The Dynamics of Indexical Belief

[This is a penultimate draft. Please quote only from the published version.]

Abstract

Indexical beliefs pose a special problem for standard theories of bayesian updating. Sometimes we are uncertain about our location in time and space. How are we to update our beliefs in situations like these? In a stepwise fashion, I develop a constraint on the dynamics of indexical belief. As an application, the suggested constraint is brought to bear on the Sleeping Beauty problem.

1   Introduction

Indexical belief constitutes a problem for epistemic dynamics right at the outset. Suppose I believe on Sunday what I would express by 'Today is Sunday'. What should I believe on Monday? In the most natural circumstances, I should not continue to believe what would be expressed by 'Today is Sunday'. Rather, I should come to accept 'Yesterday was Sunday'. Thus, continuity of indexical belief does not manifest itself in assent to the same sentence. Moreover, sometimes the evidence we acquire is indexical. On Monday, I may learn 'Today is Jacky's birthday'. How should my new credences relate to my past credences on Sunday? Clearly, I should not conditionalize my Sunday credences on 'Today is Jacky's birthday'.

The problem becomes worse. In some cases, we cannot keep track of the relevant contextual features. Around midnight on Sunday, I may be unsure whether it is still Sunday or already Monday. As a result, I become uncertain whether I should belief 'Today is Sunday' or 'Today is Monday'. How should our credence evolve when we lose track of our location in time and space?

The problem of updating in situations of uncertainty concerning our spatio-temporal location has emerged from the Sleeping Beauty problem (which will be reviewed in section 5). But, as the foregoing considerations indicate, the problem of indexical belief is much more general. In this paper, I develop in a stepwise fashion a general constraint on the dynamics of indexical belief which

1

extends the standard account of Bayesian updating. A close eye will be kept on the insights gained in the debate surrounding the Sleeping Beauty problem. Finally, I show how the account applies to this problem and defend it against some objections.

Before I start, let me briefly indicate how the present account relates to other work which has been done on the dynamics of indexical belief and the Sleeping Beauty problem. The current debate was initiated by Elga (2000), who argued that sometimes rational agents should change their credences even though they apparently do not receive new information. The major part of Elga's position, namely that Sleeping Beauty should change her credences in heads, was subsequently defended by Arntzenius (2002), Dorr (2002), Monton (2002), and many others. On the other hand, Lewis (2001) denied that the relevant change in credence would be rational by holding that one should never change one's credences without receiving relevant new information. Lewis's position was reinforced by Jenkins (2005).

Weintraub (2004) suggests that even though Sleeping Beauty knows on Sunday what she would express by 'I will be awake *tomorrow*', realizing on Monday that 'I am awake *today*' is true constitutes new information. However, it seems that 'I will be awake tomorrow' expresses on Sunday the same (or at least an intimately related and equivalent) proposition as 'I am awake today' expresses on Monday. How can a change of credence in such a case be rational? In response to this question, Monton (2002) suggests that the change in credence should be explained by the fact that relevant information is *lost*.

In a similar vein, Horgan (2004) emphasizes the role of indexicals in analyzing the Sleeping Beauty problem: although 'today' and 'yesterday' may be co-referential with respect to two successive contexts, this is not always obvious for an epistemic subject who is uncertain about the contextual changes. Sleeping Beauty cannot *retain* her indexical belief from Sunday to Monday because she loses track of the relevant contextual features. The main aim of this paper is to accommodate within a broadly Bayesian framework the possibility of losing track of relevant contextual features.

2   Continuity

Ideally, if we gain no new evidence, we should continue to believe the same. In expressing our beliefs, however, we need to make appropriate substitutions of the indexical terms we are using. Suppose I believe today what I would

express by the sentence '*Tomorrow* is Monday'. If I continue to believe the same tomorrow, I will then express my belief by the sentence '*Today* is Monday'.

In order to be sensitive to aspects of indexicality in the most perspicuous way, I will take credences to attach to sentences within contexts. The expression $P_C(A)$ will stand for the credence an epistemic subject attaches to sentence A within context C.[1] Let us think of contexts as pairs (t, s) of a time coordinate t and a space coordinate s. Define the *distance* C_1/C_2 between two contexts $C_1 = (t_1, s_1)$ and $C_2 = (t_2, s_2)$ as $(t_2 - t_1, s_2 - s_1)$.[2] For present purposes, we can assume that the substitutions we need to make when we move from one context to another depend solely on the distance between the two contexts. Given an indexical expression α and a distance x, let us write $\alpha[x]$ for a corresponding indexical expression α^* which satisfies the following constraint: given any two contexts C and C^* with distance x, the semantic value of α^* in C equals the semantic value of α in C^*.[3] For example, 'today'[1] = 'tomorrow' (measuring here the distance in days). This definition extends to sentences: given a sentence S and a distance x, let us write $S[x]$ for the sentence coming from S by substituting $\alpha[x]$ for the indexical elements α in S. Canonically, I will use notations such as $S[C/C^*]$ in which, as defined above, C/C^* is the distance between C and C^*. So, assuming that the distance between C and C^* is one day,

(1) 'Today is Monday'$[C/C^*]$ = 'Tomorrow is Monday'.

A few words of clarification. One can think of distances as operations on sentences which yield new sentences as output. For instance, one can think of distance 1 (in days) as an operation which yields the sentence 'Tomorrow is Monday' as output when applied to the sentence 'Today is Monday' as input. Equivalently, one can also decompose expressions of the form $S[C/C^*]$ by taking the sentence S as an operation which takes distances as input and sentences as output. In any case, the important feature of this operation is that the sentence $S[C/C^*]$ is a sentences which expresses in C the same proposition as S expresses in C^*. This is because the sentence $S[C/C^*]$ comes from S by making the appropriate substitutions of the indexical terms occurring in S. So,

---

[1]Formally, P_C is a function from sentences into the interval $[0, 1]$.

[2]Distances in this sense would perhaps better be called *directed distances*; for the sake of brevity, I will omit this attribute.

[3]In most cases, there will be more than one indexical expression which satisfies this constraint (think of 'tomorrow' and 'the day after today'); let us work with the most natural choice.

when I consider on Monday an indexical sentence such as 'Today is Monday', the sentence 'Today is Monday'[1], namely 'Tomorrow is Monday', is a sentence which expresses the same proposition on Sunday as 'Today is Monday' expresses on Monday.

The present picture of contexts and distances is simplified in various ways. Since contexts are modeled by a time coordinate and a space coordinate only, indexicals which are sensitive to other features than those are not dealt with. For instance, indexicals such as 'I' which are sensitive to the agent of the context or demonstrative expressions such as 'this' which are sensitive to the demonstrated object are not accommodated. In general, it may also occur that an object is demonstratively available in one context but not in another. For example, I may now be able to refer to a raven sitting on the windowsill by using 'this raven' without being in the position to refer to this raven by using a corresponding demonstrative expression tomorrow. Moreover, it should be pointed out that in order to decide whether two indexicals are co-refering in their respective contexts, sometimes more than knowing the distance is required. Sometimes our absolute position in time identified by a calendar system can matter as well. Just by knowing that three months have passed I may not be in a position to decide whether my present use of 'this year' and a certain past use of 'next year' are co-refering. However, given an indexical sensitive to a spatio-temporal position, knowing the distance between present and past context will generally put me in a position to find an indexical which co-refers in the past context with the indexical under consideration in the present context. Yet such indexicals may sound rather stilted. Looking for an expression which co-refers with 'this year' three months ago without knowing whether this year is already three months old or not, I can only come up with something like 'the year which will be the present year in three months time'. For this reason, I will here focus on indexicals such as 'today', 'tomorrow', or 'in three weeks time' whose referent in a context is determined by a specific distance to the time of the context. Doing justice to the great variety of indexical expressions in natural language will require a much more complex model. But in developing the dynamics of indexical beliefs, we may be better off by starting with the simple cases.

When we move from a context C to a context C^* and consider an indexical sentence S in the present context C^*, the sentence $S[C/C^*]$ is, as noted above, a sentence which expresses the same proposition in the earlier context C as the sentence S we are currently considering expresses in C^*. Now, when we

aim at continuity of belief between a past context C and a present context C^* and ask ourselves what our present credences in an indexical sentence S should be, it therefore seems that we should assign to it the subjective probability we assigned to $S[C/C^*]$ in C. Hence, continuity in indexical belief concerning 'Tomorrow is Monday' seems to require the following:

(2) $P_{C^*}(\text{'Today is Monday'}) = P_C(\text{'Today is Monday'}[C/C^*])$.

If C^* is a Monday-context and C a Sunday-context, the relevant distance between these two contexts is one day, and continuity in belief would then be exemplified by

(3) $P_{\text{Monday}}(\text{'Today is Monday}) = P_{\text{Sunday}}(\text{'Tomorrow is Monday'})$.

In ideal circumstances, and if no new evidence is gained, one should attach the same credence to 'Today is Monday' as one did to 'Tomorrow is Monday' one day before.

By abstracting from the details of the example, we can formulate a more general continuity principle:

(C: Continuity)
Let C and C^* be two successive contexts such that the epistemic subject gains no new evidence after C. In ideal circumstances,

$$P_{C^*}(A) = P_C(A[C/C^*]).$$

3 Continuous Conditionalization

While moving through space and time, our present credences are systematically related to our past credences. In ideal circumstances, what I think today should be a joint product of what I thought yesterday and the evidence I have gained in the meantime. As a starting point, the continuity principle describes how indexical beliefs should evolve in the absence of new information. It needs to be extended to incorporate the acquisition of new evidence.

Suppose I learn today '*Today* is Monday'. How should my new credences relate to yesterday's credences? The most straightforward suggestion would be that I should conditionalize yesterday's credences on '*Tomorrow* is Monday'. Then my Monday credence in e.g. '*Today* is Jacky's birthday' should be my Sunday credence in '*Tomorrow* is Jacky's birthday' conditional on '*Tomorrow* is Monday'.

Thus, the idea would be that we reach our present credences by conditionalizing our past credences on the newly found evidence. In order to cope with the indexicality of our beliefs, we need to substitute the indexical terms we are using in accordance with the distance between the present and the past context. In particular, if E is an indexical sentence which constitutes evidence in the present context C^*, we need to conditionalize our past credences in context C not on E itself but rather on $E[C/C^*]$.

On the basis of this idea, we can formulate a preliminary principle of how rational agents ought to update their beliefs over time:

(CC: Continuous Conditionalization)

Let C and C^* be two successive contexts and E the evidence available in C^* which was gained after C. In ideal circumstances,

1. First, set $P'_{C^*}(A) := P_C(A[C/C^*])$.

2. Second, set $P_{C^*}(A) := P'_{C^*}(A|E)$.

Combined, the two steps result in

$$P_{C^*}(A) = P_C(A[C/C^*]|E[C/C^*]).^4$$

Continuous Conditionalization is decomposed into two factors: one accommodating the change of context, e.g. the passage of time and the movement in space, and the other incorporating the newly found evidence. Given a credence function P_C in an initial context C and a succeeding context C^*, one first goes to a preliminary credence function P'_{C^*} by setting for all A: $P'_{C^*}(A) := P_C(A[C/C^*])$. This ensures that the old credences concerning indexical sentences are changed according to the distance between the two contexts. In effect, one first applies the continuity principle in order to yield preliminary credences for C^* which accommodate the change of context. In a second step, one updates the preliminary credences by conditionalizing on the new evidence E. In this way, one arrives at the final credences by setting $P_{C^*}(A) := P'_{C^*}(A|E)$. Since E is a sentence to be evaluated in C^*, $P'_{C^*}(A|E) = P_C(A[C/C^*]|E[C/C^*])$. So, a compact form of representing the constraint would be to require $P_{C^*}(A) = P_C(A[C/C^*]|E[C/C^*])$.

---

[4]As usual, $P_C(E[C/C^*]) > 0$ and so also $P'_{C^*}(E) > 0$ is assumed for the conditional probability to be defined.

4   Approximating Continuous Conditionalization

Circumstances are not always ideal. Sometimes we cannot keep track of the relevant contextual features. Consequently, we cannot make the required substitutions of the indexical terms we are using. If I lost track of what day it is, rationality does not require me to attach to '*Yesterday* was Sunday' the same credence on Monday as I attached on Sunday to '*Today* is Sunday'.

What should we do in cases in which we are uncertain about our present distance to the past context we are updating from? Even in such cases, our present credences should be sensitive to our past credences. I suggest that we update in cases of uncertainty concerning our distance to the past context by *approximating* Continuous Conditionalization. In situations in which we are uncertain about our distance to the context we are updating from, we weigh the possible results of applying Continuous Conditionalization by our credences in the relevant distances.[5]

Let us call a set of distances $x_1, \ldots, x_n$ a *partition with respect to a past context C and a present credence function P_{C^*}* iff the sentences of the form 'the distance between the present context C^* and the past context C I am updating from is x_i' form a partition with respect to the present credence function P_{C^*} (i.e. if they are mutually exclusive and jointly exhaustive). So, if I know that it is either Monday or Tuesday, and I am considering a context on Sunday, the distances (in days) $x_1 = 1$ and $x_2 = 2$ form a partition with respect to my present credences and the Sunday context. For every distance within a partition, Continuous Conditionalization tells us how to update from the past credence function. If my distance to the Sunday context is one day, and my present evidence is 'Jacky's birthday is today', I need to update my Sunday credence by conditionalizing on 'Jacky's birthday is *tomorrow*'; if the distance is two days, however, I need to conditionalize on 'Jacky's birthday is *the day after tomorrow*'. My present credences should be a weighted average of these two possible ways of updating my Sunday credences.

This proposal can be made precise:

> **(ACC: Approximated Continuous Conditionalization)**
> Let C and C^* be two successive contexts and $x_1, \ldots, x_n$ a partition
> with respect to C and P_{C^*}. Further, let E be the evidence available

---

[5]The analysis of *The Prisoner* by Arntzenius (2003: sec. II) seems to contains a similar idea.

in C^* which was gained after C. Then

$$P_{C^*}(A) = \sum_{i=1}^{n} \left[P_{C^*}(C/C^* = x_i) \times P_C(A[x_i]|E[x_i]) \right].$$

Note that Continuous Conditionalization is a special case of ACC: if one keeps track of the contextual changes, then one knows about one's distance to the context one is updating from, and the relevant partition will consist of a single distance x for which $P_{C^*}(C/C^* = x) = 1$.[6]

In some respects, ACC is structurally similar to Jeffrey's rule of conditionalization. To recall, Jeffrey conditionalization takes the following form: $P_f(A) = P_i(A|E) \times P_f(E) + P_i(A|\neg E) \times P_f(\neg E)$, where P_f represents the *final* credences after updating from the *initial* credences P_i. This rule of updating is a generalization of standard conditionalization: if $P_f(E) = 1$, then $P_f(A) = P_i(A|E)$. Jeffrey conditionalization allows to deal with uncertain evidence. After updating, the evidence E assumes credence $P_f(E)$ which may be strictly less than 1. Jeffrey conditionalization can be described as follows: given uncertain evidence E, our credence in A should be a weighted average of our initial credence conditional on E and our initial credence conditional on $\neg E$ weighted by our final credences in E and $\neg E$ respectively. In fact, Jeffrey's rule applies not only to a sentence E and its negation but to any partition of sentences E_j in general. It then takes the form $P_f(A) = \sum_j P_i(A|E_j) \times P_f(E_j)$. Now, this is structurally similar to what ACC prescribes: if we are uncertain about our distance to the context we are updating from, our credence in A should be a weighted average of the various values determined by these distances conditional on the corresponding evidence weighted by our final credences concerning the relevant distances. However, there are some relevant differences between ACC and Jeffrey's rule. Whereas Jeffrey conditionalizes on the partition forming sentences in the term $P_i(A|E_j)$, ACC does not conditionalize in a similar way. Rather, ACC conditionalizes on the evidence which will in general be distinct from the partition forming sentences. Yet the way the evidence relates to the past credence function one is updating from depends on the partition forming sentences in a systematic way. Also, as far as ACC is concerned, the new evidence always becomes certain after updating: $P_{C^*}(E) = 1$.

---

[6]A note of caution. Principle ACC is meant to describe a canonical way of how indexical belief tends to develop. It should not be expected to be able to deal with phenomena alien to indexical belief. Cp., for instance, the phenomenon of losing evidence due to strong defeaters as described in Williamson (2000: 219).

On an intuitive level, it may not come as a surprise that uncertainty about contextual changes and uncertain observation bear structural similarities to each other. In the case of uncertain evidence E, two ways of updating the initial credences, namely either by E or by $\neg E$, are possible. Similarly, in the case of uncertainty about contextual changes, various ways of updating the initial credences are possible because the epistemic dynamics of indexical belief depends on the relevant contextual changes. In a way, losing track of the contextual changes can be compared to vision getting blurred: by losing track of the relevant contextual features, our initial indexical beliefs fade out of sight. Sometimes we cannot retain our indexical beliefs.

As it stands, ACC is silent about how our credences in the sentences expressing the relevant distances are determined. Let us recall that a sentence of the form 'C/C^*=x' means that the distance between the present context and the past context one is updating from is x. Note that such statements are indexical. Typically, they allow for an equivalent expression by more mundane sentences such as 'Today is Monday' when it is clear that one is updating from a Sunday context. How are our credences in the sentences forming the relevant partition determined?

Could we perhaps simply apply CC which would suggest $P_{C^*}('C/C^*\text{=x'}) = P_C('C/C^*\text{=x'}[\text{x}])$? Unfortunately, CC will not generally be applicable. Since the partition forming sentences are indexical, they are within the realm of sentences which are affected by uncertainty about the relevant contextual changes. For instance, if 'Today is Monday' is part of the partition and I lost track of time, I should not transfer my certainty in 'Today is Sunday' on Sunday to 'Today is Monday' on Monday. Thus, for the same reason why CC does not hold in general, it cannot be applied in the case of finding the right credences in the partition forming sentences.

Even though ACC does not necessarily determine the credences in the partition forming sentences, it constrains them in various ways. Most importantly, if information about the contextual changes is part of the evidence gained by the subject, this will bear directly on the credences in the partition forming sentences. For, as noted above, ACC requires that the newly found evidence is assigned credence 1. Also, there will be cases in which the past credence function contains non-indexical information which is relevant to the expected contextual changes. Thus, ACC imposes some constraints on what our credences in the partition forming sentences should be.

The fact that ACC does not determine the credences in the partition forming sentences may be more of a virtue than a vice. Such credences should reflect the degree to which the subject was able to follow the contextual changes. What is right to think about contextual changes will depend both on the activity of the subject and on what is happening in the subject's environment. For example, whether or not the subject pays attention to the passage of time is relevant to the subject's credences concerning how much time has passed. And sometimes there are external conditions which make it hard to keep track of contextual changes: think, for instance, of the difficulty of spatial orientation in a desert-like environment. Given such dependencies, it seems unlikely that the credences in the partition forming sentences can be calculated as a product of the past credences and the newly found evidence (as long as the credences are not already determined by the new evidence). Rather, the subject's best estimate of how closely it has kept track of the contextual changes may depend on what has happened to the subject while it has undergone the contextual changes.

5   Sleeping Beauty

Recall the story of Sleeping Beauty:

> Researchers at the Experimental Philosophy Laboratory have decided to carry out the following experiment. First they will tell Sleeping Beauty all that I am about to tell you in this paragraph, and they will see to it that she fully believes all she is told. Then on Sunday evening they will put her to sleep. On Monday they will awaken her briefly. At first they will not tell her what day it is, but later they will tell her that it is Monday. Then they will subject her to memory erasure. Perhaps they will again awaken her briefly on Tuesday. Whether they do will depend on the toss of a fair coin: if heads they will awaken her only on Monday, if tails they will awaken her on Tuesday as well. On Wednesday the experiment will be over and she will be allowed to wake up. The two possible brief awakenings during the experiment will be indistinguishable [...]. However, she will be able, and she will be taught how, to distinguish her brief awakenings during the experiment from her Wednesday awakening after the experiment is over [...]. (Lewis 2001: 171)

Let us fix some abbreviations: HEADS is the sentence 'The outcome of the toss is Heads', MONDAY is the sentence 'Today is Monday', TUESDAY is the sentence 'Today is Tuesday', and finally AWAKE is the sentence 'I am awakened today'. Moreover, let P_{Sunday} be Sleeping Beauty's credence function on Sunday, and P_{Monday} her credence function on Monday before she is told that it is

Monday. Thus, the subscripts indicate the relevant contexts Sleeping Beauty is in.

Sleeping Beauty's credences on Sunday are usually taken as evident (we will come back to them below):

(4) $P_{\text{Sunday}}(\text{'Tomorrow is Monday'}) = 1$.

(5) $P_{\text{Sunday}}(\textsc{Heads}) = P_{\text{Sunday}}(\textsc{Heads}|\text{'I am awakened tomorrow'}) = 1/2$.

(6) $P_{\text{Sunday}}(\textsc{Heads}|\text{'I am awakened the day after tomorrow'}) = 0$.

Sleeping Beauty loses a piece of self-locating knowledge between Sunday and Monday. On Sunday, she knows 'Tomorrow is Monday' to be true. On Monday, however, she does not know 'Today is Monday' to be true. This is because Sleeping Beauty cannot keep track of the relevant contextual changes, i.e. she does not know on Monday morning whether one or two days have passed.

Let us therefore see how ACC applies to Sleeping Beauty. Since Sleeping Beauty knows on Monday that it is either Monday or Tuesday, she knows that her distance to the Sunday context is either one or two days; it is one day just in case MONDAY is true, and it is two days just in case TUESDAY is true. So, the distances (in days) $x_1 = 1$ and $x_2 = 2$ form a partition with respect to her credence function on Monday and the Sunday context (we can safely ignore Beauty's spatial position).

When Sleeping Beauty is awakened on Monday, she will know 'I am awakened today' to be true. Thus, AWAKE is a piece of evidence available to Sleeping Beauty within the Monday context (for present purpose, we can leave it open whether it is new or not). Since AWAKE is an indexical sentence, it is affected by the two possible distances which need to be considered. Therefore, when Sleeping Beauty updates her Sunday credences, she has to consider the two sentences:

(7) AWAKE[1]='I am awakened tomorrow'.

(8) AWAKE[2]='I am awakened the day after tomorrow'.

If her distance from the Sunday context is one day, AWAKE expresses what was expressed on Sunday by 'I am awakened tomorrow'. If her distance from the Sunday context is two days, this sentence expresses what was expressed on Sunday by 'I am awakened the day after tomorrow'.

Now, applying ACC yields the following constraint on how Sleeping Beauty's credences on Monday should relate to her Sunday credences:

(9) $P_{\text{Monday}}(\textsc{Heads}) = P_{\text{Monday}}(\textsc{Monday}) \times P_{\text{Sunday}}(\textsc{Heads}|\textsc{Awake}[1]) +$
$P_{\text{Monday}}(\textsc{Tuesday}) \times P_{\text{Sunday}}(\textsc{Heads}|\textsc{Awake}[2]).$

ACC requires that Sleeping Beauty's credence in Heads should equal her credence on Sunday conditional on the evidence available to her on Monday, namely 'I am awakened today'. Since she is uncertain about her distance in time to the Sunday context, this means that she needs to conditionalize her Sunday credence either on 'I am awakened tomorrow' (= Awake[1]) or on 'I am awakened the day after tomorrow' (=Awake[2]). Her actual credence on Monday will then be a weighted average of these two possible ways of updating her Sunday credences.

On the basis of Beauty's Sunday credences (recall (5) and (6)), (9) allows us to narrow down her credence in Heads even further:

(10) $P_{\text{Monday}}(\textsc{Heads}) = P_{\text{Monday}}(\textsc{Monday}) \times 1/2 + P_{\text{Monday}}(\textsc{Tuesday}) \times 0.$

Or simply:

(11) $P_{\text{Monday}}(\textsc{Heads}) = P_{\text{Monday}}(\textsc{Monday}) \times 1/2.$

So, why should her credence in Heads be the product of 1/2 and her credence in Monday? The answer according to ACC is the following. Sleeping Beauty considers her two possible distances to the Sunday context. If the distance is one, i.e. if it is Monday, her credence in Heads should be her Sunday credence in Heads conditional on 'I am awakened tomorrow', which is 1/2. If her distance is two, i.e. if it is Tuesday, her credence in Heads should be her Sunday credence in Heads conditional on 'I am awakened the day after tomorrow', which is 0. So, if it is Monday, she should have credence 1/2 in Heads; if it is Tuesday, her credence should be 0. Consequently, she weighs 1/2 by her credence in Monday.

This already excludes the halfers' solution to the Sleeping Beauty problem. By (11), her Monday credence in Heads can only be 1/2 if her credence in Monday is 1. But since she cannot know upon awakening that it is Monday (so much is uncontroversial), her credence in Monday will be less than 1. Consequently, her credence in Heads will be less than 1/2. For this reason, ACC excludes the halfer's solution because it dictates a change in credence

concerning HEADS given that Sleeping Beauty cannot know on Monday that it is Monday.

We arrive at the thirder's solution if we add to (11) the assumption that Beauty's credence in MONDAY should be 2/3. Then we find

(12) $P_{\mathrm{Monday}}(\mathrm{HEADS}) = 2/3 \times 1/2 = 1/3.$

Why should Beauty's credence in MONDAY be 2/3? Arguably, because in the long run two out of three awakenings would be Monday-awakenings. This information is non-indexical and independent of other indexical information Sleeping Beauty has. Therefore, it is available to Sleeping Beauty on Monday. Given that she has no more information bearing on 'Today is Monday', she should adjust her credence in 'Today is Monday' to 2/3. Since I have nothing of substance to add to the existing literature concerning this part of the thirders' solution, I will leave it here at that.[7]

Halfers (e.g. Lewis 2001 and Jenkins 2005) typically hold that Beauty's credence in HEADS should remain 1/2 as a matter of principle: Since Beauty would not gain any new relevant evidence, her credence in HEADS must not alter. For this reason, the crucial task for a rule of updating which backs up the thirders' position is to explain the change in credence towards HEADS. On the basis of ACC, the following reasoning is available. Sleeping Beauty loses a piece of self-locating knowledge, namely what day it is: On Sunday she knows what is expressed by 'Today is Sunday'; on Monday, however, she does not know what is expressed by 'Yesterday was Sunday'. Due to this loss of self-locating information, she does not know about her distance to the Sunday context she is updating from. As a paradigm of rationality, she does her best in conditionalizing her Sunday credence on the evidence available to her on Monday, namely 'I am awakened today'. If her distance to the Sunday context is one day, i.e. if it is Monday, her evidence is something she knew all along and her credence in HEADS should remain 1/2. However, if her distance to the Sunday context is two days, then 'I am awakened today' expresses what was expressed on Sunday by 'I am awakened the day after tomorrow'. This, on the other hand, would be decisive evidence against HEADS. She weighs these two credences in HEADS by her estimates of the corresponding distances to the Sunday context. As a result, her credence in HEADS decreases.

---

[7]Cp, for instance, Arntzenius (2002), Dorr (2002), Elga (2000), Horgan (2004), and Monton (2002). Hitchcock (2004) gives a dutch-book argument for the thirder's position.

Thus, the explanation for her change in credence consists in the observation that due to the loss of self-locating knowledge, 'I am awakened today' constitutes *potentially* new evidence: with some positive subjective probability, it expresses decisive evidence against HEADS. So, by losing a piece of self-locating information, a new epistemic possibility opens up which would be expressed by 'Today is Tuesday'. If this possibility obtained, 'I am awakened today' would imply 'I am awakened on Tuesday'. Given the newly opened epistemic possibility, 'I am awakened today' becomes relevant to the question of whether the coin landed heads. Although it may be held that 'I am awakened today' does not express *new* relevant evidence, for it expresses, as a matter of fact, what was expressed on Sunday by 'I will be awakened tomorrow', it becomes *relevant* evidence due to the loss of self-locating information: with some positive probability it expresses what was expressed on Sunday by 'I will be awakened the day after tomorrow'.[8] So, the change in credence is not explained through an acquisition of new evidence, but rather through the loss of old evidence possessed at an earlier time. This loss of evidence raises a previously excluded epistemic possibility in the light of which 'I am awakened today' becomes relevant to the question of whether the coin landed heads.[9]

This explanation is similar to the explanation given by Monton (2002). Monton assimilates the Sleeping Beauty problem to the phenomenon of forgetting: Sometimes we simply forget something we knew at a previous time (and sometimes we may even know in advance that we will forget it).[10] It is a difficult question whether Beauty's way of losing knowledge is best described as forgetting. It should be noted, though, that many cases of losing self-locating knowledge are not cases of forgetting.[11] Around midnight I may cease to know which day it is because my inner clock is not precise enough to be reliable in close cases. Think also of situations in which we are unable to keep track of our spatial position because the environment does not offer enough clues for our

---

[8]One may, however, reasonably hold that evidence is more finely individuated than it is assumed in this line of argument. For instance, Horgan (2004) argues that contexts of subjective probability are *intensional*; they do not allow for substitution of co-referential indexical terms *salva veritate*. Cp. also Weintraub (2004). One may even go further and argue that contexts of subjective probability are *hyper-intensional*; for such a view, see Williamson (2006).

[9]Cp. Horgan (2004: 13).

[10]Positively, Kierland & Monton (2005) propose an epistemic goal which can be described as the aim of minimizing the inaccuracy of one's beliefs. However, their proposed method is not univocal and therefore leaves it open how we should update possibly indexical beliefs in particular cases.

[11]Cp. the discussion in Arntzenius (2003).

sense of orientation to be reliable. Intuitively, we would not describe such cases as cases of forgetting.[12] Since Beauty remembers on Monday the credences she assigned to sentences on Sunday but fails to know her distance to the Sunday context, it seems open to me to model the Sleeping Beauty problem as a case of losing self-locating knowledge in a way which is not forgetting but rather the inability of keeping track of the relevant contextual features. It would be interesting to compare this approach to an account essentially based on a model of forgetting.

Some of the crucial aspects of the present position have been criticized by Jenkins (2005).[13] Although Jenkins ultimately follows Lewis (2001) in defending the halfer's solution, she seems to agree to a fairly high extent with the general approach underlying ACC. For instance, she presents Beauty's credences in HEADS and TAILS on Monday as a weighted average of her corresponding credences conditional on it's being either Monday or Tuesday (Jenkins: 199).

The thirders' solution presented above relies on three main ingredients: Sleeping Beauty's credences on Sunday, ACC, and her credences on Monday concerning what day it is. To recapitulate, an application of ACC yielded

(9) $P_{\text{Monday}}(\text{HEADS}) = P_{\text{Monday}}(\text{MONDAY}) \times P_{\text{Sunday}}(\text{HEADS}|\text{AWAKE}[1]) + P_{\text{Monday}}(\text{TUESDAY}) \times P_{\text{Sunday}}(\text{HEADS}|\text{AWAKE}[2])$.

This equation can be simplified considerably by substitution of Beauty's Sunday credences as described above. Based on this, we found

(11) $P_{\text{Monday}}(\text{HEADS}) = P_{\text{Monday}}(\text{MONDAY}) \times 1/2$,

which is already incompatible with the halfers' solution (unless one would be willing to say that Beauty can know on Monday that it is Monday which she intuitively cannot).

Thus, in order to defend the halfer's solution, it is not sufficient to deny a certain assignment to $P_{\text{Monday}}(\text{MONDAY})$. In addition, one will also need to take issue either with ACC or with the assignment of Beauty's Sunday credences. Otherwise one would be committed to (11) which is not part of any sensible halfers' position.

---

[12]As an aside, let me mention at this point that ACC is not supposed to apply to proper cases of forgetting; such cases seem to require a treatment of their own.

[13]For lack of space, I will not discuss more complex rivals to the thirders' position. For instance, Bostrom (2007) offers a hybrid model for self-locating belief by which he attempts to unify the major bits of the thirders' and the halfers' camp (however, at the cost of violating a plausible instance of conditionalization).

Of course, Jenkins is aware of this and offers reasons to resist the argument for (11). However, it is not totally clear whether her defense is based on a rejection of the relevant instance of ACC, i.e. (9) above, or whether she rather disagrees with the assignment of Beauty's Sunday credences. I will consider both possibilities in turn.

There are two Sunday credences which are used in deriving (11) from (9), namely $P_{\text{Sunday}}(\text{HEADS}|\text{AWAKE}[1]) = 1/2$ and $P_{\text{Sunday}}(\text{HEADS}|\text{AWAKE}[2]) = 0$ (i.e. the sentences (5) and (6) above), where $\text{AWAKE}[1]$ is the sentence 'I am awakened tomorrow' and $\text{AWAKE}[2]$ is the sentence 'I am awakened the day after tomorrow'. Since Beauty knows on Sunday that if the coin comes up heads, she will not be awakened the day after tomorrow, i.e. on Tuesday, she knows that if she is awakened the day after tomorrow, the coin could not have come up heads. Hence, conditional on the sentence 'I am awakened the day after tomorrow', she should on Sunday assign credence 0 to the sentence 'The coin comes up heads'. Indeed, nothing in Jenkins's defense of the halfer's solution suggests that she would doubt that the credence in 'The coin comes up heads' on Sunday conditional on the assumption that Beauty is awakened on Tuesday should be 0.

For this reason, the interesting question is whether it is possible to resist the assumption that Beauty's credence on Sunday in the sentence 'The coin comes up heads' conditional on the sentence 'I am awakened tomorrow' should be 1/2. This assignment of credence follows, though, from two plausible assumptions. Firstly, recall that it is part of the Sleeping Beauty scenario that Beauty knows on Sunday that it is Sunday and that she will be awakened on Monday. In effect, it is thus stipulated that Beauty's credence on Sunday in 'I am awakened tomorrow' is 1. Secondly, there does not seem to be any doubt that on Sunday Beauty should think it 1/2-likely that the coin which will be tossed a day later will come up heads. This assumption about Beauty's Sunday credence in 'The coin will come up heads' is undisputed. The disagreement is only about Beauty's Monday credence in HEADS: halfers argue that it should remain 1/2, thirders hold that it should shrink to 1/3. But now the claim that on Sunday Beauty's credence in HEADS conditional on 'I am awakened tomorrow' should be 1/2 already follows. Generally, if A and B are any sentences and P a probability distribution obeying the standard laws, $P(B) = 1$ implies $P(A|B) = P(A)$. In particular, since Beauty's credence on Sunday in 'I am awakened tomorrow' is 1 and her credence in HEADS is 1/2, her conditional credence in HEADS on

'I am awakened tomorrow' will be 1/2, too. Thus, the relevant assignment of Sunday credences is not only plausible in itself, it also follow from what is stipulated about the Sleeping Beauty scenario and an assumption both halfers and thirders are happy to accept.

Given that Beauty's Sunday credences are hard to dispute, it is much more plausible to assume that Jenkins would object to the relevant instance of ACC, namely (9).[14] Part of ACC's motivation in this case stems from the thought that Sleeping Beauty considers the two possible contexts she is in, i.e. that it is Monday and that it is Tuesday: if it is Monday, her credence in HEADS should be 1/2, and if it is Tuesday, her credence should be 0. This is what her Sunday credences demand from which she aims to update. Jenkins objects to the validity of this chain of reasoning. In accordance with Lewis (2001), she thinks that conditional on MONDAY, Sleeping Beauty should have a credence of 2/3 in HEADS. As a consequence, Jenkins recommends that when Sleeping Beauty is told later in the day that it is Monday, she will have a credence of 2/3 in HEADS.[15] Thus, Sleeping Beauty will have a credence of 2/3 that a fair coin still to be tossed will come up heads (recall that the coin is not tossed before Monday night and that Sleeping Beauty knows all this). This is a rather unhappy consequence which is usually considered a strong objection to the halfer's solution.[16] At the very least, it seems illegitimate to rely on it in arguing against a general principle which can be given independent motivation such as ACC.

So far, Beauty's Sunday credences and ACC as it applies to Beauty's credences on Monday have been defended. As we saw, this leads to the requirement that Beauty's credence on Monday in HEADS should be her credence in it's being Monday weighted by 1/2. This result excludes the halfers' solution. But in order to arrive positively at the thirders' solution, one needs to argue further that Beauty's credence on Monday in it's being Monday should be 2/3. As indicated above, with regards to this assumption I do not want to add anything to the considerations already contained in the thirders' arguments (e.g. Arntzenius 2002, Dorr 2002, and Elga 2000). Hence, ACC alone should be seen as a constraint on our self-locating beliefs which rules out the halfers' solution and provides a partial argument for the thirders' position. It can be supplemented

---

[14]Cp. Jenkins (2005: 199).

[15]Cf. Jenkins (2005: 198).

[16]Cp. Elga (2000). Note that this consequence is in apparent conflict with Lewis's (1980) Principal Principle. See Lewis (2001) for a defense and some discussion of the role of the Principal Principle in the Sleeping Beauty case.

by an assumption which will then allow for the derivation of the full-blown thirders' conclusion.

6   Conclusion

Indexical belief requires an extension of the standard account of bayesian updating. Even in the absence of new evidence, continuity of indexical belief depends on appropriate substitutions of the indexical terms we are using. In normal cases, updating on indexical evidence proceeds by conditionalizing the past credences according to the distance between past and present context: after learning 'Today is Jacky's birthday', I should think about 'Today is Monday' what I thought a day before about 'Tomorrow is Monday' conditional on 'Tomorrow is Jacky's birthday'. However, this account needs to be modified even further in order to accommodate cases in which we become uncertain about our spatiotemporal location. In such cases, it seems that our present credences are still related to our past credences in a systematic way. The present proposal suggests that these cases can be described as estimating the possible results of updating according to the form of conditionalization suitable for indexical belief. As an application, the thirders solution to the Sleeping Beauty problem can be described as an instance of a general constraint on our indexical beliefs applying specifically to cases in which we lose track of our location in time and space.

7   Acknowledgements

References

Arntzenius, F. 2002. Reflections on Sleeping Beauty. *Analysis* 62: 53-62.

Arntzenius, F. 2003. Some Problems for Conditionalization and Reflection. *The Journal of Philosophy* 7: 356-370.

Bostrom, N. 2007. Sleeping beauty and self-location: A hybrid model. *Synthese* 157: 59-78.

Dorr, C. 2002. Sleeping Beauty: in defence of Elga. *Analysis* 62: 292-296.

Elga, A. 2000. Self-locating belief and the Sleeping Beauty problem. *Analysis* 60: 143-147.

Hitchcock, C. 2004. Beauty and the Bets. *Synthese* 139: 405-420.

Horgan, T. 2004. Sleeping Beauty awakened: new odds at the dawn of the new day. *Analysis* 64: 10-21.

Jenkins, C. 2005. Sleeping Beauty: A Wake-Up Call. *Philosophia Mathematica* 13: 194-201.

Kierland, B. & Monton, B. 2005. Minimizing Inaccuracy for Self-Locating Beliefs. *Philosophy and Phenomenological Research* LXX: 384-395.

Lewis, D. 1980. A subjectivists guide to objective chance. In *Studies in Inductive Logic and Probability* 2, ed. R. C. Jeffrey, 263-93. Berkeley: University of California Press.

Lewis, D. 2001. Sleeping Beauty: reply to Elga. *Analysis* 61: 171-176.

Monton, B. 2002. Sleeping Beauty and the forgetful Bayesian. *Analysis* 62: 47-53.

Weintraub, R. 2004. Sleeping Beauty: a simple solution. *Analysis* 64: 8-10.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, T. 2006. Indicative versus subjunctive conditionals, congruential versus non-hyperintensional contexts. In Sosa, E. and Villanueva, E. (eds.). *Philosophical Issues* 16: 310-333.