Sleeping Beauty Revisited

Tom Goldring

November 23, 2020

Abstract

If someone flips a fair coin (evenly weighted on both sides), we say the probability of it coming up heads is $\frac{1}{2}$. If a party has 20 males and 30 females and one of them is selected at random, we say the probability that it's a male is $\frac{2}{5}$. If a pair of fair dice are rolled, we say the probability that one comes up 2 and the other 5 is $\frac{1}{18}$. In these and many other cases, we speak of *the* probability of the event, and we calculate it by dividing the number of "favorable" outcomes by the number of possible outcomes. Can we always associate a single probability with each event resulting from a repeatable experiment, or are their cases where the probability is a matter of opinion? And if the event is unique (not repeatable), is it still meaningful to talk about its probability? What if the event is repeatable, but can still be viewed as unique? The beauty of the "Sleeping Beauty" problem lies in how it provides a simple, concrete example for all of these questions and more.

1 Introduction

Normally we compute the probability of an event by counting outcomes and takiing ratios – for example, if one flips a fair coin, we say the probability of it coming up either heads (or tails) is $\frac{1}{2}$, because if it were flipped many times, approximately half of them would be heads. In the statistics of polling or of clinical trials, we sample from a large population and test hypotheses based on mathematics which presumes that this sample is only one of many possible others, and whatever statistic we derive from it has a fixed distribution. In the *frequentist* interpretation, "probability is only defined for an aggregate, while the probability of a unique (non-repeateable) event is meaningless. Nevertheless, examples of unique events abound in the real world, and we still speak about their probability. If I can't find my car keys, I "probably" left them in the car. If my date shows up late, she "probably" got stuck in traffic or had to work late. In a criminal trial, usually it's impossible to know for certain whether the defendent is innocent or guilty, so when we ask the jury to deliver a verdict, we're really asking them to assess the probability of guilt, not as a specific number, but rather "beyond a reasonable doubt", even though it's often meaningless to think about repeated trials of the defendent having committed the crime. Other examples are the interpretation of political events, historical events, or the "true" intent of the novelist, biblical author, or philosopher. The "fine tuning for life" argument posits a cosmic Designer by assuming it's just too unlikely for our life-supporting universe to have arisen by chance, even though we don't know about any other universes. The reasoning is based on the standard belief that if something we observe is more likely to occur when a given hypothesis is true, that thing ought to count as evidence for the hypothesis. The alternative to frequentist probability is the *Bayesian* interpretation, where "probability" is defined as "degree of belief" and the underlying mathematics is based on the concept of "conditional" probability. Does this make Bayesian reasoning inherently subjective? The frequentist vs. Bayesian controversy has been going on for at least a couple of centuries.

If you're thinking that all of this is getting a bit muddy, you're in good company. But surely these questions can be definitively answered in simple, well defined scenarios, right? Well, let's see.

2 Statement of the problem

Sleeping Beauty agrees to participate in an experiment. On Sunday, she is put to sleep and a fair coin is flipped. If it comes up heads, she is awakened on Monday and interviewed, ending the experiment. If it comes up tails, she is awakened on Monday, interviewed, then put back to sleep, awakened again on Tuesday and interviewed once more, ending the experiment. When awakened, she does not know if it's Monday or Tuesday, nor does she remember on Tuesday that she was already awakened/interviewed on Monday. She does however have complete knowledge of the experimental protocol. The interview consists of a single question: how do you estimate of the probability that the coin-flip was heads?

This puzzle was originally formulated in the mid 1980's and created quite a controversy, with many scholarly articles published, but no consensus was ever reached (see [PW] for a survey article containing an extensive bibliography). Most respondents tended to split into two camps: those who claim it's $\frac{1}{2}$ (the "halfers"), and those who claim it's $\frac{1}{3}$ (the "thirders"). However, there is also a "dualist" camp, which insists that Beauty's answer could be either $\frac{1}{2}$ or $\frac{1}{3}$, depending on how one interprets the interview question. In what follows, I will explain the evolution of my reasoning on this problem:

- Why I started out as a thirder:
  - the standard halfer arguments are logically incorrect while the standard thirder argument is quite reasonable
  - moreover, the halfer position has mathematical implications which at first glance seem counterintuitive.
- And yet, at some point I realized that there were better halfer arguments, according to which the counterituitive implications made sense as well.
- Finally, I saw how the entire discussion exemplified a long-standing debate over the fundamental meaning of "probability".
- The upshot is that over time I've come to appreciate the dualist point of view.

3 The basic halfer argument

The halfer argument one sees most often observes that when the coin is flipped, the probability of it coming up heads is certainly $\frac{1}{2}$, and since Beauty receives no new information, there is no reason for her to modify that number. However, I consider this argument to be incorrect because the word "new" is misleading. What we should really be asking is whether Beauty has any information *in addition to* the fact that the coin is fair, and indeed she does: her knowledge of how the experiment proceeds. To this the halfers might object that such knowledge is irrelevant because causation only goes one way – from coin-flip to awakening[1] – but that is also mistaken. It's true that *physical* causation only goes one way, however *epistemic* causation – how Beauty's answer is affected by what she knows – also goes the other way. For example, if the interview is taking place on Tuesday, the coin must have come up tails. To this, halfers will say that Beauty can reason based *only* on what she knows, and she does not know what day it is. Of course that is true, but they neglect to tell us that her reasoning should include *everything* she knows. In particular, even though she doesn't know *if* it's Tuesday, she does know it *might* be Tuesday, and indeed it *will* Tuesday be some of the time. When it is, the correct answer would be zero, so in general, one could argue that knowing all of this ought to change her estimate of the probability of heads from $\frac{1}{2}$ to something smaller.

4 The basic thirder argument

The simple thirder argument is that there are twice as many tails-awakenings (awakenings following a tails-flip) as heads-awakenings, therefore $\frac{1}{3}$ of the total number of awakenings follow a heads-flip. For example, imagine 100 trials (repetitions of the experiment). On average, the coin comes up heads in 50 of them and tails in the other 50. The heads-flips result in Beauty being awakened 50 times (all on Monday), while the tails-flips result in 50 Monday awakenings and 50 Tuesday awakenings. Out of a total of 150 awakenings, in 50 of them the coin-flip was heads, so the probability is $\frac{50}{150} = \frac{1}{3}$. In what

[1] For the purpose of this analysis I will treat "awakening" and "interview" as synonymous.

follows, I refer to this as the *counting argument*.

One might wonder how that argument could possibly be contested. Well, nobody denies that we counted correctly, but halfers can claim that it answers a different question than the one Beauty is asked. Imagine you were an independent observer of the coin-flip, and were asked to estimate the probability of the coin coming up heads. Everyone would agree that the correct answer is $\frac{1}{2}$. So what if the original question ("what is your best estimate of the probability that the coin-flip was heads?") is slightly modified to "had you been awake when the coin was flipped, what is the likelihood that you would have observed heads"? Shouldn't she answer exactly as the independent observer does? And isn't this just another way of phrasing the original question? On the other hand, thirders could object that Beauty is being interviewed after being awakened, not after the coin-flip, and what she is really being asked is "*right now*, how do you estimate the probability that you are experiencing a heads-awakening?". Should that change her answer?

5 Everyone is right

Perhaps the root of the controversy is not about how to calculate a probability, but rather about how the interview question ought to be interpreted (see [PM] for a very readable exposition of this point of view). Here are some variations on how that question might be phrased:

1. What is the probability that an arbitrary fair coin comes up heads?
2. What is the probability that the coin in this experiment came up heads?
3. What is the probability that this interview is taking place after the coin came up heads?

Obviously question 1 must be $\frac{1}{2}$. As for 2, assuming the experimenter simply reached into his pocket and took out an arbitrary fair coin, shouldn't the answer be the same, since "the coin in this experiment" *is* an arbitrary fair coin? Coming to 3, shouldn't it still be the same, since the coin-flip in "this experiment" necessarily results in a subsequent interview taking place? And yet, how can we simply answer $\frac{1}{2}$ without considering what day the interview is taking place? If it's Monday, one would think heads and tails are equally likely, but if it's Tuesday, it has to be tails. On the other hand, if the answer is anything other than $\frac{1}{2}$, it seems as though the situation in which this arbitrary fair coin is being used mysteriously affects its physical behavior!

I suppose die-hard halfers will have a strong preference for 1, and insist that the extra words in 3 just obscure the true intent, which is to count coin-flips. On the other hand, die-hard thirders will maintain that the wording in 3 is there to clarify the true intent, which they claim is to count interviews. In the elementary definition of the probability of an event, one imagines many trials and computes the probability of the event as

$$\frac{\text{the number of trials in which the event takes place}}{\text{the total number of trials}}\,. \tag{1}$$

No doubt the philosophical school of thought that emphasizes language above all else will say the controversy comes down to what constitutes a "trial". Is it a flip of the coin? If so, in 100 trials, on average the coin comes up heads in 50 of them, so the fraction in (1) is $\frac{1}{2}$. But if a trial is an interview, the fraction is $\frac{1}{3}$. According to the dualists, the interview question is ambiguous and could be interpreted either way, and so could whether an "independent observer" is an observer of the coin-flip or the interview. How you interpret it is a matter of opinion, so both sides can agree to disagree, and everyone lives happily ever after in peace and harmony. This will be made precise in Sec. 7.2.1.

6 The mathematical argument using conditional probability

The reader who is not familiar with conditional probability can read Appendix 1 before proceeding, or else take the formulas presented below on faith.[2]To make the calculation less cumbersome, we

[2]In that case, one need only know that for any event A, by $P\,[A]$ denotes "the probability of A", and for any two events A and B, the notation $P\,[A\,|\,B]$ is just shorthand for "the probability of A given B", or "the conditional probability of A under the

introduce some notation:

notation	event
h	the coin-flip came up heads
t	the coin-flip came up tails
M	the awakening takes place on a Monday
T	the awakening takes place on a Tuesday

(keep in mind that the end result of the exercise is to estimate $P[h]$). What follows is a good example of how the power of mathematical reasoning comes from the combination of intuition and formalism. To begin, we list the conditional probabilities for when Beauty is interviewed:

1. $P[M \mid h] = 1,\ P[T \mid h] = 0$ (if the coin-flip came up heads, the awakening can only be on Monday).
2. $P[M \mid t] = P[T \mid t] = \frac{1}{2}$ (if the coin-flip came up tails, Beauty is always awakened on both Monday and Tuesday, and when she is awakened, even though she doesn't know which day it is, she does know that "today" is equally likely to be either).
3. $P[h \mid T] = 0,\ P[t \mid T] = 1$ (if the awakening is on Tuesday, the coin-flip must have come up tails).

We know these values intuitively as immediate consequences of the experimental setup, therefore all camps should agree on their correctness. Note however that $P[h \mid M]$ (the probability that the coin came up heads given the interview takes place on a Monday) and its complement $P[t \mid M]$ have been omitted, for reasons which will soon become clear.

Next, we write down the "conditioning" equations that determine how the probabilities for heads and tails are related to those for Monday and Tuesday:

$$P[M] = P[M \mid h]\,P[h] + P[M \mid t]\,P[t] \tag{2}$$
$$P[T] = P[T \mid h]\,P[h] + P[T \mid t]\,P[t] \tag{3}$$
$$P[h] = P[h \mid M]\,P[M] + P[h \mid T]\,P[T] \tag{4}$$
$$P[t] = P[t \mid M]\,P[M] + P[t \mid T]\,P[T]\ . \tag{5}$$

If we let $a = P[h]$ and $x = P[h \mid M]$, then $P[t] = 1 - a$, $P[t \mid M] = 1 - x$, and substituting into (2)-(5):

$$\begin{aligned} P[M] &= 1 \cdot a + \frac{1}{2}(1-a) \\ P[T] &= 0 \cdot a + \frac{1}{2}(1-a) \\ P[h] &= x \cdot P[M] + 0 \cdot P[T] \\ P[t] &= (1-x) \cdot P[M] + 1 \cdot P[T] \end{aligned}$$

and these reduce to

$$P[M] = \frac{1}{2}(1+a) \tag{6}$$
$$P[T] = \frac{1}{2}(1-a) \tag{7}$$
$$P[h] = x\,P[M] \tag{8}$$
$$P[t] = (1-x)\,P[M] + P[T] \tag{9}$$

(the reader can check that $P[M] + P[T] = P[h] + P[t] = 1$). Substituting (6) into (8),

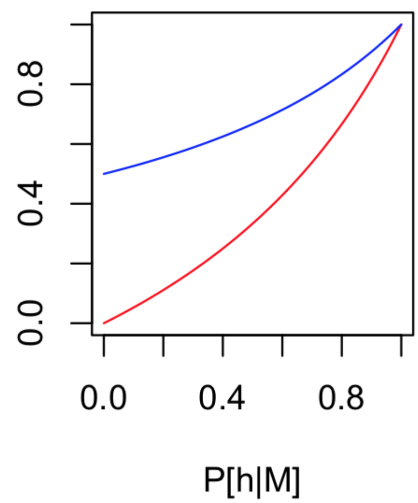
$$a = \frac{1}{2}\,x\,(1+a)$$
$$x = \frac{2a}{1+a} \tag{10}$$
$$a = \frac{x}{2-x} \tag{11}$$

assumption that B is known to be the case".

The reader can check that (9) is redundant, because substituting (6) and (7) into (9) gives the same result. Finally, substituting (11) into (6) gives

$$P[M] = \frac{1}{2-x}\,. \tag{12}$$

Referring to the diagram below, using (11) and (12) to plot $a = P[h]$ (red) and $P[M]$ (blue) as a function of $x = P[h\,|\,M]$



shows that any one of the three variables $0 \le P[h\,|\,M] \le 1$, $0 \le P[h] \le 1$, $\frac{1}{2} \le P[M] \le 1$ uniquely determines the other two. As far as the mathematics, one can choose any number $P[h]$ between 0 and 1, then compute $P[h\,|\,M]$ from (10) and $P[M]$ from (12), and the result will be consistent with the laws of probability. In other words, *the mathematics does not determine a unique value for Beauty's answer*. In particular, halfers can let $P[h] = \frac{1}{2}$ (which determines $P[h\,|\,M] = \frac{2}{3}, P[M] = \frac{3}{4}$), and thirders can let $P[h] = \frac{1}{3}$ (which determines $P[h\,|\,M] = \frac{1}{2}, P[M] = \frac{2}{3}$). Any preference for either of these two possibilitiies (or any other) will require a separate (non-mathematical) justification.

But this should come as no surprise, since nowhere in the calculation did we specify that the coin is fair, and therefore the mathematics is equally valid for biased coins. The only way to incorporate fairness would be to assume $P[h] = \frac{1}{2}$, which would beg the question. Nevertheless, keep in mind that thirders claim the value $P[h] = \frac{1}{3}$ to also be consistent with the coin being fair! The only way that can make sense is if we interpret $P[h]$ to mean something other than the physical probablity[3]. That said, the obvious interpretation for $P[h]$ (and by extension, all of the other probabilities we've computed) is epistemic probability, i.e. that Beauty's assessment of $P[h]$ is influenced by what she knows. Since this is more than the simple fact that the coin is fair, the physical and epistemic probabilities do not have to equal each other. Of course, no one will deny that the physical probability of heads or tails is $\frac{1}{2}$, so one way to summarize the debate between halfers and thirders is that the former insist on physical probability, while the latter insist on epistemic.

And yet, isn't it necessarily the case that there should only be one correct answer? After all, couldn't we theoretically perform the experiment many times, ask the experimenter how the coin actually came up each time, and at the end, divide the number of "heads" answers by the total number of experiments? Better yet, write a computer program to simulate it? Well no, that won't help at all, because we would need to specify *when* the experimenter is queried. At the end of each experiment, or after each interview? If it's the former, the answer will be $\frac{1}{2}$, but if the latter, $\frac{1}{3}$.

[3] Either the physical fact that the coin is not more heavily weighted on one side, or the empirical fact that the coin is assumed to come up heads or tails equally often in the long run.

7 Do these values make sense?

7.1 The thirder values follow immediately from the counting argument

Define a *heads interview* as an interview following a heads-flip, a *Monday interview* as an interview that takes place on a Monday, and a *Monday-heads* interview as a heads interview that takes place on a Monday. Similar definitions are given for Monday-tails, Tuesday, etc. Assuming we are computing probabilities by counting interviews, there are

- 50 heads-interviews out of a total of 150 interviews, giving $P\left[h\right] = \frac{1}{3}$
- 50 Monday-heads interviews out of 100 Monday interviews, giving $P\left[h \,|\, M\right] = \frac{1}{2}$
- 100 Monday interviews out of a total of 150 interviews. giving $P\left[M\right] = \frac{2}{3}$.

This illustrates the *frequentist* interretation of probability: one imagines many repeated trials of the experiment and calculates probabilities as ratios of counts of events. For a hard-line frequentist, the idea of the "probability" of a *unique* event has no mathematical meaning.

7.2 The halfer values

7.2.1 It's all about what you count

If what was said in Sec. 3 is correct, the standard justification for the halfer position fails, however, a better argument was offered at the end of Sec. 5, which claimed that Beauty's answer depends on what is being counted in the denominator of (1): coin-flips or interviews. To formalize that idea, consider the conditional probability of the interview taking place on a Monday, given a heads-flip:

$$P\left[M \,|\, h\right] = \frac{P\left[M \& h\right]}{P\left[h\right]}\,. \tag{13}$$

In 1 of Sec. 6 we noted that $P\left[M \,|\, h\right] = 1$, and it immediately follows that $P\left[h\right] = P\left[M \& h\right]$, expressing the fact that every heads-flip results in only a Monday interview. The dualist position as stated in Sec. 5 says that there are two ways to interpret $P\left[M \& h\right]$:

- I_1: What fraction of all coin-flips result in Monday-heads interviews?
- I_2: What fraction of all interviews are Monday-heads (i.e. occur on Monday following a heads coin-flip)?

The thirder counting argument uses I_2 to justify the thirders ($\frac{1}{3}$ Monday-heads, $\frac{1}{3}$ Monday-tails, and $\frac{1}{3}$ Tuesday-tails). If we use I_1 instead, the halfer counting argument tells us that for 100 coin-flips, on average 50 of them result in Monday-heads interviews, so $P\left[h\right] = P\left[M \& h\right] = \frac{50}{100} = \frac{1}{2}$. Out of the other 50, 25 of them result in Monday-tails interviews, so $P\left[M\right] = \frac{75}{100} = \frac{3}{4}$. Finally, 50 out of the 75 Monday interviews are the result of heads-flips, therefore $P\left[h \,|\, M\right] = \frac{2}{3}$.

7.2.2 The Bayesian perspective

As it turns out, there are other intuitive arguments to justify the halfer values $P\left[h\right] = \frac{1}{2}, P\left[h \,|\, M\right] = \frac{2}{3}, P\left[M\right] = \frac{3}{4}$, and by extension, dualism.

1. Suppose we first tell Beauty that today is Monday, and then ask her "what is the probability of a heads-flip?". Given this new information, Beauty can reason as follows: since heads is certain to produce a "today is Monday" outcome but tails produces it only half of the time, it's not unreasonable to conclude that at the very least, $P\left[h \,|\, M\right]$ should be greater than $P\left[t \,|\, M\right]$, therefore $P\left[h \,|\, M\right] > \frac{1}{2}$. This is essentially the idea behind Bayes' theorem: if the evidence (today is Monday) is more likely when the hypothesis (the coin came up heads) is true as opposed to false, our degree of belief in the hypothesis ought to increase. Furthermore, since the evidence is twice as likely when the hypothesis is true, one might guess that the hypothesis is also twice as likely (so that $P\left[h \,|\, M\right] = \frac{2}{3}$), and in 2 we will see that Bayes' theorem (14) bears this out. On the other hand, this argument is not available to the hard-line halfer who believes that it simply does not matter what day it is and there's nothiing more to discuss.

2. The reader who is unfamiliar with Bayes' theorem should read Appendix 2 before proceeding. To make what was said in 1 precise, we take the hypothesis H to be "the coin came up heads", and the evidence E to be "today is Monday". Bayes' theorem becomes

$$P[h \mid M] = \frac{1}{1 + \frac{P[M \mid t]}{P[M \mid h]} \cdot \frac{P[t]}{P[h]}} \quad (14)$$

$$= \frac{1}{1 + \frac{\frac{1}{2}}{1} \cdot \frac{P[t]}{P[h]}} \, . \quad (15)$$

The critical point here is that when we substitute $P[M \mid t] = \frac{1}{2}$ into (14) to obtain (15), by $P[M \mid t]$ we *don't* mean the probability that a tails-flip will result in a Monday interview (that probability is 1), rather we mean the probability that given a tails-flip, *today* is Monday, not Tuesday. Equivalently, if the experiment were repeated many times, half of the interviews resulting from a tails-flip would occur on Monday. That said, the argument can go either way. Substituting the thirder values $P[h] = \frac{1}{3}, P[t] = \frac{2}{3}$ for $\frac{P[t]}{P[h]}$ we obtain $P[h \mid M] = \frac{1}{2}$ as expected, but substituting the halfer values $P[h] = P[t] = \frac{1}{2}$ gives $P[h \mid M] = \frac{2}{3}$. Now there are two ways to understand this calculation:

   (a) One way is that letting $\frac{P[t]}{P[h]} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$ (resp. $\frac{P[t]}{P[h]} = \frac{\frac{2}{3}}{\frac{1}{3}} = 2$) is just plugging the halfer (resp. thirder) values into a formula to see what pops out, so it's nothing more than a cross-check on what was already calculated in Sec. 6.

   (b) However, in the *Bayesian* interpretation of (14), there's a lot more going on. The second interpretation is that since $\frac{P[t]}{P[h]}$ is the inverse prior odds, the halfer substitution $\frac{P[t]}{P[h]} = \frac{\frac{1}{2}}{\frac{1}{2}} = 1$ is an *agnostic prior*, i.e. before considering the fact that today is Monday, we have no opinion regarding the probability of heads vs. tails. In this reading of Bayes' theorem, what initially appeared to be the halfer assumption $\frac{P[t]}{P[h]} = 1$ is actually the lack of an assumption, and this is consistent with the fact that it does not contribute to the result $P[h \mid M] = \frac{2}{3}$. As we saw in Sec. 6, that result is consistent with the halfer, not the thirder position, and in that sense, Bayes' theorem can be viewed as an intuitive justification of the halfers, not merely the application of a mathematical formula.

3. Suppose Beauty is asked "what is the probability that today is Tuesday?". She observes that today is Tuesday if and only if two things happen in direct succession:

   (a) when the the coin is flipped on Sunday, it comes up tails

   (b) given the coin came up tails, today is Tuesday, not Monday.

   Given a Sunday tails-flip, "today" being Monday or Tuesday are equally likely, therefore 3b has probability $\frac{1}{2}$. What about 3a? The physical probability of a tails-flip is $\frac{1}{2}$, but if the experiment were repeated many times, $\frac{2}{3}$ of the interviews would occur after a tails-flip, so from that perspective, the probability of a tails-interview is $\frac{2}{3}$. Rewriting (3) as

$$P[T] = P[T \mid t] \, P[t]$$

$$= \frac{1}{2} P[t] \quad (16)$$

   if one substitutes $P[t] = \frac{1}{2}$ then $P[T] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ and $P[M] = 1 - P[T] = \frac{3}{4}$ (the halfer result), but substituting $P[t] = \frac{2}{3}$ gives $P[T] = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$ and $P[M] = \frac{2}{3}$, (the thirder result).

   As the physical probability of the coin-flip, $P[t] = \frac{1}{2}$ seems natural, but the epistemic probability $P[t] = \frac{2}{3}$ is also reasonable. So which should it be, physical or epistemic? Treating both as equally legitimate argues for dualism.

   Also, if Beauty sees the interview as a unique event, the probability that today is Tuesday must be $\frac{3}{4}$. Since "today" is not a single event selected at random out of an aggegate, using the physical probability $P[t] = \frac{1}{2}$ is her only choice. This shows that epistemic probability is not necessarily frequentist, but can also be Bayesian.

8 Is there a paradox here?

Historically, the frequentist vs. Bayesian controversy has centered around whether interpreting probability as degree of belief is "subjective". If that were the case, it would count against Bayesianism as unscientific. With the caveat that I have not studied this history in depth, I have two responses:

- As mentioned in Sec. 1, our day-to-day experience includes many unique events to which we routinely assign probabilities, and this can only be understood as degree of belief.
- Supposing one believes (as I do) that both frequentist and Bayesian analysis are valid ways to reason about probability, then the frequentist argument of Sec. 7.1 and the Bayesian argument in 2b are both correct, even though they give different answers. If "probability" is not *in* the world, but rather a way we try to make sense out *of* the world, then different ways of understanding things can indeed lead to different answers.

This point is made emphatically by the French statistician Georges Matheron in his book *Estimating and Choosing*:

> We shall first reject any final assertion concerning the inherent objectivity or subjectivity of "the" probability of an event. The singular form of the word "probability" and the definite article are, strictly speaking, nonsensical ...[GM, p. 27] in applications, there are many ways (all reasonable) to probabilize a given phenomenon, depending on the viewpoint adopted, the aim pursued, etc ... Finally, the notion of probability is of a mathematical, not empirical nature, and thus the notion of objectivity as it is accepted in the positive sciences is not relevant to it. It will of course be objected that what is being debated here is not the theory of probability as such, but its "application to reality." Unfortunately, nobody has ever applied either the theory of probability, or for that matter any other mathematical theory, to reality. One can only "apply" to reality real (physical, technical, etc.) operations, not mathematical operations. The latter only apply to mathematical models of the same nature as themselves. In other words, it is always to probabilistic models, and only to them, that we apply the theory of probability. And the question is to examine whether these probabilistic models may, or may not, have an objective meaning. [GM, p. 27]

Returning to the argument in 3, even though it does not directly invoke Bayes' theorem, one might argue that the reasoning is still Bayesian, since it depends on conditioning the day on the coin-flip. In fact, it seems the interviewer is adopting a Bayesian mindset from the get-go, because from Beauty's perspective, each interview is a unique event, and she is being asked for her degree of belief that this particular interview occurs on a Tuesday. Here's another quote, this one from the translator's preface to Matheron's book:

> Ever since the beginning of modern probability theory in the seventeenth century there has been a continuous debate over the meaning and area of applicability of the concept of probability ... The most popular interpretation of probability among statisticians has been for many years the frequency interpretation, which is quite satisfactory as long as the data represent outcomes of a large number of repeated trials ... My own interest in the subject started with my involvement with probabilistic design techniques and reliability calculations for large civil engineering structures. The interpretation of the concept of probability in that area presented grave difficulties for the following reasons:
>
> (a) On the one hand the objects of study were unique complex structures, so that a frequency interpretation was quite unrealistic.
>
> (b) On the other hand the essentially objective nature of engineering design made the adoption of any purely subjective interpretation repugnant.
>
> It should be noted that exactly the same problem exists in all time series analysis work, since there also the data consists in just the one unique realization of a random function, and repeated trials are usually unavailable ... There is in fact a widely held view that the only possible interpretation of a Bayesian procedure is a subjectivist one. As for myself, I remained in a state of confusion ... Reading Matheron's work was an illumination. Here was a coherent, well thought-out framework for the use of probabilistic models to describe

unique phenomena in a purely objective way. In one sweep, Matheron was able to rid the practice of probability of all the confusing philosophical overtones that had clouded it for decades, and to provide a clear guide for the determination of the suitability and range of applicability of probabilistic modelling. [GM, pp. V-VI]

Appendix 1: a brief overview of conditional probability

For a single event A, we denote the probability of A by $P[A]$, and for two events A and B, we denote the probability of both A and B occuring simultaneously as $P[A\&B]$. Then the *conditional probability* of A given the condition B is defined as

$$P[A\,|\,B] = \frac{P[A\&B]}{P[B]} \tag{17}$$

which is just the ordinary probability of A where the sample space has been resticted to only those events satisfying whatever condition defines B. For example, when rolling a pair of fair dice, let A denote the event "the sum of the two numbers is 5", and let B denote the event "at least one die comes up as 3". Then the probability of A given B is $\frac{2}{11}$:

- there are 36 possible outcomes
- if one die is 3 and the sum is 5, then the other must be 2; there are two ways this can happen - (2,3) and (3,2), therefore $P[A\&B] = \frac{2}{36}$
- there are 11 ways in which at least one die is 3 (6 ways in which the first is 3, plus 6 ways in which the second is 3, take away 1 because the pair (3,3) was counted twice), therefore $P[B] = 11/36$
- $\frac{P[A\&B]}{P[B]} = \frac{2/36}{11/36} = \frac{2}{11}$.

Note that the denominators (36) cancel, so $P[A\,|\,B]$ reduces to the number of ways in which at least one die is 3 and the sum is 5, divided by the number of ways in which at least one die is 3.

A common mistake is to assume the inverse $P[B\,|\,A]$ is the same as $P[A\,|\,B]$, which in general is false. In this example, there are four ways the sum can be 5 $(1+4, 4+1, 2+3, 3+2)$, so $P[A] = \frac{4}{36}$ and therefore

$$\begin{aligned} P[B\,|\,A] &= \frac{P[A\&B]}{P[A]} \\ &= \frac{\frac{2}{36}}{\frac{4}{36}} \\ &= \frac{1}{2}\,. \end{aligned}$$

The conditioning equations follow immediately from the definition (17). For example, if $\tilde{B}$ denotes the negation of B then

$$\begin{aligned} P[A] &= P[A\&B] + P\left[A\&\tilde{B}\right] \\ &= P[A\,|\,B]\,P[B] + P\left[A\,|\,\tilde{B}\right]P\left[\tilde{B}\right] \end{aligned}$$

by (17) and its counterpart, in which $\tilde{B}$ is substituted for B.

Appendix 2: interpreting Bayes' theorem

Suppose we are given a *hypothesis* H and *evidence* E for or against that hypothesis. We denote the negation of H by $\tilde{H}$. and define the *odds* of a given outcome as the ratio of the number of events that produce the outcome to the number that don't. There are a few different (but mathematically equivalent) ways to state Bayes' theorem - the one I prefer is

$$P\left[H \mid E\right] = \frac{1}{1 + \frac{P\left[E \mid \tilde{H}\right]}{P\left[E \mid H\right]} \cdot \frac{P\left[\tilde{H}\right]}{P\left[H\right]}} .$$

- The ratio $\frac{P[H]}{P[\tilde{H}]}$ is called the *prior odds* and is interpreted as the odds of our belief in the hypothesis *before* considering the evidence. Then its inverse $\frac{P[\tilde{H}]}{P[H]}$ will be a large positive number if belief is weak, close to zero if belief is strong, and 1 if no opinion,
- $\frac{P[E \mid H]}{P[E \mid \tilde{H}]}$ is called the *likelihood ratio* and represents the odds of seeing the evidence if the hypothesis is true as opposed to false. Its inverse $\frac{P[E \mid \tilde{H}]}{P[E \mid H]}$ gets closer to zero if the evidence is more likely when the hypothesis is true, becomes large if the evidence is less likely when the hypothesis is true, and equals 1 if the hypothesis and evidence are unrelated.
- The left-hand side $P\left[H \mid E\right]$ is called the *posterior* probability and is interpreted as the degree of belief in the hypothesis *after* considering the evidence. Therefore, (14) tells us that the larger the likelihood ratio, the larger the posterior probability. In other words, Bayes' theorem gives mathematical precision to the universally accepted principle that if one is more likely to see the evidence when the hypothesis is true as opposed to false, then the evidence counts in favor of the hypothesis, and conversely, if one is less likely to see the evidence when the hypothesis is true, then the evidence counts against the hypothesis. Note however that the posterior probability will not be affected all that much if one starts out with a very small prior, and this agrees with what we observe over and over again in everyday life. As Martin Gardner so aptly put it, "it is a rare event when believers of any stripe change their minds about anything".

References

[GM] Georges Matheron, *Estimer et Choisir*, Fascicules du CGMM no. 7, Paris School of Mines (1978), translated into English as *Estimating and Choosing: an Essay on Probability in Practice* by A. M. Hasofer, Springer-Verlag (1988)

[PM] Pradeep Mutalik, Solution: 'Sleeping Beauty's Dilemma', Quanta Magazine, Jan. 2016

[PW] Peter Winkler (2017) The Sleeping Beauty Controversy, The American Mathematical Monthly, 124:7, 579-58