Lost memories and useless coins

Lost memories and useless coins: Revisiting the absentminded driver*

Wolfgang Schwarz
12 November 2014

Abstract. The puzzle of the absentminded driver combines an unstable decision problem with a version of the Sleeping Beauty problem. Its analysis depends on the choice between "halfing" and "thirding" as well as that between "evidential" and "causal" decision theory. I show that all four combinations lead to interestingly different solutions, and draw some general lessons about the formulation of causal decision theory, the interpretation of mixed strategies and the connection between rational credence and objective chance.

1 Introduction

Sometimes it is controversial what rationality demands in a given situation. One-boxers and two-boxers disagree on the choice to make in Newcomb's problem, halfers and thirders disagree on the beliefs to have in the Sleeping Beauty problem. Such disagreements often trace back to different general perspectives on rationality. At the heart of Newcomb's problem lies the divide between causal and evidential decision theory. At the heart of the Sleeping Beauty problem arguably lies a tension between evidentialism and conservatism in epistemology. In this paper, I want to look at a case that raises both of these issues, as well as several others. The case was introduced in [Piccione and Rubinstein 1997], and goes as follows.

> An absentminded driver has to take the second exit off the highway in order to get home. If she turns off at the first exit, she reaches a desolate area and has to spend the night in her car. If she continues at both exits, she has to stay at a motel at the end of the highway. Due to her absentmindedness, she cannot tell upon arriving at an exit whether it is the first or the second (unless, of course, she knows that she turns off at the first).

* Ancestors of this paper were presented at the ANU in 2007 and the Formal Epistemology Workshop in Munich in 2012. Thanks to Alma Barner, Rachael Briggs, Kenny Easwaran, Alan Hájek, Daniel Nolan, Michael Titelbaum, David Wiens and three anonymous referees for comments and discussion.

Our main question is what the driver ought to do. The answer varies between evidential and causal decision theory, and even between different formulations of the latter. In addition, what the driver ought to do depends on what she ought to believe, and this, too, turns out to be controversial: we will find essentially the same two options as in the Sleeping Beauty problem. We will also see that if the driver makes her choice by tossing a coin, then her degree of belief in the two possible outcomes (heads and tails) does not always match what she knows to be the objective chance. Consequently, several widespread ideas about the role of chance in game theory and decision theory threaten to break down.

The point of this paper is not to take sides in the debate between causal and evidential decision theory or between halfing and thirding. In fact, I will argue that all four combinations give defensible answers to the puzzle if one keeps in mind the general perspective that motivates these combinations.

2 Absentmindedness and two types of expected utility

Before we begin, I should make some clarifications about the driver's predicament. The driver suffers from an unusual kind of absentmindedness. Her problem is not that she is likely to pass an exit without noticing it. On the contrary, she is certain to make a deliberate, rational choice at every exit she reaches. Her problem is that if she decides to stay on the highway at the first exit, then the monotony of the traffic will make her forget the whole event before she reaches the second exit, so that she arrives at that exit in the very same state of mind in which she arrived at the first exit. The two exits may look different, but the differences don't help the driver to figure out which is which. For some reason there are no signposts, and the driver can't leave marks to counteract the memory loss brought on by the traffic. For example, she can't tie a knot in her handkerchief after continuing at the first exit and thus use the handkerchief to find out where she is. Throughout her journey, the driver is aware of all these facts.

Subject to the constraints of the scenario, we assume that the driver is ideally rational, and knows that she is ideally rational. We model her beliefs by a probability measure P over possible states of affairs so that the probability assigned to a state represents the degree to which she believes that the state obtains. Similarly, the degree to which she desires the different states to obtain is represented by a utility function V. The driver would mostly like to get home, but also has a slight preference for staying at the motel over spending the night in her car. For concreteness, let's say that $V(\mathit{Car}) = 0$, $V(\mathit{Home}) = 4$ and $V(\mathit{Motel}) = 1$. We assume that due to the memory loss caused by the highway, her beliefs and desires are in all relevant respects the same whenever she gets to an exit.

It follows that if the driver's beliefs and desires determine a particular choice as uniquely

rational, then that is what she is going to do at every exit. It may therefore be reasonable for the driver to assume that whatever she does at the present exit is also what she does at the other exit (if reached). She might then reason as follows.

"I can either leave the highway here or continue. If I leave, I must be at the first exit, for I know that I make the same choice at every exit, and I couldn't be at the second exit after leaving at the first. So if I leave, I'll end up spending the night in the car. Alternatively, if I continue on the highway, then it is clear that I'll continue at both exits, so I'll spend the night in the motel. That is the slightly better outcome, so I should continue."

There is something odd about this line of reasoning. The driver is right that if she leaves the highway, then she is probably at the first exit and will spend the night in her car. Leaving the highway is bad news. Continuing is also bad news, since it equally entails that the driver won't get home, although it is not quite as bad news as leaving. But arguably the driver's aim is not to receive good news; it is to bring about good outcomes. The two aims often go together, but they come apart in situations in which a particular choice of action would be evidence for a desirable or undesirable state of affairs without having any influence over whether that state obtains. Famously, taking both boxes in Newcomb's problem is evidence that the opaque box is empty, but it doesn't cause the box to be empty. Similarly, if our driver decides to leave the highway, this is strong evidence that she is at the first exit, but it doesn't have any genuine influence over where she is. If she is actually at the second exit, then nothing she can do will bring it about that she is (right now) at the first exit. Likewise if she decides to continue: this is evidence that she would continue at the other exit, but it doesn't control the other (earlier or later) decision.

The different aims – receiving good news vs. bringing about good outcomes – show up in different formulations of decision theory. *Causal decision theory*, as formulated e.g. in [Savage 1954] and [Lewis 1981], advises agents to maximise expected utility in the sense of

$$EU(A) = \sum_{S \in W} P(S)V(S\&A),$$

where W is a suitable partition of states of affairs (which is here assumed to be finite). Roughly speaking, a partition is "suitable" if (a) it is fine-grained enough to distinguish relevantly different outcomes, and (b) the agent's choice has no causal influence over which of the states obtains.[1]

Richard Jeffrey's [1965] *evidential decision theory* instead says that agents should

1 For more precise statements, see e.g. [Lewis 1981] and [Joyce 1999: ch.5]. While it is not essential for my discussion, I assume that states, acts and outcomes are all entities of the same kind (propositions), and that outcomes can be identified with conjunctions of states and acts; see [Joyce 1999: ch.2] for discussion. See also [Skyrms 1984: ch.4] and [Joyce 2002] on the causal interpretation of [Savage 1954].

choose the option with the highest conditional expected utility, defined by

$$CEU(A) = \sum_{S \in W} P(S/A)V(S\&A).$$

Conditional expected utilities are invariant under different choices of W, so the need to define suitability disappears.

Intuitively, the conditional expected utility of a proposition represents the extent to which the agent hopes, or desires, that the proposition is true. That's because the degree to which an agent desires that a disjunction of incompatible propositions X and Y is true can plausibly be identified with the average of the degree to which she desires the disjuncts X and Y, weighted by their relative probability. That is, $V(X \vee Y) = P(X/X \vee Y) \cdot V(X) + P(Y/X \vee Y) \cdot V(Y)$. Since the act A is equivalent to the disjunction of $S\&A$, for all $S \in W$, it follows that $V(A) = CEU(A)$. The advice of evidential decision theory thus amounts to the advice to choose the option for which you have the strongest desire that it be chosen. From the perspective of causal decision theory, this is not always correct, as sometimes you may desire that you make a particular choice merely because that would be evidence for something good, without contributing at all to bringing it about.[2]

If we analyze the driver's problem in causal decision theory, we find that what the driver *should* do inversely depends on what she believes she *will* do. Suppose she is confident that she will leave the highway. She can then infer that she is probably at the first exit, in which case the best choice is to continue. On the other hand, if she is confident that she will continue, then she knows that her journey eventually brings her to both exits, in which case it would be reasonable to give equal credence to being at the first exit and being at the second. Leaving the highway then has an equal probability of bringing her to the desolate area and bringing her home. Since getting home is much better than the other two outcomes, the best choice is then to take the risk and turn off. Either way, if the driver is confident that she will do one thing, she is better off doing the other!

Let's spell this out in a bit more detail. The outcome of the driver's present choice depends on whether she is at the first exit or at the second. If she is at the second exit, then her choice settles whether she reaches home or the motel. If she is at the first exit, the outcome also depends on what she will do at the second exit. At the first exit, the driver has causal control over *whether* she'll face another decision problem at the

2 There are other ways to understand desire on which desirability isn't represented by conditional expected utility. Roughly speaking, conditional expected utilities measure the extent to which the agent would be pleased to learn that the relevant proposition is true. Something else is in play when one expresses a positive attitude towards counterfactual scenarios in which Kennedy was not killed by saying, "I wish Oswald hadn't killed Kennedy". The causal expected utility $EU(A)$ of a proposition A arguably captures this kind of *subjunctive desirability* of A. See e.g. [Etlin 2008: ch.2] for discussion.

second exit, but she doesn't have direct control over her choice at that exit. Different possibilities about that choice should therefore be represented by different states in the partition W. So we have three states: *First* & *Continue*$_2$, *First* & *Leave*$_2$, and *Second*. Here *Continue*$_2$ means that the driver either continues at the second exit or would have continued if she had reached it. The decision matrix for the driver's problem then looks as follows.

	First & *Continue*$_2$	*First* & *Leave*$_2$	*Second*
Continue	1	4	1
Leave	0	0	4

Two comments on this matrix before we continue. First, some outcomes in the matrix are ruled out by the assumption that the driver makes the same choice at every exit. For example, she then can't be at the second exit and leave. I nevertheless assume that the combination of *Second* and *Leave* has a well-defined utility of 4. If that seems problematic, we could allow that the driver reserves a very small probability for the hypothesis that she makes different choices at the two exits – small enough to make no great difference to the following calculations.

A second complication arises from the interpretation of *Continue*$_2$ and *Leave*$_2$. I said that *Continue*$_2$ means that the driver either does or would continue at the second exit if she were to reach it. One might argue that the driver should be certain that if she were to reach the second exit, she would continue, for the ("backtracking") reason that she could only reach the second exit by having continued at the first, in which case she would also continue at the second. In the opposite direction, one might worry that if the driver leaves at the first exit, then there is no fact of the matter about what she would do if she were to reach the second. Either way, the above matrix would not be an adequate representation of the driver's decision problem.

These difficulties could be avoided by moving from a Savage-Lewis type formulation of causal decision theory to Joyce's [1999] formulation in terms of subjunctive conditional probabilities. All the results of the present paper can be replicated in that framework. To fix the right interpretation of *Continue*$_2$ and *Leave*$_2$ in the Savage-Lewis framework, it may help to imagine that the first exit eventually leads to another highway with yet another exit that the driver can't tell apart from the first. (The driver has to spend the night in her car no matter what she does at that exit.) We can then read *Continue*$_2$ as the counterfactual-free proposition that the driver will continue at whatever exit she reaches after the first.

To compute the expected utilities, we need to know the probability of the three states. I will assume for the whole of this paper that conditional on reaching both exits, the

driver gives equal credence to being at the first and being at the second:

$$P(\mathit{First}/\mathit{Continue}_1) = P(\mathit{Second}/\mathit{Continue}_1) = 1/2. \tag{1}$$

Call this the *symmetry* assumption. ($\mathit{Continue}_1$ means that the driver continues at the first exit.) I will also assume for now – although not for the whole of the paper – that the driver is confident that she makes the same choice at every exit; more specifically,

$$P(\mathit{Continue}_1/\mathit{Continue}_2) = P(\mathit{Continue}_2/\mathit{Continue}_1) = 1. \tag{2}$$

Call this the *uniformity* assumption. (The precise value 1, like 1/2 in the symmetry assumption, serves mainly to simplify the calculations.)

Now let c be the driver's degree of belief that she will continue. Notice that *Continue* can be defined as $(\mathit{First}\ \&\ \mathit{Continue}_1) \vee (\mathit{Second}\ \&\ \mathit{Continue}_2)$. Symmetry and uniformity then entail that[3]

$$P(\mathit{First}\ \&\ \mathit{Continue}_2) = c/2; \tag{3}$$

$$P(\mathit{First}\ \&\ \mathit{Leave}_2) = 1 - c; \tag{4}$$

$$P(\mathit{Second}) = c/2. \tag{5}$$

Hence

$$EU(\mathit{Continue}) = c/2 \cdot 1 + (1-c) \cdot 4 + c/2 \cdot 1 = 4 - 3c; \tag{6}$$

$$EU(\mathit{Leave}) = c/2 \cdot 0 + (1-c) \cdot 0 + c/2 \cdot 4 = 2c. \tag{7}$$

The two are equal at $c = 4/5$. If the probability of continuing is greater than 4/5, leaving maximises expected utility, if it is less than 4/5, continuing is best. The grass is always greener on the other side.

In evidential decision theory, we would replace the probability of the states by their probability conditional on the relevant choice. Conditional on *Leave*, all probability goes

3 Proof: I first show that $P(\mathit{Continue}_1) = P(\mathit{Continue}_2) = c$. Observe that *Second* entails $\mathit{Continue}_1$, wherefore *Continue* entails $\mathit{Continue}_1$. More specifically, $\mathit{Continue}_1$ is the disjunction of *Continue* and $\mathit{Second}\ \&\ \mathit{Leave}_2$. By uniformity, the latter has probability 0, as does $\mathit{Continue}_1\ \&\ \mathit{Leave}_2$. Hence $c = P(\mathit{Continue}) = P(\mathit{Continue}_1) = P(\mathit{Continue}_1\ \&\ \mathit{Continue}_2)$. Since $\mathit{Continue}_2\ \&\ \mathit{Leave}_1$ also has probability 0, it follows that $P(\mathit{Continue}_2) = P(\mathit{Continue}_2\ \&\ \mathit{Continue}_1) = c$.

Now $\mathit{Continue}_2$ divides into $\mathit{First}\ \&\ \mathit{Continue}_2$ and $\mathit{Second}\ \&\ \mathit{Continue}_2$, so $c = P(\mathit{First}\ \&\ \mathit{Continue}_2) + P(\mathit{Second}\ \&\ \mathit{Continue}_2)$. By uniformity, it follows that $c = P(\mathit{First}\ \&\ \mathit{Continue}_2) + P(\mathit{Second}\ \&\ \mathit{Continue}_1)$. Moreover, by symmetry, $P(\mathit{First}\ \&\ \mathit{Continue}_1) = P(\mathit{Second}\ \&\ \mathit{Continue}_1)$, and so $P(\mathit{First}\ \&\ \mathit{Continue}_2) = P(\mathit{Second}\ \&\ \mathit{Continue}_1)$ by uniformity. Hence $P(\mathit{First}\ \&\ \mathit{Continue}_2) = P(\mathit{Second}\ \&\ \mathit{Continue}_1) = c/2$. $\mathit{Second}\ \&\ \mathit{Continue}_1$ is equivalent to *Second*. The remaining possibility $\mathit{First}\ \&\ \mathit{Leave}_2$ must then have probability $1 - c$.

to the state *First* & $Leave_2$. Conditional on *Continue*, the driver's probability is evenly divided between *First* & $Continue_1$ and *Second*. Thus

$$CEU(Continue) = 1/2 \cdot 1 + 0 \cdot 4 + 1/2 \cdot 1 = 1, \tag{8}$$

$$CEU(Leave) = 0 \cdot 0 + 1 \cdot 0 + 0 \cdot 4 = 0. \tag{9}$$

These values match the driver's informal reasoning above, on which the choice between *Continue* and *Leave* is effectively a choice between *Motel* and *Car*.

3 Rational indecision

In causal decision theory, the driver faces an *unstable decision problem*: a situation where any tendency to do one thing makes it advisable to do something else. Decision problems of this kind were mentioned in [Gibbard and Harper 1978] and have been studied in [Weirich 1985], [Harper 1986], [Skyrms 1990], [Arntzenius 2008], and [Joyce 2012], among others. One thing brought out by these discussions is that decision theory has implications not just about what agents should do, but also about what they should believe that they will do.

Consider our driver. A simplistic application of causal decision theory would assume that the driver's beliefs about what she will do are externally given as part of the decision problem. Let's say the driver is 95 percent confident that she will continue. By the calculations of the previous section, we would then conclude that she ought to leave. However, if the driver recognizes that leaving is the rational choice, shouldn't that make her reconsider her belief that she will continue? Moreover, how could she have arrived at the belief that she will continue in the first place? Knowing that she is rational, she knows that she will do whatever maximises expected utility. Her beliefs about what she will do thus can't be treated as externally fixed. They have to match her beliefs about expected utility.

So to find out what the driver ought to do, we must first find out what she ought to believe. Brian Skyrms's work on the dynamics of rational deliberation (see [Skyrms 1990]) here proves useful. Suppose the driver begins her deliberation in a state where she is 95 percent confident that she will continue. She can then figure out that leaving has a higher expected utility. Since she knows that she is rational, this should increase her degree of belief that she will leave. However, as soon as $P(Leave)$ goes up, the expected utilities change, so she has to re-assess whether leaving is still the optimal choice. If $P(Leave)$ gets too high, *Continue* becomes the better option, in which case the driver's probability in *Leave* should decrease. If the details are filled in sensibly, this process always leads to an equilibrium, a point where the probabilities no longer change. In the case of the driver, the equilibrium lies at $P(Continue) = 4/5$. Here the two options have equal expected utility, so the probability is no longer pulled in either direction.

So from the perspective of causal decision theory, the driver's degree of belief that she will continue should be 4/5. Since *Continue* and *Leave* then have equal expected utility, both acts are permissible. But they are only permissible *because* $P(\mathit{Continue}) = 4/5$. At the point when the decision is made, the driver must not have made up her mind: she must still be undecided what to do. If she had made up her mind that she will continue, then continuing would not be in line with causal decision theory. In that sense, causal decision theory primarily recommends not an act, but a deliberational state – a *state of indecision*. That state is what makes both options permissible.

To say that the driver should be in a certain state of indecision does not imply that indecision is a further option. By stipulation, the driver has only two options: leave and continue. Nonetheless, the driver's deliberation can lead to a state of indecision. The hallmark of unstable decision problems is that it *must* end in such a state.

If deliberation ends in a state of indecision, some non-deliberative cognitive mechanism must break the tie and select an action. From the agent's point of view, the tie-breaking mechanism will appear stochastic. Consider the driver at the end of deliberation, where she is 80 percent confident that she will continue. She knows that further deliberation won't settle what she will do, for she knows that the unique deliberation equilibrium in her decision problem lies at an 80 percent probability for *Continue*. So she knows that her eventual act is chosen by her tie-breaking mechanism. Yet she is 80 percent confident that she will continue. So she must be 80 percent confident that the tie-breaking mechanism will decide in favour of *Continue*. From the driver's point of view, then, it's as if her actions are chosen by an internal flip of a coin whose bias is set by her state of indecision. As William Harper [1986: 31] puts it, the driver "will have reasoned [her]self into becoming a chance device".

At this point, the uniformity assumption has become implausible. The driver knows that she faces the same decision problem at every exit, but what she will end up doing depends on the outcome of the tie-breaking, which may well be different on different occurrences of the same problem. Let us quickly compute the equilibrium state without the uniformity assumption – assuming instead that, from the driver's perspective, the outcomes of different tie-breakings are independent.

In equilibrium, the driver has figured out the extent c to which she ought to be inclined towards continuing, reflected in her present degree of belief that she will continue. She also knows that if she is at the first exit, then continuing would lead her to another instance of the same decision problem, where she'll reach the same state of indecision, so that the probability of continuing will again be c (because the tie-breakings are independent). So $P(\mathit{First} \,\&\, \mathit{Continue}_2) = P(\mathit{First}) \cdot c$. What is $P(\mathit{First})$? For a start, *First* divides into $\mathit{First} \,\&\, \mathit{Continue}_1$ and $\mathit{First} \,\&\, \mathit{Leave}_1$. By the symmetry assumption,

$$P(\mathit{First} \,\&\, \mathit{Continue}_1) = P(\mathit{Second} \,\&\, \mathit{Continue}_1) = P(\mathit{Second}) = 1 - P(\mathit{First}). \quad (10)$$

Moreover, the driver knows that no matter at which intersection she is, her inclination towards continuing is c; so

$$P(\mathit{Continue}_1/\mathit{First}) = P(\mathit{Continue}) = c. \tag{11}$$

Since $P(\mathit{First}\ \&\ \mathit{Leave}_1) = P(\mathit{Leave}_1/\mathit{First})P(\mathit{First}) = (1-P(\mathit{Continue}_1/\mathit{First}))P(\mathit{First})$, it follows that $P(\mathit{First}) = 1 - P(\mathit{First}) + (1-c)P(\mathit{First})$, which resolves to $P(\mathit{First}) = 1/(c+1)$. Hence the equilibrium probabilities of the three states are

$$P(\mathit{First}\ \&\ \mathit{Continue}_2) = c/(c+1); \tag{12}$$

$$P(\mathit{First}\ \&\ \mathit{Leave}_2) = (1-c)/(c+1); \tag{13}$$

$$P(\mathit{Second}) = c/(c+1), \tag{14}$$

which yields

$$EU(\mathit{Continue}) = \frac{c}{c+1}\cdot 1 + \frac{1-c}{c+1}\cdot 4 + \frac{c}{c+1}\cdot 1 = \frac{4-2c}{c+1}; \tag{15}$$

$$EU(\mathit{Leave}) = \frac{c}{c+1}\cdot 0 + \frac{1-c}{c+1}\cdot 0 + \frac{c}{c+1}\cdot 4 = \frac{4c}{c+1}. \tag{16}$$

In equilibrium, both options must have the same expected utility, which means that $c = 2/3$. So without the uniformity assumption, the driver's tendency towards continuing should be 2/3, not 4/5.

This is a rather satisfying result: if the driver always continues with probability 2/3, then she gets home with probability $2/3 \cdot 1/3 = 2/9$, reaches the motel with probability $2/3 \cdot 2/3 = 4/9$, and the desolate area with probability 1/3. Her expected payoff is $2/9 \cdot 4 + 4/9 \cdot 1 + 1/3 \cdot 0 = 4/3$. In general, if at every exit, c is the probability for continuing, then the expected payoff is $(1-c)\cdot 0 + c(1-c)\cdot 4 + c^2 \cdot 1 = 4c - 3c^2$, which has its maximum at $c = 2/3$.

Recall that evidential decision theory told the driver to continue, for a guaranteed payoff of 1. Agents following the causal theory therefore seem to have a greater expected payoff than agents following the evidential theory. This is noteworthy because asymmetrically unstable situations like the driver's are often presented as intuitive *problems* for causal decision theory (e.g. [Richter 1985], [Egan 2007]). In my view, causal decision theory – when properly spelled out – gets such cases exactly right. Suppose the driver follows the evidential advice and decides to continue. She then gives equal credence to being at the first exit and being at the second. So she is 50 percent confident that the present exit leads home. But she much prefers getting home to the other two outcomes – and note that it makes no difference on the evidential account whether the utility of getting home is 4 or 4 million. Shouldn't this make her reconsider her decision? Shouldn't she be tempted to take the exit and try her luck?

Confronted with the better expected payoff of causal decision theory, friends of evidential decision theory might complain that we have given the causal agent an unfair advantage by allowing her to remain in a state of indecision. However, we can grant that the evidential agent, too, has the capacity to remain in a state of indecision. But evidential decision theory doesn't seem to recommend a state of indecision. What's true is that if we gave the evidential agent a further *option* to randomise her choice, then she might decide to randomise in such a way that continuing has probability 2/3 (as we will see). But our causal agent hasn't been given any new options. She was deliberating only between *Continue* and *Leave*. There was no further possibility of delegating her choice to a stochastic mechanism, and this is not what she decided to do.[4]

On the other hand, evidential decision theory unequivocally recommends *Continue* only if we hold fixed the uniformity assumption. That assumption is warranted if the driver decides to continue; so continuing is indeed an equilibrium in the evidential deliberation dynamics. But there is also an equilibrium state of indecision in which the driver is not certain that she will choose the same act at every exit. The evidential driver can also reason herself into becoming a chance device. The conditional and unconditional expected utilities then coincide. For assume the driver is certain that at any exit she will continue with probability c. Then $P(\mathit{Continue}/\mathit{First}\ \&\ \mathit{Continue}_2) = P(\mathit{Continue}) = c$, and so $P(\mathit{First}\ \&\ \mathit{Continue}_2/\mathit{Continue}) = P(\mathit{First}\ \&\ \mathit{Continue}_2)$; similarly for the other two states. Intuitively, the hypothesis *Continue* no longer affects the probability that she continues at the other exit, nor does it shed any light on whether the present exit is the first or the second; it is merely information about the outcome of a chance process. The states have become evidentially independent of the acts, and so the conditional expected utilities coincide with the unconditional expected utilities.

In evidential decision theory, the driver's decision problem therefore has two solutions: the driver can decide to continue, or she can be in state of indecision where she is 2/3 inclined towards continuing, at which point both acts are permissible. Which equilibrium will be reached depends on the starting point and on the details of the deliberation dynamics.

I have argued that a state of indecision is not a further option. Reasoning oneself into becoming a chance device is not the same as deciding to randomize one's acts by using a chance device. Admittedly, the distinction is subtle. One might argue that if indecision is a possibility, then it should be formally treated as a further option. Instead of pursuing this matter in abstract generality, let's see what happens to the driver if we explicitly add randomisation as an option.

4 Pace [Lewis 1981: 29f.].

4 Randomisation: a puzzle and an alleged solution

The scenario we are going to look at for the rest of the paper is the same as before, except that the driver now has the additional option of tossing a coin of any bias she likes. Let's say that heads means leave, tails continue.

It then makes sense to redefine the uniformity assumption to say that the driver is confident that she chooses the same bias at every exit she reaches. As we saw at the end of the previous section, the optimal coin to choose at every exit seems to have bias 2/3 towards tails, since the expected payoff from choosing a coin with bias c at every exit is $4c - 3c^2$, which is maximal for $c = 2/3$. (Here and henceforth, 'bias' always means bias towards tails, i.e. towards continuing.) Given the redefined uniformity assumption, we might therefore expect that evidential decision theory recommends this choice. After all, conditional on choosing any particular bias, the driver is certain that this is her choice at every exit, so comparing conditional expected utilities amounts to considering which coin would be best given that it is tossed at every exit.

Oddly, this is not what we find – at least on a *prima facie* plausible way of modelling the situation. Instead, we find that causal decision theory recommends the coin with bias 2/3, while evidential theory seems to recommend a sub-optimal coin with bias around 0.53.

The causal argument mirrors the argument at the end of the previous section, but I will spell it out a bit more carefully this time. To apply causal decision theory, we first have to find a suitable partition of states. This is not entirely straightforward, because the eventual outcome depends (among other things) on the result of the present coin toss, which is *partly* under the driver's control: by choosing a certain bias, she can make it more or less likely that she will continue, but she doesn't have any further control over how the coin lands. To model this, we will follow the advice of [Lewis 1981] and [Skyrms 1984] and use a partition of states which, combined with a choice of the driver, only determines an objective chance for the relevant outcomes. Thus *Second* is still a complete state, because it determines, combined with any choice of a particular coin, the chances of getting home and of reaching the motel. In general, the cells in our decision matrix are propositions assigning objective probabilities to the three ultimate outcomes. I will call such propositions *lotteries*.

Second represents the driver as being at the second exit. In the other states, she is at the first exit; combined with the choice of a bias, this determines an objective probability for reaching the second exit. That's not yet a complete lottery, because the objective probabilities of the eventual outcomes further depend on what coin the driver is disposed to choose at the second exit. In the previous section, we divided *First* into *First* & *Continue*$_2$ and *First* & *Leave*$_2$. Similarly, we now divide *First* by all possible choices the driver could make at the second exit.

Let '$Bias_2 = x$' be the proposition that the driver either chooses bias x at the second exit or would have chosen bias x if she had reached the exit. (The counterfactuals have to be treated with the same caution as in section 2.) If there are uncountably many possible values of $Bias_2$, we get uncountably many states, which leads to minor complications further down the line. To keep things simple, let's assume – as seems realistic anyway – that the driver has only a finite number of coins at her disposal. Let B be the set of available bias values. It doesn't matter exactly which values are in B; I assume it contains at least all ratios n/m for moderately sized n and m with $m \geq 1$ and $n \leq m$.

Now we have finitely many states, each of which, combined with one of the driver's options, determines a lottery: an assignment of objective chance to the three eventual outcomes. What is the utility of such a lottery? Plausibly, it is the average of the utility of the outcomes, weighted by the chances. For example, the utility of tossing the 2/3 coin at the second exit is 1/3 times the utility of getting home (4) plus 2/3 times the utility of reaching the motel (1). Let $Bias = b$ be the proposition that the driver chooses bias b at the present exit. It follows that for all $b, c \in B$,

$$V(\text{First } \& \ Bias_2 = c \ \& \ Bias = b) = (1-b)0 + b(1-c)4 + bc1 = 4b - 3bc; \tag{17}$$

$$V(\text{Second } \& \ Bias = b) = (1-b)4 + b1 = 4 - 3b. \tag{18}$$

This fixes the utilities of all lotteries – the cells in the decision matrix. Next we need to know a few things about how the driver's degrees of belief are distributed over the states. By probability theory, $P(\text{First } \& \ Bias_2 = c) = P(Bias_2 = c)P(\text{First}/Bias_2 = c)$. As argued in the previous section, the driver's degree of belief in *First*, given that she continues with probability c at the first exit, should be $1/(c+1)$; that is, $P(\text{First}/Bias_1 = c) = 1/(c+1)$. By uniformity, we then also have $P(\text{First}/Bias_2 = c) = 1/(c+1)$. It follows that $P(Second/Bias_2 = c) = 1 - P(First/Bias_2 = c) = c/(c+1)$. Hence

$$P(\text{First } \& \ Bias_2 = c) = P(Bias_2 = c)/(c+1); \tag{19}$$

$$P(\text{Second}) = \sum_{c \in B} P(Bias_2 = c)c/(c+1). \tag{20}$$

Putting all this together, the expected utility of the driver's options are given by

$$EU(Bias = b) = \sum_{c \in B} P(Bias_2 = c)\frac{4b + 4c - 6bc}{c+1}. \tag{21}$$

As before, what the driver ought to do depends on what she believes she will do. For example, if the driver is certain that she chooses a coin with bias 1, then the expected utility of choosing bias 1 is $(4+4-6)/(1+1) = 1$, while the expected utility of choosing bias 1/2 (say) is $(2+4-3)/(1+1) = 3/2$. So $Bias = 1$ is not an equilibrium in the deliberation dynamics: the more the driver is inclined towards bias 1, the more her other options appear better. An equilibrium choice would be a value of c on which $EU(Bias = b)$

is maximal for $b = c$. The only such choice is (unsurprisingly) 2/3. If the driver is certain that she'll choose bias 2/3, the expected utility of bias b is $(4b + 8/3 - 4b)/(5/3) = 8/5$. This is constant, so the deliberation is no longer pulled anywhere else.[5]

The recommendation of evidential decision theory is now easy to compute. Conditional on any $Bias = b$, the driver can be confident that $Bias_2 = b$. The conditional expected utility of choosing bias b is therefore the "diagonal" of $(4b + 4c - 6bc)/(c + 1)$, with $b = c$, which works out to

$$CEU(Bias = b) = (8b - 6b^2)/(b + 1). \tag{22}$$

This has its maximum at $b = \sqrt{336}/12 - 1 \approx 0.53$, where the conditional expected utility is around 5/3. So evidential decision theory recommends a coin with bias 0.53.

As I mentioned above, that is a puzzling result. By using conditional expectations, evidential decision theory effectively represents the driver's choice as a choice of which bias to use not only at the present exit, but at every exit. Given that the optimal bias to use at every exit is 2/3, one might therefore have expected evidential decision theory to recommend bias 2/3. But now we find that *causal* decision theory recommends 2/3, while evidential decision theory recommends something else. As a consequence, drivers who follow the causal advice score higher utility, on average, than drivers who follow the evidential advice. This time, friends of evidential decision theory can't even complain that causal agents have secretly been given further options, in the form of states of indecision, for causal decision theory doesn't recommend a state of indecision.

The puzzle goes further. As Piccione and Rubinstein [1997] point out, drivers who follow the evidential advice seem to undergo a curious change of mind. Suppose at the start of her journey, the driver already considered what she ought to do at every exit, and saw that 2/3 is the optimal bias. (If the driver has to fix a bias once at the start of her journey, both causal and evidential decision theory say she ought to choose bias 2/3.) As soon as she reaches an exit, she would now change her mind and prefer a coin with bias 0.53. What could justify this change of opinion? It doesn't seem to be prompted by her learning any interesting facts: if she had anticipated at the start of the journey everything she would observe at the exit, she still would have regarded the coin with bias 2/3 to be optimal.

One might be tempted to put the blame on evidential decision theory. After all, causal decision theory gives the intuitively correct advice. In essence, this is the answer of [Aumann et al. 1997] to Piccione and Rubinstein's puzzle.[6] To anyone who has followed

5 There are also equilibrium states of indecision. In general, any state in which the expected bias is 2/3 is a stable solution.

6 Neither Aumann et al. nor Piccione and Rubinstein actually mention the connection to evidential and causal decision theory. Piccione and Rubinstein implicitly assume evidential decision theory by setting $c = b$. In addition, they overlook the fact that the probability of being at the first exit (evidentially) depends on the chosen bias b. Instead they assume that $P(First)$ has a fixed value of 3/5, based on

the deadlocked philosophical debate over causal vs. evidential decision theory, such a simple diagnosis should raise suspicion.

Indeed, set aside the debate between causal and evidential decision theory. On either account it is plausible that conditional expected utilities measure the agent's degree of desire, or hope, that the relevant proposition is true. By the above argument, the driver should therefore be happy to discover that she uses a coin with bias 0.53 rather than 2/3. Is this correct? It is certainly not what the driver would have thought at the start of her journey. There, she would have preferred to learn that she uses bias 2/3. Whence her change of mind, in the absence of any relevant new information?

To understand what is going on here, we need to look at another puzzle that was also introduced in [Piccione and Rubinstein 1997] (as 'example 5') and has gained fame in philosophy by the exchange between [Elga 2000] and [Lewis 2001]: the Sleeping Beauty problem.

5 Halfing and Thirding

The scenario is probably familiar. On Sunday night, while Sleeping Beauty is asleep, a fair coin is tossed. If it lands tails, Beauty will be awoken on Monday and again on Tuesday, but before the second awakening, all her memories of Monday will be erased. If the coin lands heads, Beauty will be awoken only on Monday and made to sleep through Tuesday. Beauty knows all these facts.

The parallels to the absentminded driver should be obvious. Imagine the driver decides to use a fair coin, and focus on the outcome of the toss at the first exit. If the outcome is tails, the driver reaches both the first and the second exit ("Monday" and "Tuesday"), but will have lost all memories of the first by the time she reaches the second. If the coin lands heads, she only reaches the first exit ("Monday").[7]

The "Sleeping Beauty problem" is the question what Beauty should believe about the outcome of the coin toss when she wakes up on Monday morning. Analogously, we can ask what the driver should believe about the outcome of the first toss when she arrives

the driver's prior decision to use bias $c = 2/3$ and the fact that $P(\mathit{First}) = 1/(c+1)$. Setting $c = b$ then yields the payoff function $(6b - 3b^2 + 8)/5$, which has its maximum at $b = 1/3$. The corrected evidential formula $(8b - 6b^2)/(b+1)$ appears in the appendix of [Rabinowicz 2003].

Aumann et al. [1997] object that by setting $c = b$, Piccione and Rubinstein erroneously represent the driver's present choice as controlling her choice at the other exit. They suggest that if the driver is certain that she chooses bias c, then the expected utility of $\mathit{Bias}{=}b$ is $(4b + 4c - 6bc)/(c+1)$, which agrees with our formula (21). According to Aumann et al., the optimal bias is then the value of b that maximises that function when plugged in for c: 2/3.

7 One disanalogy between the two cases is that the driver's coin is tossed after she arrives at the first exit, while Beauty's coin is usually assumed to be tossed before she wakes up on Monday. However, that timing is plausibly inessential to the Sleeping Beauty problem. Since the outcome of the toss has no effect until Tuesday morning, the coin might as well be tossed on Monday instead of Sunday.

at the first exit. *Halfers* say her credence in heads should be 1/2; *thirders* say it should be 1/3.

Numerous arguments have been given for either side, and this is not the place to recapitulate the whole debate. Broadly speaking, considerations of diachronic rationality tend to support halfing, while considerations of evidential support tend to support thirding. For example, halfing may be supported by the following principle of *doxastic conservatism*:

> If an agent rationally assigns positive credence to a proposition A which is certain not to change its truth-value, then her credence in A should remain unchanged as long as she receives no evidence which, by the lights of her own previous beliefs, has any bearing on the truth of A.

Everyone agrees that on Sunday, Beauty's credence in heads should be 1/2. On Monday morning, she then learns various facts (including "centred" facts): that she is awake, that the sky is overcast, and so on. For each of these, we can ask whether Beauty's Sunday credence in heads would have been any different if she had already known that she would learn the relevant fact when awakening. The answer is plausibly 'no'. In that sense, whatever Beauty learns on Monday is, by the lights of her previous beliefs, neutral on the outcome of the coin toss. By the principle of doxastic conservatism, her new credence in heads should therefore still be 1/2.[8]

A standard argument for thirding, by contrast, looks just at Beauty's evidence on Monday morning. Regarding the outcome of the coin toss, Beauty's evidence includes (i) her knowledge of the general setup, (ii) her experience of being awake, and (iii) the fact that she has no memories from later than Sunday. By itself, (i) lends equal support to the four combinations *Heads & Monday*, *Heads & Tuesday*, *Tails & Monday*, *Tails & Tuesday*; (iii) rules out any further possibilities (such as *Heads & Wednesday*), and (ii) excludes *Heads & Tuesday*. The remaining possibilities should therefore have probability 1/3 each – assuming that Beauty should believe these propositions to the degree to which they are supported by her present evidence.[9]

8 This line of thought occurs in [Elga 2000] and [Lewis 2001], and is further developed e.g. in [Bradley 2011], [Schwarz 2012a] and [Schwarz 2012b]. One complication here is that Beauty may be forced to violate doxastic conservatism if the coin lands tails: for example, her credence in the proposition *that the sky is overcast on either Monday or Tuesday* will be high on Monday evening, but low on Tuesday morning, although she doesn't acquire any relevant information. (At least this should be so if we assume, as we did for the driver, that her belief state on Monday morning is identical to her belief state on Tuesday morning.) One might think that this threat of diachronic irrationality somehow undermines the force of doxastic conservatism at the earlier transition from Sunday to Monday. In [Schwarz 2012b] I argue that it doesn't.

9 The argument from evidential support can be found, in different variations, e.g. in [Piccione and Rubinstein 1997], [Dorr 2002], [Arntzenius 2003], [Horgan 2004], [Horgan 2008] and [Horgan and Mahtani 2013].

These considerations apply with equal force to the absentminded driver. Of course, the driver plausibly doesn't use a fair coin, so we have to generalise the halfer and thirder positions. Suppose the driver knows that the coin she chooses at the first exit has bias c towards tails. The halfer position then suggests that her credence in tails, when she arrives at the first exit, should equal c. Thirding, on the other hand, suggests that her credence should equal $2c/(c+1)$. Let me re-run the above argument for thirding to explain why. Imagine the driver knows that she reaches the first exit at a time called "Monday" and the second (if she continues) at "Tuesday". When she arrives at the first exit, her relevant evidence consists of (i) her general knowledge of the setup, including the fact that a coin with bias c is tossed at the first exit, and (ii) her observation that she is still on the highway. (i) supports the hypothesis that the coin at the first exit lands tails to degree c; it is neutral on *Heads & Monday* vs. *Heads & Tuesday*, and on *Tails & Monday* vs. *Tails & Tuesday*; (ii) then rules out *Heads & Tuesday* as well as any further possibilities. This leaves the two tails possibilities with probability $c/(c+1)$ each, and the remaining heads possibility with $(1-c)/(c+1)$.[10]

Here I have assumed that the driver knows all along that the coin at the first exit has bias c. But the arguments plausibly carry over to the driver's conditional beliefs. Thus

$$P(\mathit{Tails}_1/\mathit{Bias}_1 = c) = \begin{cases} c & \text{on halfing} \\ 2c/(c+1) & \text{on thirding.} \end{cases} \tag{23}$$

I have added a subscript '1' to '*Tails*' because there might be a second coin toss at the second exit, which doesn't exist in the Sleeping Beauty story.

It is worth teasing out a few consequences of the two positions. Combined with the symmetry assumption (that the driver assigns equal credence to being at the first and being at the second exit conditional on reaching both), the halfer position entails that $P(\mathit{First}\ \&\ \mathit{Tails}_1/\mathit{Bias}_1 = c) = c/2$ and $P(\mathit{First}\ \&\ \mathit{Heads}_1/\mathit{Bias}_1 = c) = 1-c$; by contrast, on the thirder position, $P(\mathit{First}\ \&\ \mathit{Tails}_1/\mathit{Bias}_1 = c) = c/(c+1)$ and $P(\mathit{First}\ \&\ \mathit{Heads}_1/\mathit{Bias}_1 = c) = (1-c)/(c+1)$.[11] Thus

$$P(\mathit{First}/\mathit{Bias}_1 = c) = \begin{cases} 1-c/2 & \text{on halfing} \\ 1/(c+1) & \text{on thirding.} \end{cases} \tag{24}$$

10 For another motivation of the thirder assignment, observe that if the experiment were repeated indefinitely, the ratio of occasions where the driver finds herself at an exit and the (first) coin lands tails to such occasions where the coin lands heads would converge to $2c : (1-c)$.

11 In the case of Sleeping Beauty, the symmetry assumption is sometimes motivated by a stipulation that Beauty's two tails awakenings are "subjectively indistinguishable". Symmetry is then supposed to follow from a general principle of self-locating indifference. Halfers sometimes object to symmetry, suggesting that Beauty ought to be certain on Monday that it is Monday and on Tuesday that it is Tuesday; see e.g. [Hawley 2013]. This response gets less attractive if it is stipulated, as I have, that the driver has the same beliefs at every exit. (Whether the two situations are otherwise indistinguishable is to my mind irrelevant, but feel free to suppose so if you think it matters.)

Knowing $P(\mathit{First}\ \&\ \mathit{Tails}_1/\mathit{Bias}_1 = c)$ and $P(\mathit{First}_1/\mathit{Bias}_1 = c)$, we can compute their ratio,

$$P(\mathit{Tails}_1/\mathit{First}\ \&\ \mathit{Bias}_1 = c) = \begin{cases} c/(2-c) & \text{on halfing} \\ c & \text{on thirding.} \end{cases} \tag{25}$$

Unlike the first toss, the second toss (if it comes about) has no influence on how many exits the driver will reach. From an epistemic perspective, it is an ordinary coin toss, so halfers and thirders should agree that

$$P(\mathit{Tails}_2/\mathit{Tails}_1\ \&\ \mathit{Bias}_2 = c) = P(\mathit{Tails}_2/\mathit{Tails}_1\ \&\ \mathit{First}\ \&\ \mathit{Bias}_2 = c) = c. \tag{26}$$

I will call this the *neutral* assumption. (Conditioning on Tails_1 here only serves to ensure that the second coin toss actually takes place.)

Finally, let's compute the driver's attitude towards the proposition that the coin at the present exit, whichever it may be, will land tails. That is, define Tails as $(\mathit{First}\ \&\ \mathit{Tails}_1) \vee (\mathit{Second}\ \&\ \mathit{Tails}_2)$. Observe that Tails_1 is equivalent to $\mathit{Tails} \vee (\mathit{Second}\ \&\ \mathit{Heads}_2)$. So every coherent probability measure P must satisfy

$$P(\mathit{Tails}) = P(\mathit{Tails}_1) - P(\mathit{Second}\ \&\ \mathit{Heads}_2). \tag{27}$$

By the neutral assumption, $P(\mathit{Heads}_2/\mathit{Second}\ \&\ \mathit{Bias}_2 = c) = 1 - c$. Moreover, by (24), $P(\mathit{Second}/\mathit{Bias}_1 = c)$ equals $c/2$ on halfing and $c/(c+1)$ on thirding. Assuming uniformity, it follows that $P(\mathit{Second}\ \&\ \mathit{Heads}_2/\mathit{Bias}_1 = c)$ and $P(\mathit{Second}\ \&\ \mathit{Heads}_2/\mathit{Bias}_2 = c)$ both equal $(1-c)(c/2)$ on halfing and $(1-c)c/(c+1)$ on thirding. Combined with (23) and (27), this yields

$$P(\mathit{Tails}/\mathit{Bias}_1 = c) = P(\mathit{Tails}/\mathit{Bias}_2 = c) = \begin{cases} (c + c^2)/2 & \text{on halfing,} \\ c & \text{on thirding.} \end{cases} \tag{28}$$

So if the driver is a halfer and tosses a coin with bias 2/3 towards tails (say), then her credence in the hypothesis that the coin she tosses will land tails is not 2/3, but 5/9. This is certainly an unusual situation. The driver seems to violate a form of the "Principal Principle" according to which one's rational credence should match the known objective chances (compare [Lewis 1980]).[12] However, since $P(\mathit{Tails}) = P(\mathit{Tails}_1) - P(\mathit{Second}\ \&\ \mathit{Heads}_2)$, some such violation is unavoidable: if the driver knows that the objective chance of tails at every exit (and hence also at the present exit, whichever it is) equals c, then $P(\mathit{Tails}_1)$ and $P(\mathit{Tails})$ cannot both equal c as long as the driver is open-minded about her location and the outcome of the second toss. If one of them matches the known chance, the other one doesn't.

12 The driver does in fact not violate the Principle formulated in [Lewis 1980], since her credences aren't "initial". One might also question whether there even is an objective chance for the centred proposition *Tails*, as opposed to the uncentred Tails_1 and Tails_2.

6 The puzzle resolved

In section 4, we calculated the optimal bias for the driver's coin, getting 2/3 for causal decision theory and around 0.53 for evidential decision theory. If you look back at the calculations, you can see that we have unwittingly presupposed thirding. (To my knowledge, this presupposition is universally shared in the literature on the absentminded driver.) In particular, we assumed that $P(\mathit{First}/\mathit{Bias}_1 = c) = 1/(c+1)$. This is in line with thirding, but not with halfing. On the halfer account, $P(\mathit{First}/\mathit{Bias}_1 = c) = 1 - c/2$.

Another point where we presupposed thirding is in the calculation of the utilities. We identified the utility of a lottery with the expectation of the utility of the outcomes relative to their objective probability. For example, we assumed that

$$V(\text{First } \&\ \mathit{Bias}_2 = c\ \&\ \mathit{Bias}_1 = b) = 4b - 3bc.$$

The reasoning was that conditional on $\mathit{First}\ \&\ \mathit{Bias}_1 = b\ \&\ \mathit{Bias}_2 = c$, the driver's probability for reaching the desolate area is $1-b$, for getting home $b\cdot(1-c)$, and for getting to the motel $b\cdot c$; so the desirability of the lottery is $(1-b)\cdot 0 + b(1-c)\cdot 4 + bc\cdot 1 = 4b - 3bc$. This assumes that the driver's probability for Tails_1, conditional on $\mathit{First}\ \&\ \mathit{Bias}_1 = b\ \&\ \mathit{Bias}_2 = c$, equals b, which is a sign of thirding (compare equation (25) above). On the halfer account, $P(\mathit{Tails}_1/\mathit{First}\ \&\ \mathit{Bias}_1 = b) = b/(2-b)$. If the driver obeys halfing, her probability for reaching the desolate area conditional on $\mathit{First}\ \&\ \mathit{Bias}_1 = b\ \&\ \mathit{Bias}_2 = c$ is therefore not $1-b$, but $1 - b/(2-b)$. Similarly, the probability for getting home is $b/(2-b)\cdot(1-c)$, and for reaching the motel $b/(2-b)\cdot c$. The desirability of the lottery is therefore $b/(2-b)\cdot(1-c)\cdot 4 + b/(2-b)\cdot c\cdot 1 = (4b-3bc)/(2-b)$.[13]

So the results in section 4 are correct only if we assume thirding: *combined with thirding*, causal decision theory recommends a coin with bias 2/3, while evidential decision theory recommends a coin with bias 0.53.

Let's have another look at the conditional expected utilities that raised the puzzle in section 4. Conditional on $\mathit{Bias} = b$, the uniformity assumption makes it certain that $\mathit{Bias}_1 = \mathit{Bias}_2 = b$. So

$$\begin{aligned} CEU(\mathit{Bias} = b) &= P(\mathit{First}/\mathit{Bias}_1 = b)V(\mathit{First}\ \&\ \mathit{Bias}_1 = \mathit{Bias}_2 = b) \\ &\quad + P(\mathit{Second}/\mathit{Bias}_1 = b)V(\mathit{Second}\ \&\ \mathit{Bias}_2 = b) \\ &= \begin{cases} (1 - b/2)\cdot\frac{4b-3b^2}{2-b} + b/2\cdot(4-3b) = 4b - 3b^2 & \text{on halfing,} \\ \frac{1}{b+1}\cdot(4b-3b^2) + \frac{b}{b+1}\cdot(4-3b) = \frac{8b-6b^2}{b+1} & \text{on thirding.} \end{cases} \end{aligned}$$

The thirder function, with its maximum at around 0.53, is what we found in section 4; the halfer function is what we expected to find: it coincides exactly with our reasoning that the

13 For the second exit, we assumed that $V(\mathit{Second}\ \&\ \mathit{Bias} = b) = (1-b)4 + b1 = 4 - 3b$. This is correct on both halfing and thirding, since on either account the probability of *Tails*, given $\mathit{Second}\ \&\ \mathit{Bias} = b$, is b.

expected payoff from choosing bias b at every exit is $(1-b)\cdot 0+b\cdot(1-b)\cdot 4+b\cdot b\cdot 1 = 4b-3b^2$, which is maximal at $b = 2/3$. If the driver obeys halfing, her preferences therefore don't change between the start of the journey and the first exit. Accordingly, evidential decision theory no longer tells her to use the sub-optimal bias 0.53. So the puzzle of section 4 is raised by thirding; on the halfer perspective, it disappears.

Above I suggested that halfing and thirding may be regarded as consequences of two more general attitudes towards rational belief: halfing is motivated by the idea that rational agents should only revise their attitude towards a proposition if they receive evidence they regard as relevant to that proposition (at least if the proposition is certain not to change its truth-value over time); thirding, on the other hand, is motivated by the idea that an agent's degree of belief in a proposition should match the extent to which the proposition is supported by their present evidence. In cases like the absentminded driver, the two constraints pull in opposite directions.

To illustrate, suppose the driver is certain all along that she chooses a coin with bias 2/3. At the start of her journey, her degree of belief that she will get home is then $2/3 \cdot 1/3 = 2/9$. When she reaches the first exit, she does not receive any information that would allow her to rule out previously open possibilities in which she doesn't get home. If she had known at the start of the journey exactly what she would later learn at the first exit, her credence in getting home would still have been 2/9. This is why halfing entails that her belief should remain unchanged. On the other hand, consider the driver's evidence when she has reached the first exit. The evidence tells her that she is at one of the two exits, at each of which the chance of continuing is 2/3. Given that information, how likely is it that the driver will get home? Well, if she is at the first exit, the chance of getting home is $2/3 \cdot 1/3 = 2/9$; if she is at the second, the chance is 1/3. Going only by the driver's evidence, the net probability for getting home should therefore be some mixture of 2/9 and 2/3. Thirding says that this should also be the driver's degree of belief. More precisely, it says that if the driver is certain that she tosses a coin with bias b, then $P(\mathit{Home}) = P(\mathit{Home}/\mathit{First})P(\mathit{First}) + P(\mathit{Home}/\mathit{Second})P(\mathit{Second}) = b(1-b)\cdot 1/(b+1) + (1-b)\cdot b/(b+1) = 2(b-b^2)/(b+1)$, which is 4/15 for $b = 2/3$. Relative to her previous beliefs, the driver did not gain any information that would lend further support to getting home. Nevertheless, her new evidence, by itself, more strongly suggests that she will get home than her evidence at the start of her journey.

If the driver's beliefs change in accordance with thirding, it is also understandable why she would suddenly be happy to learn that she uses a coin with bias 0.53 rather than 2/3. The hypothesis that she uses a coin with bias 0.53 makes it slightly more probable that she is currently at the first exit and that she will reach the desolate area, but it also makes it more probable that she will get home if she is at the second exit. The latter effect outweighs the former, albeit not by very much.

It may help to imagine many repetitions of the driver's situation, with different choices

of the bias. If we look at all drivers who arrive at the first exit, those with bias 2/3 perform best. On the other hand, if we look at the drivers arriving at the second exit, those with a lower bias do better. Theorists who endorse both thirding and evidential decision theory therefore shouldn't accept that their recommendation has a worse average performance: that's true if we look at drivers at the first exit, but it's not true among drivers at the second exit – and for all the driver knows, she might be one of those.

Why, then, does thirding combined with causal decision theory recommend the coin with bias 2/3? Causal decision theory agrees that the optimal bias at the second exit is 0. But it doesn't agree that 2/3 would be optimal at the first exit. Since the driver's choice only controls her present action, the optimal bias at the first exit would be 1. The right compromise between the two values, taking into account the driver's ignorance about her location, is not 0.53, but 2/3.

To sum up: If the driver is given an option to randomize her acts by flipping a coin, then the answer to her decision problem depends not only on the choice of causal or evidential decision theory, but also on the choice between halfing and thirding. Two of the four combinations – evidential decision theory with thirding and causal decision theory with halfing – recommend choosing the coin with bias 2/3. By contrast, evidential decision theory with thirding recommends a bias of 0.53. While this answer may at first appear wrong, I have argued that it is actually in line with the ahistorical perspective on rational belief that motivates thirding.

The most curious of the four combinations is the last one: that of causal decision theory with halfing. Working it out will be the topic of the next section. Before we turn to that, I want to take another look at the coin-free scenario of sections 2 and 3. Recall that when the driver's only options where to continue or to leave, causal decision theory recommended a state of indecision in which the driver is 2/3 inclined towards continuing. The argument for the value 2/3 was analogous to the argument for choosing bias 2/3 in section 3. In particular, we assumed (in equation (11)), that if the driver knows that she will continue with probability c, then $P(\mathit{Continue}_1/\mathit{First}) = P(\mathit{Continue}) = c$. Did we presuppose thirding already in section 3? Interestingly, we did not. Let me explain.

We saw in section 3 that if the driver is in a state of indecision, then she should treat her choice as if it were the outcome of a chance process, with probabilities set in accordance with the her beliefs about what she will do. Let's imagine for the sake of vividness that in fact a little coin is tossed inside the driver's head to determine what she will do. As before let's assume that tails means continue, heads leave. The coin's bias c towards tails is set to equal the driver's equilibrium probability that she will continue: $P(\mathit{Continue}) = c$.

Now assume the driver is aware of these facts. Then $P(\mathit{Continue}) = P(\mathit{Tails}) = c$. By equation (28), this is a sign of thirding. Halfing would require that $P(\mathit{Tails}_1) = c$, and as we saw in section 5, $P(\mathit{Tails}_1)$ and $P(\mathit{Tails})$ can't be equal unless $P(\mathit{Second}\ \&\ \mathit{Heads}_2) = 0$.

So the driver can't possibly be a halfer about the coin in her head!

From a halfer perspective, this is odd. Why should the driver obey thirding about the coin in her head, but halfing about the coin in her hand? How could it make an epistemic difference whether the coin is tossed inside or outside the driver's head? But there really is an important difference. The difference does not lie in the location of the coins, but in the fact that the bias of the coin in the head is controlled by the driver's beliefs.

Consider the evolution of the driver's beliefs, assuming her only options are *Continue* and *Leave*. We can imagine that the driver has already figured out at the start of her journey that the deliberation equilibrium in the problem she will face at any exit lies at $P(\mathit{Continue}) = 2/3$. Hence she knows that a coin with bias 2/3 will be tossed in her head at the first exit, and again at the second exit if the first one lands tails. So $P(\mathit{Tails}_1) = 2/3$, and $P(\mathit{Bias}_1 = \mathit{Bias}_2 = 2/3) = 1$. Now she reaches the first exit. If she obeys the principle of doxastic conservatism that motivates halfing, she will then still be certain that a coin with bias 2/3 is tossed at any exit, and her credence in Tails_1 will still be 2/3. By the symmetry assumption, the probability of Tails_1 is evenly divided between *First* and *Second*. So $P(\mathit{First}\ \&\ \mathit{Tails}_1) = 1/3$. Moreover, $P(\mathit{Tails}_2/\mathit{Tails}_1) = 2/3$, and so $P(\mathit{Second}\ \&\ \mathit{Tails}_2) = 2/3 \cdot 1/3$. Since *Tails* is equivalent to $(\mathit{First}\ \&\ \mathit{Tails}_1) \vee (\mathit{Second}\ \&\ \mathit{Tails}_2)$, it follows that $P(\mathit{Tails}) = 5/9$ – in line with the halfer formula (28). But now the driver's credence is not in equilibrium: at $P(\mathit{Tails}) = (\mathit{Continue}) = 5/9$ and $P(\mathit{Continue}_2) = 2/3$, continuing has a higher expected utility than leaving. In the process of deliberation, the driver's probability for *Continue* will therefore increase. Deliberation will take her to the thirder function with $P(\mathit{Continue}) = 2/3$. And so the coin with bias 2/3 will indeed be tossed, just as the driver anticipated at the start of her journey.

Compare the situation where the driver can decide to toss a coin of any chosen bias. Assume again that she has figured out in advance that she will decide to use the coin with bias 2/3; that is, she has figured out that the equilibrium in her decision problem lies at $P(\mathit{Bias}{=}2/3) = 1$. As before, upon reaching an exit, the conservative belief update will result in a halfer function, with $P(\mathit{Continue}) = 5/9$. But *Continue* here doesn't stand for one of the driver's options. Her options are given by the propositions $\mathit{Bias}{=}b$. To be sure, one of her options, $\mathit{Bias}{=}1$, can be identified with the option to continue, but since the driver is certain that she chooses bias 2/3, $P(\mathit{Bias}{=}1)$ is 0, not 5/9. In fact, both before and after the belief update, $P(\mathit{Bias} = 2/3) = 1$. So deliberation does not change her beliefs, which will keep conforming to the halfer equations.

Here, then, we have a partial response to the charge, mentioned in section 3, that allowing for states of indecision is in effect to allow for randomized options. From the halfer perspective, the situation in which the driver has a stochastic tie-breaker to resolve states of indecision is not at all the same as the situation in which she can actively choose a randomized option.

Something similar is true even from the thirder perspective. At the end of section 3, we saw that the state of indecision with $P(\mathit{Continue}) = 2/3$ (and thirder attitudes towards the internal coin) is also an equilibrium solution in evidential decision theory. But when we gave the driver the option to choose a coin, evidential decision theory combined with thirding recommended a coin with bias 0.53. The reason is that in the second case, the driver's beliefs about where she is and what she will do at the other exit are affected by the hypothesis that she chooses a given bias; by contrast, in the state of indecision her beliefs about these matters are independent of the hypothesis that she will continue.

7 Useless coins

Let's return to the scenario where the driver can choose a coin. We know what she ought to do on three of the four combinations of causal vs. evidential decision theory with halfing vs. thirding. The remaining combination is that of causal decision theory and halfing.

At the beginning of section 6, we already computed the utility of the relevant lotteries from the halfer perspective:

$$V(\text{First } \& \ \mathit{Bias}_2 = c \ \& \ \mathit{Bias}_1 = b) = (4b - 3bc)/(2 - b); \tag{29}$$

$$V(\text{Second } \& \ \mathit{Bias} = b) = 4 - 3b. \tag{30}$$

What is left is to compute the halfer probability of the states, that is, of *Second* and states of the form $\mathit{First} \ \& \ \mathit{Bias}_2 = c$. Beginning with the latter, probability theory says that $P(\mathit{First} \ \& \ \mathit{Bias}_2 = c) = P(\mathit{First}/\mathit{Bias}_2 = c)P(\mathit{Bias}_2 = c)$. Halfing entails that $P(\mathit{First}/\mathit{Bias}_1 = c) = 1 - c/2$ (see equation (24)). By uniformity, we can infer that $P(\mathit{First}/\mathit{Bias}_2 = c) = 1 - c/2$. So

$$P(\mathit{First} \ \& \ \mathit{Bias}_2 = c) = P(\mathit{Bias}_2 = c)(1 - c/2). \tag{31}$$

As for $P(\mathit{Second})$, note that $P(\mathit{Second}/\mathit{Bias}_2 = c) = 1 - P(\mathit{First}/\mathit{Bias}_1 = c) = c/2$; so

$$P(\text{Second}) = \sum_{c \in B} P(\mathit{Bias}_2 = c)c/2. \tag{32}$$

The resulting expected utilities work out as follows:

$$EU(\mathit{Bias} = b) = \sum_{c \in B} P(\mathit{Bias}_2 = c)\frac{3bc^2 + 3b^2c - 20bc + 8c + 8b}{4 - 2b}. \tag{33}$$

(The diagonal, with $b = c$, is again the CEU function $4b - 3b^2$.)

As in section 4, where we considered the thirder version of these formulas, what the driver should do depends on what she believes about what she will do: the expected

utility of $Bias{=}b$ varies with the distribution over $Bias_2$. However, this time $Bias{=}2/3$ is not a stable choice. If the driver is confident that $Bias_2 = 2/3$, then the expected utility of $Bias = 2/3$ is 4/3, while the expected utility of $Bias = 1$ is 5/3.

In general, the more the driver is certain that she chooses bias c, the more the expected utility of choosing bias b converges to

$$\frac{3bc^2 + 3b^2c - 20bc + 8c + 8b}{4 - 2b}.$$

For $c < 4/5$, this function is maximal at $b = 1$; for $c > 4/5$, it is maximal at $b = 0$; and for $c = 4/5$ it is maximal at both $b = 1$ and $b = 0$. That is, for every available coin, if the driver is confident that she chooses that coin, then choosing either bias 1 or bias 0 (or both) has greater expected utility. And of course, if the driver is confident that she chooses bias 1, then choosing bias 0 has greater expected utility, and *vice versa*. There is no stable option: the driver should be undecided which coin to choose!

A well-known result in game theory states that every finite game has a Nash equilibrium if the players are allowed to use randomized ("mixed") strategies. By essentially the same reasoning one can show that every finite decision problem has a stable solution if randomisation is allowed.[14] The proof assumes that the expected utility of a lottery in a given state equals the corresponding mixture of the utility of the pure acts in that state. For example, if an agent tosses a fair coin to decide between an act A whose payoff in a given state of nature is a, and an act B with payoff b, then it is assumed that the randomized strategy has expected payoff $(a + b)/2$ in that state. This assumption is false if randomization is punished. It is also false in the present case of the absentminded driver, although for a very different reason.

To understand the reason, note that the driver would like her coin at the first exit to land tails, as heads leads to the worst possible outcome of spending the night in the car. But if the driver follows halfing, then by (25), $P(Tails_1/First\ \&\ Bias_1{=}b) = b/(2 - b)$, which is less than b whenever b is strictly between 0 and 1. For example, if $b = 1/2$, then $b/(2 - b) = 1/3$. So on the assumption that the driver is at the first exit, tossing a fair coin gives her only a subjective probability of 1/3 for reaching the second exit. It's as if the driver had an unduly pessimistic attitude towards randomizing at the first exit. By contrast, at the second exit, where leaving would be much better than continuing, the driver's credence in heads does not exceed the objective chance: by the neutral assumption (26), $P(Tails_2/Tails_1\ \&\ Bias_2 = b) = b$. So the "pessimism" about randomizing at the first exit is not compensated by optimism about the second exit.

For a concrete illustration, imagine the driver is 4/5 confident that she will choose $Bias = 0$ and 1/5 that she will choose $Bias = 1$. There are then only three states with

14 See [Harper 1986] and [Skyrms 1990] on the connection between Nash equilibria in game theory and deliberation equilibria in causal decision theory.

positive probability: *First* & $Bias_2 = 1$, *First* & $Bias_2 = 0$, and *Second*. Their probabilities are 2/5, 1/5, and 2/5, respectively. Assuming *First* & $Bias_2 = 1$, the expected utility of choosing bias b equals the probability that this choice will make the driver continue, which is $b/(2-b)$. Similarly, the expected utility of $Bias = b$ in the state *First* & $Bias_2 = 0$ is $4 \cdot b/(2-b)$. In *Second*, it is $b \cdot 1 + (1-b) \cdot 4 = 4 - 3b$. So the decision matrix for the three options $Bias = 1$, $Bias = 0$, and $Bias = 1/2$ looks as follows.

	First & $Bias_2 = 1$	*First* & $Bias_2 = 0$	*Second*
$Bias = 1$	1	4	1
$Bias = 0$	0	0	4
$Bias = 1/2$	1/3	2/3	5/2

The first two options have equal expected utility, since $(2/5) \cdot 1 + (1/5) \cdot 4 + (2/5) \cdot 1 = 8/5 = (2/5) \cdot 4$. However, the third option only has expected utility $(2/5) \cdot (1/3) + (1/5) \cdot (2/3) + (2/5) \cdot (5/2) = 19/15$. In general, every choice of bias other than 1 and 0 has an expected utility less than 8/5.

Since non-trivial choices of a bias always look worse than either $Bias = 0$ and $Bias = 1$, the driver's deliberation dynamics will carry her to a state in which she is certain that she will choose either bias 0 or bias 1. We've already found the equilibrium: it lies at $P(Bias = 1) = 4/5$ and $P(Bias = 0) = 1/5$. Here all options with positive probability have equal expected utility. Of course, tossing a coin with bias 0 or 1 is pointless: $Bias = 1$ means to continue, $Bias = 0$ to leave. In the equilibrium state, the driver therefore deems her pure options to be equally good, while tossing a coin to choose between them is strictly worse. In the end, she is torn only between continuing and leaving – exactly like in section 3.

That the equilibrium inclination towards continuing is 4/5 rather than 2/3 is due to the (revised) uniformity assumption that the driver is certain that she uses the same bias at every exit. As before, the absence of a pure solution makes that assumption implausible. If instead we assume that the driver treats different tie-breakings of her state of indecision as independent, the equilibrium state lies at $P(Bias = 1) = 2/3$ and $P(Bias = 0) = 1/3$.

Here is why. In equilibrium, the driver knows that if she is at the first exit and ends up continuing, then she'll arrive at the same state of indecision again, with the same equilibrium probabilities P. Given independence of the tie-breakings, $P(Bias = b) = P(Bias_1 = b/First) = P(Bias_2 = b)$. By halfing, $P(Tails_1/First \;\&\; Bias_1 = b) = b/(2-b)$. So $P(Continue_1/First) = \sum_{b \in W} P(Bias = b)b/(2-b)$. Let's abbreviate this quantity as t; that is, $t = \sum_{c \in B} P(Bias_1 = c)c/(2-c)$. Now we argue as in section 3. *First* divides into *First* & $Continue_1$ and *First* & $Leave_1$. By the symmetry assumption, $P(First \;\&\; Continue_1) = 1 - P(First)$. Moreover, $P(First \;\&\; Leave_1) =$

$P(Leave_1/First)P(First) = (1 - P(Continue_1/First))P(First) = (1-t)P(First)$. So $P(First) = 1 - P(First) + (1-t)P(First)$, which resolves to $P(First) = 1/(t+1)$. So the equilibrium probabilities of the states are

$$P(First \,\&\, Bias_2{=}c) = P(Bias{=}c)/(t+1); \tag{34}$$

$$P(Second) = t/(t+1). \tag{35}$$

And thus

$$EU(Bias{=}b) = \sum_{c \in B} \left(\frac{P(Bias{=}c)}{t+1} \cdot \frac{4b - 3bc}{2-b} \right) + \frac{t}{t+1} \cdot (4 - 3b). \tag{36}$$

In equilibrium, all options with positive probability must have maximal expected utility. The only such equilibrium, as far as I can tell, is given by the probability function that assigns 2/3 to $Bias = 1$ and 1/3 to $Bias = 0$.

So the driver should be 2/3 inclined towards continuing and 1/3 towards leaving, and she should be certain that she won't use any of her coins. Combined with halfing, causal decision theory advises against randomization and in favour of indecision. Once again, that indicates that indecision is not the same as randomization.

8 Conclusions

The puzzle of the absentminded driver is a treasure trove of interesting results. Let me summarize our main observations.

In sections 2 and 3, we looked at the driver's predicament assuming her only options are *Leave* and *Continue*. By the lights of causal decision theory, neither choice is a stable solution. The driver's deliberation should end in a state of indecision where she is 2/3 inclined towards *Continue* and 1/3 towards *Leave*. In evidential decision theory, there are two solutions to the driver's problem: she can decide to continue, or she can be in a state of indecision where she is 2/3 inclined towards *Continue*. The difference between halfing and thirding does not matter here: even if the driver starts out with halfer attitudes towards the resolutions of her indecision, the process of deliberation will turn them into thirder attitudes, as explained in section 6.[15]

In sections 4 to 7, we assumed that the driver can make her decision by tossing a coin. Here the answer depends not only on the choice between causal and evidential decision theory, but also on that between halfing and thirding. Two of the four combinations

15 One noteworthy aspect of the puzzle that I have not discussed concerns the form of the deliberation dynamics. In all versions of the puzzle, the dynamics can't go by the simple models described in [Skyrms 1990]. For example, these models would not preserve the symmetry assumption; they also don't take into account how the driver's choices at the two exits can become decorrelated through the process of deliberation.

– thirding with causal decision theory and halfing with evidential decision theory – recommend tossing a coin with bias 2/3. Since that is *prima facie* the optimal choice, this result mirrors arguments in [Arntzenius 2002] and [Briggs 2010] to the effect that thirders should be causal decision theorists and halfers evidential decision theorists. However, I have argued that the other two combinations are perfectly acceptable as well. The combination of halfing with causal decision theory recommends a state of indecision whose expected payoff coincides with that of choosing a coin with bias 2/3 on the thirder account. Thirding with evidential decision theory recommends a coin with bias 0.53. I have argued that this is indeed the most desirable choice given the ahistorical perspective on rationality that motivates thirding.

Along the way, we have encountered some other interesting facts. For example, we saw that the driver's rational degrees of belief about her coins cannot consistently match the known objective chances: on every account, either $P(\mathit{Tails}/\mathit{Bias}_1 = \mathit{Bias}_2 = c) \neq c$ or $P(\mathit{Tails}_1/\mathit{Bias}_1 = \mathit{Bias}_2 = c) \neq c$. As [Titelbaum 2012] points out, this casts doubt on the so-called "double-halfer" solution to the Sleeping Beauty problem. It also raises problems for Skyrms's [1984] formulation of causal decision theory in terms of expected conditional chance. On Skyrms's account, the expected utility of an option A is defined as

$$EU(A) = \sum_{S \in W} V(S) \sum_{x} x \cdot P(Ch(S/A) = x),$$

where W is a partition of possibilities whose members have uniform utility (meaning that for any $S \in W$ and any proposition B compatible with S, $V(S \,\&\, B) = V(S)$) and $Ch(S/A) = x$ is the proposition that the objective chance of S conditional on A equals x.[16] This is plausible as long as the agent satisfies a conditional version of the Principal Principle, so that her credence in S given A, on the assumption that $Ch(S/A) = x$, equals x. The absentminded driver who follows halfing does not satisfy that principle.

Another fact brought out by the absentminded driver is the difference between indecision and randomization: if agents are equipped with a stochastic tie-breaker to resolve states of rational indecision, their attitudes towards the outcomes can diverge from the attitudes they would have towards the outcomes of an explicitly chosen randomisation device. Relatedly, we saw that even if there is no punishment on randomization, allowing for randomized strategies does not guarantee that one of the options is a rational choice – or that every finite game has a Nash equilibrium. This supports the idea that mixed strategies in game theory might be better understood not as choices of a randomized option, but as states of indecision (compare [Aumann and Brandenburger 1995]).

16 Skyrms's own formulation is superficially different: he defines $EU(A)$ as $\sum_K P(K) \sum_C Ch_K(C/A)V(C)$, where K ranges over chance hypotheses, C over "consequences" and $Ch_K(C/A)$ is the conditional chance of C given A according to K.

References

Frank Arntzenius [2002]: "Reflections on Sleeping Beauty". *Analysis*, 62: 53–62

— [2003]: "Some problems for conditionalization and reflection". *Journal of Philosophy*, 100: 356–370

— [2008]: "No Regrets, or: Edith Piaf Revamps Decision Theory". *Erkenntnis*, 68: 277–297

Robert Aumann and Adam Brandenburger [1995]: "Epistemic Conditions for Nash Equilibrium". *Econometrica*, 63: 1161–1180

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

Darren Bradley [2011]: "Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty". *British Journal for the Philosophy of Science*, 62: 323–342

Rachael Briggs [2010]: "Putting a Value on Beauty". In T. Szabo Gendler and J. Hawthorne (Eds.) *Oxford Studies in Epistemology,* vol Vol. 3. Oxford: Oxford University Press

Alex Byrne and Alan Hájek [1997]: "David Hume, David Lewis, and Decision Theory". *Mind*, 106: 411–728

Cian Dorr [2002]: "Sleeping Beauty: In defence of Elga". *Analysis*, 62: 292–296

Andy Egan [2007]: "Some Counterexamples to Causal Decision Theory". *Philosophical Review*, 116: 93–114

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

David Etlin [2008]: "Desire, Belief, and Conditional Belief". PhD Dissertation

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

William Harper [1986]: "Mixed Strategies and Ratifiability in Causal Decision Theory". *Erkenntnis*, 24: 25–36

Patrick Hawley [2013]: "Inertia, Optimism and Beauty". *Noûs*, 47(1): 85–103

Terry Horgan [2004]: "Sleeping Beauty Awakened: New Odds at the Dawn of the New Day". *Analysis*, 64: 10–21

— [2008]: "Synchronic Bayesian Updating and the Sleeping Beauty Problem: Reply to Pust". *Synthese*, 160: 155–159

Terry Horgan and Anna Mahtani [2013]: "Generalized Conditionalization and the Sleeping Beauty Problem". *Erkenntnis*, 78(2): 333–351

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

James Joyce [1999]: *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press

— [2002]: "Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions". *Philosophical Studies*, 110: 69–102

— [2012]: "Regret and instability in causal decision theory". *Synthese*, 187(1): 123–145

David Lewis [1980]: "A Subjectivist's Guide to Objective Chance". In Richard Jeffrey (Ed.), *Studies in Inductive Logic and Probability* Vol. 2, University of California Press, Berkeley.

— [1981]: "Causal Decision Theory". *Australasian Journal of Philosophy*, 59: 5–30

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

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

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

Reed Richter [1985]: "Rationality, Group Choice and Expected Utility". *Synthese*, 63: 203–232

Leonard Savage [1954]: *The Foundations of Statistics*. New York. Wiley

Wolfgang Schwarz [2012a]: "Changing Minds in a Changing World". *Philosophical Studies*, 159: 219–239

— [2012b]: "Diachronic rationality". Manuscript

Brian Skyrms [1984]: *Pragmatics and Empiricism*. Yale: Yale University Press

— [1990]: *The Dynamics of Rational Deliberation*. Cambridge (Mass.): Harvard University Press

Michael G. Titelbaum [2012]: "An Embarrassment for Double-Halfers". *Thought*, 1(2): 146–151

Paul Weirich [1985]: "Decision Instability". *Australasian Journal of Philosophy*, 63: 465–472