Belief update across fission*

Wolfgang Schwarz
Draft, 05 Nov 2012

Abstract. When an agent undergoes fission, how should the beliefs of the fission products be related to the pre-fission beliefs? This question is central for the Everett interpretation of quantum mechanics, but it is of independent philosophical interest. Among other things, fission scenarios provide counterexamples to several popular accounts of belief update, showing that "self-locating" information can affect the probability of uncentred propositions even if agents are never uncertain, conditional on an uncentred world, about their location within the world. I present an alternative update rule for centred beliefs that gives sensible results in cases of fission, and show that it is supported by the same considerations that support standard conditioning in the traditional framework of uncentred propositions.

1 The problem

Fred's home planet, *Sunday*, is surrounded by two moons, *Monday* and *Tuesday*. Tonight, while Fred is asleep, his body will be scanned and destroyed; then a signal will be sent to both Monday and Tuesday where he will be recreated from local matter.

A lot of ink has been spent on how to describe scenarios like this. Can people survive teleportation? Will Fred find himself both on Monday and on Tuesday? Which of the persons awakening on the two moons, if any, is identical to the person going to sleep on Sunday? In this paper, I want to look at a different question: what should Fred's successors believe when they awaken on Monday and on Tuesday? More precisely, how should their beliefs be related to Fred's beliefs before he went to sleep on Sunday?

The issues are independent. For the present topic, it does not matter whether the two "successors" are identical to Fred, either temporally or absolutely. If you think Fred would not survive the double teleportation and that two new persons come into existence on Monday and Tuesday, it still makes sense to ask how the beliefs of these persons should be related to the beliefs of Fred. Imagine you are designing a species of intelligent

* This paper is part of an ancestor of [Schwarz 2012] circulated in 2008/2009. Upon request of a referee, I removed discussion of fission cases from that paper, leaving the proposed update rule incomplete. Here is the completion. For comments and discussion, I thank David Chalmers, Alan Hájek, Christopher Meacham, Paolo Santorio, Michael Titelbaum and the audience of a PhilSoc seminar at the Australian National University in February 2008.

amoebae that regularly undergo fission. What update process would you implement for the amoebae's beliefs so as to make optimal use of the previously collected information?

The story of Fred gets more interesting if he doesn't know what is going to happen. Suppose Fred learns that a fair coin will be tossed and the signal to Tuesday is cut iff the coin lands heads. *We* know that the coin lands tails and the signal isn't cut, but Fred doesn't know this. How confident should his successors be that they are on Monday? What should they believe about the outcome of the coin toss? What should they believe if they learn that they are on Monday?

Fred's predicament is not as far-fetched as it may at first appear. According to the Everett interpretation of quantum mechanics, what are commonly regarded as chance events are really branching events in which every possible outcome determinately occurs, although on different "branches" of the universe. If a particle is in a superposition of two locations (call them *Monday* and *Tuesday*) and you set up a detector to find out where it is, one of your successors will find the detector point at Monday, the other at Tuesday. Thus if you give intermediate credence to the Everett interpretation, your beliefs are divided between a branching hypothesis and a non-branching hypothesis, just like Fred's.

2 Conditioning and self-location

When Fred's Monday successor wonders whether he is on Monday or on Tuesday, what he lacks is not objective information about the universe, but "self-locating" information about himself. He knows that the universe contains a Monday successor and a Tuesday successor. What he doesn't know is whether he himself is the former or the latter.

I will model this kind of ignorance by assuming that degrees of belief attach to *centred propositions* whose truth value can vary between different locations within a world. A natural suggestion, due to [Lewis 1979], is to identify centred propositions with properties: Fred's successors give some degree of belief to *being on Monday* and some to *being on Tuesday*. Suitably regimented, the space of properties forms a Boolean algebra, closed under conjunction, disjunction and negation. To keep distracting technicalities at bay, I will pretend for most of this paper that this algebra is finite and hence isomorphic to the full powerset algebra on its atoms. These atoms are known as *centred (possible) worlds*. Intuitively, a centred world represents a maximally specific way a thing might be. Two centred worlds are *worldmates* if they can be instantiated in the same universe. Traditional *uncentred propositions* are propositions that never distinguish between worldmates; a maximally specific uncentred proposition is an *uncentred (possible) world*.[1]

1 Many authors take uncentred worlds as primitive and define centred worlds as triples of an uncentred world, an individual and a time. I think it is more natural to start with centred worlds or centred propositions. This also avoids the obvious problems for the triples account if the relevant individual is a time-traveler or a multi-headed dragon.

There are other ways to model self-locating beliefs. On one alternative, objects of belief are factored into uncentred "contents" and centred "modes of presentation" (see [Perry 1979], [Bradley 2007]). Another alternative postulates haecceitistic propositions involving the relevant subject and time, so that the uncertainty of Fred's Monday successor might concern the uncentred proposition that individual s is on Monday at time t – a proposition Fred's Tuesday successor cannot even entertain, for lack of direct acquaintance with s (see [Chisholm 1967], [Stalnaker 2008]). The proposal I will make can be translated into these other frameworks, but I will not spell out the translations.

Ordinary objects trace a path through the space of centred worlds. Consider Napoleon. Initially, in 1769, Napoleon had properties like *living on Corsica* and *being called Napoleone*; later he lost these properties and acquired new ones, until eventually his properties included *being 51 years old* and *living on St. Helena*. At each point in his life, the totality of Napoleon's properties constitute a centred world. The complete history of Napoleon is a sequence of possible worlds: a trajectory through logical space. On the assumption that the space of centred worlds is finite, every non-terminal point on Napoleon's trajectory has a determinate successor, unlike the points on the real number line.

The familiar Bayesian rule of *conditioning* specifies how an agent's degrees of belief should evolve over time. Let w_1, w_2 be subsequent positions on an agent's trajectory, and suppose at w_2 the agent undergoes a learning event whose direct impact is to confer certainty on some proposition E. Conditioning says how the agent's degrees of belief in other propositions should change: if P_1 is the belief function at w_1 and P_2 the belief function at t_2, then for every proposition A,

$$P_2(A) = P_1(A/E) = \frac{P_1(A \wedge E)}{P_1(E)}, \text{ provided } P_1(E) > 0.$$

The status of this rule is controversial. Some doubt that there are diachronic constraints on rational belief at all. Others question whether learning events can be modeled as conferring certainty on some proposition E. For the present paper, I want to set aside these objections. I will assume that conditioning is the right way for an ideally rational agent to change her mind, if all relevant propositions are uncentred.

But what if we allow for uncentred propositions? Suppose at w_1 you believe that A is true at your present position in logical space. Later, at w_2, you learn that E is true at your new position. Should this make you believe that your *new* position satisfies A to the extent that you previously believed that your *old* position satisfies A conditional on satisfying E? Clearly not, unless you have reason to believe that the two positions agree with respect to A and E. Conditioning does not take into account the possibility that agents might change their position in logical space.

A common reaction to this problem is to restrict conditioning to uncentred propositions and add a new rule for self-locating propositions (see e.g. [Piccione and Rubinstein 1997], [Halpern 2006], [Meacham 2008], [Titelbaum 2008], [Kim 2009] and [Briggs 2010]). The new update process then goes more or less as follows. Let P_1^* be P_1 restricted to uncentred propositions, representing the agent's uncentred belief state at w_1. At the new state w_2, this function gets conditioned on the uncentred information acquired at w_2. I will use '$\Diamond A$' to denote the strongest uncentred proposition entailed by a proposition A.[2] So $\Diamond E$ is the uncentred information acquired at w_2. Let P_2^* be P_1^* conditioned on $\Diamond E$. Every uncentred world to which P_2^* assigns positive probability contains at least one point at which E is true. Suppose they each contain exactly one such point. In this case, I will say that the evidence E is *sufficient for self-location* (relative to P_2^*): conditional on any uncentred world, E tells the agent exactly where she is. The new centred credence P_2 can then be defined by assigning the P_2^* probability of every uncentred world u to the corresponding centred world in u where E obtains. In effect, the one-one map between open centred and uncentred worlds allows the agent to "translate" centred propositions into uncentred propositions; the classical evolution of uncentred probabilities (from P_1^* to P_2^*) thereby also settles the new centred probabilities.

What if E is not sufficient for self-location? In this case, a standard procedure is to invoke a principle of self-locating indifference according to which the agent should give equal credence to each E point within the same uncentred world. Evenly dividing the P_2^*-probability of each uncentred world among its E centres leads to halfing in the Sleeping Beauty problem. A more popular alternative, which supports thirding, is to assign the whole P_2^*-probability of each uncentred world to all its E centres and then renormalise the probability distribution.

I will call models of this type *uncentred conditioning models*. Notice that in these models, the new probability of uncentred propositions depends not only on the new evidence E, but also on the previous uncentred probabilities. By contrast, the previous self-locating beliefs are discarded: all that matters to P_2 is P_1^* and E.

This has striking consequences in scenarios involving fission. Suppose you presently assign credence 1/2 to the Everett interpretation: $P_1(\textit{Everett}) = 1/2$. Then you carry out a measurement on a system in superposition between "Monday" and "Tuesday". The Everett worlds in your belief space all contain a branch on which the outcome is Monday and one on which it is Tuesday. Among non-Everett worlds, your credence is divided between worlds where the outcome is Monday only and worlds where it is Tuesday only. Suppose you now observe Monday. Following the uncentred conditioning models, we first condition P_1^* on $\Diamond \textit{Monday}$, i.e. on the proposition that your evidence *Monday* is true at some point in the universe. This rules out all and only the non-Everett worlds in which the outcome is Tuesday. So $P_2^*(\textit{Everett}) > 1/2$. Assuming your evidence is sufficient for self-location, $P_2(\textit{Everett}) = P_2^*(\textit{Everett})$. More specifically, if you previously

2 $\Diamond A$ is true at a world w iff A is true at some worldmate of w; the worldmate relation is the diamond's accessibility relation.

assigned equal credence to non-Everett Monday worlds and non-Everett Tuesday worlds, then $P_2(\mathit{Everett}) = 2/3$. Exactly the same would have happened if you had observed the alternative outcome Tuesday. By repeatedly carrying out measurements, you would become more and more confident in the Everett hypothesis, no matter what outcomes you observe. This does not look rational.[3]

In general, if a proposition A is certain not to change its truth-value, and an agent knows in advance that her new evidence will be one of $E_1, \ldots, E_n$, then rationality should not demand her degree of belief in A to increase no matter which of $E_1, \ldots, E_n$ she learns. This requirement of *dynamic stability* can be supported by Dutch books and considerations of expected accuracy. In the Everett scenario, you would initially regard as fair a deal that pays $3 if the Everett hypothesis is false and costs $3 if it is true. Afterwards, you would regard as fair a deal that pays $2 if the hypothesis is true and otherwise costs $4, irrespectively of what you observe. You are guaranteed to lose $1.[4] Similarly, we will see in section 6 that your new credence function will have lower expected accurate by the lights of your previous beliefs than a function which assigns probability 1/2 to the Everett hypothesis.

The source of the problem is the assumption that if an agent's evidence is sufficient for self-location, then their degrees of belief in uncentred propositions should evolve by standard conditioning on their uncentred evidence. The assumption does sound plausible: if your evidence is sufficient for self-location, then your only uncertainty concerns which uncentred world you inhabit; so it should be enough to consider what your new evidence has to say on this matter. But that thought is mistaken. Even if the evidence is sufficient for self-location, its self-locating aspect can be relevant to uncentred propositions. When you see observe the outcome Monday, what you learn is not just a fact about the universe as a whole, but also that you are presently looking at a Monday outcome. Conditioning on this information would exclude Tuesday possibilities in Everett worlds just as much as

3 The present phenomenon has been discussed in the literature on Sleeping Beauty, where it is usually understood as a challenge for thirders, by the supposed analogy between the Everett scenario and Sleeping Beauty (see e.g. [Lewis 2007a], [Bradley 2011b]). The more immediate consequence that uncentred probabilities should not always evolve by conditioning on uncentred information is noted in [Greaves 2007a].

4 [Briggs 2010] presents a purported proof that the uncentred conditioning rule (in its thirder version) is immune to diachronic Dutch books. Her reasoning goes as follows. If there *were* a Dutch book B for an agent who follows this rule, then we could convert B into a Dutch book B^* for an imaginary agent with only uncentred beliefs who updates by standard conditioning. But the latter is impossible by a result in [Skyrms 1987] (falsely attributed by Briggs to [Teller 1973]). If we apply Brigg's conversion recipe to the present Dutch book B, we get $B^* = B$. Since the imaginary agent with belief function P_1^* assigns credence 1/2 to *Everett*, she regards the first bet as fair. After conditioning on either $\Diamond\mathit{Monday}$ or $\Diamond\mathit{Tuesday}$, she also regards the second bet as fair. However, *pace* Briggs, this pair of bets does not constitute a Dutch book against the imaginary agent. The problem is that $\Diamond\mathit{Monday}$ and $\Diamond\mathit{Tuesday}$, unlike *Monday* and *Tuesday*, are not mutually exclusive: they are both true at Everett worlds.

in non-Everett worlds. The uncentred conditioning models let you only condition on the much weaker proposition that the universe contains some point or other where Monday is true, which rules out none of the Everett worlds.

The problem isn't limited to cases of fission. All that's needed is that several possible evidence propositions are true within the same uncentred world. Here is a general template. Your initial credence P_1 is divided between three uncentred propositions X, Y and Z. You know that you are going to observe either M or T. X worlds contain a point where M is true and another point where T is true. Y worlds only contain an M point, Z worlds a T point. If your uncentred beliefs evolve by conditioning on uncentred evidence, then your credence in X will increase no matter whether you learn M or T.

Can we spell out an update rule that lets you condition on *all* your evidence, including your centred evidence, while also taking into account that you change position in logical space?

3 Shifted conditioning

Like standard conditioning, the rule I will present determines the agent's beliefs at w_2 based on the previous beliefs at w_1 together with the new evidence E. In this context, we must assume that w_2 is an immediate successor of w_1. If we want to determine an agent's credence at w_2 based on their credence at w_1 and the evidence at w_2, we cannot in general allow there to be further points in between w_1 and w_2 – at least no points at which relevant evidence arrives. For if the agent receives evidence in between w_1 and w_2, then their credence at w_2 should be sensitive to this intermediate evidence, some of which might already have turned false by the time of w_2, so we can't simply count the intermediate evidence as part of E.

Now recall that centred worlds are maximally specific possibilities. A centred world does not only contain information about the present, but also about the past and the future. If w is Napoleon's position in logical space on New Year's eve 1805, then w settles not only what Napoleon is doing right at that time, but also what everybody else is doing at every other time; it entails that Napoleon will die on St. Helena in 1824, and that the world centred on this event lies on the same personal trajectory as w. More importantly, if a world w lies on some trajectory, then it fully determines which other worlds, in which order, lie on the same trajectory.

Given all this, there is a rather obvious way to amend standard conditioning. We simply need to add an operation to the update process that shifts the probability of all previously possible worlds to their successors on the relevant trajectory. This shifted probability is then conditioned on the new evidence.

To see how this works, imagine an omniscient agent whose credence at w_1 is concentrated on the single world w_1. The update then simply moves her credence to the successor of

w_1; the agent remains omniscient without receiving any new information. If instead her initial credence is divided 5:4:1 between three worlds w_1, w_2, w_3, and the new evidence rules out none of their successors w'_1, w'_2, w'_3, then the new credence is divided 5:4:1 between these successors. If the evidence rules out w'_3, the new credence is divided 5:4 between w'_1 and w'_2. And so on. Note that the update does not invoke the agent's *actual* change in location, but the possible changes foreseen by the agent's beliefs. If you fall asleep or enter an indeterministic time machine, not knowing whether you will awaken before or after midnight, then your new beliefs will be divided between it being before and after midnight, irrespective of how much time has actually passed.

Something like the two-stage process of shifting and conditioning has long been used in computer science.[5] In philosophy, it has only recently been rediscovered by Christopher Meacham [2010] and myself [2012].

Let me spell out the new rule a bit more precisely. Suppose, for now, that every world with positive probability at w_1 has exactly one successor. That is, every such world lies on a trajectory where it is succeeded by a unique other world. (This assumption will soon be dropped.) Define the shifting operator '$\succ$' (read 'next') so that $\succ w$ is true at a world v iff w is a successor of v. More generally, for any proposition A, let $\succ A$ be true at v iff A is true at some successor of v. Given a probability function P, define $P^{\succ}$ so that $P^{\succ}(w) = P(\succ w)$ for every world w. This is the shifted probability function under which the probability of each world has been moved to its successor. Finally, the new probability P_2 is the shifted previous probability P_1 conditional on the new evidence E:

$$P_2(A) = P_1^{\succ}(A/E) = \frac{P_1^{\succ}(A \wedge E)}{P_1^{\succ}(E)}, \text{ provided } P_1^{\succ}(E) > 0.$$

Call this amended form of conditioning *shifted conditioning*. Instead of first shifting and then conditioning on E, we could also first condition on $\succ E$ and then shift, as follows:

$$P_2(A) = P_1(\succ A / \succ E).$$

The result is the same.

I do not assume that the reader has a clear pre-theoretic grip on the concept of a "next world", and thereby on the shifting operation $\succ$. Rather, I assume that we are interested in the dynamics of belief across a certain type of trajectory, and that these trajectories can be modeled as discrete sequences of worlds, with evidence arriving at various precise points. Any such model determines a successor relation that can be plugged into the amended form of conditioning. It is mathematically routine to relax the modeling assumptions so as to allow for continuous trajectories with a continuous stream of evidence. However, the added mathematical complexity would only obscure

5 For a textbook presentation, see [LaValle 2006: part III]; see also [Boutilier 1998] for an application of the same ideas to the framework of [Alchourrón et al. 1985].

the issues I want to discuss – arguably without even making the model more realistic, since actual belief update is plausibly discrete. In practice, when we consider particular agents in particular scenarios, it is typically easy to construct a discrete model of the relevant update process. We can even choose the "next world" to be a full day in the future, as long as the agent doesn't receive any relevant evidence in between.

Shifted conditioning, as presented above (and in [Schwarz 2012]), does not work if worlds can have multiple successors.[6] Consider the story of Fred. Here the tails worlds lie on a branching trajectory that continues with one branch to Monday and with another to Tuesday. On the present model, successor worlds always inherit the full probability of their predecessor. So *Tails & Monday* and *Tails & Tuesday* both get probability 1/2, as does *Heads & Monday*. The new probabilities don't add up to 1.

To address this problem, [Meacham 2010] adds a normalisation step to shifted conditioning.[7] On his account, $P_2(A) = P_1^M(A/E)$, where the shifting transformation M is defined by

$$P^M(w) = P(\succ w)\,\frac{P(\Diamond w)}{\sum_{v\in\Diamond w} P(\succ v)}.$$

Recall that $\Diamond w$ is the strongest uncentred proposition entailed by w; so $P(\Diamond w) = \sum_{v\in\Diamond w} P(v)$. If all points in $\Diamond w$ have unique successors, the scaling factor on the right is 1. On the other hand, if $v \in \Diamond w$ has two successors w_1 and w_2, then $P(v)$ is counted only once in the numerator but twice in the denominator, first as $P(\succ w_1)$ and again as $P(\succ w_2)$. The effect is that if some points in an uncentred world have multiple successors, then the shifted probabilities of all points in that world are normalised so that their sum equals the previous probability of the uncentred world.

6 Further complications arise from terminal worlds with no successor, and fusion scenarios in which several worlds have the same successor. Terminal worlds are relatively unproblematic. For technical convenience I here assume that intuitively terminal worlds are modeled as succeeded by arbitrary worlds excluded by the new evidence. Fusion cases are more delicate. The current statement of shifted conditioning allows for such cases, but it will generally be unsatisfiable if the agents at the predecessor worlds have different beliefs. A natural generalisation is to employ a mixture of the predecessor probabilities in place of P_1, as suggested in [Meacham 2010]. On the other hand, this might not make optimal use of the available information: if one predecessor has found out that $A \vee B$ and another that $\neg A$, why not let the successor know that B (at least if the propositions are uncentred)? In this paper, I will set aside the possibility of fusion.

7 What follows is a corrected version of Meacham's "Local Predecessor Conditionalization". ([Meacham 2010] also discusses a "Global" rule that yields the same results as the uncentred conditioning models.) Meacham's own formulation of his rule, in the present notation, goes as follows.

$$P_2(A) = \sum_{w\in A} P_1(\succ w/\succ E)\frac{P_1(\Diamond\succ w/\succ E)}{\sum_{v\in\Diamond w} P_1(\succ v/\succ E)}.$$

By summing over all successors of worlds that have E-worlds as successors, this gives positive probability to worlds that are incompatible with the new evidence. My formulation avoids this problem, and matches Meacham's informal presentation of his rule. Meacham (personal communication) agrees.

Unfortunately, Meacham's rule leads to the same problems with dynamic stability as the uncentred conditioning models. Suppose a certain universe contains three agents on three different planets that might be you. Call the three planets X, Y and Z. The person on planet Y is about to find out that it is Monday. The person on Z will find out that it is Tuesday. The person on X will fission in such a way that one successor will find themselves at Monday and the other at Tuesday. (What's new is that these possibilities are all located in the same uncentred world.) If your initial credence in each of the three locations is 1/3, then Meacham's shifted probability assigns $1/3 \times \frac{1}{4/3} = 1/4$ to each of the four successor locations. The probability of being on planet X thereby increases to 1/2, and it remains there after conditioning either on Monday or on Tuesday. Your credence in coming from planet X goes up no matter what you learn.

I think there is a simpler way to generalise shifted conditioning that avoids this consequence. The problem with the original rule was that every world gives its full probability to all its successors, so that the total probability in the successor generation exceeds 1. The natural fix is to say that the probability of a world with multiple successors must be divided among its successors: you can't bequeath more than you own.

Let's apply this to Fred. Heads worlds have a unique successor, so shifting simply transfers the probability from *Heads & Sunday Heads & Monday*. On the other hand, the probability of Tails worlds is divided between *Tails & Monday* successors and *Tails & Tuesday* successors. If it is divided evenly, the new credence is divided 1/2 – 1/4 – 1/4 between *Heads & Monday*, *Tails & Monday* and *Tails & Tuesday*.

To complete the proposal, we would need to say how, in general, the probability of a world should be divided among its successors.

4 Transition probabilities

[Parfit 1984] has convinced many philosophers that survival comes in degrees. One might similarly argue there are degrees of epistemic successorhood. If w_1 is more of a successor of v than w_2, then arguably more of v's probability should be shifted to w_1. Allowing for unequal inheritance is also crucial in Everettian quantum mechanics, where the redistribution of credence should reflect the quantum mechanical amplitudes of the relevant branches.

To achieve this kind of generality, we need *transition probabilities*, i.e. a function τ that assigns to each world v a probability distribution τ_v over worldmates w of v. The idea is that $\tau_v(w)$ captures the degree to which w is a successor of v. Given a transition function τ, the shifted probability $P^{\succ}$ can be defined as the expectation of the relevant transition probabilities:

$$P^{\succ}(w) = \sum_{v \in \Diamond w} P(v)\pi_v(w).$$

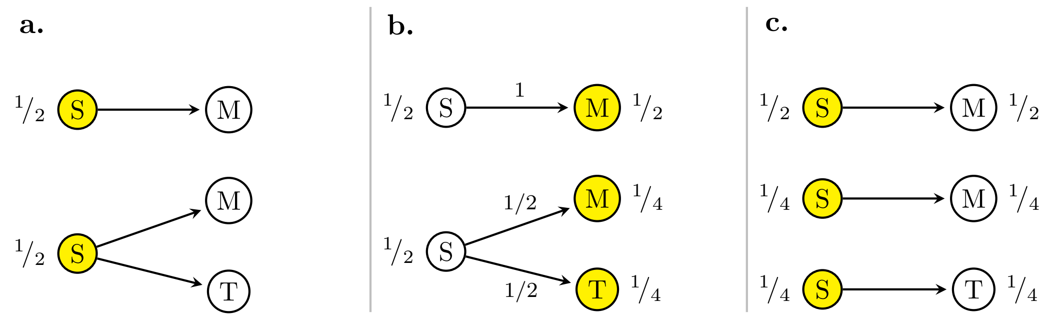

Figure 1: **shifting probabilities in branching worlds.** a. Fred's beliefs are divided between two possibilities: either he is about to be teleported to Monday or he is about to be teleported to both Monday and Tuesday. b. The degrees of belief are shifted to the successor points, weighted by the transition probabilities. c. In an Ockhamist framework, the second possibility – being teleported to both Monday and Tuesday – is treated as two distinct possibilities from the outset.

So to compute the shifted probability of w, you add up the probability of each worldmate v of w, weighted by the degree to which w is a successor of v. As before, the final probability is $P_2(A) = P_1^{\succ}(A/E)$.

Return once again to Fred. Assume the transition probabilities between *Tails & Sunday* worlds and the corresponding Monday and Tuesday worlds are 1/2. To compute the shifted probability of, say, *Tails & Monday*, we add up the old probability of all worlds with links into *Tails & Monday* (i.e. of all *Tails & Sunday* worlds), weighted by the strength of those links (1/2). Since the old probability of *Tails & Sunday* was 1/2, the shifted probability of *Tails & Monday* is 1/4. Figure 1.a–b illustrates the process.

Transition probabilities are "probabilities" because they satisfy the probability axioms. They are not degrees of belief. They are not objective chances. What are they? Well, consider an agent who knows that she is about to undergo fission, with one successor waking up on Monday, the other on Tuesday. Would it be reasonable for the successors, without any relevant evidence, to be certain that they are on Monday, or on Tuesday? I think not. Arguably, the successors should at this point be undecided between Monday and Tuesday. Norms like these are captured by the transition probabilities: the transition probabilities between the relevant Sunday worlds and their successors are about 1/2.

In general, the transition probability between v and w is the fraction of the agent's credence that should move from v to w during shifting. In easy cases, where every world has a unique successor, all transition probabilities are either 0 or 1. (In the next section, we will meet a view on which this covers all cases.) If v has multiple successors, we have to decide how probability should be divided among them when the agent updates their belief. In cases like Fred's, it is plausible that the split should be fair, so that $\tau_v(w) = 1/n$

for all v successors w, where n is the total number of successors. In less symmetrical cases, if one of the successors is a skeptical scenario, or if different successors come with different degrees of survival, the distribution should arguably be uneven, privileging non-skeptical scenarios and higher degrees of survival.

There may not always be a single right way to distribute credence among successor worlds. Often a whole range of transitions may be permissible, corresponding to a range of transition functions. In this respect, transition probabilities are similar to ultimate priors. In fact, I think it is plausible that the range of acceptable transition probabilities between a world and its successors always coincides with the range of rational prior probabilities conditional on the successors. But I have not officially built that into the model.

Very specific transition probabilities are required by the Everett interpretation of quantum mechanics, to make sense of the way physicists derive empirical predictions from hypotheses in quantum mechanics. Roughly speaking, this requires that conditional on a certain hypothesis about the wavefunction, one should expect to witness outcomes in proportion to the squared modulus of the corresponding branch amplitudes. For example, if hypothesis H_1 says that some outcome has low branch weight, while H_2 says that it has high weight, then observation of the outcome is taken to support H_2 over H_1. In the present framework, the Everettian requirement can be expressed as follows. If an agent assigns positive credence to a centred world v that entails a particular hypothesis H about branch weights, then shifting should divide the probability of v among its successors in accordance with the branch weights postulated by H. More succinctly, $\tau_v(w)$ should be the squared modulus of the amplitude of the w branch diverging from v. This ensures that from an epistemic perspective, branch weights behave just like objective probabilities.

To illustrate, suppose your credence is evenly divided between an Everettian hypothesis H_1 on which outcome O has branch weight 0.2 and a hypothesis H_2 on which the weight is 0.6. If these were statements about objective probability, we could use the usual method of plugging the Principal Principle into Bayes's Theorem to show that observation of O should raise your credence in H_2 to 0.75. The same happens if you follow shifted conditioning with the transition probabilities matching the branch weights: after shifting, $H_1\&O$ then has probability $0.5 \cdot 0.2 = 0.1$, $H_2\&O$ has $0.5 \cdot 0.6 = 0.3$, and the rest goes to $\neg O$ possibilities. Conditioning on O therefore raises the probability of H_2 to 0.75.

The result of shifting can be understood as a hypothetical stage in the update at which the relevant experiment is over but the agent has not yet looked at the outcome. Such intermediate stages are well-known in the philosophical literature on the Everett interpretation: Vaidman [1998], Tappenden [2010] and others have suggested that while the Everett interpretation leaves no room for genuine uncertainty about outcomes before a branching event, there can be uncertainty after the branching and before the observation.

At this point, they argue, the agent's credence about outcomes, conditional on some hypothesis about the wavefunction, should match the corresponding branch weights. In the present framework, we don't have to assume that these intermediate stages really exist, and we do not need to appeal to dubious counterfactuals about what *would* have been the case if there *were* such a stage.

The remaining challenge for Everettians is to explain why $\tau_v(w)$ should match the weight of the w branch diverging from v. Can this be taken as a primitive norm of rationality? Or can it be derived from pragmatic considerations, in the tradition of Deutsch [1999] and Wallace [2012]?[8] I am skeptical about either proposal. Fortunately, we don't need to wait for this issue to be settled. The basic form of the update process remains the same no matter how the transition probabilities are filled in and how they can be justified.

5 Ockhamism

Suppose for a moment that Fred knows he is about to be teleported to both Monday and Tuesday. Would it nevertheless make sense for him to wonder where he will wake up? I have effectively assumed that it would not. Uncertainty requires multiple possibilities. For Fred to be uncertain about where he is going to wake up, his doxastically possible worlds would have to divide into worlds where he wakes up only on Monday and worlds where he wakes up only on Tuesday. But then Fred would have misunderstood his situation. Perhaps he thinks he has an immaterial soul that will determinately travel to either Monday or Tuesday. Fred's actual situation is not one in which he wakes up only on Monday, nor is it one where he wakes up only on Tuesday. If Fred is aware of the relevant facts, he therefore cannot be uncertain about which of these possibilities is actual.

Some philosophers disagree and claim that Fred could meaningfully wonder whether he will awaken on Monday or on Tuesday. [Ninan 2009] supports this by an appeal to

8 [Greaves 2007b] pursues this second strategy, defending a general model for belief update across fission. The model is compatible with shifted conditioning, but its centred propositions are restricted to propositions concerning the agent's present branch in a universe. As a consequence, post-branching probabilities are defined over possibilities (new branches) that were not even in the agent's doxastic space before the branching. Greaves assumes that agents nevertheless have a "quasi-credences" defined over their successors' doxastic space. According to her model, this quasi-credence then gets conditioned on the new evidence. Quasi-credence divides into genuine credence for non-branching scenarios and a "caring measure" for branching scenarios. Following Deutsch and Wallace, Greaves argues that conditional on a branching scenario, an agent's caring measure for future branches should match the quantum weights of these branches. In the present model, the requirement that the new credence is the old quasi-credence conditioned on the new evidence translates into the requirement that the transition probabilities between an Everett world and its successors match the corresponding caring measure and therefore the branch weights.

imagination: couldn't Fred *imagine* waking up on Monday and not on Tuesday? He surely could. *Waking up on Monday* is an ordinary proposition that is true, for example, at Fred's Monday successor. But what follows from the fact that Fred can imagine this proposition, which he knows is false (since he knows he is on Sunday)? I do not see how the fact about imagination supports the idea that Fred can be uncertain about where he will wake up. The question is not whether Fred can distinguish two *future* possibilities, waking up on Monday and waking up on Tuesday, but whether he can distinguish two *present* possibilities – whether he lacks information. If Fred were omniscient, would he know which of the supposedly two possibilities is actual? I don't think so.

A different argument in support of pre-fission uncertainty starts with a certain metaphysics of personal identity. According to [Lewis 1976], a situation like Fred's really involves two persons, one of whom wakes up on Monday and the other on Tuesday. Call these two persons Fred_M and Fred_T. Before the fission, Fred_M and Fred_T are co-located: they occupy the exact same place at the same time. But then we can distinguish two Sunday possibilities: *being* $\textit{Fred}_M$ and *being* $\textit{Fred}_T$. The first is true for Fred_M, the second for Fred_T. Each of these possibilities has a unique, non-branching successor. The co-located Freds on Sunday can be uncertain about where they will be tomorrow by not knowing which Fred they are today.

This line of thought has recently been explored in the context of Everettian quantum mechanics (see e.g. [Saunders 1998], [Saunders and Wallace 2008], [Lewis 2007b], [Tappenden 2008]). The discussion is obscured not only by the metaphysics of personal identity, but also by the unfortunate choice of English sentences, rather than propositions, as the bearers of probability. A central topic in the debate is therefore the interpretation of sentences like 'I am going to be on Monday', when uttered by Fred on Sunday: is the sentence true under these conditions? If there are two Freds on Sunday, are there also two utterances? Who is the referent of 'I'? From the present perspective, it does not matter how we answer these questions. What matters is whether Fred's credence is divided between possibilities with only a Monday successor and other possibilities with only a Tuesday successor. The semantics of English is beside the point.[9]

9 One distracting factor when looking at sentences comes from linguistic indeterminacy or ignorance. Perhaps the semantics of English does not settle how to evaluate sentences about one's future in a case of fission. Even if it does, Fred may not be fully aware of the relevant rules. In either case, Fred might display a kind of uncertainty towards the sentence 'I am going to be on Monday' even if he is not at all uncertain about the relevant non-linguistic facts.

Another distraction arises from the fact that the evaluation of statements about the first-person future depends on the metaphysics of personal identity. Suppose (following [Parfit 1984]) we decide that persons cannot survive episodes of fission, so Fred will wake up neither on Monday nor on Tuesday. Then 'I am going to be on Monday' is plausibly false when uttered by Fred on Sunday. However, if we separate questions of dynamic rationality from issues of personal identity, it does not follow that Fred can't be uncertain about a relevant fact. On the view that persons can't survive fission, Fred (qua person) has the property of *not existing tomorrow*; the relevant branching trajectory

A more interesting argument in support of pre-fission uncertainty is implicit in the decision-theoretic program of Deutsch and Wallace (see [Wallace 2012]). Deutsch and Wallace argue that before a branching event, agents in Everett worlds ought to act as if they distribute their credence over future trajectories in accordance with the Everettian branch weights. If this is correct, the decision-theoretic role of rational degree of belief is realised by a probability function that distinguishes between the branching futures even before the fission.

There is an old position in tense logic according to which statements about the future in a world with branching time can only be evaluated relative to a particular branch. [Prior 1967] called this view *Ockhamism*. Since every branch determines a unique future, sentences like 'there will be a sea battle' have a determinate truth-value at every evaluation point, even if there is a sea battle only on some branch of the future. Similarly, one could say that Fred's utterance of 'I will be on Monday' must be evaluated relative to a maximal linear subset of Fred's trajectory. If Fred undergoes fission, the sentence would be true relative to one branch and false relative to another. Returning to matters of belief, let's redefine *Ockhamism* as the view that every maximally specific possibility in an agent's belief space has a determinate, linear future. Distinguishing Fred_M and Fred_T as alternative Sunday possibilities may achieve this in the scenario of Fred, but Ockhamism does not require the controversial metaphysics of [Lewis 1976]. Whatever we say about personal identity, we can ask whether the possibilities in Fred's belief space should be modeled by "disambiguating" branching structures or not.[10]

Imagine an omniscient agent whose credence goes to a single world w with multiple successors. In an Ockhamist model, w is represented as several possibilities, one for each branch. Starting from a non-Ockhamist model, Ockhamist possibilities could be identified with ordered pairs of ordinary possibilities and a branch. In the Ockhamist model, the omniscient agent is only "weakly omniscient" in the sense that her credence is divided between possibilities that differ at most with respect to the selected branch.

Formally, it is easy to translate back and forth between Ockhamist models a non-Ockhamist branching models. Every Ockhamist world is guaranteed to have at most one successor. Thus Ockhamism allows us to stick with the original form of shifted conditioning from section 3. The later revisions to account for cases of fission are redundant; all transition probabilities are either 0 or 1.

Let's model the story of Fred in an Ockhamist framework, returning to the original

is therefore "temporally incoherent" in the sense it cannot be instantiated by a persistent object, for such an object would have to exist tomorrow (at two different places) while today have the property of not existing tomorrow.

10 [Saunders 2010] and [Wilson 2012] argue for a generalisation of Lewis's metaphysics of persons on which branching events in Everett worlds are generally *divergence* events wherein previously indistinguishable branches become distinguishable. This view goes naturally with an Ockhamist epistemology, but again it is not required by Ockhamism.

case where he does not know what will happen. We now start with *three* possibilities on Sunday: a heads possibility leading to Monday, a tails possibility leading to Monday, and a tails possibility leading to Tuesday. How is Fred's Sunday credence divided between these alternatives? Since the coin is fair, he should give equal credence to heads and tails. Within the tails possibilities, he should presumably give equal credence to the possibility leading to Monday and the possibility leading to Tuesday. Applying shifted conditioning then yields the same result as before (see figure 1.c).

In general, where we previously saw a single possibility with several futures, we now see several possibilities with unique futures. These possibilities are at present indistinguishable by the agent, so the question arises how rational credence should be divided among them. This is what was previously captured by the transition probabilities. Any constraint on transition probabilities translates directly into a constraint on the division of credence between pre-fission alternatives. The result of shifted conditioning is always the same whether we use an Ockhamist model or a non-Ockhamist model with the corresponding choice of transition probabilities.

The upshot is that little hangs on the question of pre-fission uncertainty. In my view, Ockhamist models distort the doxastic situation of agents in expectation of fission by postulating uncertainty where there is nothing to be uncertain about. But it is reassuring that in the framework of shifted conditioning, this somewhat esoteric question makes practically no difference. There is, however, a technical advantage to Ockhamism, since it retains the equation $P^{\succ}(A) = P(\succ A)$. I will exploit this in the following section to streamline some arguments in support of shifted conditioning.

6 Diachronic rationality

Shifted conditioning combines two operations on an agent's degrees of belief. The first is an update step that accounts for the expected change in the agent's location; the second is standard conditioning on the agent's total new evidence, including evidence about matters of self-location.[11] As I argued in [Schwarz 2012], this account is supported by the

11 In a number of publications (including [Bradley 2011b], [Bradley 2011a], [Bradley 2012]), Darren Bradley has defended a superficially similar account on which an update consists of a step he calls "mutation" and standard conditioning. However, Bradley offers no precise or general statement of mutation. The basic idea is that mutation turns yesterday's belief that it is Sunday into today's belief that it was Sunday yesterday. (According to Bradley, there is an important sense in which the new belief is "the same" as the old one.) But this is the right mutation only if there is reason to believe that exactly one day has passed: in the case of Sleeping Beauty, Bradley assumes that Beauty's 50% credence in *Tails & Sunday* mutates into a 25% credence in *Tails & Monday* and a 25% credence in *Tails & Tuesday*, but he does not explain why (nor does he explain which of these new beliefs is the same as the Sunday belief). Another difference between Bradley's account and mine concerns "observation selection effects". In the Everett example, an observation selection effect would occur if the post-fission method of observation were more likely to yield Monday than Tuesday observations,

very same considerations that are traditionally taken to support standard conditioning when uncentred propositions are ignored. In the present section, I want to further illustrate this point by looking at some consideration not discussed in [Schwarz 2012]. To simplify the arguments, I will initially assume an Ockhamist framework, so that $P_1^{\succ}(A/E) = P_1(\succ A/\succ E)$.

As a warm-up, let's verify that shifted conditioning satisfies the stability condition from section 2.

> *Dynamic stability:* if a proposition A is certain not to change its truth-value, and an agent knows in advance that her new evidence will be one of $E_1, \ldots, E_n$, then rationality should not require her probability in A to increase no matter which of $E_1, \ldots, E_n$ she learns.

If A is certain not to change its truth-value, then $P_1(A \leftrightarrow \succ A) = 1$. By shifted conditioning, the new credence in A after learning E_i is $P_1^{\succ}(A/E_i) = P_1(\succ A/\succ E_i) = P_1(A/\succ E_i)$. Since the agent knows in advance that she will learn one of the mutually exclusive propositions $E_1, \ldots, E_n$, $P_1(A) = \sum_i P_1(\succ E_i)P_1(A/\succ E_i)$. It follows that $P_1(A/\succ E_i)$ cannot be greater than $P_1(A)$ for all E_i.

In section 2, I mentioned that uncentred conditioning models are vulnerable to Dutch books, thereby violating the following condition.

> *Dynamic coherence:* rationality should not demand an agent to update her beliefs in such a way that she can incur a sure loss if she bets in accordance with her earlier and later beliefs.

In this context, a bet may be understood as deal that pays some amount $\$X$ if a certain proposition A is true and otherwise costs some amount $\$Y$, with no further financial consequences. An agent *bets in accordance with her beliefs* if she accepts any bet with positive expected payoff. In effect, this means that we consider agents whose are only interested their net monetary profit. The idea behind dynamic coherence is that rationality should not require such an agent to accept bets which, by their own lights, amount to a sure loss.

In [Schwarz 2012] I showed that *every* systematic alternative to shifted conditioning violates this condition. That shifted conditioning itself satisfies it can be proved as follows, adapting an argument from [Skyrms 1987].

Let us model a diachronic Dutch book as a pair consisting of a (finite) number of earlier bets B_1 together with a mapping B_2 from the members of some evidence partition $E_1, \ldots, E_n$ to (finite) sets of later bets, so that $B_2(E_i)$ is offered if E_i the new evidence,

even if the branch weights are equal. I don't think anyone believes in such an effect, so I don't agree with Bradley that disagreement over observation selection effects is relevant to the debate surrounding such cases.

and where accepting B_1 together with $B_2(E_i)$ amounts to a net loss, for every E_i. Now consider one of the later bets, to be placed if the new evidence is E_i. Let's say the bet pays $\$X$ in case of A, otherwise $\$Y$. A rational agent who only cares about her net profit should accept this bet iff it has positive expected payoff according to her beliefs after learning E_i, i.e. iff $P_2(A)\$X + P_2(\neg A)\$Y \geq \$0$, where P_2 is P_1 updated on the information E_i. By shifted conditioning, $P_2(A) = P_1^{\succ}(A/E_i) = P_1(\succ A/\succ E_i)$. So the later bet has positive expected payoff iff $P_1(\succ A/\succ E_i)\$X + P_1(\succ\neg A/\succ E_i)\$Y \geq \$0$, i.e. iff a bet conditional on $\succ E_i$ that pays $\$X$ in case of $\succ A$ and $\$Y$ in case of $\succ\neg A$ has positive expected payoff at the earlier time. Since $\succ A$ and $\succ E_i$ are true at the earlier time iff A and E_i are true at the later time, the payoff is guaranteed to be the same for the original later bet and the converted earlier bet. Substituting each of the bets in $B_2(E_i)$ by a corresponding earlier bet, and combining these bets with the bets in B_1 therefore yields a synchronic Dutch book against the agent at the earlier time. But [Kemeny 1955] proved that if an agent's probabilities respect the probability calculus, then they are immune to (finite) synchronic Dutch books. It follows that an agent who obeys the probability calculus and updates by shifted conditioning is also immune to diachronic Dutch books.

Dissatisfaction with the allegedly pragmatic nature of Dutch Book arguments has recently led some epistemologists to turn to considerations of *accuracy*, representing the distance between a belief function and the actual facts. For concreteness, let's define the *inaccuracy* $I(P, w)$ of a belief function P at a world w by the Brier score $\sum_A |P(A) - w(A)|^2$, where A ranges over all propositions and $w(A)$ is the truth-value of A at w. (For a defense of this measure, see e.g. [Leitgeb and Pettigrew 2010a].) A plausible constraint on rational belief update is that if E is the new evidence, then the new belief function should have minimal expected inaccuracy by the light of the previous belief function among all functions P with $P(E) = 1$ (see [Leitgeb and Pettigrew 2010b]). However, if propositions can change their truth-value, then what should be considered is not the expected *present* inaccuracy of the candidate future function P, but its expected *future* inaccuracy. More precisely, we should weight the inaccuracy of P at w not by the probability that w is the present point, but by the probability that w is the next point, the point where E will be learnt. So the expected future inaccuracy of P by the lights of P_1 is not $\sum_w P_1(w)I(P, w)$, but $\sum_w P_1(\succ w)I(P, w) = \sum_w P_1(\succ w) \sum_A |P(A) - w(A)|^2$.

> *Accuracy Conduciveness:* If E is the evidence an agent receives at t_2, then her new belief function P_2 should have minimal expected future inaccuracy by the lights of the previous belief function P_1 among functions assigning 1 to E.

Adapting an argument in [Leitgeb and Pettigrew 2010b], it is easy to show that P_2 has minimal expected future inaccuracy by the light of P_1 iff P_2 results from P_1 by shifted conditioning. Let P be any function with $P(E) = 1$. The expected future inaccuracy of P is

$$\sum_{w \in E} P_1(\succ w) \sum_A |P(A) - w(A)|^2 = \sum_A \sum_{w \in E} P_1(\succ w)|P(A) - w(A)|^2.$$

But for every proposition A,

$$\begin{aligned} \frac{d}{dx} \sum_{w \in E} P_1(\succ w)|x - w(A)|^2 &= 2(\sum_{w \in E} P_1(\succ w)x - P_1(\succ w)w(A)) \\ &= 2(P_1(\succ E)x - P_1(\succ (A \wedge E))). \end{aligned}$$

This is zero iff $P_1(\succ E)x = P_1(\succ (A \wedge E))$, i.e. iff $x = P_1(\succ (A \wedge E))/P_1(\succ E) = P_1(\succ A/\succ E)$. So the function P with minimal expected future inaccuracy assigns to any proposition A the value $P_1(\succ A/\succ E) = P_1^{\succ}(A/E)$.

The assumption of Ockhamism allowed us to ignore various subtleties that arise in cases of fission. For example, how shall we define the expected future inaccuracy of a belief function P by the light of P_1 in a non-Ockhamist model when the agent knows that her present point has two successors? P might then be more inaccurate at one than at the other. To get the desired result, we should average the two degrees of inaccuracy, weighted by the corresponding transition probabilities.

More interesting complications arise for Dutch Books. We now have to decide whether we consider the net profit between the pre-fission point and all its successors, or the net profit for each pair of a pre-fission point and one successor. (Under Ockhamism, these coincide.) I think the pairwise consideration is more natural, in which case the Dutch book arguments go through as before. But what if we look at the result for the whole "tree" including all successors? You may have noticed that the Dutch Book against uncentred conditioning models described in section 2 does not work from this perspective: the tree does not incur a sure loss. In fact, you might argue that trees can be Dutch booked if their branches obey shifted conditioning. Consider Fred. Initially, he should accept a bet that pays \$10 in case of tails and costs \$9 on heads. Updating by shifted conditioning, his successors should accept a bet that pays \$8 in case of heads and costs \$7 on tails. If the outcome is tails, the second bet gets offered twice, and the net payoff is \$-4. If the outcome is heads, the bet is offered only once and the net payoff is \$-1.[12]

The first thing to note about this "Dutch book" is that it works just as well in a framework where all propositions are uncentred and the agent updates by standard conditioning. So if it shows that shifted conditioning is not the right update rule in the centred worlds framework, it also reveals that standard conditioning is not the right rule in the uncentred worlds framework. However, I don't think it reveals either of these things.

12 This mirrors the Dutch Book argument against Lewisian halfing in [Hitchcock 2004].

Consider the following analogy. Three agents A, B and C are offered bets on the outcome of a fair coin toss. The bet offered to A pays \$10 on tails and costs \$9 on heads. B and C are offered a bet that pays \$8 on heads and costs \$7 on tails. In addition, some or all of the players are assigned to a "group": on heads, the group consists of A together with one of B and C, chosen at random. On tails, it consists of A, B and C. If each player only cares about their own payoff, they should accept the bets; the group then incurs a sure loss of either \$1 or \$4. Interestingly, the players should even accept their bets if they only care about the net payoff for the group. The decision problem for B and C can then be represented by the following matrix, ignoring states where they are not chosen for the group, in which case it is irrelevant what they do.

	Chosen & other accepts	Chosen & other doesn't accept
Accept	Heads: \$8, Tails: \$-14 (\$-3)	Heads: \$8, Tails: \$-7 (\$0.5)
Don't Accept	Heads: \$0, Tails: \$-7 (\$-3.5)	Heads: \$0, Tails: \$0 (\$0)

The outcomes in the cells are fair lotteries, whose expected payoff is registered in parentheses. Accepting the offer is the dominant option, although each player would prefer if both rejected the offer. The situation is a Prisoner Dilemma.

The general lesson is that if a group of agents does not have a fixed number of members, then the agents should sometimes accept bets that amount to a sure loss for the group, even if all they care about is the net outcome for the group.[13]

7 Consequences and conclusions

Cases of epistemic fission illustrate that even in the absence agent of essentially self-locating uncertainty, degrees of belief in uncentred propositions should not evolve by conditioning on uncentred evidence. I have presented an alternative update rule that allows conditioning on arbitrary centred or uncentred propositions, by adding a "shifting" step that takes into account the agent's change of location. The rule not only gives intuitively sensible results in cases of fission, it is also supported by the same sorts of arguments that support standard conditioning in models with only uncentred propositions. On its own, my proposal does not solve the confirmation-theoretic problem for the Everett interpretation, but it might give us a better grip on what is needed. It is not important whether, conditional on a specific fission scenario, agents can be uncertain about the

13 In the case of Fred's successors, the columns in the above decision matrix should be interpreted as representing what the other successor *would* do in case of tails. Since the successors can be reasonably confident that they make the same choice, the Prisoner Dilemma here turns into a Twin Dilemma and thereby a Newcomb problem. This is why the alleged "Dutch book" could be blocked by using Evidential Decision Theory, as pointed out in [Arntzenius 2002] and [Briggs 2010] for the case of Sleeping Beauty.

future. Nor is it important whether there always is (or could be) an intermediate stage between branching and observation. All that matters is that rational prior probabilities conditional on Everett worlds (equivalently, transition probabilities for Everett worlds) match the corresponding branch weights.

Fission scenarios have further consequences that deserve investigation. For example, they show that dynamic coherence and stability are incompatible with the *principle of self-locating indifference* for posterior beliefs, assumed e.g. in [Elga 2000] and [Lewis 2001]: if Fred is initially undecided between two possibilities within the same uncentred world, one of which fissions into a *Monday* possibility and a *Tuesday* possibility, while the other leads only to *Monday*, dynamic stability requires that his new credence is divided 1/4 – 1/4 – 1/2 between these alternatives, although they are located in the same uncentred world. On the other hand, Fred's new evidence is arguably neutral between these three alternatives. His ultimate priors, conditional on his new evidence, might well have satisfied the requirement of indifference. Among other things, this shows that there can be disagreement between perfectly rational agents with the very same priors and the same evidence.

Relatedly, fission cases point at a neglected way in which rational degrees of belief can come apart from known objective chance. Suppose in the original story of Fred, Tuesday (the moon) is a lot further away than Monday, and that the coin that decides whether Fred gets teleported to Tuesday is actually tossed by Fred's successor on Monday. When the Monday successor tosses the coin, he knows that he is on Monday, for the Tuesday successor doesn't get to toss a coin. As we've seen, at this point, his credence should be divided 2/3 – 1/3 between *Heads & Monday* and *Tails & Monday*. So he assigns credence 2/3 to the hypothesis that the fair coin he is about to toss will land heads!

Finally, the story of Fred bears an obvious resemblance to the Sleeping Beauty problem. The 1/2 – 1/4 – 1/4 distribution between *Heads & Monday*, *Tails & Monday* and *Tails & Tuesday*, together with the 2/3 – 1/3 distribution after updating on *Monday*, is known as the "Lewisian halfer" solution, defended in [Lewis 2001]. Does shifted conditioning commit us to Lewisian halfing? It does not – at least not directly. The answer to Sleeping Beauty turns out to depend on several further questions. In this paper, I have set aside the question whether norms like conditioning or shifted conditioning should be read as literal diachronic norms or as a second-order norms linking beliefs about A to beliefs about previous beliefs about A. As shown in [Schwarz 2012], the second-order version of shifted conditioning supports the standard "thirder" solution. The diachronic version does not lead to Lewisian halfing either, but rather to the strange answer that if the coin lands tails, then Beauty should be certain on Monday that it is Monday and on Tuesday that it is Tuesday. Arguably this violates a constraint of the setup: the memory erasure on Monday night does not allow Beauty to have different beliefs on the two awakenings. The answer to Sleeping Beauty thus depends on the question how an agent should update their beliefs under circumstances in which it may not be possible to update them in the optimal way. The model defended here does not address this question. I think it does ultimately support Lewisian halfing, but this requires further argument that will have to wait for another occasion.

References

Carlos E. Alchourrón, Peter Gärdenfors and David Makinson [1985]: "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision". *Journal of Symbolic Logic*, (50): 510–530

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

Craig Boutilier [1998]: "A unified model of qualitative belief change: a dynamical systems perspective". *Artificial Intelligence*, 98: 281–316

Darren Bradley [2007]: "Bayesianism and Self-Locating Beliefs, or Tom Bayes Meets John Perry". PhD Thesis, Stanford University

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

— [2011b]: "Self-location is no problem for conditionalization". *Synthese*, 182: 393–411

— [2012]: "Four Problems About Self-Locating Belief". *The Philosophical Review*, 121: 149–177

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

Roderick Chisholm [1967]: "Identity through Possible Worlds: Some Questions". *Noûs*, 1: 1–8

David Deutsch [1999]: "Quantum Theory of Probability and Decisions". *Proceedings of the Royal Society of London*, A455: 3129–3137

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

Hilary Greaves [2007a]: "On the Everettian epistemic problem". *Studies in History and Philosophy of Modern Physics*, 38: 120–152

— [2007b]: "Probability in the Everett Interpretation". *Philosophy Compass*, 2: 109–128

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

Christopher Hitchcock [2004]: "Beauty and the Bets". *Synthese*, 139: 405–420

John G. Kemeny [1955]: "Fair Bets and Inductive Probabilities". *Journal of Symbolic Logic*, 20: 263–273

Namjoong Kim [2009]: "Sleeping Beauty and Shifted Jeffrey Conditionalization". *Synthese*, 168: 295–312

Steven M. LaValle [2006]: *Planning Algorithms.* Cambridge: Cambridge University Press

Hannes Leitgeb and Richard Pettigrew [2010a]: "An Objective Justification of Bayesianism I: Measuring Inaccuracy". *Philosophy of Science*, 77: 201–235

— [2010b]: "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy". *Philosophy of Science*, 77: 236–272

David Lewis [1976]: "Survival and Identity". In Amelie O. Rorty (Hg.), *The Identities of Persons*, University of California Press, 17–40,

— [1979]: "Attitudes *De Dicto* and *De Se*". *The Philosophical Review*, 88: 513–543. Reprinted in [**?**]

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

Peter Lewis [2007a]: "Quantum Sleeping Beauty". *Analysis*, 67: 59–65

— [2007b]: "Uncertainty and Probability for Branching Selves". *Studies in History and Philosophy of Modern Physics*, 38: 1–14

Christopher Meacham [2008]: "Sleeping Beauty and the Dynamics of De Se Beliefs". *Philosophical Studies*, 138: 245–269

— [2010]: "Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs". In Tamar Szabo Gendler and John Hawthorne (Eds.) *Oxford Studies in Epistemology, Volume 3,* Oxford University Press, 86–125

Dilip Ninan [2009]: "Persistence and the First Person". *The Philosophical Review*, 118: 425–464

Derek Parfit [1984]: *Reasons and Persons.* Oxford: Clarendon Press

John Perry [1979]: "The problem of the essential indexical". *Noûs*, 13: 3–21

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

Arthur N. Prior [1967]: *Past, Present and Future*. Oxford: Oxford University Press

Simon Saunders [1998]: "Time, Quantum Mechanics, and Probability". *Synthese*, 114: 373–404

— [2010]: "Chance in the Everett Interpretation". In S. Saunders, J. Barrett, A. Kent and D. Wallace (Eds.) *Many Worlds? Everett, Quantum Theory, and Reality,* Oxford: Oxford University Press

Simon Saunders and David Wallace [2008]: "Branching and Uncertainty". *British Journal for the Philosophy of Science*, 59: 293–305

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

Brian Skyrms [1987]: "Dynamic coherence and probability kinematics". *Philosophy of Science*, 54(1): 1–20

Robert Stalnaker [2008]: *Our Knowledge of the Internal World*. Oxford: Oxford University Press

Paul Tappenden [2008]: "Saunders and Wallace on Everett and Lewis". *British Journal for the Philosophy of Science*, 59: 307–314

— [2010]: "Evidence and Uncertainty in Everett's Multiverse". *British Journal for the Philosophy of Science*, xxx: xxx

Paul Teller [1973]: "Conditionalization and observation". *Synthese*, 26(2): 218–258

Michael G. Titelbaum [2008]: "The Relevance of Self-Locating Beliefs". *The Philosophical Review*, 117: 555–606

Lev Vaidman [1998]: "On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory". *International Studies in the Philosophy of Science*, 12: 245–266

David Wallace [2012]: *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press

Alastair Wilson [2012]: "Everettian Quantum Mechanics without Branching Time". *Synthese*, 188: 67–84