CHRISTOPHER HITCHCOCK

BEAUTY AND THE BETS

ABSTRACT. In the Sleeping Beauty problem, Beauty is uncertain whether the outcome of a certain coin toss was heads or tails. One argument suggests that her degree of belief in heads should be 1/3, while a second suggests that it should be 1/2. *Prima facie*, the argument for 1/2 appears to be stronger. I offer a diachronic Dutch Book argument in favor of 1/3. Even for those who are not routinely persuaded by diachronic Dutch Book arguments, this one has some important morals.

1. THE PROBLEM

Beauty is about to sleep for a long time. Not one hundred years or anything like that: two days will suffice. During that time she will be awakened briefly, either once (on Monday) or twice (on Monday and Tuesday). For definiteness, let us say that Beauty goes to sleep at 12:00 A.M. (midnight) on Monday morning, and sleeps until 12:00 A.M. Wednesday morning; the awakening(s) will take place at noon on the appropriate day(s). The number of awakenings depends upon the toss of a fair coin: if the result is heads, she is awakened but once; if tails, twice. The nature of her sleep is such that she will not remember being awake. In particular, when she is awakened, she will not know whether it is Monday or Tuesday. Upon awakening on Monday,[1] what should be her degree of belief in H, the proposition that the coin landed heads?

This problem was first posed in print by Adam Elga (Elga 2000), who attributes it to Robert Stalnaker. Related problems are presented in Aumann et al. (1997) and Piccione and Rubinstein (1997). The problem is interesting, because it involves an unusual form of reasoning under uncertainty. Beauty is uncertain, not merely about the outcome of the coin toss, but also about what day it is. In the language of possible worlds: Beauty is uncertain, not merely about which world she is in, but about where she currently is within the world. If we wish to represent Beauty's degrees of belief formally, it will not suffice to use a probability measure over possible worlds. Rather, we will need to use a probability measure over a space of what Quine (1969) calls 'centered worlds', ordered pairs

Synthese **139:** 405–420, 2004.

consisting of possible worlds and (spatio-)temporal locations within those worlds.

2. THE ARGUMENT FOR 1/3

If we imagine this scenario repeated a great many times, we would expect approximately half the coin tosses to result in heads, and half to result in tails. For each toss that results in heads, Beauty will experience one awakening; for each toss that results in tails, she will experience two awakenings. Therefore, she will experience two awakenings in which the most recent toss was tails for every awakening in which the most recent toss was heads. That is, in a repeated sequence of two-day slumbers, the relative frequency of heads-awakenings will be one-third. Since the heads-awakenings and tails-awakenings are qualitatively identical, her degree of belief in heads on a given awakening ought to be 1/3.

3. THE ARGUMENT FOR 1/2

Since Beauty knows that the coin is fair, her prior probability for H is 1/2. (We assume she obeys Lewis's 'Principal Principle' (1980).) She knows that she will be awakened from her sleep at least once. Therefore, upon awakening, she has gained no new information. So her degree of belief in H should remain at 1/2.

4. PRIMA FACIE

The case for 1/3 does not look very strong. Long-run relative frequencies are reliable guides to single case probabilities only when the individual trials are independent. In the Sleeping Beauty problem, the 'trials' are the individual awakenings (not runs of the whole experiment). Thus the trials are manifestly *not* independent: if the first awakening is a tails-awakening then the second awakening will also be a tails-awakening.

By contrast, the case for 1/2 appears to rest on well-grounded principles of probabilistic reasoning. The rule of updating by conditionalization says that if one's prior degree of belief in heads is $P(H)$, then, upon learning E, one should update one's degree of belief to $P(H|E)$. Violation of this rule – at any rate, deliberate, pre-meditated violation of this rule – makes one susceptible to a diachronic Dutch Book, or DDB for short (Teller 1973).[2] It follows that if one has not learned anything, then one's degrees of belief should not change. The claim that Beauty should not change her mind if she has learned nothing new is not supported merely by intuition: it follows from a very general principle about belief revision. A related rule governing probabilistic reasoning is the 'Reflection Principle' (van Fraassen 1984):

$$P_t(H|p(t', H, r)) = r$$

where P_t is the agent's degree of belief at time t, and $p(t', H, r)$ is the proposition that one's degree of belief in H at time $t' \geq t$ will be r. The Reflection Principle can also be justified by a DDB. The Reflection Principle clearly supports the answer of 1/2. For suppose that Beauty's degree of belief upon awakening will be 1/3. This new degree of belief does not depend upon her learning anything that she does not already know Sunday night (before going to sleep). So Beauty knows full well that her degree of belief will be 1/3, that is, she knows that $p(t', H, 1/3)$ is true where t' is Monday at noon. But surely her degree of belief in H should be 1/2 *before* going to sleep – after all, the coin *is* known to be fair. So if 1/3 is the correct answer to our puzzle, Beauty's degree of belief at time t on Sunday evening will be of the form $P_t(H|p(t', H, 1/3)) = 1/2$, in violation of Reflection.

If awakening with a degree of belief in heads equal to 1/3 violates these principles, and these principles are supported by DDB arguments, then it ought to be possible to show directly that waking up with that degree of belief renders Beauty susceptible to a DDB. And indeed, this appears to be the case. Before going to sleep, the bookie sells to Beauty bet #1 on H; the bet pays $30 if the result of the coin toss is heads, nothing otherwise; and the bet costs $15. (I assume throughout that Beauty's utility is linear in dollars.) Since Beauty's degree of belief in H before going to sleep is 1/2, she will find this bet fair. Upon awakening, her degree of belief in H is 1/3 (by hypothesis), so the bookie sells her bet #2, on tails, with a payoff of $30, for a cost of $20. Whatever the result of the coin toss, Beauty will win one bet, for a payoff of $30, while paying out a total of $35 for the privilege of gambling. This book of bets is shown in Table I, where the entries in the 'heads' and 'tails' columns reflect Beauty's net gain or net loss on that outcome. A similar argument could be constructed for any degree of belief different from 1/2.

All in all, then, the case for 1/2 appears to be much stronger than the case for 1/3. But the careful reader will have noted the frequent use of the word 'appears'. We shall see in the sequel that there is an important flaw in the Dutch Book argument of the previous paragraph. But first, I present a further argument in favor of 1/3.

TABLE I

Bets	Payoff	Cost	Heads	Tails
bet #1	\$30 if heads	\$15	\$15	−\$15
bet #2	\$30 if tails	\$20	−\$20	\$10
Combined	\$30	\$35	−\$5	−\$5

5. A SYMMETRY ARGUMENT FOR 1/3

Elga (2000) offers a symmetry argument favoring the answer 1/3. In this section I present a more formal argument for 1/3 that incorporates the same symmetry assumptions. This argument, I think, is highly *suggestive*, but not fully *convincing*. The argument is not fully convincing because several of the assumptions are difficult to construe, and thus hard to assess for plausibility. Indeed, our difficulty in construing these assumptions will serve to underline just how puzzling the Sleeping Beauty problem really is.

We begin by constructing a probability measure representing Beauty's degrees of belief. Since Beauty is uncertain not only about the state of the world that she is in, but also about her location within it, the probability function that represents her degrees of belief will have to be defined over sets of centered worlds. For purposes of solving our problem, it suffices that each centered world specify three things: whether the coin toss landed heads or tails; the day of the week; and whether Beauty has just been awakened from a deep sleep. So we can let our elementary centered propositions be of the form $\langle D, O, B\rangle$, where D is the day of the week (*Sun* or *Mon* or ... or *Sat*), O is the outcome of the coin toss (H or T), and B is Beauty's present status (just awakened, A, or other, $\sim A$). Thus, for example, the centered proposition $\langle Mon, H, A\rangle$ asserts that today is Monday, the coin lands heads, and Beauty has just been awakened from a deep sleep. We will write disjunctions of these basic propositions in the natural way: $\langle Mon, H\rangle \equiv \langle Mon, H, A\rangle$ or $\langle Mon, H, \sim A\rangle$; $H \equiv \langle Sun, H\rangle$ or $\langle Mon, H\rangle$ or ... or $\langle Sat, H\rangle$; and so on.

What should Beauty's degrees of belief over these centered propositions look like? Since she knows the coin toss to be fair, her 'prior' probability for H should be $P(H) = 1/2$. 'Prior' means in the absence of any further information, before she has conditionalized on any other propositions. Since this implies the absence of any *temporal* information, 'timeless' might be a better word than 'prior'. Since the outcome of the coin toss is unaffected by the passage of time, her degrees of belief in the outcome of the coin toss should be independent of her degrees of belief regarding the day of the week. Thus she should have $P(\langle H, Mon\rangle) = P(\langle T, Mon\rangle)$, and likewise for the other days of the week.

Now what about Beauty's 'priors' of the form $P(Mon)$? Here the word 'prior' is even more suspect: immediately 'prior' to the start of a new week she should assign $P(Sat) = 1$ (assuming we adopt the religion-biased convention that the week starts on Sunday). But clearly this is not what we have in mind. What we *do* have in mind is something more like this: Suppose that Beauty were to wake up after a *very* long sleep, having no idea how long she slept, nor indeed any idea what day it was when she went to sleep. Then what would her degrees of belief be in the propositions *Sun*, *Mon*, etc? But even this suggests the *conditional* probabilities $P(Sun|A)$, $P(Mon|A)$, etc., rather than the unconditional probabilities $P(Sun)$, $P(Mon)$, etc. Really, we just want Beauty's degrees of belief in the absence of any information about what day it is. And at this point, it is natural to appeal to symmetry: in the absence of any information favoring one day of the week over the others, she should assign $P(Sun) = P(Mon) = \ldots = 1/7$. The reader will be excused for finding these probabilities 'funny'; we shall see that their funniness makes it easier rather than harder to justify the symmetry assumption.

Finally, we want to give Beauty conditional credences of the form $P(A|Mon\&T)$, reflecting her knowledge of the protocol that determines her schedule of awakenings. A natural assignment would be $P(A|Mon\&T) = P(A|Mon\&H) = P(A|Tues\&T) = 1$; $P(A|Tues\&H) = P(A|Wed\&T) = \ldots = P(A|Sun\&H) = 0$. Note that A represents awakening from a deep sleep of the sort she enters on Monday at midnight. We assume that she sleeps normally the rest of the week, and that her normal awakenings are qualitatively distinguishable from her deep sleep awakenings. We might balk at the assignments $P(A|Mon\&T) = P(A|Mon\&H) = P(A|Tues\&T) = 1$; after all, there are many times during those days when she has not just been awakened. Very well then, let us have $P(A|Mon\&T) = P(A|Mon\&H) = P(A|Tues\&T) = p$, $0 < p \leq 1$. The important thing is that these probabilities are greater than zero, and that they are equal. She will definitely be woken up on each of these scenarios, and she will be awake equally often and for an equal duration in each. Nonetheless, there is still something 'funny' about all of these probabilities. Although there is a positive probability of $\sim A$ occurring on Tuesday (and on Monday as well, if $p < 1$), Beauty can never learn of $\sim A$'s being true in this scenario. Put another way, A receives a probability of less than 1, but it is, to Beauty, unfalsifiable.

Upon awakening, Beauty conditionalizes upon A. Using Bayes' theorem, Beauty's posterior probabilities can be computed from the probabilities specified above: $P(Mon\&T|A) = P(Mon\&H|A) = P(Tues\&T|A) = 1/3$. From this it follows that $P(H|A) = 1/3$. Beauty's degree of belief upon wakening should be 1/3. *QED*

One way in which a 'halfer' might respond would be by challenging the independence assumption. He might reason as follows: suppose that upon awakening, Beauty were told that it was Monday rather than Tuesday; wouldn't that lead her to raise her degree of belief that the coin landed heads? In order to reflect the evidential relevance of the day for the outcome of the coin toss, Beauty's 'priors' should be such that Mon and H are positively correlated. Lewis (2001) offers an argument along these lines. In order to arrive at the answer 1/2, however, it is not enough to simply reject the independence assumption: each day of the week would have to be evidentially relevant to H in just the right way. For example, if Beauty's 'priors' are $P(Mon\&H) = 2/21$, $P(Mon\&T) = 1/21$, $P(Tues\&H) = 2/21$, $P(Tues\&T) = 1/21$, then, upon learning that A, her posterior probability for H will be equal to 1/2 as desired. (This would also give us $P(H|Mon\&A) = 2/3$, as suggested by Lewis (2001)). So far, so good. But now let us suppose that her probability is 1/14 for every other centered proposition of the form $\langle D, O\rangle$. Then, adding over the all days, her total 'prior' probability for heads would be 23/42 $\neq$ 1/2. In order for Beauty's 'prior' $P(H)$ to be 1/2, she would also have to have $P(D\&H) = 13/210$, $P(D\&T) = 17/210$, for every other day of the week D. It is very hard to see how these degrees of belief could be independently motivated. Indeed, it is easy to show that there is no probability distribution over the defined space that satisfies all of the following conditions: (i) $P(H) = 1/2$, (ii) $P(H|A) = 1/2$, (iii) $P(Mon) = P(Tues)$, (iv) $P(Mon, T) > 0$, and (v) $P(D, H) = P(D|T)$ for every D other than Monday or Tuesday.[3] The first assumption is a constraint of the problem, and the second is the halfer's desired answer; thus the halfer must reject one of (iii)–(v), and it is hard to see any non-*ad hoc* motivation for doing so.

How did the halfer get into this mess? A fundamental mistake was made at the very first step. It is indeed true that if, *upon awakening*, Beauty were told that it was Monday rather than Tuesday, that would lead her to raise her degree of belief that the coin landed heads. From this it follows that Mon is evidentially relevant to H, *given* A. This relevance is reflected in the original probability measure constructed above: $P(H|Mon\&A) = 1/2 > 1/3 = P(H|A)$. But it is a mistake to infer from this that Mon is *unconditionally* relevant to H.[4,5] Suppose, for example, that the nature of Beauty's sleep is such that she can still receive information and reason probabilistically while sleeping. If she were given the information that it is Monday while sleeping, this should not affect her subjective probability that the coin landed heads.

If we accept the symmetry argument for 1/3 developed in this section, then we can explain what went wrong with the argument for 1/2 in Section 3. That argument wrongly claims that Beauty learns nothing new. In fact, Beauty does 'learn' A, that she has just awakened from a deep sleep. The word 'learn' is in scare quotes, because what Beauty 'learns' is not an ordinary proposition, but rather a centered proposition. Let W be the proposition that Beauty (tenselessly) awakes from a deep sleep. That is, W is the proposition that is true of every world in which Beauty awakes from a deep sleep at least once. Beauty did not learn *this* proposition – she knew this all along. But it does not follow from this that her degrees of belief did not change.

The distinction between learning a proposition and 'learning' a centered proposition does raise a serious worry, however. There are well established DDB arguments showing that one should accommodate newly learned propositions by conditionalizing upon them. In the preceding argument, however, we have assumed that one should also conditionalize upon newly 'learned' centered propositions.[6] To illustrate just how problematic this assumption is, note that before going to sleep, Beauty believes with certainty that the centered proposition 'today is Sunday' is true. Upon awakening, she believes with equal certainty that this centered proposition is false. This sort of complete turnabout is something that can't be accomplished under normal conditionalization. It is simply not clear how our ordinary canons of rational belief formation are to be extended to our beliefs about our *location* within the world. Thus, while I think that the foregoing argument *suggests* that the answer to the Sleeping Beauty puzzle is 1/3, it is also serves to make clear *why* that puzzle is so perplexing. I will attempt to bolster the case for 1/3 by developing a novel argument in the next section.

6. THE DUTCH BOOK ARGUMENT FOR 1/3

As we saw in Section 4 above, there is an apparent Dutch Book argument that upon awakening, Beauty's degree of belief in heads should be 1/2. That argument is mistaken. An essential constraint on Dutch Book arguments is that the bookie not exploit any information that is not available to the agent being booked. It is not an agent's susceptibility to a sure loss *per se* that renders the agent's degrees of belief incoherent. For one thing, she could protect herself against Dutch Books by abstaining from gambling.

Rather, susceptibility to Dutch Book is symptomatic of an underlying evaluative inconsistency.[7] On the one hand, the agent views a book of bets as fair – she judges each individual bet as yielding no expected loss or gain for either side. On the other hand, she views the book of bets as unfair – she can determine that a loss is inevitable using purely deductive reasoning, which does not presuppose probabilistic coherence. In the Dutch Book arguments, the bookie is a colorful device for bringing out this second evaluation. But the bookie's certain gain is not an appropriate stand-in for the agent's second evaluation if it depends upon information unavailable to the agent. If the bookie can achieve his certain gain only by exploiting information that is unavailable to the agent, then the Dutch Book reflects an evaluation of the system of bets that is not the agent's own.

In the Sleeping Beauty Puzzle, the bookie cannot be allowed to know the outcome of the coin toss. Moreover, when he sells the second bet to Beauty upon awakening, *he cannot be allowed to know what day it is*. If Beauty knew what day it were upon awakening, that would be relevant to her degree of belief in heads; if she knew it were Tuesday, for example, her degree of belief in heads would be zero. Thus the bookie cannot know the day of the week without being in possession of relevant information that is unavailable to Beauty. But now let us ask, how can the bookie arrange to sell the second bet to Beauty without violating this restriction? That is, how can the bookie formulate an appropriate *Dutch Strategy*? He cannot simply plan to show up on Monday at noon, for then he will be selling the bet in the full knowledge that it's Monday. Nor can he arrange to go on either Monday or Tuesday, selected at random, while somehow remaining ignorant of what day it is. In this case, he risks showing up on a Tuesday when the outcome of the coin toss was tails, in which case he loses the first bet and never gets to sell the second.

There is one way in which the bookie can ensure that he has no information that is unavailable to Beauty: he can sleep with her. That is, he can place his first bet, go into a deep sleep when Beauty does, arrange to have himself awakened under the same protocol as Beauty, and sell a follow-up bet to Beauty whenever they wake up together. The bookie, like Beauty, will awaken having no idea whether it is the first or second awakening, having no idea whether an initial follow-up bet has already been placed. Thus he must sell the same bet to Beauty whenever they both wake up. Under this arrangement, the bookie will end up selling *two* follow-up bets to Beauty if they wake up together twice; this will happen precisely if the outcome of the coin toss is tails.

If the bookie follows this strategy, he can sell a Dutch Book to Beauty if her degree of belief in heads upon awakening is 1/2, but not if it is 1/3.

TABLE II

Bets	Payoff	Cost	Heads	Tails
bet #1	$30 if tails	$15	−$15	$15
bet #2	$20 if heads	$10	$10	−$10
bet #3 only placed if tails	$20 if heads *and* tails	$10 if tails	0	−$10
Combined	$20 if heads	$25 if heads	−$5	
	$30 if tails	$35 if tails		−$5

That is, once we take care to specify just how the Bookie arranges to sell all his bets, it turns out that the Dutch Book argument favors the answer 1/3, rather than 1/2. Here is how Beauty can be made to pay if she keeps her degree of belief constant at 1/2. Before going to sleep, the bookie sells Beauty bet #1, with a cost of $15, and a payoff of $30 if the coin lands tails. Every time they wake up together, the bookie sells a follow-up bet, costing $10, and paying $20 if the coin landed heads. If they wake up once, he will sell her one such bet; if they wake up twice, he will sell her two. Thus, if the outcome is heads, Beauty will lose the first bet, and win one follow-up bet; her winnings will be $20, but she will have paid $25 for the two bets. If the outcome is tails, Beauty will win the first bet, and lose two follow-up bets; she will win $30, but will pay out a total of $35. Either way, Beauty loses $5. The Bookie has arranged it so that the combined stake of the follow-up bet(s) is correlated with the outcome. These results are summarized in Table II.

A similar story can be told without invoking a sleeping bookie. Suppose, before going to sleep, we were to ask Beauty whether she thought that the foregoing betting strategy was fair (regardless of whether she would actually take those bets). She would be unable to arrive at a single answer. On the one hand, she would find every individual bet fair in light of the information available to the agent at the time. On the other hand, she would be capable of doing the math and determining that the strategy doomed the bettor to a certain loss. Thus Beauty's degrees of belief would not allow her to consistently evaluate this betting strategy. Analogous problems arise if Beauty changes her degree of belief in heads to any value other than 1/3.

TABLE III

Bets	Payoff	Cost	Heads	Tails
bet #1	$X if heads	$X/2	$X/2	−$X/2
bet #2	$Y if heads	$Y/3	$2Y/3	−$Y/3
bet #3 only placed if tails	$Y if heads *and* tails	$Y/3 if tails	0	−$Y/3
Combined	$(X + Y) if heads	$(X/2 + Y/3) if heads	$(X/2 + 2Y/3)	
	$0 if tails	$(X/2 + 2Y/3) if tails		−$(X/2 + 2Y/3)

By contrast, if upon awakening Beauty's degree of belief in heads is 1/3, she cannot be caught by such a Dutch Strategy. Suppose that the first bet sold by the bookie pays $X if the coin lands heads; then Beauty's fair price for this bet is $X/2. If X is negative, this is equivalent to a bet on tails with payoff $|X| and cost $|X/2|. The follow-up bets will pay $Y if the coin lands heads, and will cost $Y/3. If the coin lands heads, she will 'win' the first bet (actually a loss if X is negative), and will win one follow-up bet, for a gross gain of $(X + Y), and a cost of $(X/2 + Y/3). If the coin lands tails, she will lose three bets – the first bet plus two follow-up bets – for a gross gain of zero, and a cost of $(X/2 + Y/3 + Y/3). Thus her net 'profit' (possibly negative) is $(X/2 + 2Y/3) if the coin lands heads, and −$(X/2 + 2Y/3) if the coin lands tails. These results are summarized in Table III. In order for Beauty to suffer a guaranteed loss, X and Y need to be chosen so that both $(X/2 + 2Y/3) and −$(X/2 + 2Y/3) are negative. Since the one net payoff is just the negative of the other, there is no way to make them both negative. A properly constructed DDB argument thus reinforces the argument from symmetry presented in the previous section: Beauty's degree of belief upon awakening should be 1/3.

I have made no attempt to show that there is *no* way that the bookie can arrange to make exactly one follow-up bet without violating the strictures against extra information. Here is one possibility: The bookie sells his first bet to Beauty before going to sleep. He plans to go to sleep when Beauty does, and arranges to be awakened on Monday, but not on Tuesday. Before going to sleep, Beauty is told of these arrangements. Shortly before waking up, however, both Beauty and the bookie are given a drug that makes them forget the nature of the arrangements that were made (and hence to be uncertain about whether it is Monday or Tuesday). The bookie then sells his second bet. In this manner, the bookie could sell the system of bets comprising the Dutch Book of Section 4. And it *seems* as though the bookie is never in possession of information that is lacking to Beauty. Does this strategy provide us with a rival Dutch Book argument, supporting the answer 1/2? The case is problematic. First, how does the bookie know to sell the bet on Monday? For all he knows, he has already sold the second bet, and to sell yet another would endanger his certain profit. Second, suppose that the outcome of the coin toss is tails, so that Beauty is awakened on Tuesday while the bookie continues to doze. There is one sense in which he has no information lacking to Beauty: he is fast asleep. On the other hand, he is behaving just as though he has been given the information that he is not awake, and is using this information to avoid selling a second follow-up bet on Tuesday. In general, it is just not clear what counts as 'information' in this sort of scenario; indeed, this is one of the features of the puzzle that makes it so interesting. In the absence of any detailed solution to this problem, it seems that the only way to ensure that the 'no extra information' clause is satisfied is to have the bookie undergo the same protocol that Beauty does.

7. MORALS

For those who are persuaded by DDB arguments in general, the argument of the previous section shows that the correct answer to the Sleeping Beauty puzzle is 1/3; and as I will argue shortly, even those who are normally skeptical of DDB arguments have less reason to be skeptical of this particular instance. But first let us review the other arguments that were presented in sections two through five.

The frequency argument of section two is clearly flawed. One cannot simply infer from the fact that the long-run relative frequency of heads awakenings is 1/3, to the conclusion that the probability of heads on some particular awakening – the very first awakening, as the problem is set up – is one-third. However, this argument does contain a grain of truth. The DDB argument of the previous section hinges on the fact that the bookie can place two follow-up bets if the coin lands tails, and only one if the coin lands heads. Thus the actual number of awakenings that occurs in the two different outcomes does play a central role in the solution to the problem.

The argument of section three fails, because it wrongly assumes that Beauty acquires no new information upon waking up. She undergoes a transition from believing the centered proposition $\sim A$ to believing the centered proposition A. While this may not be learning in the normal sense, it does reflect a change in Beauty's overall beliefs.

The DDB argument of section four fails, because the bookie is only able to formulate an appropriate Dutch Strategy by exploiting information that will be unavailable to Beauty. This means that the Dutch Strategy is not one that Beauty would simultaneously evaluate as being both fair and unfair.

The symmetry argument of section five is problematic for at least two reasons. First, it makes assumptions about the nature of Beauty's 'prior' degrees of belief. These are not simply Beauty's degrees of belief at some specific time, say on Sunday before the experiment begins. They are rather Beauty's degrees of belief in the absence of any *temporal* information. These are much harder to construe than temporally located degrees of belief; indeed, they are degrees of belief that Beauty may never actually possess. Second, it was assumed that Beauty assimilated newly 'learned' centered propositions by conditionalization. This cannot be right: conditionalization can never allow us to pass from certainty in the truth of a proposition to certainty in its falsehood (or vice versa), while it is possible to make this transition upon learning a centered proposition. Despite these flaws, the argument of section five seems not to have led us astray, so perhaps our assumptions are innocuous after all.

In addition to shedding light on the Sleeping Beauty problem, the argument of section six also illustrates the care that must be undertaken when wielding Dutch Book arguments. In particular, the strategy of having the bookie adhere to the same protocol as the agent is a useful way to ensure that the bookie has no information that is unavailable to the agent. Consider, for example, the following counterexample to the Reflection principle due to Talbott (1991). Let S be the proposition that Mary ate spaghetti for dinner last night,[8] and suppose that Mary is almost certain (0.99) that S is true. Let t be the present, and let t' be some time one year in the future. Mary knows that at time t', she will no longer remember what she ate for dinner last night. Perhaps at time t' her degree of belief in S will be 0.1. Then Mary's present degree of belief will be $P_t(S|p(S, t', 0.1)) = 0.99$, in violation of Reflection. Can a Dutch Book be made against Mary? Apparently so: the bookie sells to her a bet on S at time t, and then sells to her a bet on $\sim S$ at time t' so as to ensure a net loss.[9] But has the bookie exploited information that is presently unavailable to Mary? Presumably, at time t', Mary has not only forgotten what she ate for dinner on that night so long ago, but she has also forgotten which bet she bought on the following day. If she could recall that she bought a bet on S, despite being offered very unfavorable odds, she would reasonably infer that S is (very probably) true. So suppose we require that the bookie also forget which bet he sold at time t. Then he would not know which bet to sell her at time t' so as to complete the Dutch Book.[10] Thus the bookie cannot sell the Dutch Book to Mary without exploiting information that is unavailable to her. Thus this is not a case where the application of the Reflection Principle can be justified by appeal to a DDB.[11]

This discussion should provide some solace to critics of DDB arguments (see e.g., Bacchus et al 1990; Christensen 1991; Howson 1993). We do indeed have reason to be suspicious of blanket DDB justifications of principles such as Reflection. Although an agent will be vulnerable to a DDB whenever her degrees of belief violate Reflection, there may nonetheless be situations in which she violates that principle, while remaining invulnerable to a DDB *of the appropriate sort*. That is, an agent may violate Reflection without being involved in the sort of evaluative inconsistency that is normally the root cause of Dutch Book vulnerability. The case described by Talbott has just this structure. It follows that those who are skeptical of DDB arguments generally need not be skeptical of the specific DDB argument presented in Section 6. That argument was not offered in blanket support of a generic principle of rationality, but was offered to justify the appropriateness of a specific degree of belief in a specific scenario.

Some will remain skeptical of my argument anyway; even for these holdouts, there is a moral to be drawn. The argument shows that Beauty's situation is *importantly different* from one in which an agent simply learns nothing. In the latter case, an appropriate DDB argument could be constructed to support the answer 1/2. Even those who do not find DDB arguments compelling for purposes of establishing appropriate degrees of belief must grant that a belief change in which an agent is subject to a DDB is structurally different from one in which she is not. Moreover, even those who are skeptical of the idea that Beauty has 'learned' a new centered proposition must recognize that Beauty's epistemic state is interestingly different from that of an agent who has simply learned nothing.

I will conclude by drawing attention to a rather striking feature of the DDB argument that the reader might well have missed. The argument of section six did not invoke any of the symmetry assumptions of section five. In particular, the DDB argument did not make use of the fact that Beauty's 'priors' were such that $P(Sun) = P(Mon) = \ldots = 1/7$. Rather, this symmetry assumption is supported by the DDB argument.[12] Suppose, for example, that $P(Mon) = 0.2$ and $P(Tues) = 0.05$. Then, making the other assumptions in section five, Beauty's degrees of belief upon awakening will be: $P(Mon\&T|A) = 0.4$, $P(Mon\&H|A) = 0.4$,

$P(Tues\&T|A) = 0.2$. This would leave Beauty with a degree of belief in heads equal to 0.4. With this degree of belief, Beauty would be vulnerable to a DDB. Thus, if the framework of section five is the correct one for representing this problem, then the DDB argument shows that she must assign equal 'priors' to Mon and $Tues$. This is unusual: an agent's prior probabilities are not normally considered to be subject to the canons of probabilistic rationality (with the exception that tautologies and contradictions receive priors of 1 and 0, respectively). If the agent, knowing nothing at all about a coin that is about to be tossed, nonetheless has a degree of belief of 0.99 that the coin will land heads, that is her business. Such a degree of belief might be unmotivated, but it is not incoherent. Indeed, this latitude with respect to prior probability assignments is often considered a weakness in the Bayesian program. By contrast, in the Sleeping Beauty problem, symmetry with respect to degrees of belief about one's location in time is *required* for coherence. Presumably, these symmetry constraints arise because both Monday and Tuesday will eventually come to pass; there is no analogous inevitably involving the outcomes of a coin toss.

Uncertainty about one's location within a possible world adds a dimension of complexity to traditional problems in epistemology and decision theory. The Sleeping Beauty Problem and its solution illustrate this point beautifully.

ACKNOWLEDGMENTS

For discussion and comments I would like to thank three anonymous referees as well as Alan Hájek, Matthias Hild, Jim Joyce, Brad Monton, John O'Leary-Hawthorne, Bas van Fraassen, Susan Vineberg, and especially Adam Elga.

NOTES

[1] While remaining unaware that it *is* Monday, of course.

[2] For the unwashed, a *Dutch Book* is a system of bets such that the agent is guaranteed to suffer a net loss if she purchases all of them. A *diachronic* Dutch Book is one in which the bets are placed at different times. An agent is *susceptible* to a Dutch Book if she finds every bet comprising the book to be fair. It is assumed the agent finds fair a bet on proposition A with payoff Q and cost $QP(A)$, where $P(A)$ is the agent's degree of belief that A is true.

[3] Proof: By (i) and (v), we must have (*) $P(Mon, H) + P(Tues, H) = P(Mon, T) + P(Tues, T)$. By (ii), we must have (**) $P(Mon, H) = P(Mon, T) + P(Tues, T)$. It follows from these two equations that (***) $P(Tues, H) = 0$ (perhaps the halfer could motivate this value by appeal to the impossibility of Beauty's learning $\langle Tues, H\rangle$). From

(iii) and (***) we get $P(Mon, H) + P(Mon, T) = P(Tues, T)$. (Since (***) implies that $P(Tues) = P(Tues, T)$.) Subtracting (**) from (****) we get $P(Mon, T) = -P(Mon, T) = 0$, in violation of (iv). *QED*

[4] Lewis (2001) does *not* make this mistake: he very clearly distinguishes the conditional and unconditional evidential relevance, and very clearly asserts that he believes in both. Rather, it seems to me that he just has the bizarre intuition that ‘today is Monday’ is evidentially relevant to ‘the coin landed heads’, whereas ‘either today is Monday, or today is Tuesday and the coin landed tails’ is not.

[5] Bartha and Hitchcock (1999) argue that a similar mistake is implicated in Leslie’s Doomsday argument (Leslie 1996): the evidential relevance of my time of birth for doom *given* that I am born at all is mistaken for the unconditional relevance of my time of birth for doom. Indeed, there are many interesting parallels between the two puzzles, but a detailed exploration of these will have to await another occasion.

[6] See Monton (2001) for an interesting discussion of the relationship between the two different kinds of learning.

[7] This interpretation of what Dutch Book arguments show isn’t newfangled; it dates back to the origin of the Dutch Book arguments in Ramsey (1926).

[8] Where ‘last night’ picks out a particular night rigidly.

[9] The bookie also needs to sell her a side bet on $p(S, t', 0.1)$ at time$t$. This will protect him in case Mary’s degree of belief is S at time t' is *not* 0.1.

[10] Note that this only helps Mary in the case where she actually *does* forget which bet was made. If she and the bookie both do remember, and she thereby resets her degree of belief in S to 0.99, then she will lose money due to the side bet described in the previous footnote. Thanks to Susan Vineberg for pointing this out.

[11] ‘So much the worse for Reflection’ or ‘so much the worse for the putative counterexample’? On the one hand, this argument removes one sort of objection to the Reflection Principle; on the other hand, it narrows the scope of the principle. How narrow? I doubt that it is possible to provide any one simple criterion for when the principle applies and when it doesn’t. For example, the line of argument given above will not help with Christensen’s LSQ example (1991). The difference between memory loss and intoxication is that in the latter case it would be of no help to be reminded of one’s former opinions. This seems to be the sort of evaluation that needs to be made on a case-by-case basis.

[12] At any rate the assumption that $P(Mon) = P(Tues)$ is so supported.

REFERENCES

Aumann, R., S. Hart, and M. Perry: 1997, ‘The Forgetful Passenger’, *Games and Economic Behavior* **20**, 117–120.

Bacchus, F., H. Kyburg, and M. Thalos: 1990, ‘Against Conditionalization’, *Synthese* **85**, 475–506.

Bartha, P. and C. Hitchcock: 1999, ‘No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes’s Theorem’, *Philosophy of Science* **66**, S339–S353.

Christensen, D.: 1991, ‘Clever Bookies and Coherent Beliefs’, *The Philosophical Review* **100**, 229– 247.

Elga, A.: 2000, ‘Self-Locating Belief and the Sleeping Beauty Problem’, *Analysis* **60**, 143–147.

Howson, C.: 1993, 'Dutch Book Arguments and Consistency', in D. Hull, M. Forbes, and K. Okrulik (eds.), *PSA 1992*,Vol. II, East Lansing, Philosophy of Science Association, pp. 161–168.

Leslie, J.: 1996, *The End of the World*, New York, Routledge.

Lewis, D.: 1980, 'A Subjectivist's Guide to Objective Chance', in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Vol. II. Berkeley: University of California Press, pp. 263— 294. Reprinted in D. Lewis, *Philosophical Papers*, Vol. II, Oxford: Oxford University Press, pp. 83–113.

Lewis, D.: 2001, 'Sleeping Beauty: Reply to Elga', *Analysis* **61**, 171–176.

Monton, B.: 2001, 'Sleeping Beauty and the Forgetful Bayesian', *Analysis* **61**, 47–53.

Piccione, M. and A. Rubinstein: 1997, 'On the Interpretation of Decision Problems with Imperfect Recall', *Games and Economic Behavior* **20**, 3–24.

Quine, W. V. O.: 1969, 'Propositional Objects', in *Ontological Relativity and Other Essays*, New York, Columbia University Press, pp. 139–160.

Ramsey, F.: 1926/1990, 'Truth and Probability', in D. H. Mellor (ed.), *Philosophical Papers*, Cambridge, Cambridge University Press, pp. 52–94.

Talbott, W.: 1991, 'Two Principles of Bayesian Epistemology', *Philosophical Studies* **62**, 135–150.

Teller, P.: 1973, 'Conditionalization and Observation', *Synthese* **26**, 218–258.

van Fraassen, B.: 1984, 'Belief and the Will', *The Journal of Philosophy* **81**, 235–256.

Division of Humanities and Social Sciences 101-40
California Institute of Technology
Pasadena CA 91125
U.S.A.
E-mail: cricky@caltech.edu