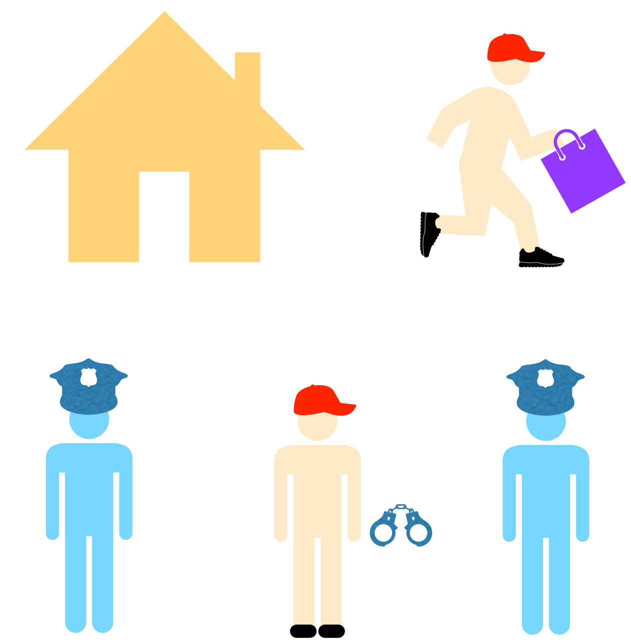
Intuition and Probability
- Monty Hall and Bayesian Reasoning

Lunch and Learn @ Jellyfish

July 21, 2020

Warm Up

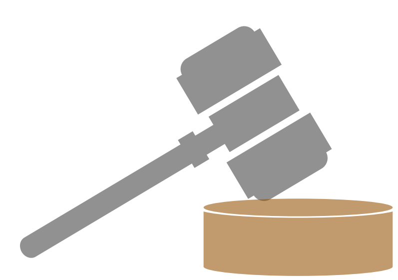
- Detectives in a city (whose **population is one million**) are working on a crime
- They have a description of the perpetrator such that only **one person in 10,000** fits the description.
- On a routine patrol, police find a person fitting the description. This person is brought into trial based **solely** on the fact that he fits the description



During the trial, the prosecutor states:

"Since only one person in 10,000 fits the description, it is highly unlikely that an innocent person fits the description. Thus it is highly unlikely that the defendant is innocent."

You are a member of the jury.
Do you cast a "Guilty" vote?



Guilty!

or

Not Guilty!

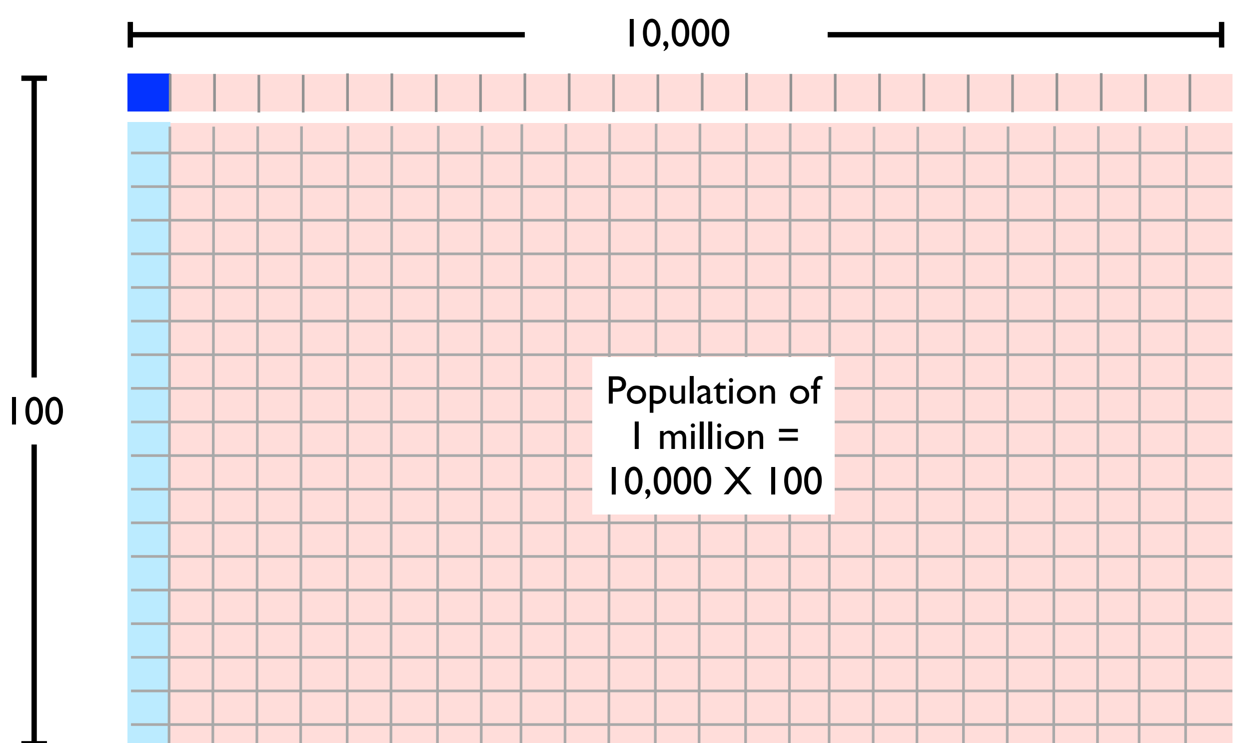
Warm Up

- Detectives in a city (whose **population is one million**) are working on a crime
- They have a description of the perpetrator such that only **one person in 10,000** fits the description.
- On a routine patrol, police find a person fitting the description. This person is brought into trial based solely on the fact that he fits the description

During the trial, the prosecutor states:

"Since only one person in 10,000 fits the description, it is highly unlikely that an innocent person fits the description. Thus it is highly unlikely that the defendant is innocent."

You are a member of the jury.
Do you cast a "Guilty" vote?



Solution:

1 out of every 10,000 people fits the description

But there are 1 million people in the city

So we can expect that $\frac{1}{10{,}000} \times 1{,}000{,}000 = 100$ people fit the description

The probability that any **one** person **out of this 100** is the perpetrator is $\frac{1}{100}$

So if arrest is based on description alone, it is actually highly likely that the defendant is innocent.

You should **not** cast a guilty vote

Intro to Bayesian Reasoning

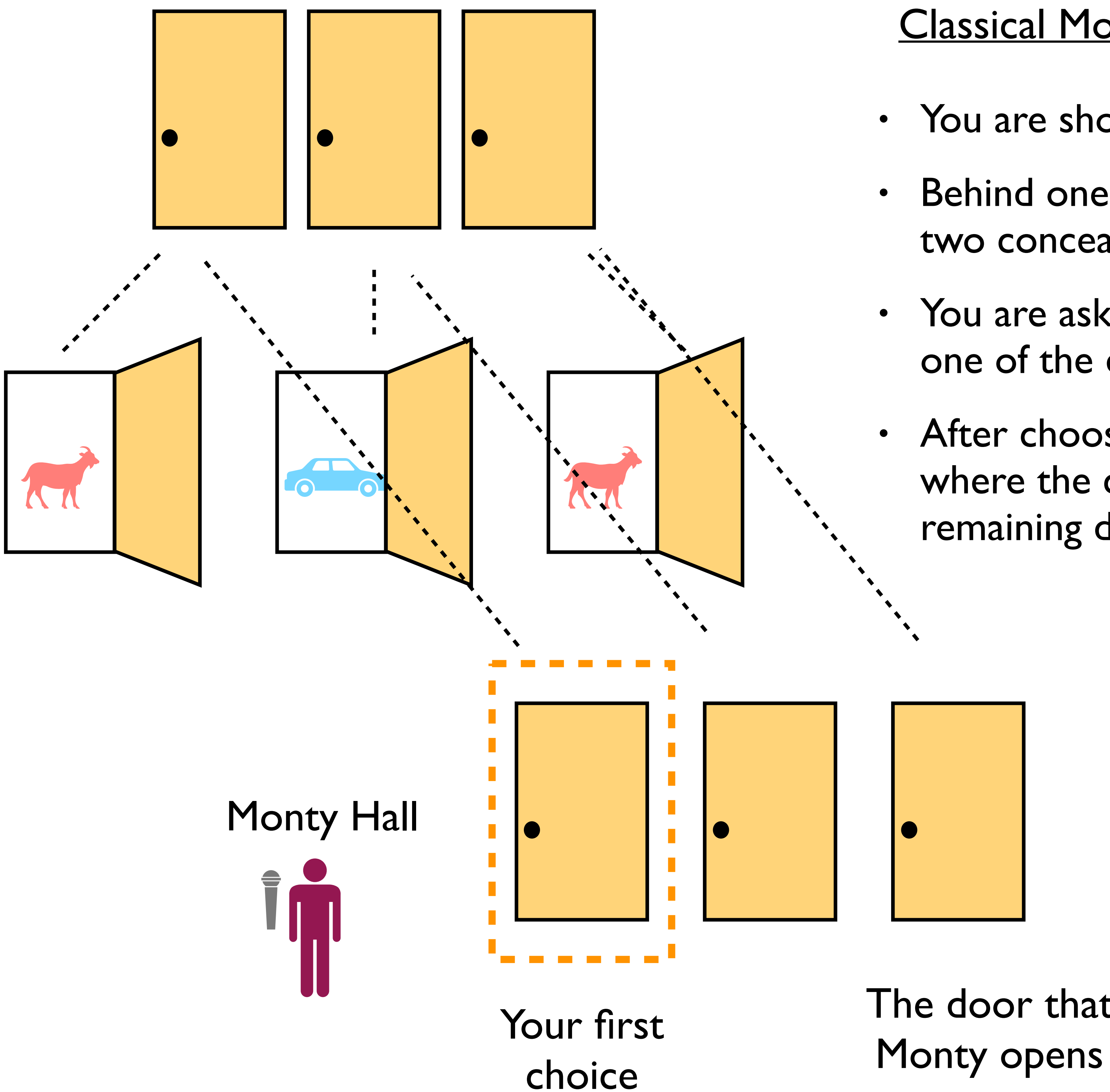
This problem exemplifies the canonical idea behind **Bayesian Reasoning**

Bayesian Reasoning (in a nutshell) : Information —changes→ Probability estimates

(e.g.,Perpetrator and Jury problem) The number of possible perpetrators in the city —changes→ Probability estimate that any one suspect is the perpetrator

Let's see how this idea applies to more famous probability problem

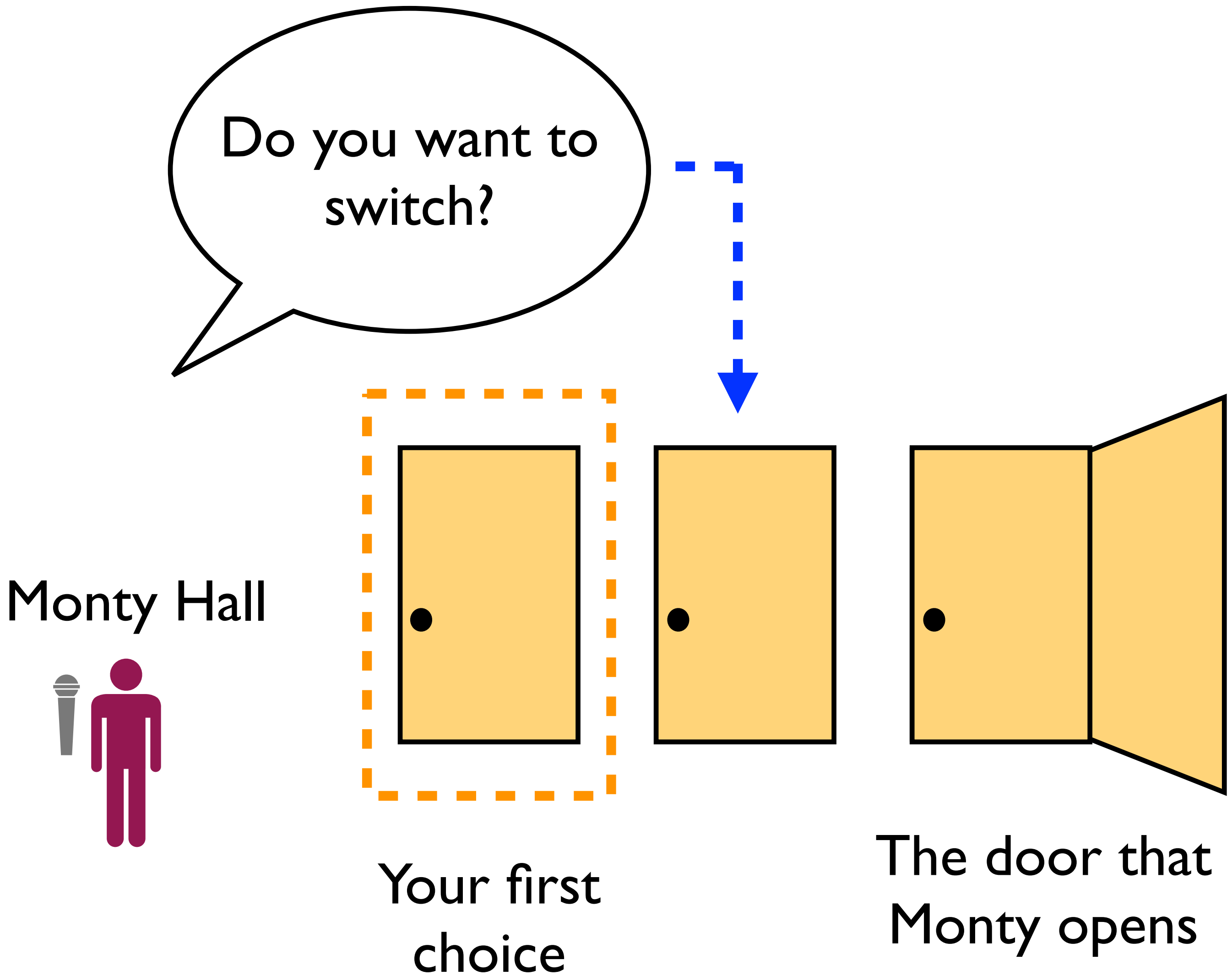
Classical Monty Hall Problem - Part I



Classical Monty Hall Problem (Part I)

- You are shown three identical doors.
- Behind one door is a car and the other two conceal goats.
- You are asked to choose, but NOT open one of the doors.
- After choosing a door, Monty (who knows where the car is) opens one of the two remaining doors.

Classical Monty Hall Problem - Part II



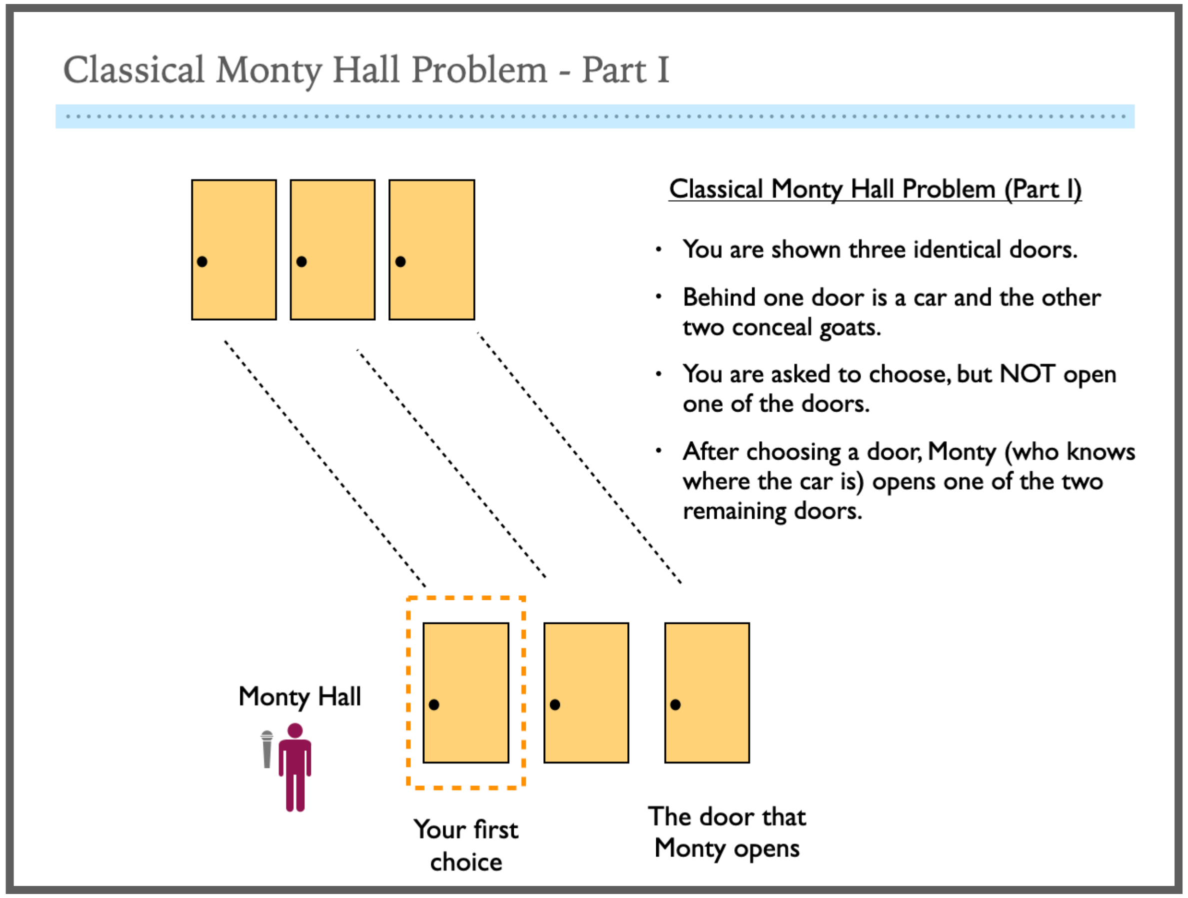
<u>Classical Monty Hall Problem (Part II)</u>

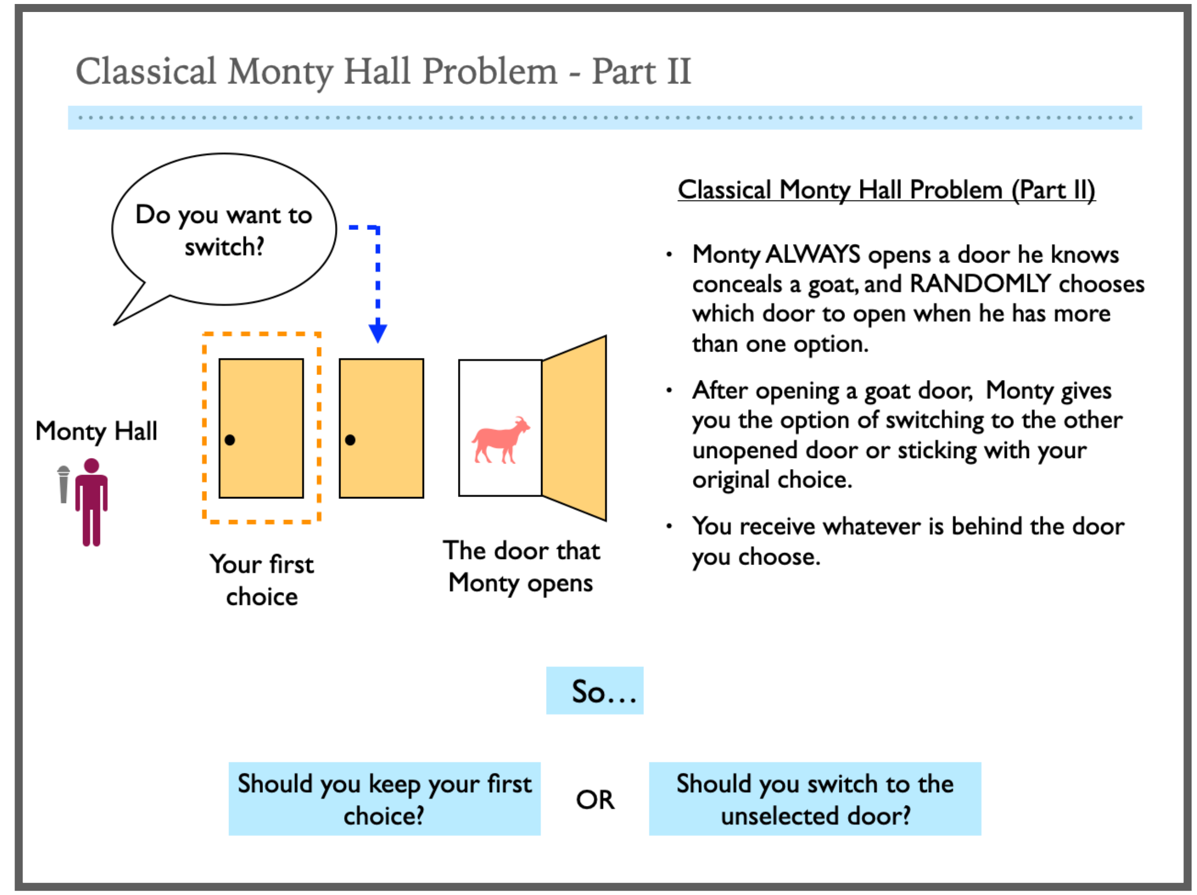
- Monty ALWAYS opens a door he knows conceals a goat, and RANDOMLY chooses which door to open when he has more than one option.
- After opening a goat door, Monty gives you the option of switching to the other unopened door or sticking with your original choice.
- You receive whatever is behind the door you choose.

So…

Should you keep your first choice? OR Should you switch to the unselected door?

Thinking…





(Jeopardy Music)

The Answer Is…

You should switch!

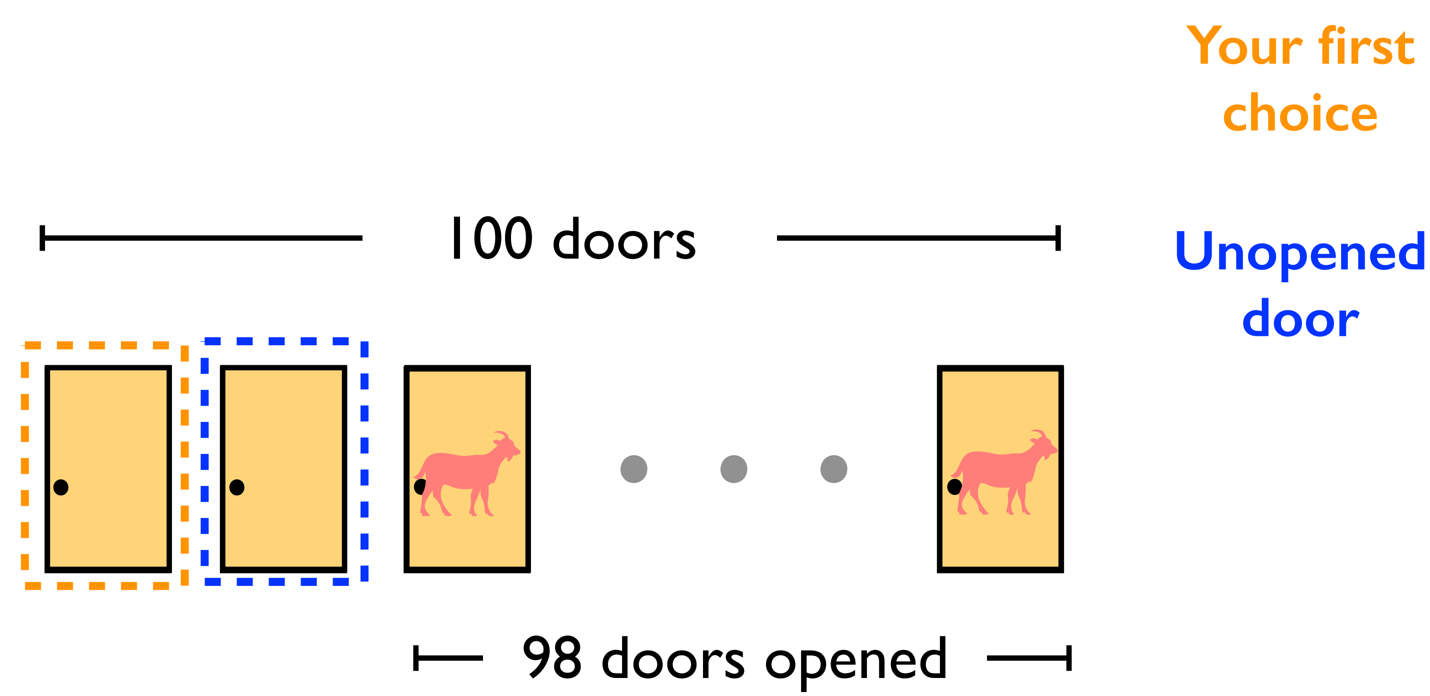
that is…

You are more likely to get the car, if you switch to the unopened door than if you keep your first choice.

Explanation #1

Let's modify the problem → Instead of three doors let's say we have **100** doors

- **One** door hides a car and the other **99** doors hide goats
- You select one door
- Monty opens 98 doors that hide goats



Should you switch to the unopened door?

YES!

In 1/100 of the scenarios, your first choice has the car

BUT!

In 99/100 of the scenarios, the unopened door has the car

The Answer Is…

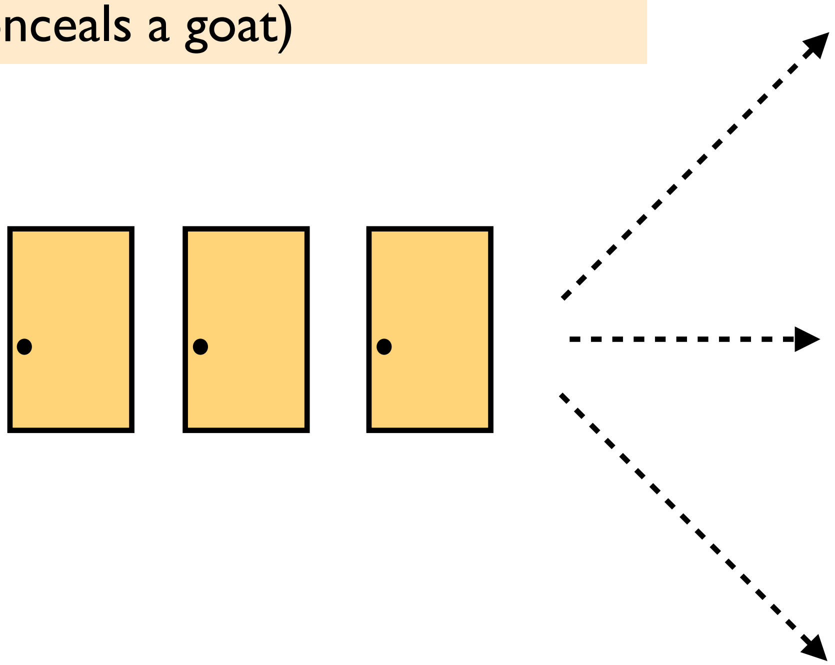
You should switch!

that is…

You are more likely to get the car, if you switch to the unopened door than if you keep your first choice.

Explanation #2

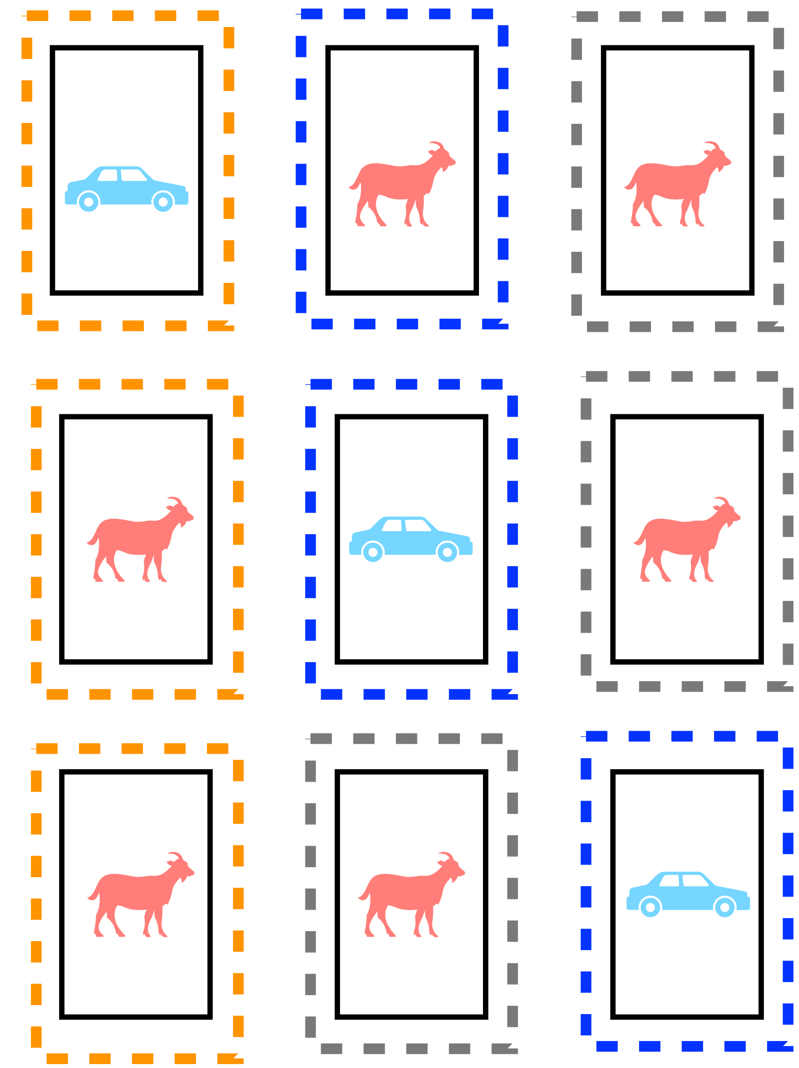
NOTE (We know that Monty ALWAYS opens a door that conceals a goat)



Scenario #1

Scenario #2

Scenario #3



Your first choice

Unopened door

Monty's reveal

Three equally likely scenarios

Why you should switch:

In 1/3 of the scenarios, your first choice has the car

BUT!

In 2/3 of the scenarios, the unopened door has the car

The Answer Is…

You should switch!

that is…

You are more likely to get the car, if you switch to the unopened door than if you keep your first choice.

“Explanation” #3

Let’s simulate the problem

Create 10,000 rounds of the game

Car is randomly placed behind one of the doors for each round

Two Strategies

Strategy 1: Always Stay

Strategy 2: Always Switch

Which strategy wins most often?

Link to Notebook

https://colab.research.google.com/drive/1WBDYt_JxusZR0bDxr_FO4PPTjt5aO7Xr?usp=sharing

How does the public fare with this problem?

In 1990, Marilyn vos Savant a Q&A columnist for *Parade* was given the Monty Hall problem by a reader.

She answered the problem correctly (with correct probabilities) and gave the "100 door" explanation for why.

Here's what followed…

vos Savant received thousands of letters.

- 92% of the letters from the *general public* disagreed with her
- 65% of the letters with a *university address* disagreed with here

Example disagreement from a university reader

> …You blew it! Let me explain. If one door is shown to be the loser, that information changes the probability of either remaining choice, *neither of which has any reason to be more likely*, to 1/2. As a professional mathematician I'm very concerned with the general public's lack of mathematical skills. Please help by confessing your error and in the future being more careful.

How does the public fare with this problem?

In 1990, Marilyn vos Savant a Q&A columnist for *Parade* was given the Monty Hall problem by a reader.

She answered the problem correctly (with correct probabilities) and gave the "100 door" explanation for why.

Here's what followed…

vos Savant received thousands of letters.

- 92% of the letters from the *general public* disagreed with her
- 65% of the letters with a *university address* disagreed with here

vos Savant wrote on the topic a **second time** and gave a new explanation (similar to Explanation #2) she received more mail:

You are utterly incorrect about the game-show question, and I hope this controversy will call some public national attention to the serious national crisis in mathematical education…

May I suggest that you obtain and refer to a standard textbook on probability before you try to answer a question of this type again?

How does the public fare with this problem?

In 1990, Marilyn vos Savant a Q&A columnist for *Parade* was given the Monty Hall problem by a reader.

She answered the problem correctly (with correct probabilities) and gave the "100 door" explanation for why.

Here's what followed…

vos Savant received thousands of letters.

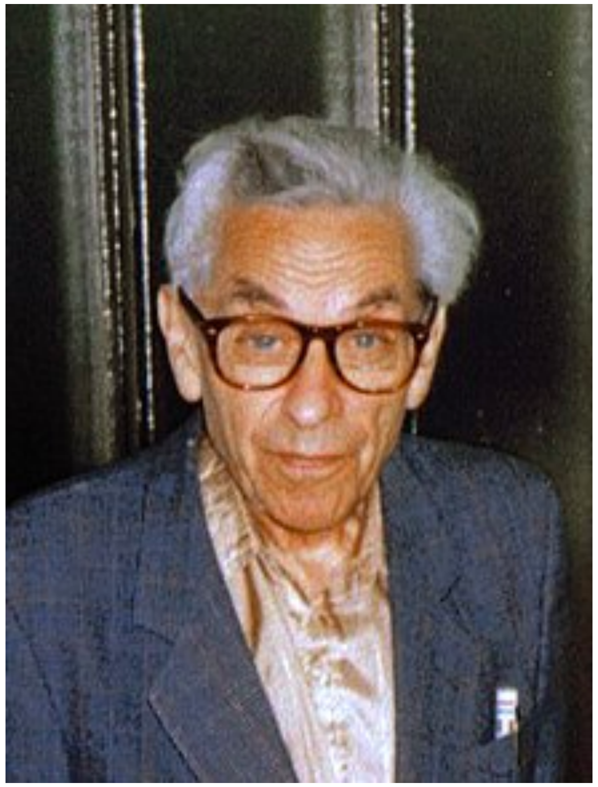
- 92% of the letters from the *general public* disagreed with her
- 65% of the letters with a *university address* disagreed with here

For her **third response**, vos Savant suggested that classrooms perform simulations of the problem. This seemed to have been more convincing.

Our class, with unbridled enthusiasm, is proud to announce that our data support your position. Thank you so much for your faith in America's educators to solve this.

I must admit I doubted you until my fifth-grade math class proved you right. All I can say is wow!

Paul Erdös gives it the old college try



Paul Erdös
(1913 - 1996)

In 1983, won the Wolf Prize in Mathematics

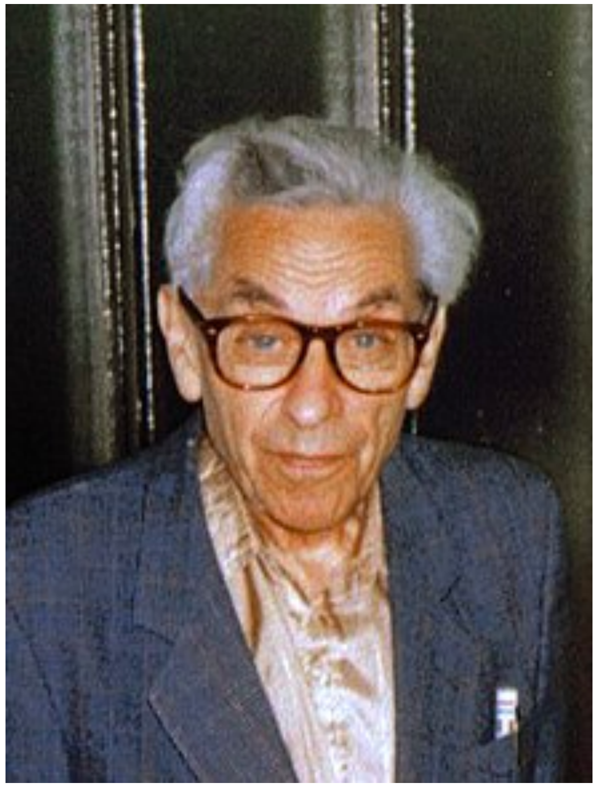
"for his numerous contributions to number theory, combinatorics, probability, set theory and mathematical analysis, and for personally stimulating mathematicians the world over"

- Lived vagabond existence
- Had 500 collaborators
- Published 1500 papers
- "Kevin Bacon of Mathematics";
  - "Erdös Number" is a mathematician's degrees of separation from Erdös by collaboration.

    (E.g., Nick has an Erdös number of 3(?))

Essentially the real-life version of a stereotypical mathematician who only cares for numbers

Paul Erdös gives it the old college try



Paul Erdös
(1913 - 1996)

Essentially the real-life version of a stereotypical mathematician who only cares for numbers

(Hoffman, *The Man Who Loved Only Numbers*)

Fellow mathematician Vazsonyi told Erdös about the Monty Hall problem.

"I told Erdös that the answer was to switch," said Vazsonyi, "and fully expected to move to the next subject. But Erdös, to my surprise, said 'No, that is impossible. It should make no difference.' At this point I was sorry I brought up the problem… An hour later he came back to me really irritated 'You are not telling me *why* to switch,' he said. 'What is the matter with you?' I said I was sorry, but that I didn't really know why…He got even more upset."

Erdös was eventually convinced by a simulation of the problem.

Why do people have so much trouble with this problem?

(Gilovich, Medvec, Chen, 1995)

(De Neys, Vershueren, 2006)

- People are afraid of switching and being wrong
- The problem taxes people's working memory
- People are bad at Bayesian reasoning (fancy term for conditional probabilities)

! Most popular reason

Bayesian Reasoning (in a nutshell) : Information → changes → Probability estimates

(e.g., Perpetrator and Jury problem): The number of possible perpetrators in the city → changes → Probability estimate that any one suspect is the perpetrator

(e.g., Monty Hall Problem): The fact that Monty always opens a goat door → changes → Probability estimate that car is behind one of the unopened doors

This has particular relevance today for Covid tests

Covid and Conditional Probabilities

Let's say you get tested for Covid and the test comes back negative?

Does that mean you **absolutely** don't have Covid?

Not necessarily!

You need to take into account two pieces of information:

- Your **base probability** for having Covid
- The "**accuracy**" of the test.

(Good, Hernandez, Smith, 2020)

Interpreting COVID-19 Test Results: a Bayesian Approach

J Gen Intern Med
DOI: 10.1007/s11606-020-05918-8

Check for updates

INTRODUCTION

As physicians care for patients with contact history and symptoms that might represent coronavirus disease 2019 (COVID-19), interpreting the results of polymerase chain reaction (PCR) assays from nasal and pharyngeal swabs is crucial. While a positive result in an acutely ill patient is straightforward, how should physicians interpret negative tests in patients with suspected COVID-19 infection?

Physicians and patients often place inappropriate confidence in test results, even when those tests are imperfect.[1] Specifically, physicians may minimize their own clinical reasoning (e.g., their pre-test probability of disease) and defer to a test result that may not be correct. With PCR testing for COVID-19, false negative tests are particularly concerning, potentially leading to an inappropriate sense of security regarding infectivity.

To accurately interpret test results, one needs to know the positive and negative predictive values of a test in the setting applied, which depend on its sensitivity and specificity, along with prevalence or pre-test probability. Although the specificity of PCR assays for COVID-19 appears to be close to 100%, documenting its sensitivity is surprisingly elusive.[2] Real-world sensitivity of the COVID-19 assay is especially impacted by difficulty in sampling technique for obtaining specimens

cough, and subjective dyspnea. She works in an emergency room that has evaluated numerous COVID-19 patients. She reports using appropriate personal protective equipment. We estimated a pre-test probability of COVID-19 infection at 90% (but varied it to as low as 70%).
Scenario 2 (low pre-test probability of COVID-19 infection): A 25-year-old male presents with subjective fevers (no temperature taken), cough, and subjective dyspnea. He has no significant exposures but lives where COVID-19 infections were reported; he has worked at home for the past month with occasional shopping for food. He reports frequent hand washing and practices social distancing. We estimated a pre-test probability of COVID-19 infection at 5% (but varied it to as high as 10%).

RESULTS

For the high-risk scenario with our estimated 90% pre-test probability, the post-test probability of a false negative test ranged from 47 to 73% (Table 1). With a 70% pre-test probability, the post-test probability of a false negative ranged from 19 to 41%. For a low-risk scenario with a pre-test probability of 5–10%, the post-test disease probability with a negative test ranged from 0.5 to 3.2%. Disease likelihood with a positive test remained > 99.9% in the high-risk scenario and > 97.4% in the low-risk patient.

Table 1

Estimates for Post-Test Probability of Acute COVID-19 Infection for Simulated Patient Scenarios

Clinical Scenarios	Pre-test probability (%)	PCR assay sensitivity (%)	Post-test probability of acute COVID-19 infection	
			Positive test (%)	Negative test (%)
Patient 1: high pre-test probability	70	70	100	41.2
		90	100	18.9
	90	70	100	73.0
		90	100	47.4
Patient 2: low pre-test probability	5	70	97.4	1.6
		90	97.9	0.5
	10	70	98.7	3.2
		90	99.0	1.1

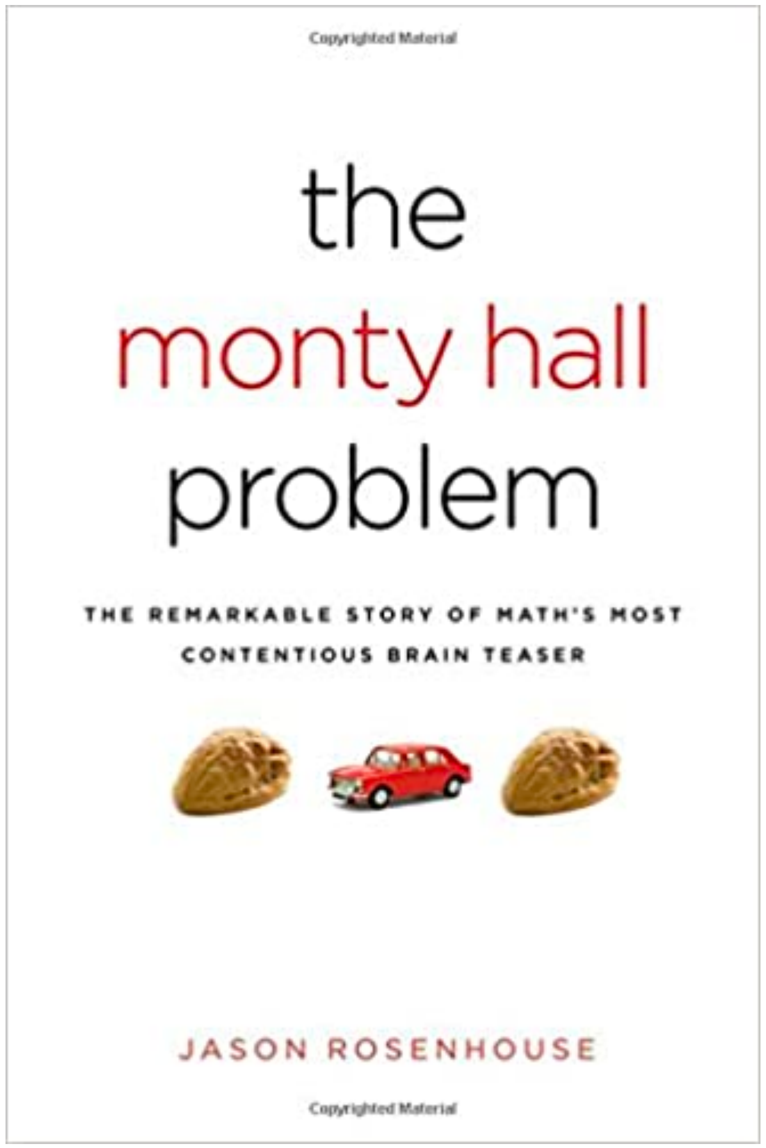
Patient 1: Healthcare worker showing symptoms

Patient 2: Person who has been sheltering in place with no symptoms

If you are a highly susceptible demographic and the test isn't perfectly accurate, **a negative test does not absolutely imply that you don't have covid**

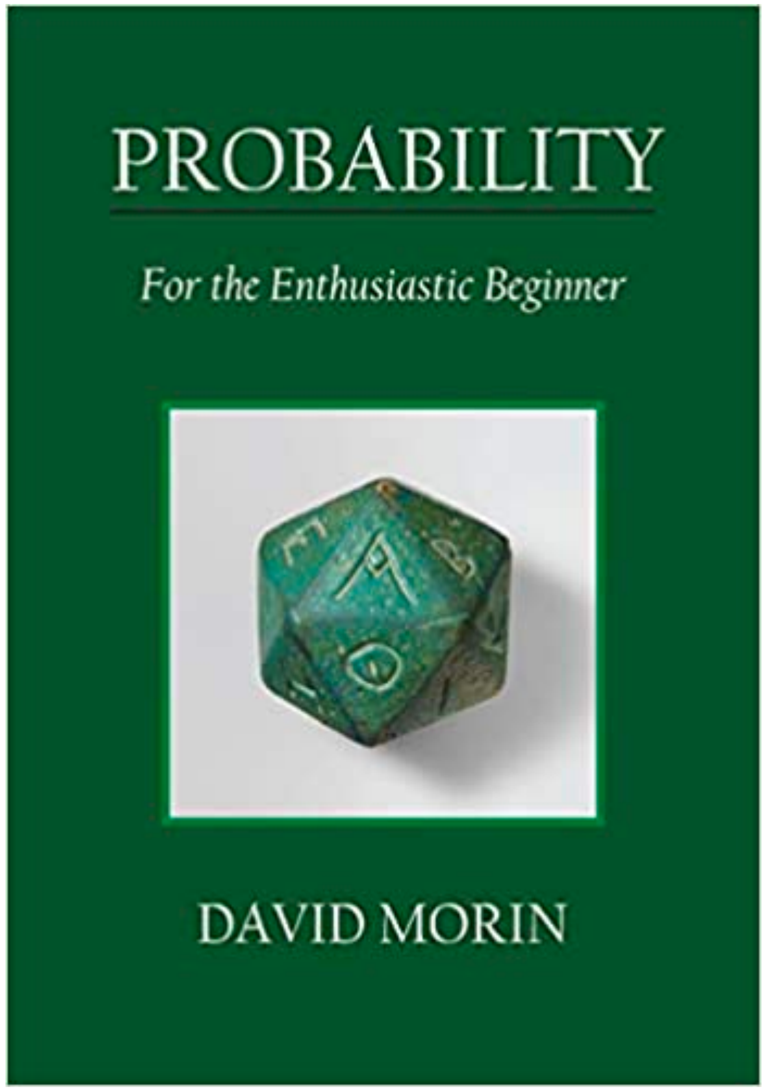
Resources

The Monty Hall Problem by Jason Rosenhouse



Has 15 variations of the problem; Discusses history, philosophy, and cognitive science of misconceptions

Probability for Enthusiastic Beginner (Chapter 2 available online) by David Morin



Chapter 2 has a great discussion on classical probability problems that illustrate how non-intuitive Bayesian reasoning is

End