Belief update across fission

Belief update across fission*

Wolfgang Schwarz
14 March 2014

Abstract. When an agent undergoes fission, how should the beliefs of the fission results relate to the pre-fission beliefs? This question is important for the Everett interpretation of quantum mechanics, but it is of independent philosophical interest. Among other things, fission scenarios demonstrate that "self-locating" information can affect the probability of uncentred propositions even if an agent has no essentially self-locating uncertainty. I present a general update rule for centred beliefs that gives sensible verdicts in cases of fission, without relying on controversial metaphysical or linguistic assumptions. The rule 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 used on the question of how to describe scenarios like this. Can people survive teleportation? Which of the persons awakening on the two moons, if any, is identical to the person going to sleep on Sunday? 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 relate to Fred's beliefs before he went to sleep on Sunday?

The two questions 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 so that two new persons come into existence on Monday and Tuesday, it still makes sense to ask how the beliefs of these persons should relate to the beliefs of Fred. Imagine you are designing intelligent 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 case 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 while he is asleep: if it lands heads, the signal to Tuesday will be cut so that he only gets teleported to Monday. In fact, the

* This paper once fissioned from [Schwarz 2012], which presents the same update rule, but (in the published version) sets aside cases of fission. Thanks to Christopher Meacham, Paolo Santorio and two anonymous referees for helpful comments and discussion.

coin lands tails and the signal isn't cut, but Fred doesn't know this. Now how confident should his successors be that they are on Monday? What should they believe about the outcome of the coin toss? What should Fred's Monday successor believe once he learns that he is on Monday?

Why care about this far-fetched situation? There are several reasons. I will argue that fission cases illustrate an important fact about the relevance of "self-locating" evidence and thereby cast doubt on a popular approach to the dynamics of rational belief. They also shed new light on the connection between objective chance and rational credence, and on the possibility of rational disagreement among agents with the same evidence and priors. Fred's predicament also bares an obvious resemblance to the Sleeping Beauty problem ([Elga 2000], [Lewis 2001]), which has caused some concern among philosophers who want to give different answers to the two problems (see e.g. [Lewis 2007a]). The model I will present can alleviate these worries.

Finally, Fred's situation 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 all possible outcomes determinately occur, although on different "branches" of the universe. Thus if a particle is in a superposition of two states M and T and you measure the relevant state, one of your successors will find the detector indicating M, the other T. 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 successor on Monday 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 useful heuristic, 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 relevant 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*. So a centred proposition is a set of centred worlds. Intuitively, each 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: if an uncentred proposition is true (or false) at a centred world w, then it is true (false) at all worldmates of w. A maximally specific uncentred proposition is an *uncentred (possible) world*.[1]

There are other ways of modelling 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 1981], [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*. For every moment in Napoleon's life, the totality of his properties (at that time) constitute a centred world. So the 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.

The familiar Bayesian rule of *conditioning* can now be interpreted as saying how an agent's degrees of belief should evolve along a trajectory. Let w_1, w_2 be subsequent positions on an agent's trajectory, and suppose that at w_2 the agent undergoes a learning event whose direct impact is to confer certainty on some proposition E. Conditioning specifies that if P_1 is the credence function at w_1 and P_2 the credence function at w_2, then for every proposition A,

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

The status of this rule is controversial. Some doubt that there are diachronic constraints on rational belief at all. On this view, conditioning should arguably be replaced by a synchronic second-order rule to the effect that $P_2(A/P_1(A/E) = x) = x$. Others question whether learning events can be modelled as conferring certainty on some proposition E

1 There are different ways of rendering the definition of worldmates more precise. For example, we might say that w is a worldmate of w' iff w entails that w' is instantiated by some individual at some time.

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.

and instead opt for what Jeffrey [1992] calls 'probability kinematics'. The issues I want to focus on are independent of these points; the model I will defend has straightforward synchronic and Jeffrey-style counterparts (see [Schwarz 2012] for details).

So assume that conditioning is indeed the right way for an ideally rational agent to change her mind if all relevant propositions are uncentred. Unfortunately, this can no longer be maintained if we allow for centred 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 fact that agents change their position in logical space.

A natural 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], [Briggs 2010][2]). The new update process might then be described as follows. Let P_1^* be P_1 restricted to uncentred propositions, so that P_1^* represents the agent's uncentred belief state at w_1. At the new point w_2, this function gets conditioned on the uncentred information acquired at w_2. So let '$\Diamond A$' denote the strongest uncentred proposition entailed by a proposition A.[3] Let P_2^* be P_1^* conditioned on $\Diamond E$. Note that each uncentred world to which P_2^* assigns positive probability contains at least one point at which E is true. Suppose for simplicity that all these worlds contain exactly *one* point where E is true; 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 at which E is true. 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 settles the new centred probabilities.[4]

2 The proposals in these papers differ in presentation, and not all of them are fully equivalent to the account I am about to describe. Most importantly, Titelbaum's model also allows conditioning on centred propositions as long as these are certain not to change their truth-value. Given an Ockhamist analysis of fission (sec. 5 below), this may block the problematic consequences in the Everett scenario discussed below; see [Titelbaum 2013], chs. 8 and 11.

3 By definition of the worldmate relation (p. 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.

4 If E is not sufficient for self-location, this simple recipe isn't applicable. Some authors here invoke a principle of self-locating indifference according to which one should give equal credence to each E point within the same uncentred world. One option is then to divide the P_2^*-probability of each uncentred world among its E centres; another is to assign the whole P_2^*-probability of each uncentred world to all its E centres and then renormalize the probability distribution. In the Sleeping Beauty

I will call models of this type *uncentred conditioning models*. Observe 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 ignored: 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 two states M and T. The Everett worlds to which you assign positive probability all contain a branch on which the measuring device says 'M' and one on which it says 'T'. Among non-Everett worlds, your credence is divided between worlds where the outcome is 'M' only and worlds where it is 'T' only. Suppose you now observe 'M'. Following the uncentred conditioning models, we first condition P_1^* on $\Diamond$'M'. Since the live Everett possibilities all contain a point where 'M' is true, this rules out all and only the non-Everett worlds in which the outcome is 'T'. As a result, $P_2^*(\textit{Everett}) > 1/2$. Assuming your evidence is sufficient for self-location, $P_2(\textit{Everett}) = P_2^*(\textit{Everett}) > 1/2$. More specifically, if you previously assigned equal credence to non-Everett 'M' worlds and non-Everett 'T' worlds, then $P_2(\textit{Everett}) = 2/3$. Exactly the same update would have occurred if you had observed the outcome 'T'. By repeatedly carrying out measurements, you become more and more confident in the Everett hypothesis, no matter what outcomes you observe. That does not seem rational.[5]

The intuition that something is epistemically amiss here can be supported by Dutch book arguments and considerations of expected accuracy. If you update your beliefs in line with the uncentred conditioning models, 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 costs \$4 if it is false, irrespective of what you observe. You are guaranteed to lose \$1.[6] Similarly, we will see in section 6 that your new credence function has lower expected accuracy according to your

problem, the former leads to "halfing", the second to "thirding"; see e.g. [Halpern 2006], [Briggs 2010].

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

6 [Briggs 2010] presents a purported proof that (a version of) the uncentred conditioning rule 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, 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] (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 $\Diamond$'M' or $\Diamond$'T', 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 $\Diamond$'M' and $\Diamond$'T', unlike 'M' and 'T', are not mutually exclusive: they are both true at Everett worlds.

previous beliefs than a function which assigns probability 1/2 to the Everett hypothesis.

The source of these problems is the assumption that if an agent's evidence is sufficient for self-location, then her degrees of belief in uncentred propositions should evolve by standard conditioning on the uncentred evidence. At first glance, this looks plausible. If your evidence is sufficient for self-location, your only uncertainty concerns which uncentred world you inhabit; to determine the new uncentred probabilities, it should then be enough to consider what your new evidence has to say on this matter. But not so. Even if the evidence is sufficient for self-location, its self-locating aspect can be relevant to uncentred propositions. When you observe the measurement outcome 'M', what you learn is not just a fact about the universe as a whole. You also learn that you are presently looking at an 'M' outcome. Conditioning on this information would exclude 'T' possibilities in Everett worlds just as much as 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 'M' is true, which rules out none of the Everett worlds.

Once we've seen the problem, it is clear that it 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 template. The prior credence P_1 is divided between three uncentred propositions X, Y and Z. The agent knows that they 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 the uncentred beliefs evolve by conditioning on uncentred evidence, then the credence in X will increase no matter whether the agent learns M or T.

Uncentred conditioning models don't work. We need an update rule that lets the agent condition on *all* her evidence, including centred evidence, while also taking into account that centred propositions can change their truth-value.

3 Shifted conditioning

We are looking for a rule that determines an agent's beliefs at w_2 based on her previous beliefs at w_1 together with the new evidence E. We can assume that w_2 is an immediate successor of w_1: if there are points in between w_1 and w_2, the agent's credence at w_2 should be sensitive to information received at these points; such information might already be false by the time of w_2, so we can't simply fold any intermediate information into the information E received at w_2.

Now recall that w_1 and w_2 are maximally specific possibilities. A centred world contains not only 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 does at that time, but also what he (and everybody else) does at every other

time. Among other things, w 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. Crucially, 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 an 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 midnight and it being 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 engineering and computer science. In philosophy, it has only recently been rediscovered by Christopher Meacham [2010] and me [2012].[7]

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:

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

7 See e.g. [LaValle 2006: part III] for a textbook presentation in computer science; see also [Boutilier 1998] for an application of the same ideas to the framework of [Alchourrón et al. 1985]. [Kim 2009], [Schulz 2010] and [Bradley 2011] also offer accounts on which yesterday's belief that it is Sunday should turn into today's belief that it is Monday, at least if the agent knows that exactly one night has passed. Unfortunately, these accounts do not give satisfactory answers if the knowledge condition isn't met.

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. (That's because on the assumption of unique successors, $\succ A \,\&\, \succ E$ is equivalent to $\succ(A \,\&\, E)$; hence $P_1(\succ A \,\&\, \succ E) = P_1(\succ(A \,\&\, E)) = P_1^{\succ}(A \,\&\, E)$ and $P_1(\succ E) = P_1^{\succ}(E)$).

I do not presuppose 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 modelled 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 modelling assumptions so as to allow for continuous trajectories with a continuous stream of evidence. However, the added mathematical complexity would only obscure the issues I want to discuss, without making the model more realistic, since actual belief update is plausibly discrete. In practice, when we consider particular agents in particular scenarios, it is usually easy to construct a discrete model of the relevant update process. Indeed, we may choose the "next world" to lie quite a bit in the future, as long as the agent doesn't receive any relevant evidence in between that isn't reflected in the later evidence.

Shifted conditioning, as presented above (and in [Schwarz 2012]), does not work if worlds can have multiple successors.[8] Consider the case of Fred. Here worlds where the coin lands tails 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.

8 Other obvious corner cases are 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 assume that intuitively terminal worlds are modelled as succeeded by arbitrary worlds excluded by the new evidence. (So if an agent is undecided between an intuitively terminal world and a non-terminal world, then her new credence is concentrated on the successor(s) of the non-terminal world because the alternatives are incompatible with 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 set aside the possibility of fusion.

To fix this, [Meacham 2010] adds a normalisation step to shifted conditioning.[9] 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. Moreover, $P(\diamond w) = \sum_{v\in\diamond w} P(v)$. So 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 normalized so that their sum equals the previous probability of the uncentred world.

Unfortunately, Meacham's rule has the same problematic consequences as uncentred conditioning models, albeit only in more far-fetched situations. Suppose a certain universe contains three agents that might be you, on three different planets. 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 $1/3 \cdot \frac{1}{4/3} = 1/4$ 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 being on planet X goes up no matter what you learn.

I think there is a simpler way to generalize 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. In Fred's doxastic space, worlds where the coin lands heads have a unique successor, so shifting simply transfers the probability from *Heads & Sunday*

9 What follows is a corrected version of Meacham's "Local Predecessor Conditionalization". [Meacham 2010] also discusses a "Global" rule which yields the same implausible results in fission cases as the uncentred conditioning models. Meacham's own formulation of his "Local" rule, adapted to 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. The corrected formulation avoids this problem, and matches Meacham's informal presentation of his rule. Meacham (personal communication) agrees.

to *Heads & Monday*. The probability of tails worlds, on the other hand, is divided between *Tails & Monday* successors and *Tails & Tuesday* successors. If it is divided evenly, the new credence is split 1/2 – 1/4 – 1/4 between *Heads & Monday*, *Tails & Monday* and *Tails & Tuesday*. (If Fred later finds out that he's on Monday, the new credence in *Heads* increases to 2/3.)

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

4 Transition probabilities

We might stipulate that whenever a world has more than one successor, shifting should evenly divide its probability among the successors. But there are reasons to strive for greater generality. For example, [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 (whatever that means), 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, and where the number of successors is arguably not even defined (see [Greaves 2004], sec. 5.3).

To achieve this kind of generality, we need *transition probabilities*, i.e. a family $\{\tau_v\}$ of probability measures, one for each world v, defined over worldmates of v. The idea is that $\tau_v(w)$ captures the fraction of v's probability that should go to w. The shifted probability $P^{\succ}$ can then be defined as the expectation of the relevant transition probabilities:

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

In words: 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 determine 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,

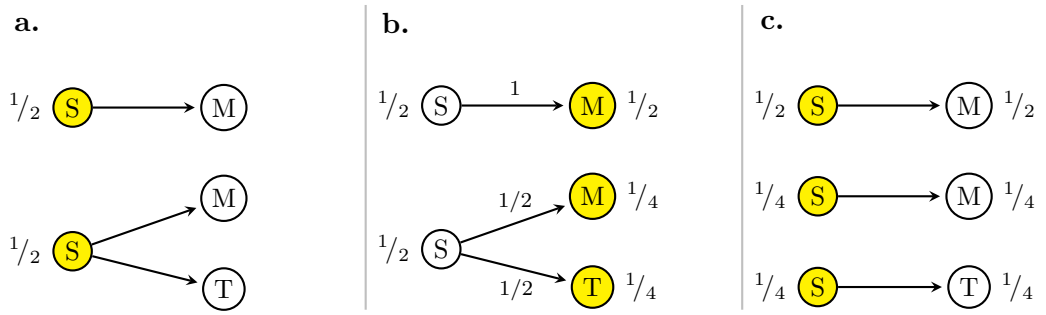

Figure 1: **shifting probabilities in branching worlds.** a. Fred's beliefs are divided between two possibilities: either he is about to be teleported to Monday or he is about to be teleported to both Monday and Tuesday. b. The degrees of belief are shifted to the successor points, weighted by the transition 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.

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 defined as the fraction of the agent's credence in v that should move 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 an approach 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 beliefs. For Fred, 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 v and its successors $w_1, \ldots, w_n$ always coincides with the range of acceptable ultimate prior probabilities conditional on $\{w_1, \ldots, w_n\}$. 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

quantum mechanical hypotheses. Roughly speaking, the requirement is 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. In the present framework, this requirement can be made precise 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 chance (as discussed in [Lewis 1980]).

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 chance, we could adopt the usual method of plugging the Principal Principle into Bayes's Theorem to show that an observation of O should raise your credence in H_2 to 0.75. The same happens if you follow shifted conditioning if the transition probabilities match the branch weights: after shifting, $H_1\&O$ 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 process 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 [2011] 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, it is argued that 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, nor do we have 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? Can it be derived from pragmatic considerations, in the tradition of Deutsch [1999] and Wallace [2012]?[10] I am skeptical about either proposal. Fortunately, we don't

10 [Greaves 2007b] pursues this second strategy. In this context, she defends a general model for belief update across fission. Her model is compatible with shifted conditioning, but the only centred propositions it considers are propositions about the agent's present branch in a universe. This actually complicates the situation, since post-branching probabilities are often 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-credence" defined over their successors' doxastic space. This quasi-credence then gets conditioned on the new evidence. Quasi-credence divides into genuine

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 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 imagination: couldn't Fred *imagine* waking up on Monday and not on Tuesday? He surely could. *Waking up on Monday* is an ordinary centred 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)? 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? Arguably not.

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* Fred_M and *being* Fred_T. The first possibility is true for Fred_M, the second for Fred_T. Each pre-fission possibility has a unique, non-branching successor.

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.

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 the pre-fission probabilities are divided between possibilities with only a Monday successor and other possibilities with only a Tuesday successor. The semantics of English is beside the point.[11]

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 realized 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 according to Ockhamism 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

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

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 modelled by "disambiguating" branching structures or not. Formally, this disambiguation is easily achieved, by construing atomic Ockhamist possibilities as ordered pairs of a possible world and a branch.[12]

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. So 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.

It is easy to translate back and forth between Ockhamist models and non-Ockhamist models (with genuine branching). In Ockhamist models, every 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 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.

12 [Saunders 2010] and [Wilson 2012] argue for a generalization of Lewis's metaphysics of persons on which Everettian branching is to be understood as *divergence*, making previously indistinguishable branches distinguishable. This view goes naturally with an Ockhamist epistemology, but again it is not required by Ockhamism.

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. However, Ockhamism has the technical advantage of preserving the equality between $P^{\succ}(A)$ and $P(\succ A)$. This will be exploited 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. As I argued in [Schwarz 2012], this account is supported by the very same arguments that are traditionally taken to support standard conditioning when uncentred propositions are ignored. In the present section, I want to illustrate this point further 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 us verify that shifted conditioning satisfies the following condition, strikingly violated by uncentred conditioning models in the Everett example.

> *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 credence in A to increase no matter which of $E_1, \ldots, E_n$ she learns.

The proof is simple. 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 agents who follow uncentred conditioning models are sometimes vulnerable to diachronic Dutch books. This should not happen to perfectly rational agents.

> *Dynamic coherence:* Rationality should not demand an agent to update their beliefs in such a way that they can incur a sure loss if they bet in accordance with their earlier and later beliefs.

To say that someone *bets in accordance with their beliefs* means that they accept any bet with positive expected payoff (if the only alternative is the status quo). Real people, of

course, don't always bet in accordance with their beliefs, but an otherwise rational agent who only cares about money arguably would. Rationality should not require such an agent to knowingly accept bets that amount to a sure loss. In [Schwarz 2012] I showed that every systematic alternative to shifted conditioning violates dynamic coherence. That shifted conditioning itself satisfies the condition can be proved as follows, adapting an argument from [Skyrms 1987].

Suppose for reductio that a diachronic Dutch book can be set up against an agent who updates her beliefs in line with shifted conditioning. A diachronic Dutch book can be represented by some (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 is the new evidence. The betting arrangement is a Dutch book if for every E_i, accepting B_1 together with $B_2(E_i)$ amounts to a net loss. Now consider one of the later bets $b \in B_2(E_i)$, to be placed if the new evidence is E_i. Let's say that b's net payoff is $\$X$ 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 b 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 another bet b' 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 bet b and the earlier bet b'. Substituting each of the bets b in $B_2(E_i)$ by a corresponding earlier bet b', 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 she is 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 apparent pragmatic nature of Dutch Book arguments has recently led some epistemologists to turn to considerations of epistemic utility or expected accuracy. Intuitively, the *inaccuracy* of a belief function represents the distance between the probabilities it assigns to all propositions and the truth-value of those propositions. A popular inaccuracy measure $I(P, w)$ for a belief function P at a world w is the Brier score $\sum_A |P(A) - w(A)|^2$, where A ranges over all propositions in P's algebra, and $w(A)$ is the truth-value of A at w. (For a defence 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, as judged by the previous belief function, among all functions P with $P(E) = 1$ (see [Leitgeb and Pettigrew 2010b]). More precisely, if propositions can change their truth-value, then what should be considered is not the expected *present* inaccuracy of the

candidate future belief function P, but its expected *future* inaccuracy. That is, 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 as judged by P_1 is not $\sum_w P_1(w)I(P,w)$, but $\sum_w P_1(\succ w)I(P,w) = \sum_w P_1(\succ w)\sum_A |P(A) - w(A)|^2$. The following constraint should be understood in this way.

> *Accuracy conduciveness:* If E is the evidence an agent receives at a given time, then her new credence function P_2 should have minimal expected future inaccuracy by the lights of the previous credence function P_1 among credence 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 relative to 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

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

For each proposition A, we can find the value $x = P(A)$ that minimizes $\sum_{w \in E} P_1(\succ w)|x - w(A)|^2$: $\frac{d}{dx}\sum_{w \in E} P_1(\succ w)|x - w(A)|^2 = 2(\sum_{w \in E} P_1(\succ w)x - P_1(\succ w)w(A)) = 2(P_1(\succ E)x - P_1(\succ (A \,\&\, E)))$ is zero iff $P_1(\succ E)x = P_1(\succ (A \,\&\, E))$, i.e. iff $x = P_1(\succ (A \,\&\, E))/P_1(\succ E) = P_1(\succ A/\succ E)$. So the function P with minimal expected future inaccuracy assigns to any proposition A the value $P_1(\succ A/\succ E) = P_1^{\succ}(A/E)$.[13]

The assumption of Ockhamism allowed us to ignore various subtleties that arise in cases of fission. How should we define the expected future inaccuracy of a belief function P as judged by P_1 in a non-Ockhamist model if the agent knows that the present point has several successors, so that P might be more inaccurate at one successor than at another? The most sensible choice is to average the degrees of inaccuracy weighted by the transition probabilities. This leads to the same result as the Ockhamism treatment. An alternative (discussed in a slightly different context in [Kierland and Monton 2005] and [Briggs 2010]) would be to use the expected *total* inaccuracy of all successors, irrespective of the transition probabilities. This has the implausible consequence that the mere number of successors becomes relevant: the more successors, the greater their total

13 The present argument generalizes to all quadratic inaccuracy measures. [Greaves and Wallace 2006] offer a related argument showing that conditioning maximizes expected epistemic utility for any utility measure U such that if E is a possible future evidence proposition and P a rational credence function, then $P(\cdot/E)$ has higher expected utility by its own lights than any alternative P'. Again the argument turns into an argument for shifted conditioning, given the requirement that $P(\cdot/\succ E)$ has higher expected utility by its own lights than any alternative P'.

inaccuracy. Intuitively, a group of people, all of whom share a mistaken belief, does not get closer to the truth just because one of them leaves.

The same issue arises for Dutch Books. In the Ockhamist framework, we effectively considered the net outcome of the pre-fission bets together with the bets accepted by any particular post-fission successor. We would reach the same conclusions if instead we considered the outcome of the pre-fission bets plus the weighted average of the outcome of all post-fission bets. As an alternative, we could simply add up the outcome of all bets, pre-fission and post-fission. You may have noticed that the Dutch book against uncentred conditioning models described in section 2 does not work from this perspective. Indeed, we can now construct a Dutch book against shifted conditioning. Imagine Fred only cares about money. On Sunday he should accept a deal that pays \$10 in case of tails and costs \$9 on heads. After updating by shifted conditioning, his successors are still undecided between heads and tails and should thus accept a deal that pays \$8 in case of heads and costs \$7 on tails. If the outcome is tails, that second deal gets offered twice, and the net payoff is \$-4. If the outcome is heads, the deal is offered only once and the net payoff is \$-1.[14]

The first thing to note about this "Dutch book" is that it only involves uncentred propositions. If it shows that shifted conditioning is not the right update rule in the centred worlds framework, it also shows that standard conditioning is not the right rule in the uncentred worlds framework. I don't think it reveals any such thing.

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.

7 Consequences and conclusions

Fission scenarios illustrate that even without self-locating ignorance, 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, after a "shifting" step that takes into account the agent's (anticipated) change of location. The proposed rule gives sensible results in cases of fission and is supported by the same general arguments that support standard conditioning in models with only uncentred propositions. My proposal does not solve the confirmation-theoretic challenge to the Everett interpretation, but it might give us a better grip on

14 This mirrors the Dutch book argument against Lewisian halfing in [Hitchcock 2004].

what is needed. It is not important whether, conditional on a specific fission scenario, agents can be uncertain about the future, nor 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 a principle of self-locating indifference for posterior beliefs as assumed e.g. in [Elga 2000] and [Lewis 2001]. Suppose Fred is initially undecided between a fission and a non-fission possibility *within the same uncentred world*. For example, suppose he knows that the universe contains two Earth-like planets that up to now have been perfect intrinsic duplicates, one of which is about to undergo fission. If his present credence in both possibilities is 1/2, then dynamic stability requires that his new credence is divided 1/4 – 1/4 – 1/2 between the resulting alternatives, although they are located within the same uncentred world, even though Fred's evidence after the fission is arguably neutral between the three locations. Self-locating indifference may still hold for ultimate priors: it is compatible with the present model that Fred's ultimate priors, conditional on his new evidence, should have satisfied the indifference principle. This suggests that there can be disagreement between perfectly rational agents with the very same priors and the same evidence.

On a related note, 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 much further away than Monday and the coin that decides whether Fred gets teleported to Tuesday is 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, what does the present account say about 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 on two further questions I have not discussed 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 second-order norms linking new degrees of beliefs to new beliefs about previous beliefs. As shown in [Schwarz 2012], the second-order version of shifted conditioning supports the standard "thirder" solution (while still supporting the "halfer" solution for Fred). 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 further question how an agent should update her 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.

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

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

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 [1981]: *The First Person: An Essay on Reference and Intentionality*. Minneapolis: University of Minnesota Press

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 [2004]: "Understanding Deutsch's probability in a deterministic multiverse". *Studies in History and Philosophy of Modern Physics*, 35: 423–456

— [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

Hilary Greaves and David Wallace [2006]: "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility". *Mind*, 115: 607–632

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

Richard Jeffrey [1992]: *Probability and the Art of Judgment*. Cambridge: Cambridge University Press

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

Brian Kierland and Bradley Monton [2005]: "Minimizing Inaccuracy for Self-Locating Beliefs". *Philosophy and Phenomenological Research*, 70(2): 384–395

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 Lewis's *Philosophical Papers*, Vol. 1, 1983.

— [1980]: "A Subjectivist's Guide to Objective Chance". In Richard Jeffrey (Ed.), *Studies in Inductive Logic and Probability* Vol. 2, University of California Press. Reprinted in Lewis's *Philosophical Papers*, Vol. 2, 1986.

— [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

Moritz Schulz [2010]: "The Dynamics of Indexical Belief". *Erkenntnis*, 72(3)

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

— [2011]: "Evidence and Uncertainty in Everett's Multiverse". *British Journal for the Philosophy of Science*, 62: 99–123

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

— [2013]: *Quitting Certainties*. Oxford: Oxford University Press

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