Diachronic Dutch Books and Sleeping Beauty

One version of the Sleeping Beauty problem (Elga 2000) is this: Beauty is the subject of an experiment that begins when she is put to sleep on Sunday night. She is awakened on Monday morning. That afternoon she is informed that it is Monday. She is put back to sleep on Monday night. The experimenters then flip a fair coin. If it comes up heads, Beauty is not awakened on Tuesday. If it comes up tails, her memories of Monday are erased and she is awakened on Tuesday morning. Her experiences on Tuesday morning (if she has any) are identical to her experiences on Monday morning. Beauty is perfectly rational and knows in advance (with certainty) that the experiment will go exactly as planned. Thus, on Sunday her credence in HEADS (the proposition that the coin toss comes up heads) is 1/2. But what is her credence in HEADS on Monday morning? Of those philosophers who have explicitly addressed this question, the more numerous group, "thirders," claim that the correct answer is 1/3, though a smaller group of "halfers" endorse 1/2.

Hitchcock (2004) attempts to settle the matter by advancing a diachronic Dutch Book argument (hereafter DDB) for 1/3. Specifically, he advances a DDB against 1/2 and then shows that no similar DDB can be used against the thirder. Bradley and Leitgeb (2006) (hereafter B&L) argue that Hitchcock's DDB fails. We demonstrate the following: (a) B&L's criticism of Hitchcock is unconvincing; (b) nevertheless, there are serious reasons to worry about the success of Hitchcock's argument; (c) however, it is possible to construct a new DDB for 1/3 about which such worries cannot be raised.

1. Hitchcock's DDB

Hitchcock's ingenious DDB for 1/3 depends on noting that if vulnerability to a Dutch book is to indicate irrationality then the bookie cannot be allowed to possess more information than his victim at the time at which each bet is placed. This constraint (which vitiates a natural DDB against 1/3) is satisfied by having the bookie undergo the same experimental protocol as Beauty. In other words, the bookie will sleep and wake exactly when Beauty does and will have no memory of Monday if he is awakened on Tuesday. As he, like Beauty, cannot determine whether it is Monday or Tuesday upon awakening and so is incapable of knowing if he has already bet in such a situation previously, Hitchcock has the bookie adopt the strategy of offering the same bet each time he and Beauty awaken. Knowing that upon awakening Beauty's credence in HEADS is 1/2, the bookie proceeds as follows. On Sunday night he sells her a bet that costs $15 and pays $30 if TAILS is true (and pays nothing otherwise). Then on each morning of an experimental awakening he sells her a bet that costs $10 and pays $20 if HEADS is true. Regardless of whether HEADS or TAILS is true, Beauty loses $5:

	Sunday Night	Monday Morning	Tuesday Morning	Net
Heads	-15	+10		-5
Tails	15	-10	-10	-5

Hitchcock also argues that the only betting odds which avoid a book of this sort are those which naturally correspond to P(HEADS) = 1/3:

	Sunday Night	Monday Morning	Tuesday Morning	Net
Heads	-15	+20		+5
Tails	15	-10	-10	-5

2. Bradley and Leitgeb on Hitchcock's DDB

B&L argue that the bet or bets offered to Beauty by Hitchcock's bookie on Monday and Tuesday are unfair because the expected utility of accepting each individual bet is negative:[1]

> In Sleeping Beauty, the Tuesday bet (if there is one) should not be accepted because it is only offered if it is a losing bet. That is, it only exists if the coin lands Tails, so the bet on Heads will lose. Beauty does not know if the bet offered to her is the Monday bet (fair) or the Tuesday bet (unfair). So she simply sums the expected utility. She adds the expected utility of the Monday bet (0) to the expected utility of the Tuesday bet (negative). The result is negative. So neither waking bet should be accepted. . . . We have shown that the waking bets should not be accepted as they have negative utility, so the Dutch book is avoided. (125)

---

[1] Hitchcock also claims that each individual bet must be judged fair by the agent and glosses this as the requirement that the agent judge each bet as "yielding no expected gain or loss for either side" (2004, 412). It should be noted that a general understanding fairness of bets in terms of expected utility seems to render Dutch book arguments unpromising as arguments for probabilism (the doctrine that a person's credences ought to accord with the probability calculus).

Let us examine this reasoning. B&L are arguing that if Beauty's credence in HEADS when she awakens is 1/2, the expected utility of the bet offered by Hitchcock's bookie (i.e. betting on HEADS at even odds) is negative. They point out that because Beauty does not know whether 'Today is Monday' (MON) is true or 'Today is Tuesday' (TUES) is true, the expected utility of betting on HEADS today (BET TODAY) is the "sum" (more precisely, the weighted average) of the respective expected utilities of BET TODAY on the assumption that MON is true and BET TODAY on the assumption that TUES is true. B&L correctly claim that the latter expected utility is negative, but their argument is unconvincing because they do not defend their claim that the former expected utility is zero.

On the assumption that MON is true, BET TODAY has two possible results. One (W1) is that because HEADS is true, Beauty wins the Monday bet. The other (L1) is that because TAILS is true, she loses the Monday bet. Thus, the expected utility of BET TODAY given MON appears to be

$$P(\text{W1/BET TODAY \& MON})U(\text{W1}) + P(\text{L1/BET TODAY \& MON})U(\text{L1}).$$

Notice that P(W1/BET TODAY & MON) = P(HEADS/MON) and P(L1/BET TODAY & MON) = P(TAILS/MON). Perhaps B&L believe that P(HEADS/MON) = 1/2 and that the expected utility of BET TODAY given MON is therefore

$$(1/2)(+10) + (1/2)(-10) = 0.$$

If so, their reasoning is clearly flawed. Given that P(HEADS) = 1/2, P(HEADS/MON) > 1/2. For P(HEADS) = P(MON)P(HEADS/MON) + P(TUES)P(HEADS/TUES). Since P(HEADS/TUES) = 0 and P(MON) < 1, it follows that P(HEADS) < P(HEADS/MON).

In fact, as the following argument demonstrates, the halfer who accepts P(MON/TAILS) = 1/2 must hold that P(HEADS/MON) = 2/3: [2]

(1) $P(\text{HEADS/MON}) = \dfrac{P(\text{HEADS})P(\text{MON/HEADS})}{P(\text{HEADS})P(\text{MON/HEADS}) + P(\text{TAILS})P(\text{MON/TAILS})}$

(2) P(MON/HEADS) = 1

(3) P(MON/TAILS) = 1/2

(4) $P(\text{HEADS/MON}) = \dfrac{P(\text{HEADS})}{P(\text{HEADS}) + (1/2)P(\text{TAILS})}$ [From (1)-(3)]

(5) P(HEADS) = 1/2

(6) P(HEADS/MON) = 2/3 [From (4)-(5)]

On the assumption that P(HEADS/MON) = 2/3, a compelling line of argument is available for the claim that, contrary to B&L, the expected utility of BET TODAY = 0. For on that assumption, the expected utility of BET TODAY given MON appears to be

(2/3)(+10) + (1/3)(-10) = +10/3.

[2] Like Lewis (2001) and many others, B&L (2006, 124) accept the restricted principle of indifference which yields P(MON/TAILS) = P(TUES/TAILS) and, because P(MON/TAILS) + P(TUES/TAILS) = 1, therefore also yields P(MON/TAILS) = 1/2.

If TUES (and therefore TAILS) is true, Beauty will lose the Tuesday bet (L2). Hence, because P(L2/BET TODAY & TUES) = 1, the expected utility of BET TODAY given TUES appears to be

$$U(L2) = -10.$$

It follows that the weighted average of these two expected utilities (i.e. the expected utility of BET TODAY) is

$$P(MON)(+10/3) + P(TUES)(-10).$$

Because P(MON/HEADS) = 1, the halfer must claim that

$$P(MON) = \frac{P(HEADS)}{P(HEADS/MON)} = \frac{1/2}{2/3} = 3/4.$$

Hence, it turns out that the expected utility of BET TODAY given MON is actually positive and *exactly cancels* the negative expected utility of BET TODAY given TUES in the weighted average:

$$(3/4)(+10/3) + (1/4)(-10) = 0.$$

3. A Problem with Hitchcock's DDB

At first glance, the argument we have just offered may seem to show conclusively that B&L are simply mistaken and that both waking bets in Hitchcock's argument are, as judged by Beauty's lights when betting, fair. In fact, however, the soundness of this defense of Hitchcock depends entirely on the correctness of causal decision theory. According to causal decision theory, the rationality of a decision is determined by the utilities of states of affairs which one takes to be causally dependent on the decision.[3] Whether Beauty makes a bet on some other day of the experiment does not causally depend on her decision to bet today. Hence, on causal decision theory, the utility of BET TODAY depends entirely on the respective payoffs of winning and losing today's bet alone. On the assumption that MON is true, then, the only relevant utilities are the respective payoffs of winning the Monday bet (U(W1)) and losing that bet (U(L1)). And on the assumption that TUES (and hence TAILS) is true, the only relevant utility is the payoff of losing the Tuesday bet (U(L2)).

However, as pointed out by Arntzenius (2002, 57-58), the proponent of evidential decision theory may plausibly hold that Beauty ought to take account of the fact that her choice to accept or to reject a given set of odds on one awakening during the experiment is conclusive evidence that she will make the same choice on any other (qualitatively identical) awakening.[4] Beauty knows that on the assumption that HEADS is true, she will not be offered a bet on

---

3 Different causal theorists have different terminology for the object of rational maximization. We gloss over those differences here.

4 The same seems to hold, as Arntzenius notes, if one accepts a "counterfactual" decision theory which permits backtracking counterfactuals (Horgan 1981), for Beauty has good reason to think that if she were to accept a bet on this awakening, she would accept such a bet on any other (qualitatively identical) awakening.

Tuesday and so she will win one bet on heads (W1) if she accepts the bookie's offer to bet today and neither win nor lose any bet on heads if she rejects that offer. However, Beauty also knows that on the assumption that TAILS is true, she will be offered and lose two bets on heads (L1 and L2) if she accepts the bookie's offer to bet today and will neither win nor lose any bet on heads if she rejects that offer. Since winning one bet at the odds specified results in a gain of $10 and losing two bets at the odds specified results in a loss of $20, the expected utility of BET TODAY given MON is

$$(2/3)(+10) + (1/3)(-20) = 0.$$

Inasmuch as the expected utility of BET TODAY given TUES is clearly negative, the weighted average of the two expected utilities (i.e., the expected utility of BET TODAY) is negative if evidential decision theory is correct. The upshot seems to be that the success of Hitchcock's DDB requires the failure of evidential decision theory. In the next section, however, we provide a new DDB for 1/3, one that appears to be telling no matter what decision theory is endorsed.

4. A New DDB for 1/3

In order to see why there is another, more compelling, DDB for 1/3, one must attend to an important element of the version of the Sleeping Beauty story we have told. Following Lewis (2001), we have built into the story the fact that Beauty will be told on Monday afternoon that it is Monday. This element of the story is missing in many versions of the tale. Though halfers other than Lewis elide the issue, it appears that the halfer is committed to holding that Beauty's credence in Heads on Monday afternoon is 2/3. For as we have argued above, the halfer is committed to the claim that on Monday morning P(HEADS/MON) = 2/3. Thus, assuming as Lewis does, that on Monday afternoon Beauty ought to conditionalize on MON when she learns that MON is true, her Monday afternoon credence in HEADS is 2/3. This is what allows for the construction of a new DDB against the halfer which satisfies all the usual restrictions on DDBs and, no matter which decision theory is endorsed, also satisfies any requirement that the bets have zero expected utility.

On Sunday, the bookie sells Beauty a bet that costs $15 and pays $30 if TAILS is true. Given that on Monday afternoon Beauty's credence in HEADS is 2/3, the bookie sells her a second bet, one that costs $20 and pays $30 if HEADS is true. If HEADS is true, then Beauty loses $15 on the first bet and gains $10 on the second bet, for a net loss of $5. If TAILS is true, then Beauty wins $15 on the first bet and loses $20 on the second bet, for a net loss of $5. Regardless of the outcome of the coin toss, Beauty suffers a net loss:

	Sunday Night	Monday Afternoon	Net
Heads	-15	+10	-5
Tails	+15	-20	-5

Unlike the later bets in Hitchcock's book, accepting the second bet in this book does not provide Beauty with evidence that she accepts and loses an identical additional bet. Hence, the reasoning utilized by the evidential decision theorist to evade Hitchcock's DDB cannot be employed here, and the expected utility of the second bet is zero on either evidential or causal decision theory. Furthermore, our DDB can be generalized so that it applies to any value other than 1/3. For the success of the bookie's strategy depends on the bookie knowing (along with Beauty) on Sunday that Beauty's Monday afternoon credence in HEADS will differ from her current credence in HEADS. But the same reasoning that leads from the premise that on Monday morning Beauty's credence in HEADS is 1/2 to the conclusion that on Monday afternoon her credence in HEADS is 2/3 leads from the premise that on Monday morning Beauty's credence in HEADS is 1/3 to the conclusion that on Monday afternoon her credence in HEADS = 1/2; and for any value n where $0 \leq n \leq 1$ and $n \neq 1/3$, this same reasoning also leads from the premise that on Monday morning Beauty's credence in HEADS is n to the conclusion that on Monday afternoon her credence in HEADS is not 1/2. Thus, the thirder is in the unique position of having no reason to suppose that Beauty's Sunday credence in HEADS differs from her Monday afternoon credence in HEADS.

5. The Lessons

It appears that halfers must pin their hopes on one of two possible responses. The first is to argue that all DDBs are unsound (Christiansen 1991). The second is to argue that our DDB for 1/3 violates some heretofore undiscovered constraint on sound DDBs.[5] Those halfers who

---

[5] A third possible response is to part company from Lewis by arguing that Beauty need not, upon learning that it is Monday, update her credence in HEADS to 2/3. It seems, however, that this response can be forestalled by modifying our DDB so that instead of the unconditional bet on HEADS on Monday afternoon, the bookie sells Beauty a conditional bet on (HEADS/MON) each morning that he and Beauty awaken. Each bet costs $20 and pays $30, but as the condition is not fulfilled on Tuesday, there is no gain or loss for Beauty on that day. Even though there are, given TAILS, two bets on HEADS made in exactly the same evidential circumstances, because the bets are conditional on MON, at most one of the bets can be lost and so, as in our DDB above, accepting one bet does not provide Beauty

accept causal decision theory and who allow that some DDBs are sound must also provide an explanation as to why Hitchcock's argument fails.

with evidence that she accepts two losing bets. (We thank [name omitted] for suggesting this alternative to our DDB.)

References

Arntzenius, F. (2002). "Reflections on Sleeping Beauty." *Analysis 62*, 53-62.

Bradley, D. & H. Leitgeb. (2006). "When betting odds and credences come apart: More worries for Dutch book arguments." *Analysis 66*, 119-127.

Christiansen, D. (1991). "Clever bookies and coherent beliefs." *Philosophical Review 100*, 229-247.

Elga, A. (2000). "Self-locating belief and the Sleeping Beauty problem." *Analysis 60*, 143-47.

Hitchcock, C. (2004). "Beauty and the bets." *Synthese 139*, 405-420.

Horgan, T. (1981). "Counterfactuals and Newcomb's Problem." *Journal of Philosophy 78*, 331-356.

Lewis, D. (2001). "Sleeping Beauty: Reply to Elga." *Analysis 61*, 171-76.