Diachronic norms for self-locating beliefs*

Wolfgang Schwarz

In *Ergo* 4 (2017): 709–738

Note (2023): In the published version of this paper, some superscripts with a plus ('+') should instead have a minus ('-'). Thanks to Michael Rescorla for pointing this out. The errors are here corrected.

Abstract. How should rational beliefs change over time? The standard Bayesian answer is: by conditionalization (a.k.a. Bayes' Rule). But conditionalization is not an adequate rule for updating beliefs in "centred" propositions whose truth-value may itself change over time. In response, some have suggested that the objects of belief must be uncentred; others have suggested that beliefs in centred propositions are not subject to diachronic norms. I argue that these views do not offer a satisfactory account of self-locating beliefs and their dynamics. A third response is to replace conditionalization by a new norm that can deal with centred propositions. I critically survey a number of new norms that have been proposed, and defend one particular approach.

1 Introduction

Two epistemic norms form the core of classical Bayesianism. The first, *probabilism*, is synchronic; it says that rational degrees of belief conform to the probability calculus. The second, *conditionalization* (or *Bayes' Rule*), is diachronic; it specifies how rational degrees of belief change as new evidence arrives. In the simplest case, where the new evidence is captured by a single proposition E that is learned with certainty, conditionalization says that the new credence Cr_{t+1} in any proposition A should equal the previous credence Cr_t conditional on E:[1]

$$\text{(C)} \qquad \mathrm{Cr}_{t+1}(A) = \mathrm{Cr}_t(A/E).$$

* Thanks to two anonymous referees, one of which was Michael Titelbaum, for helpful comments on an earlier version.

1 I will have more to say on what counts as the new belief state for a given 'previous' belief state in section 6.

If the new evidence is equivocal, determining new probabilities $x_1, \ldots, x_n$ over some evidence partition $E_1, \ldots, E_n$, (C) generalizes to (JC) (see [Jeffrey 1965]):

$$\text{(JC)} \qquad \mathrm{Cr}_{t+1}(A) = \sum_i \mathrm{Cr}_t(A/E_i)x_i.$$

Many arguments have been given in support of these rules. [Lewis 1999b] (first reported in [Teller 1973]) and [Skyrms 1987] show that probabilistically coherent agents are vulnerable to diachronic Dutch Books if and only if they do not change their beliefs by conditionalization; [Teller 1973] and [Williams 1980] show that conditionalization optimally preserves beliefs that are independent of the new evidence; [Greaves and Wallace 2006] and [Leitgeb and Pettigrew 2010] show that conditionalization maximizes the expected accuracy of the new belief state relative to the old belief state. These results all assume that the relevant propositions (A and E, or A and $E_1, \ldots, E_n$) do not change their truth-value from one time to another. They do not support conditionalization if rational credence is defined over *centred* propositions whose truth-value is not fixed once and for all. Indeed, it is widely recognised that conditionalization then becomes inapplicable.

Different authors have drawn different lessons from this observation. Some have concluded that credences should always be construed as uncentred. Others have taken the problem as (further) evidence that there are no diachronic norms on rational credence at all. Yet others have suggested that diachronic norms must be restricted to uncentred propositions. I will review these proposals in section 2 and argue that they do not provide a fully satisfactory response.

Another type of response is to develop new rules for the dynamics of "self-locating" beliefs. In sections 3–6, I will look at the main rules that have been proposed to this end, both in philosophy and in theoretical computer science. One such rule, discussed in section 3, is *imaging*; I argue that it yields sensible results only in a very limited range of cases. In section 4, I turn to a more promising approach on which updating involves a process of first *shifting* the centres of one's doxastically accessible worlds and then conditionalizing on the new evidence. This account seems to presuppose that an agent's belief updates are perfectly synchronised with objective time. In sections 5 and 6, I consider three ways of modifying the account so as to avoid this assumption. I argue that two of the modifications are problematic, but the third one seems to work. Finally, in section 7, I return to an alternative that I set aside as incomplete in section 2 and show how it may be completed. At that point, we will have two answers, but they turn out to be almost equivalent. I will not settle the choice between the two.

A caveat before we begin. The proposals I will discuss have been developed in different contexts, with different background assumptions, different notation, and a focus on different problems. Here I will abstract away from these differences, presenting the core of the relevant proposals as they bear on my topic, in my own notation. I will also ignore various subtleties in the discussed proposals that do not affect the points I will make. For the full picture, the reader is advised to consult the cited

2 The survey [Titelbaum 2016] may also be useful, as it approaches the present topic from a slightly different angle—asking how agents should "coordinate" their credences at different times—and looks at some proposals in more depth. On the other hand, Titelbaum's survey does not discuss views from theoretical computer science and treats all the proposals from sections 4–6 as one, for which [Kim 2009] is taken as representative. As I will explain in footnote 9 below, on the way I draw the lines it is doubtful whether [Kim 2009] even belongs to this family.

works.[2]

2 Self-location and conditionalization

To motivate the search for new norms on the dynamics of self-locating belief—and to clarify what exactly we are looking for—let me briefly review why we need centred objects of credence, and why we should believe that such credences are subject to diachronic norms.

The Bayesian concept of credence is a technical term. Whether we should allow for centred objects of credence therefore depends on the theoretical roles credences are meant to play. One key role lies in the theory of rational choice. According to Bayesian decision theory, rational agents choose actions that maximize expected utility relative to the agents' credences and utilities; this is how their behaviour can be explained by their (graded) beliefs and desires. Now as Perry [1977] vividly pointed out, the actions we choose often depend not only on our beliefs about the world as a whole, but also on where we locate ourselves in space and time. You and I may agree that somebody is being attacked by a bear, and on all other relevant propositions about the objective world, but if you locate yourself as the person under attack while I take myself to be an onlooker, our rational response will be very different. Similarly, we may take quite different actions if you believe that an important meeting starts in half an hour while I believe that it starts now, even if we both agree that it starts at noon. If credences are to play their standard role in guiding and explaining actions, these observations suggest that our credences are not exhausted by our views about the objective world: it also matters where we locate ourselves within the world.

Another key role for the concept of credence lies in Bayesian confirmation theory, which models how and to what extent hypotheses are supported by an agent's evidence. Here, too, allowing for centred hypotheses and centred evidence has proved useful—for example, when trying to understand how the evidence that things *around us* are such-and-such bears on the hypothesis that the universe contains many (or few) places where things are such-and-such (see e.g. [Bostrom 2002], [Sebens and Carroll 2017], [Arntzenius and Dorr 2017]). Indeed, as Lewis [1979] pointed out, one can seemingly imagine an agent who knows all objective facts about the world from a God's eye perspective but is still ignorant about who and where she is in the world, and who might learn so through further evidence.

The standard way to accommodate these phenomena, going back to [Lewis 1979], is to construe the objects of credence in such a way that their truth-value may vary not only from possible world to possible world, but also from time to time, from place to place, and from individual to individual. Suitable objects are not hard to find. Lewis suggests identifying the objects of credence with properties; a property is "true" relative to a given individual at a time and a world just in case the individual instantiates the property at the time and the world. Other authors use sentence types as objects of credence, drawing on the natural sense in which, for example, the English expression 'it is raining' is true at some times and places and false at others. Another popular idea is to construe objects of credences as sets of triples of an uncentred world, a time, and an individual; the set is "true" relative to an individual at a time and a world just in case the corresponding triple is in the set. Yet another option is to take centred propositions as theoretical primitives, or to construe them as states of affairs

in the tradition of [Plantinga 1974] and [Pollock 1984]: a state of affairs like *the sun shining* can plausibly obtain at some times and places and not at others.

For what follows, the difference between these proposals will not be important. As a neutral label, I will call any object in the domain of an agent's credence function a *(centred) proposition*. Classical, uncentred propositions can be treated as limit cases of centred propositions (as [Lewis 1979] explains).

Probability theory requires that the space of propositions is closed under conjunction, disjunction, and negation. To simplify the following discussion, I will make the slightly stronger assumption that the propositions form a complete atomic Boolean algebra, so that propositions can be identified with sets of atoms in the algebra, where an atom is a maximally consistent conjunction of propositions. I will refer to these atoms as *centred worlds*.[3]

It is sometimes claimed that the phenomena reviewed above are instances of Frege's puzzle which supposedly has been solved without departing from classical views about the nature of propositions (e.g. [Cappelen and Dever 2013], [Magidor 2015]). I disagree. The bear case, for example, does not involve ignorance of identities—the defining feature of Frege's Puzzle. But it does not matter. Even if we grant the objection, we'd arguably still need centred objects of credence. To illustrate, suppose we follow [Salmon 1986] and explain Frege's puzzle by appealing to "guises" under which classical propositions are believed. To handle the above observations in a parallel fashion, we will have to invoke indexical guises, and we will have to identify the objects of credence with propositions-under-guises, since the guises matter to confirmation and rational choice. As a result, the objects of credence—propositions in the present, stipulative sense—will be centred. We can happily allow that "propositions" in some other sense (referents of 'that'-clauses, perhaps) are uncentred, but since our topic is the dynamics of credence, that other sense is not immediately relevant.

To be sure, one might still try to account for the reviewed phenomena in some other way, without making the objects of credence centred. The most extended defence of this strategy is due to Robert Stalnaker (see [Stalnaker 1981], [Stalnaker 2008: ch.3], [Stalnaker 2014: ch.5], [Stalnaker 2016]), who suggests modelling an agent's doxastic state by an uncentred credence function together with "links" representing where the agent locates herself relative to any uncentred world she deems possible. The dynamics of these states is complicated because both the underlying probability space and the links frequently change, in ways Stalnaker does not fully explain.

Without exploring such alternatives any further, I will from now on assume that the objects of credence are centred.

We must then be careful when we think about relations between agents who locate themselves at different places or times. For example, suppose you utter 'it is raining' and thereby express high credence in the centred proposition that it is raining.[4] Even if I trust your assertion, I may not come

3 If propositions are construed as sentences or states of affairs (say), then they are of course not literally identical to sets of maximally consistent conjunctions of propositions. However, under plausible assumptions, the quotient algebra of the propositions under the relation of logical equivalence will still be a complete atomic Boolean algebra, and hence isomorphic to the power set algebra of the atoms. (Intuitively, since every proposition is logically equivalent to a disjunction of maximally consistent conjunctions, and since probability theory requires logically equivalent propositions to have the same probability, we can harmlessly use sets of maximally consistent conjunctions as proxies for propositions, interpreting the sets as disjunctions of their members.)

4 By 'the centred proposition that it is raining' I mean a centred proposition that is true at a place and a time and a

to assign high credence to the same centred proposition: if I believe that you are located 500 km to the North of me (we are talking on the phone), I will rather come to believe that it is raining 500 km to the North. So successful communication is not simply a matter of transferring beliefs (see e.g. [Weber 2013]). Similarly, if you assign high credence to the centred proposition that it is raining, and I assign low credence to that same proposition, then this does not constitute a genuine disagreement between you and me. The intuitive concepts of agreeing, disagreeing, having the same belief, etc., therefore can't be analysed simply in terms of assigning high credence to the same proposition or to incompatible propositions.[5]

Analogous problems arise for beliefs of the same agent at different times—which brings us back to conditionalization. Suppose after our conversation on the phone you fall asleep; when you later wake up, you do not receive any new information about the weather. Nonetheless, your confidence in the centred proposition that it is raining should decrease, especially if you have reasons to believe that it wouldn't rain all day. You should still be confident that it *was* raining when you fell asleep, but you should be less confident that is *is* raining. It is hard to see how this change could come about by conditionalization. Conditionalization only reduces your credence in a proposition if the new evidence is relevant to that proposition, but by assumption you do not gain any new evidence about the weather: whatever evidence E you receive, your previous credence in *rain* conditional on E may well have been high, yet your new credence is low.

Conditionalization is a rule for revising one's beliefs about the state of the world in light of new information about that same state. It fails to take into account that the world itself may change: that what was true before may have come to be false.

Some authors have taken the present difficulty as (further) support for the view that we should dispense with diachronic constraints on rational belief. According to these authors, what one should believe at any point of time is simply a matter of one's evidence at that time; the earlier beliefs don't matter. More precisely, if E captures an agent's present total evidence, then (on the present view) her credence should equal some *ur-prior* P conditional on E, where the ur-prior is something like a measure of evidential support, not the agent's previous credence function (see e.g. [Levi 1980], [Williamson 2000], [Christensen 2000], [Horgan 2008], [Moss 2014], [Hedden 2015], [Meacham 2016], [Arntzenius and Dorr 2017]).

At first glance, *ur-prior conditionalization* (as Meacham [2016] calls it) seems to elegantly avoid the problem of updating self-locating beliefs. On closer scrutiny, however, the problem is not so easily discharged.

The issue turns on the interpretation of 'evidence'. In classical Bayesianism (as employed for example in artificial intelligence), the new evidence E that is plugged into conditionalization can be understood as the information conveyed to the agent through her senses at the relevant time.[6] Such a

possible world just in case it is raining at that place at that time in that world. I will use this shorthand form throughout the present paper.

5 A popular objection to centring objects of credence is that this would yield false predictions about when two subjects agree/disagree or believe the same/different things (e.g. [Stalnaker 2008: 50], [Stalnaker 2014: 114], [Bradley 2013]). But the problematic predictions follow only under a naive and entirely optional analysis of these concepts. (Recall that credence is a technical term.)

6 What is the information conveyed to an agent through her senses? Good question. It is certainly not what philosophers of perception call the 'content' of perceptual experience—roughly, the conditions under which we would judge the

sensory conception of evidence would render ur-prior conditionalization utterly implausible: the vast majority of our beliefs are not supported by present sensory evidence, yet that does not make them irrational. Ur-prior conditionalization therefore requires a different conception of evidence on which our total evidence includes information we acquired on earlier occasions—something like Lewis's [1996: 424] notion of evidence as 'perceptual experience and memory', where 'memory' covers not only occurrent episodes of remembering (which would still leave almost all our beliefs unsupported by present evidence), but also sub-consciously retained information.

It is here where the problem of updating reappears. Before you fell asleep, you see that it is raining; upon awakening, you remember that it *was* raining. Somehow your earlier evidence that it *is* raining transformed into your new evidence that it *was* raining. What are the rules for these transformations? Clearly not anything goes. Suppose the information that it is raining had transformed into a false memory that the Earth is flat, so that you had woken up believing that the Earth is flat. Something would have gone wrong; your new belief would not be justified. So there are normative constraints on the dynamics of (non-sensory) evidence. What are these constraints?[7]

To be clear, this does not show that the ur-prior account is wrong. My claim is only that it is incomplete, and that it does not escape the problem of updating self-locating information.

The same is true for a closely related family of views on which there are normative constraints on how uncentred credences should evolve over time, but no such constraints for centred credences. On the most popular way of developing this idea, your new credences are determined by first conditionalizing your previous uncentred credences (i.e., your credences over uncentred propositions) on the uncentred information provided by your new evidence and then using the self-locating part of your new evidence to determine your location (see [Piccione and Rubinstein 1997], [Halpern 2006], [Meacham 2008], [Titelbaum 2008], [Briggs 2010], [Titelbaum 2013][8]). [Moss 2012] defends a

relevant experience to be veridical. For it would often be irrational to become absolutely certain that these conditions obtain. Jeffrey [1988] suggests that sensory information is best modelled as probabilistic information, in which case it will plausibly be sensitive to the agent's prior beliefs (see also [Field 1978], [Garber 1980], [Christensen 1992], [Weisberg 2009]). There are thorny problems here, but exploring them would take us too far afield.

7 It is tempting to think that non-occurrent memory simply is belief: I will remember that Astana is in Kazakhstan as long as I keep believing that Astana is in Kazakhstan; no second attitude is required or involved. If that is right, then taking an agent's non-occurrent memories as given amounts to taking a substantial part of their new beliefs as given; the missing story on the dynamics of memory is a missing story on the dynamics of belief.

8 The line between these accounts and ur-prior accounts is blurry because stepwise conditionalization on uncentred evidence is equivalent to conditionalizing the initial credence on the cumulative uncentred evidence. The account developed in [Titelbaum 2008] and [Titelbaum 2013] looks superficially different from the summary I've just given, but has essentially the same structure. Titelbaum notes that agents can often "translate" any given centred proposition into an uncentred proposition which is certain to have the same truth-value and therefore must have the same probability. For example, if at t you are certain that nobody else in the history of the universe ever had or ever will have the very same total phenomenal experience X that you have right now, then you can be certain that *it is raining* has the same truth-value as *it is raining at the unique time and place where someone has experience X*. Assuming that conditionalization is suitable for uncentred propositions, Titelbaum suggests that your new credence in the translated propositions should equal your previous credence conditionalized on your uncentred new evidence. To determine the new probability of centred propositions, we then have to use the centred part of your evidence to find new translations between centred and uncentred propositions. For example, if your new evidence entails that you have phenomenal experience Y, your new credence in *it is raining* might equal your updated credence in *it is raining at the unique time and place where someone has Y*. The upshot is that the new credences are determined by conditionalizing the uncentred previous credences on the uncentred evidence and then re-introducing the centres based on the new evidence.

variant approach on which your new credences are determined by first setting aside the previous self-locating beliefs, then using a certain part of your new evidence (provided by your "sense of time") to determine your location, and finally conditionalizing on the remainder of your centred and uncentred evidence.[9]

Like ur-prior accounts, these accounts plausibly require a non-sensory conception of evidence. Imagine an agent—a robot perhaps—whose senses provide only limited information about the arrangement and colour of mid-sized objects in her environment. In order to reach some goal, the agent has to pass through two rooms with identical interiors; in the first she has to exit on the left, in the second on the right. Having entered the first room, the agent's previous uncentred beliefs together with her sensory evidence do not settle whether she is in the first room or in the second. But surely the agent should not become uncertain about where she is. To deliver this verdict, the accounts just reviewed must assume that the agent's evidence includes the information that she is in the first room, determined by some transformation from her previous belief that she is about to enter that room.

Again, the problem is not so much that these views are wrong, but that they are incomplete: they do not give a full account of how rational beliefs evolve over time.[10]

Until further notice, I will henceforth reserve the term 'evidence' for sensory evidence. Our question is how an agent's credences over centred and uncentred propositions should change as time passes and new information arrives from the agent's senses. There may be reasons for tackling the question indirectly, by first defining a richer notion of evidence that includes information the agent received at earlier times, suitably transformed to take into account the passage of time. In section 7 I will explain how the relevant evidence transformations might go, and thus how the accounts just reviewed might be completed. In the meantime, I will look at accounts on which an agent's credences are determined directly by her earlier credences and her new (sensory) evidence, without appeal to a further conception of non-sensory evidence that is subject to diachronic norms of its own.

3 Imaging

In [Katsuno and Mendelzon 1991], a classical paper in theoretical computer science, Katsuno and Mendelzon distinguish two rules for updating a knowledge base. The rules are best introduced by example.

> At t you are confident that a certain basket contains either two apples or an apple and a banana; you give equal credence to both possibilities. At $t+1$ you learn E: that there is no banana in the basket.

You might treat the new information as a reason to revise your beliefs, concluding that there are two apples in the basket. But you might alternatively treat E as a message about how the world has

9 [Kim 2009] defends a similar account, but he does not say where the new beliefs about the time come from: whether they are determined by some combination of the previous uncentred beliefs and the new evidence—as in Moss's account—or whether the previous centred beliefs also play a role. I will have a little more to say on the proposals of Moss and Kim in footnotes 14, 16, 17, and 20 below.

10 In addition, most (if not all) the views I have reviewed deliver highly implausible verdicts in cases where the evidence

changed, without revealing anything new about what was the case at t. In particular, you may come to know E by learning that someone went out to remove any bananas from the basket. In that case, you should not conclude that there are two apples in the basket. Rather, your new credence should be divided between the hypothesis that the basket contains one apple and the hypothesis that it contains two.

The first kind of belief change Katsuno and Mendelzon call *revision*, the second *update*. Updating, they suggest, is "bringing the knowledge base up to date when the world described by it changes" [1991: 387]. Working in the AGM tradition of [Alchourrón et al. 1985], they characterize the two processes by axioms for operations on a qualitative (non-probabilistic) knowledge base. The probabilistic analogs of these two operations (as explained e.g. in [Kern-Isberner 2001] and [Walliser and Zwirn 2002]) are conditionalization and imaging.

Imaging was originally introduced in [Lewis 1976b] for a somewhat different purpose. It works as follows. Let P be a probability measure over some "worlds" W; let $f : W \times \mathcal{P}(W) \to W$ be a function that maps any world w and proposition E to whichever E-world is most similar to w, relative to some fixed similarity ordering; finally, let $[[A]]^w$ denote the truth-value of A at w (1 = true, 0 = false). Then the *image of P on E*, which I'll write as $P(\cdot//E)$ (with two dashes), is defined by

$$P(A//E) = \sum_{w \in W} P(w)[[A]]^{f(w,E)}.$$

Intuitively, imaging on E shifts the probability mass of every world to the "most similar" world that satisfies E.

So we have a first proposal of how an agent's credences should evolve in order to keep track of a changing world: if E is the new evidence received at $t+1$, then the new credence should equal the previous credence imaged on E:

$$\text{(I)} \qquad \mathrm{Cr}_{t+1}(A) = \mathrm{Cr}_t(A//E).$$

The rule can be generalized to cases where the evidence is equivocal or where there is more than one "most similar" world, but let's stick with the easy case.

I have put 'most similar' in scare quotes because the relevant function f need not track intuitive similarity. For the present purpose, f should rather assign to w and E whatever state of the world would result from minimally changing w into a state where E is true.

To see all this in action, return to the fruit basket scenario. At t, your credence is divided between two (coarse-grained) centred worlds or world states: a two-apples-no-banana world and a one-apple-one-banana world. Learning that somebody intervened so as to make it true that there are no bananas in the basket, your credence in the one-apple-one-banana world is moved to a one-apple-no-banana world, because that is the world that results from the one-apple-one-banana world by making 'no bananas' true. Your credence in the two-apples-no-banana world remains unchanged, because that world already satisfies the 'no bananas' constraint.

The imaging account is popular in some quarters of theoretical computer science, but has gained

does not suffice to determine a unique centre in every accessible uncentred world; see e.g. [Bostrom 2002: chs. 7 & 9], [Meacham 2008: 260–265], [Schwarz 2015: 662-665], [Arntzenius and Dorr 2017: sec.2].

almost no traction among philosophers (though see [Leitgeb 2016]). For good reasons.

A minor point first. It is not an a priori truth that removing a banana from a basket containing an apple and a banana leads to a state in which the basket contains just an apple. So either the imaging function f encodes contingent causal information about the world—in which case one would like to know how *this* information is updated over time—or the agent's probabilities must be defined over much more fine-grained states than usually assumed, so that a given world fixes not only how many fruits are in the basket but also what would happen under various interventions. Let's assume we have such a fine-grained representation.

The main problem with the imaging rule (and its non-probabilistic analog of 'updating') is that it only applies under conditions that are almost never satisfied: when the information one receives about a change in the world reveals nothing about what the world was like before.[11] Suppose you learn not that there is no banana in the basket, but more specifically that a banana has been removed from the basket. This is information about a change, but it also reveals that the basket previously contained a banana. Upon receiving the information, you should become confident that the basket contains one apple and no banana, but imaging does not apply (and would not deliver that result[12]). Here we need a different rule for "bringing the knowledge base up to date".

It is hard to think of any realistic case where the information one receives is entirely neutral on what the world was like before. The most promising examples involve decisions or commands. If you are certain all along that you could make E true and then decide to make E true, one might think that your decision carries no news about what the world was like before. But even that is not generally true. In Newcomb's Problem, for example, a decision to two-box indicates that the opaque box is empty; the epistemic impact of reaching the decision from a prior state of indecision is therefore not adequately modelled by imaging.

Even in the rare and unusual case where the information one receives sheds no light on what the world was like before, imaging can give wrong results because changes in the world are not always revealed to the agent. Suppose again at t you believe that the fruit basket contains either two apples or an apple and a banana. But suppose you also believe that a fly is sitting on the basket. Then you instruct somebody to remove all bananas, and at $t+1$ you learn that the instruction has been carried out: there are no bananas in the basket. Imaging preserves your credence in scenarios that already satisfy the new information (no bananas). Thus it would leave you certain that if the basket contains two apples then there is a fly on the basket. But if you know that flies generally don't sit at the same place for long, you should not remain certain about this.

In sum, imaging does not provide a satisfactory answer to our question. It only applies in rare and unusual cases, and even then tends to give the wrong answers.

4 Shifting

The next approach I am going to discuss is the *de facto* standard in control theory and artificial intelligence (see e.g. [Russell and Norvig 2010: ch.15]).

11 I am not the first to make this complaint: see e.g. [Boutilier 1998] and [Lang 2007].

12 The result of applying imaging depends on what counts as the closest world to a two-apples-no-banana world at which

We just saw that in order to describe how an agent's beliefs should change over time, we need to consider not only her beliefs about the present state of the world, but also her beliefs about how this state may evolve, either as the result of her actions or by itself. In artificial intelligence, such beliefs are commonly represented by a *transition model* which is added to the agent's probability measure over world states. In the simplest case—assuming that time is linear and discrete, that the immediate future depends only on the present, and that the agent doesn't intervene in the course of nature[13] —the transition model defines a conditional probability over world states at any time $t+1$ conditional on the state at the previous time t.

With the transition model in place, there is an obvious two-step process for computing the probabilities over the new world state at $t+1$ in the light of new evidence: first, the previous probability over states at time t is projected forward to $t+1$ by the transition model; then the result is conditionalized on the new evidence.

Unfortunately, presentations of this approach tend to remain unclear on points that are central to our present topic. For example, equation (15.5) in [Russell and Norvig 2010: 572]—which in this respect is representative of the entire literature—expresses the complete update by the following equation:

$$\text{(RN)} \quad P(X_{t+1}/e_1,\dots,e_{t+1}) = \alpha P(e_{t+1}/X_{t+1}) \sum_{x_t} P(X_{t+1}/x_t) P(x_t/e_1,\dots,e_t).$$

Here, $P(X_{t+1}/e_1,\dots,e_{t+1})$ is meant to be the agent's probability at time $t+1$ over possible world states X_{t+1}, after having received evidence $e_1,\dots,e_{t+1}$ at times 1 through $t+1$; $P(e_{t+1}/X_{t+1})$ is the probability of the (sensory) evidence e_{t+1} given state X_{t+1}, as specified by the agent's "sensor model"; $P(X_{t+1}/x_t)$ is the probability of X_{t+1} given a particular hypothesis x_t about the previous state, as specified by the transition model; $P(x_t/e_1,\dots,e_t)$ is the agent's previous credence in x_t; α is a normalizing constant (the denominator in the application of Bayes' Theorem to compute conditionalization).

Since we are interested in how subjective probabilities should evolve over time, it is advisable to use different labels for the probabilities at different times—'Cr_{t+1}' and 'Cr_t', rather than simply 'P'. Dealing with ideal agents, we can assume that the agent's probability measure at any time incorporates the evidence she has received up to then, so there is no need to explicitly conditionalize these measures on the history of evidence; we can write '$\mathrm{Cr}_t(X_t)$' instead of '$P(X_t/e_1,\dots,e_t)$'. Moreover, having added time indices to the probability measures, we should arguably remove them from the objects of probability, to make clear that Cr_t and Cr_{t+1} can assign probabilities to centred world states (like *there being two apples in the basket*), not just to uncentred propositions specifying the state of the world at a given time (*there being two apples in the basket at t*). This fits the informal discussion in [Russell and Norvig 2010] and elsewhere, and in any case is required for the present proposal to bear on our topic.

The transition model specifies how world states may change from one point of time to the next. We could represent this by a primitive binary probability measure over centred propositions, but it will be useful to make the content of the relevant attitudes more explicit. After all, what is captured

a banana has been removed. Presumably it's a world where a banana is first added and then removed. Imaging would then leave you undecided between a one-apples-no-banana world and a two-apples-no-banana world.

13 All these assumptions can be dropped, leading to more complicated models. I stick to the simplest case because the issues I am going to discuss do not depend on the complications.

by the transition model are ordinary conditional beliefs about how the world may change: about whether B is going to be true given that A is true now. To model these beliefs, I will ignore matters of computational tractability and assume that the agent has a joint probability measure over the present state of the world, its past, and its future. I will continue to assume that time is linear and discrete, so that any centred world in the algebra of propositions fixes all relevant facts about the present state, its ancestors, and descendants.

The following piece of notation proves useful. For any centred world w and integer n, let w^{+n} be an otherwise identical world in which the centre is shifted n units of time into the past.* How this is cashed out depends on the construction of centred worlds. For example, if centred worlds are triples of an uncentred world, an individual, and a time, then $(u, i, t)^{+n}$ may be $(u, i, t-n)$; if centred worlds are sentences, then the $+n$ operation prefixes the relevant sentence with 'in n units of time'. (We'll come to reconsider this interpretation of w^{+n} in section 6.) For any set A of centred worlds, let

$$A^{+n} = \{w^{+n} : w \in A\}.$$

The shifting operator $+n$ allows us to make explicit the agent's transition model, her beliefs about how the world may change. For example, if the agent at t is 90% confident that it is going to rain in one unit of time given that it is raining now, then $\mathrm{Cr}_t(\mathit{rain}^{+1}/\mathit{rain}) = .9$.

Now return to the two-step update described by (RN). The first step was to project the probabilities from t to $t+1$ by the transition model. The shifting operation makes this step easy to define: the new probability of any centred proposition A is simply the old probability of A^{+1}. For example, if Cr_t assigns probability .9 to rain^{+1} ('it is going to rain in 1 unit of time'), then the updated probability assigns probability .9 to *rain* ('it is raining now'). Let's denote the shifted probability measure that results from this step by 'Cr_t^{+1}'. The concept obviously generalizes to larger (and negative) intervals:

$$\mathrm{Cr}_t^{+n}(A) = \mathrm{Cr}_t(A^{+n}).$$

In the second step, the shifted probability measure is conditionalized on the evidence. If the evidence is captured by a (possibly centred) proposition E that is learnt with certainty, the whole update therefore looks as follows:

$$\text{(SC)} \quad \mathrm{Cr}_{t+1}(A) = \mathrm{Cr}_t^{+1}(A/E) = \mathrm{Cr}_t(A^{+1}/E^{+1}).$$

Since the second step is plain old conditionalization, generalizations to cases of uncertain and equivocal evidence are straightforward.

(SC) is the natural interpretation of (RN) if we assume that the objects of credence can be centred. In philosophy, (SC) has been defended in [Meacham 2010], [Schwarz 2012], and [Schwarz 2015].[14]

∗ **Correction 2023:** The published version says 'into the future'. In the next sentence, it says '$(u, i, t+n)$' instead of

14 [Schulz 2010] defends essentially the same rule, without isolating shifting as a separate step, and restricts it to cases where the agent keeps track of time. (I will turn to this issue in the next section.) [Kim 2009] defends a similarly restricted norm with a similar form but where the relative shifts are replaced by shifts to an absolute time: $\mathrm{Cr}_{t+1}(A) = \mathrm{Cr}_t(A \textit{ at } i/E \textit{ at } i)$, provided the agent is certain at $t+1$ that the present time is i. Kim does not explain where the agent's new beliefs about the time come from.

In [Schwarz 2012] and [Schwarz 2015] I show that (SC) inherits many features of conditionalization if the probability space is extended by centred propositions. For example, $\mathrm{Cr}_t^{+1}(\cdot/E)$ maximizes expected future accuracy, and agents are vulnerable to diachronic Dutch Books if and only if they violate (SC).

(SC) correctly deals with all our fruit basket examples. Recall that imaging accounts seemed to give the right result in the original case, assuming that the new information, that there are no bananas in the basket, is interpreted as information about an intervention: that somebody went ahead and removed any bananas from the basket. Since this is not equivalent to the simpler proposition that there are no bananas in the basket, we should include a corresponding event in the probability space. So the space should include events A (one apple, no banana), AA (two apples, no banana), AB (one apple, one banana), and BR (any bananas removed). To apply (SC), we also need A^{+1}, AA^{+1}, AB^{+1} and BR^{+1} so that we can model your beliefs at t about how the world might change. Let's assume that at time t you were neutral not only about whether there are two apples or an apple and a banana in the basket, but also about whether any bananas were going to be removed. Moreover, you regarded the two issues as independent, and you were certain that if the bananas won't be removed then all fruits will remain in the basket.[15] Then

$$\begin{aligned}
&\mathrm{Cr}_t(AA^{+1} \wedge BR^{+1}) = \mathrm{Cr}_t^{+1}(AA \wedge BR) = .25\\
&\mathrm{Cr}_t(AA^{+1} \wedge \neg BR^{+1}) = \mathrm{Cr}_t^{+1}(AA \wedge \neg BR) = .25\\
&\mathrm{Cr}_t(A^{+1} \wedge BR^{+1}) = \mathrm{Cr}_t^{+1}(A \wedge BR) = .25\\
&\mathrm{Cr}_t(AB^{+1} \wedge \neg BR^{+1}) = \mathrm{Cr}_t^{+1}(AB \wedge \neg BR) = .25
\end{aligned}$$

Conditionalizing the shifted probability Cr_t^{+1} on BR leaves you with $\mathrm{Cr}_{t+1}(AA) = \mathrm{Cr}_{t+1}(A) = .5$. So (SC) still gets this case right. But it also applies, and gives sensible results, in cases where imaging accounts fall silent. For example, if BT is the proposition that a banana has been taken from the basket, $\mathrm{Cr}_t(AA^{+1} \wedge BT^{+1}) = 0$, and so $\mathrm{Cr}_{t+1}(A) = 1$.

So far, so good. Unfortunately, there are cases which (SC) seems to get wrong, notably cases where the agent loses track of time.

5 Losing track of time

If I have not just looked at a clock, I usually don't know the exact time. This suggests that I am not updating my credences in accordance with (SC). To see why, imagine for the sake of vividness that the units of discrete time are minutes. Suppose at t an agent is certain that it is noon: $\mathrm{Cr}_t(\textit{12:00}) = 1$. Presumably, she can also be certain that in one minute it will be one minute past noon: $\mathrm{Cr}_t(\textit{12:01}^{+1}) = 1$. If the agent follows (SC), she will then be certain at one minute past

'$(u, i, t-n)$'. This would be correct for w^{-n}, but not for w^{+n}. I got confused by my own superscript notation. I want 'A^{+n}' to say that A will be the case in n units of time. Accordingly, 'w^{+n}' says that the world will be w in n units of time. So w's "predecessor" (the world one unit before w) is w^{+1}; the successor is w^{-1}.

15 All these assumptions intuitively should affect the result of the update; it is an advantage of (SC) over (I) that they have to be made explicit.

noon that it is one minute past noon (assuming that she does not receive evidence with probability 0): $\mathrm{Cr}_{t+1}(\mathit{12{:}01}) = \mathrm{Cr}_t(\mathit{12{:}01}^{+1}) = 1$. In general, if an agent knows what time it is at any point and they update their beliefs in accordance with (SC), it seems that they will forever know the time.

Is this a problem? The classical norms of Bayesianism are norms for ideal agents with unlimited and perfectly reliable cognitive capacities. Real agents can't conditionalize complex probability measures in an instant; they can't compute the expected utility of all their options at every moment; they can't instantaneously see through all consequences of their beliefs; they can't retain everything they ever learned. All that does not undermine the relevant norms as constraints of ideal rationality. Nor does it make contemplating such norms pointless. For one thing, ignoring cognitive limitations helps to simplify formal models (like ignoring friction and air resistance in physics). In addition, the ideal case can provide useful guidance when thinking about non-ideal cases.

So one might hold that my losing track of time is just a consequence of my cognitive limitations: cognitively ideal agents don't lose track of time.

The problem with this view is that it is false. Consider a time traveller who enters a time machine that she knows will take her either 100 or 200 years into the future. Upon arrival (without further evidence), the time traveller will be lost in time, even if she has unlimited cognitive resources. Similarly, consider an astronaut travelling to another star and back at a very fast but unknown speed. Arriving back on Earth, she can't know without further evidence how much time has passed on Earth, since that depends on the unknown velocity of her spaceship (according to Special Relativity). Or consider sleep—the poor man's version of time travel into the future. If you wake up from a surprising noise in the middle of the night, even unlimited cognitive resources won't help you to figure out the time. One might respond that cognitively ideal agents don't sleep, don't travel at unknown speed, and don't enter time machines without knowing the destination, but that is getting silly.

Note that the problem is not that (SC) gives wrong results in cases where agents lose track of time. The problem is that (SC) seems to disallow such cases from arising in the first place.

Can we generalize (SC) to make room for agents who lose track of time? One way of doing that leads to the proposal of [Santorio 2011] and [Spohn 2017]. On this approach, (SC) is the right norm only if the agent is certain that one unit of time has passed since the earlier belief state. For the more general case, we assume that the agent can at least assign a probability to different hypotheses about how much time has passed. Suppose she gives 80% credence to the hypothesis that 1 unit of time has passed and 20% to the hypothesis that 2 units have passed. Instead of moving the probability mass of each centred world w to its immediate successor w^{-1}, the shifting step then divides its mass between w^{-1} and w^{-2}, with the former receiving 80% and the latter 20%, so that the shifted credence in any centred world w is not w^{+1} but $w^{+1} \times 0.8 + w^{+2} \times 0.2$.* In general, let τ be a probability distribution over natural numbers representing the agent's belief about how many units of time have passed (still assuming that time is linear and discrete). To shift the original credence function by τ, the probability of any world w is then given by

$$\text{(GI)} \qquad \mathrm{Cr}_t^{\tau}(w) = \sum_n \mathrm{Cr}_t(w^{+n})\,\tau(n).$$

* **Correction 2023:** The publishes version wrongly says that the shifting step divides the mass of w between w^{+1} and w^{+2}.

In the second step, the agent conditionalizes on her new evidence, like before:

$$(\mathrm{SC}') \quad \mathrm{Cr}_{t'}(A) = \mathrm{Cr}_t^{\tau}(A/E).$$

(I use '$\mathrm{Cr}_{t'}$' to denote the new credence function rather than 'Cr_{t+1}' because we no longer assume that the new credence is located 1 unit of time after Cr_t.)

To complete the proposal, we need to say where τ comes from. How does the agent arrive at a probability measure over how much time has passed since the earlier credence? The simplest answer, suggested by both Santorio and Spohn, is that τ comes from the agent's sensory evidence at t', delivered by a special "sense of time".[16]

Let us grant for the sake of discussion that all rational agents have a sense of time—although that strikes me as equally implausible as the claim that rationality requires having eyes. Let us also not worry how the earlier time is represented by this sense—although I do worry about that.[17] The main problem with (SC') is that it often gives verdicts that are clearly wrong.

Before I give an example, observe that (SC') takes a step back towards the imaging approach. For if there is a proposition *that 1 unit of time has passed* (which we didn't assume in section 4, but have to assume now), then Cr_t^{+1} is the image of Cr_t on that proposition (assuming the "closest world" to w at which 1 unit of time has passed is w^{-1}*): shifting turns into imaging. Indeed, (GI) is equivalent to generalized imaging as introduced in [Gärdenfors 1982] and [Lewis 1981] for cases where there may be more than one "most similar" world. In contrast to the imaging accounts from section 3, (SC') does not image the previous credence function on the total new evidence. The previous credence is imaged only on the evidence about how much time has passed. One might hope that this avoids the problems for the imaging account, because the information about how much time has passed—understood as purely temporal information that doesn't even entail that the agent still exists—may be hoped to reveal nothing about what the world was like before.[18]

But a similar problem remains. Consider the following scenario.

> You're about to be put into an artificial coma for emergency surgery. If the surgery succeeds, you will wake up after a day. If it fails, you will wake up after 10 days. You

---

∗ **Correction 2023:** The published version wrongly says 'w^{+1}'.

16 [Moss 2012] also appeals to a sense of time for a similar purpose. Like (SC'), her proposal combines a shifting-like step with subsequent conditionalization; like Santorio and Spohn, she appeals to a sense of time for the first step. But on Moss's account, the first step does not take the form (GI). Instead, the agent completely discards her earlier self-locating beliefs and uses the sense of time (in a manner not fully explained) to locate herself in absolute time.

17 Spohn suggests that the sense of time delivers qualitative information, for example, that *2 minutes have passed since the red flag was waved* and *3 minutes have passed since the green flag was waved*. But what if the earlier time t is not distinguished by special features known with certainty to the agent? Also, how does the update process determine which of these qualitatively described times is the time of the earlier credence? According to Santorio and Moss, the earlier time is identified directly, *de re*. Here, too, I wonder how the update process knows which of the times given in that way is the time of the earlier credence.

18 Arguably, the information that so-and-so much time has passed still entails that the world hasn't come to an end in the meantime, and thus reveals that the previous state of the world was not a terminal state. Terminal world states also pose problems for (SC): what is w^{+1} if w is terminal? I will ignore these problems and assume that the agents in question give negligible credence to the hypothesis that the world (or their life) is just about to end; see [Meacham 2010: sec.4.1] and [Schwarz 2015: 667:fn.8].

know all this, and you rationally give credence 0.9 to the hypothesis that the surgery will fail. Your inner sense of time is not attuned to comas, so upon awakening it suggests to you that not much more than a day or two have passed. You have no further evidence at this point about whether the surgery succeeded.

Let's assume for concreteness that your sense of time assigns probability $1/2^n$ to the hypothesis that n days have passed; so $\tau(1) = 1/2$ and $\tau(10) = 1/1024$. Let S be the hypothesis that the surgery either will succeed or has succeeded, and let $D1$, $D2$, $D3$, …, be the hypotheses that you were in the coma for 1, 2, 3, …, days, respectively. At t, before you were put into the coma, your credence in S was 0.1. Shifting by (GI) leads to the following probabilities:

$$\begin{aligned}
&\mathrm{Cr}_t^{\tau}(S \wedge D1) = 0.1 \times 1/2 = 0.05\\
&\mathrm{Cr}_t^{\tau}(\neg S \wedge D1) = 0.9 \times 1/2 = 0.45\\
&\ldots\\
&\mathrm{Cr}_t^{\tau}(S \wedge D10) = 0.1 \times 1/1024 \approx 0.0001\\
&\mathrm{Cr}_t^{\tau}(\neg S \wedge D10) = 0.9 \times 1/1024 \approx 0.0009.\\
&\ldots
\end{aligned}$$

Notice that the shifted credence function assigns significant probability to possibilities like $\neg S \wedge D1$ in which you are not awake. (If we replace the time in the coma with genuine time travel, you wouldn't even exist at the relevant centred worlds.) These possibilities must be ruled out in the next step of the update, where you conditionalize. Your post-awakening evidence plausibly reveals to you that you are awake and that you have just woken up from the coma. Thus it concentrates all probability on $S \wedge D1$ and $\neg S \wedge D10$. As a result, you will become highly confident that the surgery was a success:

$$\begin{aligned}
&\mathrm{Cr}_{t'}(S \wedge D1) \approx 0.9827\\
&\mathrm{Cr}_{t'}(\neg S \wedge D10) \approx 0.0173.
\end{aligned}$$

This is the wrong result. Knowing that your sense of time is not to be trusted, your credence in the surgery having been a success should remain close to the previous level, at around 0.1.

The coma example is not an isolated special case. Our inner sense of time is not perfectly calibrated to the actual passage of time, and we know that its reliability, and the direction in which it errs, depends on the circumstances: during some activities, time seems to fly by, during others it almost comes to a standstill. We cannot blindly trust our sense of time.

What if we don't identify τ in (SC′) with the output of an agent's sense of time, but with the agent's all-things-considered credence about how much time has passed, based on her sense of time together with any other relevant information she may have? (SC′) then collapses into contradiction. To

19 The only way to get $\mathrm{Cr}_{t'}(D10) = 0.9$ and $\mathrm{Cr}_{t'}(D1) = 0.1$ out of (SC′) is to assume that τ assigns equal probability to $D10$ and $D1$. In general, (SC′) often gives plausible verdicts if τ is uniform over all possible time shifts. The shifting step then makes the agent completely lost in time; to find her new location in time, the agent must draw on her (remaining) evidence. With uniform τ, (SC′) effectively belongs to the family of proposals discussed in section 2 on which diachronic norms only pertain to uncentred beliefs.

illustrate, return to the coma case and suppose that at t' your all-things-considered credence in $D1$ and $D10$ is 0.1 and 0.9 respectively (as seems reasonable). (GI) then moves probability $0.9 \times 0.9 = 0.81$ to $\neg S \wedge D10$ worlds, and $0.1 \times 0.1 = 0.01$ to $S \wedge D1$ worlds. Conditionalizing on the information that you've woken up excludes all other possibilities, so $\mathrm{Cr}_{t'}(D10) \approx 0.988$ and $\mathrm{Cr}_{t'}(D1) \approx 0.012$—contradicting the assumption that your t' credence in $D1$ and $D10$ is 0.1 and 0.9 respectively.[19]

[Schulz 2010] suggests a different generalization of (SC) that gets around the problem. Let τ represent the agent's all-things-considered credence (at t') about how much time has passed. Then Schulz suggests to replace (SC) by (SC*):

$$(\mathrm{SC}^*) \quad \mathrm{Cr}_{t'}(A) = \sum_n \tau(n)\, \mathrm{Cr}_t(A^{+n}/E^{+n}).$$

In the coma example, where E is your post-awakening evidence, $\mathrm{Cr}_t(S^{+1}/E^{+1})$ is plausibly 1 and $\mathrm{Cr}_t(S^{+10}/E^{+10})$ is 0, for you are sure that the surgery succeeds iff you wake up after one day. With $\tau(1) = 0.1$ and $\tau(10) = 0.9$, we therefore get the desired result:

$$\mathrm{Cr}_{t'}(S) = 0.1 \times 1 + 0.9 \times 0 = 0.1.$$

In contrast to (SC) and (SC′), (SC*) can no longer be divided into a shifting step and a conditionalization step.[20]

Schulz's proposal escapes the coma problem, but it gives wrong results in other cases, such as the following variant of the Sleeping Beauty problem, not involving any threat of memory erasure.[21]

> After you go to sleep on Sunday, a fair coin will be tossed. If it lands heads, you will be made to sleep through until Tuesday morning, when you will be awakened by the sound of a bell. If the coin lands tails, you will be awakened on Monday by the cry of a rooster and on Tuesday by the sound of a bell. (Nobody will tamper with your memories.) You are aware of all these facts when you fall asleep on Sunday. After a deep and dreamless sleep, you find yourself waking up to the sound of a bell.

It should be uncontroversial that you ought to become confident that it is Tuesday and that the coin landed heads. After all, if the coin had landed tails, your next awakening would have been to the cry of a rooster (and you would have retained your memories of that awakening). Using days as temporal units, this means that $\tau(1) = 0$ and $\tau(2) = 1$; by (SC*), it follows that $\mathrm{Cr}_{Tue}(\mathit{Heads}) = 1 \times \mathrm{Cr}_{Sun}(\mathit{Heads}^{+2}/E^{+2})$. But you already knew on Sunday that you would be awakened by a bell in two days time. So if your relevant sensory evidence E is the sound of the bell (or your awakening

20 [Kim 2009] proposes a similar generalization to (SC) but with absolute time shifts: $\mathrm{Cr}_{t'}(A) = \sum_i \mathrm{Cr}_{t'}(\mathit{now} = i)\, \mathrm{Cr}_t(A \textit{ at } i / E \textit{ at } i)$, where i ranges over points of time. (Notice that this only makes use of the agent's uncentred opinions at t.) The following counterexample to (SC*) is also a counterexample to Kim's proposal, modulo the remark in footnote 22 below.

21 The original Sleeping Beauty problem, introduced in [Piccione and Rubinstein 1997] and [Elga 2000], is outside the scope of the present survey, as it is not a case in which an agent has unlimited and perfectly reliable cognitive capacities: if Beauty's coin lands tails, she cannot obey any substantive diachronic norms in the transition from Monday to Tuesday. The answer to the Sleeping Beauty problem therefore depends not only on the norms for ideal updates—the topic of the present paper—but also on the norms for how to compensate for a threat to one's ideal (diachronic) rationality

by that sound), then $\mathrm{Cr}_{Sun}(Heads^{+2}/E^{+2}) = 1/2$; so (SC*) falsely entails that upon awakening, you should give equal credence to heads and tails.[22]

Another obvious drawback of (SC*) is that it takes the agent's all-things-considered credence about how much time has passed as given. How is the agent supposed to arrive at these beliefs? Obviously not by (SC*). But beliefs about how much time have passed are highly constrained by the agent's previous belief state and her sensory evidence. As it stands, (SC*) is therefore at best part of the full story about ideal diachronic rationality—even setting aside that it gets some cases wrong and relies on a questionable ability to pick out the previous time.[23]

Fortunately, we can say more. To do so, let's pause to reconsider exactly what update rules like (C) or (SC) or (SC*) are supposed to tell us.

6 Shifting in personal time

Conditionalization, understood as an update rule, relates two epistemic states: an "old" state and a "new" state. Schematically, it says that the credences of an agent A_2 at time t_2 should equal the credences of an agent A_1 at t_1 conditional on A_2's evidence at t_2. But not any pair of agents and times falls in the rule's domain of application. For example, conditionalization plausibly doesn't say that *your* credence *now* should equal *my* credence *in 5 hours* conditional on your present evidence. The rule only applies if A_2 at t_2 stands in a special relation to A_1 at t_1. What is that relation?

As a necessary condition, one might suggest that A_1 must be identical to A_2: update rules like (C) are meant to tell us how an agent's later beliefs should be related to *the same agent's* earlier beliefs. But even that may be questioned. For example, a common view in the metaphysics of personal identity is that persons cannot survive episodes of fission; yet there are plausibly norms on how beliefs should change across fission (see e.g. [Meacham 2010: 93f.], [Hedden 2015: 455–8], [Schwarz 2015]).

What about the relevant times, t_1 and t_2? If we work with a sensory conception of evidence, then t_1 and t_2 can't be arbitrary times. For remember that sensory evidence is not cumulative: an agent's sensory evidence at one time is generally not part of her sensory evidence at later times. Consequently, what an agent should believe at some point t_2 is not a function of her credence at an arbitrary earlier time t_1 and her (sensory) evidence at t_2, because the new credence should also reflect the evidence received in between t_1 and t_2. That's why conditionalization is often described as relating an agent's belief state right before the information E is received to the agent's belief state right after E has been taken into account. But what exactly does that mean? How, for example, does it apply to a time traveller for whom the old state may well be in the future of the new state?

I will not explore these questions any further. I bring them up only to emphasize that stepwise diachronic norms like conditionalization are relative to an *epistemic successor* relation between epistemic states, or agents-at-times: we should not read 'Cr_t' and 'Cr_{t+1}' in (C) as invariably picking out the credence function of the same agent at successive points in objective time, given by atomic

(compare [Arntzenius 2002]).

22 To be fair, Schulz does not say whether E is to be understood as sensory evidence or as a more comprehensive kind of evidence that includes the information that you were not awakened by a rooster the day before (which, as I hereby stipulate, is not part of your sensory evidence).

23 [Titelbaum 2016] points out an analogous gap in [Kim 2009], but falsely suggests that it affects all shifting accounts.

clocks. Rather, 'Cr_t' and 'Cr_{t+1}' pick out credence functions of epistemically successive epistemic states, whatever exactly that means. In effect, the epistemic successor relation defines a kind of *personal time* (compare [Lewis 1976a]) relative to which the indices in 'Cr_t' and 'Cr_{t+1}' should be understood.

Now return to the simple shifting rule from section 4.

$$\text{(SC)} \quad \mathrm{Cr}_{t+1}(A) = \mathrm{Cr}_t^{+1}(A/E) = \mathrm{Cr}_t(A^{+1}/E^{+1}).$$

When I suggested that the rule fails to allow for agents who lose track of time, I assumed that shifting goes by objective time: that w^{-1} is w with the centre moved one unit of objective time into the future. But that makes little sense if the new belief state is not always located one unit of time in the future. (If a time traveller instantaneously travels from 4pm to 6pm, knowing at departure that it is 4pm and that she will arrive at 6pm, then her next belief update should not make her certain that it is 4.01pm.) A more sensible interpretation of (SC) re-uses the epistemic successor relation to define shifting: w^{-1} is w with the centre moved to the epistemic successor of the epistemic state on which w is centred.* This is how (SC) is interpreted in [Meacham 2010] and [Schwarz 2015]. On that interpretation, (SC) gives the right verdicts even in time travel cases and situations where agents lose track of time; there is no need to explore more general rules like (SC′) or (SC*).

The details now depend on the epistemic successor relation. But we don't need to settle every conceivable trouble case before we can start applying the rule. Assume as before that we're dealing with agents who update their belief state instantaneously and in discrete steps, and let's ignore the possibility of fission and fusion. Each update, let's assume, is a causal process that produces a new belief state based on the previous state and the current (sensory) evidence. What we want to know from an update rule is how the newly produced credences should relate to the pre-update credences and the new evidence. So we can identify the epistemic successor relation with the "update relation" that holds between states $S1$ and $S2$ just in case $S2$ is the result of an update applied to $S1$ (as suggested in [Schwarz 2012]).

Let's see how (SC), on the new interpretation, handles the coma example. The shifting step plausibly moves the probability of every pre-coma world to the next post-coma world, as that is when the next belief update takes place. For worlds in which the surgery is a success, the shift covers 1 day; for worlds in which the surgery fails, it covers 10 days. (By contrast, (SC′) indiscriminately shifted all worlds by the same amount.) The shifted credence thus still assigns probability 0.1 to the surgery having succeeded, and since the new evidence is neutral on the success of the surgery, this is the final credence.

We also get the right result in the Sleeping Beauty variant. Here the shifted Sunday credence is evenly divided between *Heads & Tuesday & Bell* worlds and *Tails & Monday & Rooster* worlds; subsequent conditionalizing on *Bell* makes you certain that it is Tuesday and that the coin landed heads.

So (SC) looks like a good replacement for conditionalization if centred propositions are allowed. It is well-motivated, correctly deals with all the examples we have considered, and displays many of the same abstract features that characterize conditionalization for uncentred credences. (SC) does

* **Correction (2023)**: The published version wrongly uses 'w^{+1}' instead of 'w^{-1}' for the successor of w.

not deny that we have a sense of time, but it also doesn't make that a requirement of rationality. If an agent has such a sense, it enters the update process like any other sense, in the conditionalization step.

The main downside of (SC), in comparison to the other rules we've considered, is that it puts more weight on the epistemic successor relation. Real agents, you may worry, do not update their beliefs discretely and instantaneously. How should we understand the successor relation, and consequently the shifting operation, for such agents?

Now, as mentioned above, there are continuous generalizations of (SC). But I doubt that this fully resolves the worry. In either form, (SC) is at best an *idealized model* of belief update, a model that does not simply and directly map onto real agents. Arguably the same is true of Bayesianism in general. We talk about agents' credences, but what does it take for a concrete lump of flesh or silicon to have a given credence function? If we make (SC) (or indeed (C)) part of the Bayesian model, then mapping the model to the world requires invoking some criterion for dividing an agent's epistemic history into discrete stages. In practise, this rarely poses serious problems. Remember that the reason why update rules need to relate immediately successive belief states is that we don't want to miss relevant evidence acquired at intermediate times. If we're only interested in modelling an agent's credence about a given subject matter, we can therefore ignore times at which the agent receives no evidence relevant to that subject matter.[24] So quite coarse-grained successor relations are often sufficient.

To illustrate, recall the application of (SC) to the original fruit basket scenario at the end of section 4. To model the update, we needed no more than twelve centred worlds and one step of shifting, from before you learned that any bananas had been removed to afterwards. It does not matter whether the learning event was really instantaneous; we can even take the "old" state to be an hour before the learning event and the "new" state to be an hour afterwards, provided you acquire no other relevant news about the fruit basket in these intervals.

You may also worry that (SC) still requires agents to keep perfect track of time, albeit only of personal time. This is true, in the following sense. Suppose at some point an agent is certain that A is true now and at no other point in time. If she updates in accordance with (SC), her successor stage will then be certain of A^{-1}; after n updates, she will be certain of A^{-n}. Informally speaking, she will know that A was true n "units of personal time" in the past. Couldn't the agent become unsure how much personal time has passed since A?

She certainly could. The question is whether this would involve a failure of ideal diachronic rationality, and it is at least arguable that it would. If before falling asleep you know that it is raining, and upon awakening you're unsure whether it was raining before, then you've lost information. Diachronically ideal agents don't lose information. To be sure, there are situations where agents can't help but lose information. (SC) is not applicable to such cases. But the same is true for (C). If there is a problem here, it is arguably independent of the problem posed by centred propositions.

Note also that personal time is not a mysterious objective quantity knowledge of which would require special sensory capacities. There is a tendency to think of belief update as a kind of reasoning

24 Some care is required because evidence that may appear irrelevant at the time (when it is received) can retrospectively become relevant in the light of later evidence.

the agent is supposed to go through at $t+1$: "Here's my earlier credence function Cr_t; here's my new evidence E; now how do I combine these to compute my updated credence?" From that perspective, the earlier credence would have to be picked out in some way or other, and the agent would have to know that it is indeed the credence of her predecessor state. But update rules like (C) or (SC) are not norms for how agents should reason at $t+1$. They rather specify how an agent's beliefs should evolve from one time to another. It does not matter if the agent recalls her earlier credences, and if so under what mode of presentation.

7 Cumulative evidence

In section 2, I mentioned that some authors have resisted introducing centred objects of credence. A common motivation for the resistance is the alleged fact that doing so would make the dynamics of rational credence intractable. Stalnaker, for example, asserts that "centred-worlds models [...] provided no resources for representing the relations between informational states across time and across persons, and so no resources for clarifying the dynamics of knowledge and belief" ([2008: 64]). We have seen that such skepticism is unfounded. With credences extended to centred propositions, the shifting rule (SC) can play essentially the same role that conditionalization played in classical Bayesianism. Ironically, the centred-worlds model allows us to present a *clearer* (or at least, more complete) picture of diachronic rationality than Stalnaker's own model.

To conclude, I want to return to another idea from section 2: that we might replace diachronic norms by (partly or entirely) synchronic norms. On the simplest proposal along these lines, an agent's credence function at any time should equal some ur-prior P conditional on the agent's evidence at that time. I have argued that these views require an unexplained non-sensory conception of evidence and thus don't provide a full answer if we want to know how an agent's credences should evolve as time goes by and new information arrives from her senses. I also suggested that the problem of updating centred beliefs reappears as a problem about updating non-sensory evidence: how does your sensory evidence that it *is* raining transform into your later non-sensory evidence that is *was* raining?

25 Aggregating inconclusive evidence raises further issues. In particular, how should we even represent an agent's history of inconclusive evidence? A convenient choice is to list the relevant evidence partitions associated not with their Jeffrey weights but with their Bayes factors (compare [Field 1978]), for the effect of stepwise Jeffrey conditioning on some partitions with associated Bayes factors can be mimicked by Jeffrey conditioning on the intersection of these partitions, with Bayes factors determined by the original Bayes factors. (With Jeffrey weights this is not possible because the final probability—the Jeffrey weight—of cells in the refined partition generally depends not just on the Jeffrey weights in the individual updates but also on the agent's prior credence.) To allow for centred propositions, the Bayes factors should be replaced by *shifted Bayes factors*

$$\frac{\mathrm{Cr}'(A)\,/\,\mathrm{Cr}'(B)}{\mathrm{Cr}(A^{+1})\,/\,\mathrm{Cr}(B^{+1})}.$$

One might object that evidence partitions with associated (shifted) Bayes factors are not an adequate representation of the agent's evidence on the grounds that the Bayes factors at a certain point in an evidence history depend not only on the given sensory experience, but also on the agent's priors and on the experiences she had earlier (see e.g. [Garber 1980], [Weisberg 2009]). I see no way to avoid the dependence on earlier experiences, but at least the dependence on prior credence can be mitigated by fixing the prior (i.e., the initial probability reflected in the Bayes factors of an evidence history) as the ur-prior.

Drawing on the lessons from sections 4–6, we can now solve the problem of updating non-sensory evidence.

For simplicity, I will focus on cases where evidence is conclusive.[25] So assume an agent receives sensory evidence $E_1, E_2, \ldots, E_{n-1}, E_n$ at successive points in her personal time. Arguably, if your sensory evidence is E and you optimally preserve that information, then the successor of your present epistemic state should be certain of E^{-1}: that E was the case "one unit of personal time in the past".[26] So we may define the agent's *cumulative evidence* at the point when she receives sensory evidence E_n as $E_1^{-(n-1)} \wedge E_2^{-(n-2)} \wedge \ldots \wedge E_{n-1}^{-1} \wedge E_n$.

How does updating by (SC) compare to ur-prior conditionalization on cumulative evidence, as just defined? In classical Bayesianism, where centred propositions are ignored, the result of successively updating on $E_1, E_2, \ldots, E_n$ is identical to conditionalizing the initial credence function on the conjunction of $E_1, E_2, \ldots, E_n$. One might similarly conjecture that successive application of (SC) to an agent's sensory evidence leads to the same result as conditionalizing the initial credence function on the agent's final cumulative evidence.

As I show in the appendix, the conjecture holds only under non-trivial assumptions. In particular, we need something like the following *stationarity* assumption (a special case of the principle of self-locating indifference defended in [Bostrom 2002], [Elga 2004], and [Arntzenius and Dorr 2017]):

> Whenever a centred world w' is a successor of a centred world w, then any rational initial credence function Cr_0 assigns equal probability to w and w'.

Given stationarity (and linearity—no fission or fusion—which I've assumed throughout), if an agent's credence function at any time equals some ur-prior P conditional on her cumulative evidence at the time, then the agent updates her beliefs by (SC). Conversely, if an agent starts out with P and from then on updates by (SC), then her credence at any time equals P conditional on her cumulative evidence. The two accounts describe the same dynamics.

There might still be reasons to prefer one over the other. One might prefer the ur-prior formulation on the grounds that it more easily generalizes to cases where agents lose information. My own view is that it does not, and that ur-prior conditionalization gives wrong verdicts in unusual cases when stationarity or linearity fail. But that is a story for another occasion.

Appendix

I will first show that if stationarity holds (as well as a minor further assumption that I will introduce in a moment), then successive updating on sensory evidence in accordance with (SC) yields the same result as conditionalizing the agent's initial credence Cr_0 on her cumulative evidence.

Let $E_1, \ldots, E_n$ be the sensory evidence the agent receives at personal times $1, \ldots, n$, respectively. With a little algebra, it is easy to show that if the agent follows (SC), then for any proposition A,

$$(1) \qquad \mathrm{Cr}_n(A) = \mathrm{Cr}_0(A^{+n} / E_1^{+1} \wedge \ldots \wedge E_n^{+n}).$$

26 As before, the interpretation of the epistemic successor relation and thus of personal time may depend on the modelling purpose; you don't need to conceptualize E^{-1} in terms of personal time.

Since the agent's cumulative evidence at point n is $E_1^{-(n-1)} \wedge \ldots \wedge E_{n-1}^{-1} \wedge E_n$, what we have to show is (2).

$$(2) \qquad \mathrm{Cr}_0(A^{+n}/E_1^{+1} \wedge \ldots \wedge E_n^{+n}) = \mathrm{Cr}_0(A/E_1^{-(n-1)} \wedge \ldots \wedge E_{n-1}^{-1} \wedge E_n).$$

The further assumption we need besides stationarity is that E_1 is incompatible with the present state being an agent's initial state—that is, her state *before* receiving any evidence. This can be captured by (3), where $\top$ is the tautology:

$$(3) \qquad Cr_0(E_1 \supset \top^{-1}) = 1.$$

To see why (3) is needed, suppose Cr_0 assigns

- probability .4 to an initial $A \wedge E_1$ world,
- probability .4 to its successor, a terminal $\neg A \wedge \neg E_1$ world,
- probability .1 to an initial $A \wedge \neg E_1$ world, and
- probability .1 to its successor, a terminal $\neg A \wedge E_1$ world.

Then $\mathrm{Cr}_0(A^{+1}/E_1^{+1}) = 0$ but $\mathrm{Cr}_0(A/E) = .8$, which violates (2). (3) could perhaps be motivated by the fact that an agent's total new evidence generally includes higher-order evidence about the agent's beliefs. Alternatively, instead of (3) we could assume that even an agent's initial credence Cr_0 is given by conditionalizing a merely hypothetical ur-prior P on some initial evidence E_0, which would play the role of $\top$ in what follows.

To prove (2), observe that for any propositions $A_0, A_1, \ldots, A_n$, any centred world that satisfies $A_0 \wedge A_1^{+1} \wedge \ldots \wedge A_n^{+n}$ is succeeded by a world that is succeeded by a world …(n times) …that satisfies $A_0^{-n} \wedge A_1^{-(n-1)} \wedge \ldots \wedge A_n$; conversely, any world that satisfies $A_0^{-n} \wedge A_1^{-(n-1)} \wedge \ldots \wedge A_n$ is n-preceded by a world that satisfies $A_0^{-n} \wedge A_1^{-(n-1)} \wedge \ldots \wedge A_n$; by linearity (no fission and fusion), this mapping from worlds to worlds is one-one; by stationarity, the worlds it pairs always have equal probability. Thus for any propositions $A_0, A_1, \ldots, A_n$,

$$(4) \qquad Cr_0(A_0 \wedge A_1^{+1} \wedge \ldots \wedge A_n^{+n}) = Cr_0(A_0^{-n} \wedge A_1^{-(n-1)} \wedge \ldots \wedge A_n).$$

Two instances of this equality are (5) and (6).

$$(5) \qquad \mathrm{Cr}_0(\top \wedge E_1^{+1} \wedge \ldots \wedge E_n^{+n}) = \mathrm{Cr}_0(\top^{-n} \wedge E_1^{-(n-1)} \wedge \ldots \wedge E_n).$$

$$(6) \qquad \mathrm{Cr}_0(\top \wedge E_1^{+1} \wedge \ldots \wedge E_n^{+n} \wedge A^{+n}) = \mathrm{Cr}_0(\top^{-n} \wedge E_1^{-(n-1)} \wedge \ldots \wedge E_n \wedge A).$$

By (3), $Cr_0(E_1 \leftrightarrow (\top^{-1} \wedge E_1)) = 1$, so we can remove $\top$ from the conjunctions in (5) and (6). By the ratio formula for conditional probability, (5) and (6) then entail (2).

Technically, (2) can be true even in the absence of stationarity. For example, different ur-priors of successive worlds could happen to balance out so as to preserve (4). So there are slightly weaker assumptions that would also do the job. But I can't think of any motivation for these assumptions that would not equally support stationarity.

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

Frank Arntzenius and Cian Dorr [2017]: "Self-Locating Priors and Cosmological Measures". In Khalil Chamcham, John Barrow, Simon Saunders and Joe Silk (Eds.) *The Philosophy of Cosmology,*

Nick Bostrom [2002]: *Anthropic bias: Observation selection effects in science and philosophy*. New York: Routledge

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

Darren Bradley [2013]: "Dynamic Beliefs and the Passage of Time". In A. Capone and N. Feit (Eds.) *Attitudes De Se,* Chicago: University of Chicago Press

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

Herman Cappelen and Josh Dever [2013]: *The Inessential Indexical*. Oxford: Oxford University Press

David Christensen [1992]: "Confirmational Holism and Bayesian Epistemology". *Philosophy of Science*, 59(4): 540–557

— [2000]: "Diachronic coherence versus epistemic impartiality". *The Philosophical Review*, 109(3): 349–371

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

— [2004]: "Defeating Dr. Evil with Self-Locating Belief". *Philosophy and Phenomenological Research*, 69: 383–396

Hartry Field [1978]: "A Note on Jeffrey Conditionalization". *Philosophy of Science*, 45(3): 361–367

Daniel Garber [1980]: "Field and Jeffrey Conditionalization". *Philosophy of Science*, 47(1): 142–145

Peter Gärdenfors [1982]: "Imaging and Conditionalization". *Journal of Philosophy*, 79: 747–760

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

Brian Hedden [2015]: "Time-Slice Rationality". *Mind*, 124(494): 449–491

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

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

— [1988]: "Conditioning, kinematics, and exchangeability". In B. Skyrms and W.L. Harper (Eds.) *Causation, chance and credence,* Dordrecht: Kluwer, 221–255

Hirofumu Katsuno and Alberto O. Mendelzon [1991]: "On the difference between updating a knowledge database and revising it". *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR-92)*: 387–394

Gabriele Kern-Isberner [2001]: "Revising and Updating Probabilistic Beliefs". In *Frontiers in Belief Revision,* Springer, 393–408

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

Jérôme Lang [2007]: "Belief Update Revisited." In *IJCAI,* vol 7. 6–12

Hannes Leitgeb [2016]: "Imaging all the people". *Episteme*: 1–17

Hannes Leitgeb and Richard Pettigrew [2010]: "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy". *Philosophy of Science*, 77: 236–272

Isaac Levi [1980]: *The Enterprise of Knowledge*. Cambridge, MA: MIT Press

David Lewis [1976a]: "The Paradoxes of Time Travel". *American Philosophical Quarterly*, 13: 145–152

— [1976b]: "Probabilities of Conditionals and Conditional Probabilities". *The Philosophical Review*, 85: 297–315

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

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

— [1996]: "Elusive Knowledge". *Australasian Journal of Philosophy*, 74: 549–567

— [1999a]: *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press

— [1999b]: "Why Conditionalize?" In [Lewis 1999a], 403–407

Ofra Magidor [2015]: "The myth of the de se". *Philosophical Perspectives*, 29: 259–283

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

— [2016]: "Ur-Priors, Conditionalization, and Ur-Prior Conditionalization". *Ergo*, 3: 444–402

Sarah Moss [2012]: "Updating as communication". *Philosophy and Phenomenological Research*, 85(2): 225–248

— [2014]: "Time-slice epistemology and action under indeterminacy". *Oxford Studies in Epistemology*, 5: 172–94

John Perry [1977]: "Frege on Demonstratives". *Philosophical Review*, 86: 474–497

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

Alvin Plantinga [1974]: *The Nature of Necessity*. Oxford: Oxford University Press

John L. Pollock [1984]: *The foundations of philosophical semantics*. Princeton: Princeton University Press

Stuart J. Russell and Peter Norvig [2010]: *Artificial Intelligence: A Modern Approach*. Cambridge (MA): MIT Press, 3rd edition

Nathan Salmon [1986]: *Frege's Puzzle*. Cambridge (Mass.): MIT Press

Paolo Santorio [2011]: "Cognitive Relocation". Unpublished Manuscript

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

— [2015]: "Belief update across fission". *British Journal for the Philosophy of Science*, 66: 659–682

Charles T. Sebens and Sean M. Carroll [2017]: "Self-locating uncertainty and the origin of probability in Everettian quantum mechanics". *The British Journal for the Philosophy of Science*

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

Wolfgang Spohn [2017]: "The Epistemology and Auto-Epistemology of Temporal Self-Location and Forgetfulness". *Ergo*, 4

Robert Stalnaker [1981]: "Indexical Belief". *Synthese*, 49: 129–151

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

— [2014]: *Context*. Oxford: Oxford University Press

— [2016]: "Modeling a Perspective on the World". In Manuel García-Carpintero and Stephan Torre (Eds.) *About Oneself: De Se Thought and Communication,* Oxford: Oxford University Press, 121–139

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

— [2016]: "Self-Locating Credences". In A. Hajek and C. Hitchcock (Eds.) *The Oxford Handbook of Probability and Philosophy,* Oxford: Oxford University Press, 666–680

Bernard Walliser and Denis Zwirn [2002]: "Can Bayes' Rule be Justified by Cognitive Rationality Principles?" *Theory and Decision*, 53(2): 95–135

Clas Weber [2013]: "Centered communication". *Philosophical Studies*, 166(1): 205–223

Jonathan Weisberg [2009]: "Commutativity or holism? A dilemma for conditionalizers". *The British Journal for the Philosophy of Science*, 60(4): 793–812

Peter M. Williams [1980]: "Bayesian conditionalisation and the principle of minimum information". *British Journal for the Philosophy of Science*, 31(2): 131–144

Timothy Williamson [2000]: *Knowledge and its Limits*. Oxford: Oxford University Press