Too clever by halving

Tim Button, Daniel Rothschild, and Levi Spectre

June 14, 2024

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 either Monday or Tuesday, what probability should she assign to *the coin lands heads*?

Anyone who answers ½ to CLASSIC SB is a *halver*.[2] In this paper, we will offer two arguments to show that halving is untenable.

In §1, we will show that halving violates the Epistemological Sure-Thing Principle, which we argue is a necessary constraint on any reasonable probability assignments. We call this the Sure-Thing Argument. In §2, we will show that halving violates solid statistical reasoning (or draws absurdly irrelevant distinctions). We call this the Statistical Argument. We reserve proofs and calculations for the appendices (§§A–B).

The sleeping beauty Problem has garnered a sizeable literature. In the interests of readability, we presented our arguments as self-contained. But they bear comparison with some extant arguments against halvers, and we explore these comparisons after presenting each argument.

1 The Sure-Thing Argument

For our first argument, we will introduce an *epistemological* version of the (decision-theoretic) sure-thing principle.[3] After motivating our Principle, we will see how halvers violate it.

1.1 The Epistemological Sure-Thing Principle

Here is a very plausible general principle about rationality:

[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] Due to Savage (1954: 21).

Epistemological Sure-Thing Principle. If an agent assigns probability r to A hypothetical on C, and also assigns probability r to A hypothetical on $\neg C$, then they should assign probability r to A (simpliciter).

We have phrased the Principle in terms of 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.

One might well think that such "hypothetical" probabilities should be *conditional probabilities*, in the strict sense defined on a probability space. If they are construed in this way, then the Epistemological Sure-Thing Principle amounts to this elementary theorem of probability theory:

$$\text{If } \Pr(A \mid C) = \Pr(A \mid \neg C) = r, \text{ then } \Pr(A) = r.$$

But we do not want to insist that "hypothetical" probabilities *must* be understood as conditional probabilities. We are happy to countenance viewpoints according to which conditional probabilities can come apart from hypothetical probabilities.[4] Our point is only that, however they are understood, the hypothetical probabilities of any reasonable agent must obey the Epistemological Sure-Thing Principle.

The motivation for the Epistemological Sure-Thing Principle is very simple. It's a sure thing that either C or $\neg C$. So if you don't know whether C, but if you found out either way, then you would assign probability r to A. So you should just assign r to A, simpliciter. It's a *sure thing*, epistemologically speaking.[5]

1.2 Tweaking the scenario

Our next aim is to show that halving violates the Epistemological Sure-Thing Principle. Admittedly, this is not obvious from considering the original sleeping beauty scenario, Classic SB. But it becomes clear when we tweak the structure of that scenario, in ways which halvers should treat as irrelevant.

If the coin lands heads in Classic SB, then Beauty is awoken only once. Consider a similar scenario, which simply varies that day of awakening. So, in Tuesday SB, if the coin lands heads then Beauty is awoken on Tuesday but not Monday. We cannot see how any halver could think there is a relevant difference between Classic SB and Tuesday SB. So halvers will insist that Beauty should answer ½, when she is asked (in Tuesday SB) about the probability of the coin's landing heads.

Now consider an extra layer of randomization: we make a second coin toss to decide whether to run Classic SB (on heads) or Tuesday SB (on tails). Here is the full protocol:

[4] Advocates of Compartmentalized Conditioning (see §2.3), and Pust (2012), will deny that probabilities which are hypothetical on self-locating beliefs are *conditional* probabilities.

[5] Compare this with Savage's (1954: 21) argument for his decision-theoretic sure-thing principle. Unsurprisingly, there is also a simple Dutch Book argument for the Epistemological Sure-Thing Principle, using three bets. Bet 1 pays out on $\neg A$; Bet 2 pays out on A, but the bet is called off if $\neg C$; Bet 3 pays out on A, but the bet is called off if C. We leave the details to the reader.

Toggled SB. A fair coin is flipped *twice*:

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

All of these possible awakenings are indistinguishable to Beauty. Beauty is fully informed about the setup. When she wakes up, knowing only that it is either Monday or Tuesday, what probability should she assign to *the first flip lands heads*?

The halver must answer ½ in Toggled SB. To see why, recall that the second flip determines whether the experiment is to be run like Classic SB or like Tuesday SB. Conditional on being in either Classic SB or Tuesday SB, the halver assigns ½ to *the first flip lands heads*. Since those are the only two options (and they are exclusive), the halver must unconditionally answer ½.

We can make the same point slightly more formally. Let H be the event that the first flip lands heads; let h be the event that the second flip lands heads; let Pr be Beauty's rational probability function just after waking in Toggled SB. Any halver should say that $\Pr(H \mid h) = ½$, for that is just the probability of H conditional on being in Classic SB. Halvers claim that H and h are probabilistically independent in the sense that $\Pr(H \mid h) = \Pr(H \mid \neg h) = ½$. So, by elementary probability theory, $\Pr(H) = ½$.

We will now argue, though, that the Epistemological Sure-Thing Principle entails that the correct answer to Toggled SB is not ½ but ⅓. Our argument depends on these two claims:

(1) If Beauty learns on Monday that it is Monday, then she should assign ⅓ to Heads.
(2) If Beauty learns on Tuesday that it is Tuesday, then she should assign ⅓ to Heads.

To establish these claims, we will 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 assign her probabilities as follows: $\Pr(\neg H) = ½$ and $\Pr(H \wedge h) = \Pr(H \wedge \neg h) = ¼$.

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 first coin landed heads and the second coin landed tails, i.e. that $H \wedge \neg h$. So, she should now assign her probabilities as follows: $\Pr(\neg H) = ⅔$ and $\Pr(H) = ⅓$. This establishes (1). And since Toggled SB is symmetric in all relevant respects between Monday and Tuesday, the same reasoning establishes (2).[6]

Having established (1)–(2), we have also established Beauty's hypothetical probabilities. Let M be the event that *today is Monday*. By (1), Beauty assigns ⅓ to H hypothetical on M. By (2), Beauty assigns ⅓ to H hypothetical on $\neg M$. Combining this with the Epistemological Sure-Thing Principle, Beauty should assign ⅓ to H (simpliciter). The correct answer to Toggled SB is therefore ⅓.

With this, we see that halving is untenable, since it conflicts with the Epistemological Sure-Thing Principle.

1.3 Relationship to Reflection

We think that our Epistemological Sure-Thing Principle should be uncontroversial. However, our Principle looks similar to van Fraassen's Reflection Principle, which *has* been the subject of much controversy. Indeed, according to Elga, Classic SB itself provides a *counterexample* to the Reflection Principle. So we should explain the important differences between our Principle and Reflection, and explain why we *reject* Rejection but *recommend* the Epistemological Sure-Thing Principle.

Here is a simple version of van Fraassen's Reflection Principle:[7]

Reflection Principle. Where Pr_s represents an agent's probability at time s and Pr_t represents that agent's probability at some later time t:

$$\mathrm{Pr}_s(A \mid \mathrm{Pr}_t(A) = r) = r$$

We can draw a link between Reflection and our Epistemological Sure-Thing Principle. Suppose we stipulate that some agent is sure to learn whether C or $\neg C$ at t, and will assign probability r to A either way, and that nothing else relevant changes between s and t. Then $\mathrm{Pr}_s(\mathrm{Pr}_t(A) = r) = 1$, so that Reflection tells us that $\mathrm{Pr}_s(A) = r$, just as our Epistemological Sure-Thing Principle would.

However, the principles are different: Reflection governs the relationship between probabilities at times; the Epistemological Sure-Thing Principle governs the relationship between hypothetical and non-hypothetical probabilities. Thinking through our scenarios can illustrate this difference.

Elga correctly notes that Reflection leads to halving in Classic SB (and treats this as a counterexample to Reflection).[8] Here's why. Let s be a moment on Sunday before

[6] The argument uses only elementary probabilistic reasoning, without any special issues concerning self-locating beliefs. Specifically, the event we are about to introduce as M (i.e. *today is Monday*) is not *purely* self-locating; it excludes all worlds where the first flip lands heads and the second lands tails (similarly for $\neg M$). So the reasoning just given is recommended by Compartmentalized Conditioning (a rule which we introduce and criticize in §2.3 and §B.2).

[7] van Fraassen (1984: 244).

[8] Elga (2000: §3). But note that *restrictions* of Reflection may not lead to halving; for example, Gallow (n.d.) suggests a restriction of Reflection which does not lead to halving.

Beauty goes to sleep; let t be a moment shortly after Beauty wakes up. Undeniably: at s, Beauty should assign ½ to the event *the coin lands heads*, i.e. $\Pr_s(H) = ½$. Moreover, Beauty is certain at s what will happen at t; so if she is rational she will select r so that $\Pr_s(\Pr_t(H) = r) = 1$; invoking Reflection, this would require that $r = ½$; i.e., Reflection recommends halving.

By contrast, the Epistemological Sure-Thing Principle does not lead to halving in Classic SB; there are no relevant hypothetical (rather than future) probabilities that could force this upon us. Indeed, our Epistemological Sure-Thing Principle *conflicts* with Reflection. This is because, as we argued in §1.2, the Epistemological Sure-Thing Principle leads to an answer of ⅓ in Toggled SB, whereas Reflection leads to the answer of ½ in Toggled SB (in almost exactly the way that it does in Classic SB).

1.4 *Antecedents of the Sure-Thing Argument*

Our Sure-Thing Argument builds on various moves in the extensive literature on sleeping beauty. Considerations about what Beauty should believe, after being told what day it is, go back to Elga's original paper.[9] We also adopt from others the idea of introducing another random event (besides the first coin flip) that Beauty can learn about after waking. Indeed, several authors, including Dorr, Meacham, Titelbaum, and Conitzer, can be seen as giving arguments with this general shape:[10]

(i) Present a tweak of Classic SB, where Beauty receives some apparently "irrelevant" but probabilistically independent information whenever she wakes up.

(ii) Note that *certain* halvers are committed—by their own favoured update-rule—to giving an answer other than ½ in this tweaked scenario.

Our argument develops this general pattern, by instead noting that *any* halver—regardless of what particular update-rule they use—conflicts with the Epistemological Sure-Thing Principle. Our point here is not to show that halvers must "bite the bullet" in insisting that some apparently "irrelevant" information is in fact relevant (for the extra information switches their answer from ½ to something else). Our point is that any halver must deny a very basic constraint on any reasonable way of assigning probabilities: the Epistemological Sure-Thing Principle.

2 The Statistical Argument

We have completed the Sure-Thing Argument. We will now present a second argument against halving: it violates basic statistical reasoning.

[9] Elga (2000: §2).

[10] Dorr (2005) offers an argument of this form against White (2006). Meacham (2008: 263), Titelbaum (2008: 591–9, 2013: 1007–8), and Conitzer (2015: 1987–8) offer arguments of this form against Compartmentalized Conditioning (a rule we discuss in §2.3).

2.1 Keeping Beauty awake

To show this, we will need to introduce further modifications to SB Classic. We start by tweaking Toggled SB. Now, instead of leaving Beauty sleeping through a day, 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. For clarity, here is the full protocol:

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

- If the first flip lands tails, then Beauty is told nothing about the coin flip on Monday or Tuesday.
- If the first flip lands heads:
  - — if the second flip also lands heads: upon awakening on Tuesday, Beauty is immediately told that the first flip landed heads (but she is told nothing on Monday).
  - — if the second flip instead lands tails: upon awakening on Monday, Beauty is immediately told that first flip landed heads (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 either flip). She is fully informed about the setup. When she wakes up and is told nothing about the outcome of the first flip, what probability should she assign to *the first flip landed heads*?

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 in Informed SB, some awakenings are distinguishable: Beauty might be told something about the coin, or she might not. And this matters. To see why, let W be the following claim:[11]
>
> *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 the protocol in Informed SB does not guarantee the same. Instead, in Informed SB, Beauty learns that W only when (and in that case only because) she is not told anything. So: when Beauty learns that W in Informed SB it seems that she *should* revise her probabilities (whereas she should *not* revise her probabilities in Toggled SB).

This challenge is revelatory. We think it is mistaken, but the mistake it makes is genuinely interesting. In particular: it ascribes epistemological significance to a mere

[11] Note that W is *centered*; see §2.3 for what this means.

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 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 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 specify that Beauty is aware of the result of the first coin flip on both Sunday and Wednesday (erasing Beauty's memory as she sleeps from Sunday to Monday). 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 first flip 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 coin lands Heads-then-heads: we end the experiment a day early, on Tuesday, letting Beauty learn on Tuesday what she will find out anyway on Wednesday, i.e. the result of the first flip.

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 coin lands Heads-then-tails: we start the experiment a day late, on Tuesday, letting Beauty remember on Monday what she already found out on Sunday, i.e. the result of the first flip

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. This shows that Toggled SB and Informed SB should, indeed, be treated in the same way.

2.2 Flipping 40,000 coins

Having shown that Toggled SB and Informed SB should be treated alike, we now consider a final change. The basic idea 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 twice. When n is between 1 and 40,000:

- If coin-n lands tails when first flipped, then no information is displayed about coin-n on the screen on either day.
- If coin-n lands heads when first flipped:
  - — if coin-n's second flip lands also heads: the numeral **n** is displayed on the screen on Tuesday (but not Monday);
  - — if coin-n's second flip instead lands tails: 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, k, between 1 and 40,000, is chosen at random; Beauty is told what k is on both Monday and Tuesday. Now: when Beauty does not see the numeral **k** on the screen, what probability should she assign to *coin-k lands heads when first flipped*?

Here is a simple statistical argument that, in Displayed SB, we must assign a probability of about ⅓ to *coin-k lands heads when first flipped*.

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 they were first flipped.[12] Now, in the case we are imagining, coin-k is undisplayed. But there is nothing special about the particular number k here: recall that k was chosen at random on Sunday. So, on statistical grounds alone, Beauty should assign a probability of about ⅓ to the claim that coin-k 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 $n \neq k$. Indeed, since all of the coin flips are independent, any information concerning any *other* coin is intuitively just *irrelevant* to the likelihood that coin-k landed heads when first flipped. So Beauty

[12] 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 $|r - 1/3| < 1/100$.

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. This is our Statistical Argument against halving.

2.3 "Irrelevant" information and Compartmentalized Conditioning

The Statistical Argument inevitably invites a kind of "bullet-biting" response (cf. our discussion in §1.4). We have presented a tweaked scenario, where Beauty has more information than she has in Classic SB; the halver can always just "bite the bullet", and insist that the scenario has changed enough to allow for a different answer. Indeed, we expect that certain halvers will say something like this:

> The answer in Classic SB is ½, because we are considering a fair coin and Beauty receives *no* new information upon awakening. That is my basic intuition. Granted, it is a surprise to find out that "irrelevant" information turns out to be relevant after all (in Displayed SB), leading Beauty to answer ⅓. But I would rather accept this *as* a surprising discovery, than abandon my basic intuition about halving.

In exactly this spirit, advocates of *Compartmentalized Conditioning* will maintain that Informed SB and Displayed SB are relevantly different—that the apparently "irrelevant" information is relevant after all—so that they can give different answers in these cases. They therefore "bite the bullet".

In fact, we think that Compartmentalized Conditioning specifically leads to a further absurdity. But, to explain this we must introduce the rule of Compartmentalized Conditioning. This is a rule for governing how a subject alters their beliefs when their state changes.[13] 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.[14]

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

[14] 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 Epistemological Sure-Thing Principle, as in §1. We think 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, as suggested above: advocates of Compartmentalized Conditioning can and will "bite the bullet" when confronted with our Statistical Argument.

However, the phrase "approximately ⅓" masks an important point, which we should bring out into the open. Simplify the setup of Displayed SB: suppose that there are only 2 labelled coins, rather than 40,000, that coin- that Beauty is asked for the probability of *coin-1 lands heads when first flipped*. Bullet-biting halvers concede that Beauty's answer is affected by learning about coin-2. This is counter-intuitive, but let us grant it. Still, naïvely, we would expect bullet-biting halvers to say that it should be *equally* relevant to Beauty's answer (about coin-1) whether she sees **2** on the display or not; after all, coin-1 and coin-2 are stipulated to be totally independent of one another. So we might expect bullet-biting halvers to say that Beauty's probability of H, hypothetical on seeing **2**, should equal her probability of H, hypothetical on *not* seeing **2**, and in both cases to be ⅓; but we would also expect them to deny the Epistemological Sure-Thing Principle (given §1), and so to deny that Beauty's un-hypothetical probability of H—i.e. her probability in Toggled SB—should be ⅓.

But this is *not* what Compartmentalized Conditioning says. Rather, in our two-coin version of Display SB, Compartmentalized Conditioning says the following:

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

That is: Compartmentalized Conditioning treats these *equally* "irrelevant" pieces of information about coin-2 as being, not just relevant, but *differently* relevant. That strikes us as genuinely absurd.

2.4 *Antecedents of the Statistical Argument*

Our Statistical Argument, like our Sure-Thing Argument, owes much to the extensive literature on sleeping beauty.

In Displayed SB, we consider tossing 40,000 coins. Such use of large numbers may call to mind Elga's Long-Run Argument against halvers. We should explain why our argument is quite different from Elga's.

Elga asks us to imagine repeating Classic SB many times. He notes that "in the long run, about ⅓ of the wakings would be Heads-wakings…" and concludes that "on any particular waking, you should have credence ⅓ that waking is a Heads-waking".[15]

[15] Elga (2000: 143–4).

This Long-Run Argument, in effect, asks Beauty to think of herself as *betting* on the flip and notes (correctly) that she should think of ⅓ as fair *betting-odds*.

Arntzenius offers a compelling rebuttal, on the halver's behalf, to the Long-Run Argument.[16] We can summarize it by considering the following scenario:

Doubled Bet. A fair coin is flipped. You are forced to bet on the outcome. You will specify the betting odds that the coin lands heads; your opponent will then specify the stakes and the direction of the bet. There is a further twist: if the coin lands tails, the bet is deemed to take place twice. *What betting odds should you specify?*

The question has uncontentious answer: you should specify ⅓ for the *betting odds* that the coin lands heads,[17] even though you know the *probability* to be ½. So betting odds and probability can come apart. And, according to the halver, this is exactly what happens in Classic SB: in Classic SB, if the coin lands tails then Beauty is forced to bet on the outcome twice (just as you are in Doubled Bet).

We have rehearsed this, so we can emphasize that our Statistical Argument is *not* a version of the Long-Run Argument. Specifically: Artnzenius's rebuttal of the Long-Run Argument does not carry over to our Displayed SB. Insofar as (fair) betting odds and (rational) probabilities can come apart, we are asking Beauty about *probabilities* in Displayed SB. Our point is that it is overwhelmingly likely that (on either day) roughly 30,000 numerals will not be displayed, of which roughly ⅓ landed heads when first flipped, and coin-k is known not to be special in any way. And we are asking about Beauty's beliefs at one moment of time, not over long-run repetitions.

We have been able to achieve the focus on Beauty's beliefs at one moment of time, thanks to our considerations about framing effects (see §2.1). This allowed us to show that there is no relevant difference between Toggled SB and Informed SB; and from there, to build up to a one-off scenario with large-numbers, i.e. Displayed SB. (It is worth noting that the device of keeping Beauty awake on both days has been used before,[18] but neither in the context of making a statistical argument, nor with a discussion of framing effects.)

Our Statistical Argument also makes use of the point that learning extra "irrelevant" information would seem to constrain Beauty's beliefs. As we noted in §1.4, this basic point has been made many times in the literature.[19] The closest antecedent to our argument is Dorr, who shows how advocates of certain update-rules may need to become "bullet-biting halvers" (i.e. accept that intuitively "irrelevant" information is relevant after all). But our Statistical Argument, like our Sure Thing Argument,

[16] Arntzenius (2002: 56–7).

[17] Let h be the betting-chance you specify; let £s be the stake. Suppose your opponent bets that the coin lands heads: if the coin lands heads you lose £$\frac{s}{h}$; but if the coin lands tails you gain £$2\frac{s}{(1-h)}$ (as the bet happens twice); and these are equally likely. The expected loss/gain is reversed if your opponent bets the other way. So to prevent your opponent from being able to inflict an expected loss on you, you should set $\frac{s}{h} = 2\frac{s}{(1-h)}$, i.e. $h = ⅓$.

[18] Karlander and Spectre (2010: 405) and Conitzer (2015: 1990).

[19] See the references in footnote 10.

is somewhat more general than its precedents: we offer a challenge to *all* halvers, and not just those who use *specific* update-rules, such as Compartmentalized Conditioning. Moreover, as we saw in §2.3, advocates of Compartmentalized Conditioning (specifically) cannot just stop at bullet-biting; they are led to a further absurdity.

3 Conclusion

We have presented the two strongest arguments against the halvers that we can construct: the Sure-Thing Argument and the Statistical Argument. Building on existing arguments in the literature, these point to the high costs that any halvers must pay. In particular, we show that halvers cannot satisfy the uncontroversial Epistemological Sure-Thing Principle, or accommodate basic statistical reasoning.

In both our arguments, we obtain these conclusions by considering tweaks of SB Classic which the halver cannot reasonable distinguish from it. So the halver's position—whilst attractive and intuitive in the original scenario—turns out to be unsustainable when a wider variety of related cases are considered.

A Calculations for 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 N 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 for 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:[20]

	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,

[20] Assuming a modest principle of indifference; she might not split things this way if she had unusual priors concerning which day is more likely.

Beauty learns that it is Monday.[21] 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 the coin's first landing Heads and ½ to it's first landing 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:

	Monday	Tuesday
Heads	¼	¼
Tails	¼	¼

So, according to Compartmentalized Conditioning, Beauty should say that $\Pr(H) =$ ½, and moreover that $\Pr(H \mid M) =$ ½.

However, if Beauty learns that it is Monday in Toggled SB, then Compartmentalized Conditioning recommends that she should revise her answer to a ⅓. (Advocates of Compartmentalized Conditioning will therefore agree with claims (1)–(2) of §1.2.) 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 ⅓ to Heads (specifically, Heads-then-heads) and ⅔ to Tails. At Step 2, she assigns all of this to Monday.

B.3 Compartmentalized Conditioning and Displayed SB

In §2.3, we discussed how Compartmentalized Conditioning treats Displayed SB. Here we will provide the relevant calculations.

[21] 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).

Let N be the number of coins involved. So $N = 40,000$ in the case described in §2, but $N = 2$ in §2.3. 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 §2.3).

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

> 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 §2.3):

Proposition 2: According to Compartmentalized Conditioning, in a Displayed SB setup with N coins, and with $1 \leq k \leq N$: if some numeral is displayed but $\mathbf{k}$ is not, then Beauty should assign ⅓ to coin-k first landing heads.

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

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

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

References

Arntzenius, Frank (2002). 'Reflections on sleeping beauty'. *Analysis* 62.1, pp.53–62. doi: 10.1111/1467-8284.00330.

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.

Gallow, J. Dmitri (n.d.). 'Two-dimensional deference'.

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.

Savage, Leonard R (1954). *The Foundations of Statistics*. John Wiley & Sons.

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.

White, Roger (2006). 'The generalized sleeping beauty problem: A challenge for thirders'. *Analysis* 66.2, pp.114–119. DOI: 10.1111/j.1467-8284.2006.00597.x.

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