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 two experiments 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 propose an experiment to determine whether—as previous pilot studies indicate—participants display likelihood neglect bias and whether the bias can be corrected by training in the mental simulations approach. I then propose another experiment to compare the approach to two other prominent approaches to the Monty Hall problem, one of which involves mental models. Pilot 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.

Introduction

We all make inferences on a daily basis, especially to guide our decision-making. The heuristics and biases research program has unearthed a range of biases that can compromise such inferences (Gilovich et al., 2002).

This paper aims to provide evidence for another one: *likelihood neglect bias*. Likelihood neglect bias occurs when one violates the law of likelihood. According to the law of likelihood, 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. The bias then occurs when an individual is *aware* that the evidence is more probable given one hypothesis in comparison to another, but they nevertheless fail to let this affect their probability judgments about the hypotheses in accordance with the law of likelihood. Put differently, it occurs when the evidence is more probable given one of the hypotheses compared to another, but the evidence does not raise the probability of that hypothesis compared to the other.

This bias is well-illustrated with the Monty Hall problem. Consequently, this paper also has several other aims: to encourage the literature on the Monty Hall problem to focus on likelihood neglect; to test whether a new method for reasoning (the mental simulations approach) corrects likelihood neglect in the Monty Hall problem; and to compare how that method fares against other methods of reasoning when confronting what I call the "new Monty Hall problem".

Much of this introduction compresses or omits information found in the extended version of this paper. I therefore refer readers to the extended version if they have objections or questions when reading this paper.

A Description of the Monty Hall problem

The Monty Hall problem is a well-known brainteaser that can be described 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 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

In probability theory, it is more probable that door B conceals the prize because the evidence has a higher likelihood if door B conceals the prize. A "likelihood" is a technical term that refers to the probability of some evidence given some hypothesis (Hacking, 2016; Hawthorne, 2018; Bandyopadhyay, 2011). 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.

We can see the importance of likelihoods when we use Bayes' theorem to calculate the probability that door B conceals the prize:

(1)

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

where A stands for "door A conceals the prize",
B stands for "door B conceals the prize", C stands for "door C conceals the prize" and c stands for "Monty Hall opens door C"

Here, the prior probabilities of the hypotheses are all the same, since $P(A) = P(B) = P(C) = 1/3$.

The only thing that differs is the likelihoods, since $P(c|A) \neq P(c|B) \neq P(c|C)$. If door A concealed the prize, then Monty Hall had an equal probability of opening door B or door C, so $P(c|A) = 0.5$. If door B concealed the prize, then Monty Hall cannot open any other door aside from door C, so $P(c|B) = 1$. Door B probably then conceals the prize because the evidence that door C was opened is twice as likely given that hypothesis.

This illustrates the law of likelihood. There are different formulations of it (Hacking, 2016, chap. 5; Hawthorne, 2018). For our purposes, though, it can be understood as follows. 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]

The law of likelihood is a truism of probability theory in general—not just Bayesian probability—and it applies anywhere where the evidence is more or less probable given competing hypotheses—not just the Monty Hall problem.

The Psychology of the Monty Hall Problem

Past studies have found that participants consistently fail to switch doors in the Monty Hall problem (Burns and Wieth, 2004). Some comprehensive literature reviews explore why this is (Saenen et al., 2018; Tubau et al., 2015). Two prominent causes are discussed by Tubau et al. (2015): 1) emotional-based choice biases where participants are averse to switching from their first choice (Granberg & Dorr, 1998; Petrocelli et al., 2011), and 2) cognitive limitations in understanding and representing probabilities since participants think the two outcomes are equally probable given the evidence (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).

However, I argue that two prominent solutions to the Monty Hall problem do not work. 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

[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.

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

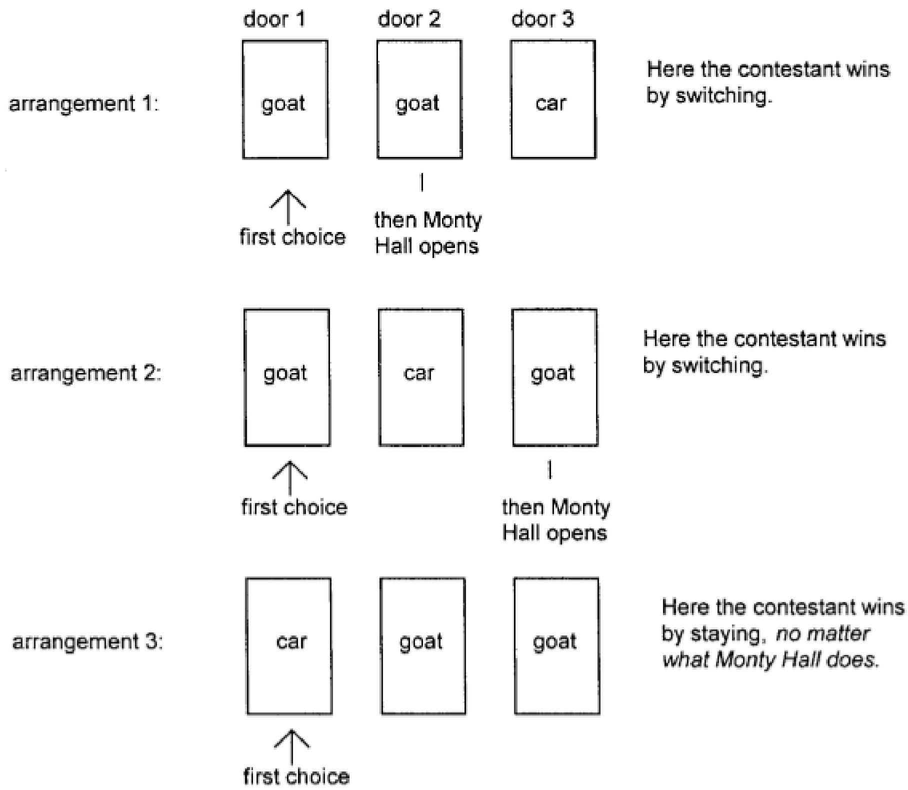


Figure 1. Explanation of the solution to the Monty Hall problem: In two out of three possible car–goat arrangements the contestant would win by switching; therefore she should switch.

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.

Another popular solution that does not work is the probability accrual approach. 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, p. 8)

The implication is that if participants were aware of these facts, they would not be susceptible to the equiprobability illusion and would instead think switching would probably be better.

Similar reasoning has been reported by participants in my pilot experiments. For example, one participant previously studied the Monty Hall problem and gave this justification for switching doors:

> 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 is that they are not sensitive to the likelihoods: one could change the likelihoods in the Monty Hall problem, thereby changing the probabilities about the location of the prize, but these approaches would still recommend the (now incorrect) two third probability that the other door conceals the prize.

Let us illustrate this with what 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' 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.*

This paper proposes an experiment to show this is indeed the case, as my previous pilot studies indicate.

The Gap in the Literature, and the Likelihood Neglect Bias

From my review of the literature, and that of Saenen et al. (2018), it appears past studies have not determined two possible mechanisms or causes of incorrect responses. One is that participants are simply unaware of what the likelihoods are, and if they were aware as such, then their posteriors would change. Another possible cause is that participants may be aware of what the likelihoods are, but this awareness does not appropriately impact their posteriors.

If the latter cause is at play, then participants are susceptible to what we can call *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.

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 I discuss later—may arise in other important contexts as well.

Nevertheless, because it is a violation of a law of probability, likelihood neglect should be regarded as a fallacy no less than other fallacies like the conjunction fallacy.

The Mental Simulations Approach to Probabilistic Reasoning

To correct likelihood bias and help participants correctly solve the Monty Hall problem, I introduce an approach to reasoning called the 'mental simulations' approach. The mental simulations approach draws on the research of Hoffrage et al. (2015) which suggests that humans often reason better in terms of natural frequencies. The approach then asks participants to imagine a number of mental simulations of the probabilistic scenario and to proportion these by the relevant probabilities to calculate the posterior probabilities.

More specifically, the approach involves these steps:

1. Generate simulations:
    - Imagine *n* number of simulations (where *n* is any number that can be divided by the priors and then the likelihoods)
    - Example: Imagine 30 mental simulations of the Monty Hall problem
2. Proportion according to prior probabilities:
    - For each possible outcome, make the proportion of simulations where that outcome is true correspond to the prior probability of that outcome

- Example: Since door A has a one third prior probability of concealing the prize, make door A conceal the prize in one third or 10 of the 30 simulations, and so on with the other two outcomes

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
    - Example: Since there is a 50% likelihood that Monty Hall would open door C if door A conceals the prize, make it so that Monty Hall opens door C in 50% or 5 of the 10 simulations where door A conceals the prize. Also make Monty Hall open door C in 100% or 10 out of 10 of the simulations where door B conceals the prize.
4. Eliminate irrelevant simulations:
    - Eliminate the outcomes where the evidence does not obtain
    - Example: Focus on only the 15 of the 30 simulations where door C is opened, including the 5 where door A conceals the prize and the 10 where door B conceals the prize.
5. Calculate probabilities:
    - Determine the proportion of simulations where a particular outcome is true; this is the probability of that outcome given the evidence
    - Example: 10 of the 15 remaining simulations are ones where door B conceals the prize, so door B has a 10/15 or 2/3 posterior probability of concealing the prize.

The approach is described in more detail in the training materials in Appendix A. T

The mental simulations approach aims to address a variety of reasoning errors. If successfully applied, the approach delivers results that conform to Bayes' theorem and thus eliminates likelihood neglect, as well the confusion of the inverse fallacy and neglect of base rates or prior probabilities. Proof of this is in appendix B. It is in principle applicable to any probabilistic problem, not just the Monty Hall problem.

The approach also integrates disparate ideas and insights about reasoning. Because its verdicts conform to Bayes' theorem, it integrates ideas from Bayesianism and probability theory more generally. Because it encourages participants to think about the causal or generative setup that produces the evidence, it resembles approaches to reasoning which emphasize causality (Krynski & Tenenbaum, 2007; Pearl, 2000). Because it encourages participants to reason in terms of natural frequencies, it resembles other approaches that also utilize natural frequencies (Hoffrage et al., 2015). And because it does all of this using mental entities to stand in for the probabilistic scenario, it resembles the mental models approach to reasoning (Johnson-Laird, 2012).

However, the mental simulations approach is not straight-forwardly the same as some prominent characterizations of the mental models approach (Johnson-Laird, 2012). This is because two or more simulations may denote the same outcome, and not all outcomes are equally probable, contra Johnson-Laird's (2012) description of the mental models approach. Additionally, the mental simulations approach does not aim to *describe* how humans actually reason; instead, it aims to *augment* reasoning by providing the simulations approach as a tool for calculating probabilities and avoiding likelihood neglect.

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. (See the discussion of the extended paper for more details here.)

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 my 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 which is 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 A conceals the prize after Monty Hall shows you the goat? *What is the probability that door B 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.

[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.

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 the experimenter presents 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.

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. Suppose 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 probable. 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 likely 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 2/3 probability that door B conceals the prize in the Monty Hall problem but a 3/4 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, a question about 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. So far, there is no evidence of social desirability response bias 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 effect size that is easiest 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 similar, say, $\left|s_e - s_c\right| > 0.75$. 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 are 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 $d \approx 0.2$, "medium" when $d \approx 0.5$ and "large" when $d \approx 0.8$ (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). Indeed, others have appraised Glass' Δ 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, p.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 groups 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

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)

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

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**.

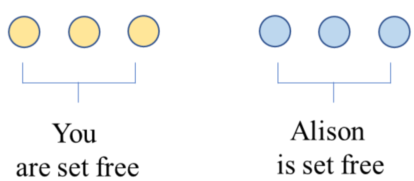
So, for instance, since you have a 25% prior probability of being set free, we will make it so that you are set free in 25% of the simulations—that is, in 3 of the 12 simulations.

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:

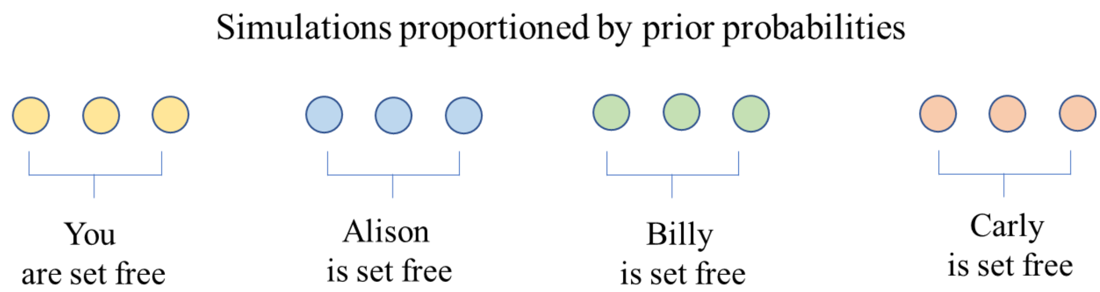
Simulations proportioned by prior probabilities

[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

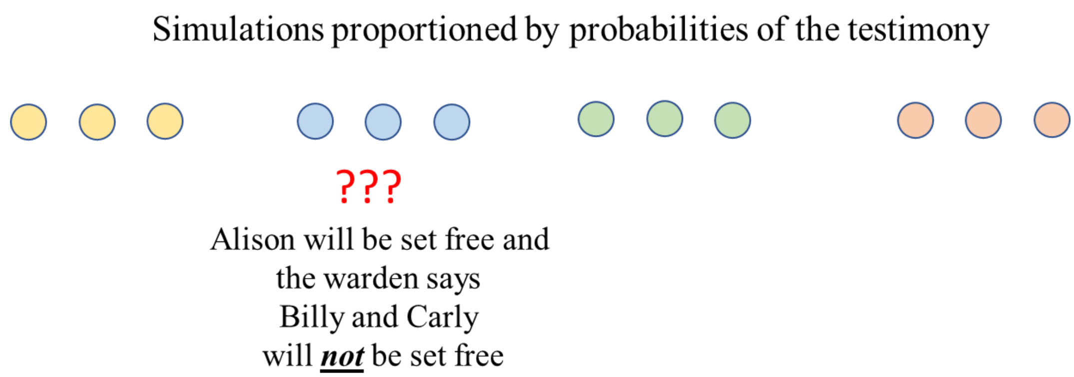
Also recall that the warden said he cannot tell you whether you will be set free, and he can tell you the names of *only* two people who would ***not*** be set free. He then gave you his truthful testimony that Billy and Carly will ***not*** be set free.

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.

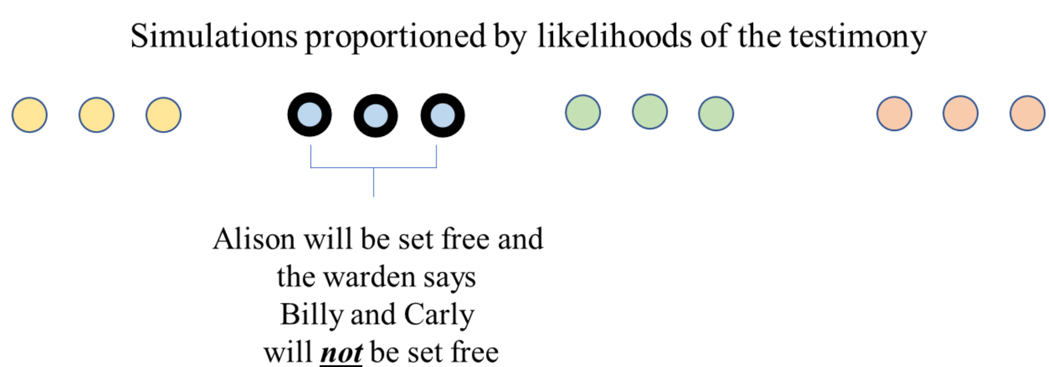
Similarly, we will make it so that the warden gives you his testimony in 0% of the simulations where Carly will be set free. Again, this is because the warden cannot lie and there is a 0% probability that the warden would give you his testimony if Carly was to be set free.

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

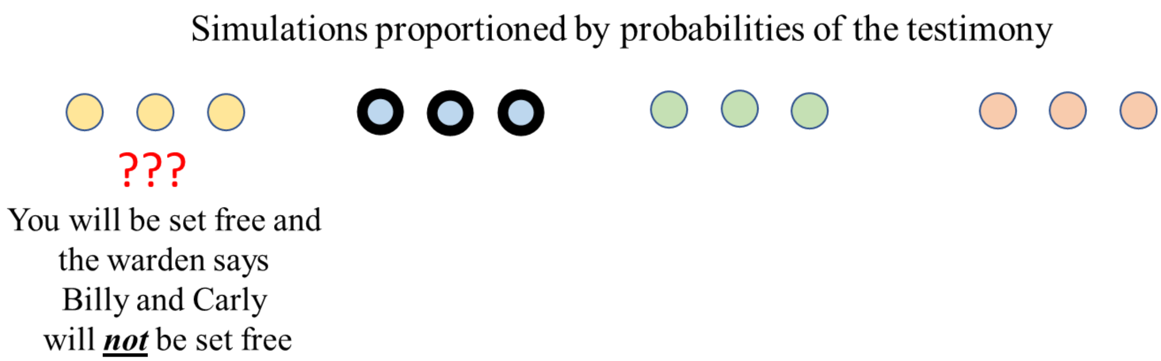


The correct answer is that the warden would give you his testimony in 100% of the simulations where Alison will be set free. This is because if Alison was to be set free, then there is a 100% probability that he would tell you that Billy and Carly will ***not*** be set free. And the reason for this is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

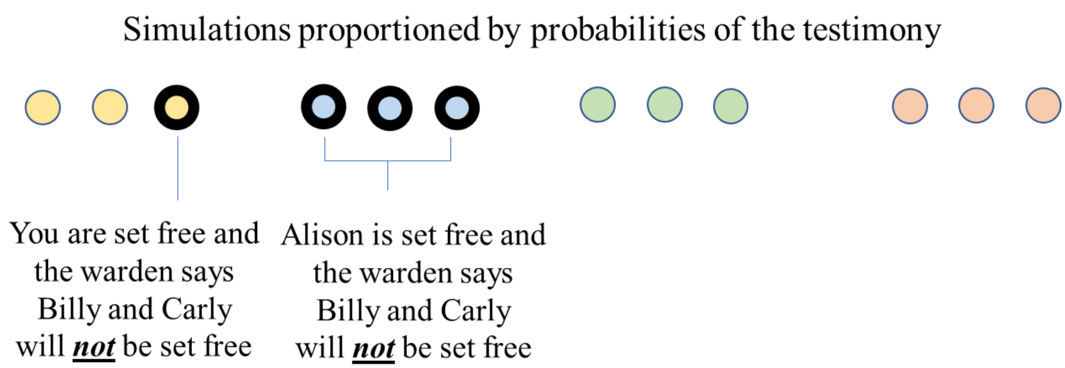
The correct answer is that the warden would give you his testimony in 1/3 or approximately 33% of the simulations where you will be set free.

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:



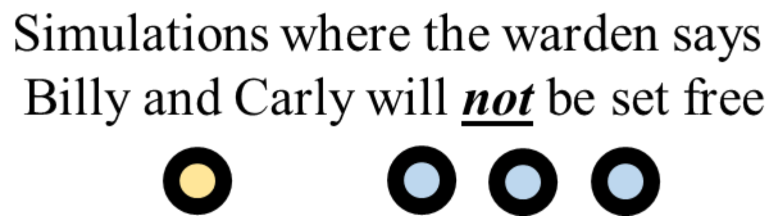
So we have seen the first three steps of the mental simulations approach: first, imagine some simulations; second, proportion the simulations by the prior probabilities; and third, then proportion the simulations by the probability of the evidence.

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.

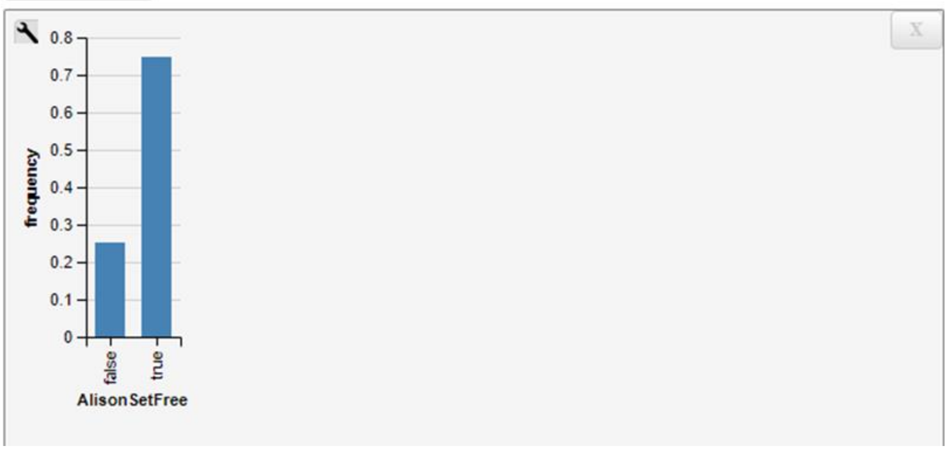
Of course, in real life, the frequency with which something happens does not always match the probability of that thing: a coin may have a 50% probability of landing heads, but if you toss it 10 times, it might land heads on 30% of those times. Nevertheless, we can correctly calculate probabilities if 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:

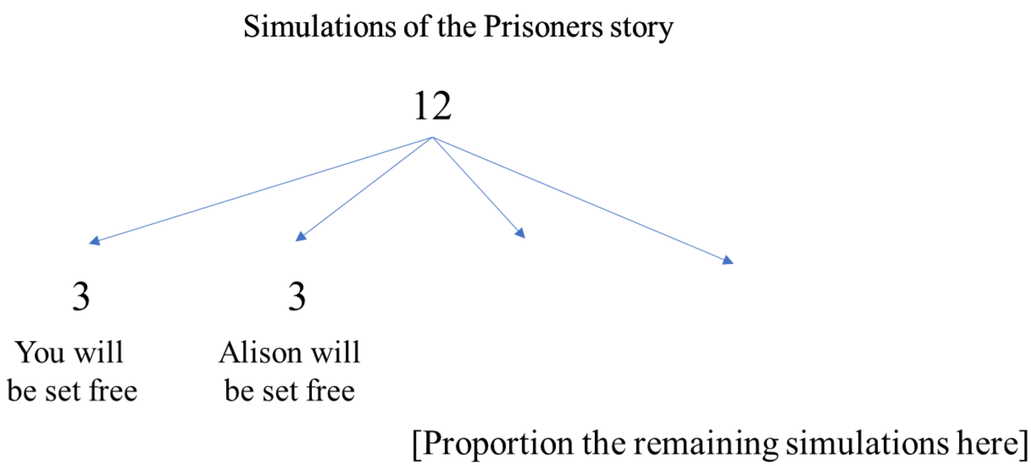
1. Imagine some simulations:
    - Imagine *n* number of simulations (where *n* is any number that can be proportioned by the prior probabilities and then by probability of the evidence)
2. Proportion according to prior probabilities of the outcome:
    - For each possible outcome, make the proportion 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.

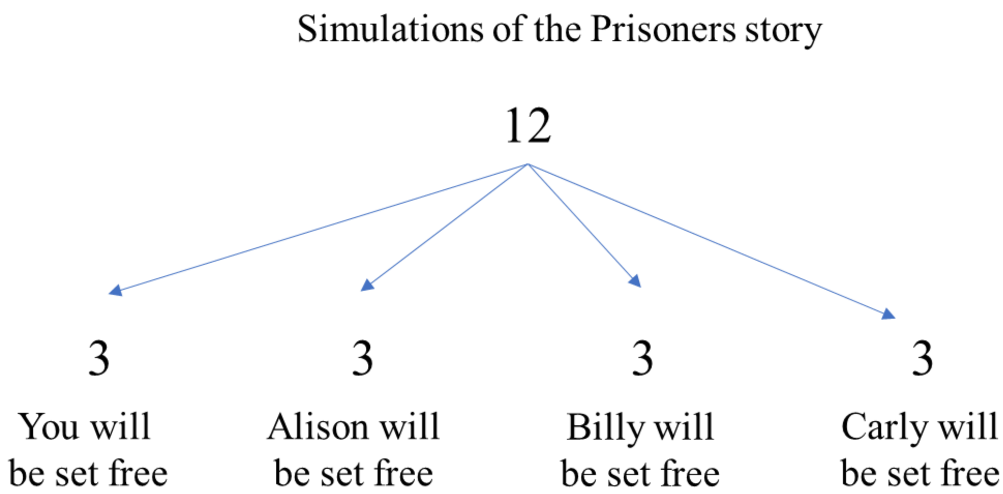
Mental Simulations Using Numbers

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.

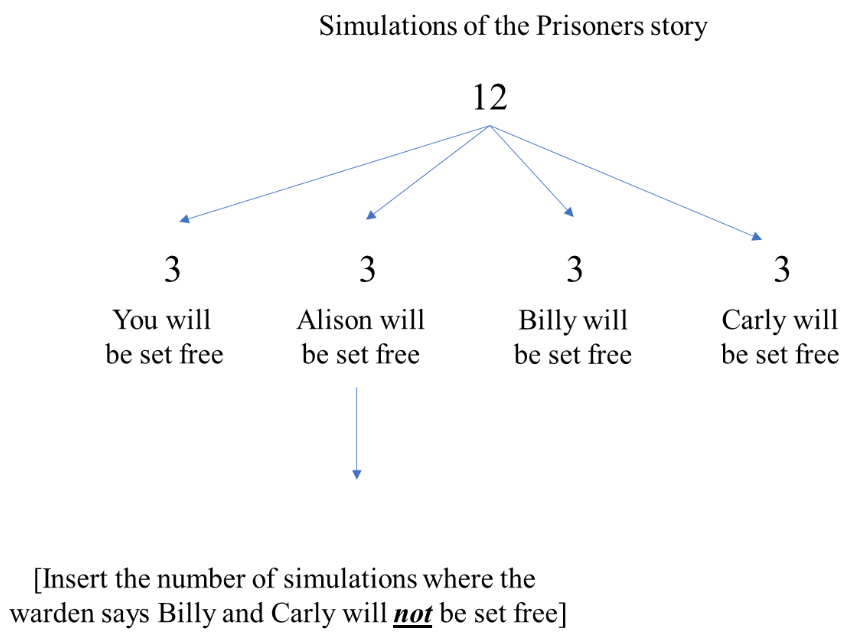
First, proportion the remaining number of simulations where an outcome is true by the prior probability of that outcome. This has already been done for two outcomes, but not for the others.



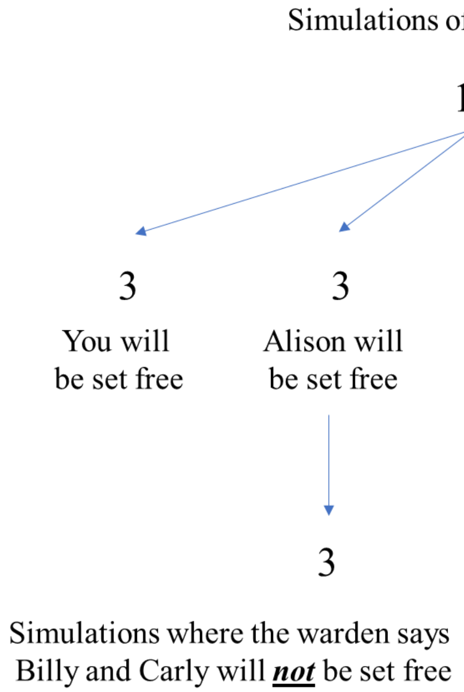
If you did that right, you should have something like what is on the following page.



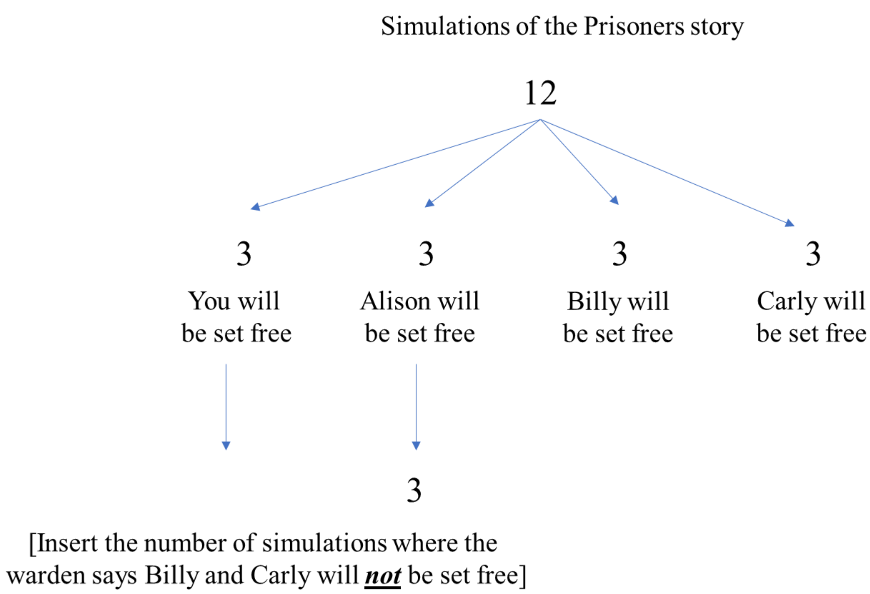
Now proportion the number of outcomes where Alison will be set free by the probability of the warden's testimony that Billy and Carly will ***not*** be free if Alison was to be set free. (Remember, the probability is 100%.)



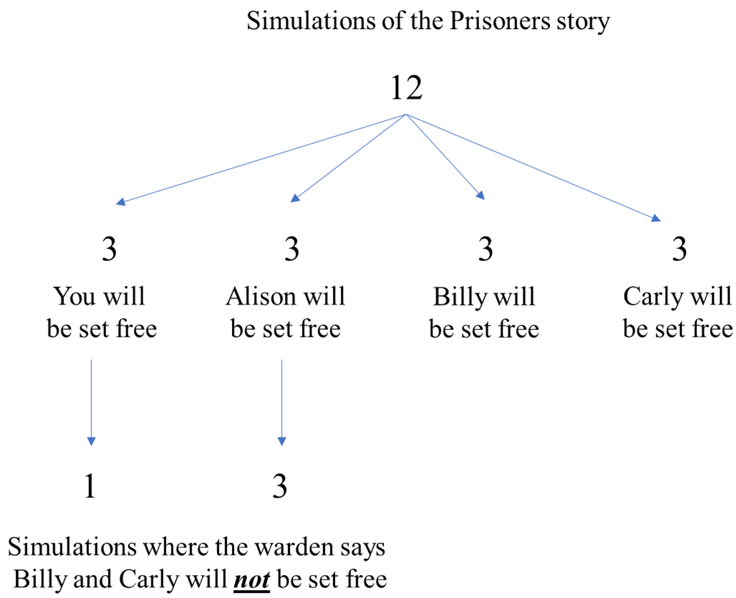
If you did that correctly, you should have something like what follows:



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.)



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:

You will be set free	Alison will be set free
1	3

Simulations where the warden says
Billy and Carly will ***not*** be set free

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 $\{h_1, \dots, h_k\}$ 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 $\{h_1, \dots, h_k\}$. 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 $P(e|h_j)P(h_j) = \frac{N_{h_j\&e}}{N_T}$ 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 $P(h_j|e) = \frac{P(e|h_j)P(h_j)}{P(e)} = \frac{\frac{N_{h_j\&e}}{N_T}}{\frac{N_e}{N_T}} = \frac{N_{h_j\&e}}{N_e}$.

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

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

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

Given these stipulations, it follows that:

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

By algebra then:

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

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

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:

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

It then follows that:

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

Yet we also know by the theorem of total probability that if $\{h_1, ..., h_k\}$ 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.

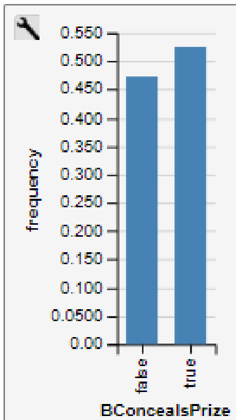
I used a probabilistic programming language called *WebPPL*—freely available at http://webppl.org/.

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.

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.

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

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. 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.

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

Johnson-Laird, P. N. (2012). Inference in Mental Models. In *The Oxford handbook of thinking and reasoning* (pp. 134–154). Oxford 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

Krynski, T. R., & Tenenbaum, J. B. (2007). The Role of Causality in Judgment Under Uncertainty. *Journal of Experimental Psychology: General*, *136*(3), 430–450. https://doi.org/10.1037/0096-3445.136.3.430

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

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge University Press.

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

Prasanta S. Bandyopadhyay. (2011). *Philosophy of Statistics.* North Holland.

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

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