Too clever by halving

Tim Button, Daniel Rothschild, and Levi Spectre

May 10, 2023

Consider the following scenario from Elga:[1]

> CLASSIC SB. A fair coin is flipped.
>
> - If it lands tails, then Sleeping Beauty is awoken twice, on Monday and Tuesday. (As she sleeps, Beauty's memory of Monday is erased, so that she has no memory of Monday's events on Tuesday.)
> - If it lands heads, then Beauty is awoken on Monday but not Tuesday.
>
> All of these possible awakenings are indistinguishable to Beauty. Beauty is fully informed about the setup. When she wakes up on Monday, knowing only that it is Monday or Tuesday, what probability should she give to the coin landing heads?

Anyone who answers ½ to CLASSIC SB is a *halver*.[2] We want to show that halving is untenable. In §1, we will show that halving violates a deeply plausible constraint on probability assignments, the Principle of Irrelevant Ignorance. In §§2–3, we will show that halving either violates solid statistical reasoning or draws absurdly irrelevant distinctions.

1 The Principle of Irrelevant Ignorance

Our first argument against halvers begins by tweaking the scenario slightly. We can easily reformulate CLASSIC SB, so that if the coin lands heads, then Beauty is awakened on Tuesday but not Monday. Call this TUESDAY SB. We cannot see how any halver would think that there is a relevant difference between CLASSIC SB and TUESDAY SB.[3] So halvers must say that Beauty should answer ½, when asked (in TUESDAY SB) about the probability that the coin landed heads.

Now let us implement an extra layer of randomization: we will toss a fair coin to decide whether to srun CLASSIC SB (on heads) or TUESDAY SB (on tails). For ease of reference, here is the full scenario:[4]

---

[1] Elga (2000); it is a variation on an example in Piccione and Rubinstein (1997).

[2] Halvers come in different varieties; some of our best friends and earlier time-slices are halvers.

[3] Indeed, in Elga's (2000: §1) initial presentation of the argument, all he mentions is the number of awakenings; the particular days come later (2000: §2).

[4] This scenario is presented in Karlander and Spectre (2010: 400), but in the service of an entirely different argument (to show that Beauty gains indexical information in CLASSIC SB). Conitzer (2015: 1987–8) has a very different setup which involves two coins; see footnote 18.

Toggled SB. A fair coin is flipped.

- If it lands tails, then Beauty is awoken twice, on Monday and Tuesday. (As she sleeps, Beauty's memory of Monday is erased, so that she has no memory of Monday's events on Tuesday.)
- If it lands heads, then the coin is flipped a second time:
  - if the second flip lands heads: Beauty is awoken on Monday but not Tuesday.
  - if the second flip lands tails: Beauty is awoken on Tuesday but not Monday.

All of these possible awakenings are indistinguishable to Beauty. Beauty is fully informed about the setup. When she first wakes up, knowing only that it is Monday or Tuesday, what probability should she give to the coin's (first) landing heads?

In each of Classic SB, Tuesday SB and Toggled SB: if the coin lands heads (when first flipped), then Beauty is awoken only once, and if it lands tails (when first flipped) then she is awoken twice. Consequently, halvers will still answer ½ in Toggled SB.[5]

That is a mistake. As we will now show, the uniquely correct answer for Toggled SB is ⅓. Our argument for this conclusion proceeds by establishing two claims.[6]

(1) On Monday or Tuesday, if Beauty learns what day it is, then she should assign ⅓ to Heads.
(2) Beauty's ignorance of what day it is in Toggled SB should not affect her credences, so she should assign ⅓ to Heads.

Our argument for claim (1) is simple. Let us tweak Toggled SB so that, whenever Beauty is awoken, she is also told what day it is (but everything else about the setup is the same, including the memory erasures). Going into the experiment, on Sunday, Beauty should have assigned her probabilities as follows:

½: the first flip lands tails
¼: the first flip lands heads and the second lands heads
¼: the first flip lands heads and the second lands tails

Suppose now that Beauty wakes up on Monday and is told that it is Monday. Her information—that she is awake on Monday—eliminates exactly one possibility: that the coin first landed heads on its first flip and then landed tails on its second. So she should now assign her probabilities as follows:

⅔: the first flip lands tails

[5] Advocates of the rule of Compartmentalized Conditioning (which we introduce and criticize in §3) will certainly answer ½ (see §B.2 for the calculations).

[6] Variants where Beauty learns what day it is are as old as the sleeping beauty problem itself (see Elga 2000: §2). We revisit this in §B.1.

⅓: the first flip lands heads

Since Toggled SB is completely symmetric with respect to Monday and Tuesday, the same reasoning applies on Tuesday. This establishes claim (1).[7]

We will now argue for (2). In establishing (1), we established what probabilities Beauty should have if she learns what day it is. It is natural to regard these as providing us with Beauty's conditional probabilities. So: let Pr be the function which provides the rational probabilities which Beauty has after awakening; let H be the event that the (first) flip lands heads; and let M be the event that today is Monday. We can take the reasoning behind (1) to show that $\Pr(H \mid M) = ⅓$ and $\Pr(H \mid \neg M) = ⅓$. We can then directly infer from the law of total probability that $\Pr(H) = ⅓$. So, after waking up, Beauty should assign ⅓ to H.

We find this a compelling reason to accept (2). However, som philosophers might be uncomfortable directly applying the laws of probability to probability spaces that include self-locating events like M. Others might deny that the conditional probabilities really are as stated above.[8] Such scruples should not, though, undermine the overall argument. For we recommend the following general principle:

Principle of Irrelevant Ignorance. If an agent who is ignorant about C assigns probability r to X hypothetical on C, and also assigns r to X hypothetical on $\neg C$, then they should assign r to X. (Here, C and X can be self-locating.)

This Principle is phrased in terms probabilities which are 'hypothetical on' a proposition. This is meant to describe probabilities which an agent would have if she were to have some information, regardless of whether the probabilities are the agent's *conditional probabilities*, in the strict sense defined on a probability space. (Hypothetical probabilities might not be definable purely dispositionally—as we might want to define them for statements like 'My partner is cheating on me but I never find out'—but the details are irrelevant to our arguments.)

Now, as noted above: if the propositions *are* described in a probability space and the hypothetical probabilities give the conditional probabilities (in the strict sense), then this Principle is a theorem of probability theory.[9] But we advance the Principle as a general norm, to constrain any plausible model of an agent's rational probabilities under ignorance (whether the ignorance involves self-location or not).[10]

---

[7] All the reasoning here is elementary probabilistic reasoning without, as far as we can tell, any special issues concerning self-locating beliefs. Indeed, this reasoning is also recommended by Compartmentalized Conditioning (a rule which we introduce and criticize in §3; see also §B.2).

[8] In §B.1, we introduce and criticize Compartmentalized Conditioning. This rule insists that $\Pr(H \mid M) = ½$, whilst agreeing that Beauty's credence in Heads should be ⅓ if she learns that it is Monday (see §B.2).

[9] i.e. if $\Pr(X \mid C) = \Pr(X \mid \neg C) = r$ then $\Pr(X) = r$.

[10] We can think of this principle as a highly restricted version of van Fraassen's (1984) Reflection Principle. The Principle of Irrelevant Ignorance constrains what credences and hypothetical credences an agent can have at a single time. It makes no further restriction on dispositions or plans to change credences over time. Those unhappy with more general reflection principles should have no objections to the Principle of Irrelevant Ignorance.

The motivation for the Principle is clear: if you don't know whether C, but regardless of whether C or $\neg C$ you would assign probability r to X, then you should assign r to X. Indeed, it is really not clear to us what 'probabilities' are meant to *represent*, for those who do not link probabilities to beliefs in the way the Principle requires.

The Principle of Irrelevant Ignorance though, allows us to re-run the reasoning from before. Beauty is ignorant about M, but assigns ⅓ to H hypothetical on M, and assigns ⅓ to H hypothetical on $\neg M$; so Beauty should simply assign ⅓ to H. The (uniquely) correct answer to Toggled SB is therefore ⅓.

With this, we see that halving—i.e. answering ½ in Classic SB—is untenable. After all, halvers cannot point to any relevant difference between Toggled SB and Classic SB, which would allow them to answer ⅓ in the former case but ½ in the latter. So halvers must deny the Principle of Irrelevant Ignorance. This is an enormous and hitherto unacknowledged cost to halving.

2 A statistical argument

We will now present a further cost of halving: it violates solid statistical reasoning.

To show this, we will need to introduce two more variant scenarios. We start by tweaking Toggled SB. Now, instead of leaving Beauty sleeping through one of the days, if the coin lands heads when first flipped, we will instead wake her on that day *and* tell her the result of the (first) coin flip.[11] For clarity, here is the full protocol:

Informed SB. Beauty is awoken on Monday and Tuesday. A fair coin is flipped.

- If it lands tails, then Beauty is told nothing about the coin flip on Monday or Tuesday.
- If it lands heads, then the coin is flipped a second time:
  - if the second flip lands heads: on awakening on Tuesday, Beauty is immediately told that the coin landed heads when first flipped (but she is told nothing on Monday);
  - if the second flip lands tails: on awakening on Monday, Beauty is immediately told that the coin landed heads when first flipped (but she is told nothing on Tuesday).

Beauty's memory is erased as she sleeps from Monday to Tuesday (iff the coin landed tails on its first or second flip). She is fully informed about the setup. When she wakes up and is told nothing about the outcome of the (first) coin flip, what probability should she give to the coin landing heads (when first flipped)?

[11] Karlander and Spectre (2010: 405) and Conitzer (2015: 1990) also consider scenarios where Beauty is awake on both days, come what may. However, we give a novel defence of the significance of these cases (see our discussion of 'framing effects'). Furthermore, our main aim here is to get to our (novel) statistical argument.

We claim that there is no effective difference between Toggled SB and Informed SB, so that everyone must give the same answer to both. However, this is not immediately obvious. Indeed, we imagine the following challenge:

> In Classic SB and Toggled SB, all possible awakenings during the experiment are indistinguishable. But they are not all indistinguishable in Informed SB: she might be told something about the coin, or she might not. To see why this matters, let W be the following claim:[12]
>
> *Either: it's Monday and the coin didn't land Heads-then-tails,*
> *or: it's Tuesday and the coin didn't land Heads-then-heads.*
>
> The protocol of Toggled SB guarantees that, whenever Beauty wakes up during the course of the experiment, she learns that W. But in Informed SB, this is not guaranteed by the protocol alone; instead, Beauty only learns that Ws when (and because) she is not told anything. So: when Beauty learns that W in Informed SB—but not in Toggled SB—it seems that she should revise her probabilities.

This challenge is mistaken, and for an interesting reason: it ascribes epistemological significance to a mere framing effect, namely, how we draw the boundaries between what is, and what is not, a 'part of the experiment'.

To explain this point, we begin by recalling the protocol for Toggled SB. In specifying it in §1, we said nothing about what happens to Beauty on Sunday or Wednesday. But, if we liked, we could have ensured that Beauty is aware of the result of the (first) coin flip on those days. In detail: Beauty would learn the result of the (first) coin flip on Sunday; her memory would be erased as she slept from Sunday to Monday; the protocol for Monday and Tuesday would then follow the specification laid down in §1; and then on Wednesday Beauty would learn the result of the (first) coin flip once again. It is obvious, though, that this slight enrichment of Toggled SB should not affect Beauty's reasoning on Monday or Tuesday.

Let us similarly enrich Informed SB. So, we now specify that Beauty is aware of the result of the (first) coin flip on both Sunday and Wednesday.[13] Again, this should not affect how Beauty should reason on either Monday or Tuesday during Informed SB. But now consider how we describe what happens when the coin lands Heads-then-heads. In this case, when she awakens on Tuesday, Beauty immediately learns that the coin (first) landed Heads. Given our enriched specification, she has exactly the same information on Wednesday. Now, we might well think of Tuesday as being 'part of the experiment', and of Wednesday as being 'after the experiment'. (Indeed, this thought motivated the challenge that we described a couple of paragraphs ago.) But we could, instead, equally well redescribe this by saying:

- if the second flip lands heads: we end the experiment a day early, on Tuesday, letting her learn on Tuesday what she will find out anyway on Wednesday, i.e. how the coin (first) landed.

[12] Note that W is *centered*; see §3 for what this means.

[13] Again, we (may) need to erase Beauty's memory as she sleeps from Sunday to Monday.

Next, consider the situation where the coin lands Heads-then-tails. Reasoning in essentially the same way (though now considering Monday and Sunday rather than Tuesday and Wednesday), we see that we could equally well redescribe the protocol by saying:

- if the second flip lands tails: we start the experiment a day late, on Tuesday, letting her remember on Monday what she already found out on Sunday, i.e. how the coin (first) landed.

Under these equivalent redescriptions, the experimental protocol of Informed SB now guarantees that, whenever Beauty wakes up 'during the course of the experiment', she finds out that W. So, by mere redescription, the apparently significant difference between Toggled SB and Informed SB has evaporated.

Having shown that we should treat Toggled SB and Informed SB in the same way, we now consider a final change. The basic idea, here, is to take the protocol of Informed SB and multiply it up 40,000 times.

Displayed SB. Beauty is awoken on Monday and Tuesday. As she sleeps, her memory of the previous day is erased. Whenever she wakes up, the first thing Beauty sees is a screen. Forty-thousand fair coins, each uniquely labelled "1" through "40,000", are flipped. When n is between 1 and 40,000:

- If coin-n lands tails, then no information is displayed about coin-n on the screen on either day.
- If coin-n lands heads, then it is flipped a second time:
  - if the second flip lands heads: the numeral **n** is displayed on the screen on Tuesday (but not Monday);
  - if the second flip lands heads: the numeral **n** is displayed on the screen on Monday (but not Tuesday);

So: for each n, seeing the numeral **n** on the screen amounts to being told that coin-n landed heads when first flipped.

On Sunday, a number, i, between 1 and 40,000 is chosen at random; Beauty is told what i is on both Monday and Tuesday. Now: when Beauty does not see the numeral **i** on the screen, what probability should she give to coin-i landing heads when first flipped?

Here is a simple statistical argument that, in Displayed SB, we must assign a probability of about ⅓ to the chance of coin-i having landed heads.[14] Based on the setup of Displayed SB, on both days, it is overwhelmingly likely that around 10,000 numerals will be displayed. It is also overwhelmingly likely that around ⅓ of the undisplayed coins (i.e. those whose numerals did not feature on the display) landed heads when

[14] Our statistical argument should be sharply distinguished from, for example, Elga's (2000: §1) 'long run' consideration.

they were first flipped.[15] Now, in the case we are imagining, coin-i is undisplayed. But there is nothing special about the particular number i here: recall that i was chosen at random on Sunday. So, on statistical grounds alone, Beauty should assign a probability of about ⅓ to the claim that coin-i landed heads when first flipped.

With this, we have a second refutation of halving. After all, from Beauty's perspective, Displayed SB is just the same as Informed SB, but with a bunch of extra information: the outcomes concerning coin-n, for each $n \neq i$. Indeed, since all of the coin flips are independent, any information concerning any *other* coin is intuitively just *irrelevant* to the likelihood that coin-i landed heads when first flipped. So Beauty should give the same answer in Displayed SB as in Informed SB, i.e. ⅓. And Beauty should give the same answer in Informed SB as in Toggled SB. So, halving is wrong, once again.

3 Compartmentalized Conditioning not to the rescue

We hope that the above two arguments will have convinced an agnostic reader to reject halving. However, a certain sub-species of halver does have a reply to our statistical argument. Specifically: advocates of *Compartmentalized Conditioning* will claim that Informed SB and Displayed SB are relevantly different, so that they can give different answers in these cases. In this section, we will explain why their insistence on a difference leads them to absurdity.

To begin, we must introduce Compartmentalized Conditioning. This is a rule for governing how a subject alters their beliefs when their state changes.[16] The rule distinguishes between *worldly* and *centered* states, and this distinction is best illustrated by example. In Classic SB there are only two possible worldly states: Heads and Tails. But there are three possible centered states, i.e. states that Beauty might be in: Heads+Monday, Tails+Monday, and Tails+Tuesday. The updating rule is then as follows.

Compartmentalized Conditioning. Updating proceeds in two steps:

Step 1. Update your probabilities concerning worldly states, following standard Bayesian conditioning, using only your worldly information.

Step 2. Keeping your worldly probabilities constant, distribute your probabilities among the centered states.[17]

[15] A little more precisely: where r is the ratio of heads to tails (for the first flip) among the undisplayed coins, Beauty should be more than 99.9% confident that $|r - ⅓| < 1/100$.

[16] Compartmentalized Conditioning was introduced by Halpern and Tuttle (1993) and advocated by Meacham (2008). We say 'state changes', to stay neutral on the question of whether they get new information or not.

[17] How this distribution is executed will not be relevant to our arguments, but we will assume it is done using either priors or some suitable indifference principle.

Compartmentalized Conditioning immediately leads to halving for Classic SB (see §B.1). Consequently, Compartmentalized Conditioning contradicts the Principle of Irrelevant Ignorance, as in §1. That is already a sufficient reason to reject Compartmentalized Conditioning. Still, Compartmentalized Conditioning *does* draw a distinction between Informed SB and Displayed SB, answering ½ in Informed SB, but approximately ⅓ in Displayed SB. So it avoids our statistical argument of §2.

That said, the phrase 'approximately ⅓' masks something rather important. Moreover, when we unpack it, we see that Compartmentalized Conditioning draws some genuinely absurd distinctions.[18] To show this clearly, we will slightly simplify the setup of Displayed SB: suppose that there are only 2 labelled coins, rather than 40,000, and that Beauty is asked whether coin-1 landed heads when first flipped. Now, according to Compartmentalized Conditioning (see §B.3):

- if neither numeral is displayed, then Beauty should answer 3⁄7;
- if only **2** is displayed, then Beauty should answer ⅓.

But recall that coin-1 and coin-2 are stipulated to be totally independent of one another. So it would be absurd for Beauty to give different answers in these two situations.[19] With this, we reject Compartmentalized Conditioning; and so our statistical argument goes through.

4 Where we are

To close, we will briefly review our arguments against halving. First, we showed that any halver must deny the Principle of Irrelevant Ignorance (see §1). Second, we showed that halvers must ignore obvious statistical truths (see §2), unless they absurdly insist on giving different answers in similar situations (see §3).[20]

[18] Dorr (2005), Titelbaum (2008: 591–5, 2013: 1007–8), and Conitzer (2015: 1987–8) have all noted (in different ways) that if we tweak Classic SB, by providing Beauty with some apparently irrelevant information whenever she wakes up, then advocates of Compartmentalized Conditioning must give an answer of (or approaching) ⅓. We agree that this is unpleasant for halvers who advocate Compartmentalized Conditioning; but they can ultimately "bite the bullet" and insist that giving Beauty extra information changes the scenario enough to allow for a different answer. Our objection does not allow for similar bullet-biting (see footnote 19).

[19] So, revisiting footnote 18: our point is not that *halvers* should be embarrassed to give an answer other than ½, when Beauty is given some apparently irrelevant information. Our point is that, in this scenario, it is absurd to think that Beauty should give *different* answers depending on whether **2** is displayed.

[20] For comments and conversation, we wish to thank: Nilanjan Das, Cian Dorr, Kevin Dorst, Adam Elga, David Enoch, J Dimitri Gallow, Joe Horton, Brian King, Finlay McCardel, Robert Trueman, Masahiro Yamada.

A Calculations concerning Displayed SB

In §2, we discussed Displayed SB, and made claims like: it is overwhelmingly likely that around ⅓ of the undisplayed coins landed heads when first flipped.

In principle, for each $\epsilon > 0$, we can calculate the exact odds that between $1/3 - \epsilon$ and $1/3 + \epsilon$ of the undisplayed coins landed heads (on either day). But if we consider something like 40,000 coins, this becomes extremely computationally demanding. Fortunately, we can make a deeper point without using vast computational resources.

Let N be the number of coins under discussion. So $N = 40{,}000$ in the original Displayed SB case; but we will now allow this to vary. Assume it is Monday; by the symmetry of our setup, exactly the same considerations will hold for Tuesday. Let R_N be the ratio of heads to tails, among those coins which aren't displayed on Monday. Then we claim:

> As N increases: (before the experiment) Beauty should become arbitrarily confident that R_N is arbitrarily close to ⅓.

This claim is a simple consequence of the weak law of large numbers. To see this, for each $1 \leq n \leq N$, define indicator random variables corresponding to events as follows:

D_n: coin-n is not displayed on Monday
H_n: coin-n is not displayed on Monday *and* coin-n landed heads when (first) flipped.

The law of large numbers tells us, for all $\epsilon > 0$, as N gets large, the value of

$$\bar{D}_N = \frac{\sum_{n=1}^{N} D_n}{N}$$

converges in probability to ¾, while $\bar{H}_N$ converges in probability to ¼. Now, where $R_N = \frac{\bar{H}_N}{\bar{D}_N}$, this implies that, for any $\epsilon > 0$, as N gets large Beauty should assign a probability arbitrarily close to 1 to the event that

$$|R_N - 1/3| < \epsilon$$

which is exactly what we claimed, in slightly more formal terms.

B Calculations concerning Compartmentalized Conditioning

In this appendix, we will provide the relevant calculations which underpin our discussion of Compartmentalized Conditioning.

B.1 Compartmentalized Conditioning and Classic SB

We begin with a familiar point: Compartmentalized Conditioning leads to halving. To see this, consider Classic SB, and reason as follows. On Sunday: Beauty should assign ½ to Heads and ½ to Tails, the two (relevant) possible states of the world. When she awakens on Monday (not knowing what day it is), she gets no new worldly information, since she knew in advance she would wake up in just this way. So there is no updating to perform at Step 1, and her probabilities should be:

Heads	½
Tails	½

However, at Step 2, she must distribute her 'Tails probability', i.e. ½, between two centered states: Monday+Tails and Tuesday+Tails. Presumably she will split her Tails probability evenly between these two centred states, obtaining:[21]

	Monday	Tuesday
Heads	½	
Tails	¼	¼

So, Compartmentalized Conditioning answers ½ in Classic SB.

Indeed, Compartmentalized Conditioning recommends *double-halving*. Specifically, consider a tweak to Classic SB whereby, a few minutes after she wakes up, Beauty learns that it is Monday.[22] According to Compartmentalized Conditioning, this does not affect Beauty's worldly probabilities: after all, there is a Monday+Tails centered state and a Monday+Heads centred state. But she will redistribute all her 'Tails probability' to Monday at Step 2, as follows:

	Monday	Tuesday
Heads	½	
Tails	½	

So Compartmentalized Conditioning leads to *double-halving*, i.e., giving ½ to Heads before and after learning it is Monday.

B.2 Compartmentalized Conditioning and Toggled SB

In §1, we argued that halvers will draw no relevant difference between Classic SB and Toggled SB. In fact, it is easy to see that Compartmentalized Conditioning answers ½ in Toggled SB. On Sunday, Beauty should assign ½ to (her coin first landing) Heads and ½ to Tails. When Beauty first awakens, she has nothing to do at Step 1; and she will then presumably split these probabilities evenly at Step 2, obtaining this distribution:

[21] This uses a principle of indifference; she might not split things differently if she had different priors concerning which day is more likely.

[22] The tweak is introduced by Elga (2000: §2). Double-halving is recommended by Bostrom (2007), Bradley (2011a,b, 2012), Cozic (2011), Halpern (2005), Leitgeb (2010), Lewis (2010), Meacham (2008), Hawley (2013), Pust (2012), and Yamada (2019).

	Monday	Tuesday
Heads	1/4	1/4
Tails	1/4	1/4

So, according to Compartmentalized Conditioning, Beauty should say that $\Pr(H) = 1/2$, and moreover that $\Pr(H \mid M) = 1/2$.

However, if Beauty learns that it is Monday in Toggled SB, then Compartmentalized Conditioning recommends that she should revise her answer to a 1/3. (So advocates of Compartmentalized Conditioning agree with claim (1) of §1.) This is because, unlike in Classic SB, the setup in Toggled SB does not guarantee that Beauty will wake up on Monday. So, on learning it is Monday, Beauty does not *merely* gain centered information; she gains the worldly information that the coin did *not* land Heads-then-tails. Accordingly, at Step 1 of the Compartmentalized Conditioning process, she assigns 1/3 to Heads (specifically, Heads-then-heads) and 2/3 to Tails. At Step 2, she assigns all of this to Monday.

B.3 Compartmentalized Conditioning and Displayed SB

In §3, we discussed how Compartmentalized Conditioning treats Displayed SB. Here we will provide the relevant calculations.

Let N be the number of coins involved. So $N = 40,000$ in the case described in §2, but $N = 2$ in §3. We represent the possible worldly states—i.e. the possible outcomes of the coin flips—using N-length strings, to record the flips associated with each of the coins. We adopt this notation system:

- T: the salient coin landed Tails (when first flipped)
- h: the salient coin landed Heads-then-heads
- t: the salient coin landed Heads-then-tails

So we are considering N-length strings with alphabet $\{\mathsf{T}, \mathsf{h}, \mathsf{t}\}$. To illustrate: if we had three coins, the string TtT would represent that coin-1 landed Tails, coin-2 landed Heads-then tails, and coin-3 landed Tails.

The priors dictate that a T is twice as likely as an h or a t. So we can assign to each string a probabilistic *weight*, given by 2^k, where k is the number of instances of T in that string. (To illustrate: if $N = 3$, then the string TtT has weight 4, indicating that it has four times the prior probability of string htt.) In this context, Step 1 of Compartmentalized Conditioning amounts to deleting certain strings, and redistributing probabilities over the remaining strings by considering their weights.

Using this framework, we can prove a Proposition which entails the oddity that, according to Compartmentalized Conditioning, where $N = 2$ and no numeral is displayed, Beauty should answer 3/7 (see §3).

Proposition 1: According to Compartmentalized Conditioning, in a Displayed SB setup with N coins, and with $1 \leq i \leq N$: if no numeral is displayed, then Beauty should assign $\frac{3^{N-1}}{3^N - 2^{N-1}}$ to coin-i first landing heads.

Proof. When no numerals are displayed, Beauty obtains this worldly information: all coins which first landed heads landed the same way as each other on their second flip. So Beauty must delete any string which contains both an h and a t. Call these the worldly-compatible strings. The worldly-compatible strings are:

- all N-length strings with alphabet $\{\mathsf{T}, \mathsf{h}\}$; and
- all N-length strings with alphabet $\{\mathsf{T}, \mathsf{t}\}$.

We now calculate the aggregate weight of the worldly-compatible strings. Elementary combinatorial reasoning shows that the aggregate weight of all N-length strings with alphabet $\{\mathsf{T}, \mathsf{h}\}$ is 3^N; similarly for all N-length strings with alphabet $\{\mathsf{T}, \mathsf{t}\}$. However, the N-length string with alphabet $\{\mathsf{T}\}$ has weight 2^N, and we must not double-count this. So the aggregate weight of the worldly-compatible strings is: $2 \times 3^N - 2^N$.

Without loss of generality, let $i = 1$. We now consider the heads-compatible strings, i.e. those strings compatible with Beauty's worldly information which correspond with coin-1 first landing heads. These are those worldly-compatible strings which start with either h or t. Those starting with h are exactly the $(N-1)$-length strings with alphabet $\{\mathsf{T}, \mathsf{h}\}$, whose aggregate weight is 3^{N-1}; similarly, the aggregate weight of those starting with t is 3^{N-1}. So the aggregate weight of the heads-compatible strings is: $2 \times 3^{N-1}$.

Dividing the aggregate weight of the heads-compatible strings by the aggregate weight of the worldly-compatible strings, we obtain $\frac{3^{N-1}}{3^N - 2^{N-1}}$. This completes Step 1 of the calculation; and, given the setup, no redistribution is required at Step 2. □

Remark 2: As N becomes arbitrarily large, the value of the formula in Proposition 1 approaches ⅓ without limit.

We have just considered the case where no numeral is displayed. However, if at least one numeral is displayed, then Compartmentalized Conditioning recommends that Beauty should reason exactly like a thirder (again, see §3):

Proposition 3: According to Compartmentalized Conditioning, in a Displayed SB setup with N coins, and with $1 \leq i \leq N$: if some numeral is displayed but i is not, then Beauty should assign ⅓ to coin-i first landing heads.

Proof. Without loss of generality: let $i = 1$ and let there be some $1 \leq k < N$ such that coin-1 through coin-k are all not displayed, but coin-$(k+1)$ through to coin-N are all displayed. The following strings now correspond to possible worldly states:

- all k-length strings with alphabet $\{\mathsf{T}, \mathsf{h}\}$, followed by $(N-k)$ instances of t.

- all k-length strings with alphabet $\{\mathsf{T}, \mathsf{t}\}$, followed by $(N - k)$ instances of h.

The aggregate weight of these strings is 2×3^k (since $k < N$, there is no double-counting). The aggregate weight of those strings starting with either h or t is $2 \times 3^{k-1}$. Dividing the latter by the former yields ⅓. This completes Step 1 of the calculation, and no redistribution is required at Step 2. □

References

Bostrom, Nick (2007). 'Sleeping beauty and self-location: A hybrid model'. *Synthese* 157.1, pp.59–78. DOI: 10.1007/s11229-006-9010-7.

Bradley, Darren (2011a). 'Confirmation in a branching world: The Everett interpretation and Sleeping Beauty'. *British Journal for the Philosophy of Science* 62.2, pp.323–342. DOI: 10.1093/bjps/axq013.

— (2011b). 'Self-location is no problem for conditionalization'. *Synthese* 182.3, pp.393–411. DOI: 10.1007/s11229-010-9748-9.

— (2012). 'Four problems about self-locating belief'. *Philosophical Review* 121.2, pp.149–177. DOI: 10.1215/00318108-1539071.

Conitzer, Vincent (2015). 'A devastating example for the halfer rule'. *Philosophical Studies* 172.8, pp.1985–1992. DOI: 10.1007/s11098-014-0384-y.

Cozic, Mikael (2011). 'Imagining and Sleeping Beauty: A case for double-halfers'. *International Journal of Approximate Reasoning* 52.2, pp.137–143.

Dorr, Cian (2005). 'A challenge for halfers'. Unpublished manuscript.

Elga, Adam (2000). 'Self-locating belief and the Sleeping Beauty problem'. *Analysis* 60.2, pp.143–147. DOI: 10.1111/1467-8284.00215.

Halpern, Joseph Y (2005). 'Sleeping Beauty reconsidered: Conditioning and reflection in asynchronous systems'. In: *Oxford Studies in Epistemology Volume 1*. Ed. by Tamar Szabo Gendler and John Hawthorne. Oxford University Press.

Halpern, Joseph Y and Mark R Tuttle (1993). 'Knowledge, probability, and adversaries'. *Journal of the ACM (JACM)* 40.4, pp.917–960.

Hawley, Patrick (2013). 'Inertia, optimism and beauty'. *Noûs* 47.1, pp.85–103. DOI: 10.1111/j.1468-0068.2010.00817.x.

Karlander, Karl and Levi Spectre (2010). 'Sleeping Beauty meets Monday'. *Synthese* 174.3, pp.397–412. DOI: 10.1007/s11229-009-9464-5.

Leitgeb, Hannes (2010). 'Sleeping Beauty and eternal recurrence'. *Analysis* 70.2, pp.203–205. DOI: 10.1093/analys/anp177.

Lewis, Peter J. (2010). 'Credence and self-location'. *Synthese* 175.3, pp.369–382. DOI: 10.1007/s11229-009-9528-6.

Meacham, Christopher J. G. (2008). 'Sleeping Beauty and the dynamics of de se beliefs'. *Philosophical Studies* 138.2, pp.245–269. DOI: 10.1007/s11098-006-9036-1.

Piccione, Michele and Ariel Rubinstein (1997). 'On the interpretation of decision problems with imperfect recall'. *Games and Economic Behavior* 20.1, pp.3–24.

Pust, Joel (2012). 'Conditionalization and essentially indexical credence'. *Journal of Philosophy* 109.4, pp.295–315. DOI: 10.5840/jphil2012109411.

Titelbaum, Michael G (2008). 'The relevance of self-locating beliefs'. *Philosophical Review* 117.4, pp.555–606.

— (2013). 'Ten reasons to care about the Sleeping Beauty problem'. *Philosophy Compass* 8.11, pp.1003–1017. doi: 10.1111/phc3.12080.

van Fraassen, Bas C. (1984). 'Belief and the will'. *Journal of Philosophy* 81.5, pp.235–256. doi: 10.2307/2026388.

Yamada, Masahiro (2019). 'Beauty, odds, and credence'. *Philosophical Studies* 176.5, pp.1247–1261. doi: 10.1007/s11098-018-1061-3.