Likelihood Neglect Bias and Mental Simulations:

An Illustration using the Monty Hall Problem

John Wilcox, wilcxoje@stanford.edu
Departments of Psychology and Philosophy
Stanford University

Acknowledgements: James L. McClelland, Nick Ogden, Deigo Solorzano, Sayeh Kohani, Maria del Mar Castaños, Thomas Icard

Contents

Abstract

Humans make inferences on a daily basis, but in doing so, we are susceptible to a range of well-known biases. This paper advocates the explicit recognition of another kind of bias: *the likelihood neglect bias*. To illustrate the bias, this paper present results from an experiment involving the Monty Hall problem. It is uncontroversial that a particular kind of probability—likelihoods—are central to the correct mathematical solution to the problem. For that reason, it is useful to examine the role of likelihoods in how people actually reason about the problem. However, previous psychological studies have not investigated whether correct reasoning about the problem is inhibited by two possible causes: i) unawareness of what the likelihoods are or ii) failure to realize the *implications* of the likelihoods. If the latter error is at play, then humans are susceptible to what, I argue, should regarded as a *likelihood neglect bias*. This bias is a fallacy which, like many others, is defined as a violation of the norms of probability theory, particularly the so-called *law of likelihood*. I also outline a new method of reasoning called the *mental simulations approach*. I then present results indicating that participants did indeed display likelihood neglect bias and that the bias can be corrected by training in the mental simulations approach. Furthermore, the majority of participants that report using the method correctly solve the Monty Hall problem the first time they encounter it. The studies also compare the approach to two other prominent approaches to the Monty Hall problem, one of which involves mental models. The results indicate these other approaches mislead participants into giving incorrect answers in a variation of the Monty Hall problem—called the "New Monty Hall Problem"—where the likelihoods change. I encourage other experimenters to attempt to independently replicate these findings, and I provide details in an appendix to try to facilitate this.

Introduction

We all make inferences on a daily basis—about what decisions to take or about what is true, for example. The heuristics and biases research program has unearthed a range of biases which can compromise such inferences (Gilovich et al., 2002). Some of these are well-known and widely acknowledged. On such example is the conjunction fallacy—a case where one regards a conjunctive statement (such as "A & B is true") as more probable than one of its conjuncts (such as merely "A is true").

This paper has five aims. The primary aim is to advocate the explicit recognition of a kind of bias that is distinct from those that have been commonly discussed. This bias, like many canonical biases, refers to a specific type of violation of the norms of standard probability theory (Gilovich & Griffin, 2002). In particular, the norm is the law of likelihood which, according to one version, is loosely the following: if a hypothesis h_1 makes some evidence more probable than another hypothesis h_2, then the evidence *raises* the probability of h_1 relative to h_2 and, by implication, it *lowers* the probability of h_2 relative to h_1. Consequently, this fallacy, which I call the *likelihood neglect bias,* is committed when an individual is *aware* that the evidence is more probable given one hypothesis in comparison to another, but the individual nevertheless fails to let this affect their probability judgments about the hypotheses in accordance with the law of likelihood. I use the terms "fallacy" and "bias" interchangeably to refer to this departure from the norms of probability theory.

A second aim is to provide, for the first time, unequivocal evidence about whether people commit this fallacy. To achieve this, the current study uses the Monty Hall problem to illustrate the fallacy, and it aims to achieve two further aims in the process.

The third aim is to re-orient the psychological study of the Monty Hall problem. In particular, this paper argues that, given the centrality of likelihoods in correctly solving the Monty Hall problem, future research should direct more attention to understanding the causes of likelihood neglect and to testing interventions to improve people's understanding of likelihoods and their probabilistic implications for the Monty Hall problem.

To this end, the fourth aim of this paper is to experimentally test one new approach to probabilistic reasoning. I call it the 'mental simulations' approach. It builds on existing research and it attempts to remedy likelihood neglect bias.

A fifth and final aim is to compare this method of reasoning to other approaches to the Monty Hall problem. In particular, I want to explore whether these other approaches deliver correct answers in a variant of the Monty Hall problem—called the "New Monty Hall Problem"—where the likelihoods change.

This paper is structured like a typical psychology report.

The rest of this introduction aims to get conceptual clarity about the relevant issues, and it proceeds as follows. First, I describe the Monty Hall problem, as well as the mathematics underpinning the correct response to it. I then briefly review the psychology of the Monty Hall problem, describing current theory about the causes of erroneous responses and outlining some interventions to improve reasoning. Then, attention is turned to a gap in the literature: examining the role of likelihoods in the psychology of responses to the Monty Hall problem. The rationale for the first experiment in this study is then discussed, the likelihood neglect bias is characterized in detail, and arguments are presented for why it should be—and has not been—explicitly recognized as a fallacy or bias. I then evaluate some salient methods of solving the Monty Hall problem. Ultimately, I argue these methods are not reliable since they are not *sensitive* to the relevant probabilities in a way which I will later clarify. I then outline a new approach to probabilistic reasoning called the *mental simulations approach*, and I explain its advantages over some alternative methods.

The introduction is relatively long and highly theoretical. Consequently, the reader might think that this is unusual for a psychology article.

However, I think this is a virtue. If psychologists aim to be experts about reasoning, about the Monty Hall problem and about methods to improve reasoning, then sufficient attention needs to be devoted to understanding the relevant theoretical issues. Consequently, I have aimed to carefully discuss the relevant issues, especially since these may not be familiar to some readers.

I then move on to describe the two main experiments.

1. Experiment 1 aims to test whether participants display the likelihood neglect bias and, if so, whether it can be corrected with the mental simulations approach
2. Experiment 2 compares the mental simulations approach to two popular approaches to the Monty Hall problem, and it aims to see whether the mental simulations approach fares better than these other approaches when participants encounter the New Monty Hall Problem

Again, the reader might think it is unusual for a psychology paper to be concerned with the efficacy of training-style interventions.

Again, however, I think this is a virtue. If psychologists aim to be experts about reasoning, about the Monty Hall problem and about methods to improve reasoning, then they should arguably also be aware of methods which, as a matter of fact, improve reasoning in the Monty Hall problem, even if these methods involve some training. Such methods are of special interest when they are relevant for improving reasoning more generally, as is the case for the method in this paper.

I now move on to discuss the theoretical issues surrounding the Monty Hall problem.

A Description of the Monty Hall problem

The Monty Hall problem is one of the most well-known psychological brainteasers. As one cognitive scientist remarks, it is “the most expressive example of *cognitive illusions* or *mental tunnels* in which even the finest and best-trained minds get trapped” (Piattelli-Palmarini, 1994, pp. 161). There are different versions of the problem, but a fairly standard one is as follows:

> Suppose you are on a gameshow where a prize is randomly placed behind one of three doors with equal probability: door A, door B and door C. Behind the other two doors are goats. You do not know which of the three doors conceals the prize. You are asked to select a door, although that door remains closed for the time being. Monty Hall, the game show host, knows where the prize is, and he will then open one of the other doors you did not chose to show that it concealed a goat. If the door you first selected conceals the prize, he will open one of the other two doors at random with equal probability. If the door you first selected does not conceal the prize, and one of the other two doors does, then he will open the one unselected door that does not conceal the prize.
>
> So suppose you play the game, selecting a door, and then Monty Hall opens one of the other doors. For example, suppose you select door A, and then Monty Hall opens door C to show you a goat.
>
> Here is the key question: should you switch doors from your initial door (door A) and instead opt to open the other unopened door (door B)?

The puzzle arises because, according to academic consensus, you should switch doors to maximize the probability of obtaining the prize: it is more probable that door B conceals the prize.

An Explanation of the Correct Response and the Law of Likelihood

The reason you should switch doors is explained by standard probability theory, the mathematics of which is uncontroversial. To explain the mathematics clearly, though, we first need to clarify some probabilistic concepts and notation.

Let $P(X)$ stand for the initial probability of some door X concealing the prize. For example, $P(B)$ is the initial probability that door B conceals the prize, and it is 1/3, just like the probabilities that the other doors conceal the prize. So we have three hypotheses: *A*, that door A conceals the prize; *B*, that door B conceals the prize; and *C*, that door C conceals the prize. Each hypothesis has a probability of 1/3, so $P(A) = P(B) = P(C) = \frac{1}{3}$.

Now, we are ultimately interested in the probability of these three hypotheses *after we receive the evidence that door C was opened and concealed a goat*. So let the symbol 'c' represent the proposition that door C was opened by Monty Hall. In this case, we are interested in $P(B|c)$—that is, the probability that door B conceals the prize given that door C was opened by Monty Hall.

According to standard probability theory, this probability can be calculated with Bayes' theorem (which we will soon see below). Bayes' theorem states that this probability is calculable via a ratio involving two kinds of probabilities.

One of these is the *prior probability*, the probability for a hypothesis prior to receipt of some evidence—such as the evidence that door C was opened. In this case, the prior probabilities are $P(A) = P(B) = P(C) = \frac{1}{3}$.

The other kind of probability is called a *likelihood*. These probabilities are the focal subject of this paper. A likelihood has acquired a specific and technical meaning in the statistical and philosophical literature: it refers to the probability of some evidence given some hypothesis (Hacking, 2016; Hawthorne, 2018; Bandyopadhyay, 2011). This usage of the term departs from convention: the common public—and many psychologists—talk of a "likelihood" as though it is interchangeable with any kind of "probability", but this is not the case in the probabilistic literature. In this sense, a likelihood is not to be confused with the probability of a hypothesis given some evidence (a confusion known as the *confusion of the inverse*).

To illustrate the difference between a probability and its inverse, consider the probability of being a human given that one is a doctor. The probability of being a human given that one is a doctor is 100% (since all doctors are humans). But the probability of being a doctor given that one is a human is much lower (since many humans are not doctors). In this technical sense, a likelihood is the probability of the *evidence given that some specified hypothesis is true*, and *not* the probability that the hypothesis is true given that evidence.

Let us now apply these concepts to the case of the Monty Hall problem.

Recall that we are interested in $P(B|c)$—the probability that door B conceals the prize given that door C was opened and concealed a goat. According to Bayes' theorem, this probability is specified by the equation whose right-most terms are prior probabilities and likelihoods:

(1)

$$P(B|c) = \frac{2}{3} = \frac{P(c|B)P(B)}{P(c|A)P(A) + P(c|B)P(B) + P(c|C)P(C)}$$
$$= \frac{1 \times \frac{1}{3}}{0.5 \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}}$$

Of course, the reader does not need to know everything about this theorem for the purposes of this paper, but they do need to note three points.

The first point is that, according to Bayes' theorem, the probability that door B conceals the prize is $\frac{2}{3}$, twice the probability that door A conceals the prize.

The second point, and one which the public is often unaware of, is that Bayes's theorem delivers the *correct* verdict that $P(B|c) = \frac{2}{3}$. There are many arguments for why this is the case (Rosenhouse, 2009), but perhaps the most straight-forward argument is that, when running numerous computer simulations of the Monty Hall problem as described above, door B will contain the prize approximately $\frac{2}{3}$ of the time—not $\frac{1}{2}$ of the time, as many think. In appendix C, I provide some code so that the reader can themselves run simulations of the Monty Hall problem.

In any case, the fact is that Bayes' theorem delivers the right verdict, but the question arises as to why it delivers that verdict, and this is the third point which the reader needs to note: Bayes' theorem delivers the verdict that it does largely because of the *likelihoods*. Notice that, in equation (1), the values of the prior probabilities of the hypotheses are all the same: $P(A) = P(B) = P(C) = \frac{1}{3}$. The hypotheses differ *only* in that the evidence is more probable under one hypothesis compared to the others.

Let us see how this is so for each hypothesis, starting with C. If door C concealed the prize, then Monty Hall would never open it. So the hypothesis that door C conceals the prize gives a probability of 0% that door C would be opened: $P(c|C) = 0$. Consequently, Bayes' theorem gives the probability that door C conceals the prize a probability of 0% conditional on door C being opened.

If door A concealed the prize, however, then Monty Hall could have opened either door B or door C with equal probability. So the hypothesis that door A conceals the prize gives a probability of 50% (or 0.5 in decimal notation) that door C would be opened: $P(c|A) = 0.5$.

In contrast, if door B conceals the prize, and you select door A, then Monty Hall *must* open door C. So the hypothesis that door B conceals the prize gives a probability of 100% that door C would be opened: $P(c|B) = 1$, twice that of the likelihood given by door A. As a result, the probability that door B conceals the prize given that door C is opened ultimately becomes twice the probability of door A: $\frac{2}{3}$ compared to $\frac{1}{3}$.

This, then, illustrates a general and uncontroversial theorem about probabilistic reasoning: the *law of likelihood*. There are various statements of the law (Hacking, 2016, chap. 5; Hawthorne, 2018), but the following is a relatively simple version of it for our purposes. Let h_1 and h_2 stand for two distinct and mutually exclusive hypotheses, meaning that they cannot be simultaneously true. Let *e* stand for some evidence. Furthermore, let $0 < P(h_1) < 1$ and $0 < P(h_2) < 1$, meaning that neither hypothesis has a probability of 0% or 100%. Then, the law of likelihood specifies that:

(2)

$$\text{If } P(e|h_1) > P(e|h_2), \text{ then } \frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)}$$

Proof of this theorem can be found in appendix C.

Informally put, what this means is that if one hypothesis h_1 makes the evidence *e* more probable than another hypothesis h_2—or, in other words, if the likelihood of the evidence given h_1 is greater than the likelihood of the evidence given h_2—then the evidence *raises* the probability of h_1 relative to h_2 and, by implication, it *lowers* the probability of h_2 relative to h_1.[1]

[1] Note that some philosophers, such as Hawthorne (2018), speak of the "likelihood of the evidence", while others, such as Nola (2013), speak of the "likelihood of the hypothesis" to refer to the same thing. I have followed Hawthorne as I think this is a less confusing way of speaking about likelihoods.

To apply this to our example in the Monty Hall problem, it is more probable that door C would be opened if door B concealed the prize than if door A concealed the prize, so $P(c|B) > P(c|A)$. Consequently, the evidence *raises* the probability of door B concealing the prize relative to door A concealing the prize, since the probability for door B goes from $\frac{1}{3}$ to $\frac{2}{3}$ while the probability for door A remains at $\frac{1}{3}$. Of course, this does not lower the probability of door A concealing the prize in absolute terms: it was $\frac{1}{3}$ both before and after door C was opened. Instead, it lowers the probability of door A concealing the prize *relative* to the probability of door B concealing the prize: both probabilities were equal, but now one probability is twice the size of the other.

It is important to dispel two misconceptions about the law of likelihood. The first is that it is a specifically *Bayesian* phenomenon—that only Bayesians endorse the principle, or that frequentists and other non-Bayesians do not adopt the principle. This is incorrect: Bayes' theorem is universally accepted in probability theory (Weirich, 2011, 234). What differentiates Bayesians and non-Bayesians is not how Bayes' theorem relates one probability to other probabilities. Rather, it is about such matters as to how probability statements are to be interpreted, how probabilities relate to degrees of belief and how Bayes' theorem is to be used. So the claims about the law that are made in this paper do not concern only Bayesian reasoning; they concern *all* probabilistic reasoning. The second misconception is that this principle only applies in toy cases, like the case of Monty Hall, but not in real world problems. Again, this is incorrect. It applies to any case where it is meaningful to talk about the probability of the evidence given competing hypotheses, and such cases arguably arise in many parts of life: in medical diagnosis, in scientific contexts and in others.

Regardless, the main point is that Bayes' theorem correctly delivers the verdict that it does because of facts about likelihoods, and it also entails the law of likelihood.

The Psychology of the Monty Hall Problem: Causes of Incorrect Responses

The preceding discussion gives an account of the correct solution to the Monty Hall problem, but we now turn to consider how people actually do reason about the Monty Hall problem.

As alluded to earlier, the correct response to the problem is to switch doors, but studies have generally reported that most participants give incorrect responses and prefer to stay with their initial selection. For example, Burns and Wieth (2004) surveyed thirteen studies of the participants responses to the standard Monty Hall problem. They found that switch rates ranged from 9% to 23%, with a mean of 14.5% (*SD* =4.5). As they note, the consistency of low switch rates is also remarkable given that the studies varied with respect to problem wording, methods of presenting the problem and the cultures and languages of the participants. Furthermore, even educated participants have faltered on the problem. Marilyn vos Savant initially popularized the problem by writing on it in the *Parade* magazine in 1990, and she found that 65% of respondents writing with university addresses disagreed with the correct solution to the problem (vos Savant, 1997). Additionally, Schechter (1998, 108-109) relates how even Paul Erdõs, a renowned mathematician of the 20th century, initially disagreed with the correct solution to the problem. Apparently, he only changed his mind after a computer simulation of the problem convinced him it was correct.

So what then are the causes of such responses? Some comprehensive literature reviews address this question (Saenen et al., 2018; Tubau et al., 2015), and two prominent causes are discussed by Tubau et al. (2015): 1) emotional-based choice biases and 2) cognitive limitations in understanding and representing probabilities. The first cause relates to how participants are averse to switching from their first choice. Some suggest this aversion this arises from an illusion that their first choice can favorably influence the outcome (Granberg & Dorr, 1998) or alternatively from a desire to reduce regret—a regret which is higher if one loses after switching than losing after sticking with their initial choice (Petrocelli et al., 2011). The second cause relates to the frequent observation that participants commonly believe that the probabilities of the two outcomes are equal given the observation that two options remain (De Neys & Verschueren, 2006; Tubau et al., 2003).

The New Monty Hall Problem: Why Some Popular Solutions Do Not Work

A range of interventions have been tested to determine whether they improve reasoning in the Monty Hall problem. These are reviewed comprehensively by Saenen et al. (2018).

Of course, there are many putative "solutions" to the Monty Hall problem—that is, attempts to 1) give the correct answers and 2) explain why those answers are the correct ones. So far, the only one which delivers uncontroversially correct answers is the solution with Bayes' theorem. But other putative solutions are also popular. While it is impossible to review all of them here, I argue that there are theoretical reasons to think that two particular approaches to the problem are unreliable.

I will first outline these two approaches and then examine how they are unreliable.

Possible Models Approach

One approach is a particular mental models approach discussed by Krauss & Wang (2003) and Tubau et al. (2003). Their solution first involves entertaining various "possibilities" about where the prize might be (or various "mental models", as they say). Then, one calculates the frequency with which switching doors would win the prize among those possibilities. In particular, these are the two sets of mental models which Krauss and Wang (2003) present in their solution to the Monty Hall problem. These are represented in their figures and tables below.

Figure 1: Krauss and Wang's (2003) reproduced explanation of why participants should switch

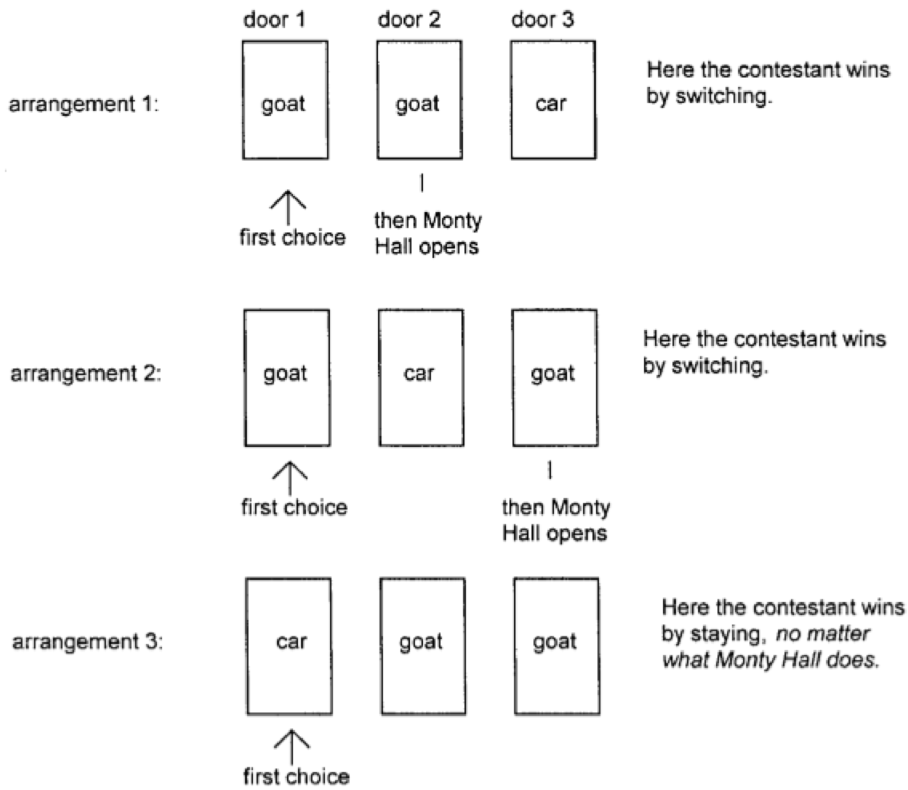


Figure 1. Explanation of the solution to the Monty Hall problem: In two out of three possible car–goat arrangements the contestant would win by switching; therefore she should switch.

Table 1: Krauss and Wang's (2003) reproduced explanation of why participants should switch

Table 1
Mental Model Representation of the Monty Hall Problem

Mental model	Door 1 (chosen door)	Door 2	Door 3
1	car	open	
2	car		open
3		car	open
4		car	open
5		open	car
6		open	car

Note. Based on mental models from Johnson-Laird et al. (1999).

The idea here is that one can both obtain the correct probabilities *and* understand why they are the correct probabilities by considering the frequency with which switching yields the prize among either of these two sets of mental models.

Tubau and Alonso (2003) also report a similar experiment that involves getting participants to consider various possibilities and count the frequency with which switching yields a favorable result.

That, then, is one prominent approach to solving the problem. For ease of reference, let us call it the *possible models* approach.

Probability Accrual Approach

There is also another way to approach the problem. Rather than entertaining all the possibilities, this approach focuses on the fact that the non-selected doors conceal the prize two thirds of the time—or, equivalently, these doors have a 2/3 prior probability of concealing the prize. Of course, once Monty Hall opens one of the unselected doors, it becomes clear that that particular door does not conceal the prize. The special part of the approach, however, is to claim that the probability of 2/3 then *accrues* to the other unselected door. Put simply, the reasoning is this: the unselected doors conceal the prize with a prior probability of 2/3, but once one of these doors is opened, there is then a 2/3 probability that the only unselected door conceals the prize. Call this the *probability accrual approach*.

Such reasoning seems to be endorsed by Tubau et al. (2015) as a promising approach. They discuss why humans are vulnerable to the "equiprobability illusion" in the Monty Hall problem (that is, the illusion that it is equally probable that both unopened doors conceal the prize):

> In particular, susceptibility to the illusion is caused by a weak representation of the facts that: (a) the non-selected doors will hide the prize 2 out of 3 times, (b) among the non-selected doors it is certain that at least one is null, and (c) this null option will always be eliminated.(Tubau et al., 2015, 8)

The implication of this statement seems to be that if participants are sufficiently aware of conditions (a)-(c), then they will not be susceptible to the equiprobability illusion. This seems to implicitly reflect the probability accrual approach: if participants were sufficiently aware that the unselected doors conceal the prize two thirds of the time and that the opened unselected door does not conceal the prize, then they would not think the remaining doors are equally likely, but would instead think that the one unselected and unopened door conceals the prize with a probability of 2/3.

Similar reasoning has been reported by participants in my pilot experiments. One such participant had previously studied the Monty Hall problem in a course well before the experiment. They correctly answered that the probability that door B concealed the prize is 2/3. Their justification for this probability appeared to reflect the probability accrual approach:

> Essentially by choosing door A and switching, I'm choosing both doors B and C. It's just that I know one of the two won't have the prize. But that means switching still increases my chances of winning from 1/3 to 2/3

The Problem with These Approaches: Safety

The problem with these two approaches, however, is that they lack *safety*.

Safety is a concept which I borrow from discussions in epistemology. There, safety is occasionally advocated as a *necessary condition* for a belief to qualify as knowledge. In brief, an agent has a safe belief in some proposition *p* when the following condition holds:

> If the agent were to believe that *p*, then *p* would not be false. (Ichikawa & Steup, 2018; Sosa, 1999)

In epistemology, some claim that a belief does not count as knowledge if it is unsafe.

Let us consider examples of safe and unsafe beliefs, examples adapted from Bertrand Russell (1948). Suppose someone looks at a functional clock which says it is 12pm, and they then form the belief that it is 12pm. This belief is safe in the sense that if the agent believes that it is 12pm, then it would not be false that it is 12pm. And it is safe as such because it is based on a reliable method of getting at the truth—namely, telling the time from a functional clock. Contrast this to the following example of an unsafe belief. Suppose someone looks at a clock which says that the time is 12pm, but the clock is *broken* and so it *never* changes with the actual time. The person does not know it is broken, and they then form the belief that it is 12pm. Coincidentally, it happens to be 12pm, and so their belief is true. However, their belief is not safe: if they used this method to arrive at their belief that it is 12pm, then it is not necessarily the case that it is 12pm rather than some other time. After all, they could have looked at the clock when it was 1pm and then falsely concluded that it was 12pm. The safety of a belief is closely related to how sensitive that belief is to the truth, or how well the belief tracks the truth (although sensitivity and truth-tracking have both taken on technical meanings and are surrounded with controversy in epistemology).

The concept of safety highlights the importance of using methods that get us to the truth, not by chance, but because those ways in some sense *track the truth*. A broken clock is generally not a good way of forming beliefs about time, even if it happens to get one to the right beliefs in some circumstances.

In this context, I claim that the aforementioned probabilistic methods are unsafe in this particular sense—they do not track the truth about the probabilities. More specifically, if someone were to use those methods to arrive at judgments about probabilities, then those judgments about probabilities are not necessarily true. Instead, those judgments about probabilities could be false.

The New Monty Hall Problem

I will illustrate this with the following scenario which I call the *New Monty Hall problem.* This version is exactly the same as the original Monty Hall problem in all respects except this: if you select a given door and it conceals the prize, then Monty Hall has a *90% probability of opening* *the right-most door* that is unselected and does not conceal the prize. In this case, if you select door A, and if door A conceals the prize, then Monty Hall is going to open door C with a 90% probability or door B with a 10% probability. Hence, the likelihoods change. Suppose you select door A and Monty Hall opens door C. By Bayes's theorem, it then follows that door A and door B have a near equal probability of concealing the prize:

$$\begin{aligned} P(B|c) &= \frac{P(c|B)P(B)}{P(c|A)P(A) + P(c|B)P(B) + P(c|C)P(C)} \\ &= \frac{.9 \times \frac{1}{3}}{1 \times \frac{1}{3} + .9 \times \frac{1}{3} + 0 \times \frac{1}{3}} \\ &= \frac{10}{19} = .53 \end{aligned}$$

So we have a case where the likelihoods change, and so too do the posterior probabilities about doors A and B concealing the prize. In this case, it *is nearly just as rational to stay* with your choice of door A as it is to switch. And we know this via *exactly the same mathematical machinery* that tells us to switch in the original Monty Hall problem scenario—namely, Bayes theorem.

To reinforce this point, I ran 100,000 computer simulations where you select door A, Monty Hall opens door C and Monty Hall had a 90% chance of opening door C if door A concealed the prize. Door B concealed the prize about 53% of the time—not 66% of the time. I have included my code in appendix C so that others can reproduce this result independently if they wish to do so.

Note, however, that even though the probabilities have changed, it is not obvious that one would get the right answer if they were to follow the possible models and probability accrual approaches above. In other words, they might *falsely still conclude that the probability that door B conceals the prize is 2/3.*

Let us see how this is so. Consider the possible models approach first. Note that *all* of the possibilities are exactly the same. Nothing has changed about the space of possible models. For that reason, the counting procedure of the possible models approach would generate exactly the same probabilities.

And consider the probability accrual approach. In the New Monty Hall problem, the following facts still obtain: "(a) the non-selected doors will hide the prize 2 out of 3 times, (b) among the non-selected doors it is certain that at least one is null, and (c) this null option will always be eliminated"(Tubau et al., 2015, 8). It seems that according to Tubau et al. (2015, 8), awareness of such facts would stop participants from being susceptible to the "equiprobability illusion".

However, in the New Monty Hall problem, there is a *near* equal probability that the doors conceal the prize, but it is doubtful that the probability accrual approach would by itself help participants realize this.

Of course, one might think that even if the possible models and probability accrual approaches do not get the right answer in the New Monty Hall problem, perhaps we could modify or supplement them so that they do.

That might be a possibility, and I make no claims about its prospects. Instead, I merely claim that, *as they currently stand*, the possible models and probability accrual approaches do not help participants get to such judgments, and I present experimental findings later on that demonstrate this is the case.

Further, I claim that any safe approach to the Monty Hall problem should be sufficiently sensitive to the *likelihoods*: if the likelihoods change, then so will the posterior probabilities that are recommended by the approach. I later outline one approach that is sensitive as such: the mental simulations approach. And I provide some experimental evidence that it improves reasoning in the Monty Hall problem—by improving both the correctness of participants' answers and their understanding of why those answers are correct.

The Gap in the Literature, and the Likelihood Neglect Bias

To review, I have briefly examined some of the literature which examines causes of incorrect judgments as well as some interventions which aim to improve reasoning, and I have introduced the New Monty Hall Problem that may allow us to probe for evidence of unsafe (potentially incorrect) reasoning and provide ways to correct it.

Before turning to experimental investigations, I note that there is an important gap in the literature, which the investigations I propose may allow us to fill. Recall that Bayes' theorem delivers the correct verdict that it does because of the likelihoods which feature in the calculation. Because of the centrality of likelihoods to the correct solution, it is sensible to ask whether likelihoods play some role in 1) understanding the causes of incorrect judgments or 2) designing potential interventions to improve reasoning. This paper explores uncharted territory in investigating these two issues.

More specifically, likelihoods may relate to incorrect responses in two distinct ways.

The first possibility is that participants might be *unaware* of the relevant likelihoods. This may be for various reasons. One is that various presentations of the problem are unclear or insufficiently specified. Perhaps it is not clear to participants that there is, for example, really is an equal chance that Monty Hall would open door B or door C if door A concealed the prize, as opposed to Monty Hall always opening the rightmost door every time the game is played. In Krauss and Wang's (2003) first experiment involving the problem, for example, participants are not specifically told that if the selected door conceals the prize, then Monty Hall would open either of the remaining doors with an equal probability in this case. A related possibility is that participants might be unaware of the relevant likelihoods, perhaps because they rush over the materials, even if the likelihoods are specified. In either of these cases, the incorrect response might be avoided if participants were simply made *aware* of what the relevant likelihoods are.

However, the second possibility is that participants do not realize the *probabilistic implications* of the likelihoods, even if they are aware of the likelihoods.

So these are two different ways that likelihoods may have an explanatory role in understanding causes of the incorrect responses:

i. participants are unaware of the likelihoods
ii. participants do not realize the probabilistic implications of the likelihoods

This then brings out a second issue about interventions to improve reasoning. In particular, if participants are simply unaware of the likelihoods (as per the first way), then a promising intervention may be to ensure they are aware of them before we obtain their response to the Monty Hall problem. Otherwise, a different kind of intervention is necessary. This study aims to investigate these two issues by testing whether participants are simply unaware of the likelihoods or, instead, do not realize their probabilistic implications, even when they are aware of them.

From my review of the literature, and that of Saenen et al. (2018), it appears that this is the first study to explicitly assess participants awareness of likelihoods.

If participants are aware of the likelihoods, but fail to realize their probabilistic implications, then they are arguably demonstrating a *likelihood neglect bias* or *likelihood neglect fallacy*. More precisely, we could define the fallacy so that it is the violation of the law of likelihood. Recall the law of likelihood:

$$\text{If } P(e|h_1) > P(e|h_2)\text{, then } \frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)}$$
$$\text{where } 0 < P(h_1) < 1,\ \ 0 < P(h_2) < 1 \text{ and } \neg(h_1 \ \&\ h_2)$$

Then, the likelihood neglect fallacy occurs just in case an individual judges that $P(e|h_1) > P(e|h_2)$, but they do not judge that $\frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)}$. For example, in the Monty Hall problem, if participants are aware that $P(c|B) = 1$ and $P(c|A) = 0.5$ (and so $P(c|B) > P(c|A)$), but they nevertheless think that $P(A|c) = P(B|c) = \frac{1}{2}$ (and so $\frac{P(B|c)}{P(A|c)} = \frac{P(B)}{P(A)}$), then they would be committing the likelihood neglect fallacy.

So that is a characterization of the fallacy, but at this point, two questions arise: first, does this really deserve to be called a fallacy and, second, if it is a fallacy, then why has it not already been recognized as such?

For the first question, my response is that it deserves to be called a bias or a fallacy no less than many widely acknowledged biases and fallacies. As Gilovich and Griffin (2002, pp. 3) state, many of the canonical biases were “defined” as “violations of basic laws of probability”, an example of this being the conjunction fallacy in probabilistic judgment. And here, the likelihood neglect fallacy is defined precisely as a violation of a law of probability, and one that reflects errors in how people may reason about the Monty Hall and perhaps other matters. Furthermore, one may think the likelihood neglect fallacy deserves its title especially since it is analogous to the base rate neglect fallacy. The base rate neglect fallacy was originally conceived of as occurring when people were *aware* of the base rates, but they did not properly "integrate" them into their judgments, although it was hypothesized that this happened because other information was deemed more relevant (Bar-Hillel, 1980, p. 211). Similarly, the likelihood neglect fallacy occurs when one is *aware* of the likelihoods, but fails to properly integrate them into their judgments about the relevant probabilities, even if this might happen for reasons which are distinct to the base rate fallacy. In any case, the name 'likelihood neglect fallacy' is presumably apt, if only because it certainly involves neglect of the implications of the likelihoods and it certainly involves an error or failure of inference in some way. So why not just call it what it is—a fallacy and a bias?

In response to the second question, it is not clear to me why this has not already been explicitly recognized as fallacy. Perhaps one answer is that, although the law of likelihood is known to numerous formal philosophers and statisticians, it is not known to many people in general, and this might explain why incorrect reasoning about the Monty Hall problem is so prevalent, even among many educated individuals. Furthermore, psychologists studying the Monty Hall problem seem to have not mentioned the law of likelihood at all, perhaps because they are more focused on how people reason about the problem rather than the mathematical details that are of primary interest to statisticians and formal philosophers. But even if we do not know why it has not been recognized as a fallacy, this should not count against it qualifying as such. Every recognized fallacy was first recognized at some point, and many of these were defined by their violation of laws of probability. Of course, it is possible that someone recognized—and provided evidence of—the fallacy earlier, but this seems improbable in this case, especially since the fallacy has not been explicitly recognized in any of the psychological literature which I have seen on the Monty Hall problem, including the comprehensive review by Saenen et al. (2018).

Of course, the likelihood neglect fallacy is not the only issue in the Monty Hall problem. Even if someone avoids the bias and conforms to the law of likelihood, they might not know how to calculate the relevant probabilities. In any case, though, the point is merely that the Monty Hall problem illustrates the likelihood neglect bias, a bias which—as a discuss later—may arise in other important contexts as well.

The Mental Simulations Approach to Probabilistic Reasoning

In this section, I introduce what I call the 'mental simulations' approach to overcoming the likelihood neglect fallacy. The approach has two purposes:

1) Facilitating the *calculation* of the correct probabilities
2) Providing an *intuitive justification* of the correct probabilities

By 2), I essentially mean helping humans to obtain an *intuitive grasp* of *why* the law of likelihood is correct and *why* the result from Bayes' theorem is correct. I will note how the approach achieves these purposes once it has been outlined.

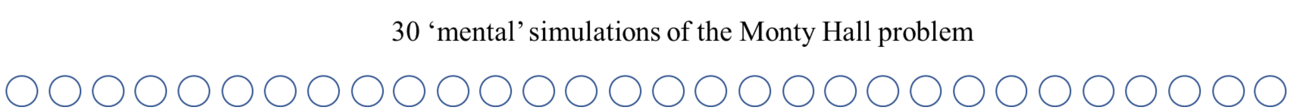
The approach relies on two insights. The first is that computer simulations have successfully convinced skeptics of the correctness of the verdict delivered by Bayes's theorem; take the aforementioned example of mathematician Paul Erdõs, for example. The second is that reasoning in the population has occasionally been improved when probabilistic reasoning is carried out in terms of natural frequencies instead of probabilities (Hoffrage et al., 2015).

The mental simulations approach then combines these two insights. It involves participants running "mental" simulations of the probabilistic set up to thereby convert the probabilities into natural frequencies. Then, the probabilities of interest can be calculated by counting outcomes among these simulations. It essentially is just an attempt to apply and generalize some of the insights of Hoffrage et al. (2015).

An Example of Mental Simulations in the Monty Hall Problem

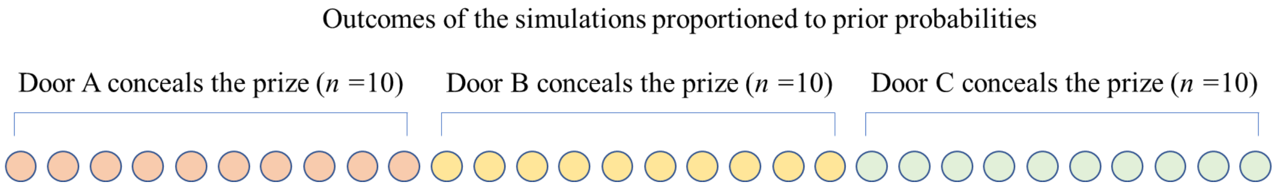
Let us go through this with an example before outlining the approach in more abstract generality. Consider the Monty Hall problem. Let us run 30 mental simulations, that is, let us imagine 30 situations in which the Monty Hall problem is played (and I will explain why I have chosen the number 30 shortly). To help visualize this, let us denote these 30 simulations with circles, imagining that each circle represents a scenario where the Monty Hall problem is played out.

Figure 2: Generating Simulations in Step One



Now we know that the prior probability of any door containing the prize is 1/3, so $P(A) = P(B) = P(C) = 1/3$. The first step, then, is to translate these probabilities into natural frequencies among the 30 simulations—that is, to divide up these simulations so that the proportion of times that a prize is in a location corresponds to the probability that that location would conceal the prize. For example, since each door has a 1/3 probability of concealing the prize, a given door conceals the prize in a third of simulations:

Figure 3: Proportioning Simulations in Step Two



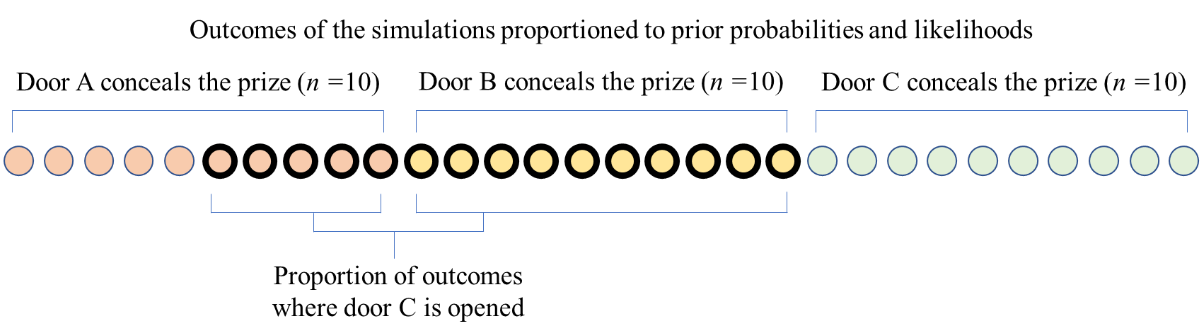
So now our simulations correspond to the prior probabilities: each outcome has a prior probability of 1/3, and each outcome is true in 1/3 of the simulations.

Now, we are going to translate the likelihoods of the evidence we observed into natural frequencies among these simulations. Recall that our evidence is *c*: that door C was opened after door A was selected, and it concealed a goat. Recall also the likelihoods:

- $P(c|A) = \frac{1}{2}$, meaning that door C has a 50% probability of being opened if door A concealed the prize.
- $P(c|B) = 1$, meaning that door C has a 100% probability of being opened if door B concealed the prize.
- $P(c|C) = 0$, meaning that door C has a 0% probability of being opened if door C concealed the prize.

We then translate these probabilities into proportions of outcomes where the evidence *c* is true. For example, door C has a 50% probability of being opened if door A concealed the prize, so door C will be opened in 50% of the simulations where door A conceals the prize. Likewise, we then note the proportion of the other simulations where door C would be opened given the respective outcomes.

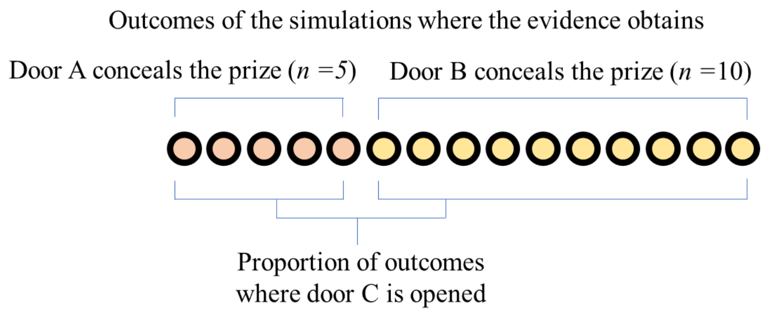
Figure 3: Proportioning Simulations in Step Three



The outcomes where the evidence c is true are denoted by the circles with bolded outlines. We can see that the proportion of outcomes where the evidence is true is proportional to the likelihoods.

So now suppose we want to calculate the probability of door B concealing the prize given the evidence. To do this, we simply eliminate the simulations where the evidence does not obtain, and we then determine the proportion of remaining simulations where door B conceals the prize:

Figure 4: Eliminating Irrelevant Simulations in Step Four



We can see that the remaining proportion of outcomes where door B conceals the prize is $\frac{10}{15}$, or $\frac{2}{3}$. This aligns precisely with the verdict of Bayes's theorem: the probability that door B conceals the prize given that door C is opened is $\frac{2}{3}$, so $P(B|c) = \frac{2}{3}$.

A Generalized Characterization of the Mental Simulations Approach

That, then, is an example of how to implement the approach.

Let us now characterize it in general detail. To calculate the probability of a hypothesis (or an outcome) given some evidence:

1. Generate simulations:
    - Imagine *n* number of simulations (and I will discuss what value *n* may take shortly)
2. Proportion according to priors:
    - For each possible outcome, make the proportion of simulations where that outcome is true correspond to the prior probability of that outcome
3. Proportion according to likelihoods:
    - For each set of simulations corresponding to a given outcome, make the proportion of simulations where the evidence obtains correspond to the likelihood of that evidence given that outcome
4. Eliminate irrelevant simulations:
    - Eliminate the outcomes where the evidence does not obtain
5. Calculate probabilities:

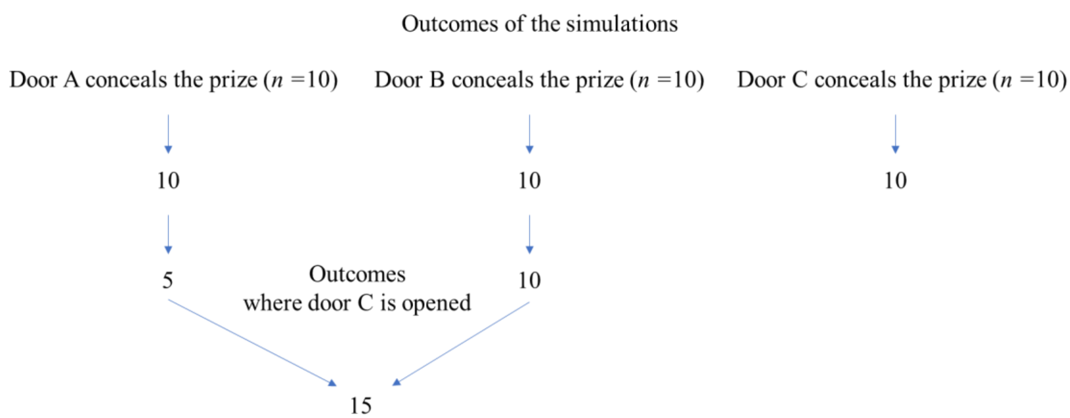
- Determine the proportion of simulations where a particular outcome is true; this is the probability of that outcome given the evidence

There are three points to note about this approach.

The first is that any choice of *n* is appropriate so long as it allows for the appropriate division of simulations by the probabilities. One will not be able to make door A conceal the prize in a third of all simulations if there are four simulations altogether: one cannot divide four into three equal sets in this case. But one could carry out the above approach with simulations of 6 or any multiple of it.

The second is that the depiction of the simulations does not matter either. For instance, the simulations could be represented with non-negative integers instead of circles corresponding to those numbers, like the following:

Figure 5: Mental Simulations Using Numbers



A third point to note is that the approach makes the idealizing assumption that the proportions of scenarios that would be observed correspond perfectly to the underlying probabilities. If an outcome has a 50% probability, for example, then it occurs in exactly 50% of the simulations at the beginning. In real life situations and in experiments, this assumption is often false: if a truly fair coin is tossed 10 times, the proportion of times it lands heads will often not be exactly 50%--it may land heads 2 out of 10 times. Regardless, this assumption is useful in the mental simulations approach because it facilitates the calculation of proportions and probabilities. The assumption also enables the simulation of finite scenarios to reflect the proportion of outcomes of a probabilistic set up if they went to infinity. In other words, as the number of simulations increases, the proportion of times that the outcome will be observed given some evidence will tend to more and more closely approach the result yielded by the finite simulations in the Mental Simulations approach.

So that is an outline of the approach.

Strengths and Limitations of the Mental Simulations Approach

This approach has four strengths.

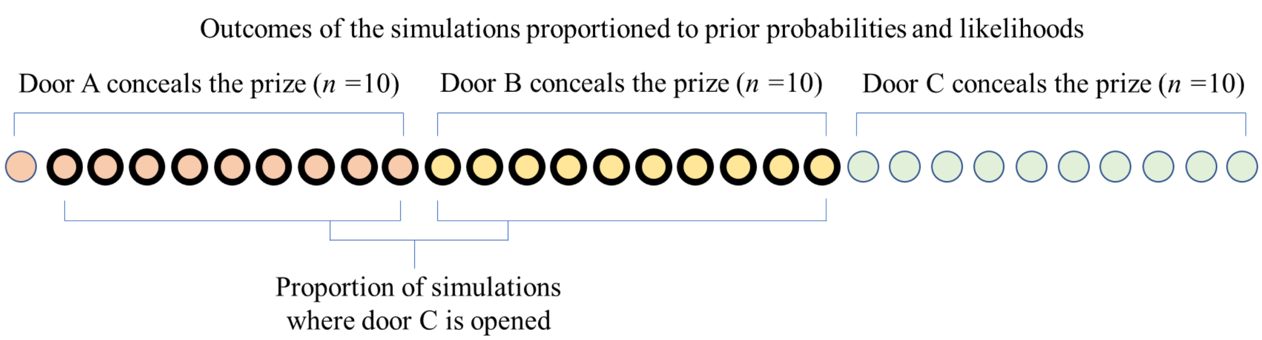
One strength is that, if carried out correctly, it ensures correct reasoning in accordance with Bayes's theorem, and it thus eliminates neglect of likelihoods and prior probabilities or base rates. As such, it also removes confusion of the inverse fallacy. Proof of this can be found in appendix B. Consequently, it achieves its first aim of helping humans to correctly calculate probabilities.

A further strength is that it can help people to understand *why* the probabilities are the correct ones. For example, from the simulations approach and the correspondence between its verdicts and what happens in the limit, we can make sense of why there is a probability of $\frac{2}{3}$ that the other door conceals the prize: because if the probabilistic set up was run an infinite number of times, then the other door would conceal the prize in two thirds of those times.

Another strength is that, as mentioned earlier, it draws on the greater facility that humans have with reasoning in terms of natural frequencies rather than probabilities, thus potentially making it easier for people to implement than calculations with Bayes's theorem.

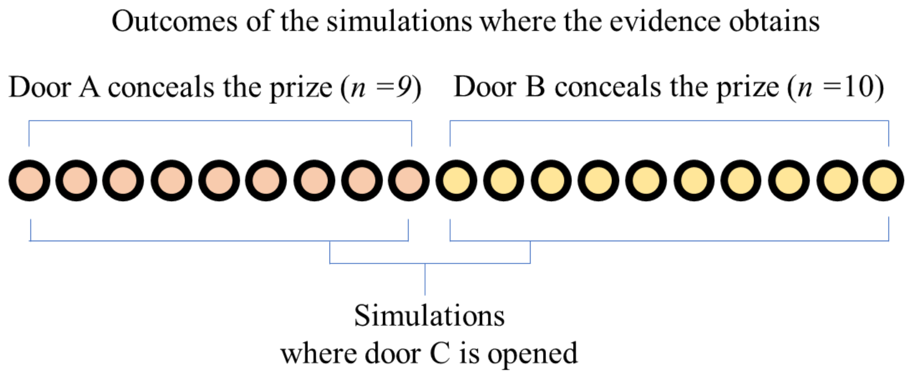
A final strength emerges relative to alternative approaches, such as the possible models and probability accrual approaches. In particular, it can be adapted when the likelihoods take on different values. Recall the New Monty Hall problem, for instance. There, the probability that door C would be opened if door A concealed the prize was instead 90% rather than 50%. Then, without having to perform laborious calculations with Bayes's theorem, we can then determine the probability that door B conceals the prize: just adjust the proportion of simulations where the evidence obtains and where door A conceals the prize to 90%.

Figure 6: Mental Simulations in the New Monty Hall Problem



Then eliminate the simulations where door C has not been opened:

Figure 6: Step Four in the New Monty Hall Problem



We can then see that the probability that door B conceals the prize is $\frac{10}{19}$, again the probability recommended by Bayes's theorem. Contrast this to the possible models approach. The possible models approach calculates only the proportion of times that switching would win the prize among a particular set of possible outcomes, and this proportion among this set is $\frac{2}{3}$ regardless of whether the likelihood $P(c|A)$ is 50% or 90% (Krauss & Wang, 2003).

But the mental simulations approach also has its limitations. For one, it requires some training: this is not a technique that would come to someone naturally without any instruction. For another, it may be difficult to run simulations in cases where the probabilities are very small or go to many decimal places, since it would require mental operations with a potentially large number of simulations. That said, one could perhaps use non-negative integers to represent the simulations in order to better handle such cases.

In any case, every approach has its strengths and weaknesses, and perhaps the strengths of this approach outweigh its limitations. Regardless, I think the strengths of the approach are such that it merits further testing.

Mental Simulations and the Mental Models Approach

One might wonder how the mental simulations approach relates to a prominent theory in the psychology of thinking: the mental models approach.

The mental models approach may mean different things to different authors.

Regardless, we can compare the mental simulations approach to at least some ways of understanding mental models. Like mental models theory, the mental simulations approach involves generating mental entities that stand in for ways that the world might possibly be—or ways that it possibly could have been before we received some evidence about it. We then call these mental entities "simulations", and they resemble mental models in this respect.

However, unlike the mental models theory (at least on some understandings of it), these simulations do not always denote *distinct* possibilities, nor is each possibility regarded as *equiprobable*. For example, consider how Johnson-Laird (2012) construes the mental models theory. He claims that, according to the theory, “Each mental model represents a *distinct* possibility” (our emphasis, 137), and “Each mental model represents an *equiprobable* possibility unless there are reasons to the contrary” (our emphasis, 144).

The mental simulations approach is not like this. Two or more simulations may represent the *same* possibility: above, for example, we have 10 simulations representing the same possibility where door B conceals the prize and door C is opened. Furthermore, not every possibility is equally probable: the possibility where door B conceals the prize after door C is opened is more probable than the possibility where door A conceals the prize after door C is opened. And the reason for this is that probabilities are calculated differently: probabilities are not calculated by counting possibilities. Instead, simulations are generated from particular probabilities, the simulations inconsistent with the evidence are then eliminated, and then other probabilities are then calculated by counting these simulations—not by counting the possibilities which they denote.

To that extent, then, the mental simulations approach differs from at least some prominent ways of understanding the mental models theory.

In saying that, one might think that the mental simulations approach involves mental models *if* mental models are understood in a different way.

In any case, though, the mental simulations approach is not intended to be a descriptive claim about human psychology. It does not intend to *describe* how humans *actually do reason*, unlike some theories involving mental models. Instead, it is a claim about one way that humans *can* reason—especially if they receive some training. And I claim at most that this approach can, in some cases at least, improve reasoning and avoid particular biases—most notably, the likelihood neglect bias.

The purpose of this study is then to provide support for this claim, and I will conduct two investigations to this end.

Hypotheses Under Test

Given the results of my previous pilot studies, I hypothesize that:

1. Most naïve participants display likelihood neglect bias (50% <)
2. Some naïve participants (30% <) when trained in the mental simulations approach will calculate the correct posteriors and avoid likelihood neglect in the original Monty Hall problem
3. Some participants (20% <) will calculate the correct posteriors in the new Monty Hall problem when exposed to the mental simulations approach
4. Few or no participants (4% >) will calculate the correct posteriors in the new Monty Hall problem when exposed to the possible models or probability accrual approaches

Experiment 1 tests the first two hypotheses, and experiment 2 tests the latter two.

Significance of the Proposed Research

If the hypotheses are confirmed, this would be significant for several reasons. First, it would reveal a new cognitive bias that may hinder reasoning in other more important contexts where information about likelihoods is present, such as medical diagnosis, legal criminal inquiries, scientific contexts, intelligence analysis and others. Second, it would furnish evidence for a new method to correct that bias that may improve judgmental accuracy and decision-making in these other contexts. Third, the existence of this bias may provide further insight into how the mind works and the extent to which Bayesian models accurately depict human cognition.

Experiment 1: Mental Simulations with MTurkers

Method & Analysis

Purpose:

Experiment 1 aims to determine whether:

1) Participants display likelihood neglect bias, and
2) The mental simulations approach corrects likelihood neglect bias and helps naïve participants correctly solve the original Monty Hall problem

Participants:

Participants will be recruited from Amazon's Mechanical Turk (MTurk). All participants will be above 18 and have English as a first language. Participants will first be given a screening survey to determine their eligibility for the experiment. They will be asked questions that collect demographic information, information about their occupational and educational backgrounds, and information about their familiarity with the Monty Hall problem. All prescreening participants will be asked: "Have you heard of the Monty Hall problem—a problem where the game show host hides a prize behind one door, you select a door and you then have the option to switch to one of the other doors?" They will also be asked whether they are familiar with other things—such as what sudoku or the Central Limit Theorem are. These other questions will be there so respondents cannot tell which question will enable them to qualify for the follow up study, and also so that they will likely indicate familiarity with some things but not others.

In my most recent pilot experiment, most respondents (139 of 202, 68.8%) reported prior familiarity with the Monty Hall problem, although a sizable minority (63 of 202 in our most recently pilot) were “naïve” participants who reported no prior familiarity.

Naïve participants will be randomly assigned to the experimental or control condition until each group is comprised of 42 valid responses.

Materials:

Participants in both conditions will complete an online survey [anonymized and attached with this submission].

Participants will first provide consent and be informed that “The **$9 base payment** requires **completing** the study” and “The **$6 bonus payment** requires **correct** answers to certain questions” (bolding original).

The bonus payment will be important. Previous piloting found that performance in both conditions strongly depended on incentives. In particular, the quality of all responses improved with generous financial bonuses for correct answers: answers in the experimental condition were accurate more often, and frequently those in both conditions more carefully described their reasoning regardless of which answers they gave. Participants were not, however, told which answers needed to be correct. They were awarded bonuses for basic comprehension questions, such as about how many doors there are in the Monty Hall problem. Thus, neither condition was less likely to be awarded the bonus (but again, participants were not aware of this).

Both groups will be presented with the same version of the problem and then asked the same questions about it. They will also answer basic comprehension questions which determine whether their responses are valid and bonused (for example, “How many doors are in the above scenario?”).

Table 2 lists the questions that will measure each central construct, as well as any additional coding or classification for the responses.

Table 2: Question and Coding Table for Measured Constructs

Question(s):	**Response coding**
Likelihood awareness	
Imagine that, unknown to you, door ***A*** *conceals the prize (and that you first selected door* ***A****). If this was the case, then what is the probability that Monty Hall would have opened door* ***C****?* *Imagine that, unknown to you, door* ***B*** *conceals the prize (and that you first selected door* ***A****). If this was the case, then what is the probability that Monty Hall would have opened door* ***C****?*	Displays “Likelihood awareness” if their answers are equal to 50% and 100% for the respective questions.

Correct posteriors	
What is the probability that door <u>A</u> conceals the prize after Monty Hall shows you the goat? *What is the probability that door <u>B</u> conceals the prize after Monty Hall shows you the goat?*	Gives "Correct posteriors" if they answer "1/3" and "2/3" to these questions respectively.[2]
Likelihood neglect bias	
[Derived from the responses to the above four questions]	Participants demonstrated "likelihood neglect bias" if they displayed likelihood awareness but did not think that B more probably concealed the prize than A
Switching vs. sticking	
If you were given the option either to switch doors (from door A to door B instead) or to stay with the door you initially chose (door A), which would you chose?	"Switchers" if they opted to switch. "Stickers" otherwise.
Justification	
Important: Please tell us the thought process you went through as you tried to determine the answers to the questions we asked you about the probability of the prize being behind different doors above. Please try to lay out your thought process sequentially if possible and provide as much detail as you can.	No additional response coding
Subjective self-confidence	
Consider your answers for the previous questions about the probabilities for the gameshow problem. How confident are you that those answers were the correct ones?	No additional response coding
Self-reported understanding	
How well do you think you understand why your answers for those questions were the correct ones?	No additional response coding

The experimental condition, however, will have the following two differences: 1) participants will be taken through a training module in the mental simulations approach, and 2) they will be asked if they used the mental simulations approach to answer the Monty Hall problem.

In the training module, participants can view either or both of two kinds of materials: reading materials or a series of videos (video materials). Both kinds of materials convey similar information about the mental simulations approach and how to use it. [The video cannot be anonymized since it is the experimenter presenting in the video. Hence it has not been included in this submission. In any case, the video materials are similar to the reading materials in the appendix, since the presenter read directly from the materials with very few departures, such as introducing themselves as a psychologist and their institutional affiliation.]

[2] Percentage or decimal answers rounded in either direction will also be accepted. For example, "33%" or "34%" will be accepted as correct, but an answer like "32%" will not.

Importantly, the materials guide participants through a problem that is analogous to the Monty Hall problem: the story of the prisoners. The story of the prisoners is an old problem (Gardner, 1959), and it is no innovation of this experiment. However, I modified the story as such. You, Alison, Billy and Carly are in prison. One of you will be set free. The rest will be imprisoned for life. A lottery random determines who will be set free. So all four of you have an equal probability of being set free—namely, 1/4. You ask the prison warden if he can tell you who will be set free. He says he can tell you only the names of two people who will not be set free. But we suppose that he cannot lie and he cannot tell you whether you will be set free or not. He then says that Billy and Carly will not be set free.

The story of the prisoners is similar to the Monty Hall problem in several ways. The possible outcomes all have equal probabilities in the beginning. Participants then get the evidence which rules out at least one of the outcomes. Participants generally conclude that the remaining possible outcomes are equally likely. But if the problems are described correctly, then one of the outcomes is actually more probable than the other. And the reason for this is that the likelihoods differ: the evidence is more probable given one possible outcome rather than another.

However, this experiment's version of the story of the prisoners is also importantly *dissimilar* to the Monty Hall problem in several ways. There are a different number of possible outcomes in the beginning: there are three in the Monty Hall problem and four in this story of the prisoners. The prior probabilities in the two scenarios are different to each other: the priors in the Monty Hall problem are all 1/3, whereas the priors in the story of the prisoners are all 1/4. The likelihoods are different too: in the Monty Hall problem, there is a 50% likelihood of the evidence if door A conceals the prize whereas, in the story of the prisoners, there is approximately a 33% likelihood of the evidence if you were to be set free instead of Alison. (This is because if you were to be set free, the warden could have given any one of three combinations of names about who will be set free: 1) Alison and Billy, 2) Alison and Carly, or 3) Billy and Carly.) Consequently, the posteriors are also different: there is a two thirds probability that door B conceals the prize in the Monty Hall problem but a three fourths probability that Alison will be set free in the story of the prisoners.

Because of these dissimilarities, participants could not solve the Monty Hall problem merely by mindlessly repeating answers to the story of the prisoners. Some additional understanding is needed.

Participants will also be asked questions to test their understanding of the mental simulations approach.

They will then be presented with the Monty Hall problem for the first time (since they will be prescreened for prior familiarity). They will then answer the above questions about the problem, as well as some obvious basic comprehension questions to screen out inattentive responses (for example, how many doors were in the Monty Hall problem).

Analysis Plan:

I will analyze various statistics of importance:

A) The frequency of likelihood neglect in the control group

B) The frequency of reported use of the mental simulations approach in the experimental group
C) Differences in particular proportions and
D) Differences in particular means

The frequency of likelihood neglect in the control group is easy to measure (recall the coding of likelihood neglect in Table 2). It is simply the proportion of participants displaying the bias p_c—that is, number of participants who display likelihood neglect over the total number of participants in the control group. This will determine whether participants do indeed display likelihood neglect (as opposed to merely, say, unawareness of what the likelihoods are).

I will also analyze data to try determine whether experimental participants used the mental simulations approach.

One way to do this is to consider the frequency with which participants reported using the mental simulations approach in the experimental condition. Recall that participants will be asked the following question:

Did you use the "mental simulations" approach to arrive at your answer?

Please answer honestly. There is no penalty for giving one answer or another.

Affirmative answers would indicate that they had indeed used the mental simulations approach. This is especially the case since some participants in the pilot experimental groups indicated that they did not use the approach, and they were told that there were no penalties for doing so. Social desirability does not appear to bias responses here.

Aside from this, two other kinds of evidence will also indicate whether participants use the approach.

One of these are the correct posterior themselves. In the pilot experiments, some people in the experimental group provided the correct posteriors while everyone in the control group provided the wrong posteriors. Something must explain why they got the correct values that they did. Since the only difference between the groups was that one was trained in the mental simulations approach and the other was not, this provides some suggestive evidence that some experimental participants used the mental simulations approach to get to the correct answer.

The other kind of evidence for use of the method comes from participants justifications for their answers. One participant in a pilot of the experimental group, for instance, gave the following answer which I have not edited.

I set up the problem like the example problem [the problem of the prisoners] with 6 circles, 2 for A, 2 for B and 2 for C to represent the prior probabilities. The host would open door C 50% of the time and door B 50% of the time when door A contains the prize so I shaded in 1 of the 2 A circles. The host would open door C 100% percent of the time when door B contains the prize because they wouldn't show me what's behind the door I picked (A) or the door with the prize so door C is the only option. I filled in both B circles to represent that. The host would open door C 0% of the time when door C contains the prize, so I *disregarded those circles and I also disregarded the unfilled circle for A. I was left with 1 filled in circle for A and 2 for B, so that is how I came to the conclusion that there was a 1/3 chance of it being A and 2/3 for B. It would make more sense to switch doors because there's a higher probability of the prize being behind door B than door A.*

To my mind, at least, it is obvious that this particular person used the mental simulations approach, and they explicitly reported that they used the approach as well. Instead of coding such justifications, I will let the data speak for itself by reproducing all of the justifications from the final experiment in an appendix or online. This has already been done for the most recent pilot study in Appendix D. I will then let the reader make up their own mind about the extent to which the justifications provide evidence of using the mental simulations approach. I think that this not only makes the data more transparent, and gives the reader more autonomy in reaching their conclusions, but it also makes the case for the approach more compelling.

Differences in Proportions: PD and Cohen's h

However, the key statistics for the purposes of this experiment are the *differences* in proportions for two variables:

1. **Likelihood neglect**: Whether participants commit the likelihood neglect fallacy
2. **Correct posteriors:** Whether participants provide correct answers for both of the posterior probabilities about door A and B concealing the prize

The effect size will be measured in two ways: unstandardized differences in proportions (PD) and Cohen's *h* (the arcsine-transformed difference between proportions).

PD will be used because it is the simplest effect size to understand. In our case, we will take the proportion in the experimental group—denoted with p_e—and subtract from it the proportion of the control group—denoted with p_c. This could then be interpreted as the percentage *improvement* in the experimental group. Note that the absolute value of PD will not be used because directional information is useful here: the non-absolute values will indicate whether the experimental procedure *increases* correct posteriors and *decreases* likelihood neglect among the experimental participants.

Cohen's *h* will also be used because it is another common measure of the distance between two proportions (Cohen, 1988). It is calculable as such:

$$h = 2[\arcsin(\sqrt{p_e}) - \arcsin(\sqrt{p_c})]$$

The absolute value will not be used for the same reasons that apply to PD.

To estimate whether the effect sizes are due to chance, I will report Wilson confidence intervals and *p*-values for each statistic too.

I will also report all of these same statistics when comparing *those who claimed to use the mental simulations approach* and *those who did not* in the experimental group. I will call those who reported using the approach "the users" and those who did not "the non-users". This would help to get a better indication of the effect of approach, since it considers the effect sizes among those who claim to use the approach (as opposed to those in the experimental group who decide not to use the approach for whatever reason).

Power Analysis and Sample Size

Since the difference in correct posteriors is a central statistic, this informed a power analysis to guide sample size selection. The sample proportion for *correct* posteriors in the experimental and control group were supposed to be 30% and 5% respectively. This is a conservative estimate of the experimental outcome based on previous piloting. The desired α level (the probability of error if the null hypothesis is true) was set to 0.05, and the desired statistical power $1-\beta$ was set to 0.9 (where β is the probability of failing to reject the null hypothesis if it is false). Assuming participants are evenly allocated to conditions, the power analysis yielded a desired sample size of 42 per condition—84 in total.

Differences in Means: PD and Cohen's h

Aside from the differences between population proportions, I will also analyze how participants in the two groups differ with respect to means and standard deviations for these two variables:

3. **Self-Confidence:** How confident a participant is in the accuracy of their answers about the posterior probabilities of the doors concealing the prize
4. **Self-Reported Understanding:** How well a participant claims to understand *why* their answers are the correct ones

Such comparisons will be made for *both* the experimental group vs. the control group, and also users vs. non-users in the experimental group.

This will help to estimate how effective the mental simulations approach is at helping participants to *understand why* the answers that it recommends are the right ones. As Saenen et al. (2018) state, improving such understanding is an unresolved challenge.

The effect sizes will again be measured in two ways: difference between the means (DM) and one of two common statistics. Which of the two common statistics will be used will depend on the variation in the data. Cohen's *d* will be used if the standard deviations for the comparison groups are the similar, say, $|s_e - s_c| > 0.75$. After all, it is not recommended for comparing two groups with standard deviations which are substantially different (or with sample sizes that are unequal or small). In any case, Cohen's *d* is calculable as follows:

$$\text{Cohen's } d = \frac{(\mu_e - \mu_c)}{\sqrt{(s_e^2 + s_c^2)/2}}$$

where μ_e and μ_c are the means of the experimental and control group respectively, and s_e^2 and s_c^2 the variances of the respective populations

If Cohen's d is used, then in keeping with convention, I will follow Cohen's (1988) rule of thumb, describing an effect size as "small" when $d = 0.2$, "medium" when $d = 0.5$ and "large" when $d = 0.8$ (although Cohen acknowledged the use of these terms should be context-dependent).

Otherwise, Glass' Δ will be used if the standard deviations are not similar for the comparison groups (Glass et al., 1981). This is especially the case since others have appraised it as an intuitive statistic when one cannot assume the standard deviations are equal (Dey & Mulekar, 2018). Glass' Δ is calculable as follows:

$$\text{Glass' } \Delta = \frac{(\mu_e - \mu_c)}{s_c}$$

where s_c is the standard deviation of the control group (as recommended by Glass himself)

Glass et al. (1981) provide no rule of thumb for labelling effect sizes as "small", "large" and the like. In fact, they explicitly rail against the context insensitive application of such adjectives (Glass et al., 1981, pp.104).

Aside from these statistics, the difference between the means is also used because it is the easiest to interpret. It does not by itself take into account the variation in responses, but it will be presented alongside standard deviations for the comparison groups.

And again, confidence intervals and *p*-values will also be reported. Which type of interval will be reported also depends on whether the variances are (near) equal. If they are equal, I will use classical confidence interval methods of the sort found for unpaired interval estimates in the fourth chapter of Altman et al. (2000). Otherwise, I will use Welch's *t*-interval method if the variances are not equal.

Bar charts will also be presented for each of comparative statistics. I will also present bar graphs for the distribution of responses among the experimental and user group with respect to their self-confidence and self-reported understanding.

I will also conduct exploratory data analysis to detect any associations between prescreen survey responses on the one hand and correct posteriors, self-confidence and self-understanding in the experimental condition on the other. My previous pilots do not have the statistical power to reveal interesting associations. I consequently have no suspicions about what interesting associations may emerge, if any, from the final dataset.

Here, then, is a summary of the analysis plan for Experiment 1:

Experiment 1: Summary of Analysis Plan

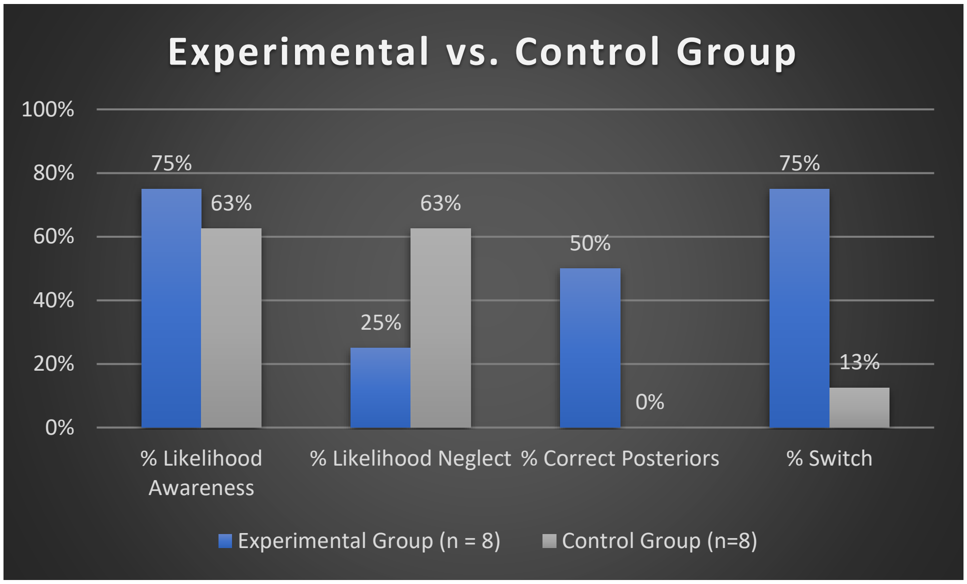
Non-Comparative Statistic:	**Measures of Significance**
- Proportion of participants committing Likelihood Neglect in the Control Group	Wilson confidence intervals and *p*-value
- Proportion of participants in the experimental group who were users	Wilson confidence intervals and *p*-value

- Distribution of responses for confidence and understanding in the experimental group		NA
- Distribution of responses for confidence and understanding among users and non-users in the experimental group		NA
Comparative Statistics:		
Comparison	**Effect Size Measures**	**Measures of Significance**
Experimental Group vs. Control Group:		
- Difference in *proportions* for *Posteriors Correct*	*PD* and Cohen’s *h*	Wilson confidence intervals and *p*-values
- Difference in *proportions* for *Likelihood Neglect*	*PD* and Cohen’s *h*	Wilson confidence intervals and *p*-values
- Difference in *means* for *Self-Confidence*	*DM* and either Cohen’s *d* or Glass’ Δ	Classical or Welch confidence intervals and *p*-values
- Difference in *means* for *Self-Reported Understanding*	*DM* and either Cohen’s *d* or Glass’ Δ	Classical or Welch confidence intervals and *p*-values
Users vs. Non-Users in the Experimental Group:		
- Difference in *proportions* for *Posteriors Correct*	*PD* and Cohen’s *h*	Wilson confidence intervals and *p*-values
- Difference in *proportions* for *Likelihood Neglect*	*PD* and Cohen’s *h*	Wilson confidence intervals and *p*-values
- Difference in *means* for *Self-Confidence*	*DM* and either Cohen’s *d* or Glass’ Δ	Classical or Welch confidence intervals and *p*-values
- Difference in *means* for *Self-Reported Understanding*	*DM* and either Cohen’s *d* or Glass’ Δ	Classical or Welch confidence intervals and *p*-values
Exploratory Analysis		
- Associations between prescreen survey responses and correct posteriors, self-confidence and self-reported understanding in the experimental condition	*Correlation Coefficients*	Confidence intervals and *p*-values
Power Analysis:		
- Assumed proportion of Posteriors Correct in Experimental Group		.3
- Assumed proportion of Posteriors Correct in Control Group		.05
- Desired Type-I error threshold α		0.05
- Desired statistical power $(1-\beta)$		0.9
- Assumed ratio of participant allocation to experimental and control groups		1
- Computed number *n* of participants per condition		42
- Total number of participants		84

Pilot Results

I have tentatively placed results from my most recent pilot study here. It is a small sample (n = 16), but these are the kinds of results I have invariably found in my numerous pilot experiments (the experimental group does better in some ways but still needs improvement in others). Also note that I will not report some of the main statistics of interest, such as Cohen's *h* or Cohen's *d*. (These will be reported in the final paper.)

On average, the experimental group performed better on particular measures.

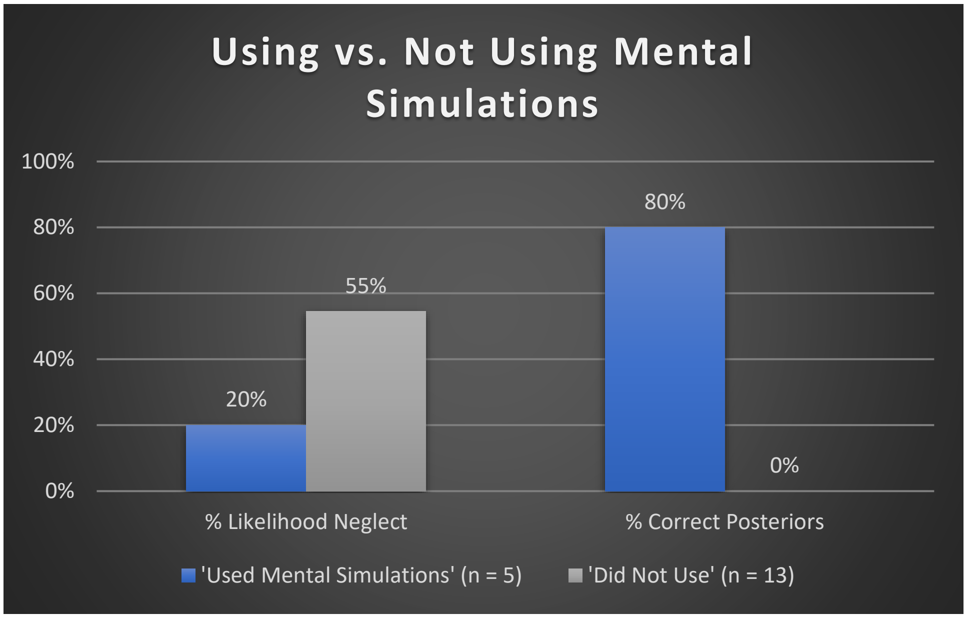


Both the experimental and the control group were similarly aware of the likelihoods. However, the experimental group was less likely to demonstrate the likelihood neglect bias and more likely to switch doors. Furthermore, half in the experimental group correctly determined the posterior probabilities of the prizes compared to none in the control group.

At first, this seems favorable for the mental simulations approach.

However, five out of eight of the participants in the experimental condition actually reported using the approach.

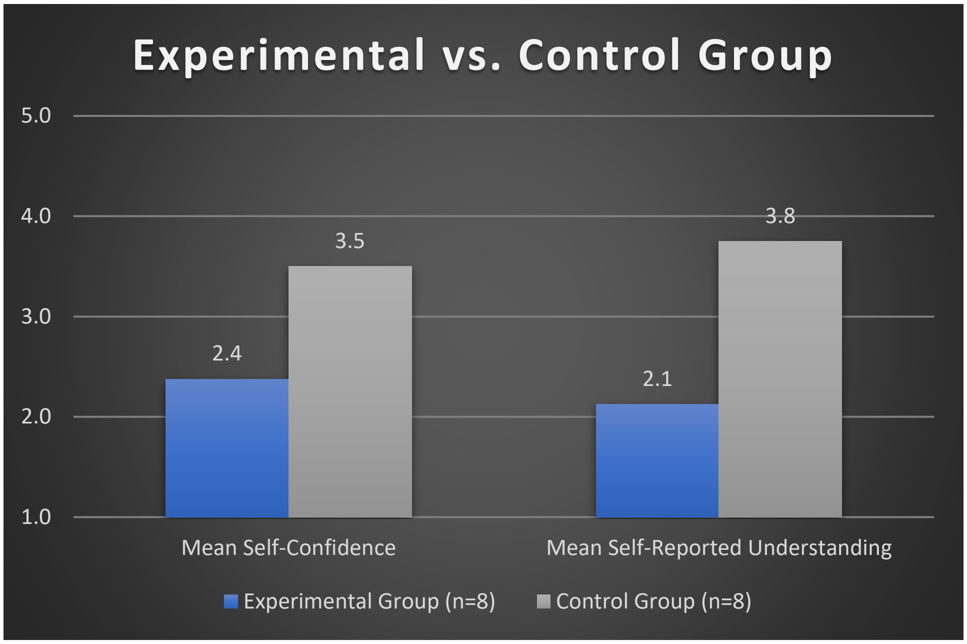
To determine the influence of the approach, it makes sense to consider the results among those who explicitly said that they used the approach:



When we consider those who said they used the approach, most correctly determined the posterior probabilities. In contrast, none of those who did not use the approach correctly determined the posterior probabilities, regardless of whether they were in the experimental or control groups.

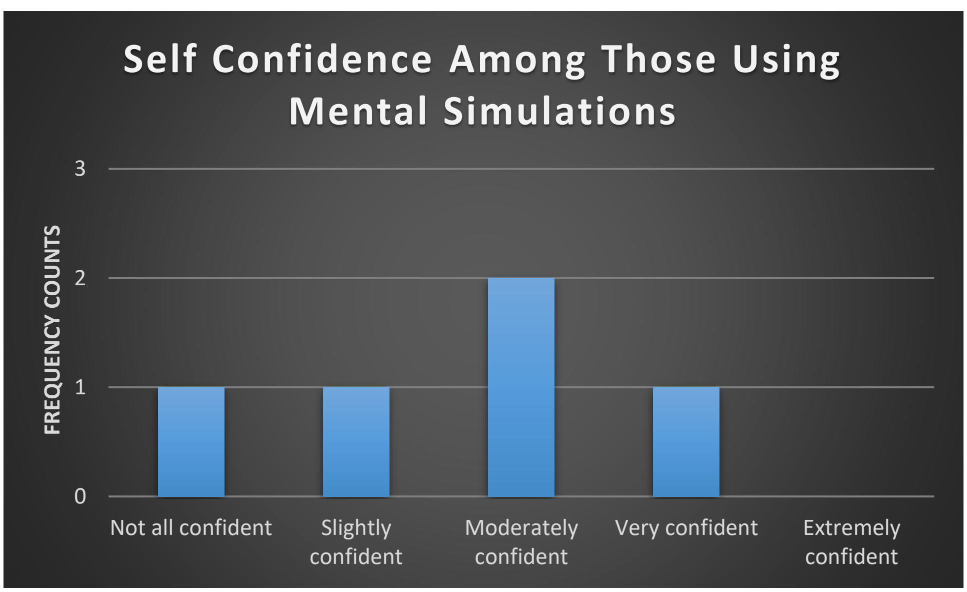
For these reasons, it is plausible that participants really did use the mental simulations approach to arrive at the correct answers. However, further insight about this comes from the participants descriptions of the reasoning process they use themselves. For that reason, these have been reproduced in Appendix D without any editing. The hope is that readers can make up their own minds about the extent to which these descriptions suggest the participants really did use the approach.

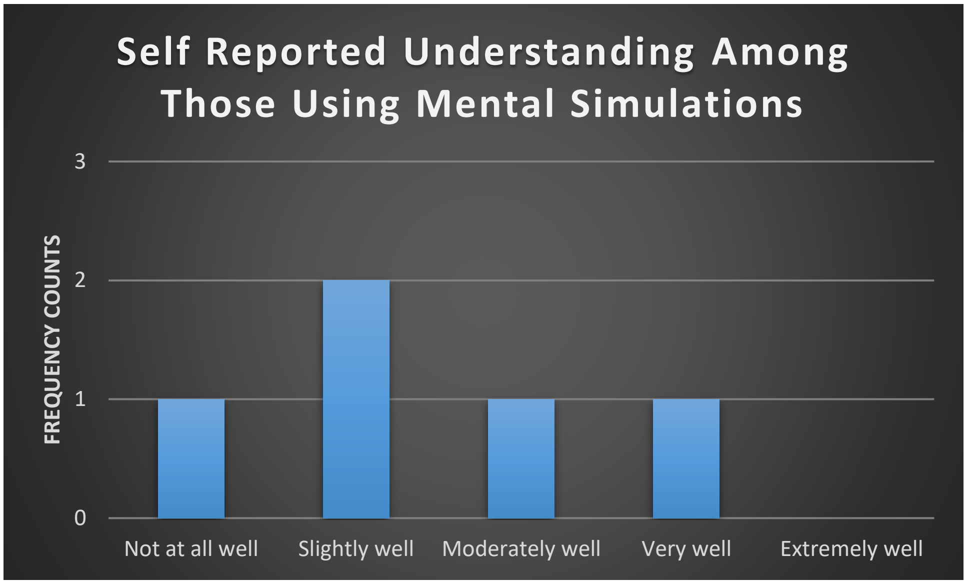
Overall, the aforementioned results seem pretty favorable for the mental simulations approach, but not all results are obviously as such. In particular, participants who were in the control group reported higher confidence in the correctness of their answers and a greater understanding of why their answers were the correct ones. On average, they reported being “Moderately” to “Very confident” in their answers and understanding the reasons for their correctness “Moderately” to “Very well”. In contrast, those in the experimental condition exhibited less confidence and self-reported understanding. On average, they reported being only “Slightly confident” in their answers and understanding the reasons for their correctness only “Slightly well”. The results are depicted here on a 5-point scale from “Not all confident/well” to “Extremely confident/well”:



Arguably, however, it is undesirable that the control group reports greater confidence and self-reported understanding. After all, their answers were incorrect. It is then no virtue to be confident that one is correct and understands why when, actually, they are incorrect and do not understand the problem at all. Consequently, this is no advantage of the control groups.

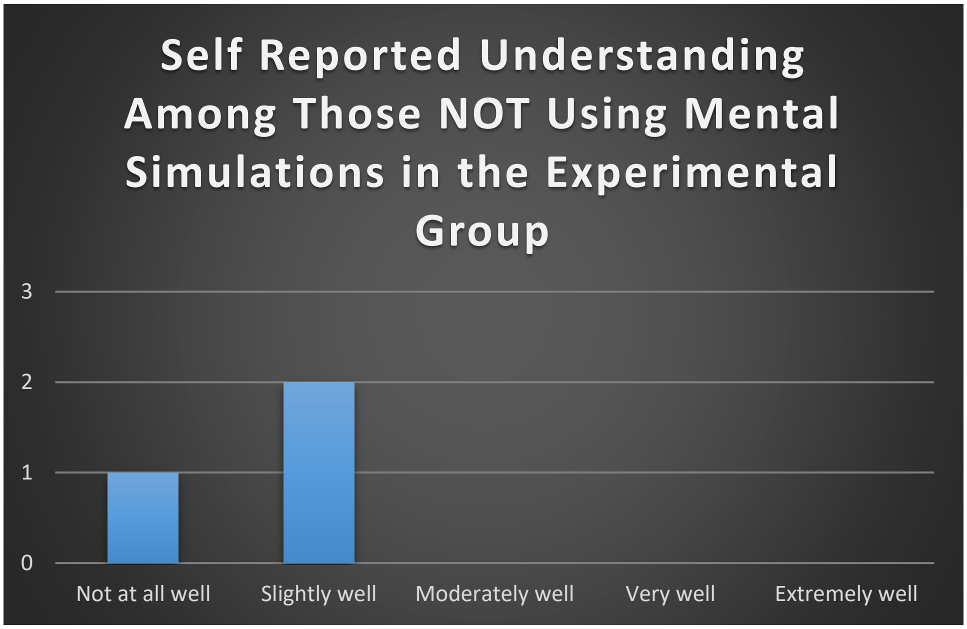
However, it is undesirable that participants in the experimental group claim on average to be only slightly confident or only slightly understanding of the correctness of their answers. Of course, the picture is more nuanced when we examine the distribution of responses among those who do and do not use the approach:





What is apparent is that there is some variation among those using the approach: some are quite confident in their responses, and report understanding the correctness of their responses fairly well. But not all are like this.

Participants in the experimental group who do not use the approach report even less understanding and confidence:



To some extent, then, those who do not use the approach bring the experimental group's average self-confidence and self-reported understanding down. Perhaps they do not use the approach for various reasons. Maybe it is too difficult for them; or maybe these participants are less likely to give the task their best effort.

That said, it is clear that enhancing confidence and understanding is also a challenge for the mental simulations approach. But this is not surprising. After all, participants are introduced to one of the most confusing math brain teasers and then to a very unfamiliar way of calculating probabilities. They are then expected to provide answers to the brain teaser after learning all of this, with little feedback and little practice, all in the space of an hour. It is not surprising if the development of confidence and understanding may take more time.

Experiment 2: Popular Solutions and The New Monty Hall Problem

Method & Analysis

Experiment 2 aims to compare how several approaches handle the New Monty Hall problem (see a description of the problem above).

Participants will be recruited from MTurk. The only prerequisite is that participants are above 18 and English is a first language. Participants will not be screened for prior problem familiarity, or anything else, since the problem they need to solve—the New Monty Hall problem—is new and not widely known. Participants will be recruited and randomly assigned to one of three conditions until each condition had 25 valid responses (as per a prior power analysis). Participants will again be offered a base payment of $9 for completing the study and then another bonus payment of $6 for giving correct answers.

Participants will first be given the old Monty Hall problem in which switching yields the prize with a probability of 2/3. They will then be given the following text:

> *Instead, the correct answer is that door B is more likely to conceal the prize: the probability that door B conceals the prize is 2/3. This may be counter-intuitive at first, but it is universally accepted as correct by experts in probability and statistics. Why is this the case, then?*
>
> *Some researchers have offered the explanation on the following page for why door B is more likely and you should switch doors.*
>
> ***PAY ATTENTION TO THIS EXPLANATION: We will ask you questions about a different version of this problem later on, and the explanation may help you answer it, even if you think you know the right answer to this problem already.*** (Bolding original, italicization added)

Participants will be given one of three explanations, depending which condition they are assigned to. Participants in the mental simulations condition will be given an explanation in terms of the mental simulations approach. Participants in the possible models condition will be given an explanation in terms of the possible models solution. Participants in the probability accrual condition will be given an explanation in terms of the probability accrual solution. All of the explanations closely resemble the explanations given earlier in this paper.

Participants will then be asked questions about the respective explanations to test their understanding of the method. In this case, incorrect answers will be rejected automatically by the survey, and they cannot continue the survey until they provided correct answers.

They will then be given the New Monty Hall problem, described in the way that it was described earlier in this paper.

Afterwards, they will be reminded that their bonus depends on correct answers and they will be asked the same questions to assess likelihood awareness and their posteriors. However, the answers for these questions will be in multi-choice format and presented in a randomly generated order each time.

Participants will then be asked to “please explain why you answered the above questions the way you did”. They will also be asked whether the earlier explanation of the original Monty Hall problem affected their answers to the New Monty Hall problem.

As with Experiment 1, they will lastly be asked to report their confidence in their answers, their self-reported understanding and their prior familiarity with the Monty Hall problem.

Analysis Plan

The key statistic of interest is the difference in *correct posteriors* for the various groups. Here, however, the posteriors for door A and door B concealing the prize are 9/19 and 10/19 respectively. Comparing the correct posteriors among the conditions will indicate whether the respective methods appropriately sensitize participants to the likelihoods so that they give the correct answers. This statistic will be reported in the same way as the earlier proportions: with PD, Cohen's *h* and p-values and confidence intervals.

Additionally, I will also explore differences in whether participants said the explanations affected their answers. This may give some insight into how convincing the explanations were, or of how easy it is to apply the explanations to the new Monty Hall problem.

I will also compare differences in reported self-confidence and understanding among the conditions. This will be done in the same way as it was for Experiment 1.

Since the correct posterior statistic is again the key statistic, it informed a power analysis to guide sample size selection. The power analysis set α to 0.05 and $1 - \beta$ to 0.9. The analysis supposed that 30% of participants in the mental simulations would answer with the correct posteriors, while none in the other conditions would. Again, this was a conservative supposition based on results from previous pilot experiments. The power analysis then suggested 25 valid responses were necessary per condition.

Experiment 2: Summary of Analysis Plan

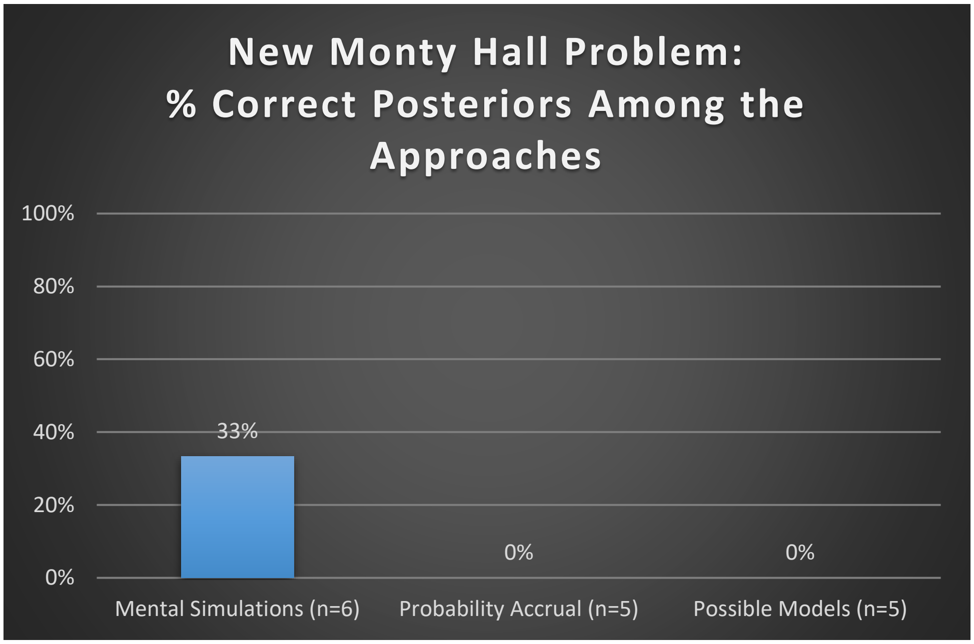
Comparative Statistics:	**Effect Size Measures**	**Measures of Significance**
Experimental Group vs. Control Group:		
- Difference in *proportions* for *Posteriors Correct*	*PD* and Cohen's *h*	Wilson confidence intervals and *p*-values
- Differences in whether the explanations affected participants answers	*PD* and Cohen's *h*	Wilson confidence intervals and *p*-values
- Difference in *means* for *Self-Confidence*	*DM* and either Cohen's *d* or Glass' Δ	Classical or Welch confidence intervals and *p*-values
- Difference in *means* for *Self-Reported Understanding*	*DM* and either Cohen's *d* or Glass' Δ	Classical or Welch confidence intervals and *p*-values
Power Analysis:		
- Assumed proportion of Posteriors Correct in Experimental Group		.3
- Assumed proportion of Posteriors Correct in Control Group		0

- Desired Type-I error threshold α	0.05
- Desired statistical power $(1-\beta)$	0.9
- Assumed ratio of participant allocation to experimental and control groups	1
- Computed number n of participants per condition	25
- Total number of participants	45

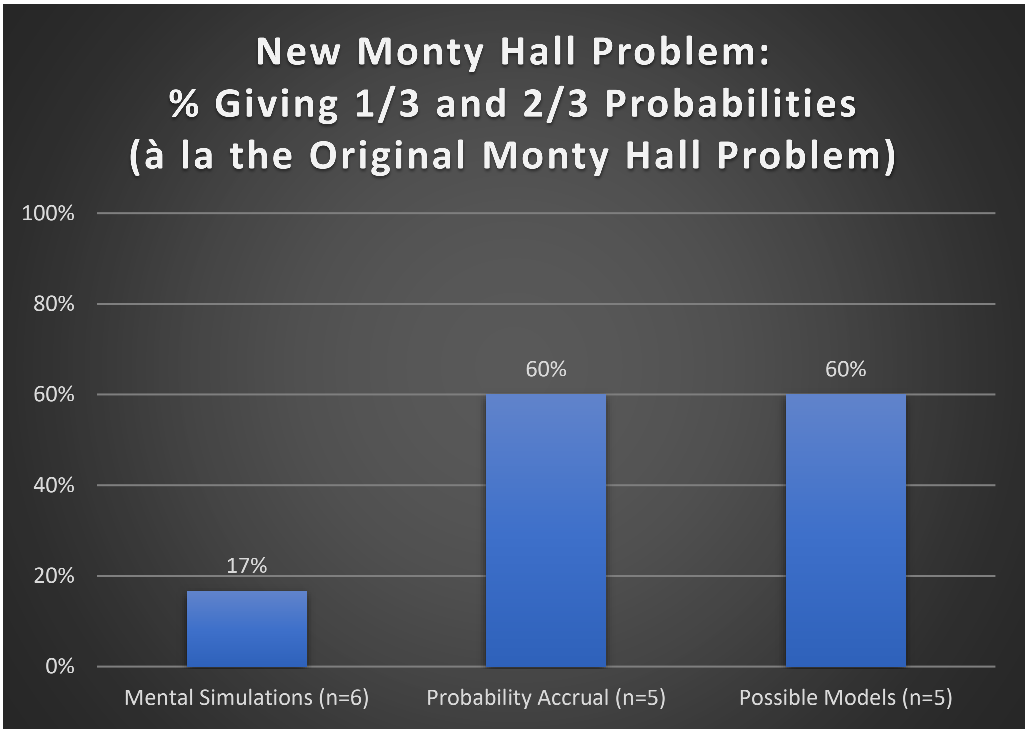
Pilot Results

I have again reported some results from the last pilot experiment, but not all of the final statistics.

As expected, the only condition in which participants correctly selected the posteriors was the mental simulations condition.



This confirms what we already had strong theoretical grounds to believe: the other approaches do not yield the correct answer in the New Monty Hall problem, and instead they lead them astray in thinking that the new problem is essentially the same as the original problem. We can consider each condition and the proportion of answers that are the same as the correct answers for the original Monty Hall problem:



In a sense, then, the alternative approaches not only generate incorrect answers, but they seem to encourage *insensitivity* to the likelihoods by generating an automatic response to the new Monty Hall problem.

General Discussion and Future Research

This paper is the first to characterize likelihood neglect bias and to provide experimental evidence that participants display the bias. Furthermore, it is the first to articulate and experimentally test the mental simulations approach against other popular solutions to the Monty Hall problem. Results indicate that the mental simulations approach helps some participants overcome likelihood neglect bias and give correct answers to both the original and new Monty Hall problems. Experiment 2 also demonstrates that neither of two popular solutions to the original Monty Hall problem help participants with the New Monty Hall problem. This, I have argued, is because these other approaches are *unsafe* and *insensitive* to the likelihoods: their verdicts do not change as the likelihoods change, and consequently they are liable to producing false judgments of probability.

That said, experimental participants in the Experiment 1 varied in their self confidence and self-reported understanding. Some were confident in the correctness of their understanding and claimed to understand why their answers were the correct ones. Others, however, were not so confident in their answers and understanding.

To some extent, this is understandable. In one hour, participants encounter one of math's most notorious and confusing brain teasers, and they are then trained in what is, for them, a completely new approach to probabilistic reasoning. It is understandable that greater confidence and understanding may require more than this just this.

Future research, then, could explore these avenues:

1. Better understanding why some participants confidently use the approach and others do not
2. Designing and testing interventions to improve confidence and understanding among those who are trained in the mental simulations approach

Why Humans Fall Prey to the Likelihood Neglect

So I have provided evidence that humans are susceptible to a likelihood neglect bias. I now turn to consider theoretical explanations for why this occurs in the Monty Hall problem.

I offer a two-part explanation of incorrect responses in the Monty Hall problem. The first is positive—about what participants *do* when reasoning. The second is negative—about what participants *do not* do when reasoning.

The first part of the explanation is that participants have equiprobabilistic intuitions, as has been noted by others. It is well-known that the majority of participants think that the unopened doors have the same probability of concealing the prize (Tubau et al., 2015). I think this represents an intuitive heuristic which humans apply: if there is no reason to favor one outcome over another, then we should regard each outcome as equally probable. This heuristic is simply a version of the *principle of indifference*—a principle that has both advocates and opponents in philosophy (Kaplan, 1996; White, 2010; Wilcox, 2019). This heuristic may in turn be explained by the availability heuristic (Tversky & Kahneman, 1973). The availability heuristic is the process whereby humans determine the probability of an outcome based on its mental availability—that is, the ease with which relevant instances come to mind. In the context of the Monty Hall problem, both the remaining options are equally available: it is just as easy to call to mind an instance of one unopened door concealing the prize as it is for the other.

Consequently, then, I posit that the humans utilize an equiprobability heuristic, and that this heuristic may itself be explained by the availability heuristic—at least in the content of the Monty Hall problem.

However, judgments of equiprobability are not inevitable, and indeed a sizable number of philosophers disavow the principle of indifference (Joyce, 2010; Kaplan, 1996; Meacham, 2014). So it is clear that judgments of equiprobability can be *overridden* if other factors are present. For example, Saenen, Van Dooren, and Onghena (2015) report that subjects were more likely to switch doors and override their intuitions if they could see statistical data that the other door concealed the prize most of the time.

This leads us to the second part of explanation—the negative part. In particular, I claim that participants often give incorrect responses because they lack the cognitive factors that would impel them to switch and, more specifically, they often do not understand the importance of likelihoods.

In this respect, humans are not born with an innate, infallible knowledge of the law of likelihood. This is plausible for several reasons. The most immediate is that, as demonstrated in the Monty Hall problem, people can violate the law of likelihood. But beyond that, the law of likelihood requires making distinctions that humans systematically blur. In particular, applying the law of likelihood requires the ability to distinguish between the likelihood of the evidence given some hypothesis on the one hand and the probability of the hypothesis given the evidence on the other. Yet humans often fail to make such distinctions, as has been well-documented with the confusion of the inverse fallacy (Villejoubert & Mandel, 2002).

That, then, is my explanation of incorrect responses in the Monty Hall problem. First, on the positive side, participants have equiprobabilistic intuitions, possibly explained by equal availability. Second, on the negative side, participants lack an understanding of the importance of likelihoods.

Of course, there is a challenge to the negative part of this explanation. In particular, there is evidence from other contexts where it looks like humans do anything *but* neglect the likelihoods: that is, contexts where humans place *too much* weight on them.

There are at least two contexts where this may be salient.

One context concerns null hypothesis testing. In particular, scientists often have the misconception that if the likelihood of the data given some hypothesis is sufficiently low, then the probability of that hypothesis is also low (Kalinowski et al., 2008). If anything, we might think this *overemphasizes* the importance of likelihoods instead of neglecting them.

The other context involves medical diagnosis. More specifically, studies of base rate neglect have repeatedly shown that people tend to believe that the probability of having a disease given some evidence is similar or identical to the likelihood of that evidence given that one has the disease (Eddy, 1982; Hammerton, 1973; Liu, 1975). Again, it looks as though people place too much importance on the likelihoods.

So what are we to make of these other cases?

I suspect that in these other cases, people are using forms of reasoning *aside from Bayesian reasoning with likelihoods*, and the information about likelihoods is merely used to inform or engage these other forms of reasoning.

Exactly how this is done, however, depends on the specific context.

In the context of null-hypothesis testing, some researchers suggest that humans are instead approximating a form of inference known as *modus tollens* (Cohen, 1994; Falk & Greenbaum, 1995). One version of modus tollens is as follows: i) if A, then not B, ii) B, so therefore iii) not A. Modus tollens is a valid argument form, and so people may be approximating it and thinking something like the following in null-hypothesis testing: "after all, if the null hypothesis is true, we probably wouldn't get the results we have, so then the null hypothesis is probably false". Here, the likelihoods are important not because they play the role envisaged by probability theory, but because they provide probabilistic support for the first step (i) in a modus tollens argument, thereby supposedly providing probabilistic support for the modus tollens conclusion. It is also noteworthy that one study found such misconceptions were reduced when participants were presented with examples of how modus tollens is not a valid argument form in probabilistic settings (Kalinowski et al., 2008). So that is one explanation for one context.

What, then, is the explanation for the other: that is, base rate neglect in medical diagnosis? Here, I hypothesize a multi-part process. First, humans judge that there is some kind of probabilistic connection between having a disease and the evidence—such as various symptoms or a positive test result. Participants may make this connection because they have what we could call *an associational history*—that is, experiences, memories or information which associate the hypothesis under consideration with the putative evidence that bears on it. In this context, people frequently associate diseases with both symptoms and test results, in part because we often experience illnesses concurrently with symptoms and positive test results. For that reason, when we are informed that a disease would probably give rise to a symptom or a test result, we make an intuitive connection between the two—we think the evidence says something about the probability of the disease. So the first step is to make an intuitive connection between the evidence and the hypothesis under consideration—to think that they have *some* kind of probabilistic connection.

The next step is to make a more *precise* judgment about what that connection is—that is, what the probability of the hypothesis is given that evidence. This, then, is where the likelihoods come in. Again, I suspect that participants utilize information about likelihoods, not because they reason in accordance with probability theory, but rather because they intuitively think there is a probabilistic connection and they use the likelihoods as a *cue* to judging what that connection is. And since, as studies have shown, humans often struggle to distinguish likelihoods from their inverse probabilities (Villejoubert & Mandel, 2002), they then think the likelihood is similar or identical to the probability of the hypothesis given the evidence.

So we have given two accounts of how undue reliance on likelihoods is really an approximation to other forms of reasoning, and likelihoods are then merely used to inform or engage these other forms of reasoning.

Given such accounts, we can summarize why likelihood neglect may occur in some circumstances while undue reliance on likelihoods may occur in others. In particular, when reasoning about a given problem, humans engage one of any number of processes to solve the problem, depending on the specific features of the problem and their knowledge, familiarity with similar problems and the like. If the problem has specific features, then it may engage undue reliance on the likelihoods and even the inverse probability confusion. Such features can include the problem resembling other problems which involve modus tollens forms of reasoning. Or they can include features with which the participant has an associational history. In such cases, the likelihoods are used to support the premise of a modus tollens argument or to inform the participant's intuitive judgment of how the evidence and hypotheses are related. However, if the problem lacks such features, then humans will use other processes to estimate probabilities, including the equiprobability or availability heuristics. In such cases, they are sometimes susceptible to the likelihood neglect bias.

So I have provided a theoretical explanation of the causes of likelihood neglect and have examined how this explanation is consistent with evidence that humans sometimes unduly rely on likelihoods.

Implications for Bayesian Cognitive Science

The previous experiments arguably provide clear evidence that humans depart from the norms of probability theory.

A question then naturally arises: what implications, if any, does this have for Bayesian cognitive science?

The answer to this question depends on how we understand the core claim of Bayesian cognitive science. I think there are at least three ways to understand its core claim, and I discuss the implications of the aforementioned results for each of them.

One way to understand the core claims of Bayesian cognitive science is as a *global claim* about *all* mental processes. More specifically, the claim is that *every* mental process conforms to Bayesian norms, including the processes involved in reasoning about the Monty Hall problem. Call this *global descriptive Bayesianism*.

Clearly our experiments provide evidence against global descriptive Bayesianism. In the Monty Hall problem, most participants deviate from the optimal Bayesian response—namely, switching. But this is old news, and the heuristics and biases research program has unearthed many deviations from Bayesian norms (Gilovich et al., 2002). In any case, it is not clear that any researchers in cognitive science endorse global descriptive Bayesianism—or, at the least, it is difficult to find them committing to it in writing.

Instead, many cognitive scientists understand the core claim of Bayesian cognitive science as being more of a *local claim*. Here, the claim is merely that *some*—but not all—mental processes conform to Bayesian norms. Furthermore, in cases where such processes do conform to Bayesian norms, one may adopt either a *realist* or an *instrumentalist* attitude towards Bayesian models of those processes. Realists believe those models provide approximately true descriptions of such processes. In contrast, instrumentalists believe only that such models are useful tools for making particular predictions about those processes, even if those models are not correct descriptions of them. Such attitudes about *some* mental processes are consistent with non-Bayesian models of other mental processes. Michael Rescorla articulates this position in a recent paper:

> Bayesian cognitive science does not regard Bayesian norms as anything like universal psychological laws. The goal is not to establish that all mental processes conform to Bayesian norms. The goal is to investigate the extent to which various mental processes conform to Bayesian norms and, in cases where they closely conform, to construct good explanations on

> that basis. Bayesian cognitive scientists can happily say that some mental processes conform to Bayesian norms while others do not. (Rescorla, 2020, 52)

Call this position *local descriptive Bayesianism*. Such a position is completely consistent with our experimental results.

Our position is also consistent with a third way of understanding the core claims of Bayesian cognitive science—that is, as more of a *normative* and *methodological* claim. In particular, the claim is that Bayesian norms provide normative standards against which to assess at least some mental processes, and one of the aims of cognitive science is to develop interventions to bring these mental processes in conformity with those standards. I call this *normative Bayesianism*.

Note that normative Bayesian is independent of—and consistent with—local descriptive Bayesianism. Normative Bayesianism can accept that some mental processes conform to Bayesian norms. To that extent, it is consistent with local descriptive Bayesianism. But normative Bayesianism can also accept that there are other processes which deviate from those norms, and the normative Bayesian then seeks to develop interventions to bring such processes into conformity with Bayesian processes. But note that this is independent of local descriptive Bayesianism: one can be a local descriptive Bayesian without endorsing Bayesian norms or developing interventions in those cases where mental processes deviate from Bayesian norms.

In any case, our results are consistent with local descriptive Bayesianism and normative Bayesianism. In fact, the very approach of this paper has been normatively Bayesian: to understand why human reasoning departs from Bayesian reasoning and to test an intervention that narrows the gap between the two.

Appendix A: Training Materials

The following page features experiment 1's reading materials for training participants in the mental simulations approach.

Participant Materials
KEEP THIS TAB OPEN THROUGHOUT THE STUDY

Important: This material is the intellectual property of Stanford researchers. Please do not distribute it in any way.

Background to the Study

We all make judgments about probabilities. You might choose one job rather than another because of your judgment that you will probably be happier in that job. Or you might take some medication because of your judgment that it is probably safe.

This study is about probabilities.

However, humans are susceptible to various *cognitive biases*—that is, errors in their judgment. This study aims to teach you a method to help you overcome a particular cognitive bias that you might fall prey to when reasoning about probabilities.

To teach you the method, we want to walk you through a hypothetical scenario. **Note that you will be asked questions about this and another problem later, and your ability to give correct answers in this study will determine whether you receive the bonus payment (if you are completing this study for payment).**

Also, this material asks you to do various tasks, such as drawing circles. We encourage you to do this with some paper and a pen or pencil, if you have these items. Otherwise, if you do not have these items, just follow these materials to the best of your ability.

Let us now consider the hypothetical scenario.

The Story of the Prisoners

Imagine that you and three other people—Alison, Billy and Carly—are in prison. Three of you will be imprisoned for life, and one of you will be set free. A lottery was used to randomly determine who will be set free. So each of you have an equal chance of being set free at the beginning of this story.

The prison warden knows who will be set free, and you ask him if he can tell you who it is. He says that he can tell you the names of ***two*** prisoners who will ***not*** be set free, but he cannot tell you whether you will be set free or not. We will also suppose he cannot lie about who will be set free.

He then tells you that Billy and Carly will ***not*** be set free. Consequently, either you or Alison will be set free.

Now, once the warden has given you this testimony—that is, his statement about who will ***not*** be set free—which of the following is true: you are more likely to be set free, Alison is more likely to be set free, or both of you are equally likely to be set free?

At this point, an intuitive answer is that you and Alison are equally likely to be set free. After all, only two options remain, and you both started off with an equal probability of being set free. This answer, however, is incorrect. It results from a cognitive bias—an error in human judgment. We want to teach you an approach to correct this bias.

Surprisingly, the correct answer is that **Alison is more likely to be set free**.

To see how this is so, we will use the *mental simulations* approach to probabilistic reasoning.

The Mental Simulations Approach

The core idea behind this approach is that we will run so-called *mental simulations* of the scenario in our mind—that is, we will imagine that the scenario with the prisoners happened a number of times. We will then ask ourselves the question: who is more likely to be set free? To correctly calculate the relevant probabilities with these simulations, we need to think about two kinds of probabilities: prior probabilities and the probability of the evidence. Let us consider these in more detail.

Prior Probabilities

We need to first consider the *prior probabilities* of who will be set free—that is, the probability of being set free *prior* to receiving some evidence. In this case, the evidence is the warden's testimony that Bill and Carly will ***not*** be set free.

We will then consider the probability of getting this testimony given the various possible outcomes for who will be set free. But for now, we are just considering the prior probabilities of the outcomes.

At the beginning of our story, then, there are four outcomes:

- Outcome 1 = You will be set free
- Outcome 2 = Alison will be set free
- Outcome 3 = Billy will be set free
- Outcome 4 = Carly will be set free

Remember that which prisoner will be set free is determined by a random lottery, so each person initially has an equal prior probability of being set free. For example, the prior probability that you will be set free is 1/4 or 25%, and it is the same with the other outcomes.

Let us then imagine a number of simulations of this situation, say, 12 simulations (we will explain exactly why the number 12 was chosen later on). We can depict these simulations in different ways.

One way to depict them is with circles, supposing that each circle represents a time that the scenario happens. You can see this here:

12 ‘mental’ simulations of the Prisoners Story

The first step of the mental simulations approach is then to image some simulations.

We will now imagine that in some of these simulations, you will be set free, while in the other simulations, the others will be set free. The second step is then to **proportion the number of simulations** where a given outcome is true by the **prior probability of that outcome**.

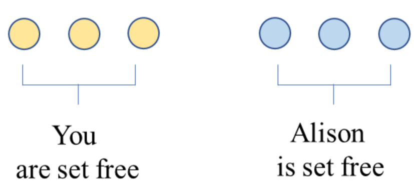
So, for instance, since you have a 25% prior probability of being set free, we will make it so that you are set free in 25% of the simulations—that is, in 3 of the 12 simulations.

To do this, if you have writing materials, go ahead and depict the 12 simulations using circles like how it was done above. Below, this proportioning has already been done for the outcomes where you or Alison are set free, but you need to do it for the other two outcomes: the outcome where Billy will be set free and the outcome where Carly will be set free.

So go ahead and proportion the 6 simulations for the remaining outcomes based on the prior probability of those outcomes. This could be done by making it so that for some circles, an outcome is true, as you can see below:

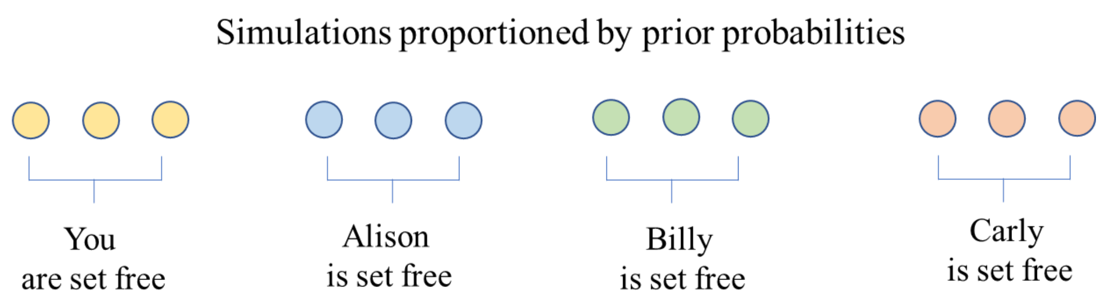
Simulations proportioned by prior probabilities

[Proportion the remaining simulations here]



Once you are done, answer the question on the webpage and continue to the next section.

If you did that exercise correctly, then a given prisoner will be set free in 25% of the simulations because they each have a 25% prior probability of being set free. What you should then have would look something like this:



So the first step of the mental simulations approach is to imagine some simulations, and the second step is to proportion those simulations by the prior probabilities. The third step is to then further proportion these simulations by the second kind of probability.

The Probabilities of the Testimony given the Outcomes

The second kind of probability that we need to consider is the probability of the testimony given the various outcomes. Recall that the testimony was this:

Warden's testimony = the warden's statement that Billy and Carly will ***not*** be set free

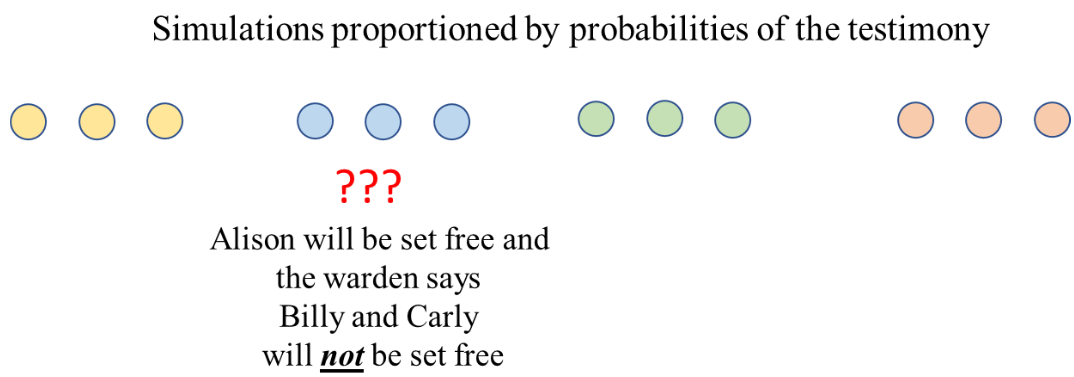
Also recall that the warden said he cannot tell you whether you will be set free, and he can tell you the names of *only* two people who would ***not*** be set free. He then gave you his truthful testimony that Billy and Carly will ***not*** be set free.

We now need to consider how probable this testimony would be given the various outcomes. Once we know how probable the evidence is for a given outcome, we then need to proportion the simulations by that probability.

For example, consider the outcome where Billy will be set free. If Billy was to be set free, then the warden would not have truthfully told you that Billy and Carly would not be set free. This is because we have supposed that the warden cannot lie. So there is a 0% probability that the warden would give you his testimony if Billy was to be set free. For that reason, we then proportion the simulations so that the warden gives you his testimony in 0% of the simulations where Billy will be set free.

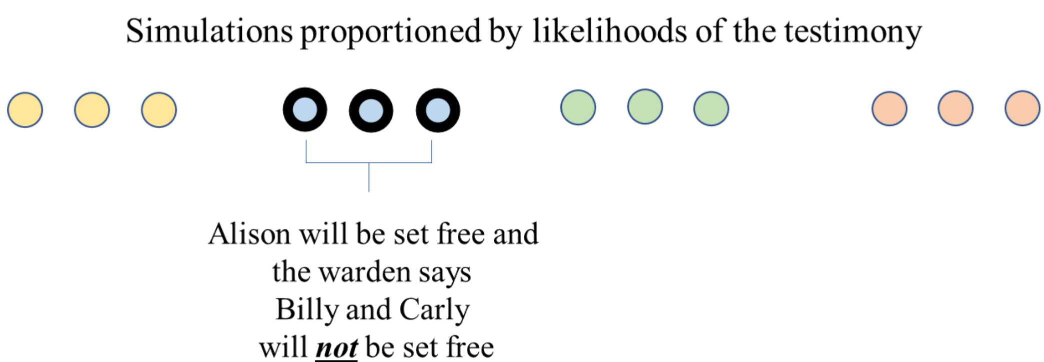
Similarly, we will make it so that the warden gives you his testimony in 0% of the simulations where Carly will be set free. Again, this is because the warden cannot lie and there is a 0% probability that the warden would give you his testimony if Carly was to be set free.

Now consider the simulations where Alison will be set free. In how many of these would the warden give you his testimony?

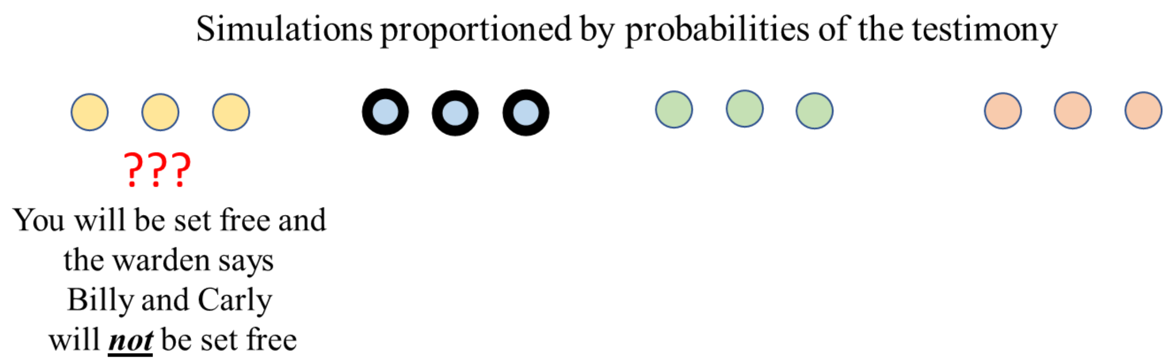


The correct answer is that the warden would give you his testimony in 100% of the simulations where Alison will be set free. This is because if Alison was to be set free, then there is a 100% probability that he would tell you that Billy and Carly will ***not*** be set free. And the reason for this is that he would **have** to tell you that Billy and Carly will ***not*** be set free: because he cannot lie, he would not say that Alison would ***not*** be set free if she was to actually be set free, and because he cannot tell you your fate, he cannot tell you that you will ***not*** be set free.

So if you have writing materials, go ahead and make it so that the warden gives you this testimony in 100% of the simulations where Alison will be set free. You can do this by circling the simulations as you can see here:



Now consider the simulations where you will be set free. In how many of these would the warden give you his testimony?



Enter your answer on the webpage and then proceed to the next section.

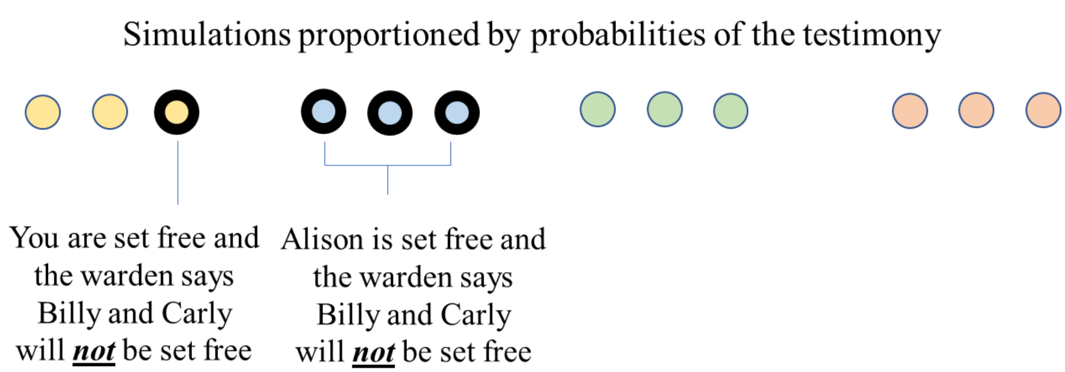
The correct answer is that the warden would give you his testimony in 1/3 or approximately 33% of the simulations where you will be set free.

To see how this is so, let us consider the probability that he would give you the same testimony if ***you*** were to be set free. So now imagine that you will be set free. Then, the warden could have given you one of any three combinations of names about who will ***not*** be set free. He could have said:

1) that Alison and Billy will ***not*** be set free
2) that Alison and Carly will ***not*** be set free, or
3) that Billy and Carly will ***not*** be set free

Since there are three combinations which the warden could have said, the probability that the warden would tell you that Billy and Carly will ***not*** be set free is 1/3 if you were to be set free. For this reason, we will make it so that the warden tells you that Billy and Carly will ***not*** be set free in 1 out of 3 of the simulations where you will be set free.

This is again depicted here:



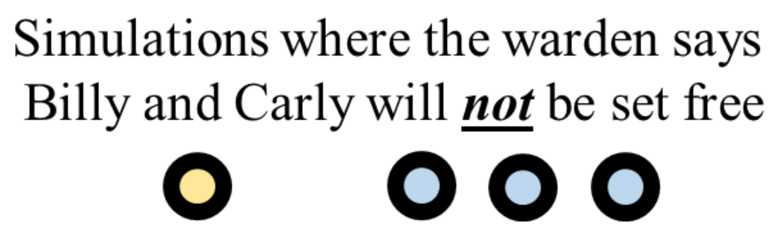
So we have seen the first three steps of the mental simulations approach: first, imagine some simulations; second, proportion the simulations by the prior probabilities; and third, then proportion the simulations by the probability of the evidence.

The fourth step is then to get rid of the simulations without the evidence. Let us explore this in more detail.

Eliminate Irrelevant Simulations and Calculate the Probabilities

Now, we can calculate the probability that you or Alison will be set free given the warden's testimony. To do that, we just consider only the simulations where the warden gave you his testimony. The rationale for this is intuitive: since, in the story, you are in a situation where you have been given this testimony, it makes sense to calculate probabilities only with reference to the simulations where the warden has given you this testimony.

We can depict the remaining simulations by crossing out or removing the circles where the evidence does not obtain, as you can see here:



Once we have eliminated the simulations without the evidence, we can carry out the fifth and final step: we can calculate the probabilities of the outcomes given the evidence by counting the remaining outcomes. Here, we can see that there are only 4 simulations where the warden gave you this testimony, and in 3 of those, Alison will be set free. For that reason, the probability that Alison will be set free is 3/4 or 75% and not 1/2 or 50%, as we might have initially thought. We can now see why it is more probable that Alison will be set free: in this case, the evidence is more probable given that outcome. In other words, the warden is more likely to give you the testimony that he did if Alison was to be set free, and this is why there are more simulations where Alison will be set free after we have eliminated the simulations that do not have the evidence.

FAQs about the Approach

We will now answer some frequently asked questions about the approach.

Why is the answer that this approach gives the correct one?

This approach not only aims to help us calculate the probabilities about who will be set free, but it also helps us to understand *why those probabilities are the correct ones.* This is because the mental simulations approach provides a snapshot of what would happen *if the scenario was to happen a large number of times*. More specifically, if the story of the prisoners was to happen, say, 100,000 times, then we can see that about 75% of the time, Alison would be released when the warden has given you that information. And we can see why this is the case. Those scenarios are first proportioned by the prior probabilities of the outcomes, and then by the probabilities of the testimony given those outcomes. This enables us to see the frequency with which Alison will be set free among the times when we have been given that information.

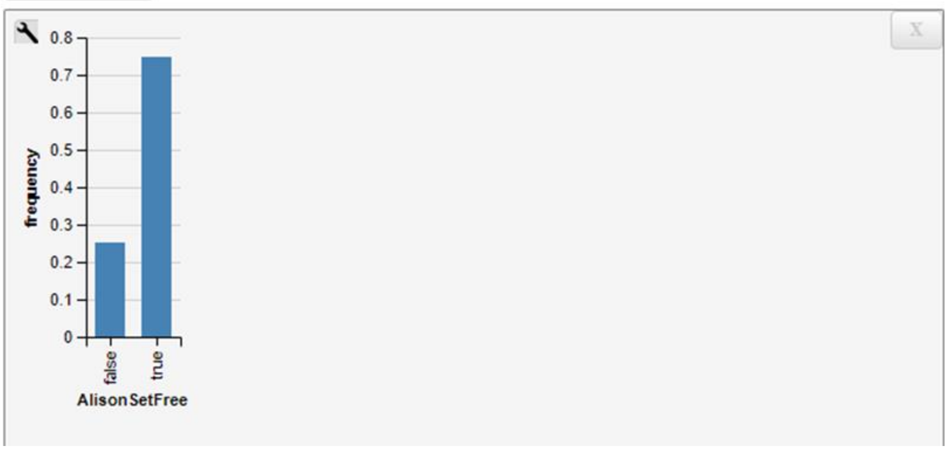
Of course, in real life, the frequency with which something happens does not always match the probability of that thing: a coin may have a 50% probability of landing heads, but if you toss it 10 times, it might land heads on 30% of those times. Nevertheless, we can correctly calculate probabilities if we suppose that the probability of something matches the frequency with which that thing happens in these simulations. For example, if an outcome has a prior probability of 25%, then we imagine that it will happen in 25% of the simulations.

We can also show that that the mental simulations approach delivers the right answers if we run computer simulations.

Here, for instance, we provide an example of a program in a programming language called *WebPPL*. We ran the program many times to generate 100,000 computer simulations where the warden gives you his testimony, and in approximately 75% of them, Alison was set free instead of you:

```
//Four prisoner story simulator

var Simulations = Infer({method: 'rejection', samples: 100000}, function (){

//Assigns prior probabilities to the possible outcomes
  var YouSetFree = flip(1/4)
  var AlisonSetFree = flip(1/4)
  var BillySetFree = flip(1/4)
  var CarlySetFree = flip(1/4)

//This is a constraint to make it so that only one prisoner will be set free
  condition(YouSetFree + AlisonSetFree + BillySetFree + CarlySetFree == 1)

//This assigns probabilities to the warden's testimony given the outcomes
  var WardensTestimony = (YouSetFree && flip(1/3))||
                         (AlisonSetFree && flip(3/3))||
                         (BillySetFree && flip(0/3))||
                         (CarlySetFree && flip(0/3))

  condition(WardensTestimony)

  return{AlisonSetFree: AlisonSetFree}
})

Simulations
```

run



Why does intuition give the wrong answer?

We can also see, from this approach, why the intuitive answer—that you and Alison are equally likely to be free—is the incorrect answer. The reason is that it fails to incorporate the probability of the evidence given the various possible outcomes: *it is more probable that the warden would give you his testimony if Alison was to be set free than if you were to be set free*. In this sense, most people do not properly account for the fact that the evidence is more probable given one possible outcome rather than another. We then display a *cognitive bias* when we fail to correctly consider these probabilities and their implications for the probability of the outcomes. The mental simulations approach helps us to avoid that cognitive bias and to see how the probability of the evidence affects the probability of the outcomes given that evidence.

So even though the intuitive answer is wrong, we can nevertheless replace the incorrect intuition with better intuitions. To do so, consider other cases where it is more obvious that some evidence favors one outcome over another if the evidence is more probable given that outcome than given the other.

Let us consider an analogy. Suppose you test positive for a disease. It is more likely that you would test positive if you had the disease than if you did not. So, intuitively, the positive test raises the probability that you have the disease!

We can now apply that same intuition to the story of the prisoners. Suppose the warden tells you that Billy and Carly will not be set free. As mentioned previously, it is more likely that he would give you that testimony if Alison was to be set free than if you were to be set free. So, intuitively, the warden's testimony should raise the probability that Alison will be set free.

How many simulations should we use with approach?

How many simulations do we need to imagine with the approach? The answer is this: whatever number lets you do the proportioning! In particular, there are two things to proportion. The first are the prior probabilities; in our story, these are each 1/4 or 25%. The second things to proportion are probabilities of the evidence given the outcomes; in our story, these varied from outcome to outcome. In our story, 12 simulations work.

But note that we did not need to run these mental simulations with *exactly* 12 simulations. For example, 36 simulations also work. We could have made each outcome true in 9 simulations before proportioning the simulations by the probabilities of the testimony so that Alison is free in 9 simulations where the warden gives you the testimony and you are set free in 3 of the simulations where the warden gives you the testimony. In this case, the probability that Alison will be set free is still $9/12 = 3/4 = 75\%$.

The only thing that matters is that the number of simulations—whatever it is—can be proportioned first by the prior probabilities and then by the probabilities of the testimony given the various outcomes.

Summary of the Approach

Here, then, is a summary of steps in the mental simulations approach:

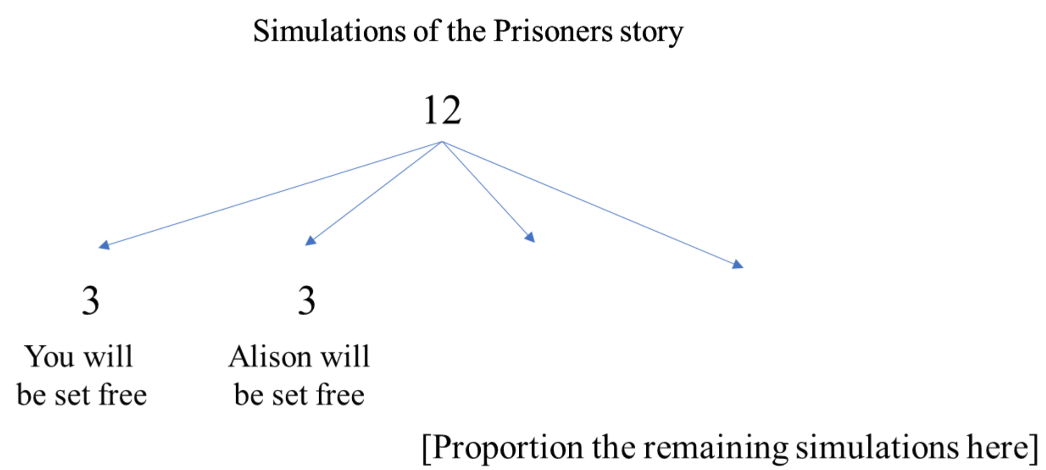
1. Imagine some simulations:
    - Imagine *n* number of simulations (where *n* is any number that can be proportioned by the prior probabilities and then by probability of the evidence)
2. Proportion according to prior probabilities of the outcome:
    - For each possible outcome, make the proportion of simulations where that outcome is true correspond to the prior probability of that outcome
3. Proportion according to probabilities of the evidence:
    - For each set of simulations for a given outcome, make the proportion of simulations where the evidence obtains correspond to the probability of that evidence given that outcome
4. Eliminate irrelevant simulations:
    - Remove the outcomes where the evidence does not obtain
5. Calculate probabilities:
    - Determine the proportion of the remaining simulations where a particular outcome is true; this is the probability of that outcome given the evidence

So that is the mental simulations approach to probabilistic reasoning.

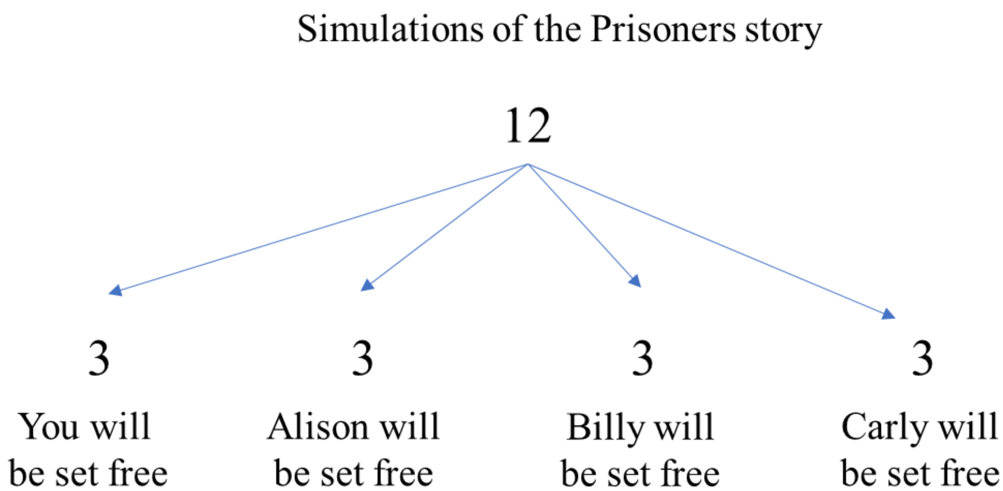
Mental Simulations Using Numbers

To help you further internalize the mental simulations approach, the below exercise asks you to repeat the above procedure, but by using an approach where the simulations are represented with numbers instead of circles.

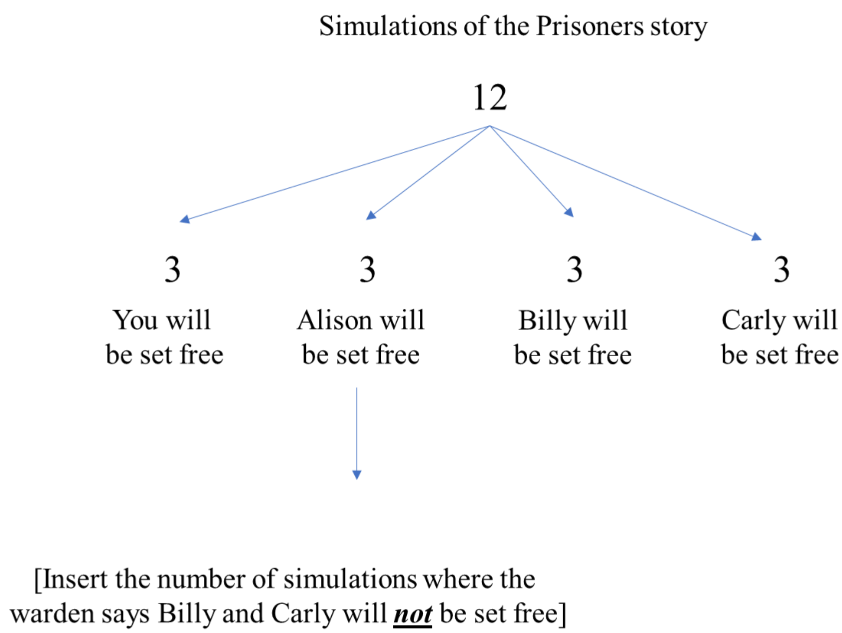
First, proportion the remaining number of simulations where an outcome is true by the prior probability of that outcome. This has already been done for two outcomes, but not for the others.



If you did that right, you should have something like what is on the following page.

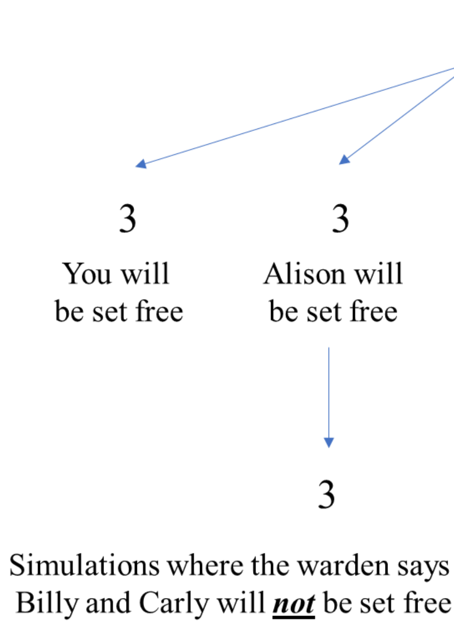


Now proportion the number of outcomes where Alison will be set free by the probability of the warden's testimony that Billy and Carly will ***not*** be free if Alison was to be set free. (Remember, the probability is 100%.)



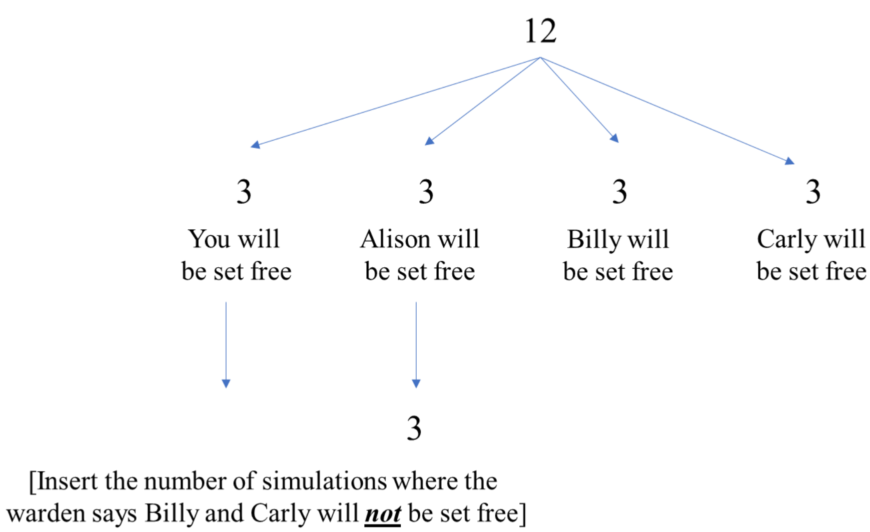
If you did that correctly, you should have something like what follows:

Simulations of the Prisoners story

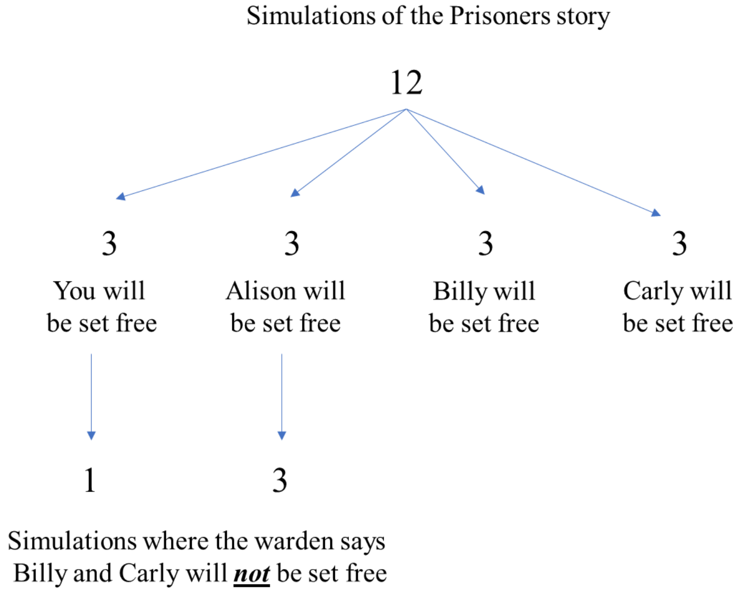


Now proportion the number of outcomes where you are set free by the probability of the warden's testimony that Billy and Carly will ***not*** be free if you were to be set free. (Remember, the probability is 1/3.)

Simulations of the Prisoners story

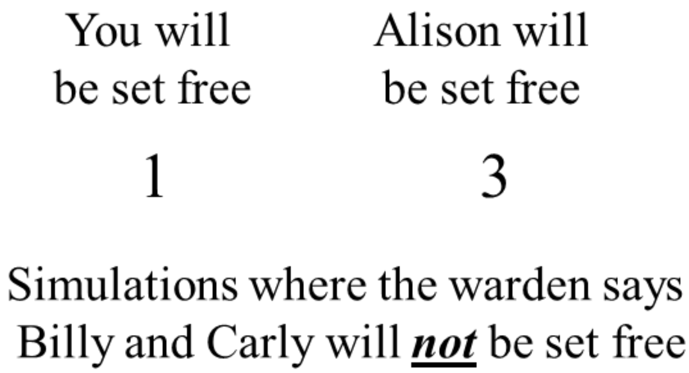


If you did that correctly, you should have something like this.



Now eliminate the irrelevant simulations, considering ***only*** the simulations where the wardens says Billy and Carly will ***not*** be set free:

If you did that correctly, you would have something like this:



Now count the number of these simulations where Alison will be set free over the total number of remaining simulations where the warden says Billy and Carly will ***not*** be set free. This will give you the probability that Alison will be set free given the warden's testimony—a probability of 3/4.

So that is one way to use the mental simulations approach—with numerical representations.

We will now present you with a final problem involving probabilities. Please complete the problem by using the survey in the link. You are free to solve the problem in whatever way you think is fitting.

Appendix B: Proofs of Main Results

In this appendix, I prove two results:

1. The law of likelihood
2. The equivalence of the results delivered by Bayes' theorem and the mental simulations approach

1. The Law of Likelihood

Our version of the law of likelihood states the following:

$$\text{If } P(e|h_1) > P(e|h_2), \text{ then } \frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)}$$

To prove this theorem of the probability calculus, I prove the stronger theorem that:

$$\frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)} \text{ if and only if } P(e|h_1) > P(e|h_2)$$

To do this, first suppose the rightmost condition holds—that is:

$$\frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)}$$

Then, by Bayes's theorem, this condition holds if and only if the following holds:

$$\frac{\frac{P(e|h_1)P(h_1)}{P(e)}}{\frac{P(e|h_2)P(h_2)}{P(e)}} > \frac{P(h_1)}{P(h_2)}$$

Next, we multiply both side by $\frac{P(e)}{P(e)}$:

$$\frac{\frac{P(e|h_1)P(h_1)}{P(e)}}{\frac{P(e|h_2)P(h_2)}{P(e)}} \cdot \frac{P(e)}{P(e)} > \frac{P(h_1)}{P(h_2)} \cdot \frac{P(e)}{P(e)}$$

Which is the same as:

$$\frac{\frac{P(e|h_1)P(h_1)}{P(e)} \cdot P(e)}{\frac{P(e|h_2)P(h_2)}{P(e)} \cdot P(e)} > \frac{P(h_1)}{P(h_2)} \cdot 1$$

Which is equivalent to:

$$\frac{P(e|h_1)P(h_1)}{P(e|h_2)P(h_2)} > \frac{P(h_1)}{P(h_2)}$$

And this is equivalent to:

$$\frac{P(e|h_1)}{P(e|h_2)} \cdot \frac{P(h_1)}{P(h_2)} > \frac{P(h_1)}{P(h_2)}$$

And when both sides of the inequality are divided by $\frac{P(h_1)}{P(h_2)}$, we have the following:

$$\frac{P(e|h_1)}{P(e|h_2)} > 1$$

Then we can multiply both sides of the inequality by $P(e|h_2)$:

$$P(e|h_1) > P(e|h_2)$$

We have then proved that:

$$\frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)} \text{ if and only if } P(e|h_1) > P(e|h_2)$$

Our version of the law of likelihood then follows:

$$\text{If } P(e|h_1) > P(e|h_2), \text{ then } \frac{P(h_1|e)}{P(h_2|e)} > \frac{P(h_1)}{P(h_2)}$$

2. Bayes' theorem/mental simulations equivalence

I now turn to prove that the mental simulations approach delivers answers that always accord with Bayes' theorem.

To do this, I first formally characterize the answer that is delivered by the mental simulations approach, and I show that this aligns exactly with the answer delivered by Bayes' theorem.

Formally Characterizing the Mental Simulations Approach:

First, mental simulations approach asks one to imagine N simulations. Let us denote the total number of simulations with N_T.

Further, it asks us to proportion the simulations where a given outcome is true by the prior probability of that outcome. Formally, let $\{h_1, \ldots, h_k\}$ be the set of *k* mutually exclusive outcomes, exactly one of which is true. Let N_{h_j} be the number of simulations where h_j is true for any h_j in $\{h_1, \ldots, h_k\}$. Then, by stipulation, the proportion of total simulations where a given outcome h_j is true is equal the prior probability of that outcome. So:

1) $\frac{N_{h_j}}{N_T} = P(h_j)$

The next step in the mental simulations approach is then to take all the simulations where a given outcome h_j is true, and to then proportion those N_{h_j} simulations by the probability of the evidence. Formally, let $N_{h_j\&e}$ be all the simulations for a given outcome h_j where e is true. Then, this step in the approach is to make the proportion of the simulations where h_j and e is true equal to the probability of the evidence given that outcome. So:

2) $\frac{N_{h_j\&e}}{N_{h_j}} = P(e|h_j)$

Then, we have to eliminate the outcomes where the evidence is not true, considering only the total number of where the evidence is true. This is number is given by the following equation:

3) $N_e = N_{h_1\&e} + \cdots + N_{h_k\&e}$ where N_e is the total number of simulations where the evidence e is true

What this says is that the total number of simulations where the evidence is true is equal to the sum of all the simulations where the evidence is true across the simulations for all the exhaustive and mutually exclusive outcomes $\{h_1, \ldots, h_k\}$.

The mental simulations approach then tells us that we can calculate the probability of a given hypothesis h_j by calculating the proportion of simulations where h_j is true among all the simulations where e is true. Put formally, the mental simulations approach claims that:

4) $P(h_j|e) = \frac{N_{h_j\&e}}{N_e}$

So we have now characterized the steps in the mental simulations approach. I now aim to prove that, given the stipulations 1) and 2) in the earlier steps of the mental simulations approach, claim 4) is indeed true—that is to say, that the mental simulations approach agrees with Bayes' theorem:

$$P(h_j|e) = \frac{N_{h_j\&e}}{N_e} = \frac{P(e|h_j)P(h_j)}{P(e)}$$

Proof of the Bayes'/Mental Simulations Agreement:

To prove this, let us start with Bayes' theorem:

$$P(h_j|e) = \frac{P(e|h_j)P(h_j)}{P(e)}$$

I will first prove that $P\left(e\middle|h_j\right)P\left(h_j\right) = \frac{N_{h_j\&e}}{N_T}$ and then that $P(e) = \frac{N_e}{N_T}$. Since, by algebra, $\frac{\frac{N_{h_j\&e}}{N_T}}{\frac{N_e}{N_T}}$ is equivalent to $\frac{N_{h_j\&e}}{N_e}$, it will follow that $P\left(h_j\middle|e\right) = \frac{P\left(e\middle|h_j\right)P\left(h_j\right)}{P(e)} = \frac{\frac{N_{h_j\&e}}{N_T}}{\frac{N_e}{N_T}} = \frac{N_{h_j\&e}}{N_e}$.

First, to prove $P\left(e\middle|h_j\right)P\left(h_j\right) = \frac{N_{h_j\&e}}{N_T}$. Recall stipulations 1) and 2) above:

1) $\frac{N_{h_j}}{N_T} = P\left(h_j\right)$

2) $\frac{N_{h_j\&e}}{N_{h_j}} = P\left(e\middle|h_j\right)$

Given these stipulations, it follows that:

$$P\left(e\middle|h_j\right)P\left(h_j\right) = \frac{N_{h_j\&e}}{N_{h_j}} \cdot \frac{N_{h_j}}{N_T}$$

By algebra then:

$$P\left(e\middle|h_j\right)P\left(h_j\right) = \frac{N_{h_j\&e}}{N_{h_j}} \cdot \frac{N_{h_j}}{N_T} = \frac{N_{h_j\&e} \cdot N_{h_j}}{N_{h_j} \cdot N_T} = \frac{N_{h_j\&e}}{N_T}$$

We have then proved the first step—that $P\left(e\middle|h_j\right)P\left(h_j\right) = \frac{N_{h_j\&e}}{N_T}$.

Now to prove that $P(e) = \frac{N_e}{N_T}$.

Recall 3) above:

1) $N_e = N_{h_1\&e} + \cdots + N_{h_k\&e}$

Then, dividing both sides by N_T, we have the following:

$$\frac{N_e}{N_T} = \frac{N_{h_1\&e} + \cdots + N_{h_k\&e}}{N_T} = \frac{N_{h_1\&e}}{N_T} + \ldots + \frac{N_{h_k\&e}}{N_T}$$

Then recall the earlier theorem we proved:

$$P\left(e\middle|h_j\right)P\left(h_j\right) = \frac{N_{h_j\&e}}{N_T}$$

It then follows that:

$$\frac{N_e}{N_T} = \frac{N_{h_1\&e}}{N_T} + \ldots + \frac{N_{h_k\&e}}{N_T} = P(e|h_1)P(h_1) + \cdots + P(e|h_k)P(h_k)$$

Yet we also know by the theorem of total probability that if $\{h_1, \ldots, h_k\}$ are exhaustive and mutually exclusive outcomes, then:

$$P(e) = P(e|h_1)P(h_1) + \cdots + P(e|h_k)P(h_k)$$

But since we have shown that:

$$\frac{N_e}{N_T} = P(e|h_1)P(h_1) + \cdots + P(e|h_k)P(h_k)$$

As desired, it follows that:

$$P(e) = \frac{N_e}{N_T}$$

We have now proved that $P(e|h_j)P(h_j) = \frac{N_{h_j\&e}}{N_T}$ and that $P(e) = \frac{N_e}{N_T}$. It then follows that:

$$P(h_j|e) = \frac{P(e|h_j)P(h_j)}{P(e)} = \frac{\frac{N_{h_j\&e}}{N_T}}{\frac{N_e}{N_T}} = \frac{N_{h_j\&e}}{N_e}$$

So we have proven that the mental simulations approach always agrees with Bayes' theorem—that is:

$$P(h_j|e) = \frac{P(e|h_j)P(h_j)}{P(e)} = \frac{N_{h_j\&e}}{N_e}$$

Appendix C: Computer Simulations of the New Monty Hall Problem

In the paper, I claimed that there is a $\frac{10}{19}$ or 53% chance that door B conceals the prize in the New Monty Hall problem. Recall that the New Monty Hall problem is the same as the original Monty Hall problem, except that Monty Hall would have a 90% chance of opening door C when door A is selected and door A also conceals the prize.

This claim should be uncontroversial since it is derived directly from Bayes' theorem—the exact same mathematical machinery that tells us to switch doors in the original Monty Hall problem.

However, to further support this claim, I ran 100,000 computer simulations of the New Monty Hall problem. The results showed that door B concealed the prize approximately 53% of the time.

I have included the code below so that the reader may replicate and verify this result for themselves.

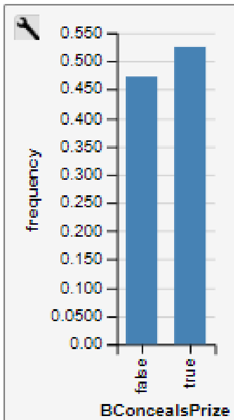
I used a probabilistic programming language called *WebPPL*—freely available at http://webppl.org/.

On it, I implemented a method of computer simulation known as *rejection sampling*. This runs numerous simulations of the probabilistic setup, and it then rejects any simulations where the specified conditions are not met. In our case, the specified conditions were that exactly one of the three doors conceals the prize, that door A is selected, and that door C is then opened. In these conditions, door B concealed the prize 53% of the time.

Here is the code which the reader can implement and modify for their purposes. Note in particular the variable *L*—where *L* strands for 'the likelihood of opening the right-most door'. Adjusting this variable changes likelihood that Monty Hall would open the right-most door that is unselected and does not conceal the prize. If it is set to .5, then the setup is equivalent to the original Monty Hall problem, and the probability that door B conceals the prize is $\frac{2}{3}$. If it is set to .9, then the setup is equivalent to the New Monty Hall problem, and the probability that door B conceals the prize is $\frac{10}{19}$ or 53%.

```
// Monty Hall problem simulator

var posterior = Infer(
//This method calculates exact probabilities
//  {method: 'enumerate'},
//This method instead runs simulation using rejection sampling
  {method: 'rejection', samples: 100000},
  function () {

// Prior probabilities
    var AConcealsPrize = flip(1/3)
    var BConcealsPrize = flip(1/3)
    var CConcealsPrize = flip(1/3)

//This is a constraint to make A, B and C exhaustive and mutually exclusive hypotheses
// i.e. that only one door conceals a prize
    condition(AConcealsPrize + BConcealsPrize + CConcealsPrize == 1)

//Door selected by participant
  var SelectedDoor = (flip(1/3) ? 'A':
                      flip(1/2) ? 'B':
                      'C')

//This variable specifies the likelihood of rightmost door
  var L = .9

//Likelihoods for phase 2
  var DoorIsOpened = (SelectedDoor == 'A' && AConcealsPrize)? (flip(L)? 'C': 'B'):
                     (SelectedDoor == 'A' && BConcealsPrize)? 'C':
                     (SelectedDoor == 'A' && CConcealsPrize)? 'B':
                     (SelectedDoor == 'B' && AConcealsPrize)? 'C':
                     (SelectedDoor == 'B' && BConcealsPrize)? (flip(L)? 'C': 'A'):
                     (SelectedDoor == 'B' && CConcealsPrize)? 'A':
                     (SelectedDoor == 'C' && AConcealsPrize)? 'B':
                     (SelectedDoor == 'C' && BConcealsPrize)? 'A':
                     (flip(L)? 'B': 'A')

  condition(SelectedDoor == 'A' && DoorIsOpened == 'C')

  return {BConcealsPrize: BConcealsPrize}
})

posterior
```

run



Appendix D: Reasoning Among Those Claiming They Used the Mental Simulations Approach

Description of Reasoning Process	Correct Posteriors
There was a 100% chance that if the prize was behind door B, Monty would open door C. There was a 50% chance that he would open door B if the prize was behind door A. I took a pen and paper and drew two sets of four circles. I labeled one set "A" and one set "B". Following the approach described earlier, I could visually see that the likelihood of the prize being behind door B was more probable. I had two dark circles under set "A" vs all four circles darkened under set "B". I then eliminated the irrelevant simulations in set "A" which left me with roughly a 66% probability (or 2/3) that the prize would be behind door B. Hopefully my thought process is correct here.	Yes
I set up the problem like the example problem with 6 circles, 2 for A, 2 for B and 2 for C to represent the prior probabilities. The host would open door C 50% of the time and door B 50% of the time when door A contains the prize so I shaded in 1 of the 2 A circles. The host would open door C 100% percent of the time when door B contains the prize because they wouldn't show me what's behind the door I picked (A) or the door with the prize so door C is the only option. I filled in both B circles to represent that. The host would open door C 0% of the time when door C contains the prize, so I disregarded those circles and I also disregarded the unfilled circle for A. I was left with 1 filled in circle for A and 2 for B, so that is how I came to the conclusion that there was a 1/3 chance of it being A and 2/3 for B. It would make more sense to switch doors because there's a higher probability of the prize being behind door B than door A.	Yes
I just sort of guessed. I used some of the information I was given previously in the document and the current information with the game scenario. I thought when it came down to switching or staying, I figured statistically I would have a higher chance of winning if I switched my answer rather than staying with my original choice.	Yes
I tried to use my math mind, and it is obvious, from the first tast that I am awful at this game. The more I learn the more I am confused.	No
After answering the multiple-choice questions based on what I had read about the scenario, it became clear that there was one outcome was twice as likely as the other.	Yes

Appendix E: Checklist for Replication Attempts

I encourage others to attempt to independently replicate the experiments in this paper. To facilitate this, I have provided the following checklist:

Item or Criterion	Check
Access to Sources: (To be included in the final)	
- Original experiment 1 dataset:	
- Original experiment 2 dataset:	
- Analysis script:	
- Prescreening survey:	
- Survey for experiment 1:	
- Survey for experiment 2:	
- Reading materials for experiment 1:	
Participants and Materials:	
- **English:** English is a first language for participants	
- **Naivety (Experiment 1 Only):** For experiment 1 only, participants lack of prior familiarity with the original Monty Hall problem (see prescreening survey)	
- **Same Materials:** Uses the same online materials as original experiments (including the additional reading and visual materials in experiment 1)	
- **Base Payment:** Participants receive a base payment for honest and attentive completion of the study ($9USD in the original experiments)	
- **Separate Bonus Payment:** Participants receive a generous bonus payment for correct answers to particular questions ($6USD in the original experiments), but they are blind to which questions these are	
Analysis:	
- **Exclusion Criteria:** Participants are excluded if they fail any of the three basic comprehension questions (and non-naïve participants excluded in experiment 1)	
- **Same Response Coding:** Responses are coded as indicated in the analysis transcript	

References

Altman, D., Machin, D., Bryant, T., & Gardner, M. (2000). *Statistics with Confidence: Confidence Intervals and Statistical Guidelines* (2nd ed.). BMJ Books.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233. https://doi.org/10.1016/0001-6918(80)90046-3

Burns, B. D., & Wieth, M. (2004). The Collider Principle in Causal Reasoning: Why the Monty Hall Dilemma Is So Hard. *Journal of Experimental Psychology: General*, *133*(3), 434–449. https://doi.org/10.1037/0096-3445.133.3.434

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.

Cohen, J. (1994). *The Earth Is Round (p < .05)*.

De Neys, W., & Verschueren, N. (2006). Working Memory Capacity and a Notorious Brain Teaser. *Experimental Psychology*, *53*(2), 123–131. https://doi.org/10.1027/1618-3169.53.1.123

Dey, R., & Mulekar, M. S. (2018). Effect Size as a Measure of Difference Between Two Populations. *Encyclopedia of Social Network Analysis and Mining*, 715–726. https://doi.org/10.1007/978-1-4939-7131-2_110195

Eddy, D. M. (1982). *Probabilistic reasoning in clinical medicine: Problems and opportunities* (pp. 249–267).

Falk, R., & Greenbaum, C. W. (1995). Significance Tests Die Hard: The Amazing Persistence of a Probabilistic Misconception. *Theory & Psychology*, *5*(1), 75–98. https://doi.org/10.1177/0959354395051004

Gardner, M. (1959). Mathematical Games: Problems involving questions of probability and ambiguity. *Scientific American*, 174–182. https://doi.org/10.1038/scientificamerican1059-174

Gilovich, T., & Griffin, D. (2002). Introduction - Heuristics and biases: then and now. In Gilovich, Thomas., D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment* (pp. 1–18). Cambridge University Press.

Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Heuristics and biases : the psychology of intuitive judgment*. Cambridge University Press. https://searchworks.stanford.edu/view/4815978

Glass, G. V, McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Sage Publications.

Granberg, D., & Dorr, N. (1998). Further Exploration of Two-Stage Decision Making in the Monty Hall Dilemma. *The American Journal of Psychology*, *111*(4), 561. https://doi.org/10.2307/1423551

Hacking, I. (2016). *Logic of statistical inference*. Cambridge University Press.

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, *101*(2), 252–254. https://doi.org/10.1037/h0035224

Hawthorne, J. (2018). *Supplement to Inductive Logic: Likelihood Ratios, Likelihoodism, and the Law of Likelihood*. The Stanford Encyclopedia of Philosophy (Spring 2018 Edition). https://plato.stanford.edu/entries/logic-inductive/sup-likelihood.html

Hoffrage, U., Krauss, S., Martignon, L., & Gigerenzer, G. (2015). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Frontiers in Psychology*, *6*,

1473. https://doi.org/10.3389/fpsyg.2015.01473
Ichikawa, J. J., & Steup, M. (2018). *The Analysis of Knowledge*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis
Johnson-Laird, P. N. (2012). Inference in Mental Models. In *The Oxford handbook of thinking and reasoning* (pp. 134–154). Oxford University Press.
Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, *24*(1), 281–323. https://doi.org/10.1111/j.1520-8583.2010.00194.x
Kalinowski, P., Fidler, F., & Cumming, G. (2008). Overcoming the inverse probability fallacy: A comparison of two teaching interventions. *Methodology*, *4*(4), 152–158. https://doi.org/10.1027/1614-2241.4.4.152
Kaplan, M. (1996). *Decision theory as philosophy*. Cambridge University Press.
Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, *132*(1), 3–22. https://doi.org/10.1037/0096-3445.132.1.3
Liu, A.-Y. (1975). Specific information effect in probability estimation. *Q Perceptual and Motor Skills*, *4*(1), 475–478.
Meacham, C. J. G. (2014). Impermissive Bayesianism. *Erkenntnis*, *79*(6), 1185–1217. https://doi.org/10.1007/s10670-013-9547-z
Nola, R. (2013). Darwin's Arguments in Favour of Natural Selection and Against Special Creationism. *Science & Education*, *22*(2), 149–171. https://doi.org/10.1007/s11191-010-9314-3
Petrocelli, J. V., Harris, A. K., & Harris, A. K. (2011). Learning Inhibition in the Monty Hall Problem. *Personality and Social Psychology Bulletin*, *37*(10), 1297–1311. https://doi.org/10.1177/0146167211410245
Piattelli-Palmarini, M. (1994). *Inevitable illusions : how mistakes of reason rule our minds /*. Wiley,.
Prasanta S. Bandyopadhyay. (2011). *Philosophy of Statistics.* North Holland.
Rescorla, M. (2020). A Realist Perspective on Bayesian Cognitive Science. In A. Nes & T. Chan (Eds.), *Inference and Consciousness* (pp. 40–73).
Rosenhouse, J. (2009). *The Monty Hall problem : the remarkable story of math's most contentious brainteaser*. Oxford University Press.
Russell, B. (1948). *Human knowledge: Its scope and limits*. Simon and Schuster,.
Saenen, L., Heyvaert, M., Van Dooren, W., Schaeken, W., & Onghena, P. (2018). Why Humans Fail in Solving the Monty Hall Dilemma: A Systematic Review. *Psychologica Belgica*, *58*(1), 128–158. https://doi.org/10.5334/pb.274
Saenen, L., Van Dooren, W., & Onghena, P. (2015). A randomised Monty Hall experiment: The positive effect of conditional frequency feedback. *Thinking & Reasoning*, *21*(2), 176–192. https://doi.org/10.1080/13546783.2014.918562
Schechter, B. (1998). *My brain is open: The mathematical journeys of Paul Erdõs*. Simon and Schuster.
Sosa, E. (1999). How to Defeat Opposition to Moore. *Nous*, *33*(s13), 141–153. https://doi.org/10.1111/0029-4624.33.s13.7
Tubau, E., Aguilar-Lleyda, D., & Johnson, E. D. (2015). Reasoning and choice in the Monty Hall Dilemma (MHD): implications for improving Bayesian reasoning. *Frontiers in Psychology*, *6*, 353. https://doi.org/10.3389/fpsyg.2015.00353

Tubau, E., & Alonso, D. (2003). Overcoming illusory inferences in a probabilistic counterintuitive problem: The role of explicit representations. *Memory & Cognition*, *31*(4), 596–607. https://doi.org/10.3758/BF03196100

Tubau, E., Alonso, D., & Alonso, D. (2003). Overcoming illusory inferences in a probabilistic counterintuitive problem: The role of explicit representations. *Memory & Cognition*, *31*(4), 596–607. https://doi.org/10.3758/BF03196100

Tversky, A., & Kahneman, D. (1973). Availability: A Heuristic for Judging Frequency and Probability122. In *COGNlTTIVE PSYCHOLOGY* (Vol. 5).

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. *Memory and Cognition*, *30*(2), 171–178. https://doi.org/10.3758/BF03195278

vos Savant, M. (1997). *The power of logical thinking*. St. Martin's Press.

Weirich, P. (2011). The Bayesian Decision-Theoretic Approach to Statistics. In *Philosophy of Statistics* (pp. 233–261). North-Holland.

White, R. (2010). Evidential Symmetry and Mushy Credence. In *Oxford Studies in Epistemology* (pp. 161–186). Oxford University Press.

Wilcox, J. (2019). An Argument for the Principle of Indifference and Against the Wide Interval View. *The Journal for General Philosophy of Science*, *50*(4).