On Sleeping Beauty Controversy

Lev Vaidman[a,b] and Simon Saunders[c]

July 17, 2001

[a] Centre for Quantum Computation
Department of Physics, University of Oxford,
Clarendon Laboratory, Parks Road, Oxford OX1 3PU, England.

[b] School of Physics and Astronomy
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel-Aviv University, Tel-Aviv 69978, Israel

[c] Faculty of Philosophy, University of Oxford,
10 Merton St. , Oxford OX1 4JJ, England.

Lewis's comment (2001) on Elga's paper (2000) has far reaching consequences. If Lewis is right, then the approach to credence for an event as the value of an "intelligent bet" on this event (e.g. Sklar 1993) does not have universal applicability. The betting approach to this question (Aumann et al. 1997) leads to Elga's result which Lewis contests. We believe, however, that Lewis's approach is untenable, and thus the universality of the betting approach to probability has not been breached.

Consider the following three experiments, the first of which is Elga's. We claim the credence of Beauty is the same in each case.

(i) Beauty sleeps during the week. On Sunday a fair coin is tossed. If the result is Heads (H), Beauty is awaken on Monday only, if the result is Tails (T), Beauty is awaken on Monday and Tuesday. On every awakening, Beauty is told nothing, but must answer the question: "What is your credence for the coin to be H?" After the conversation her memory of the awakening is erased and she sleeps again.

(ii) Beauty sleeps for 100 years, uninterrupted save for awakenings according to procedure (i), which take place every week, following a coin toss on each Sunday of each week.

(iii) Beauty sleeps for 100 years, uninterrupted save for 7827 awakenings. Using a classical random number generator to determine the order, on 2609 awakenings a coin is placed H, and on 5218 awakenings a coin is placed T. On each awakening, the same procedure is followed as in (i).

Experiment (ii) is not just a repetition of experiment (i) 5218 times. In the latter, on every awakening, Beauty knows which week it is. In (ii) Beauty does not know which week it is and, therefore, she does not know which coin in the sequence the question is about. However, since all the coins are fair, Beauty's situation is the same in all weeks and thus the lack of information as to which week it is cannot make a difference to her credence.

Experiments (ii) and (iii) are also not identical. The probability in experiment (ii) that one will obtain exactly 2609 H is very small. However, the probability that this number will be different from 2609 by more than, say, 100, is also very small. Therefore, the probability that the relative frequency of H will be significantly different from 1:2 is negligible. Thus, although we do not have an argument according to which Beauty has to give exactly the same answer in (ii) and (iii), we can argue that these answers cannot differ significantly. Since our job is to decide between credences 1/2 (Lewis) and 1/3 (Elga), this is enough.

In experiment (iii) it is obvious that Beauty should give credence 1/3 for H. If there is no significant difference in Beauty's answer in case (iii) and (ii), and no significant difference in her answer in case (ii) and (i), then in Elga's experiment Beauty should give the answer 1/3, and not 1/2 as Lewis claims.

There is a conceptual difference between (i) and (ii) (and a similar difference between (i) and (iii)). In (i) the Beauty was asked a question about an uncentered proposition: "What is the state of the coin?" In contrast, in (ii) Beauty was asked a question about a centered proposition: "What is the state of the coin of *this* week?" The unusual feature of Elga's experiment, that Beauty must alter her credence about an uncentered proposition with no new uncentered evidence, is not present here. But this does not alter the fact that Beauty must give the same answer in case (ii) and (i).

Apart from the small statistical difference between (ii) and (iii) already mentioned, the two differ in another aspect. In (ii) there is something which corresponds to Lewis's number 1/2: in half of the weeks of Beauty's sleep the coin is H, and in half it is T. In contrast, in experiment (iii), nothing corresponds to 1/2. However, knowledge of this statistical structure to her string of awakenings, in case (ii), is not knowledge that Beauty can use, since never on awakening does she learn where in the string she is located.

In order to see how the difference between the one-week experiment and the many-weeks experiment arises, and in order to show the inconsistency of Lewis's approach, consider an experiment of the kind (ii) but limited to two weeks. A fair coin is tossed twice. The credence of Beauty $p(H)$ can be calculated as the sum of conditional probabilities on the outcomes of the coin tosses:

$$p(H) = p(H|HH)p(HH)+p(H|TT)p(TT)+p(H|HT)p(HT)+p(H|TH)p(TH) \tag{1}$$

It is uncontroversial that $p(H|HH) = 1$ and $p(H|TT) = 0$. Given that one of the outcomes is H and another T, Beauty knows that there are three awakenings: one H and two T. Therefore, the conditional credences of Beauty for these cases are $p(H|HT) = p(H|TH) = \frac{1}{3}$. According to Lewis, Beauty has equal credence for all possible outcomes of the coin tosses: $p(HH) = p(TT) = p(HT) = p(TH) = \frac{1}{4}$. It then follows from (1) that Beauty's credence for H on awakening during the two weeks is $p(H) = \frac{5}{12}$. This is in contradiction with the assumption that there should be no change between Elga's one-week experiment and the similar two-week experiment. Therefore, unless Lewis rejects this very natural assumption, his approach is inconsistent.

On the analysis that we favour there is no such difficulty. On awakening, Beauty's credences for the four outcomes of the coins tosses should not be identical, they should be weighted according to the number of awakening corresponding to these outcomes. Thus $p(HH) = \frac{1}{6}$, $p(TT) = \frac{1}{3}$, and $p(HT) = p(TH) = \frac{1}{4}$. Using (1), it follows that $p(H) = \frac{1}{3}$, just as in the one-week experiment in accordance with Elga's argument.

It is a pleasure to thank Adam Elga for very illuminating comments. This research was supported in part by grant 471/98 of the Israel Basic Research Foundation and the EPSRC grant GR/N33058.

REFERENCES

Aumann, R. J., Hart, S., and Perry, M. 1997 The forgetful passenger. *Games and Economic Behaviour* **20**: 117-120.

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

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

Sklar, L. 1993 *Physics and Chance : Philosophical Issues in the Foundations of Statistical Mechanics* Cambridge : Cambridge University Press, 1993.