Restricting experiments to only being in certain states

or

How should Sleeping Beauty condition on being awake?

Giulio Katis (giulio.katis@gmail.com)
13-Nov-14

1. Introductory remarks

A lot has been written about the Sleeping Beauty (SB) problem[1], some would say too much. So why another paper? I have not written this paper out of an interest in engaging in the SB debate itself, but rather because I feel the constructs associated with resolving it are interesting in their own right. From the papers and blogs I have read, it appears that the clarification of a key concept, namely how to restrict an experiment/system to only being in certain states, would facilitate a better understanding of the underlying constructs involved in the debate. We give a definition of this simple concept in this paper and describe its relevance to the SB problem (the SB experiment seems to be the simplest example that is required to illustrate the concept).

Lack of clarification of the above concept is not the only source of confusion in the SB debate. The contention between 'halfers' and 'thirders' over the SB problem also arises for another reason, namely confusion over which exact problem or which frame of reference is being considered: the perspective of the subject SB 'waking up' inside the experiment, or the perspective of an external observer who chooses to randomly inspect the state of the experiment.

For a general experiment X (viewed as a Markov Chain), there are two natural probability measures on the set of states of the experiment which could compete for the notion of "the probability that the experiment currently is in a subset S of the states".

[1] I am assuming the reader is already familiar with the Sleeping Beauty problem, popularized by Adam Elga. In particular, I assume the reader is familiar with the terms 'thirders' and 'halfers'. For details, see the Wikipedia article. To remind the reader of the problem, SB is the subject of the following experiment: at the start of the experiment there will be a fair coin toss; tails determines she will be awoken on both Monday and Tuesday; heads will determine she will be awoken on Monday but not Tuesday; that whenever she is awoken she will be asked to calculate the probability the toss was Heads, and then after a brief period put back to sleep; and that her memories of being awoken will be erased once put back to sleep. We assume upon awakening SB is aware of the complete experimental set-up. 'Thirders' argue that upon being awoken SB should calculate the probability the toss was Heads as ⅓, while 'halfers' argue it should be ½. Of course, not all thirders (or halfers) argue for the same reasons.

We argue that one of these is appropriate for the perspective of an *external* observer. We denote it by Ξ_X, and it satisfies the property that $\Xi_X(S) < \Xi_X(T)$ if and only if the expected number of times the experiment is in S is less than the expected number of times the experiment is in T (where the expectation is taken over the probability space of behaviours of the experiment). We call Ξ_X the *counting* measure. This is the measure implicitly used by most 'thirders'.

We claim that there is another natural probability measure that is appropriate for the perspective of a *subject inside* the experiment. We denote it Π_X, and it satisfies the property that $\Pi_X(S) < \Pi_X(T)$ if and only if the expected proportion (or percentage) of times the experiment is in S is less than the expected proportion of times the experiment is in T (where again the expectation is taken over the probability space of behaviours of the experiment). We call Π_X the *proportion* measure. I will argue that the proportion measure Π is the appropriate measure for SB to use, and give examples where Ξ would instead be the appropriate measure to use.

As both Π and Ξ are probability measures on the set of states of an experiment, each admit their own conditioning (which I will call *classical* conditioning). But there is another type of 'conditioning', which I refer to as "conditioning on *only being* in a subset C of the states of the experiment" or, to avoid confusion with classical probability conditioning, "*restricting* to only being in a subset C of the states of the experiment".

This is best understood as an operation applied to an experiment, and not a construct that is inherently specific to any probability measure on the set states of the experiment: given an experiment X, viewed as a Markov Chain with specified begin and end states, we can construct another experiment $(X|C)$ which only has states C. We refer to $(X|C)$ as the experiment X *restricted to only being in the states* C. Then, to define what is meant by evaluating a probability measure, say Π, on a subset of states S of an experiment restricted to only being in the subset of states C, we first construct the experiment $(X|C)$, and then calculate $\Pi_{(X|C)}(S \cap C)$. We argue that this is the type of 'conditioning' SB should use, where S is the subset of states corresponding to Heads having been tossed, and C is the set of states where SB is awake. In this case it will turn out $\Pi_{(X|C)}(S \cap C) = \frac{1}{2}$. So, yes, I am a 'halfer' when it comes to the standard SB experiment.

To clarify the difference between this type of conditioning and the classical one, we give another example of an experiment: SB is now awoken on both Monday and Tuesday regardless of the toss, and is given orange juice when awoken except on a Tuesday after Heads was tossed, in which case she is given apple juice. Formally, this experiment is modelled by an equivalent Markov Chain to the standard SB problem (awoken$\sim$awoken and served orange juice, asleep $\sim$ awoken and served apple juice). But in this case, if awoken and served OJ, to determine the probability that Heads was tossed we argue she should use classical conditioning and calculate $\Pi(S|C)$, which turns out to be ⅓.

The distinction between this situation and the standard SB problem is that here the subject SB could potentially have been served apple juice and be asking herself about the probability of the toss; while in the standard example there is no sense in which she could be asleep and be asking herself about the probability of the toss, so she needs to *restrict the experiment to only being* in states where she is awake. Another way to view this is in terms of information: when SB is evaluating the information she has at hand, she needs to take into account the possible alternatives; and, in the standard SB example, the state of being Asleep should be removed since it is not an alternative state she could find herself in - but this should be done in a way that allows behaviours passing through that state to still be accounted for.

The mathematics presented here codifies this distinction, that is, the difference between (a) 'conditioning a probability measure on an experiment being in a state satisfying property P as opposed to $\neg P$' (which is what is achieved when conditioning with respect to a probability measure), and (b) 'calculating a probability measure when restricting an experiment to only being in states satisfying property P'.

A lot of the noise made by thirders can be interpreted in terms of the fact that this operation of "restricting to only being in a subset of states" taken together with the proportion probability measure Π does not satisfy the usual laws of conditioning (e.g. Bayes Theorem). For example, the following identity does *not* hold in general for disjoint subsets C and D:

$$\Pi_{(X|(C\cup D))}(S\cap(C\cup D)) = \Pi_X(C)*\Pi_{(X|C)}(S\cap C)+\Pi_X(D)*\Pi_{(X|D)}(S\cap D).$$

But given the restriction operation is a construct on experiments, there is no reason for us to expect it to.

In fact, in the context of applying this restriction operation on experiments and then calculating the proportion probability Π, a form of 'interference' can arise for the series *composite* of two experiments (i.e. the experiment constructed as 'first do experiment 1, then do experiment 2').

For example, as mentioned, we will see that our definition of $\Pi_{(X|C)}(S \cap C)$ gives an answer of ½ for the SB problem (i.e. restricting to the situation where SB is awake, the probability that Heads was tossed is a half). But consider a 2-week version of the SB problem, which we denote $SB^2 = SB * SB$: that is, instead of finishing the experiment after the Tuesday, we keep SB asleep, and on the next Sunday flip a coin again, and repeat the protocol. In this 2-week composite experiment, restricting to SB only ever being awake, what is the probability that the last toss was Heads? We might expect to be able to reason as follows. Restricting to SB being awake, there's an even chance she is in the first week or the second week; in each of these weeks, restricting to SB being awake, there is an even chance Heads or Tails was the result of the last toss; therefore, restricting to SB being awake, there is an even chance the result of the last toss was Heads or Tails.

However, we will show that restricting to SB only being awake, the probability that the last toss was Heads is $\frac{5}{12}$ (and the probability it was Tails is $\frac{7}{12}$). We will see that there is a type of interference involved in calculating the probabilities Π between the two sub-experiments (week 1 and week 2) that occurs because of the *renormalization along individual behaviours* involved when calculating the proportion of times a composite behaviour passes through a subset S of states. If we were interested only in counting the number of times a composite behaviour passes through a subset of states, as opposed to the percentage of times, this phenomenon would not arise.

In a purely formal manner (and, of course, not in a deeper physical sense) this 'non-locality', the failure of classical conditioning laws when restricting to sub-experiments, and the importance of considering complete behaviours of an experiment when calculating probabilities, reminds me of the interference of quantum mechanical experiments (as, say, described by Feynman in the gem of a book QED).

What is the connection between the measures Ξ and Π ?

It will be clear that they are identical when all behaviours of the experiment have the same length (i.e. pass through the same number of states).

I conjecture there is another more interesting relationship which involves the 'only being in' restriction operation, and which connects the counting measure (representing the perspective of an external observer) with the proportion measure (representing the perspective of an internal observer, i.e. a subject of an experiment).

Suppose X is an experiment. Let X^n be the experiment constructed by repeating X in succession n times (i.e. the n-fold series composite of X with itself). So the set of states of X^n is the disjoint union of X with itself n times. If S is a subset of the states of X then let S^n be the subset of states of X^n comprising the union of each copy of S in each of the n instances of X.

Conjecture: $\Xi_X(S|C) = \lim_{n\to\infty} \Pi_{(X^n|C^n)}(S^n)$.

This is just a conjecture, but I have verified it numerically in a few examples, including the case of SB. That is, in the case of the SB experiment the ‘thirders’ position can be obtained from the ‘halfers’ position by considering an experiment in which the SB scenario is indefinitely repeated in succession (the limit of the series composite of the SB experiment with itself), an experiment which I denote as SB^{∞}. We could describe this limiting case of a composite experiment as follows. You wake up in a room with a white glow. A voice speaks to you. “You have died, and you are now in eternity. Since you spent so much of your life thinking about probability puzzles, I have decided you will spend eternity mostly asleep and only be awoken in the following situations. Every Sunday I will toss a fair coin. If the toss is tails, I will wake you only on Monday and on Tuesday that week. If the toss is heads, I will only wake you on Monday that week. When you are awoken, I will say exactly the same words to you, namely what I am saying now. Shortly after I have finished speaking to you, I will put you back to sleep and erase the memory of your waking time.” The voice stops. Despite your sins, you can’t help yourself, and in the few moments you have before being put back to sleep you try to work out the probability that the last toss was heads. What do you decide it is?

To put this paper in some context, the following provides some background on my interest in this problem and some references.

My interest in this problem arose from reading an initial blog of Bob Walters. Since then he has written several posts. Before finalizing this version of the paper I noticed Bob has put one last post containing calculations that seem related to the restriction operation I describe here, but it is hard to say given the specificity of his remarks. I largely sympathize with the comments Bob has made on these blogs (perhaps not surprising as he was my PhD supervisor and scientific mentor of many years ago), but he seems to be taking the position that all that is required to make sense of the confusion around the problem is a specific calculation. In contrast to this, I believe to address the confusion certain concepts need to be identified and *explicitly* and *generally* defined.

A friend of mine Sean Carmody (a.k.a. the Stubborn Mule) then also wrote a couple blogs on the topic, including this one, which contains a link to a previous version of this paper. As a result of the exchanges in the comments on this blog, I realized that the operation of “restricting an experiment to a subset of its states” was a fundamental construction, and so I have fundamentally rewritten the paper with that concept at its centre.

One of the commenters on Sean’s blog gave a couple of interesting references. One was to a paper by D Manley. This was different to most papers I have seen on this topic in that it gave some explicit general definitions. In particular, it identified what I call the counting and the proportion probability measures. The author did not explicitly identify the construct of “restricting an experiment to only being in a subset of its states”, but rather identified the quantity $\Pi_{(X|C)}(S)$ and referred to it as the ‘invariance strategy’. Though a lot of thought went into the work, my main problem with this paper was the general approach: it put these various definitions through a battery of examples to assess which was ‘best’ or most ‘robust’ in some sense (or rather to identify ones he considered ‘worst’ in that they gave nonsensical results in more of the examples he considered).

As opposed to this, I am taking the approach that the different definitions/constructs are measuring different things, and hence the applicability of each will depend on the situation. The point of view that ‘there can only be one’ (definition that should apply in all contexts) seems to be a point of view taken implicitly by many ‘thirders’ and ‘halfers’, particularly participants in some of the blogs I have read (which tends to give some of these exchanges a ‘religious’ or dogmatic tone).

I note that although the counting measure Ξ can be thought of as being more ‘robust’ than the proportion measure Π, this ‘robustness’ can be seen as a consequence of it being ‘insensitive’ to the restriction operation on experiments. This observation is made precise in Section 4.

Another paper referenced by the same Stubborn Mule commenter was one by J. Pittard. In it there is an example of a multiple person Sleeping Beauty problem. I found this example very interesting: to me it makes clear that, to answer a problem like this, one really does need to specify a reference frame. I address this problem as an example, and specifically calculate probabilities from one subject's reference frame as well as from the joint reference frame of two subjects.

The following is a summary of the paper. In Section 2 a formal model of an experiment is given along with a number of examples. In Section 3, the counting and proportion probability measures are given. In Section 4, the operation of restricting an experiment to only being in a subset of states is defined, and applied to the standard Sleeping Beauty Problem as well as to Pittard's example of multiple Beauties. In Section 5, the definition of the series composite of experiments is given, and the multi-week Sleeping Beauty problem is addressed. The conclusion contains some speculative remarks.

2. A formal model of an experiment

There are a variety of ways experiments could be formally modelled. I have tried to pick the simplest one (i.e. a particular type of finite state Markov Chain) that will allow me to express the constructs and ideas mentioned in the introduction. However, for reasons I allude to in Section 4, I think a better theory develops if directed graphs that allow multiple transitions between states (and where the set of edges out of each state is given a probability measure) are considered.

We model an experiment X as a finite state Markov Chain. It comprises the following:

- a finite set X_S, called the underlying set of states of the experiment
- a specified state B which we call the Begin state of the experiment, and a specified state E which we call call the End state of the experiment; all the other states are called the *intermediary* states of the experiment. We denote the set of intermediary states by X_I.
- a transition function $\pi = \pi_X : X_S \times X_S \to \Re^{\geq 0}$ (a matrix of non-negative real numbers) that satisfies the condition, for all states s, $\sum_{t \in X_S} \pi(s,t) = 1$. We say that conditional on being in state s there is probability $\pi(s,t)$ of a transition from s to t occuring.

We define a *path* to be a sequence of states $s_0, ..., s_n$ with $n > 0$ where for all $i,\ 0 \leq i < n,$ we have $\pi(s_i, s_{i+1}) > 0$. In this case we write $s_0 \to ... \to s_n$ and we say this path starts at s_0 and ends at s_n.

The function π satisfies the following conditions.

- For all states s, $\pi(s,B)=0$; that is, there are no transitions into the Begin state.
- $\pi(E,E)=1$; that is, the only transition out of the End state is to itself.
- For every intermediary state s, there exists a path starting at B and ending at s (i.e. s is 'reachable' from the Begin state).
- For every intermediary state s, there exists a path starting at s and ending at E (i.e. the End state is 'reachable' from s).

Define the set $Beh(X)$ of *behaviours* of an experiment X to be the set of paths starting at the Begin state and ending at the End state. Let p be a $B=s_0 \rightarrow \ldots \rightarrow s_n = E$ be a behaviour. Define the probability of p occurring to be the product of the probabilities of each transition in the path occurring, i.e. if $n>1$ then $\pi(p)=\pi(B,s_1) * \ldots * \pi(s_{n-1},E)$.

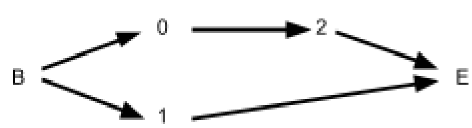
Proposition: This defines a probability measure on the set of behaviours. That is, $1=\sum_p \pi(p)$, where the sum is over all behaviours of the experiment.

Note: for a given experiment, there is at most one behaviour that does not visit an intermediate state. This behaviour exists if $\pi(B,E)>0$.

Below we will give some examples of experiments, depicting them with graphs (i.e. state transition diagrams). Usually we will want to refer to subsets of the intermediary states of the experiment, and we will typically denote them with labels on the vertices of the graph.

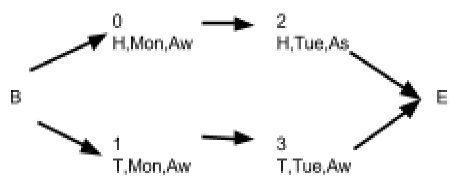
We note that most real world experiments contain effectively infinite variety and detail, and can be modelled in many ways: the specific choice of a formal model is determined by which salient features we wish to reflect. Also, two different real world experiments may be modelled by the same formal model - this would mean the two experiments can be thought of as formally similar *relative* to the features we are trying to capture *as a finite state Markov Chain*.

Example A



In this experiment, there are three intermediary states. The two transitions out of the Begin state each have a transition probability of $\frac{1}{2}$. The experiment has two behaviours: one behaviour passes through states 0 and 2; and the other behaviour passes through the state 1. Both behaviours have a probability $\frac{1}{2}$ of occurring. What real world example could this model? Imagine an experiment where initially a fair coin was tossed: if Tails, run the experiment for two days; if Heads, run the experiment for just one day.

Example B

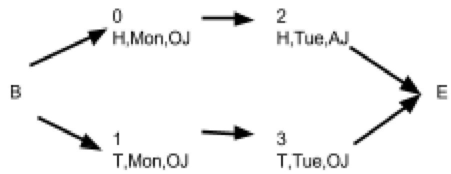


This is a model of the standard Sleeping Beauty experiment. We denote it by SB. The two transitions out of the Begin state each have a transition probability of $\frac{1}{2}$. There are 4 intermediary states, and two equally likely behaviours. The letter H denotes the subset of states where Heads had been tossed; T denotes the states where Tails had been tossed; Mon denotes the states with the property it is Monday; Tue denotes the states with the property it is Tuesday; Aw denotes the states where SB has been awakened; and As denotes the state where SB was left asleep.

Example C

This is a mild modification of the Sleeping Beauty experiment obtained by removing state 2. Of course, it is formally equivalent to Example A. I will give two different real world interpretations of this model. One is that it is the standard SB experiment but viewed from the perspective of SB's waking experience (let us say that we have restricted Example B to SB being awake). Another real world experiment which could be modelled by this example is a version of the SB experiment run in an institute under cost pressures: if Heads is tossed, then the experiment ends on the Monday so the bed can be made free for another experiment on the Tuesday.

Example D

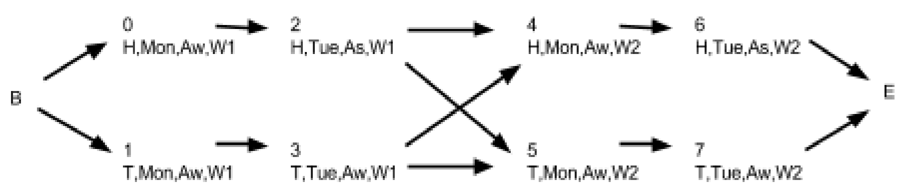


This is the modification of the Sleeping Beauty experiment that was mentioned in the introduction: SB is awoken on each day regardless of the toss, and she is served Orange Juice when awoken except on Tuesday when Heads had been tossed - in this case she is served Apple Juice. Of course, it is formally equivalent to Example B.

Note:

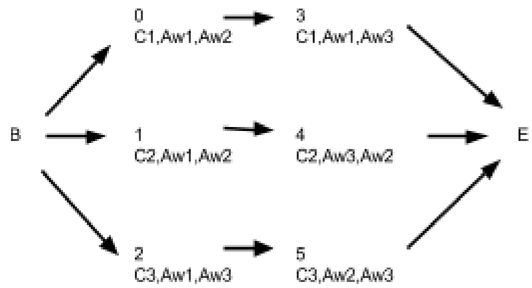
- The distinction between this experiment and the standard SB experiment is not reflected in how we have formally modelled the experiments. That is, our general definition of an experiment is not rich enough to make such a distinction. Instead the difference will be expressed by the choice of method used to calculate conditioning and probabilities. We will see this in the next Section.

Example E



This is a model of the two week Sleeping Beauty problem SB^2 mentioned in the introduction. We have added the subsets of states corresponding to week 1 (W1) and week 2 (W2). It has eight intermediary states. The Begin state, state 2 and state 3 each have a pair of outgoing transitions; these each have transition probability ½. There are four equally probable behaviours. In a later section we will show how this can be constructed as the series composite of Example B with itself.

Example F



We will use this Markov Chain to model an experiment which is a variant of one proposed in a paper by J. Pittard; it is a modification of the standard SB problem but deals with multiple subjects, and poses what I consider to be quite a curly problem.

In this experiment there are three Beauties - $B1$, $B2$ and $B3$. All three are put to sleep. One of the three is then randomly selected to be the "chosen one". There will be two waking sessions in this experiment. After each waking session, those that have been awakened will be put back to sleep and the memories of the waking session will be erased. The chosen one is awoken on both sessions, while the others are awoken only once (in numerical order). That is, if $B1$ is the chosen one, then on the first session $B1$ and $B2$ are awoken, and in the second session $B1$ and $B3$ are awoken; if $B2$ is the chosen one, then $B2$ and $B1$ are awoken in the first session and $B2$ and B3 are awoken in the second session; and if $B3$ is the chosen one, then $B3$ and $B1$ are awoken in the first session and $B3$ and $B2$ are awoken in the second session. At each session, the two awake Beauties are brought together and each is asked to estimate the likelihood that they are the chosen one. We assume that the subjects know the complete experimental set-up.

How should they answer? Suppose you are $B1$, and you have just been awoken, but have no other information. You would estimate the probability of being the chosen one as ⅓. You know that you are going to be brought together with either $B2$ or $B3$, but that isn't giving you any information beyond the experimental set-up. Suppose you are then brought together with $B2$. On the one hand, you could reason as follows: you already knew you were going to see either $B2$ or $B3$; seeing $B2$ isn't giving you more information about whether you are the chosen one than if you saw $B3$ (since you were going to see one of them); therefore you should stick to your original estimate of ⅓. On the other hand, you now know that the chosen one is either yourself or $B2$, and as you don't have any information that $B2$ doesn't, you should both conclude that there is an even chance as to who is the chosen one - that is, you should estimate the probability of being the chosen one as ½.

In the experiment as I have modelled it, there are eight intermediary states, and the three transitions out of state B each have a transition probability of ⅓. The subsets $C1$, $C2$ and $C3$ correspond to whether respectively $B1$, $B2$ or $B3$ is the chosen one. The subsets $Aw1$, $Aw2$ and $Aw3$ correspond to states where the respective Beauty is awake.

In Section 4 we will come back to this example and address the question.

3. The *counting* and *proportion* probability measures on the intermediary states of an experiment

In this section we define the two probability measures on the set of intermediary states of an experiment which compete for the notion of "the probability that an experiment is in a subset of its states".

The notion of an experiment we have given comes equipped with a probability measure on the set of its behaviours, but not on the set of its states. To proceed we need to give our experiments the extra structure of a measure (not necessarily a probability measure) on its set of states that expresses the extent of 'being' one state has versus another. The notions of probability that an experiment is in a subset of states arise from an interplay between this 'being' measure on the states and the probability measure on the behaviours.

For the purposes of this paper, we will simply take this 'being' measure to be the one that associates to a subset S of states the number of elements $n(S)$ in that subset. (So all states are thought of as having the same degree of 'being'.) But the definitions below could be modified to accommodate other measures.

Definition: The counting probability measure Ξ

Given an experiment X and a subset $S \subseteq X_I$ of its *intermediary* states, define

$$\Xi_X(S) = \frac{\sum_p \pi(p) * n(S,p)}{\sum_p \pi(p) * n(p)}$$, where

- the two sums are over all behaviours p of X
- $n(p)$ denotes the total number of times p visits an intermediary state
- $n(S,p)$ denotes the number of times p visits a state in S
- $\pi(p)$ is the probability of behaviour p occurring

This is a genuine probability measure on the set X_I of intermediary states. So we can, given another subset C of intermediary states, define the conditional probability $\Xi_X(S|C)$ in the obvious way. It turns out to be equivalent to $\Xi_X(S|C) = \frac{\sum_p \pi(p)*n(S\cap C,p)}{\sum_p \pi(p)*n(S,p)}$.

In what situations would it be reasonable to apply this construction?

One obvious application would be whenever you wanted to make a wager that depended on the expected number of times an experiment was in a particular subset of its intermediary states. Consider a variant of the SB problem (posed by many people as a justification of the thirders position): whenever SB is awoken she is asked to make odds on whether Heads was tossed; she will take bets on these odds; and she will cumulate the profits and losses on these bets over the course of the experiment. To determine the odds of Heads vs not-Heads (i.e. Tails) in order to maximize her cumulative profit, she should calculate the expected number (not proportion) of times the experiment will be in a state where Heads had been tossed, conditional on her being awake. Using the model of this experiment given in Example B of the previous section, she would calculate $\Xi(H|Aw) = \frac{1}{3}$.

There is another context in which the counting measure can be applied. Suppose you turned up to a lab and an experiment was in progress. Suppose all you knew was the set-up of the experiment (expressed in the way we formally modelled an experiment in the previous section). In particular, you did not know when the experiment started. Now, suppose v and w were two distinct intermediary states of the experiment, and you asked whether it was more likely the experiment is in state v or w. To answer this you would want to compare $\sum_p \pi(p) * n(\{v\},p)$ the expected number of times the experiment is in v to $\sum_p \pi(p) * n(\{w\},p)$ the expected number of times the experiment is in w. For instance, if the former is twice the latter then you would conclude it is twice as likely the experiment is in state v as it is in state w. From this it is follows that you would calculate that the probability the experiment is currently in a subset S of the states to be $\Xi_X(S)$.

For example, consider Example C from the previous section. The counting probability $\Xi_X(H)$ that the experiment is in state 0, i.e. that Heads had been tossed, is ⅓ ; and the counting probability $\Xi_X(T)$ that it is in the subset {1,3}, i.e. that Tails had been tossed, is ⅔.

Notice that the counting measure is *not* consistent with the probability measure associated with the set of behaviours of Example C: the behaviour corresponding to Tails having been tossed is completely characterized by the subset of intermediary states {1,3}; and the behaviour corresponding to Heads having been tossed is completely characterized by the subset {0}. The probability of the Tails behaviour occurring is ½, and the probability of the Heads behaviour occurring is ½. In this example, under Ξ_X the probability of being in (the states that characterize) the behaviour in which Tails (or Heads) had been tossed is ⅔ (or ⅓), which is *inconsistent with the probability of the behaviour occurring* (both are ½).

Is this a problem? Not if you are an external observer who has paid a visit to inspect the experiment. I say this because in this example the observer could have turned up and found that the experiment had ended, i.e. the observer could have been in a state outside the reference frame of the experiment, in which case they would have concluded it is more likely that Heads had been tossed. One could argue that the experiment as modelled does not actually capture all the possible states the observer could have found themselves in when coming to inspect the experiment - in particular, it is excluding a 'null' state representing the non-existence of the experiment.

This is why I claim this counting probability measure is appropriate for an external observer, and not for a subject such as SB that is operating within the reference frame of the experiment.

Definition: The proportion probability measure Π

Given an experiment X and a subset S of its intermediary states, define

$\Pi_X(S) = \frac{1}{1-\pi(B,E)} \sum_{p \in Beh(I)} \pi(p) * \frac{n(S,p)}{n(p)}$, where

- the sum is over the set $Beh(I)$ of all behaviours p of X that pass through an intermediary state
- $n(p)$ denotes the total number of times p visits an intermediary state
- $n(S,p)$ denotes the number of times p visits a state in S
- $\pi(p)$ is the probability of behaviour p occurring
- $\pi(B,E)$ is the probability of a transition from the Begin state to the End state, i.e. the probability of a behaviour occurring that does not visit any intermediary states

Proposition: If an experiment X has the property that all its behaviours pass through the same number of states (i.e. have the same length), then $\Xi_X = \Pi_X$.

Π_X is a genuine probability measure. So we can, given another subset C of intermediary states, define the conditional probability $\Pi_X(S|C)$ in the obvious way. That is, $\Pi_X(S|C) = \frac{\Pi_X(S \cap C)}{\Pi_X(C)}$.

For example, consider again Example C from the previous section. The proportion probability $\Pi_X(H)$ that the experiment is in state 0, i.e. that Heads had been tossed, is ½; and the proportion probability $\Pi_X(T)$ that it is in the subset {1,3}, i.e. that Tails had been tossed, is ½ . Unlike the counting probability measure, the proportion probability measure is consistent with the probability measure on the set of behaviours. This is because it involves a renormalization at the level of each behaviour.

In other words, it involves calculating the probability of being in S at the level of each behaviour, and then takes the *expected value* of these probabilities *with respect to the probability measure on the set of behaviours*. That is why I claim the proportion probability measure is appropriate for a subject in the reference frame of the experiment.

Consider Example D from the previous section. Upon being awoken and served OJ, how should the subject SB calculate the probability that the toss was Heads? As she is a subject of the experiment, she should use Π. She wants to condition on her current state being one of the OJ states *as opposed to* being a non-OJ state (i.e. an Apple Juice state). To calculate this we use classical conditioning under the measure Π, so she should calculate $\Pi(H|OJ)$, which turns out to be ⅓.

Of course, we have modelled the standard SB experiment in Example B as being formally equivalent to Example D. So we also have $\Pi_{SB}(H|Aw) = \frac{1}{3}$. Is this the right calculation SB should perform if upon waking she wants to calculate the probability the toss was Heads? As stated in the introduction there is a key difference between the standard SB example and the Orange/Apple Juice example. In the standard SB experiment she should not condition on being Awake as opposed to being Asleep (in the way we conditioned on being served OJ instead of Apple Juice), since she could never be Asleep and be asking herself this question. Instead she needs in some sense to *restrict to only being* awake. We will define what this means in the next section.

Before getting to the next section, just a note on the scope of conditioning: there is a sense in which both Ξ and Π can be conditioned on a subset of the behaviours of the experiment. For example, if $\Im$ is a subset of the set of behaviours of an experiment, then we can define Π relative to or conditional on $\Im$ as follows:

$\Pi_{\mathfrak{I}}(S) = \frac{1}{\sum_{p \in \mathfrak{I}} \pi(p)} \sum_{p \in \mathfrak{I}} \pi(p) * \frac{n(S,p)}{n(p)}$. We won't be using this construct in the examples of this paper, but the theory can accommodate utilizing information on the set of behaviours.

4. Restricting an experiment to only being in a subset of its states

We now get to the key definition of the paper. Conceptually it is quite simple, given an experiment we want to define a new experiment that starts with the original experiment and then restricts it to a subset of its intermediary states, but in a sense still 'remembers' the paths that went through the other states. In the case of the experiment Example B we have used to model the SB problem, the construction is very simple: it amounts to removing one state and one transition, so we end up with Example C. But for a general experiment it takes a little bit of work to define as we need to define the transition probabilities of the new experiment.

Definition: Restricting an experiment to only being in a subset of its states

Suppose X is an experiment and C is a subset of its intermediary states. Define $(X|C)$, the experiment X restricted to only being in C, to be the experiment defined as follows.

- It has the same Begin and End states as X, and its set of intermediate states is C. So the set of states of $(X|C)$ is $(X|C)_S = \{B, E\} \cup C$.
- Its transition probability function $\pi_{(X|C)} : (X|C)_S \times (X|C)_S \to \Re^{\geq 0}$ is defined as follows. For $v, w \in (X|C)_S$, define $Path_{\neg C}(v, w)$ to be the set of paths $v = s_0 \to ... \to s_n = w$ in X with the property that for all i, $0 < i < n$, $s_i \notin C$. We include the case where $n = 1$ here, i.e. the one step path $v \to w$ is included in $Path_{\neg C}(v, w)$. Then define

  $$\pi_{(X|C)}(v, w) = \sum_{p \in Path_{\neg C}(v,w)} \pi_X(p).$$

Proposition: $(X|C)$ is an experiment. In particular, for all $v \in (X|C)_S$,

$$\sum_{w \in (X|C)_S} \pi_{(X|C)}(v, w) = 1 .$$

Exercise: Calculate $\pi_{(X|C)}(B, E) - \pi_X(B, E)$.

Note/question for mathematicians:

- We would like to make a definition of a morphism/map of (i.e. structure-preserving function between) experiments in such a way that this operation of restriction can be viewed as a map experiments $(X|C) \to X$ - which could also be dually viewed in terms of thinking of X as an *extension* of the experiment $(X|C)$

- One way to do this would be to define a map from experiment X to experiment Y to be a function $f: X_S \to Y_S$ between the sets of states of the experiments that satisfies the following condition: for all states v, w of X,

  $\pi_X(v,w) \leq \sum_{p \in Path_{\neg image(f)}(f(v), f(w))} \pi_Y(p)$. I believe this defines a category of experiments.

  The inclusion of the sets of states $(X|C)_S \to X_S$ then is an example of such a map (and is in fact maximal in a certain sense).
- The most natural way to do define a map of experiments, though, would be to define experiments in terms of the graphs associated with the state-transition diagrams of the experiment (but where we have generalized the definition of experiment to allow more than one transition between two states). A morphism from X to Y can then be defined a map of directed graphs $f: X_G \to F(Y_G)$ where X_G is the graph underlying the state-transition diagram of X and $F(Y_G)$ is the graph underlying the free category on the graph Y_G. (So f maps a transition in X to a path in Y.) The map of directed graphs f is asked to preserve the Begin and End states of the experiments, and would need to satisfy some condition (I am not sure what) with respect to the transition probability structure.
- The above definition of restriction relies on being able to form the complement $\neg C$. What if we considered experiments where the states form a topological space, or better, can be modelled as an object of a topos? I don't ask this merely as a technical question, but moreover I am wondering whether this question has something to do with the concept of *state*, in particular, whether a certain type of discreteness or separability is implicit in a notion of being vs not-being in a state.

Proposition: The counting probability measure Ξ is insensitive to the restriction operation in the following sense. If P, C are subsets of the set of intermediary states of experiment X, then $\Xi_{(X|C)}(P \cap C) = \Xi_X(P|C)$.

This particular form of 'insensitivity' of the counting probability measure largely underlies why it appears to be 'robust', a property that appeals to many 'thirders'.

The above result does not hold for the proportion probability measure Π. But we do have the following natural interpretation of the quantity $\Pi_{(X|C)}(P \cap C)$.

Proposition: If P, C are subsets of the set of intermediary states of experiment X, then

$$\Pi_{(X|C)}(P \cap C) = \frac{1}{\sum_{p \varepsilon Beh(C)} \pi_X(p)} * \sum_{p \varepsilon Beh(C)} \pi_X(p) * \frac{n(P \cap C, p)}{n(C, p)}$$, where

- $Beh(C)$ is the subset of behaviours of X that pass through C
- $\pi_X(p)$ is the probability of behaviour p occurring in X
- $n(S, p)$ denotes the number of times p visits a state in subset S in X

Note:

- We can interpret the quantity $\frac{n(P \cap C, p)}{n(C, p)}$ as the probability of being in P conditional on being in C - just in the context of being in behaviour p.
- So we can interpret $\Pi_{(X|C)}(P \cap C)$ as the *expected value* of such probabilities, where the expectation is taken over the probability measure of behaviours passing through C.

Proposition: The restriction operation is 'associative' in the following sense. If X is an experiment, and $C, D \subseteq X_I$ then $((X|C)|C \cap D) = (X|C \cap D)$.

In the examples below, we will often just write $((X|C)|D)$ for $((X|C)|C \cap D)$.

Example 1 (How SB should condition on being awake)

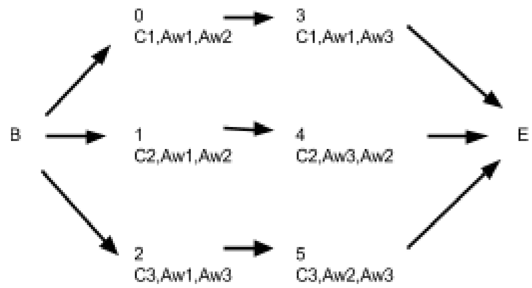
Consider the model SB of the standard SB experiment - Example B in Section 2. If we condition this on SB only ever being awake, i.e. construct the experiment $(SB|Aw)$, we get Example C from Section 2.

We argue that upon awakening, to calculate the probability Heads was tossed SB should calculate $\Pi_{(SB|Aw)}(H)$, which turns out to be ½.

Example 2 (How should multiple Sleeping Beauties condition on being awake)

Warning: don't get confused by the numbers in this example. (Most 'halfers' will probably assert the answer to this problem is ⅓; and 'thirders' will no doubt claim it is ½!)

Recall the experiment described in Example F of Section 2:

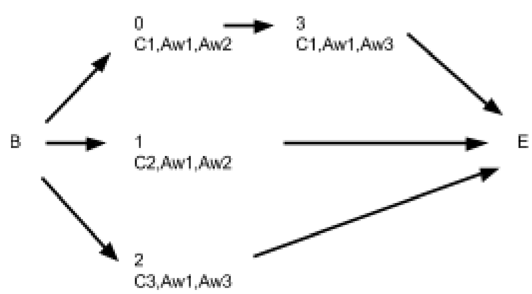


Suppose you are $B1$, you have been awoken and have been brought together with $B2$. The problem is to determine the likelihood you are the chosen one (i.e. that the experiment is in one of the $C1$ states). Recall there were two arguments: if we take just the perspective of $B1$, we got ⅓; but if we took the joint perspective of $B1$ and $B2$, we got ½.

It turns out the constructs developed here mirror this.

Since we are answering from the perspective of a subject of the experiment, we want to use the proportion probability measure. There are two perspectives we can take.

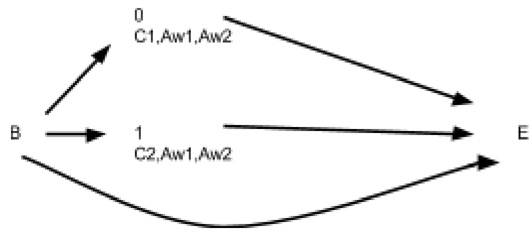
On the one hand, consider restricting the experiment to the subset of states $Aw1$ (i.e. states where $B1$ is awake). This gives:



There are three equally likely behaviours in this experiment. We then calculate the proportion probability of $C1$ conditional on $Aw2$:

$$\Pi_{(X|Aw1)}(C1|Aw2) = \frac{\Pi_{(X|Aw1)}(C1 \cap Aw2)}{\Pi_{(X|Aw1)}(Aw2)} = \frac{\frac{1}{3}*(0.5+0+0)}{\frac{1}{3}*(0.5+1+0)} = \frac{1}{3}.$$

On the other hand, consider restricting the experiment to the subset of states $Aw1 \cap Aw2$ (i.e. states where both $B1$ and $B2$ are awake). This gives:



Here there are three equally likely behaviours, only two of which pass through intermediary states. We then calculate the proportion probability of $C1$:
$\Pi_{(X|Aw1 \cap Aw2)}(C1) = \frac{1}{1-\frac{1}{3}} * (\frac{1}{3} * 1 + \frac{1}{3} * 0) = \frac{1}{2}$.

So depending on the view taken, just the perspective of $B1$, or the joint perspective of $B1$ and $B2$, we get different answers. I still have not got my head around this completely, but this approach suggests that there is not an absolute answer to the question - the answer must be relative to a perspective.

As mentioned, we have used the proportion probability as we are calculating from the perspective of the subject. But we note that if we use the counting probability measure, we get $\Xi_X(C1|Aw1) = \Xi_X(C1|Aw1 \cap Aw2) = \frac{1}{2}$.

5. Composite experiments

We now define the series composite of two experiments X and Y. This is supposed to represent performing experiment X immediately followed by experiment Y.

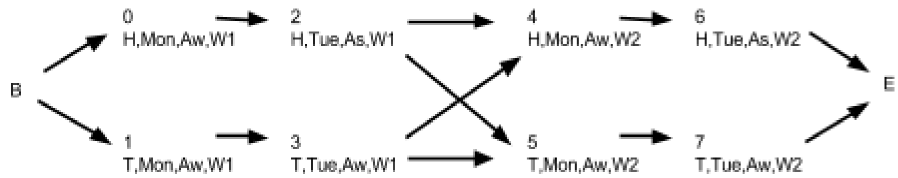
Definition

Define the *series composite* $X * Y$ of the experiments X and Y as follows. Its Begin state is the Begin state of X; its set of intermediary states is the union $X_I \cup Y_I$; and its End state is the End state of Y. Note that the End state of X and the Begin state of Y have been excluded. For any two states v, w, the transition probability $\pi_{X*Y}(v, w)$ is defined as follows:

- If $v, w \in X$ then $\pi_{X*Y}(v, w) = \pi_X(v, w)$
- If $v, w \in Y$ then $\pi_{X*Y}(v, w) = \pi_Y(v, w)$
- If $v \in Y, w \in X$ then $\pi_{X*Y}(v, w) = 0$
- If $v \in X, w \in Y$ then $\pi_{X*Y}(v, w) = \pi_X(v, E_X) * \pi_Y(B_Y, w)$

Proposition: This defines an experiment.

Recall the the 2-week SB experiment SB^2 that was [Example E](#) from Section 2:

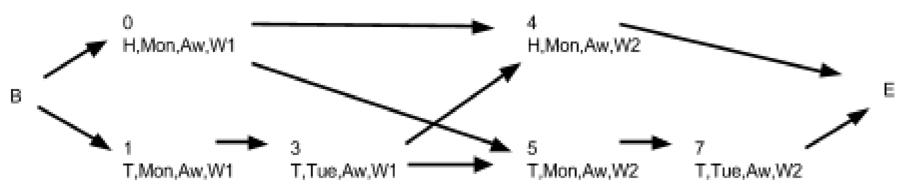


Exercise: Show that SB^2 can be obtained as the series composite $SB * SB$ of [Example B](#) with itself.

Example (Conditioning the 2-week SB experiment)

In this example we will go through the calculations described in the introduction relating to the 2 week SB experiment SB^2 . Specifically, we will see how classical conditioning laws can break-down when restricting to sub-experiments of a composite experiment due to a type of ‘interference’ in proportional probability calculations (that arises due to renormalization of probabilities along individual behaviours).

Below we depict $(SB^2|Aw)$:



Exercise: check $\Pi_{(SB^2|Aw)}(W1) = \Pi_{(SB^2|Aw)}(W2) = \frac{1}{2}$.

If you restrict $(SB^2|Aw)$ to being in week 1, i.e. calculate $((SB^2|Aw)|W1)$, you get an experiment that is formally equivalent to $(SB|Aw)$ the standard (1 week) SB experiment restricted to SB being awake. That is, $((SB^2|Aw)|W1) = (SB|Aw)$. Similarly, you can check $((SB^2|Aw)|W2) = (SB|Aw)$.

Therefore we have $\Pi_{((SB^2|Aw)|W1)}(H) = \Pi_{((SB^2|Aw)|W2)}(H) = \Pi_{(SB|Aw)}(H) = \frac{1}{2}$.

Now, when restricting this 2-week experiment to SB being awake, we want to calculate $\Pi_{(SB^2|Aw)}(H)$, i.e. the proportion probability that the last toss was heads. We might hope to be able to calculate this as follows. Restricting to SB being awake, there's an even chance she is in the first week or the second week (as calculated above); in each of these weeks, restricting to SB being awake, there is a 50% chance Heads was the result of the last toss (also calculated above); therefore, restricting to SB being awake, there is an even chance the result of the last toss was Heads or Tails. In mathematical terms:

$$\Pi_{(SB^2|Aw)}(W1) * \Pi_{((SB^2|Aw)|W1)}(H) + \Pi_{(SB^2|Aw)}(W2) * \Pi_{((SB^2|Aw)|W2)}(H) = \tfrac{1}{2} * \tfrac{1}{2} + \tfrac{1}{2} * \tfrac{1}{2} = \tfrac{1}{2}.$$

However, in this case we will see

$$\Pi_{(SB^2|Aw)}(H) \neq \Pi_{(SB^2|Aw)}(W1) * \Pi_{((SB^2|Aw)|W1)}(H) + \Pi_{(SB^2|Aw)}(W2) * \Pi_{((SB^2|Aw)|W2)}(H)$$

The experiment $(SB^2|Aw)$ has 4 equally likely behaviours which are characterized by the following subsets of intermediary states: $\{0,4\};\ \{0,5,7\};\ \{1,3,4\};\ \{1,3,5,7\}$. A straightforward calculation gives:

$$\Pi_{(SB^2|Aw)}(H) \ = \ \tfrac{1}{4} * (1 + \tfrac{1}{3} + \tfrac{1}{3} + 0) = \tfrac{5}{12}$$

What is going on? If you look closely, the difference is due to the ⅓ terms in the calculation of $\Pi_{SB^2}(H|Aw)$ (e.g. on the path where there was first a Heads and then a Tails, the probability the last toss was heads conditional on SB being awake is ⅓). These appear because of *renormalization* of probabilities along composite paths (thinking of a behaviour being a composite of the week 1 experiment and the week 2 experiment).

The moral is that, from the perspective of the subject, in general you have to look at *entire* behaviours of the experiment to calculate the probability of being in a particular type of state.

Even though there is no cancellation (as can occur when adding amplitudes viewed as complex numbers in a quantum mechanical context), it doesn't seem entirely unreasonable to refer to this effect as a type of non-locality or probabilistic 'interference' associated with composing two experiments.

You could more generally consider the experiment of running the SB experiment for n weeks in a row, i.e. form the n-fold series composite of the experiment with itself. Call this composite experiment SB^n (I won't draw a graph!), and consider the probability of the experiment being in a state such that the last toss has been heads conditional on SB being awake. What is this probability in the limit as $n \to \infty$?

Well, it is easy to see it is equal to $\lim_{n\to\infty} \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} * \frac{i}{2n-i}$. Numerical estimation shows this converges to ⅓.

To me, this concurs with intuition. In the limit as we approach the fantastical case where SB is in an eternal experiment, there is a sense in which the space of behaviours becomes increasingly crowded with behaviours for which the probability of it being heads conditional on SB being awake is close to ⅓.

Question:

- to prove the convergence to ⅓ is equivalent to showing $\lim_{n\to\infty} \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} * (\frac{i}{2n-i} - \frac{i}{n+i}) = 0$. Any suggestions on how to do this (maybe using results about the limiting case of the binomial distribution approaching a Gaussian) ?

As noted in the introduction, we generalize this to make the following conjecture relating the perspectives of an external observer and an internal observer (i.e. a subject) of the experiment.

Suppose X is an experiment. Let X^n be the experiment constructed by repeating X in succession n times (i.e. the n-fold series composite of X with itself). So the set of states of X^n is the disjoint union of X with itself n times. If S is a subset of the states of X then let S^n be the subset of states of X^n comprising the union of each copy of S in each of the n instances of X.

Conjecture: $\Xi_X(S|C) = \lim_{n\to\infty} \Pi_{(X^n|C^n)}(S^n)$.

6. Final comments

To sum up, this paper has attempted to clarify the SB and related problems by making two key conceptual distinctions: one regards internal versus external perspectives (as represented respectively by the use of the proportion probability measure Π and the counting probability measure Ξ); and the other regards the distinction between 'only being in a state satisfying property P' versus 'currently being in a state satisfying property P as opposed to $\neg P$' (as represented respectively by the use of the operation of restricting experiments to only being in a subset of states and the classical conditioning associated with probability measures).

I have not in this paper made any mathematical definitions that specify, or have in any sense proven, when one should use the proportion probability measure Π instead of the counting measure Ξ, or when one should use the classical conditioning associated with these probability measures instead of restricting experiments to only being in certain states. Instead I have used examples to indicate when/how they should be used. Moreover, the concept of experiment I have defined is so abstract that I have used the same formal model to represent two different real world situations (e.g. Example B and Example D), and argued that different constructs need to be applied in the two situations. This could be seen as a weakness in the comprehensiveness of the approach as a formal model. To address this one would have to develop a deeper mathematical theory of experiments in which concepts and distinctions such as internal vs external, and ‘currently being in’ vs ‘only ever being in’ were made explicit in the way experiments themselves are defined.

On a technical note, the series composite construction in the previous section (or a variant of it) can be understood in terms of the algebra (a monoidal bicategory) of what are known to category theorists as *cospans* of directed graphs (or more generally one could consider reflexive graphs, or even cospans of categories equipped with 'duration' functors into a monoid). One can also consider structures known as *spans* of graphs to model components working in parallel and synchronizing over actions. The idea of using spans and cospans of graphs (and in this context defining operations such as parallel and sequential composition, various forms of feedback, as well studying properties of such operations such as distributive laws) was first put forward myself some 15 years ago with Bob Walters and Nicoletta Sabadini to develop a compositional model of distributed systems (where components could share state as well as synchronize over actions). Bob and Nicoletta, with their collaborators, have continued to work with these constructs. If you are interested in learning more you can find out about them from Bob’s blog or you can peruse his publications page. Specifically, they have looked at how Markov Chains can be modelled in this context.

What I originally found interesting about the SB problem is that it forced me to define what I meant *as a halfer* by 'the probability an experiment is in a state satisfying certain conditions...'; and, in doing so, I needed to assert the following philosophical or ‘moral position’:

As a subject of an experiment, to make sense of what it means for an experiment to be in a state requires referencing the complete behaviour of which the state is being considered a part

More specifically, this ‘moral position’ manifested itself in the following way: to capture the perspective of a subject of an experiment, the renormalization associated with the conditioning of the probabilities of an experiment being in a state should be first carried out at the level of each complete behaviour, and then integrated over the probability space of behaviours of the whole experiment to obtain a quantity such as $\Pi_{(X|C)}(P \cap C)$.

Many people may well disagree with this ‘moral position’. However, I note, this seems to me to echo (again, perhaps only very weakly) Feynman's following remark made to assist readers in thinking about quantum mechanical interference (p81, QED, Princeton University Press, 1988 printing): 'let me remind you of a most important principle: in order to correctly calculate the probability of an event, one must be very careful to calculate the complete event clearly - in particular, what the initial conditions and the final conditions of the experiment are.'

There is obviously a world of difference between combinatorial structures such as a Markov Chain, and the calculations involving complex numbers and probability amplitudes. Also, I realize there is no analogy with the key physical aspect of QED (as say represented in Chapter 2 of Feynman’s book Quantum Mechanics and Path Integrals) whereby the classical behaviour can be seen as resulting from the cancellation of rapidly spinning phases on paths deviating from the classical path (in cases where the action is very large compared to Planck’s constant). So the analogy perhaps amounts to no more than simply noting that in both cases interference has something to do with the algebraic fact that in general $(a+b)(c+d) \neq ac+bd$.

But I can’t help wondering if there is something more, perhaps something to do with emergence (where descriptions of two levels of reality are involved, with one level giving rise to *events* in the other). I am imagining a situation where we couldn’t directly observe whether an experiment is in a given state, but only ask questions like ‘is the experiment X in a state satisfying a property P, conditional on the experiment only being in states satisfying C’; to which we would get an answer of yes (an event) with probability $\Pi_{(X|C)}(P \cap C)$. For this to happen, the answer can’t be generated at random in the middle of a behaviour of X (as the total number of times the behaviour will visit a state satisfying the condition C needs to be known - this is the renormalization). So the ‘time’ in which the underlying behaviour of the experiment occurs would have to be in some sense independent of the time in which we, ‘outside the experiment’, are waiting for an answer (i.e. for the event to happen). Let me put this another way. The difficulty with the concept of emergence is that you want one (lower) level description of reality to give rise in some sense to another (higher) level description of reality without the lower level trivially explaining (reducing) the higher level. Can the ideas presented here be used or extended to give a formal model of such a thing?