Sleeping Beauty, Read

Clark Glymour

“The Sleeping Beauty problem: 1 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 to back to sleep with a drug that makes you forget that waking. 2 When you are first awakened, to what degree ought you believe that the outcome of the coin toss is Heads?”[[1]](#footnote-1)

Let the two days be Monday and Tuesday. Many discussions misread the problem, assuming that if Heads, then SB is awakened on Monday[[2]](#footnote-2), or if Heads, then SB is awakened on Monday or Tuesday with equal probability or other gratuitous assumptions[[3]](#footnote-3) not warranted by Elga’s original text, taking implicit advantage of the fact that probability consequences are non-monotonic. The question presupposes, but the specification does not say, that the answer is the same for first wakening on Monday as on first waking on Tuesday. The probability that SB is awakened on Monday given that the flip is heads is unspecified in the text. Let that probability be p = Pr(M |h), with p unknown to Sleeping Beauty. The only solution that follows necessarily from the text is that Pr(heads on coin flip |FA[(first awakening]) = ½.

Proof. Let M and T denote *first awakening* on Monday or Tuesday, respectively, and let Pr(h) (Pr(t)) denote the probability of heads (tails) on the coin flip. Note that if t is the outcome of the flip, Pr(M | t) = 1.

We must determine: Pr(h | M) and Pr(h | T) or Pr(h | FirstAwakening (= M exclusive or T).

Here is a 3- line proof from Elga’s formulation that Pr(h | FA) = ½, p ≠ 0,1

*Given:*

Pr(h) = ½ = Pr(t)

Pr(M | t) = 1

Pr(M | h) = p

Pr(T | h) = 1 - p

Pr(M) + Pr(T) = 1 (SB will surely be first awakened one of these two days)

Pr(h on FirstAwakening) = Pr(hM) + Pr(hT) (M, T are mutually exclusive).

*Hence,*

Pr(hT) = ½(1 – p). (Since the probability of h is ½, and if h occurs then T occurs with probability 1 – p.)

Pr(hM) = ½p (Since the probability of h is ½, and if h occurs then M occurs with probability p.)

But, Pr(h on FirstAwakening) = Pr(hM) + Pr(hT) = ½p + ½(1 – p) = ½ (Since M, T are mutually exclusive and exhaustive).

*Done*.

1. Elga, Self-locating belief and the Sleeping Beauty problem, *Analysis*, 60(2): 143-147, 2000. [↑](#footnote-ref-1)
2. The Wikipedia entry, for example, accurately quotes Elga and then immediately misstates the problem this way. The entry gives extensive references. The “Heads implies that there is a waking is on Monday” misreading entails that the first awakening is on Monday regardless of the outcome of the coin flip, making the problem even more trivial: two hypotheses that entail the same datum have posterior probabilities on that datum in the same ratio as their prior probabilities. [↑](#footnote-ref-2)
3. Arguing for the 1/3 conclusion, Elga introduces a number of gratuitous assumptions. [↑](#footnote-ref-3)