This is the end, Beautiful friend

Tagline: Beauty died today. Or maybe yesterday; I can't be sure.

1. Introductory parable

Father Adam is adminstrator for a scholarship program. This year there are two equally qualified applicants, so he goes to Cardinal David and says "Full scholarships to both, I presume." Cardinal David says "No. Half scholarships to each." It's apparent that they disagree, so they consult scripture (Ecclesiastes 1:9): "what has been done is what will be done...there is nothing new under the sun."

It's discovered, however, that for as far back as records go, there has been a single scholarship applicant each year, who's always been awarded a full scholarship. Confounded, Father Adam and Cardinal David again consult scripture (Thessalonians 2:15): "...stand firm and hold to the traditions that you were taught...by our spoken word or by our letter."

But after a thorough perusal of paperwork and minutes turns up nothing, Father Adam finds himself back in Cardinal David's office complaining "There *is* no letter." So they wind up consulting scripture yet again (2 Corinthians 3:5-6): "God...has made us sufficient to be ministers of a new covenant, not of the letter but of the Spirit."

Meanwhile, word of the issue proliferates and many others weigh in. A Father Joseph, citing the loaves and fishes trick, preaches that though the Church should only ante up once, everyone who applies should still get a full scholarship. Eye-rolling followers of St. Peter, under secular heat to deny Josephian faith in divine conceptions, want the applicants to draw lots; St. Patrick's world-wise disciples reply "just give it to whoever asked first."

Finally a mystic (Father Jacob) has a vision on the Baltic Sea according to which two full scholarships will be given on earth but that such action will recoil in the hereafter, with apocalyptic consequences. Father Jacob, who wears a tunic emblazoned with the slogan "The wisdom of your wise men will perish," preaches that, ironically, the wisdom of the wise men need perish precisely because the wise men themselves need not.

A lay majority eventually determines to entrust the decision to Cardinal David. (Who's not the Pope or anything, but he has had more time to think about it now and *is* awfully good at this sort of thing.) But, it turns out that the wise man perishes that very night, in his sleep. Which, as far as the current crisis is concerned, proves damn inconvenient.

Oh well–they'll figure it out eventually.

2. Before dawn

Beauty's days are numbered. (But how?)

In case you've been in a coma yourself, Sleeping Beauty (popularized by Elga 2000) is a rational agent taking part in an experiment. She is awakened Monday morning, asked her credence in *heads*, told what day it is and asked her credence again. If now the toss of a fair coin lands *heads*, she's put back to sleep until Wednesday morning. If *tails*, she's put

back to sleep but, in her sleep, is given a drug that erases all memory of that morning's experiences, and then on Tuesday suffers an analogous interview. Beauty knows all of this in advance. The problem is what her initial personal credence in *heads* should be upon initial wakeup. A *halfer* says one-half. A *thirder* says one-third. For halfers, there is a second issue: Beauty's credence in *heads* upon learning *Monday*.

Long viewed as something of a dead horse by many, Beauty has, of late, almost surely proved wearisome even to those who've been making a living off of her predicament. To parrot one of the more respectable (and apparently more marginalized) papers (Dorr 2005) to emanate from this coterie, I intend to "regale you with yet another" take on Sleeping Beauty. Which is more or less this: halfer credences are to thirder credences as population proportions are to sample proportions in statistical sampling.

That's how David Lewis (2001) saw halfing as well. If it's right then it would be rather strange to refer to halfer credences simply as "personal credences", but not altogether untenable, and, however named, halfer credence would at least survive as a distinct and viable concept. But that isn't the way things have played out in the literature, where a rival form of halfing (so-called "double halfing") has been seeing a lot more ink.

Double halfing tries to be less strange than Lewisian halfing (thus more credible as unadorned "personal credence"), but thereby violates diachronic norms (conditionalization). It procurs any advantage that it has illicitly, pandering to the faulty intuitions responsible for the Monty Hall fallacy. At great cost–double halfing violates *reflection*, a maligned notion due to its appearance in various incorrect guises but surely a constraint on rationality when properly formulated. Unchallenged acceptance of practices in violation of rational constraints erodes our culture–a theme I pursue further in the second half of the paper, where I take on the claim of Jacob Ross (2010) that there is a "deep tension" (leading to "rational dilemmas") between thirder reasoning and countable additivity of credences.

Several approaches to credences are explored here, including frequencies, evidence and solutions to optimization problems. In the last case one can consider gamblers' stakes and maximize *capital* (utils), or invoke information theory and minimize *surprisal*. It's the policy governing accrual of the optimized quantity that matters. The naive view, which Ross (2010) calls[1] *Every Awakening Legitimacy* (*EAL*), is that relevant quantities accrue twice if *tails*. This supports thirding (Elga 2000; Horgan 2004; Rosenthal 2009 etc.). The competing view, *Single Awakening Legitimacy* (*SAL*), according to which relevant quantities accrue precisely once, supports halfing (Bostram 2007; Hawley 2012; Jenkins 2005; Halpern 2004; D. Lewis 2001; P. Lewis 2007; Meacham 2008; Pust 2012; White 2006 etc.) provided Beauty isn't "tipped off" as to the rate of accrual.

I assume (with apologies) familiarity with some concepts in information theory and stochastic analysis. I will however keep the arguments brief, though they are for the most part

[1]More or less. I use the term somewhat more generally than Ross, and with different emphasis. At any rate use of "legitimate" here is innocent. *EAL* is just a convention about accrual of certain decision-theoretic quantities; it's not a philosophical thesis.

novel and, I think, necessary, sometimes improving on or amending deficiencies in extant counterparts. When I've dispensed with an issue I'll move on to the next without fanfare.[2] Fleshing out is usually left as an exercise (in patience). Here's a brief synopsis:

Section 3 is an elaboration on Elga's (2000) argument for thirding, which I generally commend. In the current climate we must view *EAL* was an unformulated premise, but it would have been unrealistic to expect Elga to anticipate the somewhat bizarre rival *SAL*. I do not, however, fully commend what Elga goes on to say about his argument. He's right to say that Beauty's temporal location is relevant to *heads*, but wrong to say that Beauty receives no new information relevant to *heads* upon waking. I'll in fact argue that, though unusual, thirder Beauty's change of belief is quite classical in spirit.

In Section 4 I explain briefly how a broad range of arguments for thirding become arguments for halfing (after David Lewis 2001) when *EAL* is swapped for *SAL*. I then give two informal justifications for the plausibility of *SAL*, since that's where thirders will attack. This step is extremely important[3] (Lewis's argument is underdeveloped, as is that of Jenkins 2005, who defends Lewis): *SAL must* square with the *self indicating assumption*. This is necessary to avoid the "Doomsday" type arguments (I don't analyze these carefully, but see Meacham 2008) that defeat what one might call "overgeneralized Lewisian halfing".

Having escaped *Doomsday* there's no reason not to dispense with double halfing. I do this in Section 5, explaining along the way why the Monty Hallish example I use is more ruinous than other merely "embarassing" attacks (such as Titelbaum 2012; see however Dorr 2005 for a devastating treatment I don't much improve on here–other than by making it more clear that Lewis emerges unscathed).

In Section 6 I defend countable additivity of rational credences, both by an improved direct argument and by showing that one of the assumptions Ross needs to generate "rational dilemmas" from thirder reasoning spawns them *by itself*–i.e., independently of thirder reasoning. Reading between the lines, I may appear to press further, suggesting that Ross's scenarios can (or should) be ignored in virtue of their high rational cost and specious plausibility. I confess to the grounding temper (faith in the metaphysical possibility of Ross's scenarios as rank superstition), but loyalists can choose to stand by their nihilism about rationality for all I say to the contrary here. My point is that their pessimism can't discriminate between halfism and thirdism.

[2]I am indebted to an anonymous referee for helpful comments relating to philosophical jargon and several others for barometric readings on what can go wrong when mathematicians try to talk to philosophers. (Essentially everything–thanks to those who responded.)

[3]Having undermined extant justifications for a one-half solution, the safer bet would be to fall in with the thirders. But as explained in the text, halfing has more going for it than just coherence, elegance and similarity to population proportions in statistical sampling (or for an esoteric example, equal slice measures in combinatorial mathematics); it provides continuity (where thirding doesn't) in some aspects of decision-theoretic practice.

Finally, in Section 7 I make a brief comparison with a related puzzle (appearing in Arntzenius 2003), that of "The Prisoner".

3. Twice told tails: thirding and conditionalization on uncertain evidence

Subject to *EAL*, betting and frequency arguments capably vindicate thirding, as has been argued by many others. Another type of argument is from *surprisal*.[4] If an agent has credence p in A, her surprisal (the number of bits of information acquired) upon learning A is $-\log_2 p$. Since to know more now is to be surprised by less later, agents seek to minimize surprisal. According to *EAL*, Beauty is surprised twice if *tails*, so her expected surprisal during the experiment is $-\frac{1}{2}\log_2 p - 2 \cdot \frac{1}{2} \cdot \log_2(1-p)$, minimized at $p = \frac{1}{3}$.

Arguments from evidence have spawned the most lively debate. Some thirders (such as Elga 2000 and Arntzenius 2003) agree with Lewis that Beauty has no new evidence for *heads* upon waking, while others (such as Horgan 2004) say that she has. *EAL* (which thirders must accept) implies that she has. The road to why starts with the:

> *Self Indicating Assumption (SIA).* Let $h_1, h_2, \ldots, h_k$ be mutually exclusive events. In the absence of further evidence, an observer's credence in h_i should be proportional to the product of h_i's objective chance and the expected number of novel observations, conditional on h_i.

SIA is mandatory (modulo explication of *novel*) for rational agents. Cf. the so-called *Doomsday* argument: denial of *SIA* and near-one-half objective chance of near-term human extinction implies that, conditioned on self-locating evidence that one is an early human, near-term extinction becomes practically certain. As to how *SIA* works, to count novel observations in normal cases is to count congruence classes under the *same observer* relation. What's not clear is whether non-communicating time-slices of the same individual qualify as different observers. Applied to the accrual of novel observations as a decision-theoretic quantity, *EAL* says *yes*: two novel observations if *tails*, one if *heads*. *SIA* and Elga's indifference principle (which says that *Monday* and *Tuesday* should be taken as equally likely conditional on *tails*) together yield the familiar Elgain centered credences:

	Monday	*Tuesday*
Heads	$\frac{1}{3}$	0
Tails	$\frac{1}{3}$	$\frac{1}{3}$

So the one-third solution follows from indifference, self-indication and *EAL*.

[4]From a surprisal perspective, *EAL* seems indicated when Beauty is debriefed after each interview, whereas a single deferred surprisal associated to multiple non-communicating interviews may generate sympathy for *SAL*. Credences shouldn't depend on debriefing, so this dialectic further supports the convention hypothesis in the *EAL*/*SAL* debate. (Non-logarithmic scoring rules, incidentally, which are inappropriate as measures of information gain, have also been discussed in the Sleeping Beauty literature. I don't know why.)

But did Beauty gain evidence? Is the observation expressible as *I am awakened now* informative? In normal cases, an observation is informative precisely when one's prior probability in it is less than 1. Horgan (2004) claims that Beauty's priors are given by:

	Monday	*Tuesday*
Heads	$\frac{1}{4}$	$\frac{1}{4}$
Tails	$\frac{1}{4}$	$\frac{1}{4}$

Pust (2008) disagrees, and indeed, it doesn't seem that Beauty can have access to her previous timestep priors at all, as her final prospective centered credences are different on Monday than they are on Sunday. One could try to take a weighted average of her priors over candidate previous timesteps. Doing this, I suspect $P_{-}(wakeup)$ could come out to be strictly less than 1: if not debriefed, Beauty's final Monday credence in *wakeup* is such. Dependence on debriefing has its pitfalls, though, so I won't pursue this line.[5]

Jenkins (2005), defending Lewis, writes "Beauty's subjective experience on waking is exactly the same' (however the toss lands)....(so) Beauty has no interestingly new evidence on waking." But such an antecedent requires only that credence be uniform across awakenings. It's her *tails* experience *throughout the experiment* that reflects on *heads*, and (if we grant *EAL*) this *is* interestingly different from her *heads* experience–it consists in two experimental awakenings. That's interesting because it implies that experimental awakenings are relevant to *heads* (by virtue of confirming *tails*) to the degree that they might be second awakenings.

It's tempting to reply that although Beauty may have seen evidence sufficient to confirm *tails*, she can't remember it and, therefore, can't factor it into her credences. But it's classical (cf. so-called *Jeffrey conditionalization*) that uncertain evidence does play a role in rational credence formation.[6] That her current uncertain evidence is not grounded in any certain proposition seems neither here nor there. She can discount neither that this might be a second legitimate awakening, nor that, if it is, the coin surely landed *tails*. Rational agents (thirders and halfers alike) are required to condition on what evidence they've seen, and that includes accounting for available evidence they *might* have seen.

The familiar calculation makes clear where Beauty's information comes from.

$$P(heads) = P(2^{nd})P(heads|2^{nd}) + P(\sim 2^{nd})P(heads| \sim 2^{nd}) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}.$$

[5]Another option is to parse *now* rigidly. If today is Tuesday, *I am awakened now* means *I am awakened Tuesday*, which is informative. If *I am awakened now* might be informative, it is informative, i.e. relevant to decision theory and epistemology. (Dretske's *Knowledge and the Flow of Information* to the contrary notwithstanding, that's how information works.) This is more or less what I do in the text; see also the final section.

[6]It's ironic that Jeffrey conditionalization appears to be a superfluous addition to classicism in classical cases, where, given observations that have led one to an uncertain conviction, one can always just condition on the fact that there were such observations.

There remains the question of what to do upon learning *Monday*. In the usual case, for a potential observation e there is an uncentered proposition L_e such that L_e holds if and only if e is observed in the prevailing epistemic circumstance. In Beauty's case something similar holds, but with a change: L_e is now a *centered* proposition. Elga recommends conditionalization of centered credences on the centered event L_e upon observation of e. Pust (2012) makes a counterargument that runs somewhat as follows: one cannot condition on L_e because one has no prior credence in L_e. Say at 9:05 Beauty learns *Monday*. What then is the content of L_e? Not *I learn Monday at 9:05 on Monday*, which she already knew, but *it's now 9:05 on Monday*. However, she had prior credence 0 in the assertion expressed by those words at the previous timestep (say at 9:00).

This strikes me as more of a filibuster than a serious objection. In local coordinates, one has, at timestep $z := now$, a probability distribution P_z over future observations $\{e_{z+t} : t > 0\}$. (On a view standard in stochastic analysis, that's just what a credence function is.) Now upon observation of $B = \{e_{z+t} : 0 < t \leq q\}$, one has, e.g.,

$$P_{z+q}(\{e_{z+q+t} : 0 < t \leq r\}) = \frac{P_z(\{e_{z+t} : 0 < t \leq q + r\})}{P_z(\{e_{z+t} : 0 < t \leq q\})}.$$

Expressing new evidence in local coordinates like this avoids proliferation of confusing indexicals, skirting the aforementioned "objection". A further attempt at stonewalling is that later when $now = z+t$ the observation e_{now} won't have the same cognitive significance as the centered event $\{e_{z+t}\}$ contemplated at timestep z. By analogy with Hesperus and Phosphorus, this worry must surely be based on the idea that, although $now = z + t$, Beauty doesn't *know* that $now = z + t$. But she does, so that's another non-issue.

4. Asked and answered: halfers as dilutional self indicators

Whereas *EAL* supports thirding, *SAL* supports halfing. But *SAL* is in want of defense. The standard defense of halfing (*no new evidence*) requires *SAL* as a premise, so it can't help, but halfers have resources in the realms of frequency, information...and wagering.

Indeed, the literature is rife with betting protocols mirroring *SAL*. Bostrom (2007) proposes a thought experiment (*Beauty the high roller*) where bets are offered to Beauty on Mondays only. Adding phony bets on Tuesdays so she won't be tipped off, Beauty then follows Hawley (2012), who assigns *Monday* probability 1 conditioned on *tails*. Scheduling the one bet on Monday or Tuesday with equal likelihoods conditioned on *tails* recalls Peter J. Lewis's (2007) *quantum Sleeping Beauty* interpretation. Shaw (2013) introduces bets that Beauty can make only (and only *once*) by agreeing to them during each awakening of the experiment, a global reward perspective evoking Lewis's (2001) answer (borrowed from statistics) to tails world oversampling: sample weight dilution of the *tails* awakenings.

Still, one needs a reason to regard single-bet protocols as something other than a contrivance in the underspecified case. Nor again should the reason contradict *SIA*. Otherwise, in order to avoid *Doomsday* halfers will have to violate diachronic norms, and that's

a bad idea. (See below.) Here then are two senses (relating to betting and information, respectively) in which *SAL* naturally extends *credence*'s prior conceptual underpinnings without compromising *SIA*. (A sense relating to frequency is left to a footnote.)

1. *Day late, dollar short*: in wagering scenarios, capital goes proxy for utility, and in normal cases rational agents have access to their utility balance. (If agents can't, upon ideal reflection, figure out that they've been punished or rewarded, they haven't been.) If we preserve this, Beauty's utility can't take a double hit: her utility function gets reset along with her memory. Otherwise, thinking of betting *heads* on Tuesday, she will, upon reflection, note that she's a "dollar short", infer that she's a "day late", and not bet.

2. *Asked and answered*: in normal cases, accrued surprisal measures information acquired since initiation of scoring. Preserving this[7], one must abandon realtime surprisal as basis of the rule's explication to Sleeping Beauty scenarios. This creates an injunction against double counting of information, leading to *SAL* implemented by Lewisian dilution of Beauty's *tails* awakenings.[8]

These observations indicate that certain properties of the decision theory grounding prior use of "credence" are only preserved by *SAL* (as others are preserved only by *EAL*). These aren't skeptical nonsense properties like *quassociativity*, which one might claim grounds a rival explication (*quus*) of *plus* to heretofore untested cases. What we have here are two sets of ordinary, sensible properties of decision theoretic practice that were always coupled before but have come apart now. Nothing about our prior use determines which set we must aim to preserve. Indeed, we should expect this choice to depend on our intentions.

It's less controversial that when one incorporates *SAL* into arguments for thirding, they become arguments for halfing. It's obvious that one-wager protocols (betting paradigm versions of *SAL*) lead to halfing behavior, and that when counting just one *tails* awakening per *tails* toss the long-run frequency of *heads* awakenings will be one-half. It's equally

[7]Logarithmic scoring requires information theoretic independence of iterated questions in cases where answers are withheld, a condition violated under *EAL*. A similar point about stochastic independence in the frequency argument was made by Schervish et. al. (2004), who wrote "the repeated trials in Sleeping Beauty's game do not form an independent sequence, and her mandated forgetfulness precludes any 'feedback' about the outcome of past previsions. When repeated trials are dependent and there is no learning about past previsions, coherent previsions may be very badly calibrated in the frequency sense."

[8]Strange as dilution is, the alternatives are worse. Hawley (2012) for example uses a principle of "inertia" to peddle wholesale disenfranchisement of Tuesday's awakening, assigning credence 1 to *Monday*. Inertia solves some problems but cripples one's ability to respond appropriately to new evidence. Indeed, if Beauty buys in literally and the coin lands *tails* she'll spend the rest of her life (starting Tuesday) thinking it's a day earlier than it actually is, mounting evidence to the contrary notwithstanding. (On the other hand, quantum style disenfranchisement seems to be a viable alternative to dilution.)

straightforward how the information-theoretic argument gets changed: under *SAL*, only one *tails* surprisal should be scored, which means that the quantity to be minimized is $-\frac{1}{2}\log_2 p - \frac{1}{2}\log_2(1-p)$ (this occurs at $p = \frac{1}{2}$). The thirder argument from evidence, meanwhile, required multiple awakenings; initial awakenings aren't relevant to *heads*. Granting that *Monday tails* and *Tuesday tails* should be given equal credence (an "indifference" principle proposed by Elga 2000 and accepted in most of the literature), *SAL* cashes out as Lewisian centered credences:

	Monday	*Tuesday*
Heads	$\frac{1}{2}$	0
Tails	$\frac{1}{4}$	$\frac{1}{4}$

These numbers also serve as Lewis's priors, which explains Lewis's claim that Beauty learns nothing on waking. From the standpoint of her decision theory, Beauty's diluted halves exist in parallel. Both are viewed as immediate successors to Beauty's waning Sunday moments–no part of her experiences both *tails* awakenings.

Half-baked? Approximately. When Beauty learns *Monday*, though, watch out.

5. 'Deal' breaker: double halfers and the flouting of protocol

The most infamous artifact of dilution is that when a Lewisian Beauty learns *Monday*, her credence in *heads* jumps to $\frac{2}{3}$. Self-indication cuts both ways, and Monday's wakeup counts as just half an observation if *tails*; put another way, Monday half-awakenings confirm *heads* to the degree that they might be second half-awakenings. If that's not strange enough, consider fellow Lewisian Sleeping Gorgeous, who gets awakened once if *tails*, twice if *heads*. Gorgeous has credence $\frac{1}{3}$ in *heads* upon learning *Monday*. She and Beauty, who we can take to have been awakened in the same room, agree about how to determine credences, can talk to each other about their evidence, trust each others' judgments and yet find themselves on opposite sides of objective chance concerning a future toss of a fair coin.

That's *too* strange, say thirders and some would-be halfers. A "double halfer" is a halfer who continues, contra Lewis, to assign *heads* credence one-half upon elimination of a *tails* scenario.[9] Double halfing is halfing combined with a scheme whereby Beauty updates propositional credences in response to centered evidence by conditioning on the proposition corresponding to the set of worlds consistent with the evidence. Such updating is advocated by (Bostrom 2007; Halpern 2004; Meacham 2008; Pust 2012; White 2006).

Bostrom refers to his such brand of halfing as a "hybrid model". Indeed, double halfers appear to sport multiple personalities. Like Lewis, they start out in apparent deference

[9]Or at least tries to. As shown in Cian Dorr's refreshing unpublished manuscript (2005), if there are n equally likely, mutually discriminable ways that Beauty's awakening could go down, double halfer credence in *heads* upon observing one of them is actually $\frac{n}{3n-1}$, so that in practice double halfers are effectively thirders. Lewisian halfing doesn't suffer this amusing feature (cf. *day late, dollar short* and *asked and answered*, which aren't based on anything like identicality of awakenings).

to *SAL*, but when a *tails* scenario is eliminated, double halfers assign full weight to the remaining one, which is indicative of a switch to *EAL*. The result is a halfing scheme that looks to be a kneejerk response to Lewisian strangeness.

However, it's a scheme that fails viability by virtue of its neglect of *protocol*. On the one hand, there are no natural betting/scoring protocols under which Beauty should behave as a double halfer. (Under such a protocol, how many legitimate bets will Beauty make if *tails*? One on Monday and one on Tuesday, for a total of...one. That's not good.) The locus classicus for evidential protocol's role in updating meanwhile is Monty Hall, and in fact one defender of Lewisian halfing (Jenkins 2005) promises that a "fruitful comparison" can be made between Monty Hall and the problem of how halfers should update upon learning *Monday*. It's possible to deliver on this promise in an extremely direct way.

Suppose that a **big prize** is hidden behind one of three doors, each with equal objective chance. The hypothesis *Door i* corresponds to the state of affairs in which the **big prize** is behind Door *i*. If *Door 1*, Beauty will have a single awakening, on Monday. If *Door 2*, Beauty will have a single awakening, on Tuesday. And, if *Door 3*, Beauty will have two awakenings, on Monday and Tuesday. Halfers of course assign each of the alternatives credence $\frac{1}{3}$ upon awakening.

Suppose now that a halfer learns what day it is, and is asked for her updated credence in *Door 3*. Note: if *Monday*, *Door 1* is eliminated. If *Tuesday*, *Door 2* is eliminated. *Door 3* cannot be eliminated. Recall that our halfer has prior credence $\frac{1}{3}$ in *Door i* for each *i* and, if she accepts Elga's principle, *Monday* and *Tuesday* are equally likely conditioned on *Door 3*. Suppose our halfer learns *Monday*. Since the current protocol is isomorphic to that of the Monty Hall problem, her situation is precisely that of a Monty Hall contestant that has initially chosen *Door 3* and seen the hypothesis *Door 1* eliminated.

Accordingly, halfers who update credences by conditioning on *not Door 1* are committing the very error of those who answer $\frac{1}{2}$ in the Monty Hall problem, in defiance of the understood protocols. On the contrary, Beauty's credence in *Door 3* must remain $\frac{1}{3}$.[10]

This "embarrassment" for double halfers differs from that of Titelbaum's (2012) in an important respect. The main consequence of his observations is that if Beauty subscribes to Elga's indifference principle and performs the fateful toss herself (hence a corresponding meaningless toss on Tuesday) then in order to maintain credence $\frac{1}{2}$ in Monday's toss landing *heads* she has to assign credence $\frac{5}{8}$ to the centered proposition *today's toss will*

[10]What else? It violates reflection for Beauty to update credence in *Door 3* from $\frac{1}{3}$ to $\frac{1}{2}$ upon learning what day it is *regardless of what day it is*. In fact that's Rosenthal's (2009) argument for thirding; alter the original problem so that the single *heads* awakening occurs on either Monday or Tuesday (with equal probabilities). Rosenthal takes it as uncontentious (double halfers agree) that Beauty's credence in *heads* upon learning *Monday* is $\frac{1}{3}$. The same holds for *Tuesday*, so absolute credence in *heads* must be $\frac{1}{3}$. (The argument doesn't net Lewis, for whom Beauty's credence in *heads* upon learning *Monday* is $\frac{1}{2}$.)

land heads. As this applies to Lewis as well, Titelbaum clearly intends for his indictment to extend to other halfers, and only singles out double halfers because Lewis has already embraced similarly counterintuitive consequences in print.[11] The mishandling of Monty Hall, however, isn't merely an embarrassment...it's a deal breaker. And as Lewis responds correctly to the given protocol, it's entirely on double halfers.

6. Sleeping Methuselah: on self indication and countable additivity

Rational credences are generally taken to be constrained by:

> *Countable Additivity (CA).* For any countable, pairwise incompatible set of propositions, the sum of one's credences in the propositions in the set must equal one's credence in their disjunction.

Not by everyone. Some question the legitimacy of the several extant Dutch Book arguments in support of *CA*. An argument with finitely many stakes should answer these questions.

Let X be a random variable on the naturals and consider a credence function P such that $\sum_{n=1}^{\infty} P(X = n) = 1 - \epsilon < 1$. For a large M, let $(X_i)_{i=1}^{M}$ be independent random variables distributed as X is. An agent subscribing to P has X_i revealed to her in turn. After $X_1, \ldots, X_{i-1}$ are revealed, she may bet a dollar that $X_i > \max\{X_j | 1 \leq j < i\}$. If she wins, she gets $\frac{2}{\epsilon}$ dollars. For any k, $P(X_i > k) \geq \epsilon$, so she'll take the bets.

Next, imagine that we have $M!$ agents, all subscribing to P. Each is assigned a different permutation π of $\{1, 2, \ldots, M\}$ and is offered a series of bets like that of the previous paragraph, but with the X_i's revealed in the order $X_{\pi(1)}, \ldots, X_{\pi(M)}$ (the agent wins the ith bet if $X_{\pi(i)} > \max\{X_{\pi(j)} | 1 \leq j < i\}$). They all bet from the same account. To break even, the proportion of bets they win must be at least $\frac{\epsilon}{2}$. But if X_i is the kth largest out of $X_1, \ldots, X_M$ (ties broken arbitrarily), the probability of a randomly selected agent winning when X_i is revealed is at most $\frac{1}{k}$, meaning that the proportion of winning bets is at most $\frac{1}{M}(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{M}) \approx \frac{\log M}{M}$, which tends to zero as M increases. For large M, the P-subscribers collectively suffer a sure loss, so it's irrational to subscribe to P.

Ross (2010) doesn't reject *CA*, but he does claim that there are situations in which one is unable to subscribe to thirder reasoning while simultaneously satisfying *CA*. The one he describes is a Sleeping Beauty problem ("a problem in which a fully rational agent, Beauty, will undergo one or more mutually indistinguishable awakenings..." where the number of such awakenings is a function of a discrete random variable taking values in a set S of "hypotheses") in which the expected number of awakenings is infinite. His claims about what thirders are committed to starts with the following "indifference principle":

[11] Not everything counterintuitive is embarrassing, and I see little reason why Lewis's $\frac{2}{3}$ should be more embarrassing than his original choice of $\frac{1}{2}$, which is equally (and as intentionally) bad at conforming to rational expectations under the more natural premise *EAL*–surely the source of any intuitive insult. Ditto Titelbaum's $\frac{5}{8}$ (for Lewis, anyway).

Finitistic Sleeping Beauty Indifference (FSBI). In any Sleeping Beauty problem, for any hypothesis h in S, if the number of times Beauty awakens conditional on h is finite, then upon first awakening, Beauty should have equal credence in each of the awakening possibilities associated with h.

Note: *FSBI* is too strong. All thirders are committed to is adoption of solutions to appropriately formulated (in particular, employing *EAL* as a premise) optimization problems whenever such exist. It's easy to show that such solutions satisfy indifference, but thirder reasoning is silent in the no-solution case. Ross's thesis should be framed as tension between *EAL* and the principle that rational credences exist in all logically possible scenarios; I for one would accept the argument as a successful *reductio* against such a principle.

At any rate *FSBI*, together with some additional premises (details omitted), leads to a:

Generalized Thirder Principle (GTP). In any Sleeping Beauty problem, upon first awakening, Beauty's credence in any given hypothesis in S must be proportional to the product of the hypothesis' objective chance and the number of times Beauty will awaken conditional on this hypothesis.

A pathological example is introduced, purporting to show that *GTP* is in conflict with *CA*:

Sleeping Beauty in St. Petersburg (SBSP). Let $S = \mathbf{N}$ and suppose that Beauty awakens 2^X times, where X is a random variable with $P(X = n) = 2^{-n}$, $n \in \mathbf{N}$.

If Beauty subscribes to *GTP*, then in *SBSP* it would appear that she must assign equal credences to the exhaustive and mutually exclusive assertions $X = n$, which violates *CA*.

As mentioned, in *SBSP* the expected number of awakenings, $\sum_{h \in H} Ch(h)N(h)$, is infinite. Here $Ch(\cdot)$ denotes objective chance and $N(h)$ is the number of awakenings associated with h. So *SBSP* can't be faithfully implemented at our world, nor at any nomologically accessible world, nor for that matter at any world subject to a reasonably time stationary threat of mortality (which, arguably, includes all metaphysically possible worlds). As the example requires infinite expectation in order to do its work, it isn't clear, therefore, how to interpret Ross's reports of a deep tension between *GTP* and *CA*.

More seriously, in the context of Ross's ambitions *GTP* is a red herring. Ross argues from conflict between *GTP* and *CA* to rational dilemmas, i.e. "contexts in which full rationality is impossible". But what if the only worlds at which the conflict can arise are so crazy that *everyone* finds it impossible to achieve full rationality there? Ross takes this possibility seriously, for he briefly considers the following premise:

Sleeping Beauty Indifference (SBI). In any Sleeping Beauty problem, for any hypothesis h in S, upon first awakening, Beauty should have equal credence in each of the awakening possibilities associated with h.

Ross notes that if everyone is committed to *SBI*, then everyone should reject *CA* (hence full rationality is impossible for everyone), regardless of whether they accept *GTP*. This would undermine his thesis, and he's quick to deflate it, in particular by substituting *FSBI*

for *SBI*, which he hopes will pull halfers back in line with *CA*. But *SIA* is advisable for all (including halfers), so everyone taking *SBSP* at face value should *still* reject *CA*. For any world supporting faithful implementation of *SBSP* will also support a version with unique subjects, and self indication in such contexts still runs afoul of *CA*.

Not that it's a plausible alternative anyway, but halfers can't even get out of this by rejecting *SIA*, as face-value interpretation of *SBSP* wrecks rational decision theory entirely on its own. For consider a situation in which Beauty has been sentenced to an *SBSP*-style incarceration (without memory erasure) involving mild torture. She's free to choose between two rival detention facilities (A and B) to carry out the sentence. Each has already computed the number of days they would confine her. She's chosen A, but this choice is arbitrary. Now she will be offered a sequence of two trades that she'll have to accept if she believes that her expected time of incarceration is infinite, but which will leave her worse off. First, the judge (who doesn't know the values N) offers her a halving of her sentence to switch facilities. By indifference, she accepts, and switches to B. Next, the judge asks representatives of A to reveal their number and offers to let her switch back. At a price–the quadrupling of her previously halved sentence. This is twice as much time as she was originally going to serve. Nevertheless she accepts, as $E(N_B) = \infty$.

The upshot is that there are finite expectation constraints on rationality. For Beauty:

> *Finite Expectation (FE).* In any Sleeping Beauty problem, if $N(h)$ denotes the number of awakenings associated with h, then Beauty's credences $\{P(h) : h \in S\}$ should satisfy $\sum_{h \in S} P(h)N(h) < \infty$.

Some have interjected that adoption of *FE* is tantamount to changing the subject–avoiding issues rather than engaging with them. Only professional desperation could drive anyone to such an extreme form of irrational pessimism. *Let* Beauty reject *FE* (by taking *SBSP* at face value). According to the example, she's abandoned effective decision (and so embraced nihilism about rationality), irrespective of her views on self indication. That self indication can make trouble for *CA* now is quite beside the point. Everybody knows she's got troubles and, in point of fact, she no longer qualifies as a rational agent anyway.

The question of the steep rational toll of eschewing finite expectation constraints vs. the power-to-model or expressiveness costs, if any, of adhering to them, meanwhile, isn't necessarily uninteresting. I've stated my views but haven't argued for them. Suffice it to say that it's an old topic (see, e.g., Arntzenius and McCarthy 1997 or Gallager 2014 (Chapter 6, esp. Summary[12])) quite separate from Ross's ostensive concerns.

[12]In particular, the "paradox" regarding countable additivity for null recurrent Markov chains (*SBSP* qualifies, as described) isn't new; asymptotic "forgetfulness" as to number of transitions since initiation is a by-product of finite-state mind, so it's never been necessary to postulate *SB*-style memory erasure in order to generate "violations" of *CA*. The folklore view (canvassed by Gallager 2014) that has emerged in the sciences provides stark contrast to Ross's comparatively radical conclusions. Indeed, the latter echo McGee 1999,

7. Stop, thief! The unequivocality of ostensive indexicals and The Prisoner

A comparison of Sleeping Beauty with Arntzenius's (2003) Prisoner is enlightening. The Prisoner is waiting in his cell, where there is no clock, hoping for a stay of his scheduled execution. Right now, his credence in *I am executed* is one-half. A helpful guard will turn out the light in the cell at precisely midnight if and only if he is to be executed. Otherwise the light stays on. At 11:59, the light will surely be on but The Prisoner won't be sure whether or not it's past midnight, and will take his suspicion that it might be as partial evidence in favor of his stay having been granted. If his internal clock (apart from the light being on) assigns *after midnight* probability one-half then, like Beauty's credence in *heads*, The Prisoner's credence in *executed* will have dropped to one-third.

The analogy probably shouldn't be pressed. The Prisoner is a more typical sort of Bayesian than Beauty. Suppose he sees a clock at 6 P.M. By propogation of this evidence through his time slices he learns, for any future "internal time" x, the probability $c(x)$ that the actual time is past midnight, and it's conditionalization on the further fact that the light is on at the internal time x he experiences at 11:59 (an uncentered proposition) that causes his credence in *executed* to have fallen; Beauty's evidence is intrinsically *centered*.

On the other hand, to press the analogy–say by parsing "the light is on *now*" rigidly–as "the light is on *at 11:59*" or "the light is on *at 12:04*", as the case may be–isn't technically *wrong*. Such a move introduces uncertainty in the referent of the indexical *now*, just as we saw in Beauty's case (see footnote 5 above). What matters is whether *now* is before or after midnight–just as for Beauty what mattered was whether *now* was Monday or Tuesday. If he were to learn *before*, he would conclude that he has no evidence. If he were to learn *after*, he would have certain evidence for *stay*. So credences can be determined by averaging over possible referents, same as for Beauty.

But, whereas The Prisoner can avoid this sort of rigidity, it appears to be forced on Beauty. She has to parse "I am awakened *now*" as "I am awakened on Monday" or "I am awakened on Tuesday" because *when my internal state is so-and-so* is equivocal where the ostensive *now* is not. This, I think, is what makes the Sleeping Beauty problem vexing. "I am awakened *when my internal state is so-and-so*" is what seems uninformative.

So it is...it's just not what "I am awakened *now*" means.

---

who (observing that he would trade a jelly bean for a place in heaven, if it exists) opined that decision theory can tell us, at best, "how to comport ourselves in the gambling hall or the brokerage house". The conservative view is to reject this attitude, as (a) in the relevant sense, science independently describes our universe as a "gambling hall", (b) it's irrational to do otherwise, and incidentally (c) if one insists (inadvisably) on entertaining other-wordly consequences for worldly actions then one must account for the (more metaphysically respectable) scenario of a recurrence under which one surrenders such "beans" in perpetuity.

References

Arntzenius, Frank. 2003. Some problems for conditionalization and reflection. *Journal of Philosophy*. 7: 356-370.

Arntzenius, Frank and McCarthy, David. 1997. The two envelope paradox and infinite expectations. *Analysis*. 57:42-50.

Bostrom, Nick. 2007. Sleeping beauty and self location: A hybrid model. *Synthese*. 157:59-78.

Dorr, Cian. 2005. A challenge for halfers. Unpublished manuscript. Available at http://users.ox.ac.uk/˜sfop0257/papers/ChallengeForHalfers.pdf

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

Gallager, Robert G. 2011. *Stochastic Process: Theory for Applications*. Cambridge University Press. 2014.

Halpern, Joseph. 2004. Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems. *Oxford Studies in Epistemology.* Oxford University Press.

Hawley, Patrick. 2012. Inertia, Optimism and Beauty. *Nous*. 47:85-103.

Horgan, Terry. 2004. Sleeping Beauty awakened: new odds at the dawn of the new day. *Analysis* 63:10-21.

Jenkins, Carrie Ichikawa. 2005. Sleeping Beauty: A wake up call. *Philosophia Mathematica* 13:194-201.

Lewis, David. 2001. Sleeping Beauty: Reply to Elga. *Analysis* 61:171-176.

Lewis, Peter J. 2007. Quantum Sleeping Beauty. *Analysis* 67: 59-65.

McGee, Vann. 1999. An airtight Dutch book. *Analysis* 59:257-265.

Meacham, Christopher. 2008. Sleeping Beauty and the Dynamics of *De Se* Beliefs. *Philosophical Studies* 138:24569.

Pust, Joel. 2008. Horgon on Sleeping Beauty. *Synthese* 160:97-101.

Pust, Joel. 2012. Conditionalization and Essentially Indexical Credence. *Journal of Philosophy* 109:295-315.

Rosenthal, J. S. 2009. A mathematical analysis of the Sleeping Beauty problem. *Mathematical Intelligencer* 31:32-37.

Ross, Jacob. 2010. Sleeping Beauty, countable additivity, and rational dilemmas. *The Philosophical Review* 119: 411-447.

Schervish, M.J., Seidenfeld, T. and Kadane, J.B. 2004. Stopping to reflect. *Journal of Philosophy*. 6:315-322.

Shaw, James R. 2013. De se belief and rational choice. *Synthese* 190:491-508.

Titelbaum, Michael. 2012. An embarrassment for double halfers. *Thought* 1:146-151.

White, Roger. 2006. The generalized Sleeping Beauty problem: a challenge for thirders. *Analysis* 66:114-119.

rmcctchn@memphis.edu