for Beauty to stake her house on a bet that nets just $1 if the coin landed heads!

Although the foregoing considerations are well suited to highlight the absurdity of Beauty's near certainty in Heads, nothing in my argument depends on identifying rational credences with good betting strategies. If it did, then my argument would offer nothing more to the debate than is already offered by Hitchcock's Dutch Book Argument. Thus, it is important to recognize that the real source of the absurdity is not that Beauty's near certainty leads her to make bad bets, but that her credences are not supported by her evidence. The most forceful way to demonstrate this is to consider what I call an *evidentially identical Sleeping Beauty case*. Suppose we vary the original Sleeping Beauty case so that the experiment unfolds without any memory erasures, but is otherwise the same. In this version of the case, when Beauty is awoken on Monday, she will be certain that it's Monday after verifying that her last memory is of Sunday night. Now, how should Beauty answer (ii)? What should her credence in Heads be? In this variation of the case, it's very clear that it should be 1/2: nothing strange is going on with Beauty's memory; she's simply gone to sleep on Sunday, woken up on Monday, and been asked her credence in a fair coin's landing heads. But notice that Beauty's evidential situation in the variation case upon awakening on Monday is identical to her evidential situation in the original case after being told that it's Monday: she's certain that a fair coin has been flipped, she's certain that she'll wake up on Tuesday if and only if Tails, she's certain that today is Monday, and so on. And since it's clear that Beauty's credence in Heads should be 1/2 in the variation case, her near certainty in the generalized case is clearly absurd. Beauty's evidential situation supports a credence of 1/2 in Heads, not near certainty.

§6 A New Problem for Halfers: A Formal Presentation

So far we have been considering the halfer response informally. I have argued that, in the generalized Sleeping Beauty case, the halfer response leads to the absurd result that Beauty should become nearly certain of Heads after she learns that it is Monday. In this section, I want to make my argument more precise by taking a formal approach.

6.1 Formalizing the Halfer Response to the Original Case

I will start by following Lewis' ([22], p. 174) formal presentation of the halfer response. Both halfers and thirders agree that Beauty's credence in Heads on Sunday should be 1/2, so $P_{\text{-}}(Heads) = 1/2$. The Halfer Premise says that Beauty is rationally compelled to update her credence in Heads just in case she gets new evidence that is relevant to Heads. Since Beauty gets no new evidence that is relevant to Heads between Sunday night and first awakening on Monday, her credence in Heads upon first awakening on Monday should be 1/2.

$$P(Heads) = 1/2 = P_{\text{-}}(Heads) \tag{1}$$

And since Beauty is certain that the coin must land either heads or tails, her credence in Heads and her credence in tails should sum to 1: $P(Heads) + P(Tails) = 1$. It follows from (1), by substitution, that Beauty's credence in Tails, upon first awakening on Monday, should be 1/2.

$$P(Tails) = 1/2 = P(Heads) \tag{2}$$

Now, Lewis points out the following consequences of the halfer response. Since Heads is equivalent to H_1, and since Tails is equivalent to $(T_1 \vee T_2)$, it follows that from (2) that $P(Heads) = P(H_1) = 1/2$ and $P(Tails) = P(T_1) + P(T_2) = 1/2$ because the same credence should be assigned to equivalent propositions. By substitution, we get:

$$P(H_1) = P(T_1) + P(T_2) = 1/2 \tag{3}$$

We can also note that $P(T_1) = P(T_2)$. This follows from the fact that, assuming that the experiment is carried out as planned (e.g. Beauty does not die on Monday night), it is guaranteed that if Beauty wakes up on Monday and the coin landed tails (T_1), a Tuesday-tails (T_2) awakening occurs; similarly, if Beauty wakes up on Tuesday and the coin landed tails (T_2), a Monday-tails (T_1) awakening has occurred. T_1 entails T_2, and T_2 entails T_1. Therefore, T_1 and T_2 are equivalent propositions, and equivalent propositions have the same probability. So from (3) we get:

$$P(T_1) = P(T_2) = 1/4 \tag{4}$$

Upon first awakening on Monday, Beauty's credence that it is a Monday-tails awakening should be 1/4, and so should be her credence that it is a Tuesday-tails awakening.

Next, after Beauty is told that it is Monday she should update her credence in Heads by conditionalization. The proposition that it is Monday is equivalent to the proposition $(H_1 \vee T_1)$—it is Monday iff either it is a Monday-heads awakening or it is a Monday-tails awakening—so Beauty can conditionalize on $(H_1 \vee T_1)$ in order to determine her updated Monday credence in Heads. Recall that Beauty's credence function after learning that it is Monday is P_+. By conditionalization then, Beauty's updated Monday credence in Heads is $P_+(Heads) = P(Heads|H_1 \vee T_1)$. We can find the value of the right side of this equation by applying Bayes' theorem and plugging in the values from (2), (3), and (4).[9] This yields:

$$P_+(Heads) = P(Heads|H_1 \vee T_1) = \frac{1/2}{1/2 + 1/4} = 2/3 \tag{5}$$

As we saw with intuitive halfer reasoning, Beauty's credence in Heads should be 2/3 after she learns that it is Monday.

[9]Bayes' theorem is derived from the definition of conditional probability and says $P(A|E) = P(E|A)P(A)/P(E)$

6.2 Formalizing the New Problem

As I argued in §4, the problem for halfers arises when we consider the generalized Sleeping Beauty case with n awakenings. In order to formalize my problem, I will now generalize (3)-(5). We don't need to change (1) and (2) in the generalized case because, as I pointed out in §4, it follows from the Halfer Premise that Beauty's credence in Heads, upon first awakening on Monday, should be 1/2 in both the original and generalized cases.

In the generalized case, let T_n be the proposition that the coin landed tails and it is the nth day of the experiment. Like in the original case, Tails is equivalent to $(T_1 \vee \ldots [lowdots] \vee T_n)$. Since upon first awakening on Monday Beauty's credence in Heads and Tails is 1/2 (2), we can generalize (3) as:

$$P(H_1) = P(T_1) + \ + P(T_n) = 1/2 \tag{3'}$$

which says that Beauty's credence in Monday-heads equals the sum of her credences in each tails awakening, which equals 1/2.

Also, as in the original case where $P(T_1) = P(T_2)$, in the generalized case $P(T_j) = P(T_k)$ for all integers j and k that are between 1 and n. In other words, the probabilities of all the tails awakenings have the same value. From this, we can generalize (4) by substituting from (3'):

$$P(T_j) = 1/2n \text{ where } j \text{ is any integer such that } 1 \leq j \leq n \tag{4'}$$

Beauty's credence in any tails awakening T_j is equal to the inverse of twice the number of awakenings.

After Beauty learns that it is Monday she updates her credence in Heads by conditionalization, just like in the original case. So Beauty's updated Monday credence in Heads is given by $P_+(Heads) = P(Heads|H_1 \vee T_1)$. As in the original case, we can find the value of the right side of this equation by applying Bayes' theorem and plugging in the values from

(2), (3'), and (4'). This yields:

$$P_+(Heads) = P(Heads|H_1 \vee T_1) = \frac{1/2}{1/2 + 1/2n} = \frac{n}{n+1} \tag{5'}$$

This shows that Beauty's updated Monday credence in Heads equals $n/(n+1)$, which I pointed out informally in the previous section. Notice that as n gets very large, $n/(n+1)$ approaches 1. Let us record this as

In Sleeping Beauty cases with a large number of awakenings $n, P_+(Heads) = 1 - \delta$, where δ is an arbitrarily small, positive real number. (6)

In other words, in Sleeping Beauty cases with many awakenings, Beauty's updated Monday credence in Heads should be arbitrarily close to certainty. As I argued above, this result is absurd.

§7 Halfer Replies to the New Problem

What can we conclude from the fact that the halfer response has the absurd consequence stated in (6)? One of the halfer's assumptions must be rejected on pain of absurdity. In the remainder of the paper, I will argue that the Halfer Premise should be rejected. I will make my case by considering the only two ways that the halfer might reply to my problem without rejecting the Halfer Premise. The first reply denies that (6) is absurd after all and tries to explain why it *seems* absurd; the second reply rejects conditionalization, which underlies the emergence of the absurdity. I will argue that neither reply is tenable.

7.1 Explaining Away the Absurdity: Lewis' Appeal to the Principal Principle

Recall that Lewis' [22] halfer response differs from the thirder response in how it answers (iii): in the original case, after learning that it is Monday, Lewis thinks Beauty's credence

in Heads should be 2/3; thirders think it should be 1/2. Lewis defends his answer to (iii) by arguing that the thirder response relies on a misapplication of the Principal Principle (PP) that I introduced in §2.

Earlier I said that PP is a constraint on rational credences that says that an agents credence in the proposition that some event will occur should be equal to the expected chance of that event occurring. This characterization is incomplete, however. The complete definition of PP, as developed by Lewis [20], adds an important proviso:

PP Given that the expected chance of an experiment X (e.g. a coin flip) producing outcome O (e.g. a heads landing) is x, a rational agents credence in the proposition that X produces O is equal to x, provided that the agent has no *inadmissible evidence*.[10]

Obviously, this definition will not be tremendously helpful until we say something about what counts as inadmissible evidence. Lewis [20] himself does not provide a clear definition of inadmissible evidence, but roughly, evidence is inadmissible (relative to an application of PP) if it contains information about the outcome of a chance experiment above and beyond the information given by the chance of that outcome occurring. Inadmissible evidence might provide, for instance, direct information about the future. As an example, consider again a fair coin toss, but this time suppose that an oracle whose predictions you regard as completely reliable tells you that the coin will in fact land tails. Given that you trust the oracle, it would be irrational for your credence in Heads to be equal to the chance of a heads coin toss, or 1/2. Rather, your credence in Heads should be 0 because you are completely confident that the coin will land tails. In other words, it is rational for your credence in Heads to deviate from the chance that the coin will land heads (1/2) because the oracle's testimony provides you with inadmissible evidence.

[10]Formally, PP can be represented as follows:

PP $C(A|T\&E) = x$, where C is a rational agent's credence function, A is the proposition that O is the outcome of experiment X, T is a probabilistic theory which says that the chance of X producing O is x, and E is any admissible evidence that the agent has.

From the probability calculus, the definition in the main text follows: an agent's credence in the proposition that some chance event will occur should be equal to the expected chance of that event occurring.

Lewis uses PP's proviso regarding inadmissible evidence to argue that the thirders answer to (iii) is mistaken. On Lewis' view, when Beauty learns that it is Monday she gets inadmissible evidence, so it is not rational for Beauty to set her updated Monday credence in Heads equal to the chance of a fair coin's landing heads. Therefore, it is not rational for Beauty's updated Monday credence to be equal to 1/2, contra the thirder response.

If Lewis' argument is correct, then it can be used to explain away the absurdity of (6). Although Lewis does not consider the generalized Sleeping Beauty case that I have introduced, his argument can be applied *mutatis mutandis* to the generalized case: after Beauty learns that it is Monday, it is rational for her credence in Heads to deviate from the chance of a heads coin toss (1/2). Thus, although it may seem absurd that Beauty's updated Monday credence approaches certainty in cases with many awakenings, the absurdity is only apparent, and rests on a failure to recognize the fact that Beauty gets inadmissible evidence when she learns that it is Monday.

I will argue against Lewis' claim that Beauty gets inadmissible evidence when she learns that it is Monday. The key here is to recognize that Lewis cannot simply assert that Beauty gets inadmissible evidence when she learns that is Monday. After all, Beauty's learning that it is Monday seems very different from the oracle case in which you learn about the future outcome of coin toss: it is not as if Beauty learns that the coin in fact landed heads. What is needed is some independent motivation for the claim that Beauty gets inadmissible evidence, and Lewis fails to provide this. Lewis offers the following: "when Beauty is told during her Monday awakening that it is Monday . . . she is getting evidence . . . about the future: namely that she is not now in it" ([22], p. 175). But, in general, learning that you are not in the future is not sufficient for getting inadmissible evidence, so Lewis' attempt to motivate his claim fails—or so I'll argue.

Consider the following case.

Dice Case You are participating in an experiment that will last either 30 or 31 minutes depending on the outcome of a fair die roll. If the die shows an even number (Even),

> the experiment will last 30 minutes; if it shows an odd number (Odd), the experiment will last 31 minutes. Suppose the roll occurs during the 29th minute of the experiment. Also suppose that you are apprised of the experimental details prior to beginning the experiment. During the experiment, you have to continuously solve logic puzzles that are presented to you on a computer screen. Since this task is quite engaging, you quickly lose track of the time (you have no watch and there are no clocks in the lab). At some moment late in the experiment, the experimenters ask you for your credence in Even. You reply that it is 1/2, of course, since the chance of a fair die showing an even number is 1/2. At this point, you are uncertain exactly how long you have been solving the logic puzzles, but suppose that you are certain that you have been solving puzzles for at least 20 minutes. Next, the experimenters tell you that you are in the 26th minute of the experiment. How, if at all, should you update your credence in Even?

Intuitively, your credence in Even should still be 1/2. Learning that youre in the 26th minute is irrelevant to the future outcome of a fair die roll (recall that the die will not be rolled until the 29th minute). But if we take Lewis' argument seriously, then this answer cannot be correct. According to Lewis, when you learn that you are in the 26th minute, you get inadmissible evidence about the future, namely that you are not now in it, since before learning this information you thought that you might have been in minutes 27–31. Therefore, on Lewis' account, it is no longer rational for you to set your credence in Even equal to the chance of the die showing an even number. But that seems absurd. The die has not been rolled yet, so why think that your credence in Even should be anything other than the chance that the die shows even (1/2)?

The Dice Case shows that Lewis' claim that Beauty gets inadmissible evidence when she learns that it is Monday is not well motivated. There is no reason to think that, in general, learning that you are not in the future constitutes getting inadmissible evidence. So there is no reason to think that Beauty gets inadmissible evidence when she learns that it is Monday.

Since Beauty does not get inadmissible evidence, PP's proviso does not apply to her updated Monday credence. This means that it is not rational for Beauty's updated Monday credence in Heads to deviate from the chance of a fair coin's landing heads. Therefore, (6) really is absurd: in the generalized case, it is irrational for Beauty's credence in Heads to approach certainty.

7.2 Rejecting Conditionalization

Since the result in (6) is absurd, the halfer must reject at least one of the assumptions leading to this result. Looking back at intuitive halfer reasoning and the derivation of (6), it is clear that there are only two candidates for rejection: the Halfer Premise and conditionalization.[11] In this section, I explore whether rejecting conditionalization is a viable option for the halfer, and I ultimately argue that it is not. Thus, the Halfer Premise should be rejected.

What does the halfer give up by rejecting conditionalization? Here, we can distinguish between a *strong* and a *weak* rejection of conditionalization. On a strong rejection of conditionalization, the halfer asserts that Beauty is not rationally compelled to update her credence in Heads, after learning that it is Monday, in any manner *whatsoever*. In other words, the halfer who makes a strong rejection of conditionalization asserts that, although Beauty gets new information that is relevant to Heads when she learns that it is Monday, she is not rationally compelled to update her credence in Heads, and she may simply maintain a credence in Heads of 1/2. By making a strong rejection of conditionalization, the halfer can avoid the absurd result stated in (6). Because Beauty is never obliged to update her credence with respect to the evidence that it is Monday, she never reaches absurd results by updating. The problem with a strong rejection, however, is that it seems *ad hoc* since we expect a rational agent to update her credences *somehow* after she learns new relevant evidence. It therefore seems more promising for the halfer to make a weak rejection of condi-

[11] The only other assumption that the argument relies on is that rational credences conform to the probability calculus. I take it that halfers will not want to reject this assumption due to the plenitude of arguments that support it. See Hájek [12] for a useful summary and discussion of these arguments. It is also unclear whether an account that rejects this basic assumption would even be recognizable as a halfer response.

tionalization: after learning that it is Monday, Beauty should update her credence in Heads somehow, but not necessarily by using the conditionalization rule—that is, not necessarily by letting her current credence in Heads be equal to her prior conditional credence in Heads, given that it is Monday. At this point, then, the halfer must suggest an alternative rule that Beauty can apply to rationally update her credence in Heads.

But other updating rules that I am familiar with do not offer much help. One alternative is *Jeffrey conditionalization* (JC). Jeffrey [17, 18] proposed JC as a generalization of conditionalization in order to improve upon conditionalization in two ways. First, JC provides a rule for rationally updating one's credences in situations in which one's learning experience does not result in one becoming certain of some proposition. Second, JC can account for scenarios in which an agent learns some proposition E, which she takes to indicate that another proposition, H, is more likely, even though she did not even consider E as a possibility prior to learning it (i.e. E was not in her sample space).[12] It's obvious that this second feature of JC will not help us make sense of the Sleeping Beauty case because, by stipulation, Beauty is certain of how the experiment will proceed. She is certain that when she is woken up on Monday, at first, she will not know what day it is, but later on the experimenters will tell her that it is Monday. This means that when she learns that it is Monday, she is learning something that she considered as a possibility prior to learning it.

Formally, we can represent a simple case of JC as follows. Where C is an agent's credence function prior to changing her credence in E and C_+ is an agent's updated credence function after changing her credence in E, the agent's current credence in a proposition, H, is given by

Jeffrey Conditionalization $C_+(H) = C(H|E)C_+(E) + C(H|\neg E)C + (\neg E)$

which says that the agent's updated credence in H is the sum of two products: her prior conditional credence in H given E multiplied by her updated credence in E, and her prior conditional credence in H given $\neg E$ multiplied by her updated credence in $\neg E$.

[12]See Diaconis & Zabell [7] for a nice discussion of this point.

Now suppose that Beauty updates using JC instead of conditionalization in the original case. After learning that it is Monday, $(H_1 \vee T_1)$, the rational credence for her to assign to Heads is given by JC as

$$P_+(Heads) = \\ P(Heads|H_1 \vee T_1)P_+(H_1 \vee T1) + P(Heads|\neg(H_1 \vee T_1))P_+(\neg(H_1 \vee T_1)) \tag{7}$$

Without going into much detail, I think it's pretty clear that JC will not solve the problem that I have posed for halfers. So far, I have been assuming that Beauty becomes certain that it is Monday after the experimenters tell her so. No important changes to the problem occur if we now explicitly stipulate that Beauty becomes certain that it is Monday after the experimenters tell her so. This means that it is rational for Beauty to assign a credence of 1 to $P_+(H_1 \vee T_1)$ and a credence of 0 to $P_+(\neg(H_1 \vee T_1))$. Now, (7) can be simplified:

$$P_+(Heads) = P(Heads|H_1 \vee T_1) = 2/3 \tag{8}$$

But (8) is identical to (5) in the original case, which has Beauty updating by conditionalization. So we have returned to the problem with which we started. This is to be expected since JC simplifies to conditionalization when an agent becomes certain of some piece of evidence (here, the proposition that it is Monday, $(H_1 \vee T_1)$).

Replacing conditionalization with JC does not help the halfer solve my problem, and (as far as I know) there are no other alternative rules in the offing.[13] I will conclude this section with some general worries about rejecting conditionalization in order to maintain the Halfer Premise.

First, rejecting conditionalization is bound to be an extreme move, and, as such, needs to be supported by *argument*. Why extreme? Because conditionalization accords well with

[13]This is not to say that novel accounts of rational belief change have not been defended in response to the Sleeping Beauty case. They have (e.g. Halpern [13]; Meacham [23]). The point is that these novel accounts have not been advanced in order to vindicate the Halfer Premise, and it is doubtful that they can do so.

common sense, as demonstrated by intuitive halfer reasoning and simple examples involving drawing balls from urns, and it is supported by several arguments. For example, Teller [32] shows that agents who violate conditionalization are susceptible to a Dutch Book, and Greaves & Wallace [9] show that conditionalization uniquely maximizes expected epistemic utility—that is, agents who conditionalize can expect to have credences that are more accurate than agents who don't conditionalize.[14] Since conditionalization is well supported, halfers cannot simply reject it in order to maintain the Halfer Premise. That would be *ad hoc*. Rather, they must, like thirders, have some convincing, independent argument that supports rejecting conditionalization.

Second, rejecting conditionalization puts the halfer in a dialectically awkward position. For recall the conflict between halfers and thirders that I sketched in §2. Thirders endorse the Thirder Consequence, claiming that Beauty is rationally compelled to violate conditionalization upon first awakening on Monday; halfers deny the Thirder Consequence and endorse the Halfer Premise. But if halfers are now prepared to countenance violations of conditionalization *after* Beauty learns that it is Monday, why not also concede that violations of conditionalization are rational when Beauty first wakes up? By countenancing violations of conditionalization at any stage in the Sleeping Beauty case, halfers abandon the primary motivation behind their view.

§8 Is the Generalized Sleeping Beauty case a Problem for the Thirder Response?

At this point one might worry that the generalized Sleeping Beauty case raises a problem for thirders as well as halfers. Recall that thirders are motivated by considering the frequency of awakenings associated with Heads. In the original case ($n = 2$), one out of three awakenings is associated with Heads, so Beauty's credence in Heads should be 1/3 upon first awakening. In the generalized case, one out of n awakenings is associated with Heads, so Beauty's

[14] Other arguments supporting conditionalization can be found in Williams [39] and van Fraassen [36].

credence in Heads should be $1/n$ upon first awakening. $1/n$ approaches 0 for very large n, so, on the thirder response, Beauty should be nearly certain that Tails in the generalized case. Isn't Beauty's near certainty here just as absurd as her near certainty in Heads when she's a halfer?

If it were, then the argument advanced in this thesis wouldnt be able to resolve the debate between halfers and thirders. The generalized case would raise similar problems for both responses. Fortunately, this worry is misplaced. In this section, I'll argue that the generalized Sleeping Beauty case does not raise a problem for the thirder response. Part of the argument has already been provided by Ross [29]. But Ross only considers generalized Sleeping Beauty cases with infinitely many awakenings. So, before recapitulating Rosss argument, I will consider cases with finitely many awakenings.

8.1 Finite Sleeping Beauty Cases

In Sleeping Beauty cases with finitely many awakenings n, the thirder response says that Beauty's credence in Tails, upon first awakening on Monday, should approach certainty as n increases. Contrast this with my new problem for halfers. The halfer response says that Beauty's credence in Heads, *after learning that it's Monday*, should approach certainty as n increases. The thing to notice is that halfers reach an extreme credence (i.e. near certainty) in a different evidential situation than thirders.

My problem for halfers is not merely that there's something absurd about having extreme credences. There isn't. For example, if the experimenters tell Beauty that the coin landed heads, then, assuming Beauty expects the experimenters to be perfectly reliable, it's reasonable for Beauty's credence in Heads to be 1. The problem that I've presented, rather, is that it's absurd for Beauty to have an extreme credence *in her current evidential situation*, namely after learning that it's Monday. As noted above, the reason it's absurd for Beauty to become certain that Heads after learning that it's Monday is that it's being Monday is compatible with a tails coin toss. Beauty knows that Monday awakenings are just as likely

to be associated with Heads as with Tails, so it's absurd for her to be nearly certain that Heads upon learning that it's Monday.

Moreover, one needn't be a thirder to see that the halfer response leads to absurdity. This is illustrated by the evidentially identical Sleeping Beauty case that I presented in §5, which shows that Beauty's evidential situation supports a credence of 1/2 in Heads, not near certainty, and which does not rely on any thirder commitments. I also raised concerns that Beauty's updated Monday credence in Heads licenses patently bad reasoning and unreasonable actions—concerns that are independent of any thirder commitments. For example, Beauty will reason that she should be certain that Heads and that Tails is compatible with it's being Monday; and, after learning that it's Monday, Beauty will stake her house on a bet that nets just $1 if Heads. We might also speculate that Lewis' sense that the halfer response needs to be supplemented by an appeal to PP is indicative of his recognition that there is something odd about Beauty's updated Monday credence in Heads being 2/3. This all goes to show that it should be apparent to halfers, *by their own lights*, that their response leads to absurdity.

The same does not hold for thirders. The thirder might highlight the plausibility of her response to the generalized case by presenting the following analogy. Suppose we have n indistinguishable, red poker chips. The chips correspond to the indistinguishable awakenings that Beauty will experience. Only one awakening is associated with Heads. In the analogy, this corresponds to there being a blue dot on the backside of exactly one poker chip. Now, when Beauty wakes up, she won't be able to distinguish what awakening she is experiencing. This corresponds to looking at all n poker chips, arranged so that only their front sides are visible, and being unable to distinguish which poker chip is the one with a blue dot on its backside. Asking Beauty for her credence in Heads corresponds to asking, for any given poker chip, what credence you should have that it has a blue dot on its backside. In the poker chip analogy, it seems perfectly reasonable to have a very low credence (approaching 0 for large n) that any particular chip has a blue dot. Likewise, in the generalized Sleeping

Beauty case, it seems perfectly reasonable for Beauty to have a very low credence that any particular awakening is associated with Heads, and hence to have a very high credence that any particular awakening is associated with Tails.

The analogy illustrates how the halfer and thirder responses differ when charged with absurdity. Above, I argued that the halfer response leads to absurdity *by the halfer's own lights*. The analogy shows that, *by the thirder's own lights*, there is nothing absurd about the thirder response, even though it leads to extreme credences in the generalized case. Furthermore, there is no evidentially identical Sleeping Beauty case to suggest that the thirder response is not supported by Beauty's evidential situation. Of course, by the halfer's lights, the thirder response may still seem absurd. But to claim that this constitutes a problem for thirders simply begs the question. The point is that, although the thirder response leads to extreme credences, there is no non-question-begging reason to think that these extreme credences are not supported by Beauty's evidence. So, the generalized case with finitely many awakenings does not raise a problem for thirders.

8.2 Infinite Sleeping Beauty Cases

Things are trickier for Sleeping Beauty cases with infinitely many awakenings. In the finite case, the thirder response entails that Beauty's credence in Tails, upon first awakening, is nearly certain for very large n. In the infinite case, the thirder response entails that Beauty be *absolutely* certain that Tails upon first awakening.[15] Now, as I argued in 8.1, there is nothing inherently absurd about having extreme credences. This point applies to the infinite case as well as the finite case. However, Ross [29] demonstrates that the thirder must confront a different issue in the infinite case, namely that the thirder response comes into conflict with the norm of Countable Additivity. In this subsection, I will side with Ross in concluding that this conflict doesn't constitute a problem for thirders.

[15] "Absolutely" because there is no real number $\epsilon > 0$ such that $P(Tails) = 1 - \epsilon$.

Countable Additivity (CA) says the following:[16]

CA For any set of countably many propositions, any two of which are incompatible, rationality requires that one's credences in the propositions in this set sum to one's credence in their disjunction (Ross [29], p. 415). [17,18]

For example, suppose your credence that a Democrat will win the next presidential election is 0.5, and your credence that a Republican will win is 0.49. What credence should you have in the disjunctive proposition that either a Democrat will win or a Republican will win? Since the disjuncts are incompatible (it's not possible that a Democrat *and* a Republican will win), CA applies and says that your credence in the disjunction should be $0.5 + 0.49 = 0.99$.

Now, let's see how the thirder response conflicts with CA in generalized cases with infinitely many awakenings. In the infinite case, Beauty's credence in Tails upon first awakening is 1; that is

$$P(T_1 \vee T_2 \vee \ldots \vee T_n) = 1 \tag{9}$$

She's absolutely certain that Tails, but she doesn't know which tails awakening she's experiencing—it could be any of infinitely many awakenings. What credence should she have in each particular tails awakening $T_1, T_2, \ldots, T_n$? Recall (from 3.2) that, in the original case, $P(H_1) = P(T_1) = P(T_2) = 1/3$. Similarly, in the generalized case

$$P(H_1) = P(T_1) = P(T_2) = \ldots = P(T_n) = 1/m \tag{10}$$

[16]Note that CA applies to probability spaces with infinitely many propositions. For finite spaces, CA can be replaced by Finite Additivity.

[17]More formally:

CA For any countably infinite set of mutually inconsistent propositions $A_i : i = 1, 2, \ldots$, and any credence function C: the probability of the disjunction of all A_i is equal to the sum of the probabilities of each A_i.

$$C(\bigcup_i^{\infty} A_i) = \sum_i^{\infty} C(A_i)$$

[18]You might ask why CA is a *norm*. Stated in terms of probability functions instead of credence functions, CA is an *axiom* of the probability calculus. So, the normative force of CA derives from the thought that credences should be probabilities (see 3.1).

By CA and (9)

$$P(T_1) + P(T_2) + \ldots + P(T_n) = 1 \tag{11}$$

And by (10) and (11)

$$1 = n(1/m) \tag{12}$$

But, since we are considering infinitely many n, there is no m that can satisfy (12).[19] We've derived a contradiction. This shows that CA conflicts with the thirder response in generalized cases with infinitely many awakenings.[20]

We should note that this "problem" for thirders is different in kind from my new problem for halfers. The problem Ive raised for halfers shows that the halfer response leads to an *absurdity*—it is patently irrational for Beauty to become nearly certain that Heads after learning that it's Monday. In deriving a contradiction from CA and the thirder response, however, we have not shown that the thirder response leads to Beauty's having absurd credences. Rather, we have shown that, in infinite cases, the norms governing Beauty's credences conflict. It is an open question whether this sort of conflict is a problem.

The conflict between the thirder response and CA constitutes a problem for thirders only on the assumption that epistemic norms can never make conflicting recommendations (call this situation a *rational dilemma*). I'm inclined to side with Ross in denying this strong assumption:

> Thus, given the various motivations for CA and for the [the thirder response], the conflict between these principles provides some ground for accepting the possibility of rational dilemmas ... In this case, the lesson to be drawn from the conflict between CA and the [thirder response] is that there are contexts in which

[19] Ross notes that allowing for infinitesimal credences does not help resolve the conflict for reasons given by Hájek [11]. Also see Williamson [40].

[20] For ease of exposition, I've omitted an important part of Ross's argument. Strictly speaking, Ross shows that what he calls the Generalized Thirder Principle (GTP) conflicts with CA. He then argues (convincingly, I think) that the various arguments supporting the thirder response (see §3) all entail GTP, and so all conflict with CA. This is equivalent to my more general claim that the thirder response conflicts with CA.

> evidential considerations [the thirder response] and considerations of coherence [CA] pull in opposing directions (Ross [29] pp. 444-445).

On the view that Ross defends, and that I endorse, it is possible for epistemic norms to make conflicting recommendations. The conflict between the thirder response and CA, then, is simply an example of a rational dilemma; it's no problem for thirders.

It is beyond the scope of this thesis to provide original, independent arguments for the possibility of rational dilemmas. It is worth noting, however, that independent arguments in favor of this view are available (see especially Christensen [3, 4, 5]). The upshot of this is that halfers will have a difficult time arguing that the generalized case raises a problem for thirders. First, the argument will have to limit itself to cases with infinitely many awakenings, since cases with finitely many awakenings do not raise a problem for thirders (see 8.1). And second, the argument will have to rely on a highly controversial assumption, namely that epistemic norms can never make conflicting recommendations.

It's worth asking, at this point, whether the problem that I've raised for halfers can also be construed as an instance of conflicting epistemic norms. If it can, then the halfer can defend his position along the lines that I've just sketched for thirders. Unfortunately for halfers, this sort of defense is not possible: my new problem for halfers cannot be construed as a rational dilemma. To see this, we need only note that the halfer response does not violate any of the epistemic norms to which halfers are explicitly committed. Indeed, the very motivation behind the halfer response is to answer the Sleeping Beauty case in a way that is consistent with the norms associated with orthodox Bayesianism, such as conditionalization and PP. The problem for halfers is that they prescribe absurd credences *despite* the fact that their response does not violate any epistemic norms. That's why the halfer premise should be rejected. Note also that the halfer cannot construe my new problem as a conflict between the halfer response and CA. This is because my problem for halfers does not depend on considering cases with infinitely many awakenings; and in cases with finitely many awakenings, CA will never be violated (that is, we cannot derive a contradiction along

the lines of (9)-(12) if we limit ourselves to cases with finitely many awakenings). As the evidentially identical Sleeping Beauty case helps to illustrate, it's already absurd for halfers to be *nearly* certain that Heads in finite cases. I needn't introduce infinite cases to raise my problem, and this denies the halfer the opportunity to construe the problem with his response as a conflict with CA.

§9 Conclusion

The structure of my argument has been as follows. First, I introduced the original Sleeping Beauty case and distinguished between halfer and thirder responses to it. The distinctive feature of the halfer response is its use of the Halfer Premise. Unlike thirders, halfers think that Beauty is rationally compelled to update her credence in Heads just in case she learns new evidence that is relevant to Heads. I have argued that the Halfer Premise should be rejected because, in the generalized Sleeping Beauty case, the halfer response says that Beauty should become nearly certain of Heads after she learns that it's Monday, which is absurd. I defended my claim that this result is absurd by elaborating on and arguing against Lewis' remarks to the contrary. I also explored whether the halfer might reject conditionalization instead of the Halfer Premise, and concluded that this approach is problematic for halfers. Finally, I argued that the generalized case is problematic for halfers only; it does not raise problems for the thirder response.

Although I have not advanced any positive arguments for the thirder response, my argument lends some support to thirders. Since the Halfer Premise should be rejected, we can embrace the surprising Thirder Consequence, namely that it is rational for Beauty's credence in Heads to change between Sunday night and Monday, even though she does not learn any evidence that is relevant to Heads between those two times.

References

[1] Arntzenius, F. (2003). Some problems for conditionalization and reflection. *The Journal of Philosophy*, 356-370.

[2] Bradley, D. J. (2011). Self-location is no problem for conditionalization. *Synthese*, 182(3), 393-411.

[3] Christensen, D. (2007). 1. Does Murphys Law apply in epistemology? Self-Doubt and Rational Ideals. *Oxford studies in epistemology*, 1.

[4] Christensen, D. (2010a). Higher-order evidence. *Philosophy and Phenomenological Research*, 81(1), 185-215.

[5] Christensen, D. (2010b). Rational reflection, *Philosophical Perspectives*, 24(1), 121-140.

[6] Dorr, C. (2002). Sleeping Beauty: in defence of Elga. *Analysis*, 62(276), 292-296.

[7] Diaconis, P., & Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association*, 77(380), 822-830.

[8] Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis*, 60(266), 143-147.

[9] Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459), 607-632.

[10] Groisman, B., Hallakoun, N. A., & Vaidman, L. (2013). The measure of existence of a quantum world and the Sleeping Beauty Problem. *Analysis*, 73(4), 695-706.

[11] Hájek, A. (2003). What conditional probability could not be. *Synthese*, 137(3), 273-323.

[12] Hájek, A. (2008). Arguments for—or against—Probabilism?. *The British Journal for the Philosophy of Science*, 59(4), 793-819.

[13] Halpern, J. Y. (2004). Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. In *KR* (pp. 12-22).

[14] Hitchcock, C. (2004). Beauty and the bets. *Synthese*, 139(3), 405-420.

[15] Horgan, T. (2004). Sleeping Beauty awakened: New odds at the dawn of the new day. *Analysis*, 64(281), 10-21.

[16] Horgan, T. (2007). Synchronic Bayesian updating and the generalized Sleeping Beauty problem. *Analysis*, 67(293), 50-59.

[17] Jeffrey, R. C. (1990). *The Logic of Decision*. University of Chicago Press.

[18] Jeffrey, R. C. (Ed.). (1992). Probability and the Art of Judgment. Cambridge University Press.

[19] Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575-603.

[20] Lewis, D. (1980). A subjectivist's guide to objective chance. *Studies in Inductive Logic and Probability*, 2, 263.

[21] Lewis, D. (1994). Humean supervenience debugged. *Mind*, 103(412), 473-490.

[22] Lewis, D. (2001). Sleeping beauty: reply to Elga. *Analysis*, 61(271), 171-76.

[23] Meacham, C. J. (2008). Sleeping beauty and the dynamics of de se beliefs. *Philosophical Studies*, 138(2), 245-269.

[24] Meacham, C. J. (2010). Two mistakes regarding the principal principle. *The British Journal for the Philosophy of Science*, 61(2), 407-431.

[25] Monton, B. (2002). Sleeping Beauty and the forgetful Bayesian. *Analysis*, 62(1), 47-53.

[26] Priest, G. (2002). Rational dilemmas. *Analysis*, 62(1), 11-16.

[27] Pust, J. (2011). Sleeping Beauty and direct inference. *Analysis*, 71(2), 290-293.

[28] Reichenbach, H. (1971). *The Theory of Probability*. University of California Press.

[29] Ross, J. (2010). Sleeping Beauty, countable additivity, and rational dilemmas. *Philosophical Review*, 119(4), 411-447.

[30] Seminar, O. (2008). An objectivist argument for thirdism. *Analysis*, 68(298), 149-155.

[31] Strevens, M. (1995). A closer look at the new principle. *The British Journal for the Philosophy of Science*, 46(4), 545-561.

[32] Teller, P. (1973). Conditionalization and observation. *Synthese*, 26(2), 218-258.

[33] Thau, M. (1994). Undermining and admissibility. *Mind*, 103(412), 491-504.

[34] van Fraassen, C. (1984). Belief and the will. *The Journal of Philosophy*, 81(5), 235-256.

[35] van Fraassen, B. C. (1995). Belief and the problem of Ulysses and the Sirens. *Philosophical Studies*, 77(1), 7-37.

[36] van Fraassen, B. C. (1999). Conditionalization, a new argument for. *Topo*i, 18(2), 93-96.

[37] Weintraub, R. (2004). Sleeping Beauty: a simple solution. *Analysis*, 64(1), 8-10.

[38] White, R. (2006). The generalized Sleeping Beauty problem: a challenge for thirders. *Analysis*, 66(290), 114-119.

[39] Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *The British Journal for the Philosophy of Science*, 31(2), 131-144.

[40] Williamson, T. (2007). How probable is an infinite sequence of heads?. *Analysis*, 67(3), 173-180.