Sleeping Beauty goes to the lab: The psychology of self-locating evidence

Matteo Colombo

Tilburg Center for Logic, Ethics, and Philosophy of Science

Tilburg University

Jun Lai

Tilburg Center for Logic, Ethics, and Philosophy of Science

Tilburg University

Vincenzo Crupi

Center for Logic, Language, and Cognition

University of Turin

Abstract. The Sleeping Beauty Problem is a challenging puzzle in probabilistic reasoning, which has attracted enormous attention and still fosters ongoing debate. The problem goes as follows: Suppose that some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking. When you are first awakened, to what degree should you believe that the outcome of the coin toss is Heads? Theoretically, the two candidate answers are 1/2 and 1/3, the proponents of which are known as halfers and thirders, respectively. The present study examines for the first time the descriptive adequacy of both halfers' and thirders' analyses. Our results show that naïve reasoning does not simply fit either. Instead, they suggest that any psychologically adequate analysis of the Sleeping Beauty Problem should take account that the impact on probabilistic reasoning of information about one's spatio-temporal location in the world is systematically discounted.

Keywords: sleeping beauty problem; probability; reasoning; self-locating evidence.

Sleeping Beauty goes to the lab: The psychology of self-locating evidence

Introduction

The Sleeping Beauty Problem (SBP) is a challenging puzzle in probabilistic reasoning. In its standard formulation the problem goes as follows:

> On a Sunday, some researchers are going to put you to sleep. During the two days that your sleep will last, they will briefly wake you up either once or twice, depending on the toss of a fair coin (Heads: once; Tails: twice). After each waking, they will put you back to sleep with a drug that makes you forget that waking. When you are awakened, to what degree ought you believe that the outcome of the coin toss is Heads? (Elga, 2000)

Opinions on this puzzle are split between two camps. For so-called *halfers*, your credence in heads should be 1/2 on the probability scale. At the outset, you know all the details of the experiment, including that the coin is fair and that you will lose your memory of an earlier awakening. When you wake up, all new information you have is this: 'I am presently undergoing either a Monday awakening or a Tuesday awakening.' This type of information is often called "self-locating" information, and concerns one's spatio-temporal location in the world. Self-locating information bears no relevant relation to the outcome of the coin flip, according to the halfer. Since before the experiment you know that the coin is fair, you should then retain a credence of *P*(Heads) = *P*(Tails) = 1/2 (Lewis, 2001).

So-called *thirders* disagree and say your credence in heads should be 1/3 on the probability scale. If the Sleeping Beauty experiment were repeated many times, then, in the long run, about 1/3 of the total number of awakenings would happen on trials where the coin lands heads. Since credences should match long run relative frequencies, your credences should be *P*(Heads) = 1/3 and *P*(Tails) = 2/3 on any particular awakening (Elga, 2000).

The SBP has attracted enormous attention. It raises questions of unsuspected theoretical relevance for the foundations of probabilistic reasoning, belief update, decision-making, and beyond (Piccione & Rubinstein, 1997; Titelbaum, 2013). With the studies reported here, we test the descriptive adequacy of the standard halfer and thirder accounts: Does naïve uncertain reasoning comply with the former, the latter, or display yet some other pattern? In particular, we are interested in two questions: Do naïve reasoners acknowledge self-locating information as relevant in the SBP? If they do, does the quantitative impact of self-locating information get discounted as compared to the impact of statistical information like the outcome of urn draws?

Study 1

Method. Two hundred and thirty-two participants (Mean age, 35.88, SD = 10.14, male 135, female 97) were recruited using Amazon MTurk and randomly assigned to one of four possible experimental groups. As explained below, four conditions were sufficient to disentangle relevant predictions from standard halfer and thirder accounts, thus putting them to empirical test.

Participants read one version of the SBP, and were asked to express their belief about the outcome of the coin toss described in the vignette. Answers were collected on a 7-point Likert-scale ranging from 'Certain that it was Heads and not Tails' to 'Certain that it was Tails and not Heads.'

Across the four groups we manipulated the type of evidence available to participants. We used an adapted version of Elga's (2000) as *Basic* condition. In the *No Evidence* condition, only one waking was said to occur regardless of the outcome of the coin toss. In the *Plus* condition, multiple awakenings (five) would happen after Tails. Finally, in the *Ball* condition, the available evidence only consisted of draws from an urn (see Appendix for the stimuli we used).

Halfers and thirders agree on their predictions that P (Tails) = 1/2 in the *No Evidence* condition, and that P (Tails) = 5/6 in the *Ball* condition. For the *Basic* and *Plus* conditions, instead, halfers and thirders disagree. Halfers predict that P (Tails) = 1/2 in both the *Basic* and the *Plus* condition. Instead, thirders predict that P (Tails) = 2/3 in the *Basic* condition, and that P (Tails) = 5/6 in the *Plus* condition.

Results. A Kruskal-Wallis test showed that each of our four manipulations (Table 1) had a significant effect on participants' judgment, H (3) = 22.73, p = .000, r = .3. Across conditions, we also found significant differences concerning the degree of certainty that the outcome of the coin toss was Tails (ranging in 4-7, i.e. from "equally likely" to "certain"), H (3) = 41.19, p = .000, r = .4.

Table 1. Mean scores for each group (scores ranging from 1 to 7).

Conditions	**No Evidence**		**Basic**		**Plus**		**Ball**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Coin (Study 1)	3.83	.78	3.98	1.05	4.40	1.15	4.67	1.46
Pill (Study 2)	3.92	1.04	3.90	1.28	4.23	1.71	4.67	1.33
Combined (Study 1&2)	3.87	.91	3.94	1.17	4.32	1.46	4.67	1.39

Study 2

Method. Study 1 revealed that participants' judgments of the SBP depended on the type of evidence available. In particular, its results are consistent with the idea that naïve reasoners acknowledge self-locating information as relevant in the SBP (Plus condition). Study 2 examined whether these results may have been affected by a focus on the coin mechanism in the question participants were asked.

A new sample of two hundred and twenty-nine participants (Mean age, 33.82, SD = 10.82, male 131, female 98) was recruited from MTurk, and randomly assigned to one of four possible experimental groups as in Study 1. Unlike in Study 1, participants did not express their belief about the outcome of the coin toss. Instead, participants expressed their belief about the pill they were administered in the situation they were asked to consider (see Appendix). Responses were again collected on a 7-point Likert-scale.

Results. A Kruskal-Wallis test showed that all groups differed significantly in their answers, *H* (3) = 11.64, *p* = .009, *r* = .3. All groups also differed significantly in their certainty of a Tails outcome, *H* (3) = 14.80, *p* = .002, *r* = .3.

A Mann-Whitney Test showed that there was no significant difference between the answers of the participants of this study (M= 4.19) and the answers of participants from Study 1 (M = 4.07), *p* = .24. Aggregating data from both studies, the difference between the Basic and the NoEvidence condition did not reach significance, *p* = .42. However, a significant difference was found between the Plus and the Basic condition, *p* = .03, *d* = .29, and between the Ball and the Plus condition, *p* = .03, *d* = .25.

Discussion

Our results show that naïve reasoning in the SBP does not simply fit either the halfer or the thirder analyses. The halfer's analysis squares with the lack of a significant difference between the *Basic* and *No Evidence* conditions, but is at odds with our finding that the *Plus* and *Basic* conditions reliably differed (for the halfer, one should have P(Tails) = 1/2 in both cases). The thirder's analysis, on the other hand, is supported by the latter result, but clashes with our finding that the probability of Tails is reliably judged to be higher in the *Ball* than in the *Plus* condition (for the thirder, one should have P(Tails) = 5/6 in both cases).

Given that no standard theoretical analysis accounts for observed behavior, one might be tempted to complement a thirder framework with an appeal to cognitive limitations akin to those arising in other known puzzles of probabilistic reasoning. In particular, in a thirder perspective, the SBP may seem structurally similar to the Monty Hall problem. And Monty Hall is known to invite 1/2 as a largely dominant response because of the representational and computational difficulty of the task for the unaided human mind (Krauss & Wang, 2003).

Although initially appealing, this remark is not sufficient to explain our result. In fact, a thirder would be compelled to see the same kind of structural mathematical analogy between the Ball and the Plus variants in our experiment. However, this did not prevent participants to shift significantly towards the Tails hypothesis in the latter condition. For the same reason, a thirder would also be unable to account for our data by relying on a general tendency to conservatism in probability updating (Edwards, 1968; Tentori *et al.*, 2007).

In summary, our results are consistent with a pattern of judgment that is qualitatively different from either the halfer or thirder analyses, where self-locating evidence is acknowledged as relevant but its quantitative impact is largely discounted as compared to more standard statistical evidence as the outcome of urn draws. Other factors were previously shown to have such diluting effects on reasoning with evidence, such as second-order uncertainty about the values of a relevant statistical distribution (Tentori, Crupi, and Osherson, 2010). Although mixed or integrated models of the SBP exist (Bostrom, 2007; Cisewski *et al.*, 2015), this particular diluting effect is out of their reach so far, and should therefore be integrated in a satisfactory descriptive account of reasoning with self-locating information.

Acknowledgements. M.C. and V.C. acknowledge financial support by the Deutsche Forshungsgemeinshaft (priority program *New Frameworks of Rationality*, SPP 1516). V.C. also acknowledges support from the Italian Ministry of Scientific Research (FIRB project *Structures and dynamics of knowledge and cognition*, Turin unit: D11J12000470001).

References

Bostrom, N. (2007). Sleeping Beauty and self-location: A hybrid model. *Synthese*, *157* (1), 59-78.

Cisewski, J., Kadane, J.B., Schervish, M.J., Seidenfeld, T., & Stern, R. (2015) Sleeping Beauty's Credences. Manuscript.

Edwards, W. (1968). Conservatism in Human Information Processing. In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgment* (pp. 17–52). New York, NY: Wiley.

Elga, A. (2000). Self-locating Belief and the Sleeping Beauty problem. *Analysis, 60* (2): 143-147.

Krauss, S., & Wang, X.T. (2003). The psychology of the Monty Hall problem: discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, *132* (1), 3-22.

Lewis, D. (2001). Sleeping beauty: reply to Elga. *Analysis*, *61* (3): 171-176.

Piccione, M., & Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, *20* (1), 3-24.

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007). Comparison of confirmation measures. *Cognition*, *103*, 107-119.

Tentori, K., V. Crupi, & Osherson, D. (2010). Second-order Probability Affects Hypothesis Confirmation. *Psychonomic Bulletin & Review*, *17*, 129–34.

Titelbaum, M.G. (2013). Ten reasons to care about the Sleeping Beauty problem. *Philosophy Compass*, *8* (11), 1003-1017.

Appendix

Stimuli used in Study 1

BASIC condition

On a Sunday, you will be administered one of two pills, depending on the toss of a fair coin (HEADS: regular pill; TAILS: strong pill). You will not be told the outcome of the coin toss, and the two pills look identical. However, you know the following.

If the coin landed HEADS:
– the pill you're given on Sunday (regular pill) will make you sleep for one day;
– then you will wake up (on Monday).

If the coin landed TAILS:
– the pill you're given on Sunday (strong pill) will make you sleep for one day;
– then you will wake up a first time (on Monday), and shortly afterwards fall back asleep for another day, forgetting that you just woke up;
– then you will finally wake up a second time (on Tuesday).

Imagine you've just woken up. You don't know which day it is, and you do not know whether or not you have already woken up before. You are now asked to express your belief about the outcome of the coin toss that was made on Sunday: Do you think it was more probably HEADS or TAILS?

After waking up, I would think the coin toss on Sunday is:

[ ] Certain to have been HEADS and not TAILS
[ ] Much more likely to have been HEADS and not TAILS
[ ] Slightly more likely to have been HEADS and not TAILS
[ ] Equally likely to have been HEADS or TAILS
[ ] Slightly more likely to have been TAILS and not HEADS
[ ] Much more likely to have been TAILS and not HEADS
[ ] Certain to have been TAILS and not HEADS

NO EVIDENCE condition

[same introductory paragraph as above]

If the coin landed HEADS:
– the pill you're given on Sunday (regular pill) will make you sleep for one day;
– then you will wake up (on Monday).

If the coin landed TAILS:
– the pill you're given on Sunday (strong pill) will make you sleep for two days;
– then you will wake up (on Tuesday).

Imagine you've just woken up. You don't know which day it is. You are now asked to express your belief about the outcome of the coin toss that was made on Sunday: Do you think it was more probably HEADS or TAILS?

[same response scale as above]

PLUS condition

[same introductory paragraph as above]

If the coin landed HEADS:
– the pill you're given on Sunday (regular pill) will make you sleep for one day;
– then you will wake up (on Monday).

If the coin landed TAILS:
– the pill you're given on Sunday (strong pill) will make you sleep for one day;
– then you will wake up a first time (on Monday), and shortly afterwards fall back asleep for another day, forgetting that you just woke up;
– the same will happen on each of the following days, until you finally wake up a fifth time (on Friday).

Imagine you've just woken up. You don't know which day it is, and you do not know whether or not you have already woken up any time before. You are now asked to express your belief about the outcome of the coin toss that was made on Sunday: Do you think it was more probably HEADS or TAILS?

[same response scale as above]

BALLS condition

On a Sunday, a blue ball is placed in an empty and opaque urn in front of you, and you will then be administred one of two pills, depending on the toss of a fair coin (HEADS: regular pill; TAILS: strong pill). You will not be told the outcome of the coin toss, and the two pills look identical. However, you know the following.

If the coin landed HEADS:
– the pill you're given on Sunday (regular pill) will make you sleep for one day;
– meanwhile, one red ball will be placed in the opaque urn;
– then you will wake up (on Monday), and draw a ball from the urn.

If the coin landed TAILS:
– the pill you're given on Sunday (strong pill) will make you sleep for two days;
– meanwhile, five red balls will be placed in the opaque urn;
– then you will wake up (on Tuesday), and draw a ball from the urn.

Imagine you've just woken up. You don't know which day it is. You draw a ball from the urn: the ball is red. You are now asked to express your belief about the outcome of the coin toss that was made on Sunday: Do you think it was more probably HEADS or TAILS?

[same response scale as above]

Stimuli used in Study 2

BASIC condition

On a Sunday, you will be administred one of two pills, depending on the toss of a fair coin (heads: REGULAR pill; tails: STRONG pill). You will not be told the outcome of the coin toss, and the two pills look identical. However, you know the following.

If the coin landed heads:
– the pill you're given on Sunday (REGULAR pill) will make you sleep for one day;
– then you will wake up (on Monday).

If the coin landed tails:
– the pill you're given on Sunday (STRONG pill) will make you sleep for one day;
– then you will wake up a first time (on Monday), and shortly afterwards fall back asleep for another day, forgetting that you just woke up;
– then you will finally wake up a second time (on Tuesday).

Imagine you've just woken up. You don't know which day it is, and you do not know whether or not you have already woken up before. You are now asked to express your belief about the pill you were administered on Sunday: Do you think it was more probably the REGULAR pill or the STRONG pill?

After waking up, I would think the pill I was administered on Sunday is:

[ ] Certain to have been REGULAR and not STRONG

[ ] Much more likely to have been REGULAR and not STRONG

[ ] Slightly more likely to have been REGULAR and not STRONG

[ ] Equally likely to have been REGULAR or STRONG

[ ] Slightly more likely to have been STRONG and not REGULAR

[ ] Much more likely to have been STRONG and not REGULAR

[ ] Certain to have been STRONG and not REGULAR

NO EVIDENCE condition

[same introductory paragraph as above]

If the coin landed heads:
– the pill you're given on Sunday (REGULAR pill) will make you sleep for one day;
– then you will wake up (on Monday).

If the coin landed tails:
– the pill you're given on Sunday (STRONG pill) will make you sleep for two days;
– then you will wake up (on Tuesday).

Imagine you've just woken up. You don't know which day it is. You are now asked to express your belief about the pill you were administered on Sunday: do you think it was more probably the REGULAR or the STRONG pill?

[same response scale as above]

PLUS condition

[same introductory paragraph as above]

If the coin landed heads:
– the pill you're given on Sunday (REGULAR pill) will make you sleep for one day;
– then you will wake up (on Monday).

If the coin landed tails:
– the pill you're given on Sunday (STRONG pill) will make you sleep for one day;
– then you will wake up a first time (on Monday), and shortly afterwards fall back asleep for another day, forgetting that you just woke up;
– the same will happen on each of the following days, until you finally wake up a fifth time (on Friday).

Imagine you've just woken up. You don't know which day it is, and you do not know whether or not you have already woken up any time before. You are now asked to express your belief about the pill you were administered on Sunday: Do you think it was more probably the REGULAR pill or the STRONG pill?

[same response scale as above]

BALLS condition

On a Sunday, a blue ball is placed in an empty and opaque urn in front of you, and you will then be administered one of two pills, depending on the toss of a fair coin (heads: REGULAR pill; tails: STRONG pill). You will not be told the outcome of the coin toss, and the two pills look identical. However, you know the following.

If the coin landed heads:
– the pill you're given on Sunday (REGULAR pill) will make you sleep for one day;
– meanwhile, one red ball will be placed in the opaque urn;
– then you will wake up (on Monday), and draw a ball from the urn.

If the coin landed tails:
– the pill you're given on Sunday (STRONG pill) will make you sleep for two days;
– meanwhile, five red balls will be placed in the opaque urn;
– then you will wake up (on Tuesday), and draw a ball from the urn.

Imagine you've just woken up. You don't know which day it is. You draw a ball from the urn: the ball is red. You are now asked to express your belief about the pill you were administered on Sunday: Do you think it was more probably the REGULAR pill or the STRONG pill?

[same response scale as above]