The Absentminded Driver: no paradox for halfers

Wolfgang Schwarz <wo@umsu.de>

Draft, February 11, 2008

Abstract. I show that the Absentminded Driver paradox is based on the 'thirder' response to the Sleeping Beauty problem. There is no paradox if one adopts the 'halfer' response.

1 Introduction

The story of the absentminded driver, introduced in [Piccione and Rubinstein 1997b], describes a case where decision theoretic reasoning apparently leads an agent to a course of action she previously recognized as sub-optimal, and to expect a payoff she previously didn't expect for the same decision – despite the fact that she acquired no relevant new information in the meantime. I will argue that these conclusions only follow on a certain interpretation of the story that corresponds to the 'thirder' response to the Sleeping Beauty problem (also introduced in [Piccione and Rubinstein 1997b]). The paradox disappears if instead one employs the 'halfer' response.

2 The paradox

An absentminded driver has to take the second exit off the highway to get home. The first exit leads into a disastrous area; if she takes neither exit, she has to stay at a motel at the end of the highway. Due to her absentmindedness, she is unable to tell whether she is at the first or the second exit when she reaches an intersection. Fortunately however, she has a selection of coins with different biases at her disposal: when she reaches an intersection, she can toss a coin and exit iff it lands heads.

Ideally, she would throw different coins at the two intersections, with a bias towards tails on the first and towards heads on the second. But since she can't tell the two intersections apart, she cannot execute such a plan. She has to figure out the optimal bias on the assumption that it is used on any intersection that will be reached.

For concreteness, let's say that taking the second exit (home) has utility 4, taking the first exit has utility 0, and taking neither has utility 1. The selected coin will definitely be tossed at the first intersection, leading to payoff 0 with probability $(1 - b)$, where b is the coin's bias towards tails. With the remaining probability b, the driver reaches the second intersection, where she will use the same coin again and get payoff 4 with probability $(1 - b)$ or payoff 1 with probability b. So the expected payoff for the action $\underline{b}$ = tossing a coin with bias b is

$$\begin{aligned} U(\underline{b}) &= (1 - b) \cdot 0 + b \cdot (1 - b) \cdot 4 + b \cdot b \cdot 1 \\ &= -3b^2 + 4b. \end{aligned} \tag{1}$$

This function has its maximum at $b = 2/3$, where the expected utility is $4/3$. The optimal coin therefore has bias 2/3 towards tails. Let's call this calculation, which the driver may have carried out before starting her trip, the *simple calculation*.

Now consider what she ought to do once she reaches an intersection, assuming she is fully aware of her predicament. There are three possible outcomes – the disastrous area (D), home (H), and the motel (M). By standard decision theory, the expected utility of a coin selection is determined by the probability with which it leads to those outcomes multiplied by their respective value:

$$U(\underline{b}) = P(\underline{b} \Rightarrow D) \cdot 0 + P(\underline{b} \Rightarrow H) \cdot 4 + P(\underline{b} \Rightarrow M) \cdot 1. \tag{2}$$

$P(\underline{b} \Rightarrow D)$ is the probability for reaching the disastrous area D by choosing a coin with bias b. In evidential decision theory (à la [Jeffrey 1965]) this is taken to equal the conditional probability $P(D \mid \underline{b})$; in causal decision theory (à la [Gibbard and Harper 1978]) it is taken to be a suitable subjunctive conditional, stating that the driver *would* reach D *if* she chose a coin with bias b. I will return to this difference soon.

On either account, $P(\underline{b} \Rightarrow D)$ depends on whether the driver is at the first intersection, call it "Monday", or at the second, "Tuesday". For she can only reach the disastrous area if she is at the Monday intersection, in which case she will get to D iff her chosen coin lands heads. That is,

$$P(\underline{b} \Rightarrow D) = P(\text{Mon}) \cdot (1 - b). \tag{3}$$

Likewise, the probability for getting home (H) is $1 - b$ given that she is at the Tuesday intersection. If she is at the Monday intersection, her present coin toss only determines the probability with which she reaches Tuesday, where another coin toss will decide about her fate. So $P(\underline{b} \Rightarrow H)$ depends both on her current choice $\underline{b}$ and on the choice she will make at Tuesday if she is currently at Monday.

Now, on the assumption that the driver selects bias b at Monday and then reaches Tuesday, it is clear that she will select b there as well, since whatever reasoning and evidence leads her to $\underline{b}$ on the first intersection will still be available to her on the second. (I assume the driver knows that she has no possibility to randomize her coin selection, and that she is not in danger of suffering cognitive mishaps.) Hence in *evidential* decision theory, where $P(\underline{b} \Rightarrow H)$ is the conditional probability of *H on the assumption that* $\underline{b}$, we know that the relevant choice she will make at Tuesday given that she selects $\underline{b}$ at Monday is $\underline{b}$ as well.

Things are less straightforward in *causal* decision theory, where $P(\underline{b} \Rightarrow H)$ is the probability that the driver would reach H by choosing $\underline{b}$. Assume she is at Monday, and consider various counterfactuals about what would happen later at Tuesday depending on what coin she selects now. It is at least not obvious that her Tuesday choice would vary in the right way with her Monday choice. In particular, it doesn't seems like she would *cause* herself to make a certain choice later by making it now. One might therefore hold that when calculating the payoff for $\underline{b}$ at Monday, we cannot assume that the driver will select $\underline{b}$ again at Tuesday.[1]

For maximal generality, let us introduce a second variable c for the bias the driver assumes she will take at Tuesday if she is now at Monday and selects b. More generally, c is the bias the driver assumes she will take at any intersection she reaches – for she knows that she will certainly not choose different values at different intersections. Keeping in mind that c might depend on b (for instance, by identity), we have

$$P(\underline{b} \Rightarrow H) = P(\text{Tue}) \cdot (1 - b) + P(\text{Mon}) \cdot b \cdot (1 - c). \qquad (4)$$

Analogously,

$$P(\underline{b} \Rightarrow M) = P(\text{Tue}) \cdot b + P(\text{Mon}) \cdot b \cdot c. \qquad (5)$$

To calculate the expected utility of $\underline{b}$, we finally need to know how probable Monday and Tuesday are by the light of our driver – or, strictly, how probable they are depending on $\underline{b}$. Observe that if it is Tuesday, then the Monday coin must have landed tails. So the probability for it being Monday depends on the driver's view about what bias she would select upon reaching an intersection. This is our value c (which equals b at least in evidential decision theory). If $c = 0$, she ought to be certain that it is Monday; if $c = 1$, so that she is certain she never exits, I will assume she gives equal credence to being at Monday and being at Tuesday.

In general, since the probability of reaching Tuesday is c times the probability of reaching Monday, it seems that the driver should assign Tuesday c times the probability of Monday. This yields:[2]

$$P(\text{Mon}) = 1/(c + 1). \qquad (6)$$

---

1 The dependency assumption was made in [Piccione and Rubinstein 1997b], and critizised in [Aumann et al. 1997].

2 The fact that P(Mon) depends on c (and hence on b, if $c = b$), was overlooked by Piccione and Rubinstein [1997b], who instead use a fixed value α for P(Mon). The mistake was again pointed

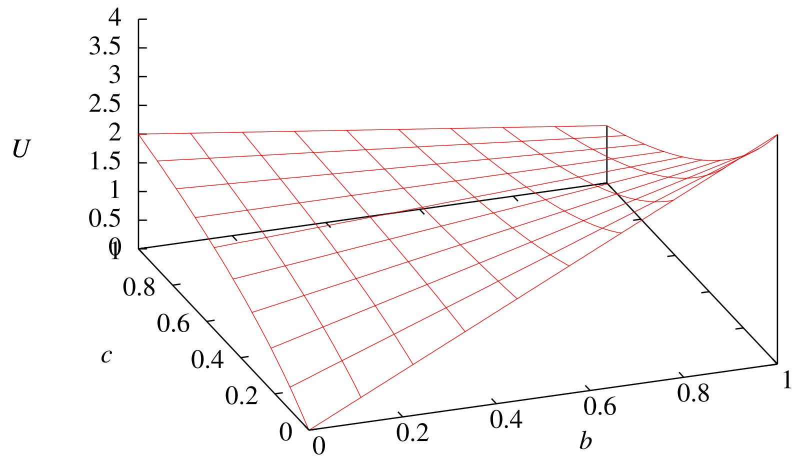
Putting (3), (4), (5), (6) and (2) together, the expected payoff for choosing bias b is[3]

$$\mathrm{U}(\underline{b}) = (4b - 6bc + 4c)/(c + 1). \tag{7}$$

Setting $b = c$, this reduces to

$$\mathrm{U}(\underline{b}) = (8b - 6b^2)/(b + 1). \tag{8}$$

A picture of (7). (8) is the diagonal from the front corner to the rear corner.



Here is the paradox: (7) and (8) contradict the payoff matrix reached by our original simple calculation, (1)!

The contradiction is plain if we use (8), as we should on evidential decision theory. The maximum of (8) lies at b = $\sqrt{336}/12 - 1$, which roughly equals 0.53; the expected payoff there is around 1.67. So should our driver toss a coin with bias 2/3, as the simple calculation said, or would she be better off with bias 0.53? Or was 2/3 indeed optimal from the perspective at the start of the trip, but is no longer optimal as soon as an intersection is reached – despite the fact that she acquired no relevant new information, and that she could easily have foreseen this change of mind? [[Maybe explain how bad this is? It violates Reflection, and can be exploited with a Dutch Book (or simply: charge for switching coins).]]

Things don't look much better on the causal account. – If not, we would have found a decisive argument for causal decision theory that has nothing to do with intuitions about Newcomb cases! Too good to be true.

out by Aumann et al. [1997], who also propose (6) as the correct probability. In [Piccione and Rubinstein 1997b], (6) is effectively used to calculate $\alpha = 3/5$, based on the prior decision by the driver to select bias $2/3$.

3 This is the payoff matrix suggested in [Aumann et al. 1997]. [Piccione and Rubinstein 1997b], setting $c = b$ and P(Mon) $= \alpha$, instead reach $\mathrm{U}(\underline{b}) = \alpha(4p - 3p^2) + (1 - \alpha)(4 - 3p)$, which for $\alpha = 3/5$ has its maximum at $b = 1/3$, with expected payoff $9/5$, compared to $8/5$ for $b = 2/3$. [Rabinowicz 2003: Appendix] considers the result of setting $c = b$ but leaving P(Mon) $= 1/(c+1)$, leading to (8) below.

The verdict of causal decision theory is not entirely straightforward, since (7) now doesn't obviously collapse onto (8), and (7) alone only tells us what the driver should pick *given* a certain prior belief about what she will pick. I will have a few more thoughts about this in the final section. To see that something is wrong with (7), however, we don't even need to settle on a precise verdict.[4]

Recall that according to our simple calculation, the expected payoff for choosing the optimal bias $2/3$ at both intersections is $4/3$. In (7), this ought to correspond to the payoff assigned to $b = c = 2/3$. But this value is $8/5$, not $2/3$. The only c-value for which selecting $b = 2/3$ yields payoff $4/3$ is 1. That is, the only way to reconcile (7) with (1) is to claim that upon reaching an intersection, i) the driver ought to select bias $2/3$, and ii) she ought to believe, based on this choice, that she would select bias 1 when reaching an intersection. (The belief has to depend on the choice, for if the driver believes that $c = 1$ anyway, then (7) tells her to choose $b = 0$, not $b = 2/3$.) But it is hard to see (to put it mildly) how any sensible decision theory could have this result.

The problem generalizes. (1) not only tells us what the optimal bias is, and how good it is, it also tells us the expected payoffs for other possible choices – as becomes important if the driver has only a limited number of coins to choose from, or if she has to pay for coins with a particular bias. For instance, (1) tells us that a fair coin has an expected payoff of $5/4$. But in (7), there is no way selecting $b = 1/2$ could have payoff $5/4$: the minimum value is $3/2$, at $c = 1$. How much, then, should our driver be willing to pay for a fair coin? (If this value changes between the outset and the first intersection, we can make a sure profit on her by first buying the coin off her

---

4 [Aumann et al. 1997] point out two constraints on the optimal choice: 1) it should satisfy $b = c$, 2) it should maximize $U(b)$ holding c fixed. As the picture shows, for large values of c, the optimal b is always 0, and for low values it is always 1. Only at $c = 2/3$ does $U(b)$ not slope in either direction, but rests at constant $8/5$. Setting $b = 2/3$ is therefore the only way to satisfy the two conditions. This is why the driver should stick to 2/3 according to [Aumann et al. 1997].

There are reasons to be dissatified with this solution. For one, it would be desirable to have not only a recommendation on the *best* option, but also a result about which other options are good, and which are worse. Which coins should our driver prefer over which others? The two constraints don't help us with this.

Moreover, the proposal only works for very specific cases and numbers. If we put another motel at the first intersection, so that we get payoff 1 there as well, [Aumann et al. 1997]'s two conditions are nowhere satisfied. Obviously, their solution also doesn't help in any case where the agent does not have a 2/3 coin at his disposal, as in the original case in [Piccione and Rubinstein 1997b]. [Aumann et al. 1997]'s response is to declase this case "uninteresting". I agree with [Piccione and Rubinstein 1997a] that this is not an adequate solution.

Moreover, assuming full independence of b and c, $2/3$ is only rational if the driver knows that $c = 2/3$; otherwise she ought to pick 0 or 1. However, if the agent knows that $c = 2/3$, then she ought to be wholly indifferent between all options, since the outcome is $8/5$ anyway. And therefore she ought to know that she ought to be equally indifferent at any intersection. And then how can she be certain that $c = 2/3$?

[[What shall I do with this? Move it to the main text? Drop it?]]

and then selling it again.)

[[Can I make another point here: if evidential and causal decision theory deliver different results, that is because the coin the driver chooses at another intersection depends probabilistically, but not causally on the coin she chooses at the present intersection. Can we change the story so that the probabilistic dependence coincides with a causal dependence?]]

I conclude that (7) is wrong.[5] This means that either we have to give up traditional decision theory, as represented by (2), when self-locating beliefs and absentmindedness are involved, or we have made a mistake somewhere in (3) – (6)?

I will argue that (3) – (6) are indeed wrong. Once they are fixed, we get the desired coincidence of answers.

3 Halfing and thirding

Suppose for a minute that the driver has only one coin to choose from, which is fair, i.e. $b = 1/2$. Let us focus on the Monday tossing of this coin. Depending on its outcome, the driver will reach Tuesday or not: if tails, yes; if heads, no. So when she reaches the Monday intersection, she can rule out one possibility: that she is at the Tuesday intersection and the coin landed heads. The three other possibilities are still open; it could be tails & Monday, heads & Monday, or tails & Tuesday.

How should she distribute her credence among these possibilities? As before, we assume that she is indifferent between tails & Monday and tails & Tuesday:

$$\mathrm{P(T\&Mon) = P(T\&Tue)}. \qquad (9)$$

The question is, how should these two values be related to the third, P(H&Mon)?

This question is known as the *Sleeping Beauty* problem.[6] Two answers have been defended in the literature. *Thirders* (including [Piccione and Rubinstein 1997b], [Elga 2000]) say that all three possibilities should have the same probability:

$$\mathrm{P(T\&Mon) = P(T\&Tue) = P(H\&Mon)} = 1/3. \qquad (10)$$

Halfers (including [Lewis 2001], [Halpern 2006]) say that P(H&Mon) should be considered twice as probable as the two others:

$$\mathrm{P(H\&Mon) = P(T\&Tue) + P(T\&Mon)} = 1/2. \qquad (11)$$

5 I am not alone with this opinion. Every paper in the 1997 issue of *Games and Economic Behavior* devoted to the Absentminded Driver paradox supports sticking to the 2/3 coin, mostly by considering various different formal models of the situation (see e.g. xxx). If this is correct, at least (8) cannot be the correct payoff matrix.

6 The current version differs from the standard one in that the coin is tossed *after* the Monday situation, as in Elga's version xxx. It is commonly assumed, and I agree, that this makes no difference to the answer.

An argument for this is that at the point under consideration, the driver has no information about the outcome of the coin toss, and therefore should align her credences with the known objective chances: P(H) = $1/2$.

In general, if there is a known objective chance b of there being multiple situations indistinguishable from the present one, and a chance $(1 - b)$ of there being just one, halfers say that, barring further information, one should assign credence $(1 - b)$ to the latter possibility, and credence b to the disjunction of the multiple others; whereas thirders maintain that one should give each of the multiple situations the value b and then renormalize the credence for all situations.[7]

Notice that on either account, the information that the driver is at Monday is relevant to heads vs tails: it rules out one of the three previously open possibilities, tails & Tuesday, which inevitably raises the probability of heads. More precisely, given indifference between tails & Monday and tails & Tuesday, halfers and thirders agree that

$$\begin{aligned} \mathrm{P(T \mid Mon)} &= \mathrm{P(T\&Mon)}/(1 - \mathrm{P(T\&Tue)}) \\ &= (\mathrm{P(T)}/2)/(1 - \mathrm{P(T)}/2) \\ &= \mathrm{P(T)}/(2 - \mathrm{P(T)}). \end{aligned} \tag{12}$$

Thus if upon reaching an intersection, the driver's credence in tails is 1/2, then conditional on being at Monday, her credence should be 1/3. If her unconditional credence in tails is 2/3, her conditional credence should be 1/2.

4 The paradox resolved

Let's return to the above calculation of $\mathrm{U}(\underline{b})$. I assumed that the probability of reaching the desastrous area D when throwing a b-coin is

$$\mathrm{P}(\underline{b} \Rightarrow D) = \mathrm{P(Mon)} \cdot (1 - b). \tag{3}$$

The motivation was this. The driver reaches D iff she is at the Monday intersection and tosses a coin that lands heads. And the probability that a b-coin will land heads if she is at the Monday intersection is $1 - b$.

This last statement seems innocent enough, but it presupposes thirding. The *objective probability* for heads is certainly $1 - b$. But what matters here is the *subjective probability* for heads given that it's Monday and a b-coin is being tossed. (3) assumes that $\mathrm{P(H \mid Mon)} = 1 - b$, and therefore, by (12), that $\mathrm{P(T)} = 2/3 \cdot b$. This corresponds to the thirder distribution, see (10). Halfers object that the unconditional credence for tails should match the known objective chance: $\mathrm{P(T)} = b$, and therefore, by (12), that $\mathrm{P(T \mid Mon)} = b/(2 - b)$. As halfers, we should therefore replace b in

7 see [Halpern 2006], xxx compare z-consistency vs time-consistency in [Piccione and Rubinstein 1997b].

(3) with $b/(2-b)$:

$$P(\underline{b} \to D) = P(\text{Mon}) \cdot (1 - b/(2-b)). \tag{3$'$}$$

Similarly for (4) and (5):

$$P(\underline{b} \to H) = P(\text{Tue}) \cdot (1-b) + P(\text{Mon}) \cdot b/(2-b) \cdot (1-c); \tag{4$'$}$$

$$P(\underline{b} \to M) = P(\text{Tue}) \cdot b + P(\text{Mon}) \cdot b/(2-b) \cdot c. \tag{5$'$}$$

The probabilities for the Tuesday intersection don't have to be adjusted, because the coin toss on Tuesday has no effect on how many indistinguishable situations there are.

However, (6), describing how the probability for Monday depends on the coin bias c, again uses thirding:

$$P(\text{Mon}) = 1/(c+1). \tag{6}$$

Note that for $c = 1/2$, (6) entails that P(Mon) $= 2/3$, in agreement with (10) and contradicting (11). In fact, (6) merely captures the thirder account for distributing an objective chance c of multiple indistinguishable situations by assigning each of them probability c and renormalizing the result: P(Mon) $= (c + (1-c))/(c + c + (1-c)) = 1/(c+1)$. On the halfer account, the disjunction of the indistinguishable situations gets probability c (corresponding to the objective chance), so that P(T&Mon) $= c/2$ and P(Mon) $= c/2 + (1-c) = 1 - c/2$. Halfers should therefore replace (6) by

$$P(\text{Mon}) = 1 - c/2. \tag{6$'$}$$

Putting (3′), (4′), (5′), (6′) and (2) together, we get

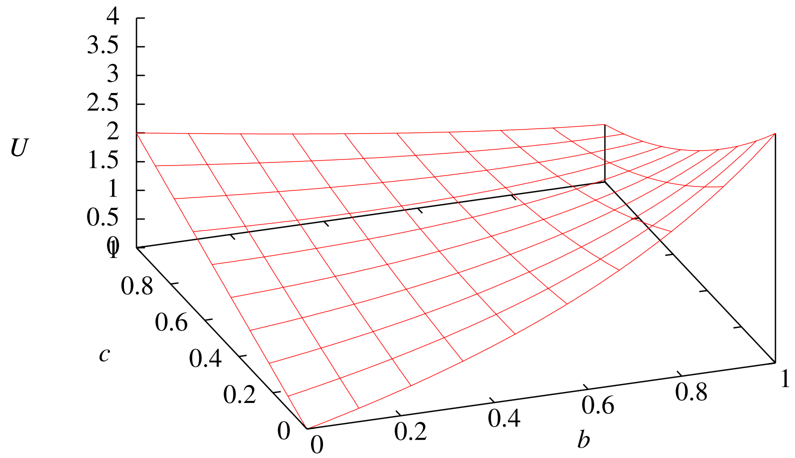
$$U(\underline{b}) = (4b - 5bc + (3/2)bc^2)/(2-b) + 2c - (3/2)bc. \tag{13}$$

For $b = c$, this reduces to

$$U(\underline{b}) = -3b^2 + 4b, \tag{14}$$

which is exactly the result of our simple calculation (1). The previously unresolvable contradiction disappears.

A picture of (13). The surface is a bit more saggy than before.



What the paradox shows is therefore not that traditional decision theory fails in cases of absentmindedness or self-locating beliefs. What it shows is that to adopt decision theory to such a case, one better be a halfer.

5 PS: deciding in causal decision theory

I haven't yet said what the payoff matrix should be in causal decision theory, as the paradox was unavoidable anyway. On the other hand, if we want to fully avoid the paradox, it is obvious what it should be: it should be (1). That is, just like evidential decision theory, causal decision theory should support setting $c = b$. The question is why.

According to causal decision theory, (13) tells us what the optimal bias b is, *given* a certain view about what the driver will choose at an intersection. For instance, given that she would choose bias $c = 0$, the optimal bias for her to choose is $b = 1$, delivering an expected payoff of 2.

To find the optimal bias, our driver thus has to first figure out what bias she would choose when she reaches an intersection. To this end, she has to reproduce her reasoning given her evidence at an intersection. Which leads into a loop: to find the optimal bias, she first has to figure out the optimal bias. (The problem, unlike in 'Death in Damascus', is not that making a decision will alter the expected payoffs, even though it might: after choosing $b = 1$, it is clear that her belief in $c = 0$ was false, and that the expected payoff will be 0, not 2. The main problem here is that there seems to be no way to calculate the payoffs so as to reach a decision in the first place. We seem to have a coordination problem of some sort.)

One might suggest that since our agent cannot complete her circular reasoning, she won't find out what she will do once she reaches an intersection and therefore should select whatever bias is best given total ignorance about c. Let this value be

x. However, if this is what she ought to select, then she could have figured that she will select x upon reaching an intersection. And if $c = x$, then (13) tells us that the optimal b value is *not* x – more precisely, that it is 0 if $c \geq 13/6 - \sqrt{73/4}/3$ and 1 if $c \leq$ that number.[8] And if by this reasoning the driver will conclude that she ought to select, say, bias 1, reproducing *this* reasoning will lead her to conclude that she ought to select bias 0; and so on.

[[Alternatively, we might look at the *ratifiability* or *stability* of the different options. That is, we look at what the driver considers to be the right choice *after* deciding. For instance, if she somehow thought she would select bias 1/4 at an intersection and therefore decides to select bias 1, promising payoff 2, this decision is now evidence that she would actually not select 1/4, but rather 1; so the payoff she would *now* expect is close to 0, and it must seem to her that selecting 1 was not a good idea. We can calculate an option's *degree of ratifiability* as the closeness between the expected payoff before and after the decision. 2/3 might be the most ratifiable option in the matrix. I have to check other stability measures. I suspect that on some measure, you actually get the desired matrix 14. – Also need to think about reasons to be skeptical about ratifiability measures, see [Rabinowicz 1989].]]

Perhaps there is a simpler way out. Note that c is not (as [Aumann et al. 1997] indicate) the bias the driver thinks she will choose at the *other* intersection, but the bias she thinks she will choose at *any* intersection, including the present one. For instance, in (6′), c is the value (assumed to be) selected at the Monday intersection – which may well be the value selected at the current intersection. Of course, the driver knows that whatever she chooses at one intersection, she will also choose at any other, so we would get the same value either way. But while it is unclear whether the bias at another intersection suitably depends on the bias chosen now, it is much more plausible that this dependency holds between the bias chosen at *any* intersection and the one chosen at the present one. In other words, the variable c is not causally independent of b. Thus when we consider the consequences of selecting b, we must not hold fixed c.

In fact, as long as we stick to (13), we have to set $c = b$: (13) tells us the expected payoff if the agent were to select b now, given that she selects c at any intersection. This makes no sense unless $b = c$. One might therefore reconsider (13): perhaps c should be broken up into the value chosen at the present intersection and the value chosen at any other intersection, if reached? [[Can we do that?]] However, there was a reason why we didn't break up c in this way: the driver is not merely confident, but absolutely certain that whatever bias she chooses now is exactly what she would choose at any intersection. When she ponders what *would* happen if she *were* to choose, say, bias 1, she therefore ought to discard any possibility in which she ends

8 The fact that the optimal b depends in this way on c entails that, as in the additional motel setup above, [Aumann et al. 1997]'s solution is inapplicable because their second condition is nowhere satisfied.

up choosing different things at different intersections: she knows for sure that this is not going to happen, whatever she decides. (This is why arguably one-boxing is rational in a Newcomb problem with a *perfectly* reliable predictor.)

[[I need to say more here. Is one-boxing agreed to be rational for absolutely reliable predictors, as can easily happen in a twin setup? If so, why? Should also look into the 'screening of self-prediction' debate.

Perhaps it wouldn't actually be too bad if causal decision theory delivers a different payoff matrix. E.g. if 2/3 is optimal, it might be okay if it says that the expected payoff for 1/2 is U(1/2) on c = 2/3. For this doesn't tell us anything about the case where the driver has no 2/3 coin, or where she has to sell or buy a coin (at least unless there is selling and buying in between the exits)?]]

References

Robert Aumann, Sergiu Hart and Motty Perry [1997]: "The Absent-Minded Driver". *Games and Economic Behavior*, 20: 102–116

Adam Elga [2000]: "Self-locating belief and the Sleeping Beauty problem". *Analysis*, 60: 143–147

Allan Gibbard and William Harper [1978]: "Counterfactuals and Two Kinds of Expected Utility". In C.A. Hooker, J.J. Leach and E.F. McClennen (Eds.) *Foundations and Applications of Decision Theory,* Dordrecht: D. Reidel, 125–162

Joseph Halpern [2006]: "Sleeping Beauty reconsidered: conditioning and reflection in asynchronous systems". In Tamar Gendler and John Hawthorne (Eds.) *Oxford Studies in Epistemology, Vol.1,* Oxford University Press, 111–142

Richard Jeffrey [1965]: *The Logic of Decision*. New York: McGraw-Hill

David Lewis [2001]: "Sleeping Beauty: Reply to Elga". *Analysis*, 61: 171–176

Michele Piccione and Ariel Rubinstein [1997a]: "The Absent-Minded Driver's Paradox: Synthesis and Responses". *Games and Economic Behavior*, 20: 121–130

— [1997b]: "On the Interpretation of Decision Problems with Imperfect Recall". *Games and Economic Behavior*, 20: 3–24

Wlodek Rabinowicz [1989]: "Stable and Retrievable Options". *Philosophy of Science*, 56: 624–641

— [2003]: "Remarks on the Absentminded Driver". *Studia Logica*, 73: 241–256