Imaging and Sleeping Beauty
A Case for Double-Halfers

Mikal Cozic
Department of Cognitive Science
Ecole Normale Suprieure
45, rue d'Ulm, F-75005 Paris
mikael.cozic@ens.fr

INTRODUCTION

(Elga 2000) introduced philosophers to the troubling scenario of Sleeping Beauty. On Sunday evening (t_0), Sleeping Beauty is put to sleep by an experimental philosopher. She is awaken on Monday morning and at this moment (t_1), the experimenter doesn't tell her which day it is. Some time later (t_2), she is told that it is actually Monday. At this point, what follows depends on the toss of a fair coin that took place on Sunday evening - Sleeping Beauty is not aware of the outcome. If the coin landed heads ($HEADS$), then Sleeping Beauty is put to sleep until the end of the week. If the coin landed tails ($TAILS$), then Sleeping Beauty is awaken on Tuesday morning. The crucial fact is that a drug that is given to her is such that she cannot distinguish her awakening on Monday from her awakening on Tuesday. Of course, Sleeping Beauty is perfectly informed of every detail of the protocol before the experiment. The question that has drawn so much attention since (Elga 2000) is the following: *what should be Sleeping Beauty's degree of belief that* $HEADS$? Actually, the question will be asked at two different times: at t_1 - when Sleeping Beauty is just awaken on Monday - and t_2 - when Sleeping Beauty has been told that it is Monday. Let us call the first question Q_1 and the second Q_2. In the sequel, P_i ($i \in \{0, 1, 2\}$) will denote Sleeping Beauty's credence at t_i, that is, her beliefs concerning the relevant propositions.

The aim of this paper is to provide a case for the *double-halfer* position, that is, the position according to which Sleeping Beauty's credence should be such that

$$P_1(HEADS) = P_2(HEADS) = 1/2$$

The double-halfer position is not new.[1] My case for it is based on the so-called *imaging rule* for probabilistic change. In what follows, I will try to argue, first, that this rule should be used by Sleeping Beauty and, second, that if it used it leads to the double-halfer position.

[1] See for instance (Meacham 2005) and (Bostrom 2006).

1 HALFERS AND THIRDERS

Let's begin with question Q_1: what should be the value of $P_1(HEADS)$? There are basically two camps: the halfers and the thirders. The thirders claim (following (Elga 2000)) that $P_1(HEADS) = 1/3$ whereas the halfers claim (following (Lewis 2001)) that $P_1(HEADS) = 1/2$. Now, the answer to Q_1 is intimately linked to the answer to Q_2. As a consequence, the two positions are best described by giving their answer to both questions. By conditionalization, one obtains $P_2(HEADS) = 1/2$ for the thirders and $P_2(HEADS) = 2/3$ for the halfers. We can sum up the positions of Lewis and Elga as follows :

	A. Elga	D. Lewis
Q1	1/3	1/2
Q2	1/2	2/3

Let's turn to the arguments. Here, I follow essentially Lewis' reconstruction of the disagreement. First, it is supposed that the underlying state space W contains three (so-called centered) worlds: $W = \{HM, TM, TT\}$ where

- in HM the coin lands heads and it's monday
- in TM the coin lands tails and it's monday
- in TT the coin lands tails and it's tuesday

W is supposed to be the relevant state space because each state of W solves all the uncertainties of Sleeping Beauty - both her temporal location and the outcome of the toss. Some propositions are, according to Lewis, "common ground" between him and Elga. Here are the most important[2]:

[2] I follow Lewis's notation. I skip propositions (3) and (4): proposition (3) unfolds proposition (2) and proposition (4) essentially equates $HEADS$ with $\{HM\}$ and $TAILS$ with $\{TM, TT\}$.

(1) $P_1(TM) = P_1(TT)$

(2) $P_2(HEADS) = P_1(HEADS|MONDAY)$

$= P_1(HEADS|\{HM, TM\})$

(5) $P_0(HEADS) = P_0(TAILS) = 1/2$

(1) is a form of the Indifference or Laplacean Principle reflecting the fact that Sleeping Beauty cannot distinguish at her awaken between Monday and Tuesday. (2) says that between t_1 and t_2, Sleeping Beauty changes her credences by conditionalization. (5) expresses the fact that at t_0 Sleeping Beauty's credence obeys to the "objective probability" of the coin landing heads or tails.

Elga's starting point is that the coin could perfectly be tossed on Monday night. If one accepts this, then, still by endorsement of objective probability, Sleeping Beauty should believe that the probability of $HEADS$ is 1/2 after she has learned that it's Monday. That is, according to Elga:

(E) $P_2(HEADS) = 1/2$

From (E) and the common ground (including crucially the rule of conditioning expressed by (2)), one has to conclude that $P_1(HEADS) = 1/3$. Elga's argument is a kind of *bottum-up* argument which starts from an answer to $Q2$ to give an answer to $Q1$.

On the opposite, Lewis provides a direct answer to $Q1$ and infers from it an answer to $Q2$. Lewis's premiss is (roughly)[3] the following one:

(L) $P_1(HEADS) = P_0(HEADS)$

Still, from (L) and the common ground (including crucially the rule of conditioning expressed by (2)), one has to conclude that $P_2(HEADS) = 2/3$.

A point stressed by Lewis is that both arguments conclude that $P_1(HEADS) < P_2(HEADS)$ - more precisely, that $P_2(HEADS) = P_1(HEADS) + 1/6$. This is a direct consequence of the fact that halfers and thirders are committed to conditionalization to go from P_1 to P_2. In the current setting, a double-halfer position and any position according to which $P_1(HEADS) = P_2(HEADS)$ are excluded.

Both Elga's and Lewis's basic intuitions are appealing. Elga's intuition is that the coin could be tossed on Monday night and that in this case, one should endorse the objective probability of $HEADS$ as her credence. Lewis's intuition is that on Monday morning, there is no new evidence that is relevant to the credence concerning $HEADS$. Therefore the credence toward $HEADS$ at t_1 should remain the same as at t_0. What is clear from the remarks above is that, *given the common ground between Elga and Lewis, these intuitions cannot be reconciled*. As a consequence, someone who finds both intuitions appealing (and accordingly who accepts both (E) and (L)) faces the following dilemma: either to give up one of the intuitions, or to give up part of the common ground.

2 CONDITIONING AND IMAGING

I will put into question neither proposition (1) nor proposition (5) but rather proposition (2), namely the use of conditionalization to go from P_1 to P_2. Let's note first that what is learned at t_2 by Sleeping Beauty ("it is Monday") is a context-sensitive information. Importantly, context-sensitive propositions are *in general* problematic for conditionalization. To be more precise, there are two central properties of conditionalization that are problematic: concentration and partiality. (i) Concentration is the fact that the beliefs of an agent who conditionalizes become more and more concentrated as she learns more and more information. Each time a non-trivial information[4] compatible with the initial probability[5] is learnt, the support of the posterior probability distribution is strictly included in the support of the initial probability. This implies preservation (Grdenfors 1988), namely that if a proposition A is believed with certainty then after having learned any information compatible with the initial beliefs, A is still believed with certainty. And this implies that if a proposition has null probability, its probability will remain null whatever information compatible with the initial probability is learnt. (ii) Partiality is the fact that when an information is incompatible with the agent's initial beliefs, the new probability distribution is undefined. These issues are general, but they give us *prima facie* reasons to look more carefully at the use of conditionalization in Sleeping Beauty's scenario.[6]

[3] As a matter of fact, Lewis's premiss is that "only new evidence, centered or uncentered, produces a change in credence; and the evidence [$\{HM, TM, TT\}$] is not relevant to HEADS versus TAILS." (Lewis 2001)

[4] That is, an information that excludes at least one of the world in the support of the initial probability distribution.

[5] That is, an information whose intersection with the support of the initial probability is not empty.

[6] For detailed discussions, see (Arntzenius 2003) and (Meacham 2005).

Conditionalization is often viewed as the only reasonable rule for changing one's credence[7]. Other rules are conceivable, however. Consider for instance the *imaging rule* introduced by (Lewis 1976) as the rule that matches Stalnaker's conditional. The basic idea is this. For each world w and each proposition A, w_A is the closest world to w where A is true.[8] Suppose that the agent is informed that A holds. In the case of conditionalization, all the weights of the $\overline{A}$-worlds are allocated to A-worlds compatible with the prior in a way that preserves the relative probabilities. In the case of imaging, the weight of a $\overline{A}$-world w is exclusively allocated to the world w_A. The rule of imaging is therefore the following:

$$P^{Im(A)}(w) = \sum_{\{w' \in W : w = w'_A\}} P(w')$$

In other words, the probability of w after imaging on A is the sum of the probabilities of the worlds w' such that w is the closest world to w' where A is true. As stressed by Lewis, imaging satisfies a form of minimality: there is "no gratuitous movement of probability from worlds to dissimilar worlds" (Lewis 1976). Here is an example that is intended to illustrate the divergent behavior of conditionalization and imaging:

Exemple 1 (Apple & Banana, partial beliefs) *A basket may contain an apple and a banana. There are four possible states :* $\{AB, A\neg B, \neg AB, \neg A\neg B\}$*:*

AB	$A\neg B$
$\neg AB$	$\neg A\neg B$

Suppose then that the initial probability, P*, is such that the agent is certain that there is at least one fruit in the basket and that the same weight is allocated to the remaining states:*

1/3	*1/3*
1/3	*0*

The agent receives the following information: $I = \{A\neg B, \neg A\neg B\}$*, that is, there is no banana in the basket. If the agent relies on conditionalization, her new belief should be this:*

0	*1*
0	*0*

But if the agent relies on imaging with $AB_I = A\neg B$ *and* $\neg AB_I = \neg A\neg B$*, one obtains this:*

0	*2/3*
0	*1/3*

3 IMAGING AND SLEEPING BEAUTY

As one would expect, the debate between halfers and thirders is dramatically transformed if one adopts a rule of belief change that is different from conditionalization. Let see what happens, for instance, if one relies on imaging. To apply the imaging rule, one needs first to make some assumption on the similarity between worlds. In the case of Sleeping Beauty, the information that Sleeping Beauty learns at t_2 ("it is Monday") excludes one world from P_1's support, namely the world TT. Therefore, the only parameter that has to be specified is the closest world to TT where it is true that it is Monday. I think it is a rather natural assumption to suppose that TM *is* the closest world to TT where it is true that it is Monday. Granting this assumption, the imaging rule is easily applied to Sleeping Beauty's scenario.

As I said before, I consider both Elga and Lewis's basic intuitions as attractive. Let's start from Lewis premiss (L) and the rest of the common ground (propositions (1) and (5)). If one relies on imaging, then $P_2(TM) = P_1^{Im(MONDAY)}(TM) = P_1(TM) + P_1(TT) = 1/2$ and $P_2(HEADS) = P_2(HM) = P_1^{Im(MONDAY)}(HM) = P_1(HM) = 1/2$. In other words from the Lewisian premiss (L) there results a *double-halfer* position: the credence of Sleeping Beauty toward $HEADS$ is the same at t_1 and t_2, namely 1/2. But we could start from Elga's intuitions as well and suppose that $P_2(HEADS) = P_2(TAILS) = 1/2$. Now, if one "backtracks" the imaging rule in the same way one "backtracks" conditionalization in Elga's original argument, one obtains $P_1(HEADS) = P_2(HEADS) = 1/2$.

What this shows is that *if* one starts either from Elga's or from Lewis's basic intuition and that one relies on the imaging rule rather than on conditionalization, *then* one obtains the double-halfer position. But what this does not show is that one should rely on imaging rather than on conditionalization. At this point, the crucial issue is to adjudicate between several rules of belief change.

4 REVISING AND UPDATING

For more than two decades, formal epistemology has developed rules of *full* belief change. It has been convincingly argued by (Katsuno & Mendelzon 1992) that one should carefully distinguish two kinds of belief change contexts: contexts of *revising* where the agent learns an information about an environment that is supposed to be stable and contexts of *updating* where the agent learns an information about a potential change in her environment. If, for instance, the agent has beliefs concerning the content of a

[7] The diachronic Dutch Book argument is the main justification for this belief.

[8] To be sure, it is not an assumption that is kept in Lewis' own semantics of conditionals. Lewis factorizes this assumption into the Limit Assumption and the Uniqueness Assumption and rejects both.

basket of fruits that may or may not contain an apple and that may or may not contain a banana, a revising information could be that there is no banana in the basket and an updating information could be that there is no more banana in the basket (if there was any). The point is that rules of belief change have to be different is these two kinds of contexts. In a revising context, the new belief set given an information that is compatible with it has to be included in the initial belief set[9] whereas in an updating context, the new belief may not be included in the initial belief set[10]. This results in two kinds of rationality postulates: the so-called AGM-postulates for belief *revision* (Grdenfors 1988) and the KM-postulates (Katsuno & Mendelzon 1992) for belief *updating*. This is illustrated by the following example:

Exemple 2 (Apple & Banana, full beliefs) *A basket may contain an apple and a banana. There are four possible states :* $\{AB, A\neg B, \neg AB, \neg A\neg B\}$. *Suppose the agent believes initially that there is at least one fruit in the basket i.e.* $K = \{AB, A\neg B, \neg AB\}$*:*

$$\begin{array}{|cc|}\hline AB & A\neg B \\ \neg AB & \\ \hline\end{array}$$

Then the agent believes a revising message according to which there is no banana in the basket. The new belief set will be: $K^r = \{A\neg B\}$. *But suppose the agent is informed that something has happened such that if there was a banana in the basket, it is no more in it. In this case, it is much more intuitive to reason in the following way: if the true world was* AB*, then it is now* $A\neg B$ *; if it was* $A\neg B$*, it is unchanged ; and if it was* $\neg AB$*, then it is now* $\neg A\neg B$*. Therefore one would obtain as a new belief set* $K^u = \{A\neg B, \neg A\neg B\}$ *which differs from* K^r*. To sum up:*

revising

"there is no banana"

$$\begin{array}{|c|}\hline A\neg B \\ \hline\end{array}$$

updating

"there is no more banana (if there was any)"

$$\begin{array}{|c|}\hline A\neg B \\ \neg A\neg B \\ \hline\end{array}$$

The Sleeping Beauty scenario involves rules of *partial* belief change. A natural question would then be the following: if one accepts the distinction between revising and updating contexts (as I do), what are the corresponding rules of partial belief change? The question was left unanswered until recently. But (Walliser & Zwirn 2002) have shown the following result, which is at the very core of my argument: conditionalization-like change rules may be derived from probabilistic transcription of AGM-postulates for belief revision whereas imaging-like change rules may be derived from probabilistic transcription of KM-postulates. This result can be interpreted in the following way: if one is guided by rationality postulates of full belief change, then, in a revising context one should rely on conditionalization whereas in an updating context one should rely on imaging.

To sum up my argument: in the previous section I have argued that if one starts either from Elga's or Lewis' basic intuitions and that one relies on imaging, then one obtains the double-halfer position. In the current section, I have argued that if the context of belief change is an updating context, then one should rely on imaging. It remains to be argued that when Sleeping Beauty learns that it is monday (at t_2), it is indeed an updating context, and not a context of updating.

5 UPDATING AND SLEEPING BEAUTY

An updating context (of belief change) is a context where an agent is informed about a potential change of her situation. Now, in so far as in Sleeping Beauty's scenario we consider centered worlds, an information bearing on a change of temporal location is an information about a change of Sleeping Beauty's situation. And it is precisely such an information that the experimenter provides to Sleeping Beauty at t_2. Therefore, it seems that this is an updating context.

But if one looks more carefully at the exact timing of information in the Sleeping Beauty scenario, things are much less clear that they appear to be. As a matter of fact, when Sleeping Beauty becomes aware at t_1 (at her awakening on Monday) that she is on Monday or Tuesday, this is a true updating context since the day it is is different from the day it was at t_0. But when she learns that is it Monday (at t_2), the information does not bear on a change that took place between t_1 and t_2. At t_1, Sleeping Beauty becomes aware that the actual (centered world) is among $I_1 = \{HM, TM, TT\}$. At t_2, the information that is given to her allows her to *refine* her beliefs since she learns $I_2 = \{HM, TM\} \subset I_1$. From this point of view, the information provided at t_2 seems to be a revising context. On the other hand, I_2 is a refinement of an updating-type information, namely I_1. This issue shows that the distinction between updating and revising contexts is *underspecified* and raises quite a general question: when an agent learns successively two informations at t_1 and t_2, which both bear on a change that took place between t_0 and t_1, should we view the second information as a revising or as an updating context?

[9] In the same way that, after conditionalization, the support of the new probability distribution is included in the support of the initial one, if the information is compatible with the latter.

[10] In the same way that, after imaging, the support of the new probability distribution may not be included in the support of the initial one, even if the information is compatible with the latter.

Note that this question is crucial for the double-halfer position: if the information that is provided to Sleeping Beauty at t_2 has to be considered as a revising context, then our case for double-halfers collapses. I won't provide a general answer to this question but I will exhibit an example with a similar structure, in particular where the agent receives two informations at two different times, and where it is more intuitive to handle the second information by updating.

Exemple 3 (Apple, Banana & Coconut) *A basket may contain three fruits: an apple, a banana and a coconut. There are eight possible worlds:*

ABC	$AB\neg C$	$A\neg BC$	$A\neg B\neg C$
$\neg ABC$	$\neg AB\neg C$	$\neg A\neg BC$	$\neg A\neg B\neg C$

Suppose first that the agent's initial beliefs can be represented by the following probability distribution P_0:

0	*1/4*	*1/4*	*1/4*
1/4	*0*	*0*	*0*

It happens that between t_0 and t_1, if there was a banana it has been removed and if there was a coconut it has been removed as well. But suppose that at t_1 the agent learns only that there is no more banana ($I_1 = \{A\neg BC, A\neg B\neg C, \neg A\neg BC, \neg A\neg B\neg C\}$) and learns at t_2 that there is no more banana and no more coconut ($I_2 = \{A\neg B\neg C, \neg A\neg B\neg C\}$). The shift from P_0 to P_1 is clearly an updating context, therefore P_1 should be $P_0^{Im(I_1)}$ i.e.:

0	*0*	*1/4*	*1/2*
0	*0*	*1/4*	*0*

Now the question is: how should the agent handle the information I_2? If he still uses the imaging rule, he will obtain for $P_2 = P_1^{Im(I_2)}$:

0	*0*	*0*	*3/4*
0	*0*	*0*	*1/4*

Note that this is what the agent would have obtained had he directly known I_2 (i.e. $P_0^{Im(I_2)} = P_1^{Im(I_2)}$). If the agent uses conditionalization, he will obtain for $P_2 = P_1^{Cond(I_2)}$:

0	*0*	*0*	*1*
0	*0*	*0*	*0*

Note that the agent would have obtained the same result had he applied conditionalization on P_0 with I_2 (i.e. $P_0^{Cond(I_2)} = P_1^{Cond(I_2)}$).

What lesson can we draw from this example? If someone is convinced that for an updating message it is appropriate to use an update rule, then $P_1^{Im(I_2)}$ is much more intuitive than $P_1^{Cond(I_2)}$. Apple, Banana & Coconut bears some similarity with Sleeping Beauty: (a) the relevant change in the world takes place between t_0 and t_1; (b) what the agent learns at t_1 and t_2 bears on the change in the world that has taken place between t_0 and t_1; and (c) the second information is a refinement of the first ($I_2 \subset I_1$). As a consequence, the example provides some support to the basic claim of the present section, namely that the information received by Sleeping Beauty at t_2 should be viewed as an updating context.

CONCLUSION

Imaging provides a way to support the double-halfer position, which may be viewed as a reconciliation of Elga and Lewis. Note that the use of the imaging rule in the Sleeping Beauty scenario rests on the same fundamental assumption as the one that underlies both Elga's and Lewis' arguments, namely that information about one's temporal location has to be treated in the same way as any other kind of information. To rigorously assess this assumption, one would need to make explicit the structural role of temporal factors in rules of belief change but this I leave for future investigation.

Acknowledgments

I would like to thank for their comments J. Baratgin, D. Bonnay, T. Daniels, I. Drouet, P. Egr, Th. Martin, B. Walliser, D. Zwirn and audiences from "Probability, Decision, Uncertainty" (IHPST, Paris), "Paris-Amsterdam Logic Meeting for Young Researchers (ILLC, Amsterdam) and the "Seminar on Belief Dynamics" (Dept. of Philosophy, Lille III). To my knowledge, the first to establish some connection between Sleeping Beauty and updating was J. Baratgin. Note that, even if I rely strongly on theoretical results by Walliser & Zwirn, their view of Sleeping Beauty is different of mine.

References

Arntzenius, F. (2003), 'Some problems for conditonalization and reflection', *Journal of Philosophy* **100**(7), 356–71.

Bostrom, N. (2006), 'Sleeping beauty and self-location: A hybrid model', *Synthese* . forthcoming.

Elga, A. (2000), 'Self-locating belief and the sleeping beauty problem', *Analysis* **60**(2), 143–7.

Grdenfors, P. (1988), *Knowledge in Flux*, Bradford Books, MIT Press, Cambridge, Mass.

Katsuno, A. & Mendelzon, A. (1992), On the difference between updating a knowledge base and revising it, *in*

P. Grdenfors, ed., ‘Belief Revision’, Cambridge UP, Cambridge, pp. 183–203.

Lewis, D. (1976), ‘Probabilities of Conditionals and Conditional Probabilities’, *The Philosophical Review* **LXXXV**(3), 297–315.

Lewis, D. (2001), ‘Sleeping Beauty: a Reply to Elga’, *Analysis* **61**, 171–6.

Meacham, C. (2005), Sleeping beauty and the dynamics of de se beliefs. Manuscript, http://philsci-archive.pitt.edu/archive/00002526/.

Walliser, B. & Zwirn, D. (2002), ‘Can bayes’ rule be justified by cognitive rationality principles’, *Theory and Decision* **53**.