Inference for Multiple and Conditional Observers

M. T. Barlow*

November 25, 2024

Abstract

We consider models for inference which involve observers which may have multiple copies, such as in the Sleeping Beauty problem. We establish a framework for describing these problems on a probability space satisfying Kolmogorov's axioms, and this enables the main competing solutions to be compared precisely.

1 Introduction

This paper is concerned with inference by observers whose existence and multiplicity is affected by the outcome of the random experiment which they are observing. One can argue that implicit in the classical model of a random experiment is the assumption that there is an Observer (which I will call a Classical Observer or CO) which makes observations and inferences, exists before the experiment is performed, and whose existence is not affected by the outcome of the experiment. Since these conditions are nearly always satisfied in practice, little attention has been paid to this requirement by the statistics community.

However there are some problems in physics and philosophy which involve non Classical Observers. One example comes from cosmology. The Standard Model gives a very complete description of physical systems in which relativistic effects can be ignored, but has 19 or so free parameters. (The exact number depends on the version of the Standard Model.) Writing Θ for this parameter space, it appears (see for example [1, 4, 37]) that only a small subset Θ_L of Θ gives rise to universes with complex chemistry. For the sake of our illustration, let us make the simplifying assumptions that complex chemistry is necessary for observers, that there is a natural probability measure μ on Θ, and that $\mu(\Theta_L) \ll 1$. Since we have to be in a universe with observers, when we measure θ we find that $\theta \in \Theta_L$. The point which is disputed is whether this observation, which could not have been different from what it was, can give any information. For more background on this see [8, 18]. This example will be discussed briefly below in Section 7.5.

The cosmological problem outlined above has conditional observers. Following the literature, and the terminology introduced by Brandon Carter in [10], I will call these *Anthropic Observers* or AO. (The term is flawed as these observers do not have to be human.) Some cosmological models also involve the possibility of multiple observers, and

*Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada. barlow@math.ubc.ca

one needs to know how the multiplicity of these observers affects inferences. (For an example of a recent paper in this area see [39].)

A 'toy model' one can use to investigate these questions is the Sleeping Beauty problem, described by A. Elga in [16]. The name is due to Robert Stalnaker, an earlier version of the problem is due to A. Zuboff (see [48]), and an essentially identical problem is given in Example 5 of [36], a paper on the Absentminded Driver problem in decision theory. It is worth noting that Elga's motivation for introducing the problem was to examine how observers update probabilities in centered and uncentered worlds – see Section A.3.

On Sunday the experiment, as described below, is explained to Sleeping Beauty (SB). On Sunday night she goes to sleep in an isolated cell. A fair coin is tossed. Whatever the outcome she is awakened on Monday morning, and then on Monday evening is given a potion which puts her to sleep and causes her to forget everything that occurred on Monday. If the outcome of the toss was Tails she is then woken on Tuesday, while if the outcome was Heads then she is not woken on Tuesday. We ask two questions:

Question 1. When SB wakes in her cell during the experiment, what probability should she assign to the event that the coin landed Heads?

Question 2. If she is then told that it is Monday, what probability should she now assign to Heads?

Philosophers' use the term *credence* to denote a subjective probability. I will frequently use that term as shorthand for "the probability law that SB uses", but without wishing to be tied to any particular philosophical interpretation of probability. There are four positions in the philosophy literature. *Halfers* claim that the answer to Question 1 is $\frac{1}{2}$. The Halfer camp then divides according to the answer to Question 2. *Standard Halfers* claim that the answer to Question 2 is $\frac{2}{3}$, while *Double Halfers* argue that (somehow) the information that it is Monday makes no difference to SB's credence that the coin landed Heads, so that the answer to Question 2 is still $\frac{1}{2}$. *Thirders* argue that the answer to Question 1 is $\frac{1}{3}$, and then deduce that the answer to Question 2 is $\frac{1}{2}$. A final group argue that the problem as stated is ambiguous, inappropriate or undetermined, or that some other answer is correct. In his excellent review [46] Winkler suggests that the Thirder view is the majority position among philosophers, but that this is not reflected in the literature due to publication bias: people who believe the question is settled are less likely to write papers on the topic.

As well as the answers to the questions above, another area of disagreement is whether or not the problem is easy. Examples of papers which tend towards the 'easy' view include [20, 38], but the majority of papers appear to regard the problem as hard, involving some deep and difficult issues. That is certainly the view of this author: the problem probes what we mean by observers and observations, with possible issues of identity lying in the background.

A third area of uncertainty is whether the problem can be adequately described using Kolmogorov's axioms for probability. The main goal of this paper is to argue that it can. The framework we construct then allows the competing views of Halfers and Thirders to be evaluated in precise mathematical terms.

The original SB experiment involves fanciful and unavailable technologies: forgetfulness potions and, in some versions, duplication machines. For the purposes of this paper, and to clarify what is meant by an 'observer', we consider inference by artificial intelligences (AI). We assume that we have an AI program which is able to make Bayesian inferences as well as the most accomplished humans. (Current chatbots are not at that level, but it would be hard to maintain that no AI can ever achieve this.) I will refer to the AI as a *shabti* – this is an ancient Egyptian automaton. A shabti is a robot with the AI program described above written on ROM; it is able to receive limited sensory information from its environment. The ROM also contains the procedure for the current experiment. The shabti has no internal RAM, but has a number of external slots into which, like USB keys, RAM can be inserted. Without a RAM the shabti cannot function. When a RAM (the RAM of the day) is inserted, the shabti is able to compute, and record data and experiences. At the end of each day a hardware switch converts the RAM of the day into ROM. All shabti have the same program in their ROM. When I say a shabti is *woken* I will mean that a new RAM is inserted, and the shabti is switched on.

To implement the Sleeping Beauty experiment with a shabti, a shabti is placed in an isolated cell. The coin is tossed. Whatever the outcome, on Monday morning the shabti is woken. On Monday at midnight, Monday's RAM, now a ROM, is removed, and the shabti is switched off. If the coin was Heads the experiment ends there, while if it was Tails then on Tuesday the shabti is woken a second time. Whenever the shabti is woken we ask Questions 1 and 2 as above.

Replacing humans by a deterministic AI allows us to be certain that different anthropic observers with the same information will make the same decisions. For the sake of continuity with earlier formulations of the problem we have kept the concept of physical entities isolated in cells. Improvements in AI mean that it is entirely possible that experiments equivalent to the original SB experiment could be performed in the next few decades.

Elga outlined a straightforward solution of the SB problem using conditional probabilities. He claimed, using a "highly restricted" principle of indifference, that if SB is told the toss was Tails, then she should view it is equally likely to be Monday or Tuesday, so

$$\mathbb{P}(\text{Mon}|\text{Tails}) = \tfrac{1}{2}. \tag{1.1}$$

Next, as SB is always woken on Monday morning the coin toss can be deferred until Monday afternoon. If SB is told at midday on Monday that it is Monday, then the (fair) coin toss is still in the future, and so SB should assign probability $\frac{1}{2}$ to Heads. (We can even ask SB to toss the coin.) Thus

$$\mathbb{P}(\text{Heads}|\text{Mon}) = \tfrac{1}{2}. \tag{1.2}$$

Since $\mathbb{P}(\text{Heads}|\text{Tue}) = 0$ and $\mathbb{P}(\text{Mon}|\text{Heads}) = 1$, we have

$$\mathbb{P}(\text{Heads}) = \mathbb{P}(\text{Heads}|\text{Mon})\mathbb{P}(\text{Mon}) = \tfrac{1}{2}\mathbb{P}(\text{Mon}),$$

while writing $\mathbb{P}(\text{Mon}\ \&\text{Tails})$ in terms of conditional probabilities gives

$$\mathbb{P}(\text{Mon}|\text{Tails})\mathbb{P}(\text{Tails}) = \mathbb{P}(\text{Tails}|\text{Mon})\mathbb{P}(\text{Mon}). \tag{1.3}$$

Using (1.1) and (1.2) we obtain $\mathbb{P}(\text{Tails}) = \mathbb{P}(\text{Mon})$, and a little algebra gives that $\mathbb{P}(\text{Heads}) = \frac{1}{3}$.

This argument is simple enough that it seems surprising that the conclusion should be disputed. Nevertheless Standard Halfers reject (1.2), while Double Halfers wish to redefine what is meant by conditional probability.

The starting point for this paper is to ask what probability space the calculations above take place in. Recall that in the the standard SB model there is only one randomization, the toss of a fair coin, which suggests that it should be possible to define the model on the space $\Omega_O = \{H, T\}$ with $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$. However it is clear this cannot be the space used in the calculation above: no event in Ω_O has probability $\frac{1}{3}$. Further the events $\{\text{Mon}\}$ and $\{\text{Tue}\}$ do not lie in this space. These events also appear ambiguous – an outside observer of the experiment will experience both Monday and Tuesday.

The set of days is $\Omega_A = \{\text{Mon}, \text{Tue}\}$, and so the state of SB on waking is described by the pair

$$\omega = (\omega_o, \omega_a) \in \Omega = \Omega_O \times \Omega_A.$$

In the terminology of the philosophy literature elements of Ω_O describe 'possible worlds', while elements of Ω_A give the observer's location in the world. The general tendency there is to regard these two components as being at fundamentally different levels, a point most clearly indicated by the introduction in [22, 33, 7] of a new form of conditioning which acts in different ways on the two components of Ω.

However, when SB is woken, she is uncertain both about ω_o and ω_a, and it is natural to place her ignorance about these at the same level. Thus in this paper we will regard the space $\Omega = \Omega_O \times \Omega_A$ as the primary object. It is viewed by two observers. The classical observer CO essentially only sees the space Ω_O with its 'objective' probability $\mathbb{P}_O$. The AO sees the space Ω, and the primary problem for the AO is to define a suitable probability P_A on Ω. This is not an 'extension problem' since we will not require that $P_A(F \times \Omega_A) = \mathbb{P}_O(F)$ for $F \subset \Omega_O$. As we cannot expect to derive P_A from $\mathbb{P}_O$ using the axioms of probability our task divides into two parts.

The first is to formulate reasonable properties (or 'Principles') which restrict the possible class of measures P_A; these Principles have to come from our real world intuitions. In mathematical terms these Principles provide links between the space $(\Omega_O, \mathbb{P}_O)$ seen by the CO and the space (Ω, P_A) seen by the AO. The second and easier mathematical problem is to identify the class of measures allowed by these Principles, and work out their properties.

One Principle mentioned in the philosophy literature is Lewis' Principal Principle (see [32]) which roughly states that unless SB has additional evidence she should keep to the objective probability $\mathbb{P}$. We will see below strong reasons for rejecting a straightforward use of this Principal Principle.

The remainder of this paper is as follows. In Section 2 we introduce a generalised Sleeping Beauty problem (GSB). We set up the extended space Ω and define a random variable S which gives the location of the Anthropic Observer. We then formulate three Principles which, if accepted, determine P_A uniquely. The first, denoted (PN), is a mild condition on null sets, and will be assumed throughout this paper. The next, the *Principle of Indifference* (PI) is an extension of (1.1). The third, the *Principle of Equivalent Information* (PEI), generalises (1.2). This principle states roughly that a Classical Observer and an

Anthropic Observer who are in communication and have shared all the information they have available should assign the same probabilities to events in the space Ω_O.

Theorem 2.4 states that these determine P_A uniquely. We write P_E for this probability, and P_L for the main alternative, which is based on Lewis' Principal Principle. In Example 2.7 we look at the standard Sleeping Beauty problem and show that $P_E(\text{Heads}) = \frac{1}{3}$ and $P_L(\text{Heads}) = \frac{1}{2}$; we will therefore refer to P_E and P_L as the *Thirder* and *Halfer* measures respectively. In Section 2.3 we show (see Theorem 2.19) that (PEI) is essentially equivalent to the assertion that SB should assign the objective probability to random events which are still in her future.

As the GSB model does not cover situations where observers are placed randomly in some space, in Section 3 we extend the model and three Principles to set valued processes.

The 'Halfer' measure P_L has counterintuitive properties as soon as one looks at conditional events such as (1.1) and (1.2). Double Halfers need to find a new approach to conditioning, and in Section 4 we examine what seems to be the most mathematically developed proposal, *HTM conditioning*, as given in [22, 33, 7]. An example of [40] shows that it takes into account what ought to be irrelevant information, and fails to satisfy the law of total probability. Another very serious flaw is that it is sensitive to the particular formulation of the probability space used to describe the random experiment – see Example A.3.

Section 5 discusses betting arguments, and we find that they do give an argument for (PEI). Further, if the AO is able to make suitable additional observations, then betting arguments suggest that the AO should use P_E.

Section 6 considers repeated experiments. While it is straightforward to find the limit of the proportion of awakenings associated with (say) Heads, it is not clear how to connect this in a mathematically precise way with the probability law P_A.

Section 7 gives some examples. Sections 7.1 looks at how SB updates her probabilities. Section 7.2 shows that one needs to be careful when one conditions on SB's observations, and Section 7.3 reviews an example of Hartle and Srednicki [23] from the physics literature where insufficient care was taken. As mentioned at the start of this paper, a key issue is how we handle the concept of 'existence'. Is Existence or Non existence of an observer an event of the same kind as an observer learning the value of a random variable which takes the values E or N? We consider a variant of the SB problem Section 7.4, and show in this case at least that if we accept (PEI) then existence is not a special property. Section 7.5 makes a model for a simple example of Peter van Inwagen [44] designed to illuminate the fine tuning problem. We look at the related Doomsday argument in Section 7.6, and Bostrom's example of the Presumptuous Philosopher (see [8, p. 124]) in Section 7.7.

The very extensive literature contains many ideas, examples and calculations. In some cases a probability space such as our extension $\Omega = \Omega_O \times \Omega_A$ is implicit, but the framework given here does not seem to have been set out precisely before, and the formulation of the key principle (PEI) seems to be new. The preprint [21] does consider probability laws on the sample space $\Omega = \Omega_O \times \Omega_A$, but rejects, at least for some probability laws on Ω, the possibility of conditioning on events like $\{\text{Mon}\}$. Thus (PEI) would be regarded as inadmissible in the setup given there. The approach of this paper, which is natural from a probabilistic viewpoint, concentrates on information and conditional probabilities, and owes much to Elga's paper [16]. An alternative approach, as seen in for example Bostrom's book [8], is to start with the idea of an observer randomly chosen from some collection of observers. Bostrom's *Self Sampling Assumption* (SSA) tends to lead to Halfer type conclusions, while his *Self Indication Assumption* (SIA) tends to give Thirder answers.

One reason philosophers do not agree on the solution to the Sleeping Beauty problem is that all the approaches seem to lead to bad or paradoxical conclusions. For the Halfers the problems arise quickly – see Example 2.7 for details. Very roughly, Standard Halfers find themselves holding that the SB's position gives her information about the future, while Double Halfers have to redefine conditional probabilities. I suspect that few probabilists or statisticians would be willing to drop Kolmogorov's definition of conditional probability without seeing much stronger evidence for its inadequacy than has been presented so far.

The difficulties for Thirders do not arise so immediately. However, the example of the Presumptuous Philosopher suggests that the methods of inference used by Thirders, together with some reasonably plausible extensions, lead to the conclusion that, without needing to perform any measurements at all, we can assign a credence of 1 to the event that the universe is infinite. We note that while the problem is most severe for Thirders, Section 7.7 shows that a milder version of the same difficulty arises for Double Halfers.

2 The generalized Sleeping Beauty model

2.1 Probability spaces and observers

We generalize the Sleeping Beauty problem as follows. Let $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O)$ be a probability space which carries a bounded integer valued r.v. $X^O : \Omega_O \to \{0, \dots, M\}$ with

$$\mathbb{P}_O(X^O = k) = q_k, \quad k = 0, \dots, M.$$

We will call this the *objective probability space*. We can assume that $q_M > 0$; if not we redefine M to be $\max\{k : q_k > 0\}$. Most of our arguments generalize to the case when X^O is an unbounded random variable with finite expectation. But since the main challenges of the problem are already present in the case when X^O is bounded, we will restrict to that case.

We have available M identical cells $C_1, \dots, C_M$. The cells are labelled, but on the outside, so that an occupant does not know its cell label. Each cell is also equipped with a telephone which permits only incoming calls. If $X^O \geq 1$ then X^O shabti or 'anthropic observers' (AO) are deployed, each one being placed and set running in a distinct cell, starting with cell 1 and continuing until cell C_{X^O} is filled. A shabti knows the distribution (q_k) but not the value of X^O or what cell it is in. If $X^O = 0$ then no shabti are deployed. A Classical Observer (CO) is situated outside the collection of cells, and does not initially know the value of X^O. The CO has a telephone which can be used to call cell C_j for any $j \in \{1, \dots, M\}$. (The shabti have uneventful lives, and always answer the telephone.)

We can consider three variants of this GSB model.

Simultaneous Duplicated shabti. This is the version described above: the whole experiment takes place on one day, and the cells are occupied by distinct shabti. (This version is almost identical with the Incubator problem given in [8, p. 64].)

Serial Single shabti. The experiment takes place over a period of M days, with just one shabti, which is successively placed in cell C_k on day k, for $1 \leq k \leq X^O$, and by appropriate removal of RAMs has no memory of previous days. (This is the original Sleeping Beauty problem.)

Serial Duplicated shabti. This is the same as Serial Single shabti, except that distinct shabti are used on each occasion.

Yet another disputed point in the philosophy literature is whether or not these three problems have the same answer. See [29, Section 4] for the view that the single and duplicated problems differ. It may appear that the Duplicated problem requires some further randomization. This would be the case if there were a warehouse containing M shabti, X^O of which are then deployed in the experiment. Even though the shabti are identical, one still has to decide which shabti will be deployed in which cell. However, one can instead suppose that one has a shabti fabricator, which makes a new shabti in each of the cells C_j for $j = 1, \dots, X^O$.

The question as to whether these different versions of the experiment have different solutions probes hard questions on the meaning of observers and identity, which are beyond the scope of this paper. See Remark 2.20.

To handle the experience of the AO we extend the probability space, and define

$$\Omega_A = \{1, \dots, M\} \cup \{\partial\}, \quad \Omega = \Omega_O \times \Omega_A.$$

Let $S : \Omega \to \Omega_A$ be defined by $S((\omega, x)) = x$ for $\omega \in \Omega_O$, $x \in \Omega_A$. Set

$$\mathcal{F}_W = \{F \times \Omega_A, F \in \mathcal{F}_O\}, \quad \mathcal{F} = \sigma(\mathcal{F}_W, S) = \sigma(F \times \{x\}, F \in \mathcal{F}_O, x \in \Omega_A). \tag{2.1}$$

It is straightforward to check that $F \in \mathcal{F}$ if and only if there exist $F_x \in \mathcal{F}_W$ such that

$$F = \bigcup_{x \in \Omega_A} \{S = x\} \cap F_x. \tag{2.2}$$

We call $\mathcal{F}_W$ the *objective σ-field*, and events in $\mathcal{F}_W$ *objective* events. In terms of the terminology in the philosophy literature, events in $\mathcal{F}_W$ tell us about possible Worlds. We extend $\mathbb{P}_O$ to a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}_W)$ by setting $\mathbb{P}(F \times \Omega_A) = \mathbb{P}_O(F)$ and write X for the obvious extension of X^O to $(\Omega, \mathcal{F}_W)$, given by $X((\omega, j)) = X^O(\omega)$. Throughout this paper we will use a superscript 'O' to denote random variables on the base space Ω_O, and remove it to denote the extension of the random variable to Ω.

Definition 2.1. Given a probability space $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O)$, and a random variable $X^O : \Omega_O \to \{0, \dots, M\}$, we call the collection $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P}, X, S)$ the *anthropic extension* of $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O, X^O)$.

The space $(\Omega, \mathcal{F}_W, \mathbb{P})$ describes the experience of the classical observer. (It is simplest here to consider the Simultaneous Duplicated version of the problem.) If multiple cells are occupied the CO does not confer a distinguished status to any particular occupied cell, so that the random variable S is not observed by the CO.

We now look at the viewpoint of an AO, which wakes in an initially unknown cell. If $G \in \mathcal{F}_O$ and $j \in \{1, \dots, M\}$ then the interpretation of the event $G \times \{j\} \in \mathcal{F}$ is that the objective event G occurs, and the AO is situated in cell C_j. As $S((\omega, j)) = j$ we see that the random variable S gives the location of the AO. The point ∂ is needed to give the value of S when $X = 0$, and there are no anthropic observers.

Let P_A be a probability on $(\Omega, \mathcal{F})$. In the theory of Markov chains on a countable state space I one can derive the law $\mathbb{P}^x$ of the process started at any point $x \in I$ by conditioning from a single probability $\mathbb{P}^\mu$ where μ is a probability measure on I with $\mu(\{x\}) > 0$ for each $x \in I$. Similarly the law P_A can cover all possible locations of the AO, and the laws $P_A(\cdot|S = j)$ give the credences of the AO if it learns that it is in cell C_j. This definition does not impose any constraint on how the law P_A should be chosen. One straightforward choice is the probability P_L which, when Ω_O is countable, divides the $\mathbb{P}$-probability of a point $\omega \in \Omega_O$ equally among the atoms (ω, j), $1 \le j \le X^O(\omega)$. More precisely for $F \in \mathcal{F}_W$ let

$$P_L(F \cap \{X = k\} \cap \{S = j\}) = \begin{cases} k^{-1}\mathbb{P}(F \cap \{X = k\}|X \ge 1) \text{ for } 1 \le j \le k \le M, \\ 0 \text{ for } j > k. \end{cases} \tag{2.3}$$

This probability agrees with $\mathbb{P}$ conditioned on $\{X \ge 1\}$ on the σ-field of objective events, so is certainly in the spirit of the Principal Principle. We have $P_L(\text{Heads}) = \frac{1}{2}$ in the original SB problem, so we will refer to P_L as the 'Halfer measure'. (However, not all Halfers would agree that this is always the appropriate credence – see Footnote 5 on p. 2891 of [12].) While this measure may seem a natural choice, we will see below that it has some significant flaws.

We now give three 'Principles' which specify apparently desirable properties of P_A. The first, which places restrictions on null sets, is quite mild, while the last two give extensions of (1.1) and (1.2). The law P_L satisfies the first two but not in general the third.

Definition 2.2. We define the following three principles.

Principle of null sets. (PN).
(a) If $F \in \mathcal{F}_W$ has $\mathbb{P}(F) = 0$ then $P_A(F) = 0$.
(b) $P_A(S > X, S \in \mathbb{N}) = 0$.
(c) $P_A(S = \partial) = 0$.

Principle of Indifference. (PI). For $k \in \{1, \dots, M\}$ with $P_A(X = k) > 0$,

$$P_A(S = j|\, X = k) = \frac{1}{k} \quad \text{for } j = 1, \dots, k. \tag{2.4}$$

Principle of Equivalent Information. (PEI). Let $F \in \mathcal{F}_W$. For $k \in \{1, \dots, M\}$ with $P_A(S = k) > 0$,

$$P_A(F|S = k) = \mathbb{P}(F|X \ge k). \tag{2.5}$$

We will assume (PN) throughout, usually without further mention. Condition (b) states that an AO cannot find itself in an empty cell. This does not follow from (a), since the random variable S is not $\mathcal{F}_W$-measurable. It follows immediately from (b) and (c) that $P_A(X = 0) = 0$.

The intuition for (PI) is that if an AO is told the value of X, and has no other information, then it should decide it is equally likely to be in each of the possible cells $C_1, \ldots, C_X$. It is difficult to formulate any reasonable alternative distribution.

To see the motivation for (PEI), we fix $k \in \{1, \ldots, M\}$ before the start of the experiment; this value is known to the AO. Note that cell C_k is occupied if and only if $X \geq k$. After the cells have been populated the CO phones cell C_k. If the phone is answered than the CO knows that the event $\{X \geq k\}$ occurs. If the AO receives a phone call, it knows that it is in cell k. The AO and CO are in contact, and are allowed to share any information they have on the experiment. (In fact neither has anything to add to what the other already knows.) Since they then have the same information they should have identical views on the probability of any $\mathcal{F}_W$-measurable event F, and hence (2.5) should hold. We note that the point of the phone conversation is just to emphasize the fact that the CO and AO have the same information, and it is still reasonable to assume (PEI) in contexts when such a conversation does not occur. We also note that both [22, 35] reject assertions similar to (PEI).

The restriction in (PEI) that $F \in \mathcal{F}_W$ is necessary, since otherwise the right hand side is undefined. Note that while the events $\{S = j\}$ are disjoint, the events $\{X \geq j\}$ are not. This asymmetry may cause one to ask whether (PEI) is the correct mathematical formulation to describe the equivalent information of the CO and AO, but we will see below that there are some strong additional reasons for accepting (PEI).

Definition 2.3. An *anthropic probability* is a probability P_A on $(\Omega, \mathcal{F})$ which satisfies (PN).

Set

$$Q_n = \sum_{r=n}^{M} q_r = \mathbb{P}(X \geq n) \text{ for } 0 \leq n \leq M. \tag{2.6}$$

Since $Q_M > 0$ we have $Q_j > 0$ for each $j \in \{0, \ldots, M\}$.

We follow the same procedure as was used by Elga in [16], and use these Principles to describe P_A.

Theorem 2.4. *(a) There is a unique anthropic probability P_E on $(\Omega, \mathcal{F})$ which satisfies (PI) and (PEI). Writing $\lambda = 1/\mathbb{E}(X)$, if $F \in \mathcal{F}_W$, $1 \leq j \leq r \leq M$, then*

$$P_E(F \cap \{X = r\} \cap \{S = j\}) = \lambda \mathbb{P}(F \cap \{X = r\}), \tag{2.7}$$

$$P_E(F \cap \{S = j\}) = \lambda \mathbb{P}(F \cap \{X \geq j\}). \tag{2.8}$$

In particular for $1 \leq r \leq M$,

$$P_E(S = r) = \lambda Q_r, \quad P_E(X = r) = \lambda r q_r. \tag{2.9}$$

(b) If $F \in \mathcal{F}_W$ then

$$P_E(F) = \frac{\mathbb{E}(X1_F)}{\mathbb{E}(X)}, \tag{2.10}$$

and so if F is independent of X then $P_E(F) = \mathbb{P}(F)$.

Proof. (a) Let P_E satisfy the hypotheses, and write

$$s_j = P_E(S = j), \quad t_r = P_E(X = r). \tag{2.11}$$

(PN) implies that $t_0 = s_\partial = 0$, and so

$$\sum_{j=1}^{M} s_j = \sum_{r=1}^{M} t_r = 1. \tag{2.12}$$

Let $1 \le j \le r \le M$. We claim that

$$\frac{t_r}{r} = \frac{q_r s_j}{Q_j} \quad \text{whenever } 1 \le j \le r \le M. \tag{2.13}$$

Using (PI) and (PEI) we have

$$P_E(\{X = r\} \cap \{S = j\}) = P_E(S = j | X = r) t_r = \frac{t_r}{r} \quad \text{if } t_r > 0, \tag{2.14}$$

$$\begin{aligned} P_E(\{X = r\} \cap \{S = j\}) &= P_E(X = r | S = j) s_j \\ &= \mathbb{P}(X = r | X \ge j) s_j = \frac{q_r s_j}{Q_j}, \text{ if } s_j > 0. \end{aligned} \tag{2.15}$$

Thus (2.13) holds if either $t_r > 0$ and $s_j > 0$, or if $t_r = s_j = 0$. If $t_r > 0$ then (2.14) implies that $s_j > 0$, and if $s_j > 0$ then (2.15) implies that $t_r > 0$. Thus (2.13) holds in all cases.

Since $q_M > 0$, setting $c_0 = t_M/(Mq_M)$ it follows that $s_j/Q_j = c_0$ for each j. Hence $t_r = c_0 r q_r$ for each r, and the condition (2.12) implies that $c_0 = \lambda$.

Using (PEI) and $s_j = \lambda Q_j$,

$$\begin{aligned} P_E(F \cap \{X = r\} \cap \{S = j\}) &= P_E(F \cap \{X = r\} | S = j) s_j \\ &= \lambda \mathbb{P}(F \cap \{X = r\} | X \ge j) Q_j = \lambda \mathbb{P}(F \cap \{X = r\}), \end{aligned}$$

proving (2.7). It follows immediately that P_E is unique. Summing (2.7) over $r \ge j$ gives (2.8).

(b) Summing (2.7) first over j and then over r we have

$$P_E(F) = \sum_{r=1}^{M} \lambda r \mathbb{P}(F \cap \{X = r\}) = \lambda \sum_{r=1}^{M} \mathbb{E}(1_F X 1_{(X=r)}) = \lambda \mathbb{E}(X 1_F).$$

The final assertion is immediate from (2.10). □

Notation. We will write (s_j), (t_r) for the probabilities given by (2.11).

Remark 2.5. (2.9) implies that for an anthropic observer (AO) which accepts (PI) and (PEI) the P_E-law of X is the size-biased distribution associated with (q_r). This reweighting of the law of X is well known – see for example [8, p. 122]. For the standard SB problem the calculations are the same as in [16], and we have $P_E(\text{Heads}) = \frac{1}{3}$. We will therefore refer to P_E as the *Thirder measure*.

Remark 2.6. In some contexts it might be reasonable to replace the uniform distribution given by (PI) with a weighted distribution. Let $a_1, \dots, a_M$ be strictly positive, write $A_k = \sum_{j=1}^k a_j$, and suppose P_A satisfies

$$P_A(S = j | X = r) = \frac{a_j}{A_r} \quad \text{whenever } q_r > 0. \tag{2.16}$$

Then as in Theorem 2.4 we have

$$P_A(X = r) = \frac{q_r A_r}{\sum_{k=1}^M q_k A_k}, \quad P_A(S = j) = \frac{a_j Q_j}{\sum_{k=1}^M q_k A_k}. \tag{2.17}$$

In particular if we discount the future at rate e^{-b} and so set $a_j = e^{-bj}$ then (A_r) are bounded, and we no longer have size-biasing in the P_A-law of X.

Example 2.7. *The original Sleeping Beauty problem.* We take $M = 2$ and $q_1 = q_2 = \frac{1}{2}$. We set $\Omega_O = \{1, 2\}$, with $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \frac{1}{2}$, and $\Omega = \{(i, j) : 1 \le i, j \le 2\}$. We set $X((i, j)) = i$, $S((i, j)) = j$. Let P_A be an anthropic probability; (PN) gives that $P_A(X = 1, S = 2) = 0$. Let

$$a = P_A(\{(1, 1)\}), b = P_A(\{(2, 1)\}), 1 - a - b = P_A(\{(2, 2)\}), \tag{2.18}$$

and write Heads $= \{X = 1\}$, Tails $= \{X = 2\}$, Mon $= \{S = 1\}$, Tue $= \{S = 2\}$. Then

$$P_A(\text{Heads}|\text{Mon}) = \frac{a}{a + b}, \text{ and } P_A(\text{Mon}|\text{Tails}) = \frac{b}{1 - a}. \tag{2.19}$$

If we assume (PEI) then $P_A(\text{Heads}|\text{Mon}) = \frac{1}{2}$, which implies $a = b$. If we assume (PI) then $P_A(\text{Mon}|\text{Tails}) = \frac{1}{2}$, so that $a + 2b = 1$. Thus (PI) and (PEI) together imply $a = b = \frac{1}{3}$. Each of (PI) and (PEI) imposes a linear relation between a and b, and the two principles together fix a and b. It is clear from this example that the two conditions (PI) and (PEI) are independent, that is neither implies the other.

The probability measure for the standard Halfer position is the probability P_L defined by (2.3), which satisfies $a = \frac{1}{2}$, $b = \frac{1}{4}$. We have $P_L(\text{Heads}|\text{Mon}) = \frac{2}{3}$, so this measure does not satisfy (PEI). Even though the coin toss can take place on Monday afternoon, on the standard Halfer view SB at midday on Monday should hold that this coin toss has probability $\frac{2}{3}$ of coming down Heads, while the CO still views it as $\frac{1}{2}$. We can imagine that the CO meets SB in her cell for this coin toss; they appear to have the same information but they have different views on the probability of Heads. One can then ask a standard Halfer which of the two has the better insight? If it is the CO, why does SB choose an inferior credence? And if it is SB, how is that her forgetfulness potion, to be taken on Monday evening, gives her superior insight into a coin toss taking place on Monday afternoon?

The Double Halfer position is that $P_A(\text{Heads}) = \frac{1}{2}$, and the probability of Heads given that it is Monday is also $\frac{1}{2}$. If we use (2.18) and (2.19) then we obtain $a = \frac{1}{2}$ and $a/(a+b) = \frac{1}{2}$, so that $a = b = \frac{1}{2}$ and $P_A(\text{Tue}) = 0$. Thus Double Halfers would appear to be committed to the view that on waking SB is certain that it is Monday: [24] does adopt that position. But most Double Halfers wish to allow $P_A(S = 2) > 0$, and some wish to keep (PI) and hold that $P_A(S = 1|\text{Tails}) = \frac{1}{2}$. (Thus they are really 'Triple Halfers'). It is clear from the calculations above that this position cannot be described by the usual axioms of probability. Double Halfers are well aware of this difficulty, and the literature contains a number of strategies for dealing with it. One of these, a revised definition of conditional probability, is discussed in Section 4 below.

2.2 Auxiliary processes and Technicolour Beauty

It is important to allow the AO to be able to make additional observations. Let $Z = (Z_j, 1 \le j \le M)$ be an $\mathcal{F}_W$ measurable process taking values in a finite set $\mathcal{Z}$: we will call this an *auxiliary process*. We call Z_i the *colour* of cell C_i and assume that it is visible somewhere in the cell C_i, so that (maybe after a short pause) the AO is able to see this value. This model is often called *Technicolour Beauty* – see [40, 11, 38]. For $y \in \mathcal{Z}$ let

$$L_y = |\{i \le X : Z_i = y\}| = \sum_{j=1}^{M} 1_{(j \le X)} 1_{(Z_j = y)} \tag{2.20}$$

be the number of occupied cells with index y. Set

$$G_y = \{L_y \ge 1\} = \bigcup_{i=1}^{M} \{Z_i = y, i \le X\}; \tag{2.21}$$

this is the event that at least one of the AO sees the index y. The event that the AO sees the value y in its cell on the particular awakening modelled by S is

$$H_y = \{Z_S = y\} = \bigcup_{i=1}^{M} \{Z_i = y, S = i\}. \tag{2.22}$$

Clearly $H_y \subset G_y$, $G_y \in \mathcal{F}_W$, and in general $H_y \notin \mathcal{F}_W$.

Lemma 2.8. *For $F \in \mathcal{F}_W$ and $y \in \mathcal{Z}$, $P_E(H_y) = \lambda \mathbb{E}(L_y)$ and*

$$P_E(F|H_y) = \frac{\mathbb{E}(1_F L_y)}{\mathbb{E}(L_y)}. \tag{2.23}$$

Proof. We have using (2.8)

$$P_E(F \cap H_y) = \sum_{j=1}^{M} P_E(F \cap \{Z_j = y\} \cap \{S = j\})$$
$$= \lambda \sum_{j=1}^{M} \mathbb{P}(F \cap \{Z_j = y\} \cap \{X \geq j\}) = \lambda \mathbb{E}(L_y 1_F).$$

Setting $F = \Omega$ gives $P_E(H_y) = \lambda \mathbb{E}(L_y)$, and (2.23) follows. □

Definition 2.9. We say that an auxiliary process Z is *injective* if $\mathbb{P}(L_y \leq 1) = 1$ for each $y \in \mathcal{Z}$. Setting $Z_j = j$ for $j = 1, \ldots, M$ gives an injective auxiliary process, so such a process always exists.

Lemma 2.10. *Suppose that Z is injective. Then*

$$P_E(F|H_y) = \mathbb{P}(F|G_y) \quad \textit{for all } F \in \mathcal{F}_W. \tag{2.24}$$

Proof. As Z is injective $L_y = 1_{G_y}$ and (2.24) follows from (2.23). □

Definition 2.11. Suppose that Z is injective. We say that an anthropic probability P_A satisfies (PZ) if

$$P_A(F|H_y) = \mathbb{P}(F|G_y) \quad \text{for all } F \in \mathcal{F}_W. \tag{2.25}$$

(PZ) can be justified in a similar fashion to (PEI). The CO has a device which phones the cell j with $Z_j = y$ if it exists and is occupied. If the call takes place then the CO knows that G_y occurs, and the AO knows that H_y occurs. The equation (2.25) states that they then have the same credence for events in $\mathcal{F}_W$. We will also see below a justification for (PZ) using betting arguments.

Remark 2.12. Note that if $Z_j = j$ for each j then (PZ) is the same as (PEI).

Example 2.13. *Technicolour Beauty.* (See [40].) We attach to the standard SB model an independent auxiliary process $Z = (Z_1, Z_2)$ which is a random permutation of $\{0, 1\}$. We set

$$\Omega_O = \{1, 2\} \times \{0, 1\},$$

and as usual we take $\Omega_A = \{1, 2\}$ and $\Omega = \Omega_O \times \Omega_A$. If $(i, j, k) \in \Omega$ then we define $X((i, j, k)) = i$, $Z_1((i, j, k)) = j$, $Z_2 = 1 - Z_1$, and $S((i, j, k)) = k$. We can write Ω as the disjoint union $\Omega = \{X = 1, S = 2\} \cup H_0 \cup H_1$, where the first set contains two points, and $|H_k| = 3$ for $k = 0, 1$.

Now let P_A be an anthropic probability which satisfies (PZ). Since (PN) holds we have $P_A(X = 1, S = 2) = 0$. Set $p_k = P_A(H_k)$ so that $p_0 + p_1 = 1$. Easy calculations give that $\mathbb{P}(G_k) = \frac{3}{4}$ for $k = 0, 1$. Then $\{X = i, Z_1 = k, S = 1\} = \{X = i, Z_1 = k\} \cap H_k \subset G_k$ and so using (PZ)

$$P_A(X = i, Z_1 = k, S = 1) = p_k \mathbb{P}(X = i, Z_1 = k|G_k) = \frac{p_k \mathbb{P}(X = i, Z_1 = k)}{\mathbb{P}(G_k)} = p_k/3.$$

A similar calculation gives $P_A(X = 2, Z_1 = k, S = 2) = p_{1-k}$. Thus for $k = 0, 1$ each of the three points in H_k have probability p_k. Furthermore, any probability of this form satisfies (PZ), so that (PZ) is not enough to determine P_A.

Let Z be an injective auxiliary process. By reducing the set $\mathcal{Z}$ if necessary we can assume that $\mathbb{P}(G_y) > 0$ for all y. Since $\mathbb{P}(Z_j = y, X \geq j) = \mathbb{P}(Z_j = y, L_y = 1)$ it follows that

$$\mathbb{P}(Z_j = y | X \geq j) Q_j = \mathbb{P}(Z_j = y | G_y) \mathbb{P}(G_y). \tag{2.26}$$

Write $a(j, y) = \mathbb{P}(Z_j = y, X \geq j)$. Set $I_M = \{1, \ldots, M\}$ and define the graph (I_M, E_M) by $\{i, j\} \in E_M$ if there exists $y \in \mathcal{Z}$ with $a(i, y)a(j, y) > 0$.

Definition 2.14. We say the auxiliary process Z is *spanning* if the graph (I_M, E_M) is connected.

Theorem 2.15. *Let Z be an injective spanning auxiliary process. If P_A satisfies (PN), (PEI) and (PZ) then $P_A = P_E$.*

Proof. Set $\tilde{s}_j = P_A(S = j)$, $\tilde{h}_y = P_A(Z_S = y)$. Since $L_y \leq 1$ we have

$$P_A(Z_j = y, S = j) = P_A(Z_j = y, Z_S = y) \text{ for all } j \text{ and } y. \tag{2.27}$$

Suppose that $a(j, y) > 0$. If $\tilde{s}_j > 0$ then using (PEI) the left side of (2.27) can be written as

$$P_A(Z_j = y | S = j)\tilde{s}_j = \mathbb{P}(Z_j = y | X \geq j)\tilde{s}_j, \tag{2.28}$$

while if $\tilde{h}_y > 0$ then using (PZ) the right hand of (2.27) equals

$$P_A(Z_j = y | H_y)\tilde{h}_y = \mathbb{P}(Z_j = y | G_y)\tilde{h}_y. \tag{2.29}$$

If follows that either $\tilde{s}_j = \tilde{h}_y = 0$ or that both $\tilde{s}_j$ and $\tilde{h}_y$ are strictly positive. In both cases we obtain

$$\mathbb{P}(Z_j = y | X \geq j)\tilde{s}_j = \mathbb{P}(Z_j = y | G_y)\tilde{h}_y,$$

and comparing with (2.26) we deduce that

$$Q_j \tilde{h}_y = \tilde{s}_j \mathbb{P}(G_y) \quad \text{whenever } a(j, y) > 0.$$

Since both Q_j and $\mathbb{P}(G_y)$ are strictly positive it follows that

$$\frac{\tilde{s}_j}{Q_j} = \frac{\tilde{s}_i}{Q_i} \text{ whenever } \{i, j\} \in E_M ,$$

and so as (I_M, E_M) is connected we have $\tilde{s}_j = \theta Q_j$ for all $j \in \{1, \ldots, M\}$ and some $\theta > 0$. Since $\sum_j \tilde{s}_j = 1$ we have $\theta = \lambda$, and so $P_A(S = j) = P_E(S = j)$ for all j. Since P_A and P_E both satisfy (PEI) it follows that $P_A(F \cap \{S = j\}) = P_E(F \cap \{S = j\})$ for all $F \in \mathcal{F}_W$ and $1 \leq j \leq M$, and thus $P_A = P_E$. $\square$

2.3 Future information

In the introduction we noted that for the standard SB problem, one argument for (1.2) is that as the coin can be tossed on Monday evening, on Monday morning it is a future event, and so the AO and CO should assign it the same probability. We now show that, subject to a mild restriction on the probability space, (PEI) is equivalent to a condition which states that the AO and CO have the same views on the probability of future events. In essence this argument proceeds by pushing every event not needed to determine whether or not $X = n$ for some time n into the post-n future.

We consider a serial variant of the model. To avoid measure-theoretic technicalities, we make the assumption that the objective probability space $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O)$ has the following structure.

Assumption 2.16. (1) There exist independent Bernoulli random variables $(V_n^O, 0 \leq n \leq M)$ and U^O, such that U^O is uniform on $[0, 1]$ and

$$\mathbb{P}_O(V_n^O = 1) = 1 - \mathbb{P}_O(V_n^O = 0) = q_n/Q_n, \quad 0 \leq n \leq M. \tag{2.30}$$

(2) $\mathcal{F}_O = \sigma(U^O, V_n^O, 0 \leq n \leq M)$.
(3) The r.v X^O satisfies

$$X^O = \min\{n : V_n^O = 1\}. \tag{2.31}$$

Remark 2.17. This assumption is much less restrictive than it might appear. A *standard probability space* is a space which is isomorphic to the unit interval with Lebesgue measure. In [26, p. 61] Itô remarks that "all probability spaces appearing in practical applications are standard", which can be restated as saying that any stochastic process appearing in practical applications can be defined as a measurable function of a single r.v. U with uniform distribution on $[0, 1]$. In particular, if $\widetilde{X}$ is a random variable taking values in $\{0, \dots, M\}$ and $\widetilde{Z} = (\widetilde{Z}_1, \dots, \widetilde{Z}_M)$ is a stochastic process taking values in a complete separable metric space $\mathcal{K}$, then there exists a probability space $(\Omega_O, \mathcal{F}_W, \mathbb{P}_O)$ satisfying Assumption 2.16 carrying X^O, Z^O such that $(X^O, Z_1^O \dots, Z_M^O)$ has the same distribution as $(\widetilde{X}, \widetilde{Z}_1, \dots, \widetilde{Z}_M)$. See [26] for more details.

We write (V_n), U for the extensions of $(V_n^O, n \in \mathbb{Z}_+)$ and U^O to the space $(\Omega, \mathcal{F}, \mathbb{P})$. Assumption 2.16 implies that $\mathcal{F}_W = \sigma(U, V_n, 0 \leq n \leq M)$. From (2.31) we have

$$\{X \geq n\} = \{V_k = 0, 0 \leq k \leq n - 1\}.$$

We regard the randomization V_k as being made at the end of day k, and the randomization U being made at the end of day M. Note that $\mathbb{P}(V_M = 1) = 1$.

Definition 2.18. Principle of no future information. (PNFI).
Let $0 \leq n \leq M$. Then if $G \in \sigma(V_n, \dots, V_M, U)$

$$P_A(G|S = n) = \mathbb{P}(G). \tag{2.32}$$

We assume that the CO is able to observe the random variables V_k as they occur. The value n is fixed at the start of the experiment. On day n the CO knows the values of $V_0, \dots, V_{n-1}$, and thus whether or not $X \geq n$. If $X < n$ the CO takes no action. If $V_0 = \dots = V_{n-1} = 0$, so that $X \geq n$, then as G is independent of $V_0, \dots, V_{n-1}$ we have $\mathbb{P}(G|X \geq n) = \mathbb{P}(G)$, and the right hand side of (2.32) gives the CO's conditional probability that G will occur. If $X \geq n$ then the CO telephones cell n and talks to the AO there. The AO thus knows that the event $\{S = n\}$ holds, and hence also knows that $\{X \geq n\}$. The event G lies in the future for the AO also, and the left side of (2.32) is the AO's conditional probability that it will occur. (PNFI) then states that the two observers agree on the probability of this future event.

Theorem 2.19. *Suppose that Assumption 2.16 holds. Then (PEI) and (PNFI) are equivalent.*

Proof. Suppose that (PEI) holds, and let $G \in \sigma(V_n, \dots, V_M, U)$. Then $G \in \mathcal{F}_W$ and is independent of $\{X \geq n\}$. Hence (2.32) follows immediately from (PEI).

Now suppose that (PNFI) holds. Let $(e_0, \dots, e_{n-1}) \in \{0,1\}^n$, $H = \{V_k = e_k, 0 \leq k \leq n-1\}$, and write $H_1 = \{V_k = 0, 0 \leq k \leq n-1\}$. Let $G \in \sigma(V_n, \dots, V_M, U)$. If the AO knows that $\{S = n\}$ then it knows that H_1 has occurred, and so $P_A(G \cap H|S = n) = 0$ if $H \neq H_1$, while

$$P_A(G \cap H_1|S = n) = P_A(G|S = n). \tag{2.33}$$

Similarly we have $\mathbb{P}(G \cap H|X \geq n) = 0$ if $H \neq H_1$, and $\mathbb{P}(G \cap H_1|X \geq n) = \mathbb{P}(G)$. We therefore deduce that

$$P_A(G \cap H|S = n) = \mathbb{P}(G \cap H|X \geq n). \tag{2.34}$$

Using the π-λ Lemma (see [15, p. 447]) and property (2) of Assumption 2.16 it follows that (PEI) holds. □

Remark 2.20. At the beginning of this Section we described three distinct variants of the GSB model. All can be described by the anthropic extension $(\Omega, \mathcal{F}, P_A)$, but the strengths of the arguments for (PEI) and (PI) differ according to the variant. The 'future information' argument for (PEI) given above is most persuasive in the context of a serial model, while the justification for (PI) is clearest for the simultaneous model.

3 A set valued GSB model

In the model given above, the cells were filled starting with cell C_1 and continuing to cell C_X. In some examples, such as the one considered in Section 7.3, one needs to allow more general possibilities for the filled and unfilled cells, and in this section we extend our model to cover this situation.

Let K be a finite set; each $x \in K$ is associated with a cell C_x which is the possible location of an AO. Let $\mathcal{P}(K)$ be the set of all subsets of K, and let $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O)$ be a probability space with a random variable $\mathcal{X}^O : \Omega_O \to \mathcal{P}(K)$. (Thus $\mathcal{X}^O(\omega) \subset K$ for each

ω). We assume that $0 \le \mathbb{P}_O(\mathcal{X} = \emptyset) < 1$. We now proceed very much as before. Let ∂ be a point not in K, $\Omega_A = K \cup \{\partial\}$, and $\Omega = \Omega_O \times \Omega_A$. Set $\mathcal{F}_W = \{F \times \Omega_A, F \in \mathcal{F}_O\}$, and extend $\mathbb{P}_O$ to a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F}_W)$ by setting $\mathbb{P}(F \times \Omega_A) = \mathbb{P}_O(F)$. We write $\mathcal{X}((\omega, x)) = \mathcal{X}^O(\omega)$, and define $S : \Omega \to \Omega_A$ by $S((\omega, x)) = x$ for $\omega \in \Omega_O$, $x \in \Omega_A$. Finally we set $\mathcal{F} = \sigma(\mathcal{F}_W, S)$. As in Section 2 we call $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P}, \mathcal{X}, S)$ the *anthropic extension* of $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O)$.

Now set

$$q_B = \mathbb{P}(\mathcal{X} = B), \text{ for } B \subset K, \tag{3.1}$$

$$Q_x = \mathbb{P}(x \in \mathcal{X}) = \sum_{B : x \in B} q_B. \tag{3.2}$$

We now rewrite our principles for this more general setup. For clarity we label these principles as (SP...).

Principle of null sets. (SPN).
(a) If $F \in \mathcal{F}_W$ and $\mathbb{P}(F) = 0$ then $P_A(F) = 0$.
(b) $P_A(S \in K \setminus \mathcal{X}) = 0$.
(c) $P_A(S = \partial) = 0$.

(SPN) gives that

$$P_A(\mathcal{X} = \emptyset) = P_A(\mathcal{X} = \emptyset, S = \partial) + P_A(\mathcal{X} = \emptyset, S \in K) = 0.$$

Principle of Indifference. (SPI). For $x \in K$, $B \subset K$, with $P_A(\mathcal{X} = B) > 0$,

$$P_A(S = x|\ \mathcal{X} = B) = |B|^{-1} 1_B(x). \tag{3.3}$$

Principle of Equivalent Information. (SPEI). If $F \in \mathcal{F}_W$ and $P_A(S = x) > 0$ then

$$P_A(F|S = x) = \mathbb{P}(F|x \in \mathcal{X}). \tag{3.4}$$

As before we will always assume that (SPN) holds. We continue to use the term *anthropic probability* to mean a probability on $(\Omega, \mathcal{F})$ which satisfies (SPN).

We define a graph structure on K by defining $\{x, y\}$ to be an edge if there exists a set $A \subset K$ with $q_A > 0$ such that $x, y \in A$. Let E be the set of edges.

Theorem 3.1. *Suppose that (K, E) is connected. There is a unique anthropic probability P_E on $(\Omega, \mathcal{F})$ which satisfies (SPN), (SPI) and (SPEI). Writing $\lambda = 1/\mathbb{E}(|\mathcal{X}|)$ we have*

$$P_E(F \cap \{\mathcal{X} = B\} \cap \{S = x\}) = \lambda \mathbb{P}(F \cap \{\mathcal{X} = B\}) \text{ for } F \in \mathcal{F}_W, B \subset K, x \in B. \tag{3.5}$$

In particular

$$P_E(S = x) = \lambda Q_x, \quad P_E(\mathcal{X} = B) = \lambda |B| q_B, \quad \text{for } x \in K,\ B \subset K, \tag{3.6}$$

and if $F \in \mathcal{F}_W$ then

$$P_E(F) = \frac{\mathbb{E}(1_F |\mathcal{X}|)}{\mathbb{E}(|\mathcal{X}|)}. \tag{3.7}$$

Proof. Let P_E be a probability which satisfies the hypotheses, and write

$$t_B = P_A(\mathcal{X} = B), \text{ for } B \subset K, \quad s_x = P_A(S = x). \tag{3.8}$$

The conditions on K and $\mathcal{X}$ imply that $Q_x > 0$ for each $x \in K$.

Let $B \subset K$ and $x \in B$. If $t_B > 0$ then by (SPI)

$$P_E(S = x, \mathcal{X} = B) = P_E(S = x|\mathcal{X} = B)t_B = \frac{t_B}{|B|}. \tag{3.9}$$

If $s_x > 0$ then by (SPEI)

$$P_E(S = x, \mathcal{X} = B) = P_E(\mathcal{X} = B|S = x)s_x = \mathbb{P}(\mathcal{X} = B|x \in \mathcal{X})s_x = \frac{s_x q_B}{Q_x}. \tag{3.10}$$

Thus we have

$$\frac{t_B}{q_B|B|} = \frac{s_x}{Q_x} \tag{3.11}$$

for any pair (x, B) with $q_B > 0$ and $x \in B$. If $\{x, y\}$ is an edge in the graph (K, E) then it follows that $s_x/Q_x = s_y/Q_y$. As the graph (K, E) is connected, the function s_x/Q_x is equal to a constant λ on K. We then have $t_B = \lambda|B|q_B$ and since $\sum_{B\subset K} t_B = 1$ it follows that $\lambda^{-1} = \mathbb{E}(|\mathcal{X}|)$.

A further application of (SPEI) proves (3.5), which gives the uniqueness of P_E. $\square$

The case when (K, E) is not connected does not arise much in applications, but for the sake of completeness the details are given in the Appendix.

Example 3.2. *Four Beauties.* This elegant example is due to Pittard [35]. The experiment lasts for one night only, with four participants ('Beauties') A, B, C, D. One of the four will be chosen at random to be the 'victim'. The four participants all go to sleep. Three times in the night a pair of participants are woken at the same time for a period (say 15 minutes) during which they can converse. Then they go back to sleep, and as usual are given a memory erasing drug so that they forget the awakening. The victim is woken all three times, and paired with each other participant exactly once.

To analyse this model we need to describe the awakenings of a particular participant, A say. We take $\Omega_O = \{A, B, C, D\}$, and let $K = \Omega_A = \{B, C, D\}$ be the set of possible partners of A during an awakening. We write $\omega = (v, x)$ for points in Ω, and with ω of this form we define $V(\omega) = v$, $S(\omega) = x \in K$. So V gives the victim, and S describes A's partner for that awakening. $\mathbb{P}$ is the probability which makes V uniform on Ω_O, and $\mathcal{X}$ is defined by $\mathcal{X} = K$ if $V = A$, and $\mathcal{X} = \{V\}$ if $V \neq A$.

As $\mathbb{P}(\mathcal{X} = K) > 0$ the model satisfies the hypotheses of Theorem 3.1. We have $Q_x = \frac{1}{2}$ for $x = B, C, D$, and $\mathbb{E}|\mathcal{X}| = \frac{3}{2}$, so that $\lambda = \frac{2}{3}$. Thus $P_E(S = x) = \lambda Q_x = \frac{1}{3}$ for $x = B, C, D$. Suppose that A wakes and sees that the other person woken is B. Then, using (SPEI) the probability that A is the victim is

$$P_E(V = A|S = B) = \mathbb{P}(V = A|B \in \mathcal{X}) = \frac{\mathbb{P}(V = A)}{\mathbb{P}(B \in \mathcal{X})} = \tfrac{1}{2}.$$

This answer accords with common sense – A and B are in the same situation and by symmetry each regards herself as equally likely to be the victim.

On the other hand, if we use the Halfer measure P_L then $P_L(V = A) = \mathbb{P}(V = A) = \frac{1}{4}$, while by symmetry $P_L(S = B) = P_L(S = B|V = A) = \frac{1}{3}$. Hence

$$P_L(V = A|S = B) = \frac{P_L(S = B|V = A)P_L(V = A)}{P_L(S = B)} = \tfrac{1}{4}.$$

Thus if A and B are woken together then as A has credence $\frac{1}{4}$ that she is the victim, she must have credence $\frac{3}{4}$ that B is the victim. B, of course, has the opposite credences. A and B appear to have the same information, and can share all they that know, but still have different credences for an objective event. Although this example would seem to pose a severe challenge for Halfers, [35] does nevertheless adhere to the Halfer position.

Some experiments may have more than one natural choice for the collection of states (or cells) for the AO. For example [11] studies the standard Sleeping Beauty problem, but include her states on Sunday, before the experiment begins, and on Wednesday when it is finished.

It is thus natural to look at *model restriction*. We start with a r.v. $\mathcal{X}^O : \Omega_O \to \mathcal{P}(K)$; we then define $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P}, \mathcal{X}, S)$ as above. Note that $\Omega_A = K \cup \{\partial\}$, and $\Omega = \Omega_O \times \Omega_A$. We assume that the graph (K, E) is connected, and write P_E for the unique anthropic probability which satisfies (SPI) and (SPEI). Let $K' \subset K$, and suppose that the AO knows that it is K'. We can consider restriction to K' in two ways.

The first is by conditioning on the observer being in K', and so to look at

$$\widetilde{P}_E(F) = P_E(F|S \in K'),\ F \in \mathcal{F}. \tag{3.12}$$

The second is by looking at a reduced model. Define $\mathcal{X}'^O = \mathcal{X}^O \cap K'$, and let $(\Omega', \mathcal{F}', \mathcal{F}'_W, \mathbb{P}', \mathcal{X}', S')$ be the anthropic extension of $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O, \mathcal{X}'^O)$. Note that $\Omega' = \Omega_O \times \Omega'_A$, where $\Omega'_A = K' \cup \{\partial\}$. If $F \in \mathcal{F}_W$ then there exists $G \in \mathcal{F}_O$ such that $F = G \times \Omega_A$. We define $F' = G \times \Omega'_A$, and note that

$$\mathbb{P}'(F') = \mathbb{P}_O(G) = \mathbb{P}(F).$$

Let (K', E') be the graph obtained by taking $\{x, y\} \in E'$ if $x, y \in K'$ and there exists $A \subset K'$ with $x, y \in A$ and $\mathbb{P}'(\mathcal{X}' = A) > 0$. The graph (K', E') need not be connected, but if it is then we write P'_E for the unique anthropic probability associated with this system which satisfies (SPI) and (SPEI). $\widetilde{P}_E$ and P'_E cannot be the same, since they are defined on different spaces. But they do give rise to the same probabilities for objective events and for the location of the AO.

Proposition 3.3. *Assume that (K', E') is connected, and let $x \in K'$, $G \in \mathcal{F}_O$. Then writing $F = G \times \Omega_A$, $F' = G \times \Omega'_A$,*

$$\widetilde{P}_E(F \cap \{S = x\}) = P'_E(F' \cap \{S' = x\}). \tag{3.13}$$

Proof. If $y \in K'$ then $Q'_y = \mathbb{P}'(x \in \mathcal{X}') = Q_y$. Let $\lambda' = 1/\mathbb{E}'(|\mathcal{X}'|)$, and note that

$$\frac{1}{\lambda'} = \sum_{y \in K'} Q_y.$$

Then using (SPEI) we obtain

$$\begin{aligned} \widetilde{P}_E(F \cap \{S = x\}) &= P_E(F \cap \{S = x\} | S \in K') = \frac{P_E(F \cap \{S = x\})}{P_E(S \in K')} \\ &= \frac{P_E(F|S = x)s_x}{\sum_{y \in K'} s_y} = \frac{\mathbb{P}(F|x \in \mathcal{X})Q_x}{\sum_{y \in K'} Q_y} \\ &= \lambda' \mathbb{P}(F \cap \{x \in \mathcal{X}\}) = \lambda' \mathbb{P}_O(G \cap \{x \in \mathcal{X}^O\}). \end{aligned}$$

Noting that $P'_E(x \in \mathcal{X}) = \lambda' Q_x$, we have using (SPEI)

$$P'_E(F' \cap \{S' = x\}) = \mathbb{P}'(F'|x \in \mathcal{X}')\lambda' Q_x = \lambda' \mathbb{P}_O(G \cap \{x \in \mathcal{X}'^O\}).$$

If $x \in K'$ then $\{x \in \mathcal{X}'^O\} = \{x \in \mathcal{X}^O\}$, so we obtain (3.13). □

Remark 3.4. Not all the models in the philosophy literature satisfy this natural property. See for example [9] and Section 6.

4 Double Halfer conditioning

We now review one explicit proposal from Double Halfers for a new definition of conditioning – see [22, 33, 7]. This is called the *Halfer rule* in [7], and *HTM conditioning* in [33]. For simplicity we will just look at the GSB model in Section 2 with $q_0 = 0$, and will assume further that Ω_O, and hence Ω, is countable.

Let P_A be an anthropic probability on Ω. We define the projection of P_A onto Ω_O by

$$\overline{P}_A(G_o) = P_A(G_o \times \Omega_A) \text{ for } G_o \in \mathcal{F}_O.$$

To avoid unnecessary clutter we will write $\overline{P}_A(\omega)$ for $\overline{P}_A(\{\omega\})$, etc. For $F \in \mathcal{F}$, $\omega \in \Omega_O$, set

$$\begin{aligned} F_\omega &= F \cap (\{\omega\} \times \Omega_A), \\ \mathcal{S}(F, P_A) &= \{\omega : P_A(F_\omega) > 0\}. \end{aligned}$$

Note that $\Omega_\omega = \{\omega\} \times \Omega_A$ and

$$P_A(F_\omega) \leq P_A(\Omega_\omega) = \overline{P}_A(\omega).$$

Hence

$$\overline{P}_A(\mathcal{S}(F, P_A)) = \sum_{\omega \in \mathcal{S}(F,P_A)} \overline{P}_A(\omega) \geq \sum_{\omega \in \mathcal{S}(F,P_A)} P_A(F_\omega) = P_A(F). \tag{4.1}$$

Definition. Let $F \in \mathcal{F}$ satisfy $P_A(F) > 0$. Let $(\omega, k) \in \Omega$. We define the *Double Halfer (DH) conditioning* $P_A((\omega, k)||F)$ as follows.
Case 1. If $P_A(F_\omega) = 0$ we set $P_A((\omega, k)||F) = 0$.
Case 2. If $P_A(F_\omega) > 0$ we set

$$P_A((\omega, k)||F) = \overline{P}_A(\omega|\mathcal{S}(F, P_A))\, P_A((\omega, k)|F_\omega). \tag{4.2}$$

Note that in Case 2 we have $\overline{P}_A(\mathcal{S}(F, P_A)) \geq \overline{P}_A(\omega) \geq P_A(F_\omega) > 0$, so that both conditional expectations in (4.2) are defined. We then write for $G \in \mathcal{F}$,

$$P_A(G||F) = \sum_{(\omega,k)\in G} P_A((\omega, k)||F). \tag{4.3}$$

If $P_A(F_\omega) > 0$ we can write

$$\begin{aligned} P_A((\omega, k)||F) &= \frac{P_A(\Omega_\omega)}{\overline{P}_A(\mathcal{S}(F, P_A))} \frac{P_A((\omega, k) \cap F)}{P_A(F_\omega)} \\ &= P_A((\omega, k)|F) \frac{P_A(F)}{\overline{P}_A(\mathcal{S}(F, P_A))} \frac{P_A(\Omega_\omega)}{P_A(F_\omega)}. \end{aligned} \tag{4.4}$$

Remark 4.1. (a). The definition looks (and is) cumbersome, but it arises naturally if one regards the two coordinates of a point $(\omega, k) \in \Omega = \Omega_O \times \Omega_A$ as being at fundamentally different levels. In the terminology of philosophers, elements $\omega \in \Omega_O$ are regarded as 'possible worlds', while $k \in \Omega_A$ gives the observer's location in that world.

Suppose that an AO with initial credence P_A on $(\Omega, \mathcal{F})$ learns that the event $F \subset \Omega$ holds. HTM conditioning gives priority to the new information F gives about possible worlds, so the first step is to reduce to the set of possible worlds which are compatible with the observation F, and up to a null set this is $\mathcal{S}(F, P_A)$. HTM conditioning proceeds by first conditioning on $\mathcal{S}(F, P_A)$, and then dividing the probability $\overline{P}_A(\omega|\mathcal{S}(F, P_A))$ among the points in F_ω in proportion to their original P_A-probabilities.
(b) If Ω_O is uncountable one would need a different definition. Given the questionable utility of HTM conditioning we will not explore that issue.

We now explore the properties of this definition.

Lemma 4.2. *Let* $P_A(F) > 0$.
(a) $P_A((\omega, k)||F) > 0$ *if and only if* $P_A((\omega, k)) > 0$ *and* $(\omega, k) \in F$.
(b) We have

$$P_A(F_\omega||F) = P_A(\Omega_\omega||F) = P_A(\omega|\mathcal{S}(F, P_A)). \tag{4.5}$$

(c) $P_A(\cdot||F)$ *is a probability measure on* $(\Omega, \mathcal{F})$.

Proof. (a) This is immediate from (4.4).
(b) This follows immediately on summing (4.2) over k.
(c) Using (b) we have

$$\sum_{\omega\in\mathcal{S}(F,P_A)} \sum_{k\in\Omega_A} P_A((\omega, k)||F) = \sum_{\omega\in\mathcal{S}(F,P_A)} P_A(F_\omega||F) = \sum_{\omega\in\mathcal{S}(F,P_A)} \overline{P}_A(\omega|\mathcal{S}(F, P_A)) = 1.$$

The countable additivity of $P_A(\cdot||F)$ is immediate from (4.3). $\square$

The next Lemma, which is a straightforward consequence of Definition 4.2, shows that HTM conditioning iterates correctly.

Lemma 4.3. *If $F, G \in \mathcal{F}$ with $P_A(F \cap G) > 0$, then writing $\widetilde{P}_A(\cdot) = P_A(\cdot||F)$, we have*

$$\widetilde{P}_A(H||G) = P_A(H||F \cap G) \text{ for } H \in \mathcal{F}. \tag{4.6}$$

HTM conditioning takes a simpler form if one of the events is in $\mathcal{F}_W$.

Lemma 4.4. *Let $F, G \in \mathcal{F}$ with $P_A(F) > 0$.*
(a) If $G \in \mathcal{F}_W$ then

$$P_A(G||F) = P_A(G|\mathcal{S}(F, P_A) \times \Omega_A).$$

(b) If $F \in \mathcal{F}_W$ then

$$P_A(G||F) = P_A(G|F).$$

Proof. (a) We can write $G = G_o \times \Omega_A$ where $G_o \in \mathcal{F}_O$. So $G_\omega = \Omega_\omega$ for $\omega \in G_o$ and

$$\begin{aligned} P_A(G||F) &= \sum_{\omega \in G_o} P_A(G_\omega||F) = \sum_{\omega \in G_o} \overline{P}_A(\omega|\mathcal{S}(F, P_A)) \\ &= \overline{P}_A(G_o|\mathcal{S}(F, P_A)) = P_A(G|\mathcal{S}(F, P_A) \times \Omega_A). \end{aligned}$$

(b) If $F \in \mathcal{F}_W$ then $F = (\mathcal{S}(F, P_A) \times \Omega_A) \cup D$, where $D \in \mathcal{F}_W$ and $P_A(D) = 0$. Then $P_A(F) = \overline{P}_A(\mathcal{S}(F, P_A))$ and $F_\omega = \Omega_\omega$ for all $\omega \in \mathcal{S}(F, P_A)$. The result then follows from (4.4). $\square$

Rather remarkably, in a number of important cases HTM conditioning of the law P_L gives the same answer as standard conditioning of P_E.

Proposition 4.5. *Consider a countable probability space $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P})$ carrying a GSB model together with an auxiliary process $Z = (Z_j, 1 \le j \le M)$ which takes values in a finite set $\mathcal{Z}$.*
(a) If $F \in \mathcal{F}_W$ and $1 \le j \le M$ then

$$P_L(F||S = j) = P_E(F|S = j). \tag{4.7}$$

(b) If $F \in \mathcal{F}_W$, $1 \le j \le M$, $y \in \mathcal{Z}$ and $\mathbb{P}(Z_j = y, X \ge j) > 0$ then

$$P_L(F||Z_S = y, S = j) = P_E(F|Z_S = y, S = j). \tag{4.8}$$

(c) If Z is injective, $\mathbb{P}(G_y) > 0$, and $F \in \mathcal{F}_W$ then

$$P_L(F||Z_S = y) = P_E(F|Z_S = y). \tag{4.9}$$

Proof. We can assume that $\mathbb{P}(\Omega_\omega) > 0$ for all $\omega \in \Omega_O$. Recall that $P_L(F) = \mathbb{P}(F)$ for $F \in \mathcal{F}_W$.
(a) As $q_M > 0$ and P_E and P_L both satisfy (PI) it follows that $P_L(S = j) > 0$ and $P_E(S = j) > 0$. We have $\omega \in \mathcal{S}(\{S = j\}, P_L)$ if and only if $X^O(\omega) \geq j$. So by Lemma 4.4(a) $P_L(F||S = j) = P_L(F|X \geq j) = \mathbb{P}(F|X \geq j)$. As P_E satisfies (PEI) this equals $P_E(F|S = j)$.
(b) Set $D = \{S = j, Z_j = y\}$. Then $\mathcal{S}(D, P_L) \times \Omega_A = \{X \geq j, Z_j = y\}$. It follows that $P_L(Z_j = y, S = y) > 0$. Hence

$$P_L(F||Z_j = y, S = j) = \mathbb{P}(F|Z_j = y, S = j).$$

Similarly we have $P_E(Z_j = y, S = y) > 0$, and using (PEI) we obtain $P_E(F|Z_S = y, S = j) = \mathbb{P}(F|Z_j = y, X \geq j)$, which proves (4.5).
(c) This does not follow from (b), since, as we will see below, HTM conditioning does not in general satisfy the law of total probability. Recall the definition of G_y and H_y from Section 2. Then $\mathcal{S}(H_y, P_L) = G_y$, and the condition on L_y implies that $P_L(H_y) > 0$. Hence

$$P_L(F||H_y) = P_L(F|G_y) = \frac{\mathbb{P}(F \cap G_y)}{\mathbb{P}(G_y)}. \tag{4.10}$$

As Z is injective $L_y = 1_{G_y}$ and the right side of (4.10) equals $\mathbb{E}(1_F L_y)/\mathbb{E}(L_y)$, which by Lemma 2.8 is equal to $P_E(F|H_y)$. □

We now look at some examples

Example 4.6. *Sleeping Beauty.* Using the notation of Example 2.7, by Proposition 4.5(a) we have

$$P_L(\text{Heads}||\text{Mon}) = P_L(X = 1||S = 1) = P_E(X = 1|S = 1) = \tfrac{1}{2}.$$

Thus HTM conditioning does indeed do its desired job in this case.

Example 4.7. *Technicolour Beauty.* We use the notation of Example 2.13, and recall that $\Omega = \Omega_O \times \Omega_A = \Big(\{1, 2\} \times \{0, 1\}\Big) \times \{1, 2\}$. The probability P_L on this space assigns probability $\frac{1}{4}$ to the points $(1, j, 1)$ for $j = 0, 1$, probability zero to $(1, j, 2)$ for $j = 0, 1$, and probability $\frac{1}{8}$ to the remaining points $(2, j, k)$ for $j = 0, 1$, $k = 1, 2$. We have

$$P_L(X = 1) = \tfrac{1}{2},\ P_L(X = 1|S = 1) = \tfrac{2}{3},\ P_L(X = 1||S = 1) = \tfrac{1}{2}.$$

The first two equalities above come from straightforward computations, while the final one comes from Proposition 4.5(a). However by Proposition 4.5(c)

$$P_L(X = 1||Z_S = 0) = P_E(X = 1|Z_S = 0) = \tfrac{1}{3}.$$

This example, which shows that HTM conditioning is sensitive to what should be irrelevant information, is given in [40]. Further, it is easy to verify that $P_L(Z_S = 0) = P_L(Z_S = 1) = \frac{1}{2}$, so that

$$\tfrac{1}{2} = P_L(X = 1) \neq \sum_{j=0}^{1} P_L(X = 1||Z_S = j)P_L(Z_S = j) = \tfrac{1}{3}.$$

Thus HTM conditioning fails to satisfy the law of total probability. According to the Double Halfer, after she wakes up, but before she opens her eyes, SB assigns a credence of $\frac{1}{2}$ to Heads (i.e. $X = 1$). She then opens her eyes, observes Z_S, and whatever she sees she then changes her credence to $\frac{1}{3}$.

Example 4.8. *A large universe version of Technicolour Beauty.* This example will be used again in Section 7.7. Let $1 \ll N_1 \ll M \ll N_2$, and consider the GSB model with $q_{N_1} = q_{N_2} = \frac{1}{2}$, and an auxiliary process $Z = (Z_1, \ldots, Z_{N_2})$ where Z_j are i.i.d., independent of X, and uniformly distributed on a space $\mathcal{Z}$ with $|\mathcal{Z}| = M$. (Note that this auxiliary process is not injective.) Let $y \in \mathcal{Z}$. We wish to calculate

$$P_L(X = N_1 || Z_S = y).$$

We therefore set $\Omega_O = \{1, 2\} \times \mathcal{Z}^{N_2}$, and write points in Ω_O as $\omega = (\omega_O, z)$, where $z = (z_1, \ldots, z_{N_2}) \in \mathcal{Z}^{N_2}$. We take $\Omega_A = \{1, \ldots, N_2\}$, and write points in Ω as (ω_O, z, k). We define the random variables X, Z_j, S by $X(\omega_O, z, k) = N_{\omega_O}$, $Z_j(\omega_O, z, k) = y_j$ and $S(\omega_O, z, k) = k$. P_L is the probability such that

$$P_L(\omega_O, z, k) = \begin{cases} \frac{1}{2} M^{-N_2} N_1^{-1} & \text{if } \omega_O = 1,\ 1 \le k \le N_1, \\ 0 & \text{if } \omega_O = 1,\ N_1 \le k \le N_2, \\ \frac{1}{2} M^{-N_2} N_2^{-1} & \text{if } \omega_O = 2,\ 1 \le k \le N_2. \end{cases}$$

Fix $y \in \mathcal{Z}$. The proof of (4.10) in Proposition 4.5 does not require that $L_y \le 1$, so we obtain

$$P_L(X = N_1 || Z_S = y) = \frac{\mathbb{P}(\{X = N_1\} \cap G_y)}{\mathbb{P}(G_y)} = \frac{\mathbb{P}(G_y | \{X = N_1\})}{\mathbb{P}(G_y)}.$$

Since $\mathbb{P}(G_y | X = N_i) = 1 - (1 - 1/M)^{N_i}$, we have $\mathbb{P}(G_y | X = N_1) \simeq N_1/M$ and $\mathbb{P}(G_y | X = N_2) \simeq 1$. Thus

$$P_L(X = 1 || Z_S = a) \simeq \frac{N_1}{M}. \tag{4.11}$$

Since $P_L(X = N_1) = \frac{1}{2}$, the observation $Z_S = y$ by SB causes a huge decrease in her credence that $X = N_1$, irrespective of the value of y. (This calculation still works if for each j the random variable Z_j is uniformly distributed on a set $\mathcal{Z}_j$ with $|\mathcal{Z}_j| = M$.)

HTM conditioning has two further undesirable properties. It is not not stable under convergence in total variation norm, and it is sensitive to 'hidden' properties of the probability space. For details see Appendix A.2.

Remark 4.9. Suppose that SB has woken up in her cell and chosen her credence P_A. She has to decide how she will update her credence for Heads if she is told later that it is Monday. She could use either $P_A(\text{Heads}|\text{Mon})$ or $P_A(\text{Heads}||\text{Mon})$. But, as we have seen, the second quantity behaves badly in a number of ways, and it is hard to see why she should prefer it to the classical definition of Kolmogorov.

5 Betting

One way to determine credences is to use betting arguments. We consider bets of the following form between two observers A and B. Let F_1 and F_2 be disjoint events, and let $a, b \in \mathbb{R}_+$ with $a + b > 0$. If F_1 occurs then B pays A \$$a$, and if F_2 occurs then B pays A \$$b$. The gain of A is

$$W^{(A)} = a 1_{F_1} - b 1_{F_2}, \tag{5.1}$$

and the bet is (weakly) favourable for A if $\mathbb{E}(W^{(A)}) \geq 0$. The bet is fair if the expected gain is zero, so that

$$a\,\mathbb{P}(F_1) = b\,\mathbb{P}(F_2). \tag{5.2}$$

If this bet such is fair, then it is fair for both players. (This might not occur if A and B have different information, or use different probability measures.) It is easy to check that the condition in (5.2) is equivalent to

$$\mathbb{P}(F_1 | F_1 \cup F_2) = \frac{b}{a+b}. \tag{5.3}$$

All the bets will be between the CO and an AO, and will be proposed by the CO. We write $W^{(CO)}$ for the payout to the CO – the payout to the AO is $-W^{(CO)}$. We will need to be careful that the arrangements for the bet do not lead to any information disparity between the CO and AO. To simplify terminology by *favourable* I will mean favourable or fair, and to avoid the need for small perturbations of bets, I will assume that the AO is a compulsive gambler and so will accept any bet which it perceives to be favourable. I also assume linear utility for both the CO and AO.

There is a substantial literature on betting arguments for the Sleeping Beauty problem, which includes [3, 25, 6, 14, 7, 12, 2, 47], and the size of this literature is in itself an indication that betting arguments are not conclusive in resolving the problem. Several issues arise, the first of which is the possible need for multiple bets.

The simplest bet in the standard SB problem is for the CO to phone the AO and offer the bet $\mathcal{B}_1$ with payout $W_1^{(CO)} = a 1_{(\text{Heads})} - b 1_{(\text{Tails})}$. But this bet immediately faces difficulties. If the bet is offered just once (say on Monday) then the offer of the bet will give the AO further information. The CO therefore has to offer the bet every time the AO wakes, and in this case one needs to decide how the CO and AO will take those other possible bets into account.

On Heads the CO gains a, but on Tails the CO has to offer the bet twice, so will lose $2b$. Thus the CO's gain over the week is $W_2^{(CO)} = a 1_{(\text{Heads})} - 2b 1_{(\text{Tails})}$, and so the CO will regard the bet as fair if $a = 2b$. (A minor difficulty is that the CO may have to offer the bet on both Monday and Tuesday, but without knowing the day: see [25] for one way of dealing with this problem.)

In the context of this paper, where the AO is a shabti, we have seen that there are Single shabti or Duplicated shabti versions of the problem. For the bet $\mathcal{B}_1$ one needs to ask how a shabti regards possible gains by itself in the past or future, or by another shabti.

In the single shabti case one can ask if it is an evidential decision theorist (EDT) or a casual decision theorist (CDT). A shabti knows that its past or future self will make the same decisions as it does. An EDT will take the effect of those decisions into account, while a CDT will not. Briggs [7] argues that SB should be a Thirder or Halfer according to whether she is a CDT or an EDT, though Arntzenius in [3] had already remarked that "it seems rather odd that SB's degrees of belief would depend on the decision theory that she accepts". Armstrong [2] looks further into the decision theory problem, and identifies five different types of possible agents.

Given these difficulties we will only consider bets which are offered at most once during the experiment.

The argument we use is based on the 'Late bet' in [47]. We consider the GSB experiment in the single serial shabti version. Write $I_M = \{1, \dots, M\}$, and recall that $\Omega_A = \{\partial\} \cup I_M$.

Definition 5.1. Let $F \in \mathcal{F}_W$, and $T : \Omega \to \Omega_A$ be a $\mathcal{F}_W$-measurable random variable which satisfies $\mathbb{P}(T \in I_M) > 0$ and $T \le X$ whenever $T \in I_M$. Set $D_T = \{T \in I_M\}$, let $a, b \in \mathbb{R}_+$ with $a + b > 0$ and set

$$W = W(F, T, a, b) = a 1_{F \cap D_T} - b 1_{F^c \cap D_T}. \tag{5.4}$$

The bets $\mathcal{B}_0(W(F, T, a, b))$ and $\mathcal{B}_T(W(F, T, a, b))$ both have payouts W to the CO. The bet $\mathcal{B}_O$ is offered by the CO before the experiment – we have to assume that the AO has a pre-experimental existence. The bet $\mathcal{B}_T$ is offered by the CO to the AO at time T if D_T occurs. Note that the conditions on T imply that the cell C_T is occupied if D_T occurs. We call $\mathcal{B}_0$ and $\mathcal{B}_T$ *betting proposals*, and will say that they are $\mathbb{P}$-fair if $\mathbb{E}(W) = 0$.

Note that as the bet $\mathcal{B}_T$ and the time that it is offered is offered is fixed in advance, the CO cannot take advantage of the AO by using information, such as the value of T, which it may have gained during the experiment.

To use this bet to give information on the AO's credences, we need two steps:
(1) Determine for which a, b the AO will regard this bet as fair,
(2) Use these values of a, b to give information on P_A.

(1) We proceed, as in [47], by a diachronic Dutch Book argument. (This differs from the standard Dutch Book argument as given in [28].) This takes place in three time periods: Before, During and After the experiment. Bets are settled the the third (After) period. Let F, T, a, b be as in Definition 5.1, and suppose that $\mathbb{E}(W(F, T, a, b)) = 0$. Before the experiment the CO offers the AO the bet $\mathcal{B}_0(W(F, T, a, b))$. It is almost universally agreed (see [42, p. 1004]) that at this point the AO should have the same credences as the CO, and so both will regard this bet as fair.

Now let $a' < a$ and choose $1 < \lambda < a/a'$. During the experiment, if D_T occurs, the CO offers the bet $\mathcal{B}_T(-W(F, T, \lambda a', \lambda b))$ at time T. Suppose that the AO will regard this bet as fair. Then the final payout After the experiment will be

$$W^{(CO)} = W(F, T, a, b) - W(F, T, \lambda a', \lambda b) = (a - \lambda a') 1_{F \cap D_T} + b(\lambda - 1) 1_{F^c \cap D_T}.$$

This satisfies $W^{(CO)} \ge 0$ and $\mathbb{E}(W^{(CO)}) > 0$: the CO cannot lose and may gain, and this suggests that the AO should not regard the bet $\mathcal{B}_T(-W(F, T, \lambda a', \lambda b))$ as fair. Using linearity the AO should also regard the bet $\mathcal{B}_T(W(F, T, a', b))$ as unfair. A similar argument handles the case $a' > a$, and it follows that the AO should regard $\mathcal{B}_T(W(F,T,a,b))$ as fair.

(2) Suppose that the AO regards the bet $\mathcal{B}_T(W(F,T,a,b))$ as fair. We wish to use this to obtain information on the AO's credence P_A. When the bet is offered the AO knows that $S = T$ but has no other information. The natural criterion is that P_A should satisfy $E_A(W(F,T,a,b)|S = T) = 0$. We will assume that this is the case, but one should note that some papers, such as [3, 6], dispute the connection between fair bets and credences.

These arguments support the following Principle.

Principle of Fair Betting (PFB). Let P_A be an anthropic probability. We say that (PFB) holds for P_A if whenever F, T, a, b are as in Definition 5.1, and $\mathcal{B}_T(W(F,T,a,b))$ is a $\mathbb{P}$-fair betting proposal, then $E_A(W(F,T,a,b)|S = T) = 0$.

Theorem 5.2. *Let P_A be an anthropic probability which satisfies (PFB).*
(a) P_A satisfies (PEI).
(b) If an injective spanning auxiliary process Z exists then P_A satisfies (PZ) and hence $P_A = P_E$.

Proof. Let Z be an injective auxiliary process with values in a set $\mathcal{Z}$. We can assume that $\mathbb{P}(G_y) > 0$ for each $y \in \mathcal{Z}$. For $y \in \mathcal{Z}$ let $T_y = \partial$ on G_y^c, and

$$T_y = \min\{j \geq 1 : Z_j = y\} \text{ on } G_y.$$

Then $D_{T_y} = G_y$ and $\{S = T_y\} = H_y$.

Fix y and let $F \in \mathcal{F}_W$, and choose a, b so that $\mathcal{B}_T(F, T_y, a, b)$ is $\mathbb{P}$-fair. By (PFB) we have

$$E_A(W(F, T_y, a, b)|S = T_y) = E_A(a1_{F\cap G_y} - b1_{F^c\cap G_y}|H_y) = 0. \tag{5.5}$$

It follows that $aP_A(F|H_y) = b/(a+b)$ and hence that $P_A(F|H_y) = \mathbb{P}(F|G_y)$, so that (PZ) holds.

(a) Taking $Z_j = j$ for $1 \leq j \leq M$ and using Remark 2.12 we deduce that (PEI) holds.

(b) As Z is spanning and (PZ) holds we have $P_A = P_E$ by Theorem 2.15. □

Remark 5.3. The Theorem above shows that betting arguments support (PEI) and (if a suitable spanning process Z exists) that the AO should take $P_A = P_E$. Note though that the Dutch Book argument given here relies on the AO having a pre-experimental existence.

6 Long run frequencies

Suppose that the SB experiment is repeated over a period of N weeks. We assume that when SB is awoken she is ignorant both of the week and the day. We need to use the model in Section 3, and take $\Omega_O = \{H,T\}^N$, $\mathcal{F}_O$ to be the set of all subsets of Ω_O and $\mathbb{P}_O$ to be the probability which assigns mass 2^{-N} to each point in Ω_O. We take $\Omega_A = \{1,\dots,N\} \times \{1,2\}$ and let $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P}^N, \mathcal{X}, S)$ be as defined in Section 3: we write $\mathbb{P}^N$ to emphasize the dependence on N. We write the r.v. $S = (S_W, S_D)$; here S_W specifies the week and S_D the day. Let $\mathcal{X}_j \subset \{(j,1),(j,2)\}$ be the set of occupied cells in week j, and $\mathcal{X} = \cup_{j=1}^N \mathcal{X}_j$.

Write $X_j \in \{H,T\}$ for the outcome of the jth toss; under $\mathbb{P}^N$ the r.v. X_j are independent and identically distributed with $\mathbb{P}^N(X_j = H) = \mathbb{P}^N(X_j = T) = \frac{1}{2}$. With Q_x as given by (3.2), we have $Q_{(j,1)} = 1$, $Q_{(j,2)} = \frac{1}{2}$. Since $\mathbb{E}^N(|\mathcal{X}|) = 3N/2$ we have $\lambda = 2/3N$.

We say that week n is a *Heads week* if $X_n = H$, and let

$$Z_N = \sum_{n=1}^N 1_{(X_n=H)}$$

be the number of Heads weeks. We say that an awakening by the AO on day j of week n is a *Heads awakening* if $X_n = 1$ (and so one must have $j = 1$). The total number of awakenings is is $2N - Z_N$, the number of Heads awakenings is Z_N, and $Z_N/(2N - Z_N)$ and Z_N/N are respectively the proportions of Heads awakenings and Heads weeks. The weak law implies that

$$\frac{Z_N}{2N - Z_N} \to \frac{1}{3}, \quad \frac{Z_N}{N} \to \frac{1}{2} \text{ in } \mathbb{P}^N \text{ - probability as } N \to \infty. \tag{6.1}$$

(If we had an infinite sequence of experiments then the weak limits in (6.1) could be replaced by a.s. convergence.)

The existence and values of these limits is not disputed in the literature. What is disputed is how to use (6.1) to find the P_A probability of Heads. Do we use the proportion of Heads-awakenings, or of Heads-weeks? Since we are asking for the credences of the AO on an awakening, I agree with [46] in thinking that the first is more reasonable, but I do not regard this argument as overwhelming.

The connection between probabilities and long run frequencies is made precise in the classical theory by the weak (or strong) laws of large numbers. In the context here one would like to connect the limits in (6.1) with the law P_A in a mathematically precise way.

Unfortunately this does not seem to be easy. To explore the difficulties we look at the Halfer and Thirder measures for this model. If $a = (a_1, \dots, a_N) \in \{H,T\}^N$ and $B \subset \Omega_A$ then the Halfer probability is defined by

$$P_L^{(N)}(\{a\} \times B) = 2^{-N} \frac{|B \cap \mathcal{X}^O(a)|}{|\mathcal{X}^O(a)|}. \tag{6.2}$$

Since $P_L^{(N)}(F) = \mathbb{P}^N(F)$ for any $F \in \mathcal{F}_W$, it is immediate that under $P_L^{(N)}$ the random variables $(X_i, 1 \le i \le N)$ are independent with $P_L^{(N)}(X_i = H) = \frac{1}{2}$. Hence

$$\frac{Z_N}{2N - Z_N} \to \frac{1}{3}, \quad \frac{Z_N}{N} \to \frac{1}{2} \text{ in } P_L^{(N)} \text{ - probability as } N \to \infty. \tag{6.3}$$

Let $P_E^{(N)}$ be the unique probability such that (SPN), (SPEI) and (SPI) hold. Since $\mathbb{P}(\mathcal{X} = \Omega_A) = 2^{-N} > 0$ the graph (K,E) given in Section 3 is connected, and by Theorem

3.1 we have

$$P_E^{(N)}(S=(j,1)) = \frac{2}{3N}, \tag{6.4}$$

$$P_E^{(N)}(S=(j,2)) = \frac{1}{3N}, \text{ for } 1 \le j \le N, \tag{6.5}$$

$$P_E^{(N)}(F) = \frac{\mathbb{E}(|\mathcal{X}|1_F)}{\mathbb{E}(|\mathcal{X}|)} \text{ for } F \in \mathcal{F}_W \ . \tag{6.6}$$

Lemma 6.1. *Assume that (SPN), (SPEI) and (SPI) hold. Then*

$$P_E^{(N)}(X_j = H) = \frac{3N-1}{6N}, \tag{6.7}$$

$$P_E^{(N)}(X_{S_W} = H) = \tfrac{1}{3}, \tag{6.8}$$

$$P_E^{(N)}(X_j = H | S_W \neq j) = \tfrac{1}{2}. \tag{6.9}$$

Proof. Setting $F = \{X_j = H)$ we have $\mathbb{E}(|\mathcal{X}|1_F) = (3N-1)/4$, so (6.7) follows immediately from (6.6). Using (SPEI) and (3.6) we have

$$P_E^{(N)}(X_j = H, S_W = j) = \mathbb{P}(X_j = H|(j,1) \in \mathcal{X}) P_E^{(N)}(S=(j,1)) = \frac{1}{3N}.$$

Summing over j gives (6.8). Since $P_E^{(N)}(X_j = H, S_W \neq j) = P_E^{(N)}(X_j = H) - P_E^{(N)}(X_j = H, S_W = j)$, a little algebra gives (6.9). □

This Lemma shows that in the N week experiment with the law $P_E^{(N)}$ the AO is a Thirder for the current week and a Halfer for the experiments in all the other weeks. It follows from (6.7) that $E_E(|\mathcal{X}|) = 3N/2 + \frac{1}{6}$, so that the size-biasing effect from using $P_E^{(N)}$ rather than $\mathbb{P}$ is very mild.

A straightforward calculation using (6.6) gives that under $P_E^{(N)}$ the r.v. X_1, X_2 are not independent. However, X_j for $j \neq S_W$ are independent.

Lemma 6.2. *Let $a = (a_1, \dots, a_N)$ be a sequence in $\{H,T\}^N$. Then*

$$P_E^{(N)}(X_i = a_i, i \neq j | S_W = j) = 2^{-(N-1)}. \tag{6.10}$$

Consequently

$$\frac{Z_N}{2N - Z_N} \to \frac{1}{3}, \quad \frac{Z_N}{N} \to \frac{1}{2} \text{ in } P_E^{(N)} \text{ - probability as } N \to \infty. \tag{6.11}$$

Proof. Since $\mathcal{X} \setminus \mathcal{X}_j$ is independent of $\mathcal{X}_j$ under $\mathbb{P}$, we have by (SPEI) that

$$P_E^{(N)}(X_i = a_i, i \neq j | S_W = j, S_D = k) = \mathbb{P}(X_i = a_i, i \neq j | (j,k) \subset \mathcal{X}_j) = 2^{-(N-1)},$$

and (6.10) is immediate.

Let $Z'_N = Z_N - 1_{(X_{S_W}=H)}$. Then by the weak law of large numbers $(N-1)^{-1}Z'_N \to \frac{1}{2}$ in $P_E^{(N)}$-probability. As $|Z_N - Z'_N| \le 1$ (6.11) follows. □

As the limits in (6.3) and (6.11) are the same, we cannot use them to distinguish between the probabilities $P_E^{(N)}$ and $P_L^{(N)}$.

Finally, let us calculate SB's credence that the coin toss in the current week is Heads under $P_L^{(N)}$. Write $N_T(a) = |\{j : a_j = T\}|$ be the number of tails in the sequence a. If $N_T(a) = k$ then $P_L^{(N)}(X_{S_W} = H|X = a) = (N-k)/(N+k)$, and so

$$P_L^{(N)}(X_{S_W} = H) = \sum_{k=0}^{N} 2^{-N} \binom{N}{k} \frac{N-k}{N+k}. \tag{6.12}$$

Using symmetry we also have

$$P_L^{(N)}(X_{S_W} = H) = P_L^{(N)}(X_1 = H|S_W = 1) = N P_L^{(N)}(S_W = 1, X_1 = H).$$

Writing y_N for the left side of (6.12) we find, as in [9], that $y_1 = \frac{1}{2}$, $y_2 = \frac{5}{12}$ and that $y_N \to \frac{1}{3}$ as $N \to \infty$. Since $P_L^{(N)}(X_1 = H|S_W = 1)$ depends on N the model restriction property described in Section 3 fails for $P_L^{(N)}$.

7 Examples

7.1 Sleeping Beauty: life outside the experiment

A variant [11] of Sleeping Beauty looks at her history over the four days Sunday to Wednesday; for convenience let us label these days 1 to 4. If the coin is Heads she sleeps through Tuesday, and is not given the forgetfulness potion. If it is Tails she is woken on both Monday and Tuesday, and given the potion once, on Monday evening. Thus on waking on Wednesday she recalls events on Sunday, as well as one day in a cell, which might be either Monday or Tuesday.

We can analyse this using the set version of the model from Section 3, by taking $K = \{1, \dots, 4\}$ and taking $\mathcal{X} = \{1, 2, 3, 4\}$ on {Tails}, $\mathcal{X} = \{1, 2, 4\}$ on {Heads}. We assume that (SPN), (SPEI) and (SPI) hold. The probability P_E describes the credence of SB on waking, and before she receives any information at all. So in this case we assume that when she wakes each day there is a short period before she can access her memory from previous days.

Using Theorem 3.1 we deduce that $\lambda^{-1} = \mathbb{E}(|\mathcal{X}|) = 7/2$, and $P_E(S = j) = \frac{2}{7}$ for $j = 1, 2, 4$ and $P_E(S = 3) = \frac{1}{7}$. By Proposition 3.3 if we condition on $S \in \{2, 3\}$ we obtain the measure for the standard Monday–Tuesday SB model.

Easy calculations give that $P_E(H|S \in \{2, 3\}) = \frac{1}{3}$, $P_E(H|S = j) = \frac{1}{2}$ for $j = 1, 2, 4$, and $P_E(H|S = 3) = 0$. If we follow SB through the experiment, on day 1 she knows $\{S = 1\}$, on days 2 and 3 she knows $\{S \in \{2, 3\}\}$, and on day 4 she knows $\{S = 4\}$. Her credence for Heads on days 1,2,4 follows the sequence $(\frac{1}{2}, \frac{1}{3}, \frac{1}{2})$. In standard probability theory the probability of an event for an observer follows a martingale, but this sequence is not a martingale – for one thing it is deterministic and non-constant.

It is straightforward to calculate P_L for this model. Since $P_L(H) = \frac{1}{2}$, we obtain $P_L(H, S = j) = \frac{1}{6}$ for $j = 1, 2, 4$, and $P_L(T, S = j) = \frac{1}{8}$ for $j = 1, \ldots, 4$. Thus

$$P_L(H|S \in \{2,3\}) = \frac{P_L(H, S = 2)}{P_L(S = 2, 3)} = \frac{4}{10}.$$

Hence the model restriction property described in Section 3 fails for P_L if we use standard conditioning.

One of Lewis' main arguments in [32] for the standard Halfer position is that SB has the same information when she wakes on Monday as when she went to sleep on Sunday. Thirders such as [25, 45, 13] have argued that she loses information: on Sunday she knows her location, and on Monday she does not. In the model in this paper the change is described not by a change in information, but by a change in the conditioning event: on Sunday SB uses the probability $P_A(\cdot|S = 1)$, and during the experiment she uses $P_A(\cdot|S \in \{2, 3\})$.

For more on updating in centred worlds see Section A.3.

7.2 Improper conditioning events and symmetry breaking

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be with a a probability space, and F, G, H be events with $H \subset G$ and $\mathbb{P}(G \setminus H) > 0$. The usual conditional probability of F given H is

$$\mathbb{P}(F|H) = \frac{\mathbb{P}(F \cap H)}{\mathbb{P}(H)}, \tag{7.1}$$

which is often informally described as being "the probability that F will occur given that the observer knows that H occurs". However if H occurs then G also occurs, so that the informal description suggests that it would also be valid to use $\mathbb{P}(F|G)$. This is so obviously incorrect that few if any introductions to conditional probability consider this possibility. (Some works are quite careful in their explanations of conditional probability and avoid the informal description above – see [17, p. 115].) In terms of Kolmogorov's definition of conditional expectation, one wants to look at

$$\mathbb{E}(1_F|\sigma(1_H)) = 1_H \mathbb{P}(F|H) + 1_{H^c} \mathbb{P}(F|H^c). \tag{7.2}$$

If one replaces H by G in (7.2) then the resulting random variable is not measurable with respect to the agent's observation, which is of 1_H. I will call conditioning events of the kind that G is here *improper conditioning events*. Easy examples show that using improper conditioning events can give wildly incorrect answers.

In the standard SB experiment she has identical awakenings on Monday and Tuesday, and we have seen how one can break this symmetry by introducing an auxiliary process Z as in Section 2.2. This suggests that one can describe SB's experience simply by using the objective probability $\mathbb{P}$. One paper which adopts this approach is [11]. Using the notation of this paper, and recalling from Section 2 the definitions of G_y and H_y, they define

$$P_{Z,y}(F) = \mathbb{P}(F|G_y) \text{ for } G_y \in \mathcal{F}_W. \tag{7.3}$$

The interpretation in [11] is that if the AO observes that the process Z in its cell takes the value y (i.e. the event H_y) then it knows that G_y occurs, and so (7.3) gives the AO's credence that F occurs. Since G_y is an improper conditioning event this procedure is of questionable validity. If Z is injective then P_E satisfies (7.3) by Lemma 2.10, but as [11] allows more general Z it is not surprising that by varying the process Z they are able obtain a variety of values for $P_{Z,y}(\text{Heads})$. The framework of [11] does not allow them to combine the laws $P_{Z,y}$ into a single probability P_Z which gives SB's credences before she makes an observation. As the authors of [11] remark, they cannot use the law of total probability since the events G_y are not disjoint.

The next section looks at an example of the use of improper conditioning events in the physics literature.

7.3 A model of Hartle and Srednicki

The paper [23] questions arguments, such as those in [34], which assume that we are 'typical' observers. To help clarify the issues they introduce a model of a universe which has N successive cycles, and assume that observers in cycle i have no knowledge of the state of the universe in any other cycle. Each cycle has one of two global properties: red (R) or blue (B). In each cycle the probability that an 'observing system' (such as 'us') exists is $p \in (0, 1)$, and is independent of whether the global state is red or blue.

There are two theories of the universe, which are equally likely. The first, AR is that all the cycles are red, and the second, SR, is that exactly M of the cycles are red, and $N - M$ are blue. Here $1 \le M \le N - 1$. Hartle and Srednicki do not specify which cycles are red or blue in the SR case; for simplicity we will assume that under SR each possible arrangement of M red and $N - M$ blue is equally likely. For each theory we write $N_R(T)$ and $N_B(T)$ for the number of red and blue cycles according to the theory T. (Thus $N_R(AR) = N$ and $N_R(SR) = M$.)

The authors of [23] wish to calculate the probability of SR given that 'we' (situated in one of the cycles) observe that our cycle is red. Writing H_{ER} for the event "we exist and observe red", and H_E for the event "we exist", they claim that

$$P(H_{ER}|T) = 1 - (1 - p)^{N_R(T)}, \tag{7.4}$$

and then use Bayes' formula to compute $P(T|H_{ER})$. Writing $f(p, n) = 1 - (1 - p)^n$, they obtain

$$P(SR|H_{ER}) = \frac{f(p, M)}{f(p, N) + f(p, N)}. \tag{7.5}$$

This conclusion is counterintuitive. If for example $M = 1$, $N = 1000$ and p is very close to 1, then (7.5) gives that $P(SR|H_{ER}) \approx \frac{1}{2}$: our observation of red has given very little information even though, on the hypothesis SR, red universes are very rare.

However (7.4), and hence (7.5), are incorrect. To see that (7.4) cannot be true, note that by red/blue symmetry we would also have $P(H_{EB}|T) = 1 - (1-p)^{N_B(T)}$. Since $P(H_{ER}|T) + P(H_{EB}|T) = P(H_E|T) \le 1$, we obtain

$$2 - (1-p)^{N_R(T)} - (1-p)^{N_B(T)} \le 1, \tag{7.6}$$

and this fails for many values of p, $N_R(T)$ and $N_B(T)$.

The error which led to (7.4) is confusion between the events $H_{ER} = \{$we exist and observe red$\}$ and $G_{ER} = \{$there is a cycle in which an observer exists and observes red$\}$. Since $H_{ER} \subset G_{ER}$ if we observe that H_{ER} occurs then we also observe that G_{ER} occurs. The events G_{ER} and G_{EB} are not disjoint, which is why the sum on the left hand side of (7.6) can be greater than 1. I suspect that the cause of this error is that the authors of [23] were (rightly) uneasy about the event H_{ER}, and wished to find a suitable objective event which they could work with. In the terminology of the previous section, G_{ER} is an improper conditioning event.

We now apply our formalism to this problem. We define

$$\Omega_O = \{AR, SR\} \times (\{R, B\} \times \{0, 1\})^N.$$

An element $\omega \in \Omega_O$ is a sequence

$$\omega = (y, a_1, n_1, a_2, n_2, \dots, a_N, n_N), \tag{7.7}$$

where $y \in \{AR, SR\}$, and for $1 \le i \le N$ we have $a_i \in \{R, B\}$ and $n_i \in \{0, 1\}$. Thus a_j gives the state (red or blue) of cycle j, and $n_j = 1$ if there are observing systems in cycle j, and $n_j = 0$ otherwise. This model is in the framework given in Section 3, with $K = \{1, \dots, N\}$. Since it is possible in this model to have no observers, we define $\Omega_A = \{\partial, 1, \dots, N\}$. With ω given by (7.7) we define random variables on $\Omega = \Omega_O \times \Omega_A$ by

$$U(\omega, k) = y, \quad S(\omega, k) = k, \quad Z_i(\omega, k) = a_i, \quad \xi_i(\omega, k) = n_i, \text{ for } 1 \le i \le N. \tag{7.8}$$

The random set of occupied universes is $\mathcal{X} = \{i : \xi_i = 1\}$. Under the objective probability law $\mathbb{P}$ the random variables U, $(\xi_j, 1 \le j \le N)$ are independent with $\mathbb{P}(U = AR) = \mathbb{P}(U = SR) = \frac{1}{2}$, and $\mathbb{P}(\xi_j = 1) = 1 - \mathbb{P}(\xi_j = 0) = p$. If $U = AR$ then $Z_i = R$ for all i, while if $U = SR$ then exactly M of the Z_i are equal to R, and the remainder are equal to B, with each ordering of reds and blues being equally likely. We assume that (SPN), (SPI) and (SPEI) hold and write write P_E for the anthropic probability which satisfies these conditions. Note that $H_{ER} = \{Z_S = R\}$. For $y \in \{R, B\}$ let

$$L_y = |\{j : \xi_j = 1, Z_j = y\}| = |\mathcal{X} \cap \{i : Z_i = y\}|.$$

The set analogue of Lemma 2.8 gives

$$P_E(F|Z_S = y) = \frac{\mathbb{E}(1_F L_y)}{\mathbb{E}(L_y)}. \tag{7.9}$$

Taking $F = \{U = SR\}$ we have $\mathbb{E}(1_F L_y) = \frac{1}{2}\mathbb{E}(L_y|U = SR) = \frac{1}{2}pM$, and similarly $\mathbb{E}(L_y) = \frac{1}{2}p(M+N)$. Hence

$$P_E(U = SR|Z_S = R) = \frac{M}{M+N}. \tag{7.10}$$

In the case noted above, when $M = 1$, $N = 1000$, we obtain the natural and unsurprising result that $P_E(U = SR|Z_S = R) = 1/1001$.

7.4 Is existence a special property?

As mentioned in the Introduction, one reason to study the SB problem is to understand how inference should be handled in situations where observers may not exist, or may have multiple copies. One can just proceed anyway, treating existence or non existence of an observer as the same kind of event as whether an observer sees 'e' or 'n' written on a randomly selected card. However it does not seem evident that it is correct to do this. An observer can see 'n' on the card, but cannot observe that it doesn't exist.

Consider two experiments.

Model 1. This is the standard SB problem; a fair coin is tossed and if the result is H then a shabti is woken once, while if it is T then it is woken twice. Using the superscript $M1$ to indicate that these probabilities are for Model 1, we have $P_E^{M1}(H) = \frac{1}{3}$ and $P_L^{M1}(H) = \frac{1}{2}$.

Model 2. In this case a fair coin is tossed, with outcomes H and T. Whatever the outcome, SB is woken twice, so that $\mathbb{P}(X = 2) = 1$. An auxiliary process Z^O with values in $\mathcal{Z} = \{e, n\}$ is defined by setting $Z_1^O(H) = Z_2^O(T) = Z_2^O(T) = e$, $Z_2^O(H) = n$. The AO wakes in its cell and is able to observe the value Z_S^O. Is seeing 'e' in Model 2 the same as existing in Model 1? (A similar model is considered in [27] and is used to support a Thirder conclusion.)

Model 2 is described by taking $M = 2$ and $q_2 = 1$. Let $L_\cdot$ be as in (2.20); we have $\mathbb{E}(L_e) = \frac{3}{2}$. Since $\mathbb{E}(L_e 1_H) = \frac{1}{2}$, by Lemma 2.8

$$P_E^{M2}(H|Z_S = e) = \tfrac{1}{3}. \tag{7.11}$$

Thus an AO which uses P_E in Models 1 and 2 will assign the same probability to Heads. This is not a surprise: (PEI) states that a CO which observes that cell C_j is occupied assigns the same probabilities to objective events as an AO which knows it is in cell C_j.

We can contrast this with what happens with the Halfer measure. The space $\Omega = \{H, T\} \times \{1, 2\}$ consists of four points, and P_L^{M2} assigns each of these a probability of $\frac{1}{4}$. Thus $P_L^{M2}(Z_S = e) = \frac{3}{4}$, and so for Model 2

$$P_L^{M2}(H|Z_S = e) = \tfrac{1}{3}.$$

As $P_L^{M2}(H) = \frac{1}{2}$, the measure P_L (with standard conditioning) does distinguish between the two models.

7.5 Fine tuning and a single conditional observer

The following is based on an example of Peter van Inwagen – see [44, p. 135] and [8, p. 25-28], which was given to clarify issues in the fine tuning debate. A model of this kind is also a component in the argument of [38].

Reframing the model in terms of shabti, rather than the colourful story of an execution given in [44], it is described as follows. We have two r.v. $Y_j^O \sim \text{Ber}(p_j)$, $j = 1, 2$ and we assume that $p_2 \ll p_1 < 1$. We set $X^O = \max(1, Y_1^O + Y_2^O)$, so that $X^O \in \{0, 1\}$, and if $X^O = 1$ a single shabti is placed in cell 1 and woken. If $X^O = 0$ then no shabti is woken. If the AO exists there are two possible causes for its existence: either $Y_2^O = 1$ and $Y_1 \in \{0, 1\}$, or else $Y_2^O = 0$ and $Y_1^O = 1$. What probabilities should the AO give to these two possibilities?

We model this using the framework of Section 2, with $\Omega_A = \{0, 1\}$. As usual we write Y_j, X for the extensions of Y_j^O and X^O to Ω. We assume that (PN) and (PEI) hold, and write P_A for the anthropic probability. Using (PN) we have $P_A(S = 1) = 1$. Then

$$\begin{aligned} P_A(Y_2 = 1) &= P_A(Y_2 = 1 | S = 1) = \mathbb{P}(Y_2 = 1 | Y_1 + Y_2 \geq 1) \\ &= \frac{p_2}{p_1 + p_2(1 - p_1)} \simeq p_2/p_1. \end{aligned}$$

This is the same as the probability the CO would assign to this event if it sees that the cell is occupied.

We now turn to the interpretation of this in terms of fine tuning. One should note that in this application, as well as in the remaining examples in this Section, the real universe differs from the GSB model in possibly significant ways. The GSB model supposes a CO, and that the AO, whether human or a shabti, knows the precise rules of the experiment. These conditions are used to justify Principles such as (PI) and (PEI). These conditions do not hold in the real world, and so we are less secure in using these principles than in the tightly constrained GSB model.

Having noted this point, in terms of fine tuning the event $\{Y_2 = 1\}$ corresponds to a single universe just happening to have the parameters suitable for complex chemistry (and so observers), while $\{Y_1 = 1\}$ corresponds to another explanation, such as the Multiverse or a Creator. Not surprisingly much more has been said on this question – see [18] for an introduction.

7.6 The Doomsday argument

The essence of this argument is that anthropic principles imply that we should be more pessimistic about the future of the human race than is suggested by objective probability estimates. An early published version is in [30]; see also [19, 31], and p. 89-90 of [8] for a brief history of the argument. A footnote on p. 89 of [8] lists a number of distinguished philosophers and physicists who accept some version of this argument. (The list does not mention any probabilists or statisticians.)

There are several versions of the argument, and one simple version, based on [30], is as follows. Suppose there are two possible futures of the human race: F_1 it becomes extinct around 2100, and F_2 it survives its current challenges, lasts for several million years, and colonizes the galaxy. Assume that initially we regard these two as equally probable. Let N_i be the total number of human beings which ever exist under future F_i, and take $N_1 = 10^{12}$ and $N_2 = 10^{24}$. The Doomsday Argument (DA) notes that under F_1 my birth order (among all the humans who will ever live) is fairly typical, while under F_2 I will be in the first 10^{-12} of humans. It then infers that as the second event is very unlikely, we should revise our estimates, and assign F_2 a much lower probability, in fact of the order of 10^{-12}.

There is an extensive literature on the Doomsday argument. It is well known that if one adopts a Halfer or SSA type of analysis, then the argument does have force, while using a Thirder or SIA analysis it does not – see for example [5].

The discussion below, which owes much to [13], exhibits these points using the framework of this paper. We consider a 'toy' DA based on the GSB model of Section 2. Let $(q_1, \dots, q_M)$ be a probability distribution on $\{1, \dots, M\}$, and look at the serial distinct shabti version of the problem, which takes place over M days. We use the same notation as in Section 2: in particular $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P})$ is the probability space constructed there, and $X : \Omega \to \{1, \dots, M\}$ is the total number of shabti. Let P_A be an anthropic probability; we assume it satisfies (PN) and (PI). Let $\lambda \in (0, 1)$, and let $k_n = \lfloor \lambda n \rfloor$. By (PI)

$$P_A(S \le \lambda X | X = n) = \frac{k_n}{n} \le \lambda.$$

Summing over n it follows that

$$P_A(X > \lambda^{-1} S) \le \lambda. \tag{7.12}$$

The equation (7.12) appears to give the AO information about the future, and this is the basis for the DA.

As we know, roughly, our birth position in the human race, we actually need to consider $P_A(X > \lambda^{-1} S | S = j)$. If P_A satisfies in addition (PEI), so that $P_A = P_E$, then we have

$$P_E(X > \lambda^{-1} S | S = j) = \mathbb{P}(X > \lambda^{-1} j | X \ge j) = Q_j^{-1} \mathbb{P}(X > \lambda^{-1} j). \tag{7.13}$$

Thus, on learning its cell number, the AO loses the apparent future information promised by (7.12) and reverts to the common sense view of the future.

If instead we use the measure P_L then $P_L(X = n) = q_n$ and $P_L(S = j | X = n) = 1/n$ if $j \in \{1, \dots, n\}$. So

$$P_L(S = j) = \sum_{n=j}^{M} \frac{q_n}{n},$$

and we find

$$P_L(X = n | S = j) = \frac{n^{-1} q_n}{\sum_{k=j}^{M} k^{-1} q_k}. \tag{7.14}$$

So under P_L and standard conditioning the anthropic law undergoes a reverse size biasing effect, and in the context of the Doomsday argument favours pessimistic outcomes. If however one uses P_L and Double Halfer conditioning then by Proposition 4.5 we find $P_L(X > \lambda^{-1}S||S = j) = P_E(X > \lambda^{-1}S||S = j)$, so that Double Halfers should join Thirders in rejecting the Doomsday argument.

For the simple example given at the start of this section we take $1 \ll N_1 \ll N_2$, and set $q_{N_1} = q_{N_2} = \frac{1}{2}$. If $j \leq N_1$ then (7.13) and (7.14) give

$$P_E(X = N_2|S = j) = \tfrac{1}{2}, \quad P_L(X = N_2|S = j) = \frac{N_2^{-1}}{N_1^{-1} + N_2^{-1}} \simeq \frac{N_1}{N_2}.$$

7.7 The Presumptuous Philosopher

We have seen some strong arguments for (PEI). (PN) appears innocuous, and (PI) also appears very plausible. Given all this, it may seem surprising that there are objections to the measure P_E. One reason is the following example, described in [8, p. 124].

Suppose there are two theories T1 and T2 for the size of the universe, *a priori* equally likely. According to T1 the universe contains $N_1 = 10^{24}$ galaxies, while according to T2 it contains $N_2 = 10^{10^{24}}$ galaxies. Which theory is correct depends on the mass m of some fundamental particle. Let us also suppose that the probability that a galaxy contains intelligent life is $p \geq 1000N_1^{-1}$ (independently of all other galaxies.) An experiment is proposed to measure m, which will give the correct result with probability greater than 99.9%, but as the final grant application is being prepared a Presumptuous Philosopher wanders into the Chief Scientist's office and explains that the experiment is not necessary: using PI and PEI we can be almost certain that T2 is correct.

Writing X for the number of intelligent civilizations, under the objective probability $\mathbb{P}$ we have $\mathbb{P}(X \simeq pN_j) \simeq \frac{1}{2}$ for $j = 1, 2$. But P_E selects the size biased distribution, so since $\mathbb{E}(X) \simeq \frac{1}{2}pN_2$, we deduce that $P_E(X \simeq pN_1) \simeq N_1/N_2$. So it is overwhelmingly likely that T2 is correct. In fact the probability that T1 is correct is so small that, even if the experiment did end up finding for T1, it would be very likely that this was due to experimental error. If one also accepts that it is legitimate to look at the limit as $N_2 \to \infty$, then it follows that one should accept that the universe is infinite with probability one.

Bostrom's SIA (see [8, p. 66]) also gives the same conclusion, and this leads Bostrom to reject SIA. But as we have seen the alternatives to P_E also encounter difficulties, which appear to me to be more immediate and more fundamental than the example of the Presumptuous Philosopher. This example might also be interpreted as showing that one needs to be cautious in applying the GSB model or Bayesian analysis to some real world situations.

It is worth noting that HTM conditioning also leads to a similar difficulty. We consider the same situation as above, but now assume that the Chief Scientist and Presumptuous Philosopher are both Double Halfers. The Presumptuous Philosopher points out that to distinguish T1 and T2 all we need to do is to make an observation which has a probability much less than 10^{-24}. Looking out of the window he sees a plot of grass 10m by 10m with about 1000 daises on it, and suggests paying some graduate students to record the position of each daisy to within 1cm. Writing $n = 10^6$ and $m = 10^3$, the number of ways of arranging m daisies in n 1cm square plots is $M = \binom{n}{m}$, which is around 10^{3000}. If all possibilities were equally likely then the probability of any one combination is of the order of $1/M$. Even taking into account clustering effects any arrangement should have probability p much less than 10^{-24}. Example 4.8 then shows that the probability of T1 given this observation is $pN_1 \ll 1$. (Since this probability does not converge to 0 as $N_2 \to \infty$, the example does take a less severe form than arises for Thirders.)

8 Conclusions

The purpose of this paper has been to provide a rigorous framework, based on Kolmogorov's axioms, in which to study the Sleeping Beauty problem, as well as related ones. A key feature of the model is that we have a probability space $(\Omega, \mathcal{F})$ with two probabilities – an objective probability $\mathbb{P}$ which is defined on a subset $\mathcal{F}_W$ of $\mathcal{F}$, and a second probability P_A on $\mathcal{F}$ which takes account if the location of the Anthropic Observer.

As P_A cannot be derived directly from the objective space $(\Omega, \mathcal{F}_W, \mathbb{P})$ one needs additional principles in order to calculate it. Suppose first that we assume the mild (PN). and also the Principle of Indifference (PI). Then we have two main alternative possibilities. The first is to adopt Lewis' Principal Principle, and this suggests that the AO should use the Halfer/SSA probability P_L defined by (2.3). The second is to choose the Principle of Equivalent Information (PEI), which if one also assumes (PI) leads to the Thirder probability P_E given in Theorem 2.4.

The paper gives several strong arguments for (PEI). As well as its intuitive reasonableness, it allows model restriction (Proposition 3.3), and is supported by betting arguments (Section 5). Finally, the strongest argument of all, PEI implies that P_E gives the same probability to future events as $\mathbb{P}$ – see Theorem 2.19. The measure P_L fails all these, at least if one the calculates conditional probabilities in the usual way.

The third main view, that of the Double (or Triple) Halfers uses P_L, but requires a new way of conditioning. We have reviewed one proposal, HTM conditioning, in Section 4 and seen that it has significant defects: it is sensitive to apparently irrelevant information, and fails to satisfy the law of total probability. Finally, one of the main arguments against (PEI) and P_E is the example of the Presumptuous Philosopher, but we have seen that Double Halfers encounter a similar difficulty.

A Appendices

A.1 The set model when (K, E) is not connected

In this case (K, E) can be partitioned into a finite number of non-empty connected components $H_1, \dots, H_k$. To avoid minor difficulties we assume that $\mathbb{P}(\mathcal{X} = \emptyset) = 0$, and that $\mathbb{P}(\mathcal{X} \subset H_j) > 0$ for each j. The same analysis as in the proof of Theorem 3.1 gives that if P_A satisfies (SPN), (SPI) and (SPEI) then there exist λ_j so that

$$P_A(S = x) = \lambda_j Q_x, \quad P_A(\mathcal{X} = B) = \lambda_j |B| q_B \quad \text{whenever } x \in H_1,\ B \subset H_1. \tag{A.1}$$

Let

$$p_j = \mathbb{P}(\mathcal{X} \subset H_j), \quad r_j = P_A(\mathcal{X} \subset H_j), \quad e_j = \mathbb{E}(|\mathcal{X}|\,|\mathcal{X} \subset H_j).$$

Summing (A.1) over $B \subset H_j$ we obtain $r_j = \lambda_j e_j p_j$, and thus (λ_j) must satisfy

$$\sum_{j=1}^{k} \lambda_j e_j p_j = 1. \tag{A.2}$$

However, neither (SPI) nor (SPEI) give any further constraint on the λ_j.

There appear to be two natural choices for (λ_j). The first to to take $\lambda_j^{-1} = e_j$, so that $P_A(X \subset H_j) = \mathbb{P}(X \subset H_j)$. In this case there is no size biasing effect when we look at the P_A-probability that X lies in some component. The second is to take $\lambda_j = \lambda$ for all j, so that λ satisfies

$$\lambda^{-1} = \sum_{j=1}^{k} e_j p_j = \mathbb{E}(|\mathcal{X}|).$$

We consider the following extensions of (SPN) and (SPEI).

Two point Principle of Equivalent Information. (SPEI2). If $F \in \mathcal{F}_W$, and x, y are in different connected components of (K, E) with $s_x > 0$, $s_y > 0$ then

$$P_A(F|S \in \{x, y\}) = \mathbb{P}(F|\{x, y\} \cap \mathcal{X} \neq \emptyset). \tag{A.3}$$

SPN2. Let H be a connected component of (K, E). If $\mathbb{P}(\mathcal{X} \subset H) > 0$ then $P_A(\mathcal{X} \subset H) > 0$.

Proposition A.1. *Suppose that $\mathbb{P}(\mathcal{X} = \emptyset) = 0$ and (K, E) has connected components $H_1, \dots, H_k$, with $k \geq 2$ and $\mathbb{P}(\mathcal{X} \subset H_j) > 0$ for each j. Assume that P_E satisfies (SPN), (SPI), (SPEI), (SPEI2) and (SPN2). Then $P_E(S = x) = \lambda Q_x$, $P_E(\mathcal{X} = B) = \lambda |B| q_B$ for all $x \in K$, $B \subset K$, where $\lambda = 1/\mathbb{E}(|\mathcal{X}|)$.*

Proof. Let $j \in \{2, \dots M\}$, and $x \in H_1$, $y \in H_j$ with $Q_x > 0$, $Q_y > 0$. Let (s_x), t_B be given by (3.8). Using (PI) and (PEI) we have $s_x = \lambda_1 Q_x$ and $s_y = \lambda_j Q_y$. By (PN2) we have that λ_1 and λ_j are strictly positive, and so s_x and s_y are also strictly positive. Since $\mathbb{P}(x, y \in \mathcal{X}) = 0$, we have $\mathbb{P}(\{x, y\} \cap \mathcal{X} \neq \emptyset) = Q_x + Q_y$. Set $F = \{\mathcal{X} \subset H_1\}$. Then since $\mathbb{P}(\mathcal{X} \not\subset H_1, x \in \mathcal{X}) = 0$,

$$\mathbb{P}(F|\{x, y\} \cap \mathcal{X} \neq \emptyset) = \frac{\mathbb{P}(\mathcal{X} \subset H_1, x \in \mathcal{X})}{Q_x + Q_y} = \frac{Q_x}{Q_x + Q_y}.$$

Also, since (up to null sets) $S = y$ implies $\mathcal{X} \cap H_1 = \emptyset$, and $S = x$ implies $\mathcal{X} \subset H_1$,

$$P_E(F|S \in \{x, y\}) = \frac{P_E(\mathcal{X} \subset H_1, x \in \mathcal{X})}{s_x + s_y} = \frac{s_x}{s_x + s_y}.$$

Using (SPEI2) we obtain $s_y/s_x = Q_y/Q_x$, and therefore $\lambda_1 = \lambda_j$. □

A.2 Two undesirable properties of HTM convergence

Example A.2. *HTM conditioning and convergence.* Given two probability measures P and Q, the total variation norm is defined by

$$||P - Q||_{TV} = \sup_F |P(F) - Q(F)|.$$

It is easy to verify that conditional probability is stable under convergence in this norm: if $P(F) > 0$ and $||P_n - P||_{TV} \to 0$ then $\lim_n P_n(G|F) = P(G|F)$.

HTM conditioning is not stable in this way. Consider the GSB problem with $q_2 = q_3 = \frac{1}{2}$, on the space $\Omega = \Omega_O \times \Omega_A = \{2,3\} \times \{1,2,3\}$. Let $n \geq 3$ and P_n be given by $P_n((2,1) = \frac{1}{2} - n^{-1}$, $P_n((2,2)) = n^{-1}$, $P_n(2,3)) = 0$, and $P_n(3,j)) = \frac{1}{6}$ for $j = 1,2,3$. We define $X(i,j) = i$, $S(i,j) = j$. Let P_∞ be the pointwise limit of the P_n. It is clear that P_n converge to P_∞ in $||\cdot||_{TV}$. However as $\mathcal{S}(\{S = 2\}, P_n) = \{2,3\}$ and $\mathcal{S}(\{S = 2\}, P_\infty) = \{3\}$ we have by Lemma 4.4(a) that

$$P_n(X = 2||S = 2) = P_n(X = 2) = \tfrac{1}{2}, \tag{A.4}$$

$$P_\infty(X = 2||S = 2) = P_\infty(X = 2|X = 3) = 0, \tag{A.5}$$

and thus $P_n(X = 2||S = 2)$ fails to converge to $P_\infty(X = 2||S = 2)$.

Let X, S be integer valued random variables on a probability space $(\Omega, \mathcal{F}, P)$, with

$$P(X = i, S = j) = p_{ij} \text{ for } i, j \in \mathbb{N}.$$

Then the conditional probability $P(X = j|S = i)$ can be calculated purely in terms of the numbers p_{ij}, and it follows that if X' and S' are random variables on another probability space $(\Omega', \mathcal{F}', P')$, with $P'(X' = i, S' = j) = P(X = i, S = j)$ for all (i,j), then $P'(X' = j|S' = i) = P(X = j|S = i)$ for all (i,j).

Example A.3. *HTM conditioning and probability spaces.* We can adapt Example A.2 to show that in general HTM conditioning does not satisfy the property above. Consider the space $\Omega' = \{1,2,3\} \times \{1,2\}$, set $X'(1,j) = X'(2,j) = 2$ and $X'(3,j) = 2$ for $j = 1,2,3$, and set $S'(i,j) = j$. Fix $n \geq 3$, and set $P_n(1,2) = n^{-1}$, $P'_n(2,2) = \frac{1}{2} - n^{-1}$, and $P'_n(3,j) = \frac{1}{6}$ for $j = 1,2,3$, with the remaining values of $P'_n(i,j)$ being 0. Then it is easy to check that with X, S as in Example A.2

$$P'_n(X' = i, S' = j) = P_n(X = i, S = j) \text{ for } 1 \leq i, j \leq 3.$$

However since $\mathcal{S}(\{S' = 2\}, P'_n) = \{1,3\}$, we have

$$P'_n(X = 2||S = 2) = P'_n(X = 2|\{1,3\} \times \Omega_A) = \frac{1}{n}, \tag{A.6}$$

which is different from the value in (A.4).

In mathematical physics random variables such as S, X are called *observables*. In standard applications of probability, to make predictions one needs to know the joint probability law of the observables, but further knowledge is unnecessary. This example shows that in general to calculate the HTM conditional probability one needs not just the joint law of the observables, but a much more intimate knowledge of the whole probability space: one needs in effect access to 'hidden variables'.

A.3 Centered and uncentered worlds

As mentioned in the introduction, Elga's main reason for introducing the SB problem was to explore how probabilities are updated by a centered observer. Let us first recall the classical situation.

We have a probability space $(\Omega_O, \mathcal{F}_O, \mathbb{P}_O)$. Let $T = \{1, \dots, M\}$ and suppose that the information available to a (classical) observer at time $t \in T$ is given by the σ-field $\mathcal{F}^O_t$. We can assume that there is a stochastic process $Z = (Z_t, t \in T)$ with values in a space $\mathcal{Z}$ such that $\mathcal{F}^O_t = \sigma(Z_1, \dots, Z_t)$. As the CO forgets nothing, $\mathcal{F}^O_s \subset \mathcal{F}^O_t$ if $s \le t$. If $F \in \mathcal{F}_O$ and

$$M_t = \mathbb{E}_O(1_F | \mathcal{F}^O_t) \tag{A.7}$$

then (M_t) is a martingale and M_t gives the CO's probability that F occurs, given the information available to it at time t.

A "centered event" is an event such as "Z will be in the set D tomorrow". It is clear that in general these events are not in $\mathcal{F}_O$, and that their probabilities may update in a different way from that of events in $\mathcal{F}_O$.

We can handle centered events and observers using the formalism of this paper, by taking X^O to be deterministic with $\mathbb{P}_O(X^O = M) = 1$, and assuming that the AO on day t knows its location. So we set $\Omega_A = T = \{1, \dots, M\}$, and define the probability space $(\Omega, \mathcal{F}, \mathcal{F}_W, \mathbb{P})$ and the random variable S exactly as in Section 2. Using (2.2) we can write any event $F \in \mathcal{F}$ in the form

$$F = \bigcup_{k=1}^{M} \{S = k\} \cap F_k, \tag{A.8}$$

where $F_k \in \mathcal{F}_W$ for each k.

As in Section 2 we need to choose a probability P_A on $(\Omega, \mathcal{F})$. We still assume (PN), but as the AO knows its location (PI) is no longer relevant. The reasons for accepting (PEI) remain as strong as before, and as $X = M$ it implies that for $F \in \mathcal{F}_W$

$$P_A(F | S = t) = \mathbb{P}(F) \text{ whenever } P_A(S = t) > 0. \tag{A.9}$$

Writing $\mu(\{t\}) = P_A(S = t)$ it follows that if F is given by (A.8) then

$$P_A(F) = \sum_{t=1}^{M} \mu(\{t\}) P_A(F_t). \tag{A.10}$$

As T is finite it is natural to take $\mu(\{t\}) = 1/M$, in which case $P_A = P_L = P_E$. However, in this setting we are only interested in the measures $P_A(\cdot | S = t)$, so that the only condition we need impose on μ is that $\mu(\{t\}) > 0$ for all t.

We assume that at time t the AO has available the information in $\mathcal{F}^O_t$, and as it also knows that $\{S = t\}$, its information is described by the σ-field

$$\mathcal{G}_t = \Big\{ (G_1 \cap \{S = t\}) \cup (G_2 \cap \{S \neq t\}), G_1, G_2 \in \mathcal{F}^O_t \Big\}.$$

Write $P_t(\cdot) = P_E(\cdot | S = t)$, and E_t for expectation with respect to P_t.

Lemma A.4. *Suppose that F is given by* (A.8). *Then*

$$E_t(1_F|\mathcal{G}_t) = \mathbb{E}(1_{F_t}|\mathcal{F}_t^O), \ \ P_t \ - a.s.. \tag{A.11}$$

Proof. Write $N_t = \mathbb{E}(1_{F_t}|\mathcal{F}_t^O)$. To prove (A.11) it is sufficient to show that

$$E_t(1_F 1_G) = E_t(1_G N_t) \text{ for all } G \in \mathcal{G}_t. \tag{A.12}$$

Let $G \in \mathcal{G}_t$ and write $G = (G_1 \cap \{S = t\}) \cup (G_2 \cap \{S \neq t\})$. Then $1_F 1_G 1_{(S=t)} = 1_{F_t} 1_{G_1} 1_{(S=t)}$, and thus

$$E_t(1_F 1_G) = \mu(\{t\})^{-1} E_E(1_{F_t} 1_{G_1} 1_{(S=t)}) = \mathbb{E}(1_{F_t} 1_{G_1}) = \mathbb{E}(N_t 1_{G_1}).$$

Similarly $E_t(1_G N_t) = \mathbb{E}(N_t 1_{G_1})$, proving (A.12). □

Remark A.5. (a) The general form of a $\mathcal{G}_t$ measurable random variable is given by $Y_1 + Y_2 1_{(S \neq t)}$, but as conditional expectation is only defined up to sets of probability zero, and $P_t(S \neq t) = 0$, we can omit the Y_2 term.
(b) The Lemma shows that in general updating of the conditional probability of a centered event by a centered observer involves updating not only the conditioning σ-field $\mathcal{F}_t^O$ but also the conditioned event F_t. If however $F \in \mathcal{F}_W$ is an objective event then we can take $F_k = F$ for all k in (A.8) and then $E_t(1_F|\mathcal{G}_t)$ is given by the martingale M defined by (A.7).

Acknowledgment. I wish to thank Richard A. Johns for telling me about the Doomsday argument, Don Page for pointing me to the paper [23], and Chris Burdzy, Jason Swanson and Zoran Vondraček for helpful conversations on this topic.

References

[1] F. C. Adams. The degree of fine-tuning in our universe ? and others. *Physics Reports* **807**, 1–111 (2019).

[2] S. Armstrong. Anthropic decision theory for self-locating beliefs, v3. arXiv:1110.6437 (2017).

[3] F. Arntzenius. Reflections on Sleeping Beauty. *Analysis*, **62**,No. 1, 53–62. (2002).

[4] S. M. Barr and A. Khan. Anthropic tuning of the weak scale and of m_u/m_d in two-Higgs-doublet models. *Phys. Rev. D* **76**, 045002 (2007).

[5] P. Bartha, C. Hitchcock. No One Knows the Date or the Hour: An Unorthodox Application of Rev. Bayes's Theorem. *Philosophy of Science. Suppl.* **66**: S229–S353. 1999.

[6] D. Bradley and H. Leitgeb. When Betting Odds and Credences Come Apart: More Worries for Dutch Book Arguments. *Analysis* **66**, No. 2, 119–127 (2006).

[7] R. Briggs. Putting a value on beauty. In T. S. Gendler and J. Hawthorne (Eds.), Oxford studies in epistemology (Vol. 3, pp. 3-34). Oxford: Oxford University Press, 2010.

[8] N. Bostrom. *Anthropic bias: observation selection effects in science and philosophy.* Routledge, 2002.

[9] N. Bostrom. Sleeping beauty and self-location: A hybrid model. *Synthese* **157**, 59–78 (2007)

[10] B. Carter. Large Number Coincidences and the Anthropic Principle in Cosmology. *Confrontation of Cosmological Theories with Observational Data.* M. S. Longair. Dordrecht, Reidel: 291–298 (1974).

[11] J. Cisewski, J. B. Kadane, M. J. Schervish, T. Seidenfeld, R. Stern. Sleeping Beauty's Credences. *Philosophy of Science*, **83**, No. 3, 324–347 (2016).

[12] V. Conitzer. A Dutch book against sleeping beauties who are evidential decision theorists. *Synthese*, **192**, No. 9, 2887–2899 (2015).

[13] D. Dieks. Reasoning about the Future: Doom and Beauty. *Synthese* **156**, No. 3, 427–439 (2007).

[14] K. Draper and J. Pust. Diachronic Dutch Books and Sleeping Beauty. *Synthese* **164** No. 2, 281–287 (2008).

[15] R. Durrett. *Probability: Theory and Examples.* 2nd ed. Duxberry 1996.

[16] A. Elga. Self-locating belief and the Sleeping Beauty problem. *Analysis* **60**, 143–147 (2000).

[17] W. Feller. *An Introduction to probability theory and its applications, Vol. 1, 3rd edition.* Wiley, 1968.

[18] S. Friederich. Fine-Tuning. The Stanford Encyclopedia of Philosophy. Ed. E. Zalta, U. Nodelman. Available at: `https://plato.stanford.edu/archives/win2023/entries/fine-tuning/`

[19] J.R. Gott. Implications of the Copernican principle for our future prospects. *Nature* **363** 315–319 (1993).

[20] B. Groisman. The End of Sleeping Beauty's Nightmare *British J. Phil. of Science*, **59**, No. 3, 409–416 (2008).

[21] U. Grömping. The Sleeping Beauty Problem Demystified. Preprint 2019. Available at: `https://www1.beuth-hochschule.de/FB_II/reports/Report-2019-002.pdf`

[22] J. Y. Halpern. Sleeping Beauty reconsidered: Conditioning and reflection in asynchronous systems. In T. Gendler and J. Hawthorne (Eds.), Oxford studies in epistemology (Vol. 1, pp. 111–142). Oxford: Oxford University Press (2005).

[23] J.B. Hartle, M. Srednicki. Are we typical? *Phys. Rev. D* **75** 123523 (2007).

[24] P. Hawley. Inertia, optimism and Beauty. *Noûs* **1** no. 1 85–103 (2013).

[25] C. Hitchcock. Beauty and the Bets. *Synthese*, **139**, No. 3, 405–420 (2004).

[26] K. Itô. *Introduction to probability theory.* Cambridge Univ. Press (1984).

[27] K. Karlander, L. Spectre. Sleeping Beauty meets Monday. *Synthese* **174** 397–412 (2010).

[28] J. G. Kemeny. Fair Bets and Inductive Probabilities *J. Symbolic Logic* **20**, 263–273 (1955).

[29] B. Kierland, B. Monton. Minimizing inaccuracy for self-locating belief. *Philosophy and Phenomenological Research* **70**(2), 384–395 (2005).

[30] J. Leslie. Risking the world's end. *Interchange* **21**, 49–58 (1989).

[31] J. Leslie. *The end of the world: The ethics and science of human extinction.* London: Routledge 1996.

[32] D. Lewis. A subjectivist guide to objective chance. In R. C. Jeffrey, ed. *Studies in Inductive logic and Probability*, vol II Berkeley, U. of California Press. (1980).

[33] C.J.G. Meacham. Sleeping beauty and the dynamics of de se beliefs. *Philosophical Studies*, **138**, 245–269 (2008).

[34] D. N. Page. Is our Universe likely to decay within 20 billion years? *Phys. Rev. D* **78**, 063535 (2008).

[35] J. Pittard. When Beauties Disagree: Why Halfers Should Affirm Robust Perspectivalism. In Tamar Szabó Gendler, and John Hawthorne (eds), Oxford Studies in Epistemology Volume 5, (2015).

[36] M. Piccione, A. Rubenstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* **20** (3), 3–24 (1997).

[37] M. Rees. *Just six numbers: The deep forces that shape the universe.* Weidenfeld and Nicholson, London 1999.

[38] J. S. Rosenthal. A mathematical analysis of the Sleeping Beauty problem. *Math. Intelligencer.* **31**, 32–37 (2009).

[39] D. Sorini, J. A. Peacock, L. Lombriser. The impact of the cosmological constant on past and future star formation. *Monthly Notices of the Royal Astron. Soc.* **535** (2) 1449–1474 (2024).

[40] M. G. Titelbaum. The relevance of self-locating beliefs. *Philos. Rev.* **177**, 555–605 (2008).

[41] M. G. Titelbaum. An Embarrassment for Double-Halfers. *Thought: A Journal of Philosophy* **1**, 146–151 (2012).

[42] M. G. Titelbaum. Ten reasons to care about the Sleeping Beauty problem. *Philosophy Compass* **8**, 1003–1017 (2013).

[43] M. G. Titelbaum. *Quitting certainties.* Oxford: Oxford University Press. (2013).

[44] P. van Inwagen. *Metaphysics.* Oxford Univ. Press. 1993.

[45] S. Weintraub. Sleeping Beauty: A simple solution. *Analysis*, **64**, 8–10. (2004).

[46] P. Winkler. The sleeping beauty controversy. *The American Mathematical Monthly*, **124**, 579–587 (2017).

[47] M. Yamada. Beauty, odds, and credence. *Philos. Stud.* **176**, 1247–1261 (2019).

[48] A. Zuboff. One self: The logic of experience. *Inquiry - An Interdisciplinary Journal of Philosophy* **33** 39–68 (1990).