Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

The Epistemology and Auto-Epistemology of Temporal Self-Location and Forgetfulness

WOLFGANG SPOHN
Department of Philosophy
University of Konstanz

This paper deals with the epistemology and auto-epistemology of temporal self-location and forgetfulness in probabilistic terms. After explicitly stating the underlying algebraic or propositional framework, it proposes two rules of probability change through our inner sense of time and generally describes how conditionalization works with respect to indexical information. It suggests a rule for rearranging beliefs after forgetting (and other unfavorable epistemic changes). After rehearsing standard auto-epistemology in terms of the reflection principle and its consequences, it moreover studies the auto-epistemology of those non-standard epistemological changes. Thus, it generalizes the reflection principle to the indexical case and to an even more general version that is free from the informal restrictions that are commonly assumed. All these principles are illustrated with various examples: the prisoner, the new riddle of induction, Sleeping Beauty, and finally Shangri-La.

1. Introduction

Sleeping Beauty is a lovely problem, first mentioned in Piccione and Rubinstein (1997) and introduced into the philosophical discussion by Elga (2000).[1] It is lovely because it concocts various prima facie unrelated epistemological issues: how to epistemically deal with temporal self-location, with forgetfulness,

1. The name "Sleeping Beauty" goes back to a fairy tale by Charles Perrault in 1697. I would much prefer the German name "Dornröschen" ("little briar rose"), which the Grimm Brothers used in their German adaptation. That name is more charming, and it would suggest that the problem concerns thorny matters. Alas, it would be futile to try to change the label.

Contact: Wolfgang Spohn < wolfgang.spohn@uni-konstanz.de>

http://dx.doi.org/10.3998/ergo.12405314.0004.013

and with auto-epistemology. It does so in quite an imperspicuous way und thus provokes a varied and fruitful bunch of possible responses and arguments. Not many philosophical problems have this fertility.

However, this paper is not another paper about Sleeping Beauty. It is a paper about these epistemological issues, which seem to me to be treated in the literature either insufficiently or not at all. It is also about their interaction and thus takes up all three of them. This will be a comprehensive enterprise. I hope that in the end it will have been worth the efforts. For illustration, the paper will return to various familiar examples, among them Sleeping Beauty. However, the effect will at best be a still clearer theoretical view on those examples and not any novel treatment. The interest lies in the epistemological issues behind the examples.

Still, in order to better understand the motivation in the first paragraph, a quick retake of Sleeping Beauty may be in order: On Sunday night Sleeping Beauty is put to sleep for three days till Wednesday, but is occasionally woken up for a few minutes. This is determined by a throw of a fair coin after she starts sleeping. If the coin lands heads, she will be woken only on Monday. If the coin lands tails, she will be woken both on Monday and on Tuesday. In the latter case, however, she will receive another drug after the first awakening that completely erases all memories of it. Thus, at no time does she have any clue or idea whether she has been woken on Monday or Tuesday. On Sunday she is fully informed about the entire set-up, believes in it with certainty, and continues to do so during those days. After being woken (for the first or the second time—this does not seem to make any difference) she is asked how likely she thinks it is that the coin fell heads. How should she respond?

The problem is that there seem to be at least two good answers. It is not the place to rehearse all the sophisticated arguments in their favor and to discuss their possible confusions. It is, however, important to sense the strong intuitions behind the answers, even if they should be flawed:

The *thirders* see that there are three possible awakenings, one on Monday with heads, one on Monday with tails, and one on Tuesday with tails, and then they argue, as seems most plausible, that the first two awakenings are equally likely and that the last two awakenings are also equally likely. Hence, all three are equally likely, and the answer should be: one third.

The *halfers* point out that by being woken up Sleeping Beauty does not learn anything whatsoever; she knew since Sunday that she will be woken at least once. However, learning nothing, or learning something that was expected with certainty, anyway, should not change the subjective probabilities. Since they clearly are fifty-fifty on Sunday, the answer should be the same later on.

The problem is not simple because, as I said, it is an ingenious concoction of at least three apparently unrelated epistemological issues, which need to be disentangled in order to understand their entanglement. Those issues have not

been fully in the focus of epistemology, although they need to be addressed. It is obvious what the issues are:

First, Sleeping Beauty clearly is a problem about *indexical belief* and its dynamics, about how to integrate uncertainty about one's temporal self-location into one's epistemic picture. Second, the assumption about having forgotten the first awakening at the potential second awakening is an essential ingredient of the story. Thus, Sleeping Beauty also seems to be a problem about the epistemic management of (presumed) *forgetting*. Third, Sleeping Beauty is obviously a problem for *auto-epistemology*. The story is not only about the evolution of her beliefs. Her foresight of this evolution, the knowledge that it might involve forgetting, is crucial; thus she needs to integrate her present and her possible future beliefs. This is what auto-epistemology is about.[2] Finally, it goes without saying that these are issues for *normative* epistemology. We are not interested in the psychology of Sleeping Beauty; we want to know how rational subjects ought to deal with these issues.

All three issues are urgent, but they are non-standard. Standardly, epistemology is about propositional, non-indexical belief, evidence, and learning. However, indexical belief is non-propositional (in a narrow sense of "propositional"), and forgetting is the opposite of learning. Hence, both topics have, undeservedly, been marginalized in standard epistemology. This is changing, fortunately. Finally, auto-epistemology has been put on the agenda by van Fraassen's (1984) reflection principle, but it has remained a philosopher's specialty. It has hardly been extended to indexical belief and not at all to forgetting. Initially, the three issues seem independent, and it is advisable to start treating them separately. Then, however, it will be most interesting to see whether and how they combine. I attempt to provide both the separate treatment as well as the integration. Thus the plan of the paper is this:

In Section 2 we have to prepare the algebraic preliminaries. To do so explicitly is an indispensible basis of the entire business. On this basis, Section 3 will deal with the issue of temporal self-location in a systematic way. This will be illustrated with two applications. In Section 4 I will discuss 'The Prisoner', a nice example invented by Arntzenius (2003); and in Section 5 I will consider an indexical version of enumerative induction, which will display a new facet of Goodman's (1946) 'New Riddle of Induction'. Section 6, then, will briefly deal with the rational management of forgetting. The only surprise, perhaps, is that something substantial and reasonable can be said about it at all. On the basis of Sections 3 and 6 we will be in a position to return to Sleeping Beauty in Sec-

2. Moreover, it has been suggested that Sleeping Beauty is a problem of *biased evidence*, which is a familiar and huge problem in statistics. At least Bradley (2012: 168–170) conceives of the experience of getting awakened as biased evidence concerning self-location, and therefore, he argues, it does not count (this would favor the halfers' solution). I will not expand on this issue.

tion 7, where I shall unambiguously defend the thirders' position and explain the mistakes in the halfers' arguments. The point will not be to present new arguments, but rather to make the principles used in the argument as explicit as possible.

So much for the epistemological extensions. The second part of the paper will turn to the auto-epistemological extensions. Section 8 will rehearse auto-epistemology along the lines of Hild (1998a), including his rule of auto-epistemic conditionalization, the fundamentality of which is little acknowledged. Section 9 will extend auto-epistemology to self-locating and indexical beliefs. This does not hide any surprises. Section 10 will explore the auto-epistemology of (anticipated) forgetting and other arational doxastic changes. Thereby we will enter entirely uncharted waters, but if I am right, substantial and interesting things can be said about such situations as well. Section 11 will finally illustrate this with 'Shangri-La', another nice example of Arntzenius's (2003).

In treating all those examples, I will be breaking a fly on the wheel. Somehow, it may seem clear enough what to think about them without all the machinery to be introduced. However, only by introducing the machinery we get at the general principles lying behind those examples. It is those principles that are in the focus of this paper.

In all those fields the issue is how to deal with one's uncertainty. We know by now that there are many different measures to represent uncertainty, and there are many arguments for preferring one measure to the others for various purposes (see, for example, Halpern 2003 and Spohn 2012). This is a huge debate. However, it is not one that specifically concerns Sleeping Beauty or the other examples. Hence, on this score, I will strictly stick to the traditional account of uncertainty in terms of subjective probabilities, which is the most familiar, widespread, and successful one.

2. Algebraic Preliminaries

Since Castañeda (1966), Perry (1979), and Lewis (1979) it is commonly accepted that indexical beliefs, that is, beliefs about who I am and when is now, are irreducible, that is, not reducible to non-indexical beliefs.[3] I share the common opinion. Moreover, I shall adopt here the wise policy of conceiving of belief as a propositional as opposed to a sentential attitude. Thereby, I neglect problems of hyperintensionality, and by attending directly to the belief contents I avoid getting dangerously involved in the ambiguities and intricacies of the linguistic expressions of those contents. This stance is indeed forced upon us by probabil-

3. Stalnaker (2014: Chapter 5) is one of the few who still disagree.

ity theory itself; I do not know of any sensible way of doing hyperintensional probabilistic epistemology.

No probabilistic considerations without explicitly introducing the probability space within which to move! Neglecting this imperative is usually not a venial sin. A probability space consists of a sample space or a set of possibilities (or whatever it is called), an algebra of subsets of that set, which represent belief contents or propositions, and a probability measure defined on that algebra. How those sets, the propositions, are linguistically specified is not relevant. So, what's the appropriate probability space for our purposes?

For non-indexical belief the possibilities are usually conceived as possible worlds and the relevant contents as sets of possible worlds or propositions in the narrow sense. Lewis (1979) proposed dealing with indexical belief simply by taking its contents to be sets of centered possible worlds, that is, centered propositions, or, as he said, (self-ascribed) properties. I shall call centered propositions also propositions in the wide sense, and I shall use only the wide sense in the rest of the paper. Thus, the objects of belief are uniformly propositions, though we will soon distinguish various kinds of them.

Lewis (1979) suggests that this is all the change that is required.[4] He was almost right. There are some special epistemological features of indexical belief, though, to which we have to attend. First, let's restrict our topic. Usually, the center of a centered world consists of a subject and a time. We might also add a location in space in order to supply a reference for "here" directly and not only as "the place where the speaker is now". However, uncertainty about who I am and where I am is not our issue, though the formalism below might easily be extended accordingly.[5] Hence, it suffices here to take as a 'center' only a time t, which represents that *now* is t and thus allows treating uncertainty about when is now (namely t? or some other time t'?). So, in principle, a centered world is just a pair $\langle t, w\rangle$ of a time t and a world w, and a proposition is a set of such pairs. That's it? No, we should exercise still more care, even if it is somewhat tedious:

First, *time*: we will consider only discrete, linear time; there is no need to indulge in extravagancies. So, we may represent time by the set T of (non-negative and negative) integers. This already makes for infinitely many possibilities. If

4. He says, ". . . it is interesting to ask what happens to decision theory if we take all attitudes as *de se*. Answer: very little. We replace the space of worlds by the space of centered worlds . . . All else is just as before" (1979: 534).

5. I am not sure, though, how interesting the first extension would be. One point is that not knowing who I am is by far a more extraordinary condition than being (somewhat) lost in time. Another point is that times are ordered and have nice arithmetical properties to be exploited below. By contrast, the set of subjects is not structured in any way. Location in space is different in both respects. Not knowing where I am is not unusual. And space comes with a rich structure. So, I think all the epistemic considerations below about temporal self-location can be transferred to spatial self-location. However, I won't pursue this line of thought.

you prefer finite models, you may enforce this by requiring that there is a finite set of temporal locations receiving probability 1, so that the infinity of other times is epistemically irrelevant. In any case, the epistemic subject has to locate herself somewhere within T.

Second, *worlds*: usually, just a set W of possible worlds is assumed. They may be Lewis's grand worlds, spatiotemporally maximal possible objects; but we better conceive of them as small worlds, as indeed we must do, given our discrete conception of time. In fact, wherever probability theory is applied, the probability space consists of what philosophers call small worlds.[6]

There is a problem, though, with this usual procedure. Of course, those worlds have a temporal extension, a time, and things are going on in them in time. Again, let us avoid extravagancies and assume that all worlds have the same time; we do not discuss beliefs about different possible structures of time. The problem is that the time of a world is only implicit in the world; any w in W is unstructured so far. However, we need a way of representing when is happening what in a world. And our modeling itself must provide this way; we can't do this by saying that, for example, "Newman is born at January 1, 2000" is true in w, because we have abstracted from sentences and temporal references therein.

Hence, let us rather assume a set S of (momentary) *world states.* Then we can represent each world in W as a temporal succession of world states, that is, a function from T into S. Thus, $W = S^T$. This formalization is standard within mathematics and physics in dealing with stochastic processes. Again, if you think this is too large and not any wild succession of world states can be a possible world, you may assume that there is a small subset of S^T receiving probability 1, so that our epistemic considerations are in effect restricted to this subset. Moreover, since the worlds are small ones, the world states might be few, for example, just the possible results of a throw of a die (so that the worlds are just sequences of such results).

To resume, each epistemic possibility is a pair $\langle t, w\rangle$ consisting of a time $t \in T$ and a world $w \in W = S^T$, and the epistemic space on which the epistemic states of the subject operate is $T \times W = T \times S^T$. Note that time is represented twice here. The first T represents indexical time, the possibilities for the present temporal location of the subject; it roughly corresponds to McTaggart's (1908) A-series. The second T represents objective time in the possible worlds; it roughly corresponds to McTaggart's B-series. In a way, though, both times are subjective, in so far as they figure only in the epistemic space of the subject. Therefore, we are bound to assume that both times have the same structure T; of course, the subject thinks that her presence moves within what she takes to be the objective time T.

6. Or the propositional algebra is so coarse-grained that most of Lewis's grand worlds remain undistinguished.

We may leave it open, though, whether her objective time agrees with how time really is. (Maybe we are presently all wrong and mythical conceptions of cyclic time are right.)

Note, moreover, that by assuming small worlds we can avoid all ontological issues concerning the identity of times or about transworld identity; in small worlds these issues are solved per fiat. Surely, these issues are legitimate and serious. But we need not burden our investigation with them.

We will have to talk a lot about *time shift*. Hence it is important to note that, if time plays a double role in epistemic possibilities, there are also two kinds of time shift, indexical and objective ones. *The indexical shift* of the possibility $\langle t, w\rangle$ by $z \in T$, denoted by $\langle t,w\rangle^{i}_{z}$ is the possibility $\langle t + z, w\rangle$, in which now is simply z units of time later (or earlier, if z is negative).

The objective shift is slightly more difficult to define. The idea is that after the shift all things happen objectively later than before the shift. After we have explicitly built objective time into the worlds, this may be explicated in the following way: for any world $w \in W = S^T$ and any time $z \in T$, the world w_z, the *objective shift* of w by z, is to be the function $w_z(t) = w(t - z)$. So, the state of w_z at t is just the state of w at $t - z$, and thus in w_z exactly the same things happen as in w, only z units of time later (or earlier, if z is negative). Then, the *objective shift* of the possibility $\langle t, w\rangle$ by $z \in T$, denoted by $\langle t,w\rangle^{o}_{z}$ is simply $\langle t, w_z\rangle$.

Note that this definition of objective time shifts essentially relies on explicitly conceiving worlds as temporal sequences of world states and on conceiving objective time in the same arithmetical way as indexical time. If we had left objective time implicit in worlds (as in Lewis's original centered worlds), those objective time shifts could not have been defined so straightforwardly.

One may wonder in which sense the world w_z ($z \neq 0$) differs from the world w. This worry seems particularly apt given our representation of time by integers. There is no objective zero time, and hence no genuine difference between w and w_z. Again, though, we can avoid these metaphysical issues. The small worlds we are considering are implicitly embedded into bigger or even grand worlds, and their time is embedded as well. And relative to the reference time in the bigger worlds, the temporal shifts in the small worlds clearly make a difference.

Next, we have to build propositions from epistemic possibilities. To begin with, a *proposition* (in the wide sense) is simply a subset of $T \times W$. For simplicity I shall assume that the algebra $\mathcal{A}$ of propositions is the power set of $T \times W$. Measurability issues, for example, whether there are non-trivial σ-additive probability measures on $\mathcal{A}$, are not relevant here.

With our shift operations we can also shift around propositions. This will be required for classifying propositions in relevant ways. For any proposition $A \in \mathcal{A}$ the *indexical shift* of A by $z \in T$ is defined as $A^{i}_{z} = \{ \langle t,w\rangle^{i}_{z} \mid \langle t, w\rangle \in A\}$ and the *objective shift* A^{o}_{z} of A by $z \in T$ as $A^{o}_{z} = \{ \langle t,w\rangle^{o}_{z} \mid \langle t, w\rangle \in A\}$. The shifts obviously

commute: we have $(\langle t,w\rangle_y^o)_z^i = (\langle t,w\rangle_z^i)_y^o$, and hence also $(A_y^o)_z^i = (A_z^i)_y^o$. The effect of those shifts on propositions is best seen when we have a look at various special kinds of propositions.

There is, first, the algebra $\mathcal{W}$ of eternal propositions only about W and not concerning self-location. Formally, a proposition A is *eternal* iff for all $t, t' \in T$ and $w \in W$: if $\langle t, w\rangle \in A$, then $\langle t', w\rangle \in A$. Thus, equivalently, we might represent an eternal proposition simply by a subset of W, a set of worlds. Note that I shall use both representations, for the sake of simplicity. For instance, "Newman is born at January 1, 2000" or "all emeralds are green" or, more explicitly, "for all x and t, if x is an emerald at t, x is green at t" express eternal propositions.

If A is an eternal proposition, then $A_z^i = A$ for all $z \in T$; eternal propositions are not changed by indexical shifts. Indeed, we may define eternal propositions by this feature. However, the objective shift A_z^o is usually a different proposition. For example, if we objectively shift "Newman is born at January 1" by three days, we get "Newman is born at January 4". There are also eternal propositions A for which $A = A_z^o$ for all $z \in T$. We might call them *stationary* or *time-invariant*. "Everyday the sun rises", or "all emeralds are green" in the above explicit version, are examples. They will play a role only in section 5.

Secondly, there is the algebra $\mathcal{T}$ of (purely) self-locating propositions only about T and not about what goes in the world. Formally, the proposition A is *self-locating* iff for all $t \in T$ and $w, w' \in W$: if $\langle t, w\rangle \in A$, then $\langle t, w'\rangle \in A$. So, we might conceive of a self-locating proposition simply as a set of times, that is, a subset of T. Accordingly, $\{t\}$ represents the proposition that now is t, which I often denote by '*now* is t' for the sake of vividness. If times are days, "today is March 1, 2016", for example, expresses a self-locating proposition, and "tomorrow is March 2, 2016" expresses the very same proposition. Given the construction of our calendar, both sentences are analytically equivalent and describe the same set of possibilities.[7]

For $z \neq 0$ the indexical shift A_z^i of a self-locating proposition A differs from A, because in A_z^i all possibilities when *now* may be according to A are shifted by z units of time. By contrast, the objective shift does not change anything. For self-locating propositions A we always have $A = A_z^o$. However, this feature does not define self-locating propositions; as just observed, time-invariant propositions have it as well. Obviously, $\mathcal{A}$, the algebra of all propositions, is the product algebra of $\mathcal{T}$ and $\mathcal{W}$.

All the other propositions not in $\mathcal{T}$ and $\mathcal{W}$, the large majority, are neither eternal nor self-locating. Let's call them *mixed propositions*. One might call them

7. In the discussion of examples, propositions have to represented by sentences or utterances, and then we must pay attention to what is being believed or getting the probability: the sentence/utterance or the proposition expressed? If the former, this would mean treating belief as hyperintensional, something strictly avoided here, as mentioned above.

indexical propositions. However, I want to reserve this label for a much narrower class. For, mixed propositions in general might be quite strange. In principle, each proposition $A \in \mathcal{A}$ can be represented in this form: if now is t_1, then $A(t_1) = \{w \mid \langle t_1, w\rangle \in A\}$, if now is t_2, then $A(t_2) = \{w \mid \langle t_2, w\rangle \in A\}$, and so on, where each $t_i \in T$ and $A_i \subseteq W$. The $A(t_i)$ can be any eternal propositions. Hence, among the mixed propositions we find such odd things as "if today is Monday, all emeralds are green, and if today is Tuesday, some emeralds are blue, if today is Wednesday, Newman is born at January 1, 2000, and . . ." Still, since they are in the propositional algebra, we have an epistemic attitude towards such odd propositions as well, at least in principle.

However, they don't fit our notion of an indexical proposition. This notion rather refers to propositions expressed by sentences like "today Newman is born", "it will rain tomorrow", etc. That is, an indexical proposition says that something specific happens now or yesterday or next year, etc., whichever time is now. We have already provided everything required for explicating this notion in our framework:

The idea is that each indexical shift is accompanied by an objective shift of the same size, so that, in a sense, the same is claimed for each indexical time. More precisely, $B \in \mathcal{A}$ is an *indexical proposition* iff there is a time $t \in T$ and an eternal proposition $A \subseteq W$ such that $B = \bigcup_{z \in T} \{t + z\} \times A^o_z = \{\langle t + z, w_z\rangle \mid z \in T, w \in A\}$. Thus, B says: for all z, if now is $t + z$, then A^o_z (and thus, in particular, if now is t (+ 0), then A (= A^o_0)). Obviously, the indexical propositions form an algebra, too. Indexical propositions may be shifted in turn indexically and objectively. Indeed, it is worth noting that for any indexical proposition B, $(B^i_z)^o_z = (B^o_z)^i_z = B$ for all $z \in T$. Indexical propositions are distinguished by this feature. Note that this explication of indexical propositions presupposes that indexical and objective time have the same structure and that the shift operations are defined as above.

Perhaps, this notion of an indexical proposition is still too wide. For, time-invariant propositions are also indexical according to this definition; they change neither by indexical nor by objective shift. We might say that an indexical proposition is *proper* if it is not time-invariant, that is, if it changes by some kind of shift. However, the proper indexical propositions do not form an algebra; for instance, $\emptyset$ and $T \times W$ are not properly indexical. So, we better accept time-invariant propositions as a limiting sub-algebra of indexical propositions. The point will be relevant in Section 5.

There is a still narrower kind of proposition, which will prove useful. Indexical propositions need not say that something happens *now*. "It will rain tomorrow" expresses an indexical proposition; this is what we want to say. Clearly, though, we can also define propositions that only describe the present state of the world. Let $R \subseteq S$ be some set of world states. Then $R(t) = \{w \mid w(t) \in R\}$ is

the eternal proposition that some state in R realizes at t; it does not say anything about other times. So, we may define $\mathcal{W}_t$ to be the algebra of *eternal propositions A about t* such that $A = R(t)$ for some $R \subseteq S$. Then, we may finally say that B is a *present* (*time*) *proposition* iff there is a $t \in T$ and a $R(t) \in \mathcal{W}_t$ such that $B = \{\langle t + z, w_z\rangle \mid z \in T, w \in R(t)\}$. In short, such a B states "*now R*". Obviously, t could be replaced by some other t' without affecting B. Each present proposition is indexical, but not reversely.

These algebraic considerations were a bit tedious. However, they will prove beneficial. Indeed, I find this explicitness indispensible. And we hardly had a choice. Probabilities only refer to propositions, sets of epistemic possibilities. Following Lewis, epistemic possibilities are (temporally) centered worlds. And then we added an objective temporal structure to the worlds by conceiving them as temporal sequences of world states. Indeed, this structure had to be the same as that of indexical time. That's it. We might have left all of this implicit; this would have been worse. We could use branching time world models, which are more general.[8] However, the advantages this formalization might have will not be relevant in our context.

In any case, we are now well armed to approach our epistemological issues. Let us first turn to the epistemological complications entailed by uncertain temporal self-location.

3. Temporal Self-Location

Our subject has beliefs about all those propositions in the algebra $\mathcal{A}$, or rather some probability measure P on $\mathcal{A}$, with which the probability space is completed. There is not much to say about P, except that is has to satisfy the axioms of probability. Uncertainty may hide everywhere, also in self-locating, indexical, and mixed propositions. Within a purely static perspective, no novel epistemic phenomena emerge by adding the self-locating component to the epistemic possibilities.

Indeed, we can show that, if there is certainty about self-location, the entire indexical extension collapses. More explicitly: if there is a $t \in T$ such that the subject is certain that now is t, that is, $P(\{t\} \times W) = 1$, then for any $B \in \mathcal{A}$ there is an eternal proposition A, namely $A = \{w \mid \langle t, w\rangle \in B\}$, such that $P(B) = P(B \cap \{t\} \times W) = P(A)$. In this case the full measure P is determined by its restriction to eternal propositions. Hence, standard epistemology is justified in neglecting the indexical extension to the extent that certainty about self-location may be presupposed

8. See, for example, Rumberg (2016), in particular Section 4. Müller (2016) applies this framework to Sleeping Beauty.

(as may normally be done in our quite recent neighborhoods where clocks are everywhere).

Whatever is interesting in the indexical extension shows up only in the dynamic perspective. And it indeed does. In this perspective we must ask how the subject's probabilities change over time. Let P be her probability measure on $\mathcal{A}$ at some prior time τ and P' her probability measure at some later, posterior time τ'. We need not be more specific about τ and τ'. The diachronic question then is, how do P and P' relate?

Note that a third kind of time is involved here, the real time, as we might say, in which the subject actually moves and which we, the observers, state. Hence the new symbols τ and τ'. Of course, the subject tries to epistemically track real time. However, we only want to capture her epistemic business with her indexical and objective time. Whether she succeeds in tracking real time, whether she can refer to τ and τ' objectively or only by "now" and "then", is not our issue. Even if she is certain in P that now is t, we are not interested in checking whether or not t is the real time τ.

The change from P to P' may have many causes—moods, drugs, forgetfulness, experience, etc. However, only changes through experience seem rationally assessable, and we will first focus on them. These changes are well described by well-justified conditionalization rules. There is so far no reason to doubt that they apply to the indexical extension as well. Let me only state the two most familiar rules, which we shall use later on. I think there is no need to enter a justificatory discussion.

The basic rule is *simple conditionalization*: if $E \in \mathcal{A}$ is all of the information the subject receives between τ and τ', if the subject accepts E with certainty, and if there is no further cause of epistemic change, then for all $A \in \mathcal{A}$

(1) $\quad P'(A) = P(A \mid E)$ (provided that $P(E) \neq 0$).

Of course, if other causes for change intrude, we cannot expect (1) to hold. So far, E may be any proposition in $\mathcal{A}$. One may object that not any proposition can be a content of experience. Yes, but let's not now dig into the difficult nature of experience. This is why the rule speaks of information. Any proposition E can be the content of information. And the subject can receive it, for instance, simply by being told that E (if the subject takes the informant to be trustworthy). One may insist that in this case the experience is being told that E, rather than E itself. Note, however, that the epistemic possibilities may be small and coarse-grained and may thus represent only E and not the telling of E. In this case, E is still the only information represented within the conceptual framework. So, let's be content with (1) for the time being. We will return to the issue.

There is another ground for discontent with simple conditionalization,

namely its assumption that the information E is received with certainty. There are many reasons for not being certain about the information received: bad observation conditions, unreliable measurement devices, etc. In particular, given my appeal to coarse-grained epistemic possibilities, I may be certain that I was told that E; but this does not entail that I become certain of E itself. And again I find no certain input explicitly represented in the coarse-grained possibilities.

For this reason, Jeffrey (1983: Chapter 11) has proposed the rule of *generalized Jeffrey conditionalization*: let $\mathcal{E} = \{E_1, \ldots, E_n\}$ be a partition of $T \times W$, called the *informational* or *experiential partition*, let the informational process between τ and τ' result in some posterior probabilities $Q(E_i)$ ($i = 1, \ldots, n$) for this partition, and assume again that there is no further cause of epistemic change. Then for all $A \in \mathcal{A}$

$$(2) \qquad P'(A) = \sum_{i=1}^{n} P(A \mid E_i) \cdot Q(E_i) \text{ (provided that } P(E_i) \neq 0 \text{ for } i = 1, \ldots, n).^{9}$$

In particular, this entails that $P'(E_i) = Q(E_i)$ for all $E_i \in \mathcal{E}$. The idea here is that the probabilities conditional on the members of the informational partition remain unchanged and thus determine all posterior probabilities together with the new probabilities for those members according to (2). This idea has found many justifications; Teller (1976) is still my favorite. Again, the informational partition may be so far any partition whatsoever. Moreover, it is important that there is no rational constraint on the new $P'(E_i) = Q(E_i)$; they are just the contingent result of the informational process.

So much for the standard conditionalization rules. We may take them for granted and need not enter any discussion of alternatives or generalizations. However, we still have to scrutinize them within our extended framework. Before doing so, we must first attend to the fact that these rules can't be all there is to belief change within the indexical extension. This is accepted by everyone participating in the discussion; divergences are only about how to describe this in an appropriate way.

What is it that is missing? For instance, I believe now that today is Thursday. A day passes, and then I no longer believe that today is Thursday; rather I believe that today is Friday and yesterday was Thursday. Also, I believe now that it rained today, but then I instead believe that it rained yesterday. So, some belief change must have occurred. The point is slightly confusing because it seems also correct to say, in a sense, that no belief change has occurred. Now I believe that it rained on Thursday, and this is what I keep believing; later on I express this

9. Often, Jeffrey conditionalization is understood as dealing with ineffable experience, which is not representable by any proposition whatsoever. This understanding is indeed suggested by Jeffrey's (1983) example of the observation by candlelight. However, as I have presented the matter, ineffability is not the point at all. The point is only whether or not the content of experience or information is represented in the *given* propositional algebra.

only in a different way by saying "it rained yesterday". Hence, one might say that there was no genuine belief change. Well, there is no belief change in eternal propositions in this case, and one is free, but perhaps not well-advised, to say that only such change is genuine. By all means, there is a belief change in indexical propositions, and this must count as belief change, too.

Could this belief change be modeled by some rule of conditionalization? It does not seem so. One could advance the formal argument that, if the prior probability of "now is Thursday" is 1, no conditionalization can lower this, let alone, make it vanish. Intuitively more convincing, I find, is the fact that I have not received any material piece of information, I have not had any experience that brings about that change. I could have closed my eyes and could have shut my ears, and still that belief change would have come about, and rationally so. No proposition is provided to conditionalize on.

It would be problematic, though, to say that no experience at all is involved. I could have failed to realize that some time, or a day, has passed, and then I would not have changed my belief from "now is Thursday" to "now is Friday". How do I realize this? Of course, I receive a lot of external signals that help me to align; this is experience proper. We still have to discuss how this might work.

However, even in the absence of any such signals I am able to realize this, and I do so simply by my inner sense of time. Surely, this sense cannot be denied. It is the base of Kant's pure inner intuition of time. We could now explore the phenomenology of this sense. Time seems to run fast and slow; sometimes we attempt to calibrate by inner counting; etc. But this is not my interest. The only point is that this sense exists, and it roughly works.

Should we say that this inner sense of time provides experience? This may seem so, even though it is not provided by a proper sense organ. It does help keeping track of time, if only unreliably; it is quite poor without external help. We might as well say, though, that the sense of time underlies our none too successful fight against losing track of time, a fight that would be completely lost after a few days without external calibration. Thus described, it resembles partial forgetting instead of experience. In any case, external calibration is amply provided by nature. For this reason, I presume, there was no evolutionary need to develop a more perfect sense of time. Below we will study how the external calibration works.

Section 9 will forward an argument to the effect that epistemic change induced by the sense of time resembles forgetting rather than learning through experience. However, there is no more than a resemblance. The sense of time is obviously special. And we have to describe how it works. In particular, we have to address uncertainty about temporal self-location. The above examples of indexical belief change are striking, but by assuming that I know which time it is they obscure the issue of change in uncertain probabilities.

So, let us return to our subject with her prior measure P at τ and her posterior measure P' at τ'. And let us suppose that between τ and τ' she undergoes nothing but the passage of time tracked by her inner sense of time and that she incurs no other kind of epistemic change. How do P and P' relate in this case?

At τ we may produce a certain signal, so that she knows that this is the time of her prior epistemic state. However, she may already be uncertain when that is. τ is our term for the prior time, not hers. She is not uncertain about when is τ (though we may say so in a *de re* way); rather at τ she is uncertain about when is now.[10] This uncertainty is represented by the restriction of P to $\mathcal{T}'$, her distribution over the self-locating propositions.

At τ' we may produce another signal, so that she knows that this is the time of her posterior epistemic state. She may still be uncertain when that is. However, the posterior uncertainty may not take any form whatsoever. Rather, the prior uncertainty persists, and it is superimposed by some new uncertainty about how much time has passed from τ to τ'. Again, she need not be able to express this uncertainty in those terms. At τ' she may be able to refer to τ only by "then", or by "the time of the first signal", and she may be unsure whether this was yesterday, the day before yesterday, or earlier. So, in general we can only assume that at τ' the subject has some *incremental* probability distribution p' for how much time has passed since τ; p' is a distribution over T^+, the set of non-negative integers (including 0). Thereby it is assumed that the subject is at least not confused about the direction of time and does not think that the posterior time is earlier than the prior time.

The new incremental uncertainty combines with the prior uncertainty to a posterior uncertainty according to the following *core rule of time shift*, which, to repeat, applies only when the sense of time is the sole cause of epistemic change:

(3) There is a distribution p' over T^+ such that for all $t' \in T$:
$P'(now = t') = \sum P(now = t) \cdot p'(z)$, where the sum is taken over all $t \in T$ and $z \in T^+$ such that $t' = t + z$.[11]

For instance, the subject may be uncertain whether the prior day was June 6 or 7. And she may be unsure whether one or two days have passed since. Thus, now,

10. Here, "now" is taken in a *de dicto* sense. Hence, strictly speaking, it is an abuse of language, since "now" has only a *de re* or wide scope reading in English. Still, the abuse is suggestive, and I will occasionally slip into it.

11. A reviewer has suggested that this process should rather be described in terms of generalized imaging, where the prior probability of a centered world $\langle t, w\rangle$ is shifted to the corresponding worlds $\langle t + z, w\rangle$ with probability $p'(z)$. I prefer to stick to the standard Bayesian ways as far as possible and not to refer to additional notions like that of proximity or similarity of (centered) worlds, as imaging does. In the given case, however, there is no difference. Applying generalized imaging in this way precisely results in (3).

the posterior day, may be June 7, 8, or 9. There are hence two ways for now to be June 8; either two days have passed since June 6 or one day since June 7. So, the probability that today is June 8 is the sum of the probability of these two possibilities. And in determining the probability of each possibility, (3) assumes that the prior uncertainty and the incremental uncertainty are *independent*. I have not much to offer as justification of this assumption—except that I have no idea how the incremental uncertainty produced by our inner sense of time could depend on the prior uncertainty.[12]

Of course, (3) also works in the case of certainty. Return to the above case where I am first sure that it is Thursday, that is, $P(now = \text{Thursday}) = 1$; then I am sure that one day has passed $p'(1) = 1$; and hence I am sure later on that it is Friday, that is, $P'(now = \text{Friday}) = P'(now = \text{Thursday} + 1) = 1$. Consequently, $P(now = \text{Friday}) = 0$ and $P'(now = \text{Thursday}) = 0$. This demonstrates that the core rule of time shift cannot be a form of conditionalization, since it changes probability 1 into 0 and 0 into 1.

(3) poses a severe restriction on P'. Not any P' can come from P via (3). Hence it is a substantial rationality constraint. However, I don't see how rationality could pose stricter demands on belief change through time shift. In particular, there do not seem to be further rationality constraints on the incremental probability p' (except the one that it is defined on T^+). I may have a good or a bad sense of time, just as I may have good or bad eyes. If I have bad eyes, it is certainly rational to wear glasses. And if I have a bad sense of time, I am well advised to look at my watch often. But having a bad sense of time is not irrational as such. So, I don't see how to further constrain p'. Of course, it is factually constrained because our sense of time is usually not so bad; therefore, the literature resorts to fancy examples where our sense of time leads us badly astray.

The core rule of time shift (3) is still incomplete. So far it only says how the probabilities for self-locating propositions shift. How does the rest of P' change? Recall that our presupposition was that the subject has no experience between τ and τ' except the one mediated by her sense of time (if this is to be called an experience). So, the situation seems to exactly fit the description of Jeffrey con-

12. A reviewer has also raised the following question: p' is part of the characterization of the subject's posterior epistemic state. So, could it not be integrated into the posterior measure P'? Yes, it could. This would require introducing suitable propositions as objects of p' to which P' should be extended, propositions of the kind "from then to now z units of time have passed", where the reference of "then" must be somehow internally given, for example, as the time of the first signal. One might say then that the inner sense of time provides information precisely about this kind of proposition. However, these propositions play a role only in rule (3) and the next rule. Hence, I prefer not to generally complicate our algebraic apparatus with these additional propositions. Note also that (3) does not turn p' into an explicit part of the posterior epistemic state. (3) only says that there must be some p' such that P' comes from P by (3). Thereby, (3) is able to treat time shift within the given algebraic framework.

ditionalization: we have some new probabilities, in this case for the self-locating propositions according to (3). The relevant conditional probabilities seem to remain unchanged through time shift. So we keep the prior conditional probabilities and apply (2).

I said "seem" twice because we have neglected an important point. When time shifts, the contents of belief shift as well. What is a belief in A at τ is a belief in something else at τ'. In order to capture this precisely, recall that any proposition A is of the form "for all $t \in T$, if now is t, then $A(t)$", where each $A(t) \subseteq W$ is an eternal proposition. So, when we look at the posterior P', we have $P'(A \mid now = t') = P'(A(t') \mid now = t')$. Which prior probability does this preserve? Well, if t' results from shifting t by z, then what is preserved must be the probability of the very same eternal proposition $A(t')$ given that now is the earlier time t. I conclude that $P'(A \mid now = t') = P(A(t') \mid now = t)$. Moreover, we have that $A(t') = A^i_z(t)$. Hence, finally, $P'(A \mid now = t') = P(A(t) \mid now \quad t) = P(A^i_z \mid now = t)$ —provided that t' results from shifting t by z. This is how conditional probabilities are preserved under indexical time shift.

This enables us to extend the core rule of time shift (3) by the derived adaptation of Jeffrey conditionalization (2). For any proposition $A \in \mathcal{A}$ we have

$$(4) \qquad \begin{aligned} P'(A) &= \sum_{t' \in T} P'(A \mid now = t') \cdot P'(now = t') \\ &= \sum_{t' \in T} \sum_{t+z=t'} P(A^i_z \mid now = t) \cdot P(now = t) \cdot p'(z) \end{aligned}$$

(where the sums are, respectively, taken only about those t' and t for which $P'(now = t') > 0$ and $P(now = t) > 0$, so that the conditional probabilities are defined). Let's call this the *general rule of time shift*, which again applies only when belief change is exclusively due to the inner sense of time which estimates with probability $p'(z)$ that z units of time have passed from τ to τ'.[13]

This is a fully general rule of belief change for this type of situation. Note that probabilities of eternal propositions do not change according to (4):

13. In his survey article, Titelbaum (2016) classifies the many proposals for accounting for self-locating credences according to three schemes: shifting schemes, stable base schemes, and demonstrative schemes. Clearly, my rules (3) and (4) fall under the shifting schemes. He mentions the story of Rip who has credence .7 on July 4 for "it rains today", falls asleep and wakes up much later, not knowing which day it is. Titelbaum asks, to what proposition should Rip now assign a credence of .7? And he senses trouble for the shifting schemes and thereby motivates the other schemes. However, there is no trouble according to (3) and (4). If Rip knows on July 4 that it is July 4, he should have a credence of .7 for "it rained on July 4", on July 4 and later on. However, whether or not he knows this, he won't have a credence of .7 for any indexical proposition (like "it rained the day before yesterday"). Rather, his posterior credences of indexical propositions will be determined by (4) and his uncertainty p' about how long he slept.

(5) If P' comes from P according to (4), then $P'(A) = P(A)$ for all eternal propositions $A \in \mathcal{W}$.[14]

This observation might explicate the view that there cannot be any 'genuine' belief change through mere time shift.

This concludes my account of belief change merely due to the inner sense of time. Let's consider change through learning or experience. The first thing to observe is that usually both forces are operative in a belief change from τ to τ', experience and the inner sense of time. So, neither (2) nor (4) by itself will do; we have to combine the two rules. I want to suggest that we best do this by cascading the two changes: first the change (4) due to mere time shift, and then the change through experience according to (2). This is clearly preferable to the reverse order, since experience helps calibrating our sense of time. According to the order suggested, the two-step change ends up with that calibration, whereas according to the reverse order, the uncertainty produced by the unaided sense of time would be reintroduced in the second step.[15]

Of course, the two-step procedure is artificial. Usually, the two steps continuously go hand in hand and cannot be clearly separated. In fact, experience will be the strongly dominating force, and we would get along well without our inner sense of time (in our modern vicinities with ubiquitous clocks). Still, within our artificially discrete framework, we can do no better than with the two-step belief change proposed.

If so, we may deal with the two steps separately. We have already treated the first step. So, let's turn, for the rest of this section, to the second step, belief change induced by information or experience proper (excluding the inner sense of time). It is well described by (1) and (2). There is no reason why these conditionalization rules should not apply within our more general framework as well. However, this framework allows us to say more about the nature of evidence or information as used in (1) and (2). I have explained why, in principle, I take any proposition or any partition as admissible evidence in the rules (1) and (2). The basic reason was the coarse-graining of the underlying propositional algebra. However, the indexical extension assumed some (temporal) fine-graining; hence, we may have a more specific conception of possible evidence and should study its consequences.

We may start by considering the special case in which we receive certain evi-

14. This confirms the conclusion of Bradley that belief mutation cannot "produce a shift in credence in eternal beliefs" (2011: 397). Belief mutation is a "belief change in virtue of a change in the truth-value of the content of belief" (2011: 395), something only indexical contents can do. So, it is clear that his belief mutation is precisely captured by my rules (3) and (4) of time shift.

15. In his Continuous Conditionalization, Schulz (2010: 341) proposes exactly the same two-step procedure.

dence about a self-locating proposition. However, this is not an interesting case. In contrast to (4), where uncertainty concerning self-location spreads further, this case removes the uncertainty and thus returns to the trivial case referred to at the beginning of this section where the indexical extension collapses. We might assume that evidence about self-location is uncertain in itself. Then learning proceeds via Jeffrey conditionalization (2), and the first equation of (4) applies, with the only difference that the $P'(now = t')$ are not produced by the sense of time, but by the new piece of evidence.

However, a more interesting observation is that, within our fine-grained framework, evidence never really comes in form of a self-locating proposition. This might seem doubtful. I look at my watch. Don't I immediately learn the self-locating proposition that it is 10 a.m. now? Yes, one may thus describe the case in this coarse-grained way. However, within our more fine-grained perspective we would have to describe the case in the following way: I look at my watch. I see that its hands are *now* pointing to 10 a.m. I know, or believe, that the watch is correct, that is, that it shows the correct time. So, I believe that the hands are pointing to 10 a.m. only at 10 a.m. and infer from what I see that is 10 a.m. now.[16]

Moreover, within our fine-grained framework, evidence never comes in form of an eternal proposition. This mirrors the previous point. Just as I never perceive which objective time it is now, I never perceive that something is going on at a specific objective time. Rather I learn that something is going on *now*, and then I use my information and my background knowledge in order to infer what time now may be and at what time those things were going on. This is how it works since the times of Stonehenge, when people saw that the sun has *now* a certain position and inferred that *now* is equinox, a specific time of the year. (Forgive again the ungrammatical *de dicto* use of "now".)

Hence, evidence comes in form of a mixed proposition. Well, not any mixed proposition. Rather, evidence consists in an indexical or, more specifically, a present time proposition: *now* the world is in a certain state. This form of evidence will be crucial when we apply our framework to Goodman's new riddle of induction in Section 5.

It is instructive to study how such evidence may promote our self-location as well as our eternal beliefs. Contrary to my suggestion, let the evidence E_{now} consist in any mixed proposition. The reason for considering the general case is that I do not see particularly simple results following from natural restrictions on the evidence E_{now}. As explained above, we can represent E_{now} as $\bigcup \{t\} \times E(t)$ where each $E(t)$ is an eternal proposition. In case E_{now} is a present time proposition, $E(t)$ is the eternal proposition that at t some state from a set E of world states obtains;

16. Note that "10 a.m." is ambiguous in this description. One usage refers to a certain time, the other usage refers to a certain position of the hands of my watch.

in this case E_{now} could indeed be expressed by "now E". In the general case $E(t)$ might be any eternal proposition, for each $t \in T$, and then E_{now} might have no simple linguistic expression. Formally, however, it does not matter whether $E(t)$ is simple or varies with t in complex ways.

Our set-up is still that our subject has P at τ, which is now the prior probability of the second step we are just discussing, and P' at τ', the posterior probability of the second step, which comes from P via simple conditionalization on E_{now}; that is, $P' = P(. \mid E_{now})$. Let's study only simple conditionalization; the generalization to Jeffrey conditionalization is straightforward. And let's assume that the evidence E_{now} works instantaneously, as it were, that is, that $\tau' = \tau$ in our coarse-grained conceptualization of time. This agrees with our two-step procedure, which applies first the general rule of time shift (4) to the time elapsed during the first step and then the learning rules to be stated now. Our first question is, how might such evidence help the subject to self-locate? In the following way:

$$
\begin{aligned}
(6) \qquad P'(now = t) &= P(now = t \mid E_{now}) = P(\{t\} \times W \cap E_{now}) \,/\, P(E_{now}) \\
&= P(\{t\} \times E(t)) \,/\, \sum_{z \in T} P(\{z\} \times E(z)) \\
&= P(E(t) \mid now = t) \cdot P(now = t) \,/\, \sum_{z \in T} P(E(z) \mid now = z) \cdot P(now = z)\,.
\end{aligned}
$$

This is, in a way, a variant of Bayes' theorem. The 'hypotheses' are of the form "now is t", and for each hypothesis t the evidence results in a different eternal proposition $E(t)$. Thus, the posterior self-location is proportional to the prior self-location $P(now = t)$ and to the 'likelihood' $P(E(t) \mid now = t)$ of the evidence $E(t)$ given the hypothesis '$now = t$'.

Let us look at the most ordinary example: let E_{now} be "the hands of my watch now show 10 o'clock". Assume that my prior says it's morning, anyway. Then $E(z)$ is "the hands of my watch show 10 o'clock at z o'clock". So, under normal circumstances, my prior probability for $E(z)$ given it is z o'clock will be roughly 1 if $z = 10$ and roughly 0 if $z \neq 10$. Hence, according to (6), my posterior probability for its now being 10 a.m. will also be roughly 1.

However, circumstances need not be normal. Suppose that I look at a church clock and that I know that it is broken. My evidence E_{now} is the same as before. But now my prior for $E(z)$ is the same for all z; whatever time the clock shows, it shows it all the time. Hence, I don't learn anything about self-location by looking at that clock, and this is what (6) says for this case.

Clocks are designed for this inference, and when the clock is reliable, the inference is so as well. In principle, though, any state of the world which is likely to realize at certain objective times rather than others is more or less well suited. Indeed, we make permanent use of this fact.

How does the inference extend to other and in particular to eternal propositions? Let A be any proposition in $\mathcal{A}$ (and let's again idealize away the time difference between τ and τ' or between P and P' and assume $\tau = \tau'$). Then:

$$(7) \qquad P'(A) = P(A \mid E_{now}) = \sum_{t \in T} P(A \mid \{t\} \times W \cap E_{now}) \cdot P(now = t \mid E_{now}) =$$
$$= \sum_{t \in T} P(A \mid \{t\} \times E(t)) \cdot P'(now = t) .$$

This looks like a variant of Jeffrey conditionalization; we keep the prior probabilities of A not just given $\{t\}$ (= '$now = t$'), but rather conditional on $\{t\} \times E(t)$ (= '$now = t$ and $E(t)$') and mix them according to the posterior weights $P'(now = t)$.[17]

I don't see how (7) could be further simplified, even by special assumptions about the form of the evidence E_{now}. In particular, learning about eternal propositions need not merely be a consequence of learning about the $E(t)$ (and their logical combinations). The prior may be set up in such a way that we may draw any wild conclusions from our self-location. For instance, Lewis's (1979) story of the two gods, who are propositionally, that is, eternally omniscient, but don't know who or where they are, can be modified such that they are even propositionally ignorant, but become propositionally or eternally omniscient by finding out about their self-location. Suppose God is to create the world, but Arch God determines by lot which god God will be. God is omniscient concerning the lottery and the plan of creation. Thus he knows that if he is Zeus he will create the 'Greek' world, if he is Jahveh he will create the 'Hebrew' world, etc. But he can't see the result of his creation. So he knows how the world actually is only when he knows which god he was determined to be.

This concludes my statement of general rules of probability change in the extended indexical framework. The only new rules are the rules (3) and (4) of time shift. Concerning learning by experience, however, the conditionalization rules (1) and (2) still apply. In particular, simple conditionalization (1) takes the forms (6) and (7) in our extended setting, which specify how calibration of self-location and belief in other and especially eternal propositions changes through indexical or, more specifically, present time experience. As mentioned in Section 2, I think a parallel account could be developed concerning epistemic problems with spatial self-location (where we also have an unreliable inner protocol of how we move in space, which is continuously controlled by external observation).

17. The rule called Approximated Continuous Conditionalization by Schulz (2010: 342) comes closest to my (7) (in his alternative notation). He also observes the similarities to Jeffrey conditionalization. The rule (SC), shifted conditioning, of Schwarz (2012: 222), is intended to have the same effect. However, he captures the indexical component only by a shifting operator $\succ$ ("next"), which underformalizes the indexical component in my view. For instance, I don't see how he could state the rules (3) and (4) of time shift in his framework.

4. First Application: The Prisoner

Let me illustrate all of this through two very different applications, and let me first discuss an example that was precisely invented for the present purpose and nicely combines both, the rule of time shift and the rule of simple conditionalization. It is the story of the prisoner from Arntzenius (2003): you are captured by a whimsical dictator and you know that, depending on the throw of a fair coin, you will either be set free within a day or imprisoned for three years. The first night in prison is agonizing. At 6 pm you are jailed in a completely empty cell; only a light is burning. And you have just your clothes and nothing else. You are told that the light will be turned off a minute before midnight if you will have to stay in prison and it will keep burning if you will be set free. Of course, you stay awake and watch the light. You have two cues as to what time it is: your inner sense of time and the light. And our two rules tell how the cues mix.

Let us keep things simple and consider only four real points of time at 6 p.m., 11 p.m., and 1 a.m. and 6 a.m. the next day. The continuous case is too complicated for a rigorous treatment on a page.[18] Let E_{now} be the present-time proposition that the light is on now, $E(t)$ be the eternal proposition that the light is on at t, and F be the eternal proposition that you will be set free. Let 0 = midnight and let the time units simply be hours; so, 6 p.m. = –6, 11 p.m. = –1, and so on. Let's look at your probabilities P_{-6}, P_{-1}, P_1, P_6 at, respectively, 6 p.m., 11 p.m., 1 a.m., and 6 a.m. You know the rules of the game; so $P_\tau(F \mid E(t)) = 1$ and $P_\tau(\overline{F} \mid \overline{E}(t)) = 1$ for $\tau = -6, -1, 1, 6$ and $t \geq 0$. Let's finally suppose that the light continues burning all night (so that you will in fact be set free soon).

Then P_{-6} and P_6 are easy to describe. When you are brought into the cell, you are told that it's 6 p.m. So, $P_{-6}(now = -6) = 1$, $P_{-6}(E(t)) = 1$ for $t < 0$, and $P_{-6}(F) = P_{-6}(E(t)) = \frac{1}{2}$ for $t \geq 0$. Similarly, for P_6: twelve hours later, at 6 am, you are very unsure which time it is; it might be anything, say, between 2 a.m. and 10 a.m. But you are very sure that it is past midnight; hence, $P_6(now \geq 0) = 1$, $P_6(E(t)) = 1$ for $t \geq 0$, and $P_6(F) = 1$ (since you have seen the light burning all the time).

The interesting thing is what happens in between. Let P^-_{-1} be the probability you would have at 11 p.m., if you would have to rely only on your sense of time. This sense is not accurate; let's assume that $P^-_1(now = -1 \pm k) = (3 \pm k) / 9$ for $k = 0, 1, 2$. That is, at that time it is most likely for you that it is now –1 = 11 p.m., namely to the degree 1/3; but with probability 1/9 it may be –3 (= 9 p.m.) or +1 (= 1 a.m.). Your probabilities for the eternal propositions $E(t)$ and F are not changed thereby. However $P^-_{-1}(E_{now}) = 5/6$, since the light is on for sure with probability

18. Arntzenius (2003: 357–362) and Bradley (2011: Section 3) consider the continuous case, though only qualitatively. And they are right in their qualitative description. Only Schwarz (2012: 225) gives a precise account of the case in terms of his shifted conditioning. Here I give a similar description, in my terminology and with somewhat different figures.

$P^-_{-1}(now < 0) = 6/9$ and with a 50% chance with the probability $P^-_{-1}(now \geq 0) = 3/9$.

Now, you need not only rely on your sense of time, you also see that the light is still on. So, $P_{-1} = P^-_{-1}(. \mid E_{now})$, according to the above proposal of cascading the two changes. This changes your self-location as well as your eternal beliefs, according to (6) and (7). More precisely, we have:

$$(8) \qquad P_{-1}(now = -3) = P^-_{-1}(now = -3 \mid E_{now})$$

$$= P^-_{-1}(E(-3) \mid now = -3) \cdot P^-_{-1}(now = -3) \,/\, P^-_{-1}(E_{now}) = 1 \cdot 1/9 \cdot 6/5 = 2/15.$$

In the same way, we get $P_{-1}(now = -2) = 4/15$, $P_{-1}(now = -1) = 6/15$, $P_{-1}(now = 0) = 2/15$, and $P_{-1}(now = 1) = 1/15$. $P_{-1}(now = 1)$ is only half of $P_{-1}(now = -5)$ because $P^-_{-1}(E(1) \mid now = 1)$ is only half of $P^-_{-1}(E(-3) \mid now = -3)$. All other times get probability 0. This means that your probability for it still being before midnight is raised from $P^-_{-1}(now < 0) = 6/9$ to $P_{-1}(now < 0) = 12/15$. And finally we have

$$(9) \qquad P_{-1}(F) = (F \mid E_{now}) = \sum_{t=-3}^{1} P^-_{-1}(F \mid t \times E(t)) \cdot P_{-1}(now = t)$$

$$= 1/2 \cdot (2/15 + 4/15 + 6/15) + 1 \cdot (2/15 + 1/15) = 3/5.$$

So, observing at that time that the light is still on makes it a bit more probable that it is still before midnight, but it also slightly raises the probability that you will be set free; after all, it may already be midnight or later. These effects are as expected.

Just for illustration let us briefly look how your probability P_1 may be at 1 = 1 a.m. From 11 p.m. to 1 a.m. two hours have passed. However, you are uncertain. Let us still stick to our coarse description and assume that your incremental probability p_1 that one or two or three hours have passed is, respectively, $p_1(+1) = 1/4$, $p_1(+2) = 1/2$, and $p_1(+3) = 1/4$. Now there are two different ways to calculate your later P_1. The first one is that you incrementally build up your later uncertainty about which time it is according to the core rule (3) of time shift. This includes the assumption that you have not aligned at 11 p.m. by looking at the light. So, your uncertainty about your self-location immediately before checking at 1 a.m. whether the light is still on is this (please ignore the numbers in the brackets for now): $P^-_1(-2) = P^-_{-1}(-3) \cdot p_1(+1) = 1/36$ [2/60], $P^-_1(-1) = P^-_{-1}(-3) \cdot p_1(+2) + P^-_{-1}(-2) \cdot p_1(+1) = 4/36$ [8/60], and, similarly, $P^-_1(0) = 8/36$ [16/60], $P^-_1(1) = 10/36$ [18/60], $P^-_1(2) = 8/36$ [11/60], $P^-_1(3) = 4/36$ [4/60], and $P^-_1(4) = 1/36$ [1/60]. So, on the one hand your uncertainty is still more dispersed, on the other hand it's pretty likely that it is already after midnight: $P^-_1(\geq 0) = 31/36$ [50/60]. And as above we find that $P^-_1(E_{now}) = 1 \cdot 5/36 + 1/2 \cdot 31/36 = 41/72$ [35/60].

This allows us to determine the posterior probability $P_1 = P^-_1(. \mid E_{now})$ after

seeing at 1 a.m. that the light is still on, again by applying (6) and (7). By the same calculation as above we get $P_1(now = -2) = 2/41$ [4/70], $P_1(now = -1) = 8/41$ [16/70], $P_1(now = 0) = 8/41$ [16/70], $P_1(now = 1) = 10/41$ [18/70], $P_1(now = 2) = 8/41$ [11/70], $P_1(now = 3) = 4/41$ [4/70], and $P_1(now = 4) = 1/41$ [1/70]. And so finally $P_1(F) = 1/2 \cdot P_1(now < 0) + 1 \cdot P_1(now \geq 0) = 1/2 \cdot 10/41 + 1 \cdot 31/41 = 36/41$ [60/70]. So, you are almost 90% sure that you will be set free.

Don't let us carry the example too far. Let me only note that we have just calculated the less plausible variant of P_1. The more plausible variant would have been that you see the light burning at 11 p.m. and thereby arrive at P_{-1}. Thus, your new uncertainty builds upon P_{-1}, and not upon P^{-}_{-1}, as we have assumed in the first variant. This is a different epistemological position and hence results in slightly different, though qualitatively similar figures, which are shown in the brackets. This makes clear at the same time that it would be quite a mathematically sophisticated task to construct a theoretically well-founded continuous model of the prisoner's situation.

5. Second Application: Indexical Enumerative Induction

Let us turn to a very different example, which has nothing to do with the rules of time shift, but does involve issues of temporal self-location, namely Goodman's (1946) new riddle of induction. This is a riddle for enumerative induction: from the fact that all past or observed emeralds have been found to be green, infer inductively that all emeralds are green. Now define an emerald to be grue iff it is first observed before year 3001 and green or first observed after year 3000 and blue. So, why not infer then that all emeralds are grue, since all green emeralds observed so far are also grue by definition? The probabilistic transformation of this inductive inference is this: the given green emerald confirms, that is, makes it more probable, that the next emerald is also green. Why not say, then, that the given green emerald, which is also grue, confirms that the next emerald is grue again?[19] It is clear that there are countless variations of "grue", which all agree on the observed or given emeralds (and which need not necessarily refer to a first time of observation).

There is no point in reviewing the rich literature on Goodman's new riddle. My lump-sum summary is that it is not clear whether the many alleged solutions of the riddle do full justice to the riddle and provide general reasons that justify breaking the symmetry between green and grue.[20] Maybe you have a

19. In ranking theory the obvious tension between the original formulation of enumerative induction and its probabilistic transformation vanishes; see Spohn (2016).

20. See Stalker's (1994) excellent collection, which is complete up to 1994. One emphasis here is on "justify". Of course, there are nice *explanations* of the *de facto* asymmetry in evolutionary

more sanguine view of the existing literature. Of course, Goodman himself has already proposed to break the symmetry by reference to our past practice and its entrenchment of "green", but not of "grue". However, that's very much like Hume's reference to our habits of thought. The only point I would like to explain here is that our present framework allows a very clear break of the symmetry as well; I shall not discuss, though, its viability as a general solution.[21]

In order to do so let's first focus on enumerative induction as such, apart from the new riddle. And let's change the example. Usually, laws quantify over objects that behave such and such in time. In our context it is more vivid, and easier to translate into our framework, when we consider examples that directly quantify over times. So, let's ponder the simple statement: "The sun rises everyday".

The crucial point is that "the sun rises everyday" is ambiguous in our context. It can refer to indexical or to objective time; that is, it can mean "for each day t in objective time the sun rises at t" or "for each day t in indexical time holds: the sun rises today". These are two different quantifications over different kinds of instances. Formally, the ambiguity is still clearer: let the times in t be days, and let $R \subseteq S$ be the set of states of the world in which the sun rises. Then $R(t) = \{w \mid w(t) \in R\}$ is the eternal proposition that the sun rises at t, and $G^o = \bigcap_{t \in T} R(t)$ is the objective generalization that the sun rises everyday; this is a quantification over objective times. Moreover, as explained at the end of Section 2, we may define $R(now) = \{\langle t^* + z, w_z\rangle \mid z \in T, w \in R(t^*)\}$ as the present proposition that the sun rises today, and $R(now + t) = \{\langle t^* + z, w_z\rangle \mid z \in T, w \in R(t^* + t)\}$ as the indexical proposition that the sun rises t days from now on. Then we can state the indexical generalization G^i that for all t the sun rises t days from now on, that is, $G^i = \bigcap_{t \in T} R(now + t)$; this is a quantification over indexical times.

Surprisingly, the two readings of "the sun rises everyday" come to the same thing. The indexical dependence vanishes in G^i; it remains the same under all indexical shifts. Hence $G^i = G^o$. How is this possible, given that G^o is an eternal proposition (or a conjunction of eternal propositions) and G^i is an indexical proposition (or a conjunction of indexical propositions)? The explanation is that G^o and G^i are time-invariant propositions, and, as explained in Section 2, those are exceptional in being the only ones that are eternal and indexical at the same time.

So the ambiguity lies in the instantiations. The instances of G^o are eternal

terms, for example, by Quine (1969). The other emphasis is on "general reasons". Freitag (2015) has given a beautiful account of the asymmetry in terms of direct or indirect epistemic dependence of the evidence on so-called discriminating predicates, which prevents projectibility. See also Freitag and Zinke (2016). Maybe this is a sufficiently general justification of the asymmetry; however, I can't discuss this issue here.

21. The subsequent ideas were provoked by Anna Kästle, who suggested the scenario to be introduced below and argued in my undergraduate course that it would make a difference.

propositions of the form $R(t)$, $R(t + 1)$, etc., whereas the instances of G^i are proper indexical propositions of the form $R(now)$, $R(now + 1)$, etc. Hence, the inductive inference from the instances to the generalization is a different one, even if the conclusion is the same. Indeed, G^i reminds me of what Quine called (observation) categoricals, the instances of which are given by occasion sentences (modulo something) as opposed to standing sentences.[22]

The probabilistic transformation makes matters still clearer. Thereby, the objective inference turns into "$R(t)$ confirms $R(t + 1)$", whereas the indexical inference turns into "$R(now)$ confirms $R(now + 1)$". Carnap's principle of positive instantial relevance endorses the first claim (1971: 161–165). For the same reason we should also accept its indexical version, the *indexical principle of positive instantial relevance*:

(10) $\quad P(R(now + 1) \mid R(now)) > P(R(now + 1).$

And this principle holds not only for R = "the sun is rising", but for any property R of times, for which $R(t)$ is a well-defined eternal proposition and $R(now)$ is a well defined indexical proposition.

Now, is this of any help with Goodman's new riddle? Yes. Let's invent an analogous riddle about sunrises and define: the sun *frises* at day t iff t is before year 3001 and the sun rises at t or t is after 3000 and the sun fails to rise at t. Or formally: $F(t) = \{w \mid t \leq 3000$ and $w(t) \in R$, or $t > 3000$ and $w(t) \notin R\}$. Since we have so far always seen the sun rising and frising at the same time, the riddle is, of course, why we should expect the sun rising rather than frising every day.

Or in more formal probabilistic terms, why not say that $F(t)$ confirms $F(t + 1)$ and $F(now)$ confirms $F(now + 1)$? We might well say this. Still, there is an asymmetry. It does not show in the eternal propositions $R(t)$ and $F(t)$. However, it does show in the propositions $R(now)$ and $F(now)$: $R(now)$ is a present proposition, whereas $F(now)$ is not a present, but a mixed proposition.

This has consequences for learnability. In Section 3 I stated that within our indexical framework evidence may be assumed to consist in present propositions. Hence, I may get the evidence $R(now)$, but $F(now)$ cannot be evidence for anyone. This point is highlighted by the following scenario:[23] Suppose Friser (= the frising hypothesizer) and I are exhausted by our dispute and conclude that we have to wait till 3000. A lady passes by and takes us on board of her time travel ship, from which we disembark at an unknown time. (Or, alternatively, we are deep

22. For instance, Quine writes, "A generality that is compounded of observables in this way – 'Whenever this, that' – is what I call an *observation categorical*. It is compounded of observation sentences. The 'Whenever' is not intended to reify times and quantify over them. What is intended is an irreducible generality prior to any objective reference" (1990: 10).

23. Suggested by Anna Kästle.

frozen, because we can't live till 3000, and then we are unfrozen at an unknown time.) We wake up and see the sun rise. We both receive the evidence $R(now)$. But does Friser also receive the evidence $F(now)$? No. Since he does not know when now is, he can't tell whether or not he sees $F(now)$. He is as uncertain about this as he is about when is now. So, he can't say whether what he sees conforms to his expectations. Even if someone tells him that he has indeed seen $F(now)$, so that his probability for $F(now + 1)$ is raised, he can't tell what to expect the next day. Suppose we see the sun rise the next day. I can say then that my expectation has been confirmed, but he can't. This day might have been the first of 3001; and in that case his expectation would have been disconfirmed.

So, this is how the assumption that evidence consists in present propositions makes a difference. The difference is not about principle (10). It applies to F no less than to R. The difference is about learning and the application of simple conditionalization (1). I can simply conditionalize on $R(now)$, but Friser cannot simply conditionalize on $F(now)$. Surely, his predicament vis à vis grue emeralds would have been exactly the same.

The difference does not exist necessarily. Friser might get help from a friendly demon who produces a characteristic sound in Friser's head whenever he meets something frising.[24] Then Friser would be in an equally comfortable position as I am facing something rising. However, it is clear, by all means, that *we* do not have any such external help.

Odd things may happen, though. Seeing the sun rising at that unknown time, Friser might claim self-assuredly that he has just seen the sun frising. This would be surprising, and it could only be explained either by some such external help or by the fact that "frising" doesn't mean in Friser's mouth what we took it to mean. Reversely, seeing the sun rising I might suddenly be uncertain whether the sun is rising now. Maybe a strange insanity has befallen me, or maybe "rising" no longer means what we thought it means. However, in such scenarios we are raising second-order inductive doubts concerning our own (or Friser's) linguistic behavior, as Kripke (1982: 58f.) did in his discussion of "grue". This is a different game. Above, I discussed only first-order inductive issues and presupposed that it is clear and fixed which propositions we are dealing with (and hence what the words expressing those propositions mean). And then it is clear that $R(now)$ is a present proposition and $F(now)$ is not.

This kind of asymmetry between Friser and me has been voiced in various ways. Already Carnap (1947) has argued that "grue", unlike "green", is a mixed predicate, that is, a mixture of purely positional and purely qualitative predicates and that presumably only purely qualitative predicates are projectible. But then it seemed that this distinction is language-dependent. Maybe Friser speaks

24. This possibility was pointed out to me by Wolfgang Freitag.

a different language in which "frising" is qualitative? There have been various attempts to make this distinction absolute.[25] We need not assess them here. Certainly, though, being part of present propositions is at least a necessary condition for purely qualitative properties. So, the point I wanted to make here was to discuss enumerative induction within our indexical framework and to argue that this has already some force vis à vis Goodman's new riddle.

6. Forgetfulness

After all this business with indexical self-location, let us turn to our second focal epistemological issue, forgetfulness, which seems almost forgotten in the literature. Because there is not much to say about it? Maybe. So, this section will be quite short as well.

Is there anything to say at all? We cannot be praised for rationality or criticized for irrationality in our forgetting. Forgetting just happens to us. To be sure, we can do a lot to prevent forgetting, if we like. We can memorize things, we can train our memory, or we can write things up before we forget them. However, intentionally forgetting something, if we should want to, is a more difficult exercise, and perhaps not really possible. Forgetting is not like deleting. If we could delete things from our minds in the same way as from a hard disc, then, if required, we should arguably delete illegitimate information first, and then irrelevant information, and if this does not suffice, continue with unimportant information (however this is measured), and so on. But as I said, forgetting is not deleting. There seems no point in setting up and trying to justify rules for forgetting.

Hence, the issue is not to assess (probabilistic) belief change through forgetting, but simply to find a schematic description for it. A pertinent point is that one might think that such a change is always from (more) certainty to (more) uncertainty. Yesterday I knew that the bus leaves at 11 a.m., but now I have forgotten it and wonder whether it is 10 a.m. or 11 a.m. However, the point does not hold; I may also become (almost) certain through forgetting. Yesterday I was unsure whether the bus leaves at 10 a.m. or 11 a.m. Today I have forgotten that 10 a.m. is a live possibility; I only recall 11 a.m. as a possible departure time; and so I come to believe through forgetting that the bus leaves at 11 a.m. Thus it

25. For instance, Hetherington argues that, although the experience *of* an emerald being green is the same as that *of* an emerald being grue (before 3001), experiencing the emerald *as* being green is different from experiencing it *as* being grue (2001: 130). And he insists that this distinction "does not open the door to . . . an unfortunate relativism about the quality of inductive inference" (2001: 133). The explanations he gives of that difference also appeal to the instantaneous character of experience.

seems that forgetting may take any form whatsoever, and no rule can be stated with confidence, not even any descriptive rule.[26]

Also, forgetting is not restricted to a special kind of proposition. In particular, I may forget about indexical propositions. I may be unsure about which week it is, though yesterday I at least knew that I have a date with my dentist next week. Today I may have forgotten or repressed that date.

Moreover, it should be clear that forgetting stands *pars pro toto* for any kind of arational belief change, which may befall us through drugs, alcohol, brain washing, or whatever, and which is opposed to belief change through evidence, learning, or experience, which can be rationally assessed. This seems to prevent any generalizations. Arationality may take any direction.

Should we conclude then that the posterior forgetful doxastic state may be, within our framework, any probability distribution over $\mathcal{A}$ whatsoever? This sounds much too radical. Forgetting is a local affair; only small parts of our doxastic state are concerned, and the rest remains unaffected. The only way I see to formally account for this locality is to say that forgetting directly concerns only some subalgebra $\mathcal{F}$ of propositions or, what comes to the same, the partition $\mathcal{F}^*$ of the atoms of $\mathcal{F}$. If I change my degree of belief in A and in B through forgetting, then the Boolean combinations are affected as well: $\overline{A}$ necessarily, and if $A \cap B$ and $A \cup B$ happen to keep their probabilities, this is so only accidentally.

However, if the propositions in $\mathcal{F}$ change their probabilities, the other propositions in $\mathcal{A}$ must be affected as well; only propositions probabilistically independent of $\mathcal{F}$ are guaranteed to remain unchanged. Can we say something from the rational point of view at this point? Yes, I think so. I want to suggest treating this as a case of generalized Jeffrey conditionalization; that is, at least all the probabilities conditional on the atoms in $\mathcal{F}^*$ remain unchanged. This is the *rule of forgetting*, which applies when the only epistemic change the subject undergoes between between τ and τ' is one of local forgetting:

$$(11) \qquad P'(A) = \sum_{F \in \mathcal{F}^*} P(A \mid F) \cdot P'(F) \text{ for all } A \in \mathcal{A} \text{ (provided } P(F) \neq 0 \text{ for all } F \in \mathcal{F}^*).$$

Here, the prior conditional probabilities $P(A \mid F)$ remain unchanged; provably, $P'(A \mid F) = P(A \mid F)$. And the $P'(F)$ are the posterior probabilities for the partition $\mathcal{F}^*$ somehow resulting from forgetting.

Why should one follow this rule? The dynamic Dutch book argument, which

26. Arntzenius (2003: 367) builds up an opposition between narrowing down one's distribution and spreading and shifting it. His diagnosis is that conditionalization always does the former and cannot describe belief changes that do the latter. While he is right, of course, that there are other belief changes besides conditionalization, I disagree with that opposition. Jeffrey conditionalization may spread one's distribution, and other changes may well narrow it down. Forgetting, for example, may well lead to false certainties, as just illustrated.

essentially refers to conditional bets, has been successfully generalized to Jeffrey conditionalization; see, for example, Teller (1976). It looks doubtful, though, whether that argument carries over to the current use of Jeffrey conditionalization. At least, it seems that I should be wary to accept bets on conditions that I know or suspect to be subject to my forgetfulness. This is a lesson of Bradley and Leitgeb (2006). Similarly, non-pragmatic or epistemic justifications of Jeffrey conditionalization via accuracy measures as given in Leitgeb and Pettigrew (2010) do not seem pertinent, since they are inspired by the idea that our beliefs and degrees of belief aim at truth. Forgetting, though, aims at nothing.

However, one might put forward principles of minimal change. If I can't help forgetting, I should see to it that I change as little as possible. In the probabilistic context this arguably means minimizing relative entropy (i.e., the Kullback-Leibler divergence). And it is well known that given the posterior $P'(F)$ ($F \in \mathcal{F}^*$) as the constraint that the posterior measure has to satisfy, P' as defined in (11) uniquely minimizes relative entropy. The argument would work for many other measures for the size of change as well.[27]

There is a further argument. According to (11), $P'(A)$ may differ from $P(A)$, while $P'(A \mid F) = P(A \mid F)$ for all $F = \mathcal{F}^*$. Now one might wish that $P'(A)$ does not change. However, given the changed $P'(F)$ this is only possible if (some of) the $P'(A \mid F)$ have changed as well. Why should this not be possible? Can't we also forget about conditional probabilities? Sure we can. However, this means that not only F, but also $A \cap F$ is somehow affected by forgetting. So, A should be included in the algebra of propositions directly affected by forgetting. We might escape this argument by having a more fine-grained conception of the set of propositions directly affected by forgetting; this set might be less than a full subalgebra. In particular, we could try to allow that only some conditional probability is forgotten, that is, takes on some new probability. However, it is already quite contested how to understand *learning* a conditional, that is, how to give some conditional probability a new value.[28] In view of this open problem we should not try to do better on the still more obscure side of change through forgetting. Thus, it seems that we have no choice but (11), as long as we describe forgetting in a coarse-grained way, that is, as directly affecting an entire subalgebra $\mathcal{F}$ of propositions.

Let me briefly sum up what we have achieved so far. We have stated rules for three kinds of pure epistemic change: the familiar conditionalization rules (1) and (2) for learning through experience or information, which imply the special rules (6) and (7) for the case of indexical information; the rules (3) and (4) of time

27. See Diaconis and Zabell (1982: Theorem 6.1) for a proof of the fact that all so-called *f*-divergences will do.

28. This is the infamous Judy Benjamin problem raised by van Fraassen (1981). See, for example, Douven (2012) for some of the intricacies turning up in this topic.

shift, which apply to epistemic change solely induced by the inner sense of time; and the rule (11) of forgetting, which also applies to other local changes of an arational nature.

It is a good question whether extended epistemic change can always be decomposed into several steps of those pure kinds. In Section 3 I had suggested that we can cascade changes through time shift and through information, though perhaps only artificially. Maybe we can do the same with respect to forgetting. I think that the question has a positive answer. If so, we get a more general account of extended epistemic change, insofar as it is rationally assessable, by iteratively applying those three kinds of rules. If the answer should be negative, we can at least deal with the three pure kinds of change.

We will put all of this into an auto-epistemological perspective in Sections 8–10. In particular, the case of forgetting will become much more interesting then. There I will also briefly return to the issue of iterated change. However, let us first study how the machinery developed so far helps with the story of Sleeping Beauty.

7. Third Application: Sleeping Beauty

To begin with, let me once more emphasize that, given the many papers that have been provoked by the apparently small problem of Sleeping Beauty, we cannot be heading for novel arguments or insights. The only goal we can aim to achieve, and will achieve, I hope, is that our more explicit machinery helps to get clearer about existing arguments and about the principles applied in those arguments.

The propositions involved in the problem are: H = heads, T = tails, "*now* = *Su*" = "today is Sunday", "*now* = *Mo*" = "today is Monday", "*now* = *Tu*" = "today is Tuesday", $A_0 = A_{now}$ = "I am awake now", A_1 = "I will be awake tomorrow", A_2 = "I will be awake the day after tomorrow", A_{Mo} = "I am awake on Monday", and A_{Tu} = "I am awake on Tuesday".

Next, I think that no less than five probability measures are involved: first P, Sleeping Beauty's initial probability on Sunday; secondly P^-, her probability briefly before she is awakened (this will be crucial, and therefore we will have to discuss whether it can be legitimately considered at all); then P', her probability immediately after being awakened (where she has no idea whether it is the first time on Monday or possibly the second time on Tuesday and is supposed to be in exactly the same epistemic state at both times); and finally, P_{Mo}, the probability she reaches when being informed that the awakening has taken place on Monday ("*now* = *Mo*"), and, alternatively, P_{Tu}, the probability she would reach when possibly being informed about a Tuesday awakening. What should P' look like? In particular, what is $P'(H)$? That's the issue.

How do those probabilities relate? First, the transition from P to P^- is somehow fishy and results in a loss of time; in P^- (if it is legitimately considered) and also in P' she no longer knows when is now. In any case, we have $P' = P^-(. \mid A_{now})$. In the transition from P^- to P' Sleeping Beauty just learns A_{now}. So, this is a case of simple conditionalization (1) or rather its indexical version (7). Of course, we will have to discuss whether she learns anything thereby; this has been prominently denied. Finally, by definition, $P_{Mo} = P'(. \mid now = Mo)$ and $P_{Tu} = P'(. \mid now = Tu)$. This is again a case of simple conditionalization through which all uncertainty about temporal self-location dissolves.

What do we know about the probabilities? Well, the initial probabilities are clear: we have $P(H) = P(T) = 1/2$, $P(now = Su) = 1$, $P(A_1) = P(A_{Mo}) = 1$, and $P(A_2) = P(A_{Tu}) = 1/2$, since $P(A_{Tu} \mid T) = 1$ and $P(A_{Tu} \mid H) = 0$. Also $P(H \mid A_{Mo}) = 1/2$ and $P(H \mid A_{Tu}) = 0$.

How, though, do we get at P'? We may reason forwards from P through the fishy change to P^-, or we may reason backwards from P_{Mo} and P_{Tu}. Let's try the latter way first. This seems promising because we know that P' is the mixture of P_{Mo} and P_{Tu}, the weights being, respectively, $P'(now = Mo)$ and $P'(now = Tu)$. Both weights are positive, since there is genuine uncertainty in P' as to when is now. we shall see that the weights are not so easily determined.

However, it should be clear what P_{Mo} and P_{Tu} look like. Obviously, $P_{Tu}(H) = 0$. Moreover, Sleeping Beauty seems justified to return in P_{Mo} to her original uncertainty about H and T in P, whatever it has been. In P_{Mo} she has first learned A_{now} and then $now = Mo$. These two changes by simple conditionalization (1) commute. So, learning first $now = Mo$ in P^-, which became uncertain about temporal self-location, is the same as applying the rule (4) of time shift to P moving from the certain $now = Su$ to the certain $now = Mo$. According to (5), the probabilities of eternal propositions are not affected by that shift. So, after learning $now = Mo$, A_{Mo} is as certain as in the initial P, so that learning A_{now} (which is equivalent to A_{Mo} given $now = Mo$) is indeed learning a certainty. (This is not to say, though, that learning A_{now} in P^- would be learning a certainty in P^-.) Hence, H or T are as uncertain as in the initial P, that is, $P_{Mo}(H) = 1/2$.

If we accept this, we can already exclude the halfers' solution. Since P' is a genuine mixture of P_{Mo} and P_{Tu}, we must have $0 < P'(H) < 1/2$. However, Lewis (2001), who triggered the entire debate by contradicting Elga (2000), turns the tables and states that precisely because of this argument its premise $P_{Mo}(H) = 1/2$ must be wrong. His suggestion, as elaborated by Bradley (2011: 403–406), is this: In assuming $P_{Mo}(H) = 1/2$ Elga (2000: 145) apparently relies on the Principal Principle according to which your credence for H should agree with the chance of H—provided, Lewis warns, the information one has in P_{Mo} is admissible in the sense of Lewis (1980). And Bradley argues that it is indeed inadmissible: if the information that now is Tuesday is inadmissible, as it clearly is, then the

information that now is Monday must be inadmissible as well. Moreover, Lewis suggests that the information one has in P' is admissible, because one has no new information at all in P', and hence by the Principal Principle $P'(H) = 1/2$. Bradley leaves it open whether the last step is correct. We will find below that it is incorrect. However, Bradley's challenge is well taken; the thirders cannot rely on Elga's reasoning for their premise that $P_{Mo}(H) = 1/2$.

They might rely instead on the reasoning given above. If that is called in question as well, though, the attempt at reasoning backwards is inconclusive. Therefore we better turn to the forward reasoning. It will further justify the contested premise $P_{Mo}(H) = 1/2$ and provide the missing weights $P'(now = Mo)$ and $P'(now = Tu)$ of the backward reasoning. At the same time the forward reasoning makes the backward reasoning superfluous, although it's mandatory, of course, that they turn out to agree.

Let me first remark, though, that the reference to events like coin tosses which clearly have chances or objective probabilities is a misleading feature of Sleeping Beauty's story. The feature may strengthen the intuition of the halfers: "I know the chance of heads and tails, don't I? And I continue knowing this, even if I forget about other things. So, there is nothing in the story to change this." Thus we run into issues about the Principal Principle, the relation between objective and subjective probabilities, and the hidden ways of admissible and inadmissible information in the sense of Lewis (1980).

However, this is a red herring, for both sides. We can simply abandon the reference to coin tosses and make the second awakening dependent on events H and T, say, something showing up *here* or *there*, for which chances make no sense and for which Sleeping Beauty has merely some initial subjective assessment $P(H) = x$ and $P(T) = 1 - x$. Then the above intuition of the halfers and the issues about objective probabilities simply vanish. Still, the question about $P'(H)$ must have some equally good and justifiable answer. I didn't find that subjectivized case discussed in the literature. Note, by the way, that my above argument in favor of $P_{Mo}(H) = P(H)$ holds in the subjective case as well.

The point also makes an important reasoning in favor of the thirders unattractive. It is initially introduced by Elga (2000: 143–144) and says that in the long run, in an infinite repetition of the experiment, only one third of all awakenings will be accompanied by heads and hence the subjective probability for heads should align with the long run relative frequency. That's a nice argument as far as it goes.[29] However, it has no force vis à vis the subjective variant just presented. Hence we should not rely on it at all. Let's simply stay away from objective probabilities.

29. Arntzenius (2002: 60) uses this argument only for arguing that Sleeping Beauty's betting odds should conform to the thirders, while her degrees of belief might diverge (surprisingly depending on the kind of decision theory she accepts).

And let's turn to the forward reasoning from P via P^- to P'. Is it at all legitimate to refer to P^-, the epistemic state (on Monday or Tuesday) immediately before the awakening? This brings me to another misleading feature of the story, its guise in terms of sleep and awakenings. It invites the false idea that only the epistemic states P on Sunday and P' after the awakening are at issue and no others, since one has no beliefs while asleep. However, there exists also that epistemic state P^- while still asleep, even though it is only dispositional. We do not lose our beliefs during sleep. This point is already emphasized by Weintraub (2004) in her story about flashing lights and by Karlander and Spectre (2010: 404) in their story about the bell, in which the awakening is replaced by the ringing of a bell (and in which all memories of the bell ringing on Monday are erased in case it comes to a bell ringing on Tuesday). In this version, P^- clearly makes sense. You are in some epistemic state P^- immediately before hearing the bell (the same state on Monday and on Tuesday). And, as stated above, the move from P^- to P' is clear: the only change is actually hearing the bell or actually getting awakened, and so $P' = P^-(. \mid A_{now})$.

Thus, let's ponder about P^-. I said that the transition from P to P^- is somehow fishy; Sleeping Beauty's epistemic state is supposed to be the same on Monday and Tuesday and hence she must have lost track of time. Section 3 should enable us to deal with such changes.

However, let me first emphasize that precisely because of that fishy transition I am not impressed by the halfers' intuition already cited in Section 1. It is not even prima facie convincing, even if we (falsely) grant the premise that Sleeping Beauty receives no new evidence by waking up. Lewis argues with this intuition. In his (L1), he outright postulates that "only new evidence, centered or uncentered, produces a change in credence" (2001: 174). But this is simply false. Only new evidence produces a rational change in credence—okay. This does not make changes caused in other ways irrational, but only arational. And we have seen in Sections 3 and 6 ways of responding rationally to such an arational change as (partially) losing track of time or forgetting. It is this rationality that is required of Sleeping Beauty, after having to envisage an arational belief change. Sleeping Beauty may and must change her credence according to other principles than those for learning.[30]

30. Hence, I also see no reason to meet the halfer solution half way, as do Bradley and Leitgeb (2006). They argue that Sleeping Beauty should have subjective probabilities according to the halfers and betting quotients according to the thirders, the latter being convincingly argued by Hitchcock (2004). Their point is that betting quotients need not necessarily correspond to subjective probabilities, because bets may be unfair and are so in Sleeping Beauty's case. Maybe. However, the need to explain an alleged discrepancy between subjective probability and betting quotients in her case arises only because they outright accept the argument that "her credence should be 1/2, because she has learnt no new evidence that is relevant to the coin landing heads" (2006: 121). This is the same false premise as Lewis's.

Let's ask: which beliefs about its now being Monday or, respectively, Tuesday should be contained in P^-? One may say that before awakening Sleeping Beauty has no hint in either direction; her sense of time is mute, and she has no memories of a possible prior awakening. Of course, we might allow her to have hunches (before getting awakened) that it is, say, rather Monday than Tuesday; she may somehow feel that only little time has passed. And we may respect such hunches with our rules (3) and (4) of time shift. Then, however, we have a different story. As the story is told, such hunches are at least implicitly excluded. So, we might plausibly assume:

(12) $P^-(now = Mo) = P^-(now = Tu) = 1/2.$

This is indeed the crucial assumption that will carry us to the thirders' solution. Why accept it?

As far as I know, Karlander and Spectre (2010) were the first to explicitly introduce this assumption. In fact, though, they don't have much of an argument. "Without the waking information"—that is, in our P^-—"a one half prior is warranted" (2010: 406), and then they continue rejecting some considerations that might undermine this warrant. Well, I think that there can't be much of a positive argument. The indifference of P^- between the two possible self-locations is rock-bottom, just a special case of a general symmetry principle or principle of insufficient reason, which is assumed to hold for a priori or initial probabilities. And the story was told in such a way that in P^- Sleeping Beauty is in such an a priori state vis à vis her self-location, because she has no hunches and no information whatsoever about it. There is only a negative absence of reasons for deviating from indifference. The novel aspect of Sleeping Beauty is that such a symmetry principle also applies to self-locating or indexical propositions. Below I will add indirect support for the symmetry assumption (12).

The symmetry (12) may seem similar to the symmetry used by Elga (2000) in his argument. However, there is a crucial difference. Elga assumes a "highly restricted principle of indifference" (2001: 144) entailing that *given T* your P' should assign equal probability to "$now = Mo$" and to "$now = Tu$". We will see that this follows from the more general symmetry (12). In my view, though, the principle of indifference applies to P^- and not to P'. In modified set-ups the change from P^- to P' may well induce asymmetries.

So, let's accept (12) and see where it leads us. First, the law of total probability yields:

(13) $P^-(A_{now}) = P^-(A_{now} \mid now = Mo) \cdot P^-(now = Mo) + P^-(A_{now} \mid now = Tu) \cdot P^-(now = Tu) = 1 \cdot 1/2 + 1/2 \cdot 1/2 = 3/4.$

This is the denominator of the formula (2) given by Karlander and Spectre (2010). Here, we have used $P^{-}(A_{now} \mid now = Mo) = 1$ and $P^{-}(A_{now} \mid now = Tu) = 1/2$, which are beyond doubt.

(13) is a most interesting conclusion; it says that A_{now} is *not* certain in P^{-}. (This would result for any $P^{-}(now = Mo) < 1$.) This is indeed intelligible and not counter-intuitive. As far as P^{-} knows, it may be Tuesday now, and then the awakening is not guaranteed. So, Sleeping Beauty does learn something in P' of which she was uncertain in P^{-} before, as Weintraub (2004) had already convincingly argued.[31]

Let me emphasize the point; it is important to see why the idea that Sleeping Beauty wouldn't learn anything by awakening is wrong. From P^{-} to P' she learns A_{now}; so $P'(A_{now}) = 1$. In which sense, though, did she know A_{now} beforehand? There is no proposition known on Sunday that is equivalent to the proposition A_{now} she learns later on. She knows on Sunday that she will be awakened tomorrow on Monday; we have $P(A_1) = P(A_{Mo}) = 1$. However, her $P'(A_{now}) = 1$ is not the shifted preservation of that previous knowledge, simply because in P' she does not know that now is Monday, that is, what she would have referred to by "tomorrow" on Sunday. She is relieved of her uncertainty about self-location only in P_{Mo}. This is why P_{Mo} should be the same as P regarding eternal propositions, as argued above within the backward reasoning. Similarly, Sleeping Beauty is sure on Sunday to be awakened at some day, that is, $P(A_1 \cup A_2) = P(A_{Mo} \cup A_{Tu}) = 1$. Again, though, $P'(A_{now}) = 1$ is not the preservation of that certainty.

The certainty of A_{now} does not come as a surprise only in the sense that the auto-epistemic proposition "I will know A_{now}" is certain on Sunday and is preserved to be so at the unknown time of P'. Again, though, this auto-epistemic proposition is not the content of the expected and the actual experience. It would be wrong to put this auto-epistemological fact as saying that Sleeping Beauty has learned something she knew before.[32]

Weintraub (2004) has emphasized that in the dialectical situation vis à vis Sleeping Beauty it is also important to convincingly mark the flaws in the opponent's reasoning. The flaw in my view is that it's not the case that there is no rational belief change for Sleeping Beauty on her way to P'. Rather, there *is* arational loss of time from P to P^{-} that requires and has a rational response, and there *is* learning from P^{-} to P'.

31. Karlander, Spectre (2010: 401) argue the same point. However, they argue that Sleeping Beauty, upon awakening on Monday or Tuesday, acquires new *de re* knowledge *of* Monday or, respectively, Tuesday that it is a waking day. I find this phrasing unhappy because it is characteristic of belief *de re* that it might change without being noticed, that is, without any internal change. Still, their point holds since the beliefs *de re* they are referring to actually are just indexical beliefs.

32. Therefore, I do not agree with Bradley (2012: 170), who classifies Sleeping Beauty as a case where one learns something (= A_{now}) certain to be instantiated with a biased procedure, which can therefore not be used for (dis)confirmation. He thereby attempts to support the halfers. However, as just explained, A_{now} is not certain in P^{-}, and hence it can be used to disconfirm H.

Once we have determined $P^{-}(A_{now})$, we might directly proceed to calculating $P'(H) = P^{-}(H \mid A_{now})$. That's the path of Karlander and Spectre (2010). However, let me make a small detour via the calculation of $P'(now = Mo)$, the value of which is interesting in itself:

$$(14) \qquad P'(now = Mo) = P^{-}(\, now = Mo \mid A_{now})$$

$$= P^{-}(A_{now} \mid now = Mo) \cdot P^{-}(now = Mo) \,/\, P^{-}(A_{now}) = 1 \cdot 1/2 \,/\, 3/4 = 2/3.$$

This follows from (12) and (13) via our indexical learning rule (6) (which was just simple conditionalization applied to indexical propositions).

I would like to suggest that this conclusion is very reasonable independently of its derivation. Within P', after being awakened, Sleeping Beauty might or should reason as follows: "I have no idea when is now; it's either Monday or Tuesday. I am awake now. Being awake on Monday (= being awake now when now is Monday) is twice as probable as being awake on Tuesday (= being awake now when now is Tuesday). Hence, now being Monday is twice as probable as now being Tuesday. Of course, it's now either Monday or Tuesday. Hence, $P'(now = Mo) = 2/3$ and $P'(now = Tu) = 1/3$."

Let me underscore the plausibility of this argument by an apparently quite different, but structurally identical example. Suppose six experienced mountaineers $b_1, \ldots, b_6$ start a very dangerous expedition. You are not sure whether they will return alive. Let's assume your probability that b_i will return is x_i. The numbers may differ, since you may have varying trust in the skills of the mountaineers. Of course, you may also have some positive probability for any two or all of them to return. So, $\Sigma\, x_i$ may well be > 1 (but it is ≤ 6). Now, after days of anxious waiting, you see the first one returning. In their suits the mountaineers all look the same. You have no further cue at all who this mountaineer, this b you see, might be. Your only cue consists in your prior return probabilities. So, what should your probability be that this b is, say, b_1? By the same reasoning as above you should argue, it is x_1/x_i as likely that b is b_1 rather than b_i, and so for all the other comparisons. Hence, the probability that this b is b_1 should be $x_1 / \Sigma\, x_i$. This seems highly plausible to me. Similarly, in the subjective variant of Sleeping Beauty: if $P(H) = x$, we have $P(A_{Mo}) = 1$ and $P(A_{Tu}) = P(T) = 1 - x$, and hence $P'(now = Mo) = 1 / (2 - x)$.[33]

33. White (2006) also plays around with the waking probabilities. He replaces her awakenings by the activation of a waking device that wakes her up with probability c each time. However, he assumes that P' is the conditionalization of P^{-} by W, where W = "Sleeping Beauty is awake at least once during the experiment" (= $A_1 \cup A_2$ in our notation). This is not her evidence, though. Her evidence is A_{now}. Then we have $P^{-}(A_{Mo}) = c$ and $P^{-}(A_{Tu}) = c/2$, and hence $P'(now = Mo) = 2/3$ and $P'(H) = 1/3$ according to the principle just introduced. So, we arrive at a confirmation of the argument,

The general rule behind these inferences may be stated in the following somewhat sloppy way indicating its applicability beyond our specific framework:

(15) Let $E_1, \ldots, E_n$ be any n eternal propositions with probabilities $Q(E_i) = x_i$. Let $this = i$ ($i = 1, \ldots, n$) be mutually disjoint and exhaustive indexical or demonstrative propositions about which Q has no evidence or information, and let $E(this)$ be another indexical proposition such that $Q(E(this)$ iff $E_i \mid this = i) = 1$. Then $Q(this = i \mid E(this)) = x_i / \Sigma_j x_j$ for $i = 1, \ldots, n$.

Note that the probability specified in (15) is *not* the probability of E_i given that exactly or at least one of the E_j is true. I see no way to reduce it to (conditional) probabilities of eternal propositions. Rather this rule infers probabilities for self-locating or, more generally, indexical propositions (like $now = Mo$ given A_{now} or "this b is b_1" given "this b returns") from probabilities for eternal propositions—provided there is no other information about those indexical propositions. As such it is something unprecedented; up to now we had no such connection between eternal and indexical propositions. One might also call the novel rule (15) a *rule for direct probabilities of self-locating propositions*.

The proviso that Q provides a priori probabilities for the indexical propositions $this = i$ by having no evidence about them is somewhat indeterminate. One might think that it is best explicated by a symmetric distribution over them, that is, by $Q(this = i) = 1/n$. Given the assumption that the proposition $this = i$ is probabilistically independent from the eternal proposition E_i, this symmetry is indeed equivalent to (15). In this perspective, (15) is not a new principle, but simply derived from the basic symmetry. However, this is not my perspective. I sense an intuitive difficulty in having a priori probabilities for these indexical propositions $this = i$, since there is no a priori reference for "this". *This* is given with something like an act of demonstration already providing information about *this*. Therefore, I emphasize the intuitive plausibility of (15) independent of an underlying symmetry assumption. (15) may be derived from the symmetry, but it can reversely be used to justify the symmetry.

So, we have two ways to arrive at $P'(now = Mo) = 2/3$. We can directly apply the rule (15). Or we can reason from the indifference principle (12) via (13) and (14). My suggestion is to take the two ways to mutually support each other. This finally entails:

(16) $P'(H) = P^-(H \mid A_{now}) = P^-(H \mid A_{Mo}) \cdot P'(now = Mo) + P^-(H \mid A_{Tu}) \cdot P'(now = Tu)$
$= 1/2 \cdot P'(now = Mo) = 1/3$.[34]

which he attributes to Elga (2000) and to Arntzenius (2003) and Dorr (2002) and which he wanted to attack with his variation.

34. This is equation (11) of Schulz (2010), from which he rightly concludes that the halfers could only be right if $P'(now = Mo) = 1$ and are hence wrong. However, unlike Karlander and Spec-

This is the thirders' solution, of course. Let me make clear where we have used the principles developed in the previous sections. The first equation of (16) is an application of our principle (7), which again is only conditionalization (1) as applied to indexical propositions. And the second equation follows from the assumptions $P^{-}(H \mid A_{Mo}) = 1/2$ and $P^{-}(H \mid A_{Tu}) = 0$, which are beyond doubt. In P^{-} and in P' Sleeping Beauty has lost track of time. According to the observation (5), however, the attitude towards eternal propositions is not affected and the same for P, P^{-}, and P'.

To resume, all we need is the a priori symmetry (12) in P^{-} or, via the rule (15), the relevant distribution in P'. Then the forward reasoning with the rule (3) and (4) of time shift (entailing (5)) and with the indexical conditionalization rules (6) and (7) carries us to the thirders' solution. I have made explicit where we have used those rules. Recall, finally, that the assumption $P_{Mo}(H) = 1/2$ would have sufficed as well to derive the thirders' solution via backwards reasoning. However, it seemed not cogently justified and was indeed doubted by Lewis (2001). Now we can see that it follows as well, since we have $P_{Mo}(H) = P'(H \mid now = Mo) = 1/2$.

So much about Sleeping Beauty. Let me repeat, though, that the point of carefully going through the exercise was not another rebuttal of the halfers and affirmation of the thirders. It was rather to analyze where precisely the non-standard principles of epistemology developed in the previous sections are required for reasoning about Sleeping Beauty.

Still, one might object that this is not a proper case of loss of time; Sleeping Beauty does not literally lose track. Rather, in P^{-} and in P' she does not recall to have awakened before; indeed she thinks she recalls *not* having been woken up before. Hence, one might think that she could be sure that it is Monday. (She may be unsure how many hours she has slept, but she is not unsure how many days she has slept.) And so there is no subjective loss of time. This is what Hawley (2013) argues.[35]

This reasoning seems correct, as far as it goes. Sleeping Beauty's uncertainty about temporal self-location is not brought about by her unreliable sense of time. It is rather generated by an auto-epistemological consideration. The loss of time is due to the information that her memories of the first awakening will be erased before the potential second awakening. Thus, whatever her memory tells her, she knew on Sunday (and has not forgotten later on) that this would be exactly the same whether her P^{-} is located on Monday or on Tuesday. This is what pro-

tre (2010) he does not go on specifying $P'(now = Mo)$.

35. Hawley does so on the basis of what he calls the inertia principle. This applies in case Sleeping Beauty neither gains nor loses relevant evidence when she awakes. He explains that the best arguments that she gains or loses evidence are unconvincing and thus applies his principle. He does not address, though, the reasons given above for her doing indeed both.

duces the indifference (12) of P^- or what makes the a priori principle (15) applicable to P'.

This informal appeal to auto-epistemology raises the need to formally study the auto-epistemological extensions of the fields of non-standard epistemology we have considered so far. So, let me turn to this study in the second part of this paper. As before, Sleeping Beauty is only the trigger of those extensions, which are important and interesting by themselves. These extensions have in fact scarcely been treated in the literature. Of course, these extensions build on standard auto-epistemology. This is why we need to briefly rehearse the basics of the standard account.

8. Standard Auto-Epistemology

Auto-epistemology deals with one's beliefs about one's future, present and past beliefs, or more generally, with how one represents one's future, present and past doxastic states in one's present doxastic state. We have beliefs about many parts of the world. We are also part of the world, indeed a part we are very well acquainted with. So, of course, our own mental and especially our doxastic states are an important topic of our theorizing. This makes auto-epistemology an important part of epistemology. It has one central principle, the reflection principle; I think there is no further independent principle of auto-epistemology in the literature. As mentioned, it has been studied mostly within the standard framework of eternal propositions. This is what we have to look at next.[36]

We still only talk about one doxastic subject, which we need not make explicit in the notation. And we still need to consider her doxastic states P and P' only at two real times τ and τ'. This unburdens the notation. We will discuss whether this suffices for studying longer doxastic evolutions. Moving within a standard framework means conceiving P and P' to be about the algebra $\mathcal{W}$ of eternal propositions and not about the indexically extended algebra $\mathcal{A}$. This restriction will hold throughout this section.

However, we now have to take the auto-epistemic extension of this framework into account, that is, not only the actual doxastic states P and P' at τ and τ', but also the possible states of our subject at those times. Therefore, let Π and Π' be the set of the subject's possible probability measures at, respectively, τ and τ'; I will use π and π', respectively, as variables ranging over these sets. Note that those possible measures are assumed to be only about ordinary eternal propositions in W; they are not auto-epistemic. Now we can describe a lot of auto-epistemic propositions: $\{\pi \mid \pi(A) = x\}$ is the proposition that her probability for

36. Hild (1998a) will be my reference text for this section.

A at time τ is x, $\{\pi' \mid \pi' = Q'\}$ is the proposition that her probability measure at τ' is Q', etc. As usual, I will slightly abbreviate these expressions as $\{\pi(A) = x\}$, $\{\pi' = Q'\}$, etc. So, the epistemic possibilities to be considered are not (small) worlds in W, but triples of a world and a prior and a posterior measure in $W^* = W \times \Pi \times \Pi'$. Let $\mathcal{W}^*$ denote the algebra over W^* generated by $\mathcal{W}$ and all the above auto-epistemic propositions.

Auto-epistemology then requires all these auto-epistemic propositions to be in the domain of the subject's actual doxastic states. That is, the actual states P and P' are not only about $\mathcal{W}$, but in fact about the whole of $\mathcal{W}^*$. However, as long as we consider only one temporal step, we need not complicate things by allowing the reflected doxastic states to be auto-epistemic in turn; they may well be restricted to $\mathcal{W}$. Let us call the whole construction resulting in $\mathcal{W}^*$ and the measures P and P' an *auto-epistemic set-up*.

Given a probabilistic auto-epistemic set-up, the *reflection principle* introduced by van Fraassen (1984) is usually presented as the following probabilistic principle: for all $A \in \mathcal{W}$ $P(A \mid \{\pi'(A) = x\}) = x$. Here I will only refer to the slightly stronger, but equally intended version:

(17) $\quad P(. \mid \{\pi' = Q'\}) = Q'$.[37]

The weaker version says that, given your future probability of $A \in \mathcal{W}$ is x, your present probability of A should be x, too. The slightly stronger version (17) extends this to your entire future probability measure. More colloquially, the principle says that you should now trust your future judgment. In this way the principle is a dynamic one. A qualitative version of this is Binkley's (1968) principle: if you now believe that you will believe A tomorrow, you should believe A already now (which he applies in order to explain the surprise examination paradox). Gaifman (1988) has introduced a more general perspective. He thinks of π' not as your own future probability measure, but of anybody's at any time. And he defines such a π' to be an *expert function* for you at τ just in case your P at τ satisfies (17). In this reading, (17) says that you should trust the experts. However, it is true by definition, since an expert is defined for you as one you trust. Hájek (2003: 192–195) points to difficulties arising when you accept several experts on the same issue.

In this terminology, the reflection principle (17) says that you should consider your future self as an expert. This makes clear that the reflection principle cannot be a universal maxim. Given that you will have forgotten things you should not trust in your forgetful state; given that your usual clear-headedness

37. Let's ignore the possibility that $P(\{\pi'(A) = x\}) = 0$ or $P(\{\pi' = Q'\}) = 0$, in which case these conditional probabilities are undefined. See Goldstein (1983) and Gaifman (1988) for how to avoid this problem.

will be clouded by (too much) alcohol, you should get sober again and not let your sober self be guided by your drunken self. This is a familiar objection (see, e.g., Maher 1993: Section 5.1), and the reflection principle needs to be restricted accordingly. We shall remove those restrictions in Section 10.

Despite such restrictions one must realize the fundamental importance of the reflection principle. The usual case is that one's future doxastic state is better informed than the present one; one learns and improves one's point of view in the course of time. Unlike the conditionalization rules (1) and (2) the reflection principle (17) as such does not specify any details about how exactly this learning is to work; it only says that the improvements must be such as to conform to (17). This looks like a very abstract constraint, but it has specific consequences.

Spohn (1978: 161–162) and Goldstein (1983) endorsed another principle called the *iteration principle* by Hild (1998a):

$$(18) \qquad P = \sum_{\pi' \in \Pi'} \pi' \cdot P(\pi') \,.$$

The iteration principle says that your present opinion is a weighted mixture of your possible future opinions, the weights being your present probabilities for arriving at the various possible future opinions. It is easily shown to be equivalent to the reflection principle (see Hild 1998a: Theorem 2.2). Hence, it is subject to the same restrictions, as already noted in Spohn (1978: 166).

These principles have quite a few precursors. Actually, I think that the iteration principle and the reflection principle are obviously suggested by de Finetti's representation theorem and his philosophy of probability (see de Finetti 1964). Still, the latter received its name and its first extensive philosophical discussion only by van Fraassen (1984), which thus serves as the common reference text.

Hild (1998a: 328) refers the reflection principle also to your present probability measure. Your present opinion is trivially an expert for your present opinion, presently you do not know better than you presently know:

$$(19) \qquad P(. \mid \{\pi = Q\}) = Q \text{ and } P'(. \mid \{\pi' = Q'\}) = Q'.$$

Hild (1998a) then shows (in his Theorem 2.1) that the reflection principles (17) and (19) together entail *auto-epistemic transparency* at τ:

$$(20) \qquad P(\pi = Q\}) = \left\{ \begin{array}{l} 1, \text{ if } Q = P \text{ restricted to } \mathcal{W} \\ 0, \text{ otherwise} \end{array} \right\}, \text{ and similarly for the posterior .}$$

Proof: One instance of (19) is $P(\{\pi = Q\} \mid \{\pi = P\}) = P(\{\pi = Q\})$. And this is 1 if $Q = P$ and 0 if $Q \neq P$.

(17) and (19) also entail *perfect memory* (with a little caveat):

$$(21) \qquad (\{\pi = Q\}) = \begin{Bmatrix} 1, \text{ if } Q = P \text{ restricted to } \mathcal{W} \\ 0, \text{ otherwise} \end{Bmatrix}.$$

Proof: If your prior P is auto-epistemically transparent, as (20) asserts, and if your prior P is also a mixture of your possible posterior expert opinions according to (18), then your possible posteriors π' and thus also your actual posterior P' can't fail to be sure about your prior opinion. The caveat here is that this reasoning has tacitly assumed that the possible π' are also about auto-epistemic propositions in W* and not only about ordinary eternal propositions in $\mathcal{W}$.

Hence, (21), unlike (18)–(20), requires a slight extension of our above set-up. (21) also highlights the restricted use of the reflection principle (17). Memory loss is simply not among the circumstances to which it applies.

How does learning or updating work in the auto-epistemological perspective? Let us continue to follow Hild (1998a). The crucial notion is that of a *protocol*, which was forcefully reintroduced by Shafer (1985). As suggested by Shafer (1985: 266) with the notion of a *subjective protocol*, Hild puts it into an auto-epistemological perspective. Let me explain:

We assume the auto-epistemic reasoner to envisage a further set Γ of possible pieces γ of total evidence (acquired between τ and τ'). According to simple conditionalization (1) Γ may be a set of propositions from $\mathcal{W}$. Let E^Γ denote the piece of total evidence the content of which is E. So E^Γ is a value of Γ, while E is a proposition in $\mathcal{W}$. According to generalized Jeffrey conditionalization (2), though, Γ rather consists of pairs $\langle \mathcal{E}, Q \rangle$, where $\mathcal{E}$ is some (evidential) partition of W and Q a probability distribution over $\mathcal{E}$. Again, let $\langle \mathcal{E}, Q \rangle^\Gamma$ denote the value of Γ corresponding to the pair $\langle \mathcal{E}, Q \rangle$. Maybe Γ has yet another structure. We may leave this open.

Thus, Γ provides another auto-epistemic extension. Now let $W^* = W \times \Pi \times \Pi' \times \Gamma$, let $\mathcal{W}^*$ denote the algebra over W^* generated by $\mathcal{W}$, the above auto-epistemic propositions, and the new propositions about Γ, and assume that P and P' distribute over that extended algebra $\mathcal{W}^*$, while all the values π and π' in Π and Π' are still restricted to $\mathcal{W}$. So we have further auto-epistemic propositions of the form $\{\langle w, \pi, \pi', \gamma \rangle \mid \gamma = \gamma^*\}$ or $\{\gamma = \gamma^*\}$ or even γ^* for short. Note again, if γ has the form E^Γ, $P(E)$ and $P(E^\Gamma) = P(\gamma = E^\Gamma)$ are two different probabilities. $P(E)$ is the prior probability for E to obtain, while $P(E^\Gamma)$ is the auto-epistemic prior probability of receiving total evidence with content E.

An *objective protocol* would be any function from W into Γ telling which total evidence γ in Γ the subject receives in the world $w \in W$. And a *subjective protocol* is just what the subject believes about the objective protocol. It is contained in her prior probability measure P that specifies how the auto-epistemic proposi-

tions about Γ are probabilistically related to eternal propositions in $\mathcal{W}$. The naïve expectation, which fortunately is usually satisfied, is that, when *E* happens, we see *E* and hence come to believe (or know) *E*. However, each of us has rich experience with many exceptions and thus a sophisticated subjective protocol. This is indeed what biased evidence is all about.

How should the epistemic subject respond to some piece of total evidence? The most abstract answer is that she should conform to some update rule *u* that assigns to any prior probability measure π on W and each piece γ of total evidence a posterior probability measure $\pi' = u(\pi, \gamma) = \pi^{u,\gamma}$ on W. In particular, if the subject actually obeys the update rule *u* and receives evidence γ, then $P' = P^{u,\gamma}$ (restricted to $\mathcal{W}$). So, the subject obeys the update rule *u* and expects or considers only probabilistic changes that are driven by evidence according to that rule, if and only if

(22) $P(R) = 1$, where $R = \{\langle w, \pi, \pi', \gamma\rangle \in W^* \mid \pi' = u(\pi, \gamma) = \pi^{u,\gamma}\}$.

Let us call an auto-epistemic set-up *evidence-driven* iff it satisfies (22).

Now one obvious update rule is the rule *u* according to which the subject simply conditionalizes on the auto-epistemic proposition that she has received total evidence *e*. Hild (1998a: 332) calls this the *rule of auto-epistemic conditionalization*:

(23) $P' = P^{u,\gamma} = P(. \mid \gamma)$.

Theorem 2.3 of Hild (1998a) then says:

(24) For evidence-driven auto-epistemic set-ups satisfying (22) auto-epistemic conditionalization (23) is equivalent to the reflection principle (17).

In other words, when the only probabilistic changes envisaged are those driven by evidence according to a certain update rule, the only update rule compatible with, and indeed tantamount to, the reflection principle (17) is auto-epistemic conditionalization (23). Hence, if we accept the reflection principle within the auto-epistemic perspective, updating reduces to a very special case of simple conditionalization or, in effect, of supposing as specified in (23). Suppose you were to learn γ: that is exactly the doxastic state you arrive at when you actually learn γ.

By all means, auto-epistemic conditionalization (23) must not be confused with simple conditionalization (1) or Jeffrey conditionalization (2). Hild (1998b) argues that they may even come into conflict, in which case (1) and (2) have to give way. They agree only under special conditions, which may be expected, but

are not guaranteed to hold. Evidence may be biased. I take this to be a fundamental lesson: in an auto-epistemological perspective, (23) is the basic probabilistic learning rule and not (1) or (2). For further details I refer to Hild (1998b).

9. Indexical Auto-Epistemology

So much about auto-epistemology regarding learning about eternal propositions in $\mathcal{W}$. Let us see now how this extends to the topics discussed in Sections 3 and 6, indexical propositions and forgetting and other arational changes, and hence to the full algebra $\mathcal{A}$. Let's first deal with indexical propositions; forgetting will be the topic of the next section. We will see that the indexical extension runs in a smooth, almost automatic way. We are not heading for exciting news. So, let me be as brief as possible without sacrificing the aim of stating the matter at least once explicitly.

It is obvious that the reflection principle (17) cannot survive in an unmodified way within this extended setting. Given that tomorrow I will be sure of the proposition "today is Friday", I am sure today, not of that proposition, but rather of the proposition "tomorrow is Friday" (which is the same as "today is Thursday"). However, the reflection principle is easily adjusted. Since we want to allow for uncertainty about temporal self-location, the subject should refer to her possible future assessment not in an eternal, but in an indexical way, as being in some distance $z \geq 0$ from now.

So, let P again be the subject's present probability measure at τ, where she need not know when τ is; possibly, she can refer to τ only by "now". Let P_z, for some $z \geq 0$, be her doxastic state at $\tau + z$, that is, z units of time later. Auto-epistemically, she has to consider her doxastic state at some fixed interval z later (though she may consider the case that she does not know then that it is z times later). And let Π_z be the set of her possible probability measures about the basic algebra $\mathcal{A}$ over $T \times W$ of eternal, self-locating and mixed propositions at $\tau + z$, that is, z units of time later. So, Π_0 denotes the set of her possible doxastic states at τ. Hence, the set of epistemic possibilities now to be considered is not $T \times W$, but rather $T \times W^*$, where $W^* = W \times \Pi_0 \times \Pi_z$ includes the auto-epistemic components. Let $\mathcal{A}^*$ be the algebra of propositions over $T \times W^*$ that is generated by the old $\mathcal{A}$ and the new auto-epistemological propositions of the form $\{\pi_z(A) = x\} = \{\langle t, w, \pi_0, \pi_z\rangle \mid \pi_z(A) = x\}$, $\{\pi_z = Q\} = \{\langle t, w, \pi_0, \pi_z\rangle \mid \pi_z = Q\}$, etc. As before, we assume that the subject's actual probability measures P and P_z apply to the whole of $\mathcal{A}^*$, in contrast to all the possible measures in Π_0 and Π_z, which apply only to $\mathcal{A}$. Then the appropriate modification of the reflection principle (17) in its stronger form is as follows:

(25) $P(. \mid \{\pi_z = Q^i_z\}) = Q$, where Q^i_z is defined by $Q^i_z\ (A) = Q(A^i_z)$ for all $A \in \mathcal{A}$.

Let's call (25) the *indexical reflection principle*. Obviously, for $z = 0$ (25) embraces the special case (19) of the reflection principle. If we apply (25) to the above example about Friday, we see that it fits perfectly.[38]

We must again ponder the restrictions applying to (25). As far as eternal propositions are concerned, the restrictions are clearly the same as those of the original reflection principle (17), even though they were only vaguely described. How do they carry over to the more general case? The guiding idea was that the subject must be able to see her future point of view, whatever it is, as somehow superior, as better informed or having a better expertise (or at least as equally good) as the present one. We have seen that this idea is fulfilled if the epistemic change is due to learning via simple or generalized conditionalization (1) and (2). But how about epistemic changes involving temporal self-location?

Let's first attend to the core rule of time shift (3). It seems clear to me that the idea guiding the reflection principle is not satisfied thereby. According to that rule I add an incremental uncertainty about my temporal self-location, as measured by some p'. So I don't learn about my self-location, I am rather incrementally losing track. The sense of time does not perfectly follow the objective passage of time. This is not a case of learning. Neither is it a case of forgetting, but it is similar.

This assessment is confirmed by looking at (25). Suppose that $P(now = t) = 1$; I am sure that now is t. Suppose further that I fear losing track of time so that for some π_z, $\pi_z(now = t + z) < 1/2$; that is, z times later I may be very unsure that now is $t + z$. However, given this fear I am not now less sure about what time it is now. Thus, (25) does not hold in this case. I take this to be a telling reason for denying that the inner sense of time provides proper information, even if it generates epistemic change. If the reflection principle is about improving one's epistemic position, such improvement is not provided by that sense. This is the ultimate reason for treating problems of self-location and problems about forgetting jointly in this paper.

Could it be that the incremental uncertainty about self-location cancels the initial uncertainty and thus creates a fake certainty about self-location? No, this is not possible, given the assumption that initial and incremental uncertainty are independent. There is no way of getting better informed about self-location merely through the inner sense of time.

If the reflection principle is inapplicable to the core rule (3) of time shift, this holds all the more for the general rule (4) of time shift. Hence, we must exclude

38. Schwarz (2012: 224–225) presents a similar generalization called shifted reflection in terms of his shifting operator "next". This is the only place I know where indexical auto-epistemology has been discussed.

belief change through the inner sense of time from the range of application of (25). Rather, the indexical reflection principle (25) presupposes a perfect sense of time, in the sense that all the possible measures π_z are certain that they are located z times later than the actual P.

This entails that we may generalize the iteration principle (18) to the following *indexical iteration principle*:

(26) $$P_z^i = \sum_{\pi_z \in \Pi_z} \pi_z \cdot P(\pi_z), \text{ where } P_z^i \text{ is defined by } P_z^i(A) = P(A_z^i) \text{ for all } A \in \mathcal{A}.$$

Like the principle (18), (26) says that your present opinion is a weighted mixture of your possible future opinions, the weights being your present probabilities for arriving at those future opinions. However, it now applies to all propositions. For example, you are unsure what time it is, and you expect to learn it with certainty by looking at your watch. Then your probabilities for what you will learn are precisely given by your uncertainty about which time it is. Note that in (26) z is some fixed interval of time; it is not variable and not uncertain. The fact that z is fixed and the same for all possible future opinions reflects the observation above that the π_z do not contain any incremental uncertainty about self-location.

Again, we may prove that the indexical versions of the reflection principle (25) and the iteration principle (26) are equivalent. Hence, they stand under the same informal restrictions. And again, we may prove that (25), which we noticed to comprise (19), entails indexical versions of *auto-epistemic transparency* (20) and *perfect memory* (21) (with the caveat mentioned there). I omit the proofs, since they seem to be only notational variations of the proofs provided by Hild (1998a) for the standard case.

I have argued that epistemic change due to mere time shift does not conform to the reflection principle (25). However, as is also emphasized by the iteration principle (26), it should hold for all cases of learning, of receiving any kind of (indexical) information. This is still clearer when we study how the account of auto-epistemic learning or updating given in the previous section generalizes to our extended setting.

In fact, as far as I see, this account immediately carries over without hardly any change. We can again assume a set Γ of possible pieces γ of total evidence (acquired between τ and $\tau + z$). It should only be added that in the subject's perception it takes exactly z units of time from the prior time τ to the posterior time $\tau + z$. And the definition (22) of an evidence-driven auto-epistemic set-up can be maintained as well, except that π' should rather be denoted as π_z. Like (25) and (26), *indexical auto-epistemic conditionalization* should take the indexical shift of propositions into account. Thus auto-epistemic conditionalization (23) should be modified as follows

(27) for all $A \in \mathcal{A}\ P_z(A)\ P^{u,\gamma}(A^i_z) = P(A^i_z \mid \gamma)$.

Theorem (24) then also holds with respect to the indexical versions (25) and (27). (Again, the proof is essentially identical to that of (24).)

So much for indexical auto-epistemology. Let me only add that the warning that auto-epistemological conditionalization (23) trumps ordinary conditionalization (1) and (2) applies to the indexical case as well.

10. Auto-Epistemology Generalized

Let us finally turn to our second non-standard extension, the auto-epistemology of forgetting. Thereby we will be entering uncharted water. It is clear that forgetting still stands *pars pro toto* for any kind of unfavorable doxastic change through drugs, brain washing, etc. The label, though, appears to be an oxymoron. How can there be an auto-epistemology of forgetting? Forgetting befalls us; there seems to be nothing to manage reflectively.

Well, not so. We can ask, and answer, the same questions as does the reflection principle for its intended applications. Do we trust our future epistemic states, and what does this mean and entail? In the case of forgetting it seems clear what to say: we do *not* trust in our future states. Today, I am sure that I have a date with my dentist next week. I might well have forgotten (or repressed) it by tomorrow. So, given that tomorrow I have forgotten it and believe I don't have a date, should I do so now as well, as the reflection principle would have it? Of course not. Even given the future forgetting I should stick to my present beliefs. And maybe I should do something to minimize the danger of forgetting, say, make a notice.

Formally, this means that for all $A \in \mathcal{A}\ P(A^i_z \mid \{\pi_z(A) = x\} = P(A^i_z)$, in case π_z is reached from P through forgetting; that is, in this case I dismiss the wisdom or rather the ignorance of my future state and stick to my present assessment. How, though, can we state the new principle? It could be saved from plainly contradicting the reflection principle only by vaguely claiming a disjoint range of application.

I want to suggest that we should not state the conditions restricting these principles as external conditions. We should rather internalize them and state them as subjective conditions set by the subject herself. It's not we who decide whether an epistemic change is a favorable one towards a better informed state or an unfavorable one towards a poorer state. In the first place, it's the subject herself who should tell; she should make her conditional judgment dependent on the kind of change she expects. Of course, she may be in error and take as evidence what is actually a hallucination. We may then criticize and tell her

that she shouldn't take the hallucination as evidence, and she may concur. Still, the primary standard for an epistemic rationality assessment is her subjective condition.

So, the subject has to assess the quality of her potential epistemic changes. Does the change improve my epistemic position? Or does it worsen it? As far as I see, this coarse assessment is all that it is required for auto-epistemological purposes. The precise kind of change need not be detailed. Normally, we would say that experience and information improve the position. Some persons might include epiphanies and gut feelings. I have sufficiently exemplified worsenings of the position. The additional insight of the previous section was that merely relying on the inner sense of time should count as a worsening as well. However this may be, what counts is only the classification as a favorable or unfavorable change.

In order to represent the subject's assessment, we have to introduce a further auto-epistemic variable Θ that describes the kind of epistemic change moving the subject from Π_0 to Π_z. The possible values $\langle \pi_0, \pi_z, \theta \rangle$ of Θ represent not only the numerical change from π_0 to π_z, but also the way θ in which this change comes about; so θ specifies something like learning with or without certainty and the pertaining piece of total evidence Γ, or forgetting, which may come in various kinds, or the quantity of alcohol influencing the epistemic state, etc. Let $(\pi_0 \to \pi_z)$ denote the set of numerical changes from π_0 to π_z in any way θ whatsoever. We may also use $(\pi_0 \to \pi_z)$ to represent the proposition $\{\Theta \in (\pi_0 \to \pi_z)\}$ that the numerical change from π_0 to π_z comes about in some way or other.

Now I have suggested that the crucial parameter for the auto-epistemological reflection is not the specific way θ of epistemic change, but only whether the subject perceives that change as an improvement or a worsening of her epistemic position. Hence, let's distinguish two kinds of changes: $(\pi_0 \uparrow \pi_z) = \{\langle \pi_0, \pi_z, \theta \rangle \mid$ the change from π_0 to π_z in the way θ is an improvement$\}$ and $(\pi_0 \downarrow \pi_z) = \{\langle \pi_0, \pi_z, \theta \rangle \mid$ the change from π_0 to π_z in the way θ is a worsening$\}$, so that $\{\Theta \in (\pi_0 \uparrow \pi_z)\}$ and $\{\Theta \in (\pi_0 \downarrow \pi_z)\}$ represent the corresponding propositions (which may again be abbreviated as $(\pi_0 \uparrow \pi_z)$ and $(\pi_0 \downarrow \pi_z)$). So, we are dealing now with the auto-epistemic extension $W^* = W \times \Pi_0 \times \Pi_z \times \Theta$, and the set of epistemic possibilities to be considered is $T \times W^*$, with $\mathcal{A}^*$ being the pertinent algebra over $T \times W^*$.

Do $\uparrow$ and $\downarrow$ generate an exhaustive disjunction? Yes, almost. The only exception I can think of is that there is no change at all so that $\pi_z = \pi_0$. However, we may decide this case *per fiat*. We can say it's a limiting case either of an improvement or of a worsening; this won't make a difference. What I cannot think of is a genuine change that is not evaluated either way. Could I undergo an epistemic change and say afterwards that the posterior opinion is just as fine as the prior one, neither better nor worse? I don't care which one I have? I think that my prior stance should be conservative and judge such a change to be willful, arbitrary,

and an unjustified worsening of my present point of view.[39] *There can't be any neutrality*. Therefore we seem justified in assuming the *no neutrality condition*:

(28) $(\pi_0 \uparrow \pi_z) \cap (\pi_0 \downarrow \pi_z) = |$ and $(\pi_0 \uparrow \pi_z) \cup (\pi_0 \downarrow \pi_z) = (\pi_0 \rightarrow \pi_z)$.

(28) includes the assumption that for any way θ of change from π_0 to π_z the subject has a determinate assessment of θ as favorable or unfavorable. This is not to say, though, that the subject may not be uncertain which kind θ of change she will undergo.

Thus we are finally prepared to state the following

(29) *Full Reflection Principle*: Let P_0 be the restriction of P to the non-auto-epistemological part $\mathcal{A}$. Then

$P(. \mid \{\pi_z = Q_z^i\} \cap (P_0 \uparrow \pi_z)) = Q$, and

$P(. \mid \{\pi_z = Q_z^i\} \cap (P_0 \downarrow \pi_z)) = P_0$,

where Q_z^i is defined as in (25), that is, by $Q_z^i(A) = Q(A_z^i)$ for all $A \in \mathcal{A}$.

The first conjunct is the original reflection principle (17) or rather its indexical version (25); here, the future opinion is accepted as an expert. By contrast, the second conjunct realizes my above proposal for how to treat expected forgetting or other impairments of one's epistemic situation: namely to ignore them. (29) makes clear, by the way, that it does not matter how we classify the limiting case of no change at all; in this case the alternatives come to the same.

I should emphasize that the consistency and the completeness of (29) depends on the assumption (28). I should also stress that the full reflection principle (29) holds absolutely; it is not subject to any restrictions or conditions. We may put to one side the old discussion about such restrictions, since we have internalized or subjectivized them. Of course, the discussion reappears when we want to develop standards for criticizing the subjective views. But this does not concern the principle (29), which is thus perfectly general.

And let me emphasize once more the crucial consequence of the no neutrality condition (28). One might say that compromising is ubiquitous. Whenever there are two points of view, maybe each one is partially right, and theory should provide ways for weighing and compromising. Not so in (29) due to (28). There

39. At least, this holds under our basic assumption that doxastic states are represented by determinate probability measures. If subjects are unsure about their own subjective probabilities so that their state is better represented by a lower probability, or a convex set of probability measures, things may be different. Then, perhaps, there is change in indeterminate probabilities that is neither to the better nor to the worse. However, then the presuppositions of our discussion would be completely different ones.

is no compromise between the present and the future point of view. The future perspective is either better or worse (or identical) and hence either accepted or rejected by the present perspective; there is no half-reliance on the future perspective. This is an important feature of auto-epistemology.

I just claimed that the full reflection principle (29) is perfectly general. It is so in comparison with the original reflection principle (17), which is constrained by external conditions. However, generality seems imperfect in another way. Couldn't there be a change to the better and to the worse at the same time so that the result is incomparable? For example, I learn that I have a date with the president and, embarrassingly, forget at once that I already arranged a date with the vice-president at the same time. Such cases must not be denied. My response is the same as in Section 3. We should distinguish two changes in such cases, one to the worse and one to the better, even though the temporal separation of the changes may be artificial in real situations. Thereby we can stick to the conclusion that each single epistemic change is uniquely evaluated, even if a chain of changes may be ambiguous. Hence, the account of single epistemic changes is indeed perfectly general.

However, this points to the real issue. What about iterated change? This is a large issue.[40] I am not entirely clear about it, and I should not add here another sub-paper trying to treat it. Let me only add a few reflections in order to illustrate that the issue is indeed intricate.

First I should say that dealing with actual iterated change is not a problem at all (beyond problems with conditionalizing by null propositions, which restrict the rules (1), (2), and (11), anyway). All the rules we have considered, conditionalization, time shift, and forgetting rules, are iteratively applicable. This holds for steps of the same kind. It is a familiar fact that two steps of simple conditionalization (1) add up to one such step. Similarly for two steps of Jeffrey conditionalization (2). Two applications of the rules of time shift (3) and (4) can be summarized in one such application. And likewise for the rule of forgetting (11). We can even iteratively apply different kinds of rules.

Problems start when we reflect on such changes and have to assess them as favorable or unfavorable. The problems already emerge with two steps of change. Well, not immediately: if I reflect in P first to reach P' and then P'' under conditions in which the original reflection principle (17) applies, I trust P'' already in P. Reversely, if I reflect in P first to reach P' and then P'' under unfavorable conditions, then I accept neither P' nor P'' in P. Hence, in these cases the full reflection principle (29) seems to have an easy extension.

However, the mixed cases are not so easy. And I just appealed to such cases in order to keep pure at least the single steps. Consider the case where I reflect in

40. Thanks to an anonymous referee for pressing me at this point.

P on undergoing first an improvement to P' and then a worsening to P''. Given this development my probabilities in P should agree with those in P' and ignore those in P''. So much seems clear.

Now, let's reverse the steps and let the step from P to P' be a worsening and the step from there to P'' an improvement. Reflecting on this in P, I don't trust in P', of course. But I cannot trust in P'', either, because P'' is infected by the inferior P'. So, given this development, what should I believe in P? This is not arbitrary. It seems to me that in this case I should enter a counterfactual consideration and ponder at which state P^* I would have arrived, had I undergone the favorable change from P' to P'' with my actual P as starting point. For instance, what should I believe now, given that I first forget about my date with the vice-president and then learn about my simultaneous date with the president? Surely, I should conditionally believe that I have two dates at the same time. That is, I should conditionally believe what I would believe when learning about the president without forgetting about the vice-president.

If this seems plausible, then iterated reflection becomes terribly involved. With each cycle of an unfavorable change followed by a favorable one, counterfactual considerations pile up. Perhaps this complexity may be simplified in an elegant way. Certainly, the complexity is amenable to theoretical grasp. It is clear, however, that such theoretical grasp would require a lot of additional counterfactual machinery. I won't even start with this now. The above remarks about two-step changes must suffice here.

We have seen in Section 8 that the original reflection principle (17) entails various other principles. How are those principles affected by the generalization and modification (29) of the reflection principle? First, it should be clear that the principle (19), the trivial expertise of an opinion for itself, is contained in (29), due to my arbitrary subsumption of the case of no change under $\uparrow$ or $\downarrow$. Second, it should be clear that the iteration principle (18), also in its indexical form (26), holds only in the case of $(P_0 \uparrow \pi_z)$; indeed, this must hold for all π_z appearing in the sum of (26). Hence, the term $P(\{\pi_z\})$ in (26) should be replaced by the term $P(\{\pi_z\} \mid (P_0 \uparrow \pi_z))$. Under this further internalized condition, the iteration principle holds unconditionally.

What about auto-transparency (20)? It is universally valid. Recall that we had derived (20) from (19) alone. Hence, both have the same range of validity. The case is different with perfect memory (21). We saw that its proof relied on the iteration principle (18); hence, perfect memory (21) shares its restricted applicability. And it does not generalize. If the posterior probability measure is not required to be an expert for the prior one (because it has moved to an inferior epistemological perspective), it cannot be expected to preserve the (auto-transparent) certainties of the prior measure.

Finally, auto-epistemic conditionalization (23) stands unshaken. If each π_z comes from P_0 by auto-epistemic conditionalization, then π_z is an improved

point of view so that $(P_0 \uparrow \pi_z)$ holds. If this is also the assessment of the subject (as it should be), then the equivalence stated in (24) extends to the full reflection principle in the first half of (29).

Let me conclude this section by extending our considerations into still another direction. So far, all reflections were forward-looking: given my future epistemic state is such and such, what should I believe now? But we may as well ask the backward-looking question: given my past state was such and such, what should I believe now? To my knowledge this backward-looking question has not been discussed in the literature. The answer seems as evident as in the forward-looking case. If my present state is an improvement of my past state, my past judgment does not count; even given the past judgment, I stick to my present judgment. However, if I think my present state is poorer than my past state, I should listen to my past state. That is, usually I will also have forgotten my past state and will be unsure what it was; but given it said this and this, I should now say the very same. I have forgotten my date with the dentist. Given, though, that yesterday I believed that the date is next Tuesday, I should now believe so as well.

This leads me to the following reverse principle, which is about the *posterior* credence P auto-epistemically reflecting upon possible prior credences π_{-z} z times *earlier* than P (for $z > 0$) and upon the possible changes $\langle \pi_{-z}, P_0, \theta \rangle$ from π_{-z} to P_0 in in various ways θ (where P_0 is again the restriction of P to $\mathcal{A}$):

(30) The (*Full*) *Reverse Reflection Principle*:
$P(. \mid \{\pi_{-z} = Q^i_{-z}\} \cap (\pi_{-z} \uparrow P_0)) = P_0$, and
$P(. \mid \{\pi_{-z} = Q^i_{-z}\} \cap (\pi_{-z} \downarrow P_0)) = Q$,
where Q^i_{-z} is defined as in (29).

Thus, backward-looking auto-epistemology is just as strong as forward-looking auto-epistemology and works in a perfectly parallel way. Therefore, it has analogous consequences. Is this really plausible? For instance, from (30) we can derive a *reverse iteration principle* in analogy to (18) and (26) (which I spare writing down). I think, rightly so. If you have somehow moved to an inferior epistemic state and ponder from where you did so, then your speculations about your possible prior states must be in agreement with your present posterior state. For instance, I cannot be very unsure now about A (because I have forgotten about it now) and at the same time be quite sure that I have been quite sure of A. If the latter should really be the case, I should use it as a guideline to remove my present uncertainty. (Of course, if my present uncertainty about A is due to some learning, some new piece of evidence, this auto-epistemic state would be

perfectly consistent. Then I can well remember that I have been certain about *A* and know now that this certainty was unjustified.)

You may feel uneasy about that reverse principle, because your speculations about the past states are hanging in the air, as it were. Recall, though, that the same may hold in the forward-looking case. If you expect to learn by Jeffrey conditionalization, your present probabilities for your future seemings that get expressed in some posterior distribution over the evidential partition are hanging in the air in a similar way. Only in case you expect to learn by simple conditionalization, your expectations about your future states are directly correlated with your expectations about the future events. I conclude that the reverse iteration principle is perfectly acceptable.

What about the other principles? Auto-transparency (20) is unaffected, since it is not an instance of backward looking auto-epistemology. However, perfect memory (21) seems now to commute into a principle of perfect auto-epistemic foresight: at the prior time I already knew about my posterior inferior state. This is absurd. Let's check, though, whether this really follows, and let's translate our proof of (21) into the reversed terms. It says then: if your posterior P is auto-epistemically transparent, as just asserted, and if your posterior P is also a mixture of your possible prior superior opinions according to the reverse iteration principle, then your possible priors π_{-z} and thus also your actual prior P_{-z} can't fail to be sure about your posterior opinion P. However, there was a caveat in the proof of (21). The proof made use of an extended iteration principle and presupposed that it applies also to auto-epistemic propositions. And this is certainly not justified for the reverse iteration principle. Your auto-transparency is not the effect of a worsening of your epistemic state and hence not something required to agree with your prior superior opinions. Hence, the reverse iteration principle cannot be auto-epistemically extended in order to prove something like perfect auto-epistemic foresight. Moreover, no auto-epistemic conditionalization rule like (27) is associated with the reverse reflection principle. This makes sense only for forward-looking principles. Finally, iterating the reverse principle is presumably as intricate as iterating the ordinary principle (29). There is no point in trying to deepen the issue.

Are the principles (29) and (30) related? Does the favorable part of (29) entail the favorable part of (30), and is the unfavorable part of (29) implied by the unfavorable part of (30)? Well, the first entailment holds. If perfect memory (20) holds on the basis of the favorable part of (29), then conditionalizing your posterior state on your prior state is conditionalizing your posterior state on a proposition with probability 1, and therefore the favorable part of (30) holds as well. However, this again presupposes the auto-epistemic extension of the iteration principle just discussed. Therefore the entailment does not carry over to the corresponding unfavorable parts.

This concludes my twofold generalization of auto-epistemology concerning unfavorable worsenings in addition to favorable improvements of epistemic situations and concerning backward-looking in addition to forward-looking reflections. I am happy to have thus arrived at general and unconditional principles.

Is all of this idle play just in order to satisfy philosophical phantasies? No, not at all. It plays an important role everywhere in everyday life. We continuously fight not only to improve our epistemic situation, but also to avoid worsenings. We fight forgetting on a small scale everyday and on a large scale with expensive historical institutions; we fight against drugs, because they ruin ourselves and our epistemic perspective; and so forth. This would be the business of a suitably extended decision theory, which is no longer our present concern.[41] It is clear, though, that such a decision theory presupposes generalized auto-epistemology as explored here. If we were not aware of our forgetfulness and other possible worsenings of our perspective, we would simply stumble into these traps without being able to gain some control.

11. Final Application: Shangri-La

Let me finally illustrate the full reflection principles (29) and (30) with an artificial and simple example: with the story of Shangri-La, which was invented by Arntzenius (2003) and which involves only problems with potential forgetting, but none with temporal self-location. It goes like this:

You are elected to enter the beautiful country of Shangri-La. There are two routes, one via the Golden Mountains and one via the Misty Sea. A fair coin is thrown to select the route: heads = the Golden Mountains, tails = the Misty Sea. So, initially you believe with probability 1/2 in each of the paths. You are guided along the path determined. Soon you see that it leads along the Golden Mountains, and you become sure that the coin showed heads. Of course, you might as well have sailed over the Misty Sea, in which case you would have become sure of tails.

Now, the tricky point is this: you finally enter the country of Shangri-La. You were told right at the beginning that, if you arrive via the Misty Sea, your memories of that travel will be completely erased upon entering Shangri-La and replaced by beautiful memories of the Golden Mountains precisely as you actually have and which you retain from your trip over the mountains. Again you are asked for your probabilities of heads and tails. Intuitively, it seems clear what you should say: you return to your initial assessment, 1/2 for each side of

41. I am not aware that my beginnings in Spohn (1978: Section 4.4) have been substantially elaborated.

the coin. I have not seen any doubt about this in the literature (except by Hawley 2013).

Still there is an obvious problem for epistemological theory. There are three stages. You start with $P_0(H) = P_0(T) = 1/2$. You see the Golden Mountains, and simple conditionalization with respect to this experience leads you to $P_1(H) = 1$. Finally, you enter Shangri-La, nothing happens, you do not learn anything, you do not forget anything, and still it seems reasonable that you change again and return to $P_2(H) = 1/2$. As Arntzenius (2003) emphasizes, this flies in the face of all our rules for changing probabilities, in particular the reflection and iteration principles (17) and (18). How are we to account for this? It seems that the mere unactualized possibility that you arrive at P_2 through some other path makes for the crucial difference. But how?

Let H = heads, T = tails, M = you take the mountain route, and S = you take the sea route. Then the set Γ_i of possible pieces of evidence at stage i contains just two elements: M_i^Γ = you have 'mountain' experiences/memories at stage i, and S_i^Γ = you have 'sea' experiences/memories at stage i. We are dealing with four epistemic states concerning those (and other) propositions: your initial state P_0, your final state P_2, and two intermediate states, $P_1 = P_{1M}$, the state you actually have traveling through the mountains, and P_{1S}, the state you would have been in, had you traveled by the sea. We surely have $P_0(H) = P_0(T) = 1/2$, $P_i(M \text{ iff } H) = P_i(S \text{ iff } T) = 1$ for all four states, that is, $i = 0, 1M, 1S, 2$; and we also have $P_i(M_i^\Gamma \text{ iff } M) = P_i(S_i^\Gamma \text{ iff } S) = 1$ for $i = 0, 1M, 1S$. Indeed, $P_{1M} = P_0(. \mid M_i^\Gamma)$ and $P_{1S} = P_0(. \mid S_i^\Gamma)$. So far for the clear part of the story.

The unclear part is about P_2. There is a brief account in terms of the notion of a subjective protocol (introduced after (21) in Section 8). Which protocol governs the (non-)change from P_0 to P_2? The story is that at the final stage you have received total evidence M_2^Γ however things go, and hence $P_0(M_2^\Gamma) = 1$. So, if you apply auto-epistemic conditionalization (23) to P_0, you get $P_2 = P_0(. \mid M_2^\Gamma) = P_0$. This already accounts for our intuition of what P_2 should be like.

However, this account omits the details of the story and hides them in the subjective protocol. In particular, it does not explain the change from P_1 to P_2. So let us look more closely. One crucial point is how much or how little you have forgotten when arriving at Shangri-La. The intuitive solution tacitly presupposes that you are still fully aware of the auto-epistemological part of the story. If you did not know or had forgotten that part, you should rationally end up believing that you came via the mountains, whether or not you actually came that way.[42]

In fact, having arrived in Shangri-La the only thing you no longer know is

42. This is indeed the way in which Hawley (2013: 96–97) argues. He insists that you don't lose information when entering Shangri-La. Indeed, you don't. However, he does not do full justice to the auto-epistemological aspects of the story. In particular, he misses that auto-epistemological conditionalization (23) may outstrip the ordinary conditionalization rules (1) and (2).

whether in between you had been in P_{1M} or P_{1S}. The story presupposes, however, that you still know with which P_0 you started. This is shown by a move we have already applied in Section 7 about Sleeping Beauty. Assume that H and T do not stand for random events having the same subjective probability from each perspective not possessing inadmissible evidence (such as M_1^Γ). Let H and T rather represent whether, say, something shows up *here* or *there*, and assume that your initial P_0 just assigns some subjective probability x to this H and $1 - x$ to this T. What should you believe then about H and T after having entered Shangri-La? You have no better assessment of H and T than in the beginning. Hence, $P_2(H) = x$ would again be the reasonable credence. However, this presupposes that you still remember your initial probabilities. If you don't, $P_2(H)$ may be anything. Hence, the reference of the original story to the random outcomes of the throw of a coin was a vivid, but inessential and potentially confusing ingredient, as it was in Sleeping Beauty.

What, then, is the auto-epistemic reasoning embedded in your final P_2? Let me sketch it only in an informal way. In P_2 you know (with probability 1) that $P_{1M}(M) = 1$ and $P_{1S}(S) = 1$, and moreover that $(P_{1M} \downarrow P_2)$ (no change being a limiting case of $\downarrow$) and $(P_{1S} \downarrow P_2)$. Hence, according to the reverse iteration principle mentioned after (30), P_2, as applied to ordinary propositions, must be a mixture of P_{1M} and P_{1S}, mixed by your P_2 probabilities for having been, respectively, in P_{1M} and P_{1S}. What are these?

They are the same as your P_2 probabilities for M and S, since you still know how P_{1M} and P_{1S} have come about. And what are your P_2 probabilities for M and S? The same as those for H and T. In the original story, these are random events about which P_2 has no inadmissible information; hence, $P_2(H) = P_2(T) = 1/2$. However, we need not and should not rely on the assessment of random events on the basis of admissible information. We may instead rely on the above assumption that P_0 is still remembered in P_2. And since P_2 is assessed as inferior to P_0, the judgment of P_0 is respected in P_2 according to the reverse reflection principle (30), not conditionally on P_0, but unconditionally, since in P_2 it is known what P_0 was. And P_0 told that $P_0(H) = P_0(T) = 1/2$. (The same reasoning works in the subjective case where $P_0(H) = x$ and $P_0(T) = 1 - x$.) In this way, the story of Shangri-La nicely demonstrates the auto-epistemology of forgetting and the appertaining reversed reflection principle (30).

Note, however, that there was a little gap in my argument. Shangri-La is a story with three stages, whereas the auto-epistemic principles discussed above applied only to two stages. So, in order to account for Shangri-La, I first applied them to stages 1 and 2 of Shangri-La and then to stages 0 and 2, assuming that in both applications your epistemic situation worsens. However, if we want to fully analyze the story as a three-stage case, then we run into the problems mentioned in the previous section concerning the iteration of our principles. Those

problems arose only with mixed cases. However, Shangri-La is such a mixed case, first an improvement from P_0 to P_{1M} or P_{1S} and then a worsening to P_2. So the above analysis is still incomplete. Here, I will rest content with my analysis of the two-stage cases and of Shangri-La as such a case.

This finishes my paper. We have used the problem of Sleeping Beauty as a motive for making a big round trip through at least three obscure areas of formal epistemology outside the standard fields, which we found hardly addressed, although they are relevant and interesting in themselves. If my explorations have made progress in those areas, this paper has reached its aims.

Acknowledgments

I am indebted to four anonymous referees for many insightful remarks and to the area editor for her or his care. Their urging helped me to substantially improve my paper. I am also grateful for critical discussions of the paper within the DFG research unit "What if?" at the University of Konstanz.

References

Arntzenius, Frank (2002). Reflections on Sleeping Beauty. *Analysis*, *62*(1), 53–62. https://doi.org/10.1093/analys/62.1.53

Arntzenius, Frank (2003). Some Problems for Conditionalization and Reflection. *Journal of Philosophy*, *100*(7), 356–370. https://doi.org/10.5840/jphil2003100729

Binkley, Robert W. (1968). The Surprise Examination in Modal Logic. *Journal of Philosophy*, *65*(5), 127–136. https://doi.org/10.2307/2024556

Bradley, Darren J. (2011). Self-Location is no Problem for Conditionalization. *Synthese*, *182*(3), 393–411. https://doi.org/10.1007/s11229-010-9748-9

Bradley, Darren J. (2012). Four Problems about Self-Locating Belief. *Philosophical Review*, *121*(2), 149–177. https://doi.org/10.1215/00318108-1539071

Bradley, Darren and Hannes Leitgeb (2006). When Betting Odds and Credences Come Apart: More Worries for Dutch Book Arguments. *Analysis*, *66*(2), 119–127. https://doi.org/10.1093/analys/66.2.119

Carnap, Rudolf (1947). On the Application of Inductive Logic. *Philosophy and Phenomenological Research 8*(1), 133–148. https://doi.org/10.2307/2102920

Carnap, Rudolf (1971). A Basic System of Inductive Logic, Part I. In Rudolf Carnap and Richard C. Jeffrey (Eds.), *Studies in Inductive Logic and Probability* (Vol. 1, 33–165). University of California Press.

Castañeda, Hector-Neri (1966). 'He': A Study in the Logic of Self-Consciousness. *Ratio*, *8*(2), 130–157.

De Finetti, Bruno (1964). Foresight: Its Logical Laws, Its Subjective Sources (Henry E. Kyburg, Jr., Trans.). In Henry E. Kyburg, Jr. and Henry E. Smokler (Eds.), *Studies in Subjective Probability* (93—158). John Wiley & Sons.

Diaconis, Persi and Sandy L. Zabell (1982). Updating Subjective Probability. *Journal of the American Statistical Association*, *77*(380), 822–830. https://doi.org/10.1080/01621459.1982.10477893

Douven, Igor, (2012). Learning Conditional Information. *Mind and Language*, *27*(3), 239–263. https://doi.org/10.1111/j.1468-0017.2012.01443.x

Dorr, Cian (2002). Sleeping Beauty: In Defence of Elga. *Analysis*, *62*(4), 292–296. https://doi.org/10.1093/analys/62.4.292

Elga, Adam (2000). Self-Locating Belief and the Sleeping Beauty Problem. *Analysis*, *60*(2), 143–147. https://doi.org/10.1093/analys/60.2.143

Freitag, Wolfgang (2015). I Bet You'll Solve Goodman's Riddle. *Philosophical Quarterly*, *65*(259), 254–267. https://doi.org/10.1093/pq/pqu093

Freitag, Wolfgang and Alexandra Zinke (2016). Ranks for the Riddle? Spohn Conditionalization and Goodman's Paradox. In Wolfgang Freitag et al. (Eds.), *Von Rang und Namen: Philosophical Essays in Honour of Wolfgang Spohn* (107–125). Mentis.

Gaifman, Haim (1988). A Theory of Higher Order Probabilities. In Brian Skyrms, William L. Harper (Eds.), *Causation, Chance, and Credence* (191–219). Kluwer. https://doi.org/10.1007/978-94-009-2863-3_11

Goldstein, Matthew (1983). The Prevision of a Prevision. *Journal of the American Statistical Association*, *78*(384), 817–819. https://doi.org/10.1080/01621459.1983.10477026

Goodman, Nelson (1946). A Query on Confirmation. *Journal of Philosophy*, *43*(14), 383–385. https://doi.org/10.2307/2020332

Hájek, Alan (2003b). Conditional Probability Is the Very Guide of Life. In Henry E. Kyburg, Jr. and Mariam Thalos (Eds.), *Probability Is the Very Guide of Life* (183–203). Open Court.

Halpern, Joseph Y. (2003). *Reasoning about Uncertainty*. MIT Press.

Hawley, Patrick (2013). Inertia, Optimism and Beauty. *Noûs*, *47*(1), 85–103. https://doi.org/10.1111/j.1468-0068.2010.00817.x

Hild, Matthias (1998a). Auto-Epistemology and Updating. *Philosophical Studies*, *92*(3), 321–361. https://doi.org/10.1023/A:1004229808144

Hild, Matthias (1998b). The Coherence Argument against Conditionalization. *Synthese*, *115*(2), 229–258. https://doi.org/10.1023/A:1005082908147

Hetherington, Stephen (2001). Why There Need Not Be Any Grue Problem About Inductive Inference as Such. *Philosophy*, *76*(295), 127–136. https://doi.org/10.1017/S0031819101000080

Hitchcock, Christopher (2004). Beauty and the Bets. *Synthese*, *139*(3), 405–420. https://doi.org/10.1023/B:SYNT.0000024889.29125.c0

Jeffrey, Richard C. (1983). *The Logic of Decision* (2nd ed.). University of Chicago Press.

Karlander, Karl and Levi Spectre (2010). Sleeping Beauty Meets Monday. *Synthese*, *174*(3), 397–412. https://doi.org/10.1007/s11229-009-9464-5

Kripke, Saul A. (1982). *Wittgenstein on Rules and Private Language*. Blackwell.

Leitgeb, Hannes and Richard Pettigrew (2010). An Objective Justification of Bayesianism II: The Consequences of Minimizing Accuracy. *Philosophy of Science*, *77*(2), 236–272. https://doi.org/10.1086/651318

Lewis, David (1979). Attitudes *De Dicto* and *De Se*. *Philosophical Review*, *88*(4), 513–543. https://doi.org/10.2307/2184843

Lewis, David (1980). A Subjectivist's Guide to Objective Chance. In Richard C. Jeffrey

(Ed.), *Studies in Inductive Logic and Probability* (Vol. 2, 263–293). University of California Press. https://doi.org/10.1007/978-94-009-9117-0_14

Lewis, David (2001). Sleeping Beauty: Reply to Elga. *Analysis*, *61*(3), 171–176. https://doi.org/10.1093/analys/61.3.171

Maher, Patrick (1993). *Betting on Theories*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527326

McTaggart, John M. E. (1908). The Unreality of Time. *Mind, 17*(68), 457–474. https://doi.org/10.1093/mind/XVII.4.457

Müller, Thomas (2016). Sleeping Beauty in Branching Time. In Wolfgang Freitag et al. (Eds.), *Von Rang und Namen: Philosophical Essays in Honour of Wolfgang Spohn* (307–325). Mentis.

Perry, John (1979). The Problem of the Essential Indexical. *Noûs*, *13*(1), 3–21. https://doi.org/10.2307/2214792

Piccione, Michele and Ariel Rubinstein (1997). On the Interpretation of Decision Problems with Imperfect Recall. *Games and Economic Behavior*, *20*(1), 3–24. https://doi.org/10.1006/game.1997.0536

Quine, Willard V. O. (1969). Natural Kinds. In Nicholas Rescher et al. (Eds), *Essays in Honor of Carl G. Hempel* (1–23). D. Reidel. https://doi.org/10.1007/978-94-017-1466-2_2

Quine, Willard V. O. (1990). *Pursuit of Truth*. Harvard University Press.

Rumberg, Antje (2016). Transition Semantics for Branching Time. *Journal of Logic, Language and Information*, *25*(1), 77–108. https://doi.org/10.1007/s10849-015-9231-6

Schulz, Moritz (2010). The Dynamics of Indexical Belief. *Erkenntnis*, *72*(3), 337–351. https://doi.org/10.1007/s10670-010-9209-3

Schwarz, Wolfgang (2012). Changing Minds in a Changing World. *Philosophical Studies*, *159*(2), 219–239. https://doi.org/10.1007/s11098-011-9699-0

Shafer, Glenn (1985). Conditional Probability. *International Statistical Review*, *53*(3), 261–277. https://doi.org/10.2307/1402890

Spohn, Wolfgang (1978). *Grundlagen der Entscheidungstheorie*. Scriptor. (Out of print. Available online at http://www.uni-konstanz.de/FuF/Philo/Philosophie/philosophie/files/ge.buch.gesamt.pdf)

Spohn, Wolfgang (2012). *The Laws of Belief: Ranking Theory and Its Philosophical Applications*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199697502.001.0001

Spohn, Wolfgang (2016). Enumerative Induction. In Christoph Beierle, Gerhard Brewka, and Matthias Thimm (Eds.), *Computational Models of Rationality: Essays Dedicated to Gabriele Kern-Isberner on the Occasion of Her 60th Birthday* (96–114). College Publications.

Stalker, Douglas (Ed.) (1994). *Grue! The New Riddle of Induction*. Open Court.

Stalnaker, Robert C. (2014). *Context*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199645169.001.0001

Teller, Paul (1976). Conditionalization, Observation and Change of Preference. In William L. Harper, Clifford Alan Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, (205–259). D. Reidel. https://doi.org/10.1007/978-94-010-1853-1_9

Titelbaum, Michael G. (2016). Self-Locating Credences. In Alan Hájek, Christopher Hitchcock (Eds.), *The Oxford Handbook of Probability and Philosophy* (666–680). Oxford University Press.

Van Fraassen, Bas C. (1981). A Problem for Relative Information Minimizers in Probability Kinematics. *British Journal for the Philosophy of Science*, *32*(4), 375–379. https://doi.org/10.1093/bjps/32.4.375

Van Fraassen, Bas C. (1984). Belief and the Will. *Journal of Philosophy*, *81*(5), 235–256. https://doi.org/10.2307/2026388

Weintraub, Ruth (2004). Sleeping Beauty: A Simple Solution. *Analysis*, *64*(1), 8–10. https://doi.org/10.1093/analys/64.1.8

White, Roger (2006). The Generalized Sleeping Beauty Problem: A Challenge for the Thirders. *Analysis*, *66*(2), 114–119. https://doi.org/10.1093/analys/66.2.114