


EMPIRICAL ARTICLE

Likelihood neglect bias and the mental simulations approach: An illustration using the old and new Monty Hall problems

John E. Wilcox 

Columbia University, New York, NY, USA

Email: john.wilcox@columbia.edu

Received: 15 May 2023; **Revised:** 16 February 2024; **Accepted:** 26 February 2024

Keywords: likelihood neglect bias; new Monty Hall problem; mental simulations approach; law of likelihood

Abstract

This article introduces a new violation of a law of probability, *likelihood neglect bias*. The bias occurs when one is (i) aware that some evidence is more likely given one hypothesis rather than another but (ii) that evidence does not cause them to raise their probability for the former hypothesis relative to the latter. The Monty Hall problem illustrates this bias, although the bias is potentially present in other contexts as well. Unlike previous studies, the present study shows that incorrect responses to the problem are attributable to failure to realize the *implications* of likelihoods (i.e., likelihood neglect) instead of unawareness of what the likelihoods are. Likelihood neglect is also distinguished from other biases which it is occasionally confused with, and a theoretical explanation of likelihood neglect is proposed. Experimental results further indicate that a new approach—the mental simulations approach—showed a large effect in reducing likelihood neglect and increasing correct probabilities when encountering the Monty Hall problem. The approach also increased correct probabilities for a new modification called the ‘new Monty Hall problem’, unlike two prominent alternative approaches to the original Monty Hall problem. The new Monty Hall problem also illustrates that humans can fail to recognize objectively strong evidence for a hypothesis if they neglect likelihoods.

1. Introduction

1.1. Likelihoods, likelihood neglect bias, and the Monty Hall problem

We all make inferences on a daily basis—about what decisions to take or about what is probably true, for example. However, we are susceptible to a range of biases that can compromise these inferences (Gilovich et al., 2002).

This article introduces a new bias—likelihood neglect bias—and a new method for correcting it—the mental simulations approach. In the statistical literature, a ‘likelihood’ is the technical term for the probability of some evidence or data given some hypothesis (Bandyopadhyay, 2011; Hacking, 2016; Hawthorne, 2018), an example of which will be given shortly. It is the conditional probability of something we are more certain about (the evidence) given something we are less certain about (the hypothesis)—not the probability of something we are less certain about (the hypothesis) given something we are more certain about (the evidence), even though people often confuse the two (Villejoubert and Mandel, 2002). This article adopts the technical definition of a ‘likelihood’, although the terms ‘likelihood’ and ‘probability’ are often used interchangeably by people who might have less familiarity with the probabilistic literatures. We can then define likelihood neglect bias as a violation of

what is known as the ‘law of likelihood’ (Hacking, 2016; Hawthorne, 2018). According to one version of the law, if a hypothesis h_1 makes some evidence more likely than another hypothesis h_2 , then the evidence *raises* the probability of h_1 relative to h_2 and it *lowers* the probability of h_2 relative to h_1 (proof of the law is in Appendix B of the Supplementary Material). Put simply, the law states that if the evidence is more likely given one hypothesis compared to another, then the evidence raises the probability of the former hypothesis relative to the latter.

Likelihoods play an important role in the infamous Monty Hall problem—a problem which, according to one cognitive scientist, is ‘the most expressive example of *cognitive illusions* or *mental tunnels* in which even the finest and best-trained minds get trapped’ (Piattelli-Palmarini, 1994, p. 161). In the problem, a prize is placed randomly behind one of three closed doors on a gameshow, so each door has an equal prior probability of concealing the prize. (Here, a ‘prior probability’ is the probability *prior* to receiving additional evidence about the location of the prize.) The gameshow contestant selects one of the doors and the gameshow host, called ‘Monty Hall’, opens one of the unselected doors to reveal that there is no prize behind it. If the contestant’s selected door conceals the prize, then there is a 50% likelihood that Monty Hall would open the door he did (since he could have also opened the other door). But if the selected door does not conceal the prize, then there is a 100% likelihood that Monty Hall would open the door he did (since he cannot open any door that conceals the prize or is selected by the contestant). The problem thereby exemplifies the law of likelihood: if the evidence (the door being opened) is more likely given one hypothesis (that the unselected and unopened door conceals the prize) compared to another hypothesis (that the selected door conceals the prize), then the evidence raises the probability of the former hypothesis compared to the latter. Because the prior probabilities were equal, and because Monty Hall’s choice of door is twice as likely if the unselected and unopened door conceals the prize, it is twice as probable that this door conceals the prize (with a probability of 2/3) compared to the selected door (which has a probability of 1/3).¹ This is why participants should switch doors.

Note also that, as mentioned earlier, the likelihood of the evidence given a hypothesis is distinct from the probability of the hypothesis given the evidence. For example, if you select door A, then the likelihood of Monty Hall opening door C given that door B conceals the prize is 100%, but the probability of door B concealing the prize given that Monty Hall opens door C is 2/3. A likelihood should then not be confused with the probability of some hypothesis given the evidence.

Such likelihoods explain why it is more probable that the unselected and unopened door conceals the prize. However, most untrained participants erroneously stick with their initial choice and think the remaining two doors have equal probabilities of concealing the prize (Burns and Wieth, 2004).

Erroneous responses to the Monty Hall can have at least two possible theoretical explanations:

1. *Likelihood unawareness theory*: participants are unaware of what the likelihoods are.
2. *Likelihood neglect theory*: participants are aware of the likelihoods, but simply do not let this affect their other probabilities in the right way.

Let us call the latter error ‘likelihood neglect’, somewhat similarly to base rate neglect. As originally conceived, base rate neglect occurs when people might be *aware* of the base rates, but do not properly ‘integrate’ them into their judgments (although it was hypothesized that this happened because other information was deemed more relevant [Bar-Hillel, 1980, p. 211]). Similarly, likelihood neglect bias occurs when one is *aware* of the likelihoods, but they fail to properly integrate them into their judgments about the relevant probabilities (even if this might happen for reasons which are distinct to base rate neglect bias). More specifically, likelihood neglect is a violation of the law of likelihood: it occurs when someone is (i) aware that the evidence is more likely given one hypothesis compared to another hypothesis, but (ii) this evidence does not cause them to raise their probability for the hypothesis relative to the latter. In the Monty Hall problem, likelihood neglect would occur if participants were

¹This can be demonstrated with Bayes’ theorem: $P(B|c) = \frac{2}{3} = \frac{P(c|B)P(B)}{P(c|A)P(A)+P(c|B)P(B)+P(c|C)P(C)} = \frac{1 \times \frac{1}{3}}{0.5 \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}}$, 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’, c stands for ‘Monty Hall opens door C’ and the likelihoods are $P(c|A)$, $P(c|B)$, and $P(c|C)$.

(i) aware it was more likely that the opened door would be opened if the unselected and unopened door concealed the prize, but they (ii) did not think the unselected and unopened door more probably concealed the prize as a result.

Previous research has not explored or provided evidence for either of these two theoretical explanations. Instead, erroneous responses are explained primarily in other terms, including (1) emotion-based choice biases where participants are averse to switching from their first choice (Granberg and Dorr, 1998), and (2) various other cognitive limitations in understanding and representing probabilities that are involved in why participants think the two hypotheses are equally probable given the evidence (De Neys and Verschueren, 2006; Tubau et al., 2003). A comprehensive description of all competing theories is impossible here, but see Saenen et al. (2018) and Tubau et al. (2015) for some comprehensive reviews of the existing literature. This study investigates the hypothesis that likelihood neglect exists and partially explains erroneous responses to the Monty Hall problem.

More generally, a literature review indicates this bias has not been previously characterized nor demonstrated experimentally in a population. The review included a search through the 1,903 results returned by a PsychInfo search using the keywords ‘likelihood’ and ‘bias’. It also included a search through the first 500 results of a Google Scholar query using ‘bias’, ‘likelihood’, and ‘psychology’ as keywords. Based on the review, the closest contribution in the literature appears to be an article from Fischhoff and Beyth-Marom (1983). There, they discuss four errors involving likelihoods: (1) unawareness of the likelihoods (p. 246), (2) biased and erroneous evaluation of the likelihood ratio (p. 247), (3) failure to assess the likelihood for the alternative hypothesis (p. 247), and (4) the use of incorrect aggregation procedures (p. 248). None of these phenomena are the same as likelihood neglect bias. Likelihood neglect occurs when one is aware of the likelihoods (contra 1), correctly evaluates them (contra 2), includes an assessment of the likelihood given the alternative hypothesis (contra 3), and simply uses no aggregation procedure with the likelihoods (contra 4) because the likelihoods are not seen as relevant. Furthermore, unlike phenomena 1–4 above, and like many other cognitive biases (such as the conjunction fallacy), likelihood neglect bias directly contradicts a law of probability—the law of likelihood. Likewise, likelihood neglect also differs from other phenomena like ‘conservatism bias’ (Phillips and Edwards, 1966) since conservatism bias need not imply an awareness of the likelihoods coupled with a failure to regard them as relevant *at all*.

Consequently, it appears from the aforementioned literature review that this bias is novel since it has not been previously documented in the Monty Hall problem or psychology literatures more generally. Some pilot results initially confirmed the hypothesis that likelihood neglect bias exists and partially explains erroneous responses to the Monty Hall problem, and so a method of reasoning called the ‘mental simulation approach’ was developed to correct the bias.

1.2. The mental simulations approach

The mental simulations approach draws on the research of Gerd Gigerenzer, Ulrich Hoffrage, and colleagues which suggests that humans can sometimes reason better in terms of natural frequency formats (Gigerenzer and Hoffrage, 1995; Hoffrage et al., 2015b). Gigerenzer and Hoffrage (1995) often discuss frequencies that are ‘based on observations’ that occur in a normal or ‘ecological’ setting based on a ‘raw’ count of events (Gigerenzer and Hoffrage, 1999, p. 426). As a result, their interventions typically involve naturally observed frequencies in ecological settings, such as the proportion of people with a disease given some test result (Bramwell et al., 2006; Gigerenzer et al., 2007; Hoffrage et al., 2002, 2015a; Kurzenhäuser and Hoffrage, 2002), the proportion of people matching a DNA test outcome (Hoffrage et al., 2000) or the proportion of blue cabs identified correctly by witnesses and so on (Gigerenzer and Hoffrage, 1995).

However, some reasoning tasks are not as obviously susceptible to frequency formats because the relevant probabilities are not naturally ‘observed’ as ‘raw’ counts in a similarly ‘ecological’ way. For example, when someone plays the Monty Hall problem for the first and only time, they ordinarily do not observe raw counts about which door conceals the prize when another door is opened, at least not in

the same way that medical scientists would sample a population of people to determine the raw counts of a disease or of false positives. There is a big difference between a *one-off* event with probabilities and no explicitly observed frequencies (e.g., the Monty Hall scenario) and a sample of *many* people with naturally observed frequencies (e.g., the medical sampling scenario).

Consequently, to extend Gigerenzer and Hoffrage's insights and convert the Monty Hall problem—which does not necessarily involve naturally observed frequencies—into a similar frequency format, the mental simulations approach then encourages participants to run 'mental simulations' of the Monty Hall problem. This explicit simulation procedure is what builds on Gigerenzer and Hoffrage's (1995) approach and enables it to capitalize on their insights about the advantages of some frequency formats. The approach then encourages participants to calculate the posterior probabilities based on these *mentally simulated* frequencies. (Here, a 'posterior probability'—or 'posterior' for short—refers to a probability *after* the receipt of some evidence—such as the evidence that door C is opened in the Monty Hall problem.) The mental simulations approach thereby aims to help participants calculate the correct posterior probabilities and to also help them to understand *why* those probabilities are correct.

More generally, the approach involves these steps:

1. Generate simulations:
 - Imagine n number of simulations (where n is any positive integer that can be multiplied by the prior probabilities and then the likelihoods such that the result is a positive integer).
 - 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 the selected door A conceals the prize, make it so that Monty Hall opens door C in 50% (or five) 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 simulations 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 posterior 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 probability of concealing the prize.

The mental simulations approach aims to address a variety of reasoning errors. If successfully applied, the approach delivers results that provably conform to Bayes' theorem and thus potentially eliminates likelihood neglect, as well the confusion of the inverse fallacy and neglect of base rates or prior probabilities. (Proof of the conformity is found in Appendix B of the Supplementary Material.) The approach is in principle applicable to any probabilistic problem where some finite number of simulations can be proportioned by the probabilities, not just the Monty Hall problem.

Training in the approach may also help people understand *why* the probabilities are the correct ones. For example, from the simulations approach, we can make sense of why there is a probability of $\frac{2}{3}$ that the other door conceals the prize: because if the probabilistic set up was run many times, then the other door would conceal the prize in two-thirds of those times (at least in the limit, as explained in Appendix

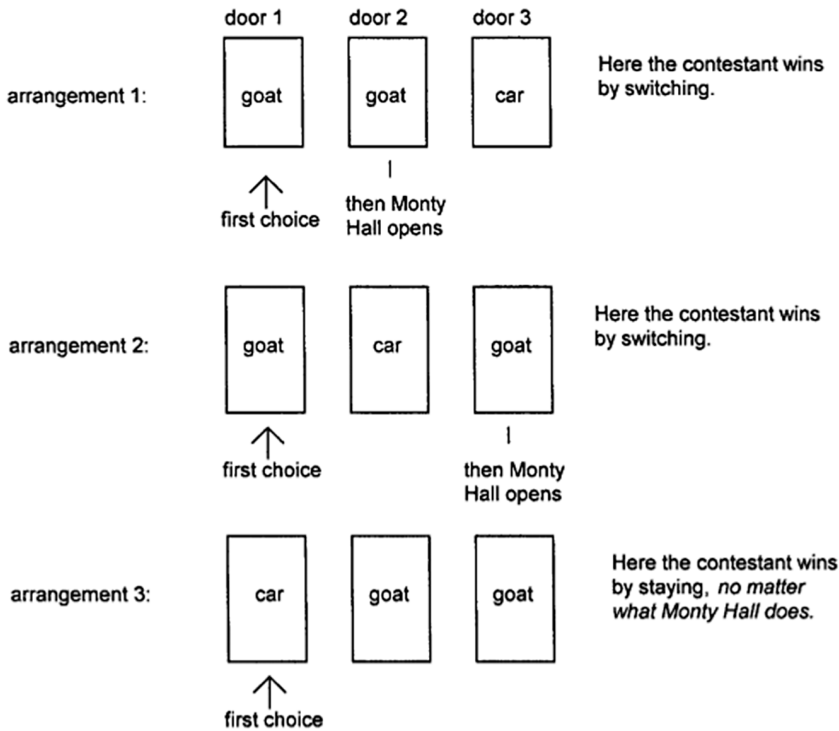


Figure 1. Krauss and Wang's (2003) reproduced explanation of why participants should switch

A's training materials of the Supplementary Material). However, the mental simulations approach is not the same as Johnson-Laird's (2012) prominent mental models approaches because, unlike this approach, two or more simulations may denote the same outcome, not all outcomes are equally probable and the mental simulations approach aims to *improve* reasoning instead of *describing* how it naturally occurs.

The first experiment presented in this article aimed to determine (1) whether untrained participants in the experimental group correctly solve the Monty Hall problem when trained in the mental simulations approach and (2) whether the control group displays likelihood neglect (as opposed to mere likelihood unawareness).

1.3. Two prominent alternative solutions

The second experiment aimed to compare the mental simulations approach to two prominent approaches to solving the Monty Hall problem.

We can call the approach from Krauss and Wang (2003) and Tubau et al. (2003) the 'possible models' approach. Their solution first involves entertaining various 'possibilities' about where the prize might be (or various 'mental models', as they say). Then, one calculates the frequency with which switching doors would win the prize among those possibilities.

Consider Krauss and Wang's (2003) possible models in Figure 1 and Table 1.

In Krauss and Wang's (2003) article, Figure 1 also comes with the caption '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' (p. 5). The idea here is that—somehow—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.

A second prominent approach can be called the 'probability accrual' approach. Tubau et al. (2015) suggest the following:

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

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 Tabossi et al. (1999).

In particular, susceptibility to the illusion [of equal probabilities in the Monty Hall problem] 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 above quote reflects the core of the probability accrual approach: participants would not have equal probability illusions and would realize that the other unopened door has a $2/3$ probability of concealing the prize if they were more aware that (a) the non-selected doors conceal the prize two-thirds of the time and (b–c) once Monty Hall opens one of the unselected doors, that particular door does not conceal the prize. Put simply, once one of the two unselected doors are opened, the probability of $2/3$ then *accrues* to the now one unselected and unopened door. Participants in this article's pilot experiments have also demonstrated this reasoning; for instance, one participant said:

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 $2/3$.

A limitation with the possible models and probability accrual approaches is that they are insensitive to the likelihoods: if the likelihoods change, then so too do the posterior probabilities change, but the approaches would still recommend the same (now incorrect) posterior probabilities. To use an analogy from Russell (1948), these approaches would be like broken unmoving clocks that coincidentally give the correct times when the circumstances are favorable (when the actual time happens to be the same time that is displayed on the broken clock), but these clocks do not track the truth via tracking changes in those circumstances (since they are not sensitive changes in the actual time). That is, the mental models and probability accrual approaches coincidentally give the correct posteriors when the circumstances happen to be favorable (when the prior probabilities are equal and the relevant likelihoods happen to be 50% and 100%), but they do not track the truth via tracking changes in those circumstances (since they are not sensitive to various changes in the priors or likelihoods). Since these approaches do not give the right answer in virtue of tracking the features of the situations which make it the right answer (the priors and likelihoods), these approaches coincidentally give the right answer in the original Monty Hall problem, but for the wrong reasons. So the advantage of the mental simulations approach over these approaches is that it gives the correct answer, not just because the circumstances happen to be favorable, but because—by its very definition—it necessarily tracks those features that make the answer correct and that also generalize to particular other situations.

Of course, one might claim these considerations do not indict either approach because they aim to describe how participants *do* think, not how they *should* think. It is debatable whether humans naturally do think in either of these ways, but at the very least, this article can then be interpreted as (1) a critique in the limitations of how humans would think *if* they reasoned according to these approaches—naturally or not—and (2) an illustration of an alternative approach that overcomes these limitations.

1.4. The new Monty Hall problem

Let us illustrate the limitations of the alternative approaches with what we can 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 10% likelihood 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 10% likelihood or door B with a 90% likelihood. It can be mathematically proven that the posterior probability that door B conceals the prize is now 10/11 or approximately 91%, not 2/3 as in the original Monty Hall problem.² Indeed, the reader can also demonstrate that this is true using computer simulations, as is described in Appendix C of the Supplementary Material.

However, the possible models and probability accrual approaches are not sensitive to the change in likelihoods: therefore, we can predict that if confronted with the new Monty Hall problem where the likelihoods change, many participants following those approaches would give the now incorrect 2/3 probability that they would have given in the original Monty Hall problem. To be fair, proponents of the possible models and probability accrual approaches do not claim to be able to solve the new Monty Hall problem; after all, that problem is introduced for the first time in this article.³ Rather, the point is just to illustrate how these approaches are insensitive to likelihoods and how, for that reason, these approaches deliver the right answer for the wrong reasons in the original Monty Hall problem.

Experiment 2 aims to show this is the case by asking participants to specify their posterior probabilities in the new Monty Hall problem. These participants were randomly assigned to four groups. The control group encountered only the new Monty Hall problem. The three other groups also encountered the *new* Monty Hall problem, but they *first* encountered the *original* Monty Hall problem *and* an explanation of it in terms of the respective approaches: the mental simulations approach, the possible models approach and the probability accrual approach.

1.5. Hypotheses under test

The following hypotheses were proposed and preregistered on the Open Science Framework (OSF) prior to the data collection:

1. Even participants in the control conditions who are aware of the likelihoods will nevertheless judge the posterior probabilities to be equal, thereby displaying likelihood neglect.
2. Participants in the mental simulations experimental condition in Experiment 1 are more likely to give the correct posteriors and less likely to commit likelihood neglect.
3. Participants in the mental simulations condition are more likely than those in any other condition to choose the correct posteriors for the new Monty Hall problem.
4. No participants will choose the correct posteriors in the new Monty Hall problem when exposed to the possible models or probability accrual approaches.
5. Participants in the possible models and probability accrual conditions of Experiment 2 will be more likely to incorrectly state that door A and door B have 1/3 and 2/3 posterior probabilities of concealing the prize.

The first hypothesis concerns two competing theoretical explanations of incorrect responses of the Monty Hall problem: (i) that incorrect responses are attributable to *unawareness* of the likelihoods and (ii) that incorrect responses are attributable to *neglect* of the likelihoods. The second hypothesis is a causal hypothesis: that the mental simulations training causes an improvement in correct responses and a reduction of likelihood neglect. The third and fourth hypotheses amount to a comparative causal

²Similarly to footnote 1, this can be proved using Bayes’ theorem: $P(B|c) = \frac{1 \times \frac{1}{3}}{\frac{1}{10} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{10}{11}$.

³While, so far as the author can tell, the new Monty Hall problem as presented here is novel, other scholars have also noted that the likelihoods can be adjusted to give different versions of the Monty Hall problem. See, for example, the ‘Monty Small problem’ variant in Rosenthal (2008).

hypothesis: that exposure to the mental simulations training will result in correct answers to the new Monty Hall problem while exposure to the two alternative prominent solutions will not. This fifth hypothesis is consistent with the possibility that the possible model and probability accrual approaches desensitize participants to changes in likelihoods. Experiment 1 tests the first two hypotheses, and Experiment 2 tests primarily the latter three.

The OSF preregistered hypotheses, methods, materials, and analysis procedures are accessible at the following link: https://osf.io/kdme9/?view_only=516d7baaf57b452b9fe958c83365ac6c.

1.6. Significance of the proposed research

This article focuses on the Monty Hall problem for two reasons. First, the Monty Hall problem has its own sizable literature, and this article aims to make the already substantial contribution of re-orienting this literature so as to focus on likelihood neglect, its causes and its solutions. Second, the Monty Hall problem provides a clear illustration of likelihood neglect and how the mental simulations problem can address it.

But it is a mistake to think the significance of this article is confined solely to the Monty Hall problem. If the hypotheses are confirmed, this would have more general significance 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 relevant, such as medical diagnosis, legal criminal inquiries, scientific contexts, intelligence analysis, and others. That said, this article is committed to the claim that the bias is present only in *some* contexts, such as the Monty Hall problem, but not in *all* contexts, such as those where base rate neglect occurs. (See the ‘Discussion’ section for more details about how likelihood neglect bias is reconcilable with phenomena like base rate neglect.) 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. In particular, if humans display likelihood neglect, then this is another departure from Bayesian norms of cognition, providing further evidence that humans think in (at least partially) non-Bayesian ways.

That said, this article does not provide evidence for the existence of likelihood neglect and for the efficacy of mental simulations beyond the Monty Hall problem. This is an exciting topic for future work, ideally including work from other researchers who seek to replicate these phenomena.

2. Experiment 1

2.1. Method

2.1.1. Purpose

Experiment 1 aimed to determine whether:

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

2.1.2. Participants

Participants were recruited from Amazon’s Mechanical Turk (MTurk). All participants were above 18, reported English as a first language, were based in the United States and had a ‘HIT Approval Rate (%) for all Requesters’ HITs’ that is between 80% and 100%. Participants were first given a screening survey to determine their eligibility for the experiment. They were asked questions that collected demographic information, information about their occupational and educational backgrounds, and information about their familiarity with the Monty Hall problem. All prescreening participants were 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 were also asked whether they are familiar with other things—such as what Sudoku, Stanley Milgram’s

experiments or the Central Limit Theorem are. These other questions were there so participants could not tell which question would enable them to qualify for the follow up study. For Experiment 1, a total of 952 participants completed the eligibility survey, 283 (30%) of which indicated no prior familiarity with the Monty Hall problem and were invited to the follow up survey.

An a priori power analysis and previous piloting informed sample size selection. A Fisher exact test power analysis was calculated using G*Power, with $\alpha = 0.05$, $(1 - \beta) = 0.9$, a 1-to-1 allocation ratio and assumed proportions of correct posteriors for the experimental and control group of 0.3 and 0.05, respectively. This 90% power analysis' results indicated 42 participants were needed per condition to detect an effect. Consequently, participants were randomly assigned to the experimental and control conditions until each group was comprised of 50 valid responses.

2.1.3. Materials

Participants in both conditions completed an online survey. They first provided consent and were informed that the \$9 base payment required completing the study and the \$6 bonus payment required correct answers to certain questions.

The bonus payment was important. Previous piloting indicated 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 correctly specifying the likelihoods (the probability of door C being opened given that door A concealed the prize or given that door B concealed the prize). Thus, participants in both conditions were able to obtain the bonus (but again, participants were not aware of this).

Both groups were presented with the same version of the Monty Hall problem and then asked the same questions about it. They also answered basic comprehension questions: 'How many doors are in the above scenario?' (Correct answer: '3'), 'If you select door A, and door B conceals the prize, then Monty Hall will open:' (Correct answer: 'door C') and 'If you select door A, and door A conceals the prize, then Monty Hall will open:' (Correct answer: 'either door B or door C'). Valid responses were those which correctly answered all the comprehension questions and indicated no prior familiarity with the Monty Hall problem. Table 2 lists the questions that measured each central construct, as well as any additional coding or classification for the responses.

Experimental condition training module. The experimental condition differed from the control group in the following two ways: (1) participants in this condition were first taken through a training module in the mental simulations approach, and (2) they were asked if they used the mental simulations approach to answer the Monty Hall problem (see the final question in Table 2).

In the training module, participants viewed either or both of two kinds of materials: reading materials or a series of videos (video materials). Both kinds of materials conveyed similar information about the mental simulations approach and how to use it, and both are available at the OSF website (specified in sub-section 1.5.). The video materials are similar to the reading materials in Appendix A of the Supplementary Material, since the presenter read directly from the materials with very few departures, other than 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, the story was modified 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 randomly 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 (or hypotheses) all have equal probabilities in the beginning. Participants then get the

Table 2. Question and coding table for measured constructs.

Question(s):	Response coding:
<p>Likelihood awareness (open-ended, text entry)</p> <p>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?</p> <p>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?</p>	<p>Displays ‘Likelihood awareness’ if their answers are equal to 50% and 100% for the respective questions. Coded by algorithm (not directly by experimenter).</p>
<p>Correct posteriors (open-ended, text entry)</p> <p>What is the probability that door A conceals the prize after Monty Hall shows you the goat?</p> <p>What is the probability that door B conceals the prize after Monty Hall shows you the goat?</p>	<p>Gives ‘Correct posteriors’ if their answers are (strictly or approximately) equal to 1/3 and 2/3 for these questions, respectively.⁴ Coded by algorithm.</p>
<p>Likelihood neglect bias</p> <p>(Derived from the responses to the above 4 questions)</p>	<p>Participants demonstrate ‘Likelihood neglect bias’ if they display likelihood awareness but did not think that B more probably concealed the prize than A. Likelihood awareness was coded by algorithms, but an experimenter coded whether B more probably conceal the prize than A. (A blinded, independent rater agreed with the experimenter’s coding on all cases except two, but this did not significantly affect the results.)</p>
<p>Answer confirmation (closed-ended, multi-choice)</p> <p>To confirm, which of the following do you agree with? [Response options are listed in ‘Likelihood Citation’ question immediately below.]</p>	<p>No additional response coding.</p>
<p>Likelihood citation (open-ended, text entry)</p> <p>You indicated above that you think [Door A has a higher probability of concealing the prize than door B/Door B has a higher probability of concealing the prize than door A/Door A and door B have an equal probability of concealing the prize]. What part of the earlier description of the game show makes you think this? (You may copy and paste parts of that description if you wish, but do not feel obligated to do so.)</p>	<p>Participants were coded as ‘Citing a Likelihood’ if they mentioned one of the two likelihoods for the evidence given door A or door B concealing the prize, or if they copied parts of the description mentioning such a likelihood. Coded by experimenter blinded to conditions.</p>

Continued

⁴Percentage or decimal answers rounded in either direction were also accepted. For example, ‘33%’ or ‘34%’ were accepted as correct, but an answer like ‘32%’ was not.

Table 2. (Continued).

Switching vs. sticking (closed-ended, two choices)	
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 choose?	‘Switchers’ if they opted to switch. ‘Stickers’ otherwise.
Self-reported rationale (open-ended, text entry)	
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 (closed-ended, multi-choice)	
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 (closed-ended, multi-choice)	
How well do you think you understand why your answers for those questions were the correct ones?	No additional response coding.
[Experimental condition only] Use of mental simulations (closed-ended, two choices)	
Did you use the ‘mental simulations’ approach to arrive at your answer?	No additional response coding.
Please answer honestly. There is no penalty for giving one answer or another.	

evidence which rules out at least one of the outcomes. Untrained 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 hypothesis 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. The problems have different numbers of possible outcomes or hypotheses 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 from 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 not 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 were also asked questions to test their understanding of the mental simulations approach.

They were then presented with the Monty Hall problem, presumably for the first time (since they were prescreened for prior familiarity). They then answered the above questions about the problem,

as well as the aforementioned basic comprehension questions to screen out inattentive responses (e.g., the question about how many doors were in the Monty Hall problem). Experiment 1 departed from the preregistered method in one way. During the experiment, it was noticed many participants seemed to rush the survey and provide inattentive responses that were invalid (i.e., failing to answer basic comprehension questions). Consequently, in an effort to reduce inattentive responses, the consent statement was updated mid-way through the experiment to say ‘The **\$9 base payment** requires **honest and attentive** completion of the study (payment may be **refused** in cases of dishonesty or inattentive completion)’ [bolding original]. Otherwise, there were no departures in method.

2.1.4. Procedure

To summarize the procedure implicitly described above, participants were first prescreened for prior familiarity with the Monty Hall problem. Eligible participants were then invited to participate in a follow-up Qualtrics survey where they provided informed consent. Participants were then randomly assigned to either the experimental condition or the control condition. At that point in the survey, participants in the experimental condition were taken through a training module in the mental simulations approach, using the story of the prisoners—but with significant structural differences—to illustrate the approach, and they also answered questions about the approach (with automatic feedback provided). Then, participants in both conditions were presented with a description of the Monty Hall problem. Each group was asked the same questions about the problem, including those in this order: (1) some basic comprehension questions as attention checks, (2) the questions in Table 2 (in that order), and (3) a question to confirm their prior lack of familiarity with the problem. Prior to the familiarity question, participants in the experimental condition were also asked if they used the mental simulations approach.

2.2. Results

A total of 196 MTurk participants completed the survey, while 36 started but did not finish (96 of 124 completed the experimental condition compared to 100 of 108 in the control condition). A total of 102 passed the attention checks and indicated that they had no prior familiarity with the Monty Hall problem, so they were included as valid participants (although each condition was comprised of only the first 50 participants who had valid responses, meaning two valid responses were omitted from the dataset).

Table 3 and Figure 2 summarize the main differences between the mental simulations experimental group and control group, as measured using Cohen’s h (Cohen, 1988) and the difference in proportions (PD).

The results support the first two hypotheses. Hypothesis 1 predicted that even participants in the control condition who are aware of the likelihoods will nevertheless judge the posterior probabilities to be equal, thereby displaying likelihood neglect bias. Fulfilling this prediction, 40 out of 50 (80%) participants in the control condition did indeed judge the posteriors to be equal even though they were aware of the likelihoods, thereby displaying likelihood neglect bias. Hypothesis 2 predicted that participants in the mental simulations experimental condition would be more likely to give the correct posteriors and less likely to commit likelihood neglect. This was indeed the case since 16 participants (32%) in the experimental condition gave the correct posteriors whereas none in the control group did (Cohen’s $h = 1.2$, 95%CI [0.81, 1.59]) and 40 participants (80%) displayed likelihood neglect in the control group whereas only 13 (26%) in the experimental group did (Cohen’s $h = -1.14$, 95%CI [-0.1.54, -0.75]). These qualify as large effect sizes according to Cohen’s (1988) classification.

Additionally, another five participants in the experimental group—but none in the control group—qualify as giving the correct posteriors if more lenient coding criteria are used whereby participants’ posteriors are coded as correct if they specify probabilities of 30% and 70% for doors A and B, respectively ($n = 3$), or if they specify probabilities of 66% and 33% for doors A and B, respectively, meaning they potentially got the right answers but confused which fields to put them in for the survey

Table 3. Experiment 1: Comparison of mental simulations experimental group and control group.

	Mental simulations (n = 50)	Control (n = 50)	Differences between groups		
	Percentage (count)	Percentage (count)	PD (95% CI)	Cohen's <i>h</i> (95% CI)	<i>p</i> -Value
Likelihood awareness	64% (32)	84% (42)	-20% [-39%, -01%]	-0.08 [-0.48, 0.31]	0.04
Likelihood neglect	26% (13)	80% (40)	-54% [-72%, -36%]	-1.14 [-0.154, -0.75]	< 0.001
Correct posteriors	32% (16)	0% (0)	32% [17%, 47%]	1.2 [0.81, 1.59]	< 0.001
Switch	52% (26)	8% (4)	44% [26%, 62%]	1.04 [0.64, 1.43]	< 0.001
	Mean (<i>SD</i>)	Mean (<i>SD</i>)	<i>DM</i> (95% CI)	Cohen's <i>d</i> (95% CI)	<i>p</i> -Value
Confidence ^a	2.56 (0.91)	3.16 (1.17)	-0.6 [-1.01, -0.18]	-0.68 [-1.09, -0.26]	0.005
Understanding ^b	2.42 (0.97)	3.18 (1.24)	-0.76 [-1.2, -0.32]	-0.57 [-0.97, -0.15]	< 0.001

^aResponses ranged from 1 = 'Not at all confident', 2 = 'Slightly confident', 3 = 'Moderately confident', 4 = 'Very confident', and 5 = 'Extremely confident'.

^bResponses ranged from 1 = 'Not at all well', 2 = 'Slightly well', 3 = 'Moderately well', 4 = 'Very well', and 5 = 'Extremely well'.

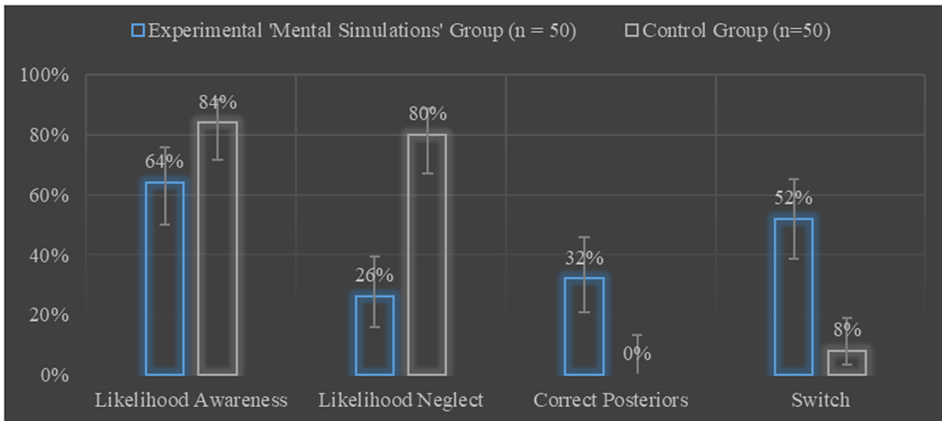


Figure 2. Comparison of experimental vs. control outcomes in Experiment 1.

(*n* = 2). No other more lenient and relevant criteria affected the outcomes. If these lenient criteria were adopted, this would provide even stronger support for the second hypothesis that the mental simulations approach made correct posteriors more likely. However, these lenient criteria were not employed so that the experiment provided more conservative and stringent evidence in favor of the hypothesis instead of more favorable analyses that were adjusted post-hoc.

That said, perhaps surprisingly, participants in the experimental group were also less likely to display awareness of the likelihoods (*PD* = -20%, 95%CI [-39%, -01%]). This may be because respondent fatigue from the additional training module could make participants less likely to correctly specify the likelihoods. Since they showed less likelihood awareness, they also show less likelihood neglect, but only *partly* for this reason (since likelihood neglect requires likelihood awareness). Therefore, the reduction in likelihood awareness could be removed from the overall reduction of likelihood neglect to thereby isolate the effect of the intervention in reducing likelihood neglect among those *who are still*

aware of the likelihoods; measured as such, the effect of the experimental intervention was Cohen's $h = -0.70$, 95%CI [-1.09, -0.31] and $PD = -34\%$, 95%CI [-54%, -14%].

Other differences between the groups were observed. Participants in the experimental group were more likely to opt to switch doors ($PD = 44\%$, 95% CI [26%, 62%]) and on average had lower ratings for their self-reported confidence ($DM = -0.6$, 95% CI [-1.01, -0.18]) and understanding ($DM = -0.76$, 95%CI [-1.2, -0.32]). Control group participants then had higher confidence in their posterior answers and higher self-reported understanding of their correctness. However, this confidence and understanding was mistaken since their answers were, unknown to them, factually incorrect.

An exploratory correlation analysis also examined relationships between responses and participant features, such as age, sex, and number of science courses taken. This analysis failed to reveal any statistically significant variables that predict correct posteriors in the experimental group (smallest $p = 0.14$). This, however, may be due to a small sample size with relatively few respondents (16 of 50) who correctly specified the posteriors in the experimental group. In any case, the most predictive variables were participants' level of science education ($r = -0.21$, $p = 0.14$), whether they had played Sudoku ($r = 0.19$, $p = 0.18$) and their familiarity with Stanley Milgram's experiments ($r = -0.18$, $p = 0.21$), although these results should obviously be interpreted with caution.

Thirty-eight of 50 (76%) participants in the experimental condition reported using the mental simulations approach. Some of the participants' rationales also indicate that they used the mental simulations approach and that this approach helped them produce correct answers. Consider the following example:

Participant 199229318: 'I imagined six simulations, two for each door. Then I figured out how many of those two would likely result in the evidence (that C was not the prize). For A, this was 1/2. For B it was 2/2 (because A could not be revealed), and for C it was 0/2. So eliminating the simulations that did not support the evidence, we get 1 case of A being the prize and 2 cases of B being the prize. So it looks like B is more likely, with a 2/3 probability'.

All the rationales with correct posteriors are reproduced in Appendix E of the Supplementary Material. Readers are encouraged to peruse the rationales in the Supplementary Material and make up their own minds about what proportion display evidence of using the approach.

Furthermore, citations were coded regarding whether they cited a likelihood. If participants mentioned the likelihood of the evidence given door A concealing the prize or given door B concealing the prize, or if they copied the text from the problem description about either likelihood, then they were coded as citing a likelihood. To avoid bias, the Random IDs and citations were exported to an excel sheet and coded without awareness of any other identifying information of the responses, including which experimental group they were assigned to. According to the coding, 19 participants (38%) cited likelihoods in the experimental group and 17 participants (34%) cited likelihoods in the control group, but this difference failed to reach levels of statistical significance ($p = 0.835$).

Thus, there was no significant difference between the groups with respect to their citation of likelihoods. This was surprising, since one might have expected participants in the experimental group to be more sensitive to likelihoods and thus more likely to explicitly cite them. It is not clear what explains this lack of difference. One possibility is that participants in the control group may have mentioned the likelihoods, or pasted parts of the text mentioning likelihoods, merely because they did not know how else to answer the question: 'What part of the earlier description of the game show makes you think [door A and B are equally likely to conceal the prize]? (You may copy and paste parts of that description if you wish, but do not feel obligated to do so.)' It is especially unclear why control group participants would cite likelihoods since the likelihoods do not by themselves seem to provide any good reason to think each door is equally likely to conceal the prize once door C has been opened. Thus, these specific results may be confounded since participants in the control group might have cited likelihoods merely because they did not know what else to cite from the description. Future research could further investigate and test competing explanations.

Table 4. Self-reported confidence and understanding among participants with correct posteriors who reported using mental simulations ($n = 13$).

Question: How confident are you that your answers were the correct ones?				
Not at all confident	Slightly confident	Moderately confident	Very confident	Extremely confident
2 (15%)	4 (31%)	3 (23%)	4 (31%)	0 (0%)
Questions: How well did you understand why your answers were the correct ones?				
Not at all well	Slightly well	Moderately well	Very well	Extremely well
3 (23%)	1 (8%)	7 (52%)	2 (15%)	0 (0%)

Table 4 also depicts the distribution of self-reported ratings of confidence and understanding among those who with correct posteriors who reported using the mental simulations approach.

When comparing those in the experimental group who reported using the mental simulations approach and those who did not, no statistically significant differences were detected with respect to likelihood awareness, likelihood neglect, correct posteriors, switching behavior, self-reported confidence, and self-reported understanding. Interestingly, three participants in the experimental group (6%) correctly reported the posterior probabilities but reported *not* using the mental simulations approach. While this number is very small, it suggests these participants might have used online information to correctly answer the Monty Hall problem instead of the mental simulations approach. Their rationales have been reproduced in Appendix E of the Supplementary Material for the reader to examine if they wish. If these three participants did not use the mental simulations approach, then the true effect of the mental simulations approach in increasing correct posteriors may be better represented by omitting them with Cohen’s $h = 1.07$, 95% CI [0.68, 1.46] and $PD = 26\%$, 95% CI [12%, 40%]. In either case, the effect size is still large according to Cohen’s (1988) classification.

Experiment 1 made multiple comparisons, but only two comparisons were from the preregistered confirmatory hypotheses, while the rest were additional relationships in exploratory analysis. These two comparisons are from Hypothesis 2: namely, participants in the mental simulations experimental condition in Experiment 1 are more likely to give the correct posteriors and less likely to commit likelihood neglect. A Bonferroni correction of $\frac{\alpha}{2} = 0.025$ was applied to these 2 comparisons, both of which were still statistically significant after the correction: the participants in the mental simulations conditions were still more likely to give correct posteriors (Cohen’s $h = 1.20$, 97.5% CI [0.75, 1.65], $p = 0.00004284$) and less likely to commit likelihood neglect (Cohen’s $h = 1.14$, 97.5% CI [-1.59, -0.70], $p = 0.0000001895$). In conclusion, the support for Hypothesis 2 is not a statistical artifact.

3. Experiment 2

3.1. Method

While Experiment 1 aimed to determine the effect of the mental simulations approach in the *original* Monty Hall problem, Experiment 2 aimed to show how *several alternative approaches* compare with the mental simulations approach in handling the *new* Monty Hall problem (see a description of the problem in Section 1.4).

Participants were recruited from Amazon’s MTurk. All participants were above 18, had reported English as a first language, were based in the United States, had a ‘HIT Approval Rate (%) for all Requesters’ HITs’ that is between 80% and 100% and must not have participated in Experiment 1. Participants were first given a screening survey to determine their eligibility for the experiment as per

the procedure in Experiment 1. Participants were again offered a base payment of \$9 for completing the study and then another bonus payment of \$6 for giving correct answers.

An a priori power analysis and previous piloting informed sample size selection. A Fisher exact test power analysis was calculated using G*Power, with $\alpha = 0.05$, $(1 - \beta) = 0.9$, a 1-to-1 allocation ratio and assumed proportions of correct posteriors for the mental simulations and other groups of 0.3 and 0, respectively. The 90% power analysis results indicated that 25 participants were needed per condition to detect an effect. Consequently, participants were randomly assigned to the conditions until each group was comprised of 25 valid responses.

For Experiment 2, a total of 1,009 participants completed the eligibility survey, 313 (31%) of which indicated no prior familiarity with the Monty Hall problem and were invited to the follow up survey where they first provided their informed consent. In one of the four conditions, participants were just presented with the new Monty Hall problem where door B conceals a prize with a probability of 10/11. They then answered some questions. However, in the other three conditions, participants were first presented with the original Monty Hall problem where the probability that door B conceals the prize is 2/3. They were then 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)

These participants were given one of three explanations depending on which condition they were assigned to: an explanation in terms of the mental simulations approach, an explanation in terms of the possible models solution or an explanation in terms of the probability accrual solution. These explanations can be viewed in the experimental materials at the aforementioned OSF link.

Participants were then asked questions about the respective explanations to test their understanding of the explanations. In this case, incorrect answers were rejected automatically by the survey, and participants could not continue the survey until they provided correct answers. They were then given the new Monty Hall problem, described similarly to how it was described in section 1.4 of this article.

All participants were reminded that their bonus payment depended on correct answers, and they were then asked the same questions to assess likelihood awareness and their posterior probabilities. However, for each of these questions, the four to five possible answers were in closed-ended, multi-choice format and presented in a randomly generated order each time. The multi-choice format facilitates algorithmic coding, reduces the possibility of erroneous or biased experimenter coding of responses and also gives participants in each condition the opportunity to see what the right answers may be, especially because the calculations are potentially more difficult in the new Monty Hall problem compared to the original.

Participants were then asked: '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'. Participants in the three explanation conditions were also asked whether the earlier explanation of the original Monty Hall problem affected their answers to the new Monty Hall problem.

They were also asked whether, after seeing the explanation of the original Monty Hall problem, they agreed that door B probably concealed the prize. This is a measure of how convincing the explanation of the original Monty Hall problem is. As with Experiment 1, they were lastly asked to report their confidence in their answers, their self-reported understanding and their prior familiarity with the Monty Hall problem. Valid responses were those which correctly answered all the comprehension questions (as in Experiment 1) and indicated no prior familiarity with the Monty Hall problem.

Like Experiment 1, Experiment 2 also departed from the preregistered method in some respects. In an effort to reduce inattentive responses, the consent statement for Experiment 2 was updated before the experiment to say ‘The **\$9 base payment** requires **honest and attentive** completion of the study (payment may be **refused** in cases of dishonesty or inattentive completion)’ [bolding original]. The coding of all variables was performed similarly to Experiment 1, with one exception. The preregistered analysis plan coded participants as displaying likelihood neglect only if they correctly specified the likelihoods. After conducting this experiment, however, it was realized that, as per the definition in the introduction, likelihood neglect occurs if the likelihoods are unequal in a particular way, regardless of whether they are correct or not. Consequently, Experiment 1 stuck to the preregistered plan while Experiment 2 adhered to the more technically correct coding of likelihood neglect (since this was computationally easier to code than for Experiment 1). However, this coding of the results does not significantly affect the main conclusions of this article; other researchers can confirm this by investigating the dataset and analysis script, or by conducting their own independent replications. Otherwise, there were no departures in the method or analysis plan.

3.2. Results

A total of 217 MTurk participants completed the survey for Experiment 2, while 29 started but did not finish. A total of 102 of these responses were valid, while 2 were omitted from the analysis (as per the sample selection plan of only including the first 25 valid responses in each condition).

Figure 3 and Table 5 summarize significant differences between the conditions. The results support hypotheses 3 and 5. Hypothesis 3 predicted that participants are more likely than any other condition to choose the correct posteriors for the new Monty Hall problem when exposed to the mental simulations approach. The results support this hypothesis since participants in the mental simulations were significantly more likely to give the correct posteriors compared to any other condition (Cohen’s $h = 0.88$, 95% CI [0.33, 1.44]). Unlike Experiment 1, no other more relevant and lenient coding of responses could have affected the support for this hypothesis (since the closed-ended response options precluded this). Hypothesis 5 predicted that participants in the possible models and probability accrual conditions would be more likely to incorrectly state that door A and door B have 1/3 and 2/3 posterior probabilities of concealing the prize. The results again support this prediction, since participants were more likely to give incorrect 1/3 and 2/3 posteriors for these conditions, and this was statistically significant compared to the control group (Cohen’s $h = 1.53$, 95% CI [0.98, 2.09]) but not the mental simulations group (Cohen’s $h = 0.51$, 95% CI [-0.05, 1.06]). Participants were less likely to display likelihood neglect in these conditions only because they were more likely to give these erroneous posteriors, thereby replacing one error with another.

Hypothesis 4 was falsified, however. It predicted that no participants would choose the correct posteriors in the new Monty Hall problem when exposed to the possible models or probability accrual approaches. Despite this, one participant in each condition did select the correct posteriors. That said, the number of correct answers in the possible models or probability accrual conditions was not significantly different from the control group since one participant also selected the correct posteriors in the control condition too. Ultimately, then, these two conditions did not perform better than the control condition which featured no explanation of the original Monty Hall problem. Again, the rationales indicate both that some participants used the mental simulations approach in the mental simulations condition and that this approach helped them produce correct answers, as can be seen in Appendix E of the Supplementary Material.

Results also indicated that the mental simulations approach was the most convincing of any explanation for the original Monty Hall problem. As Table 5 shows, 80% of participants in the mental simulations condition reported thinking that door B more probably concealed the prize than door A, compared to 60% for the possible models approach ($PD = 20\%$, 95% CI [-8%, 49%], $p = 0.22$) and 52% for the probability accrual approach ($PD = 28\%$, 95% CI [-1%, 57%], $p = 0.07$). That said, these differences failed to reach statistical significance at the $\alpha = 0.05$ level.

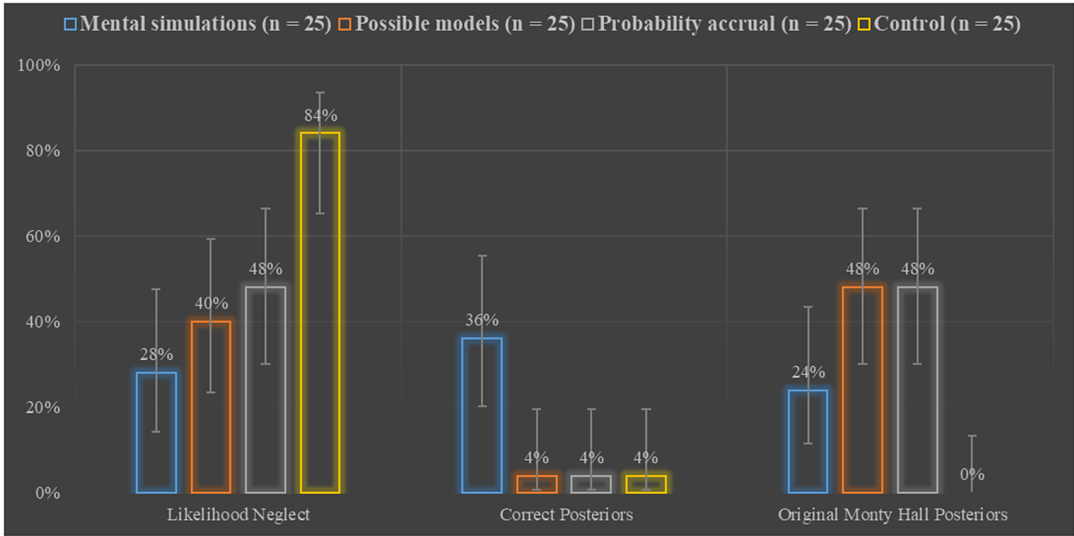


Figure 3. New Monty Hall problem—Differences for measured constructs.

Table 5. New Monty Hall problem—Differences for measured constructs.

	Experimental condition			
	Mental simulations (n = 25)	Possible models (n = 25)	Probability accrual (n = 25)	Control (n = 25)
	Percentage per condition (count)			
Likelihood awareness	68% (17)	64% (16)	52% (13)	72% (18)
Likelihood neglect	28% (7)***	40% (10)**	48% (12)*	84% (21)
Correct posteriors	36% (9)*	4% (1)	4% (1)	4% (1)
Convincingness in original problem	80% (20)	60% (15)	52% (13)	NA
Original Monty Hall problem posteriors	24% (6)*	48% (12)***	48% (12)***	0% (0)
	Mean (SD) per condition			
Confidence ^a	2.92 (1.04)	3.08 (1.15)	2.88 (1.2)	3.12 (1.3)
Understanding ^b	2.8 (.87)	2.76 (1.13)	2.96 (1.21)	3.04 (1.17)

^aResponses ranged from 1 = ‘Not at all confident’, 2 = ‘Slightly confident’, 3 = ‘Moderately confident’, 4 = ‘Very confident’, and 5 = ‘Extremely confident’.

^bResponses ranged from 1 = ‘Not at all well’, 2 = ‘Slightly well’, 3 = ‘Moderately well’, 4 = ‘Very well’, and 5 = ‘Extremely well’.

*Indicates significantly different from control at the $p < 0.05$ level.

**Indicates significantly different from control at the $p < 0.01$ level.

***Indicates significantly different from control at the $p < 0.001$ level.

Table 7 also depicts the distribution of self-reported confidence and understanding among those who correctly specified the posteriors for the new Monty Hall problem and who indicated that the mental simulations explanation of the original Monty Hall problem affected their answers. The levels of self-

Table 6. *New Monty Hall problem—Distribution of participant answers.*

Question: What is the probability that door B conceals the prize after Monty Hall shows you the goat?

Possible answer	Percentage per condition (count)			
	Mental simulations	Possible models	Probability accrual	Control
91% (correct answer)	44% (11)**	4% (1)	4% (1)	4% (1)
80%	0% (0)	0% (0)	0% (0)	4% (1)
66% (original Monty Hall problem)	32% (8)*	68% (17)***	72% (18)***	4% (1)
50% (common answer)	12% (3)***	24% (6)***	20% (5)***	84% (21)
33%	12% (3)	4% (1)	4% (1)	4% (1)

Question: What is the probability that door A conceals the prize after Monty Hall shows you the goat?

Possible answer	Percentage per condition (count)			
	Mental simulations	Possible models	Probability accrual	Control
66%	16% (4)	20% (5)	28% (7)	8% (2)
50% (common answer)	16% (4)***	28% (7)**	20% (5)***	76% (19)
33% (original Monty Hall problem)	28% (7)	48% (12)**	48% (12)**	8% (2)
20%	4% (1)	0% (0)	0% (0)	4% (1)
9% (correct answer)	36% (9)*	4% (1)	4% (1)	4% (1)

*Indicates significantly different from control at the $p < 0.05$ level.
 **Indicates significantly different from control at the $p < 0.01$ level.
 ***Indicates significantly different from control at the $p < 0.001$ level.

Table 7. *Self-reported confidence and understanding among participants with correct posteriors who said the mental simulations explanation affected their answers ($n = 9$).*

Question: How confident are you that your answers were the correct ones?

Not at all confident	Slightly confident	Moderately confident	Very confident	Extremely confident
0 (10%)	0 (0%)	2 (22%)	6 (67%)	1 (11%)

Questions: How well did you understand why your answers were the correct ones?

Not at all well	Slightly well	Moderately well	Very well	Extremely well
0 (0%)	0 (0%)	4 (44%)	5 (56%)	0 (0%)

reported confidence and understanding are somewhat high, but these are not significantly different from the other conditions, including the control.

Experiment 2 made multiple comparisons, but only three comparisons were comparisons from the preregistered confirmatory hypotheses, while the rest were additional relationships in exploratory analysis. These confirmatory comparisons are from Hypothesis 3 (participants in the mental simulations condition are more likely than those in any other condition to choose the correct posteriors for the new Monty Hall problem) and Hypothesis 5 (participants in the possible models and probability accrual

conditions of Experiment 2 will be more likely to incorrectly state that door A and door B have 1/3 and 2/3 posterior probabilities of concealing the prize). A Bonferroni correction of $\frac{\alpha}{3} = .0167$ was applied to these 3 comparisons, all of which were still statistically significant after the correction: participants in the mental simulations condition were more likely than those in any other condition to choose the correct posteriors (Cohen's $h = 0.88$, 98.3% CI [0.21, 1.56], $p = 0.013 < \frac{\alpha}{3} = 0.0167$) and participants in the possible models and probability accrual conditions of Experiment 2 were more likely to incorrectly state that door A and door B have 1/3 and 2/3 posterior probabilities (Cohen's $h = 1.53$, 97.5% CI [0.85, 2.21], $p = 0.0002$). Furthermore, the rationales in Appendix E of the Supplementary Material are consistent with Hypothesis 3, since some of them display clear indications of using the mental simulations approach to reach correct posterior probabilities. In conclusion, the support for hypotheses 3 and 5 is not a statistical artifact.

4. Discussion

The aforementioned experiments showed that likelihood neglect bias exists and that training in the mental simulations approach can reduce this bias and help participants correctly solve the new and old Monty Hall problems. It falsified the theory that likelihood unawareness explains incorrect responses and it supported the theory that likelihood neglect explains incorrect responses. Results also indicate that two prominent approaches to the old Monty Hall problem do not help participants solve the new Monty Hall problem, presumably because the approaches are insensitive to the likelihoods, and they increase the risk that participants will give erroneous 1/3 and 2/3 posterior probabilities.

The new Monty Hall problem also further illustrates the power of Bayesian reasoning to deliver unintuitive yet highly accurate probabilities. In Experiment 2, the large majority of participants (21 or 84%) in the control group assigned door B a 50% probability of concealing the prize. However, the correct answer is that door B has a 91% probability of concealing the prize, as mentioned in the introduction. It is gratifying that 44% or 11 participants in the mental simulations condition provided the correct probability for door B (although only 36% or 9 participants correctly specified the probability for door A). Since the mental simulations approach accords with Bayesian reasoning, it can yield a 41% increase in accuracy and high confidence in the far more probable outcome in this case where participants would otherwise have very little confidence. More generally, then, the new Monty Hall problem illustrates how humans can fail to recognize objectively strong evidence for a hypothesis if they neglect likelihoods.

4.1. Improving understanding of the Monty Hall problems

Furthermore, Saenen et al. (2018) state that improving understanding in the Monty Hall problem is an unresolved challenge; however, the experiments reported in this article indicate that perhaps this challenge is resolvable. Across both experiments, 22 participants in total gave correct posterior probabilities and reported using—or being affected by—the mental simulations approach. Of these 22 participants, 7 (32%) reported understanding ‘Very well’ why their answers were correct while 11 (50%) reported understanding this ‘Moderately well’. This suggests that some participants may understand why their answers are correct when using the mental simulations approach (see Appendix A of the Supplementary Material for more details about how the training materials attempted to foster understanding).

The results show that the mental simulations approach is more likely to foster understanding of the new Monty Hall problem than either the possible models or probability accrual approaches. Of course, the mental simulations participants self-reported lower or similar levels of understanding relative to the other conditions in Experiment 2, including the control condition. But the other participants did not actually understand the correct answers better on average because the vast majority of those participants actually gave *incorrect* answers to the new Monty Hall problem. Put simply, a participant cannot understand the *correctness* of their answers if their answers are actually *incorrect*.

Furthermore, the mental simulations approach is more likely than the other approaches to provide a better understanding of the original Monty Hall problem too. This is for the following reasons. The correct answers for the posterior probabilities in the original Monty Hall problem are one-third and two-thirds for doors A and B, respectively. This is largely because of the likelihoods, as experts in the relevant statistics know. Thus, if a participant truly understands why the answers are correct, then they will understand the likelihoods and their importance, since these are, as a matter of uncontroversial statistical fact, the exact reason why the answers are correct (i.e., alongside the prior probabilities). However, when the likelihoods change, as with the new Monty Hall problem in Experiment 2, so too do the correct answers change. Therefore, if people understood why the original Monty Hall problem answers were correct in virtue of the likelihoods—that is, in virtue of the actual reason why those answers were correct—then they would understand that the new Monty Hall problem answer would be different in virtue of the different likelihoods—in virtue of the original reason now changing. But only the participants in the mental simulations approach condition were more likely to give the correct answer in the new Monty Hall problem—not the possible models or probability accrual conditions. It appears that this is because participants in the mental simulations condition better understood the importance of likelihoods, since this is the only relevant experimental feature that differed and could have led them to the correct answer for the new Monty Hall problem. For these reasons, if they understood the importance of likelihoods better across these problems, then they understood better why the original Monty Hall problem answer was correct too.

Ultimately, then, there are three lines of evidence suggesting that some participants in the mental simulations conditions genuinely understood why their answers were correct. First, they were more likely to give the correct posteriors when the likelihoods change, unlike any other condition, where these likelihoods are provably part of why the relevant answers are correct in both the new and original Monty Hall problems. Second, some of these participants self-report a relatively high level of understanding when they also display this other evidence of understanding (that is, correct posteriors when the likelihoods change). Third, the rationales for some of these participants suggest they have understanding because they appeal explicitly to likelihoods or to hypothetical frequencies when justifying their answers (see Appendix E of the Supplementary Material).

In any case, as mentioned in the introduction, likelihood neglect may exist in other more important contexts where information about likelihoods is relevant, such as medical diagnosis, legal criminal inquiries, scientific contexts, intelligence analysis and others. For that reason, replication attempts are encouraged in contexts both within and beyond the Monty Hall problems.

4.2. *Why humans fall prey to likelihood neglect*

So we have seen evidence that humans are susceptible to a likelihood neglect bias. What follows is a proposed theoretical explanation of why this might occur in the Monty Hall problem.

This article offers a two-part explanation of incorrect responses in the Monty Hall problem. The first is positive: it is about what participants *do* when reasoning. The second is negative: it is about what participants *do not* do when reasoning.

The first part of the explanation is that participants have equiprobabilistic intuitions, as has been noted by others. It is well-known that the majority of participants think that the unopened doors have the same probability of concealing the prize (Tubau et al., 2015). This represents an intuitive heuristic which humans apply: if there is no reason to favor one outcome over another, then we should regard each outcome as equally probable. This heuristic is simply a version of the *principle of indifference*—a principle that has both advocates and opponents in philosophy (Kaplan, 1996; White, 2010; Wilcox, 2020). This heuristic may sometimes be explained by the availability heuristic (Tversky and Kahneman, 1973). The availability heuristic is the process whereby humans determine the probability of an outcome based on its mental availability—that is, the ease with which relevant instances come to mind. In the context of the Monty Hall problem, both the remaining options are equally available: it is just as easy to call to mind an instance of one unopened door concealing the prize as it is for the other.

Consequently, then, humans utilize an equiprobability heuristic, and this heuristic may itself be explained by the availability heuristic—at least in the context of the Monty Hall problem. However, judgments of equiprobability are not inevitable, and indeed a sizable number of philosophers disavow the principle of indifference (Joyce, 2010; Kaplan, 1996; Meacham, 2014). So it is clear that judgments of equiprobability can be *overridden* if other factors are present. As an additional example, Saenen et al., (2015) report that subjects were more likely to switch doors and override their intuitions if they could see statistical data that the other door concealed the prize most of the time.

This leads us to the second part of explanation—the negative part. In particular, participants often give incorrect responses because they lack the cognitive factors that would impel them to switch and, more specifically, they lack knowledge of the importance of likelihoods. In this respect, humans are not born with an innate, infallible knowledge of the law of likelihood. This is plausible for several reasons. The most immediate is that, as demonstrated in the Monty Hall problem, people can violate the law of likelihood. But beyond that, the law of likelihood requires making distinctions that humans systematically blur. In particular, applying the law of likelihood requires the ability to distinguish between the likelihood of the evidence given some hypothesis on the one hand and the probability of the hypothesis given the evidence on the other. Yet humans often fail to make such distinctions, as has been well-documented with the confusion of the inverse fallacy (Villejoubert and Mandel, 2002).

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

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

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

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

This article offers the theory that, in these contexts, people are using forms of reasoning *aside from Bayesian reasoning with likelihoods*, and the information about likelihoods is merely used to inform or engage these other forms of reasoning. Exactly how this is done, however, depends on the specific context.

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

What, then, is the explanation for the other—that is, for base rate neglect in medical diagnosis? Here, the present article postulates a multi-part process. First, humans judge that there is some kind of probabilistic connection between having a disease and the evidence—such as various symptoms or

a positive test result. Participants may make this connection because they have what we could call *an associational history*—that is, experiences, memories or information which associate the hypothesis under consideration with the putative evidence that bears on it. In this context, people frequently associate diseases with both symptoms and test results, in part because we often experience illnesses concurrently with symptoms and positive test results. For that reason, when we are informed that a disease would probably give rise to a symptom or a test result, we make an intuitive connection between the two—we think the evidence says something about the probability of the disease. As per Krynski and Tenenbaum (2007), one might think participants see a special *causal* connection between the evidence and the hypothesis. So the first step is to make an intuitive connection between the evidence and the hypothesis under consideration—to think that they have *some* kind of probabilistic connection.

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

So we have two accounts of how undue reliance on likelihoods may really be an approximation to other forms of reasoning, and likelihoods are then merely used to inform or engage these other forms of reasoning. Given such accounts, we can summarize the aforementioned theory about why likelihood neglect may occur in some circumstances while undue reliance on likelihoods may occur in others. In particular, when reasoning about a given problem, humans engage one of any number of processes to solve the problem, depending on the specific features of the problem and their knowledge, familiarity with similar problems and the like. If the problem has specific features, then the problem may engage undue reliance on the likelihoods and even the inverse probability confusion. Such features can include the problem approximating modus tollens-like problems (as in null hypothesis testing). Or they can include features of the situation whereby the person has a causal or historical association between the evidence and hypotheses (as one does with diseases and symptoms). In such cases, the likelihoods are used to support the premise of a modus tollens argument or to inform the participant's intuitive judgment about how the evidence and hypotheses are related. However, if the problem lacks such features, then humans will use other processes to judge probabilities, possibly including the equiprobability or availability heuristics. In such cases, they are sometimes susceptible to the likelihood neglect bias.

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

4.3. Limitations and future research

However, this study has several limitations, each of which give rise to promising avenues for future research.

First, the present study has showed the existence of the bias in only two contexts: the new and old Monty Hall problems. Future research could test for whether the bias occurs in other contexts, including the problem of the prisoners, as well as naturalistic settings where, say, two or more hypotheses might be equally mentally available and hence regarded as equiprobable, even though some hypothesis makes the evidence more likely than others.

A second limitation is that Experiment 2 tested the efficacy of alternative (possible models and probability accrual) approaches using a particular presentation of those approaches. The presentation of these approaches was simple, partly because of the inherent simplicity of those approaches (compared to the mental simulations approach) and because any potential addition to those presentation faced

particular risks, such as not being endorsed by any proponent of the alternative approaches (thus potentially being a strawman) or unduly lengthening the alternative approaches (thus potentially increasing respondent fatigue, confusion and confounds). However, proponents and sympathizers of those approaches may wish to run follow-up studies with modified presentations to determine whether they fare better on the new Monty Hall problem. That said, as mentioned in the introduction, there are theoretical reasons to doubt that any such modification will be successful unless supplemented in ways that are attentive to the likelihoods, in which case they would then cease to be the very simplistic approaches that others have endorsed and that this article discusses exactly because they ignore the likelihoods.

A third limitation is that neither experiment revealed much about *why* some participants successfully used the mental simulations approach whereas others did not. Indeed, most participants in the mental simulations condition of both experiments did not specify the correct posterior probabilities, potentially indicating a limitation stemming from the approach, the training materials used to teach it, the Mechanical Turk participant population or some other as yet undetermined source. However, it is perhaps worth noting that this article's pilot experiments indicated that Stanford students were more likely to successfully understand and use the approach compared to a sample of adults from the general population, potentially indicating that some cognitive or other variables predict better performance. What variables those might be, however, is a topic for future research.

A final limitation is that some correct posteriors in Experiment 1 may have resulted from two particular causes aside from the mental simulations approach: (1) participants looking up answers online and (2) participants merely analogizing from the problem of the prisoners (but without understanding the mental simulations approach). These two explanations are discussed in much more depth in Appendix D of the Supplementary Material, and there are reasons to doubt both of them: for example, the highlighted rationales in Appendix E of the Supplementary Material show clear use of the mental simulations procedures and a third of participants in Experiment 2 got the correct posteriors despite not being exposed to the problem of the prisoners nor being able to search the new Monty Hall problem online (since it is an innovation of this experiment). That said, to further experimentally test these alternative explanations, future research could replicate the experiments but with (1) concurrent experimenter supervision (to ensure participants do not do internet searches) and (2) an additional experimental condition providing the answer to the problem of the prisoners but without the mental simulations explanation. These two innovations should be implemented together to ensure that the problem of the prisoners does not prompt participants to search for answers to the Monty Hall problem online. In fact, the experimental protocol has been designed for both of these innovations (and is available at the OSF link provided in section 1.5), but the present investigation was unable to carry it out for practical reasons. Regardless, other researchers are welcome and indeed encouraged to implement these innovations (with or without the present author's collaboration).

4.4. Implications for Bayesian cognitive science

So the aforementioned results suggest humans do not reason with likelihoods as normative Bayesian models of cognition prescribe. Such results then provide additional evidence to refute Bayesian cognitive science *if* 'Bayesian cognitive science' is understood as the *global* claim that all human thinking processes conform to Bayesian norms. Clearly human thinking processes often violate Bayesian norms, but this was already known from the study of other biases (Gilovich et al., 2002), so it is not clear if any scholars endorse this claim.

However, the results are consistent with Bayesian cognitive science *if* 'Bayesian cognitive science' is understood as the more *local* claim that only *some* but *not* all human thinking processes conform to Bayesian norms. After all, other cognitive processes may be Bayesian, even if the ones explored here are not, and this is precisely the position of scholars like Rescorla (2020).

The results are also consistent with what we may call *normative Bayesian cognitive science*—the claim that Bayesian norms provide normative standards against which to assess at least some

mental processes, and one of the aims of cognitive science is to develop interventions to bring these mental processes in conformity with those standards. In fact, the very approach of this article has been normatively Bayesian: to understand when and why humans depart from Bayesian reasoning and to test an intervention that narrows the gap between the two. In this case, that intervention—the mental simulations approach—appears to be somewhat successful.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/jdm.2024.8>.

Data availability statement. Data and code are available in the OSF repository at https://osf.io/kdme9/?view_only=516d7baaf57b452b9fe958c83365ac6c.

Funding statement. These studies were funded with generous support from the Stanford Interdisciplinary Graduate Fellowship, the Stanford University Department of Philosophy and James L. McClelland.

Competing interest. The author has no competing interests to declare.

References

- Bandyopadhyay, P. S. (2011). *Philosophy of statistics*. Amsterdam: North Holland.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211–233. [https://doi.org/10.1016/0001-6918\(80\)90046-3](https://doi.org/10.1016/0001-6918(80)90046-3)
- Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: Experimental study. *BMJ*, 333(7562), 284.
- 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.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, 49(12), 997–1003.
- 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>
- Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In: D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge: Cambridge University Press.
- Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*, 5(1), 75–98. <https://doi.org/10.1177/0959354395051004>
- Gardner, M. (1959). Mathematical games: Problems involving questions of probability and ambiguity. *Scientific American*, 201, 174–182. <https://doi.org/10.1038/scientificamerican1059-174>
- Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90(3), 239–260. <https://doi.org/10.1037/0033-295X.90.3.239>
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8(2), 53–96.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4), 684.
- Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review*, 106(2), 425–430.
- Gilovich, T., Griffin, D. W., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.
- Granberg, D., & Dorr, N. (1998). Further exploration of two-stage decision making in the Monty Hall dilemma. *American Journal of Psychology*, 111(4), 561. <https://doi.org/10.2307/1423551>
- Hacking, I. (2016). *Logic of statistical inference*. Cambridge: Cambridge University Press.
- Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology*, 101(2), 252–254. <https://doi.org/10.1037/h0035224>
- Hawthorne, J. (2018). Supplement to inductive logic: Likelihood ratios, likelihoodism, and the law of likelihood. In *The Stanford encyclopedia of philosophy* (Spring 2018 Edition). <https://plato.stanford.edu/entries/logic-inductive/sup-likelihood.html>
- Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, 84(3), 343–352.
- Hoffrage, U., Hafenbrädl, S., & Bouquet, C. (2015a). Natural frequencies facilitate diagnostic inferences of managers. *Frontiers in Psychology*, 6, 642.
- Hoffrage, U., Krauss, S., Martignon, L., & Gigerenzer, G. (2015b). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Frontiers in Psychology*, 6, 1–14. <https://doi.org/10.3389/fpsyg.2015.01473>

- Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, 290(5500), 2261–2262.
- Johnson-Laird, P. N. (2012). Inference in Mental Models. In *The Oxford handbook of thinking and reasoning* (pp. 134–154). Oxford University Press.
- Joyce, J. M. (2010). A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, 24(1), 281–323. <https://doi.org/10.1111/j.1520-8583.2010.00194.x>
- Kalinowski, P., Fidler, F., & Cumming, G. (2008). Overcoming the inverse probability fallacy: A comparison of two teaching interventions. *Methodology*, 4(4), 152–158. <https://doi.org/10.1027/1614-2241.4.4.152>
- Kaplan, M. (1996). *Decision theory as philosophy*. Cambridge: Cambridge University Press.
- Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, 132(1), 3–22. <https://doi.org/10.1037/0096-3445.132.1.3>
- 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>
- Kurzenhäuser, S., & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher*, 24(5), 516–521.
- Liu, A.-Y. (1975). Specific information effect in probability estimation. *Perceptual and Motor Skills*, 4(1), 475–478.
- Meacham, C. J. G. (2014). Impermissive Bayesianism. *Erkenntnis*, 79(6), 1185–1217. <https://doi.org/10.1007/s10670-013-9547-z>
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72(3), 346.
- Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. New York: Wiley.
- Rescorla, M. (2020). A realist perspective on Bayesian cognitive science. In A. Nes & T. Chan (Eds.), *Inference and consciousness* (pp. 40–73). New York: Routledge
- Rosenthal, J. S. (2008). Monty hall, monty fall, monty crawl. *Math Horizons*, 16(1), 5–7
- Russell, B. (1948). *Human knowledge: Its scope and limits*. Simon and Schuster.
- Saenen, L., Heyvaert, M., Van Dooren, W., Schaeken, W., & Onghena, P. (2018). Why humans fail in solving the Monty Hall dilemma: A systematic review. *Psychologica Belgica*, 58(1), 128–158. <https://doi.org/10.5334/pb.274>
- Saenen, L., Van Dooren, W., & Onghena, P. (2015). A randomised Monty Hall experiment: The positive effect of conditional frequency feedback. *Thinking & Reasoning*, 21(2), 176–192. <https://doi.org/10.1080/13546783.2014.918562>
- Tabossi, P., Bell, V. A., & Johnson-Laird, P. N. (1999). 13 Mental models in deductive, modal, and probabilistic reasoning. In *Advances in psychology* (Vol. 128, pp. 299–331). North-Holland.
- 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>
- Placeholder Text Tubau, E., Alonso, D., & Alonso, D. (2003). Overcoming illusory inferences in a probabilistic counterintuitive problem: The role of explicit representations. *Memory & Cognition*, 31(4), 596–607. <https://doi.org/10.3758/BF03196100>
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. *Memory and Cognition*, 30(2), 171–178. <https://doi.org/10.3758/BF03195278>
- White, R. (2010). Evidential symmetry and mushy credence. In T. Szabó Gendler and J. Hawthorne (Eds.), *Oxford studies in epistemology* (pp. 161–186). Oxford: Oxford University Press.
- Wilcox, J. (2020). An argument for the principle of indifference and against the wide interval view. *Journal for General Philosophy of Science*, 51(4), 65–87.