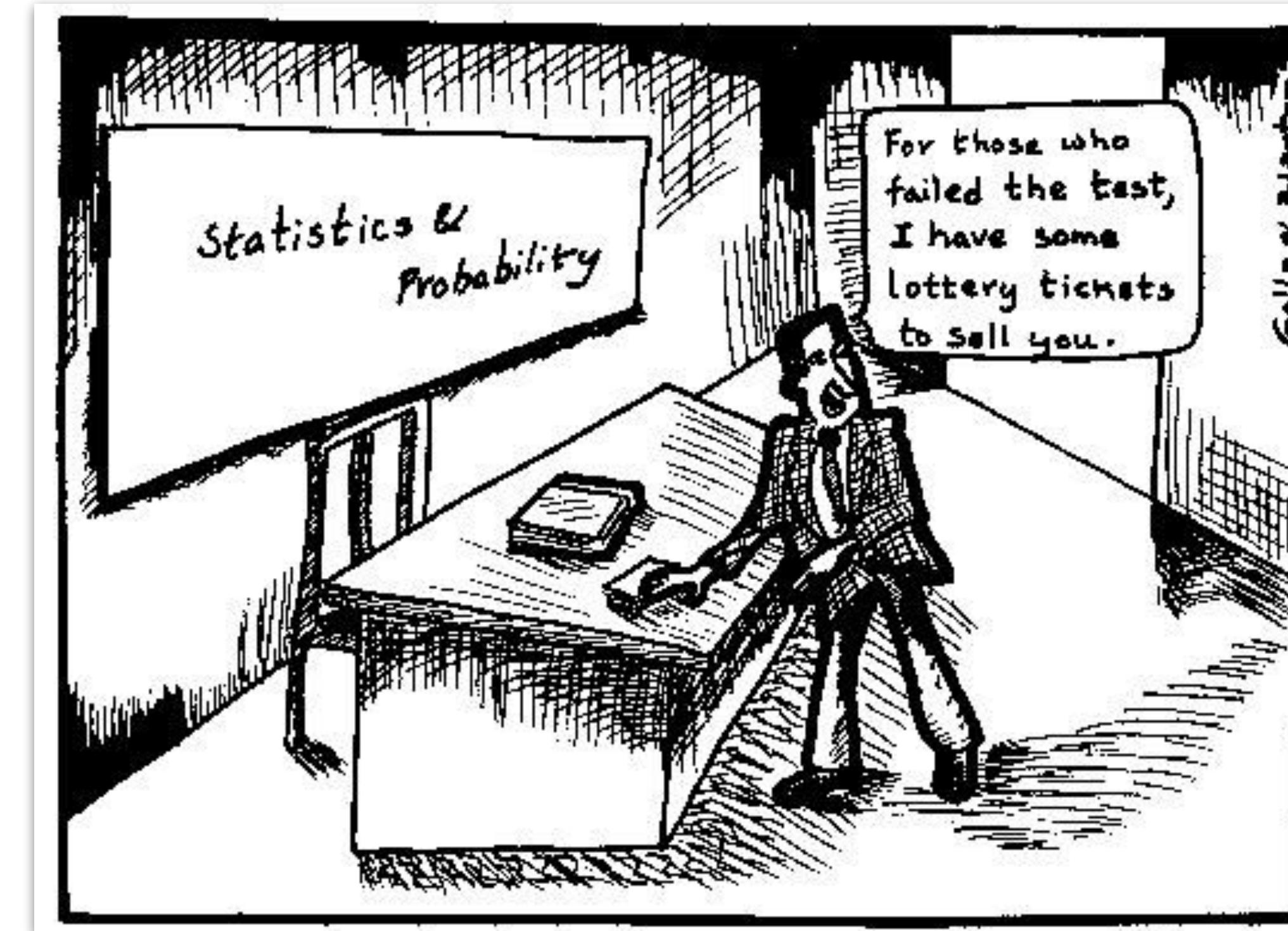


Probability



COLLABORATIVE PLAYLIST
psych252
<https://tinyurl.com/psych252spotify25>

01/16/2025

Logistics

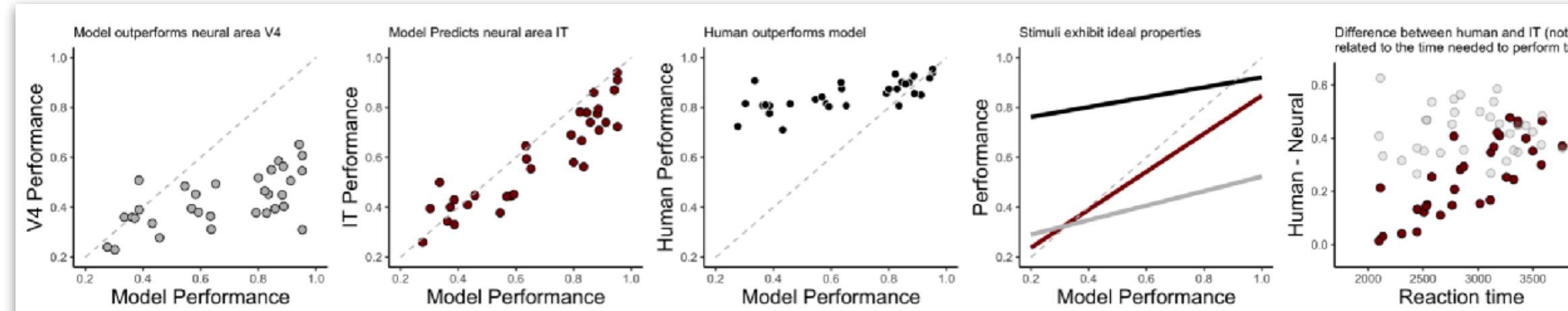
Homework 2



cool experiment

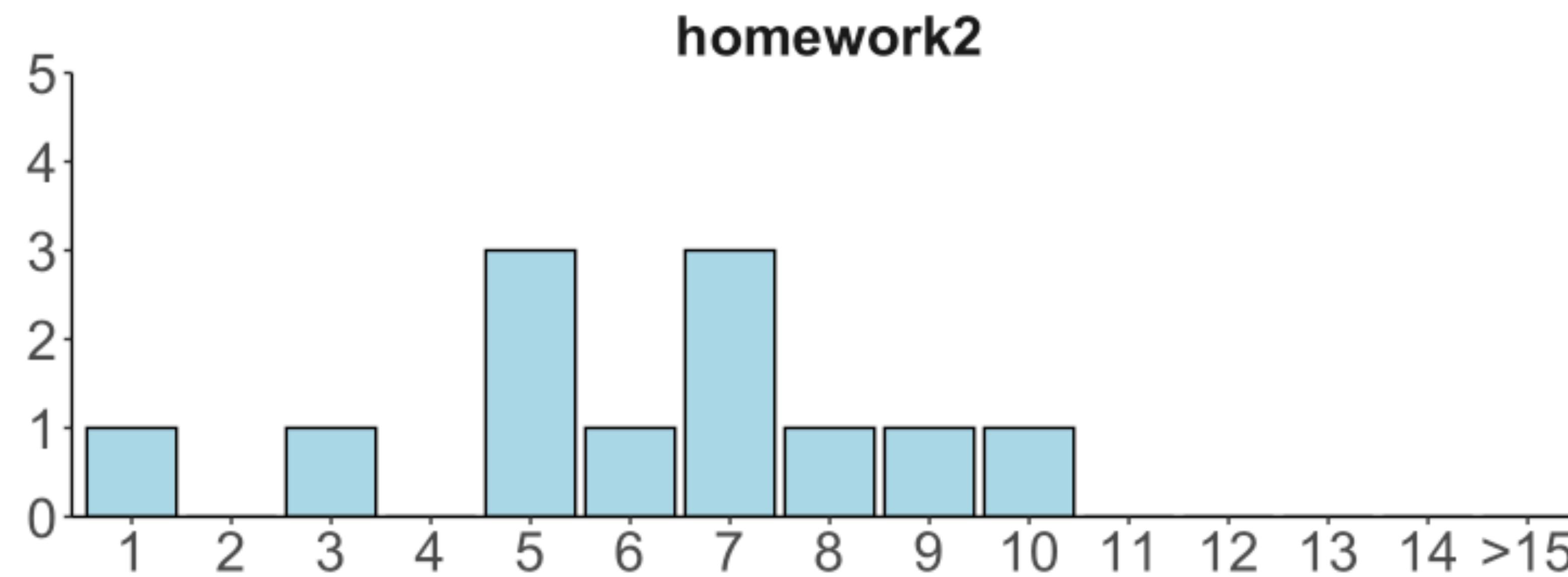
NAME	Relationship of each person whose place of residence on April 1, 1880, was in this family and connection with the head of the family	SEX	SOME DATA		FAMILY DESCRIPTION	EDUCATION	PLACE
			AGE	IN MONTHS			
Turner, Lee	Son	M	10	120	Black hair, blue eyes, 5 ft 7 in., 140 lbs.	Can't read	East
Mabel	Daughter	F	8	96	Black hair, blue eyes, 5 ft 2 in., 110 lbs.	Can't read	East
George C.	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East
Lena	Daughter	F	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
John	Son	M	14	168	Black hair, brown eyes, 5 ft 7 in., 160 lbs.	Can't read	East
Catharine	Daughter	F	12	144	Black hair, brown eyes, 5 ft 2 in., 120 lbs.	Can't read	East
Coker, George E.	Son	M	14	168	Black hair, brown eyes, 5 ft 7 in., 160 lbs.	Can't read	East
Harriet	Daughter	F	12	144	Black hair, brown eyes, 5 ft 2 in., 120 lbs.	Can't read	East
Hodge	Sister	F	11	132	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
Samuel	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East
Henry	Daughter	F	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
John	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East
Albert	Daughter	F	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
Ezra	Son	M	14	168	Black hair, brown eyes, 5 ft 7 in., 160 lbs.	Can't read	East
Lizzie	Daughter	F	12	144	Black hair, brown eyes, 5 ft 2 in., 120 lbs.	Can't read	East
George	Son	M	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
John	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East
Mary	Daughter	F	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
William	Son	M	14	168	Black hair, brown eyes, 5 ft 7 in., 160 lbs.	Can't read	East
John W.	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East
Edward	Son	M	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
Albert	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East
George	Son	M	10	120	Black hair, brown eyes, 5 ft 2 in., 110 lbs.	Can't read	East
Henry	Son	M	12	144	Black hair, brown eyes, 5 ft 7 in., 150 lbs.	Can't read	East

messy dataset



neat plot!

Homework 2



how long it took people last year

Homework 2

- Due **Thursday 25th, at 8pm**
- Don't wait until the very last moment to knit your RMarkdown file into a pdf. It may not compile and debugging takes time ...
- You can upload earlier versions of your homework on Canvas and still update until the deadline.
- Get and give help via Ed Discussion!

Homework 2

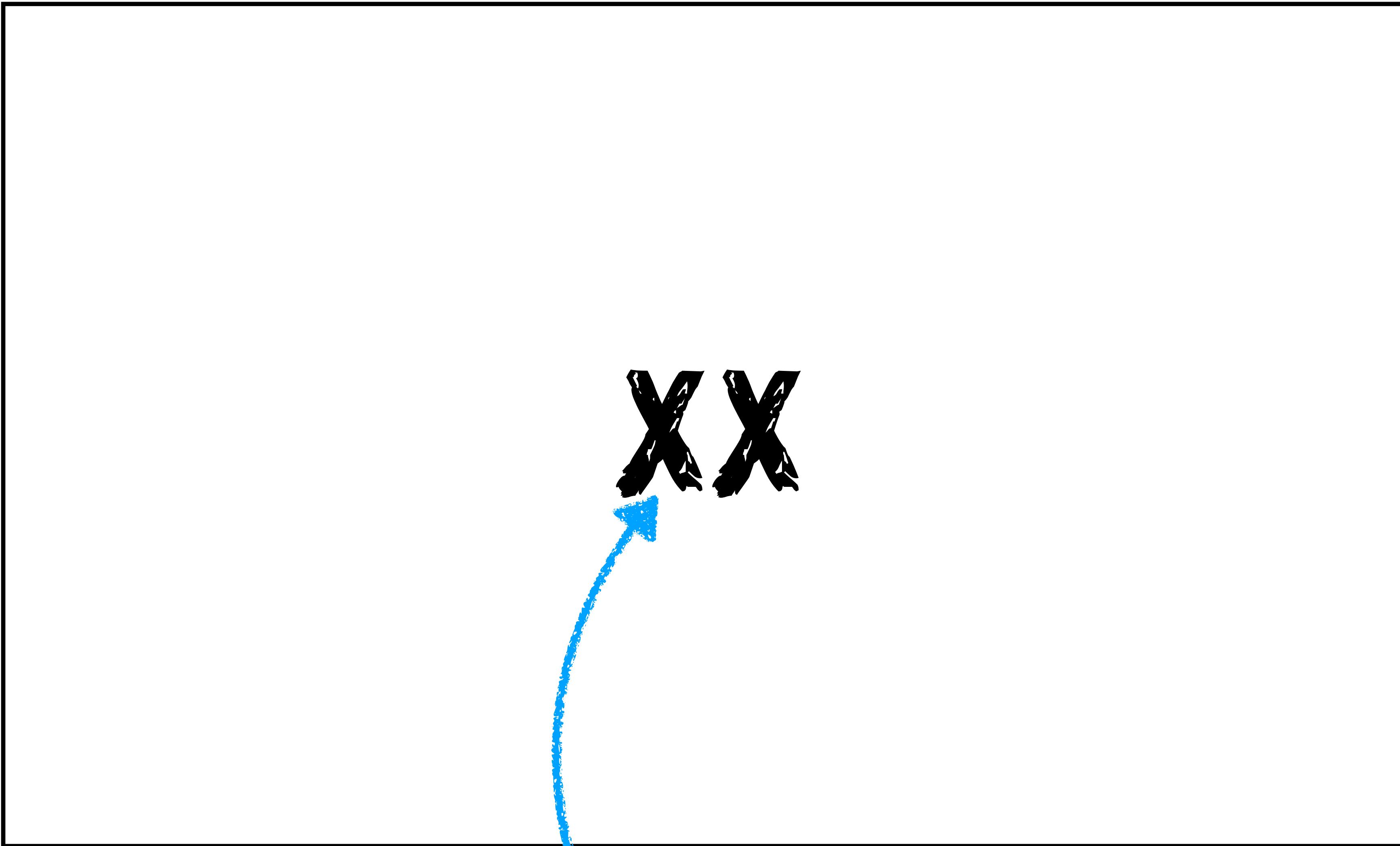
- **We encourage you to work in groups!**
- It's more fun
- You learn more (learning through explaining)

Homework 2

- You can adjust the figure size in your output by using the code chunk options



It's a mystery

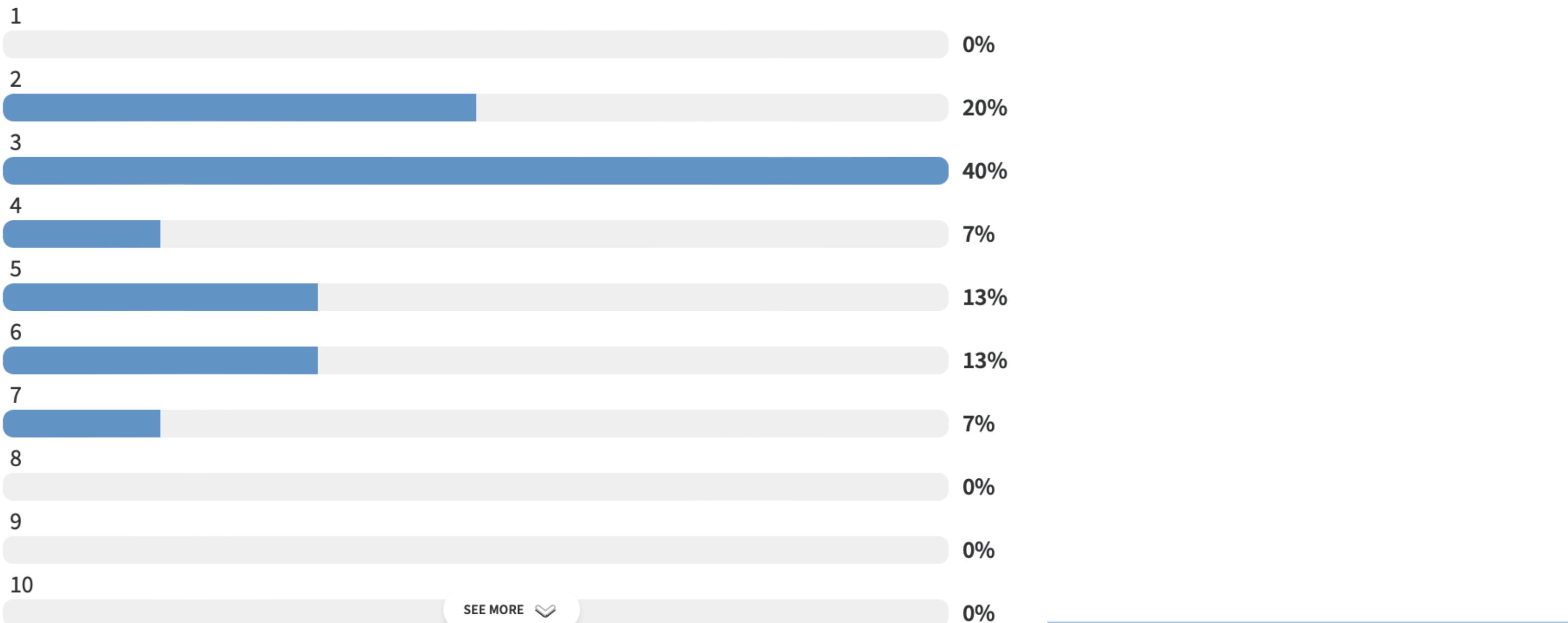


when I want to hide stuff from you

Your feedback



How many hours did it take you to complete Homework 1?

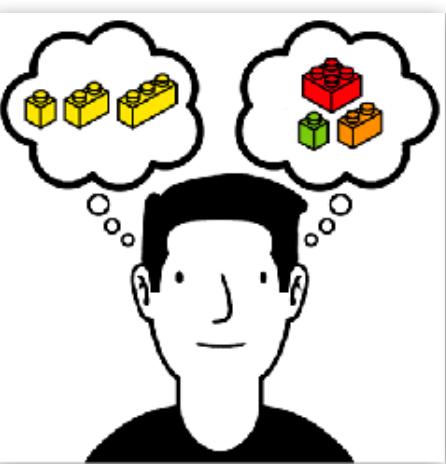


Outline

- Introduction to probability / Recap
 - Motivation
 - Counting possibilities
 - **Clue** guide to probability
 - Understanding Bayes' Rule
 - Getting Bayes' right matters!
 - Building a Bayesis

Motivation

What does statistics have to do with probability?



Theory

Our goal is to develop theories. In psychology, theories of how the mind works.



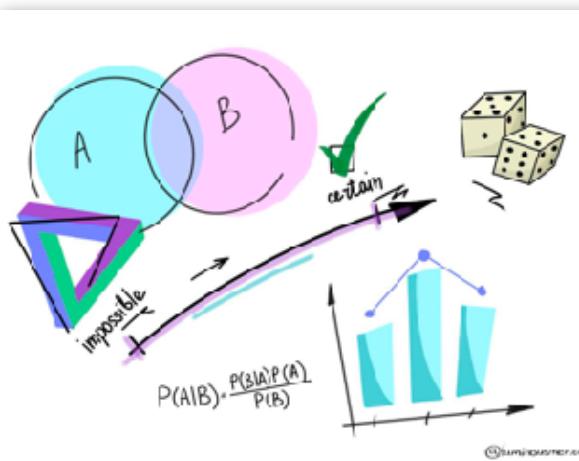
Prediction

Our theories need to make testable/falsifiable predictions.



Uncertainty

Because the domains that we are interested in are fundamentally uncertain (e.g. we want to say something about people generally but can only test a sample), we formulate and test these predictions using statistical models.

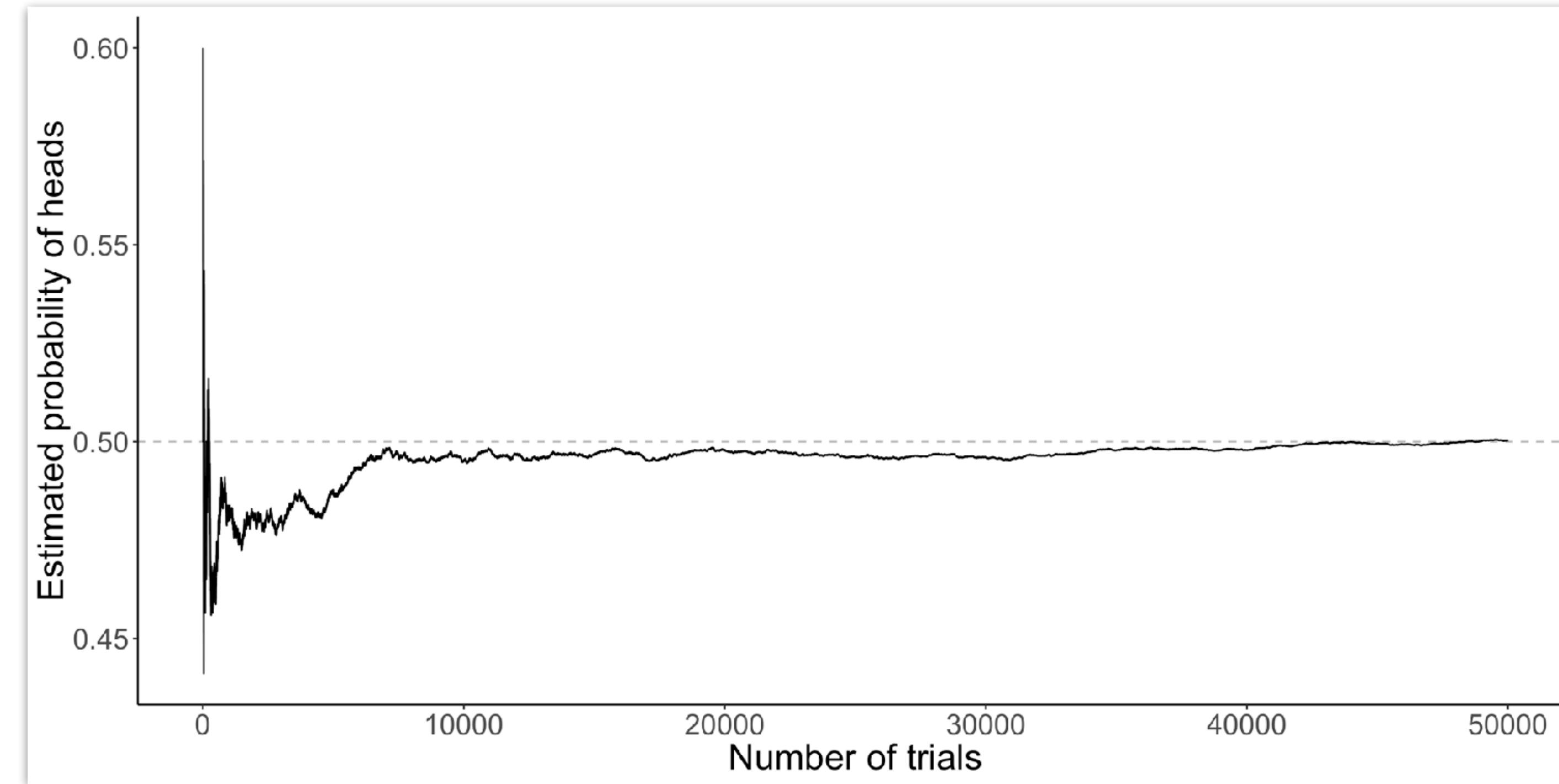


Probability

Probability theory is the formal language for dealing with uncertainty.

Frequentist interpretation

Probabilities = **long-range frequencies**



law of large numbers = empirical probability will
approximate the true probability as
the sample size increases

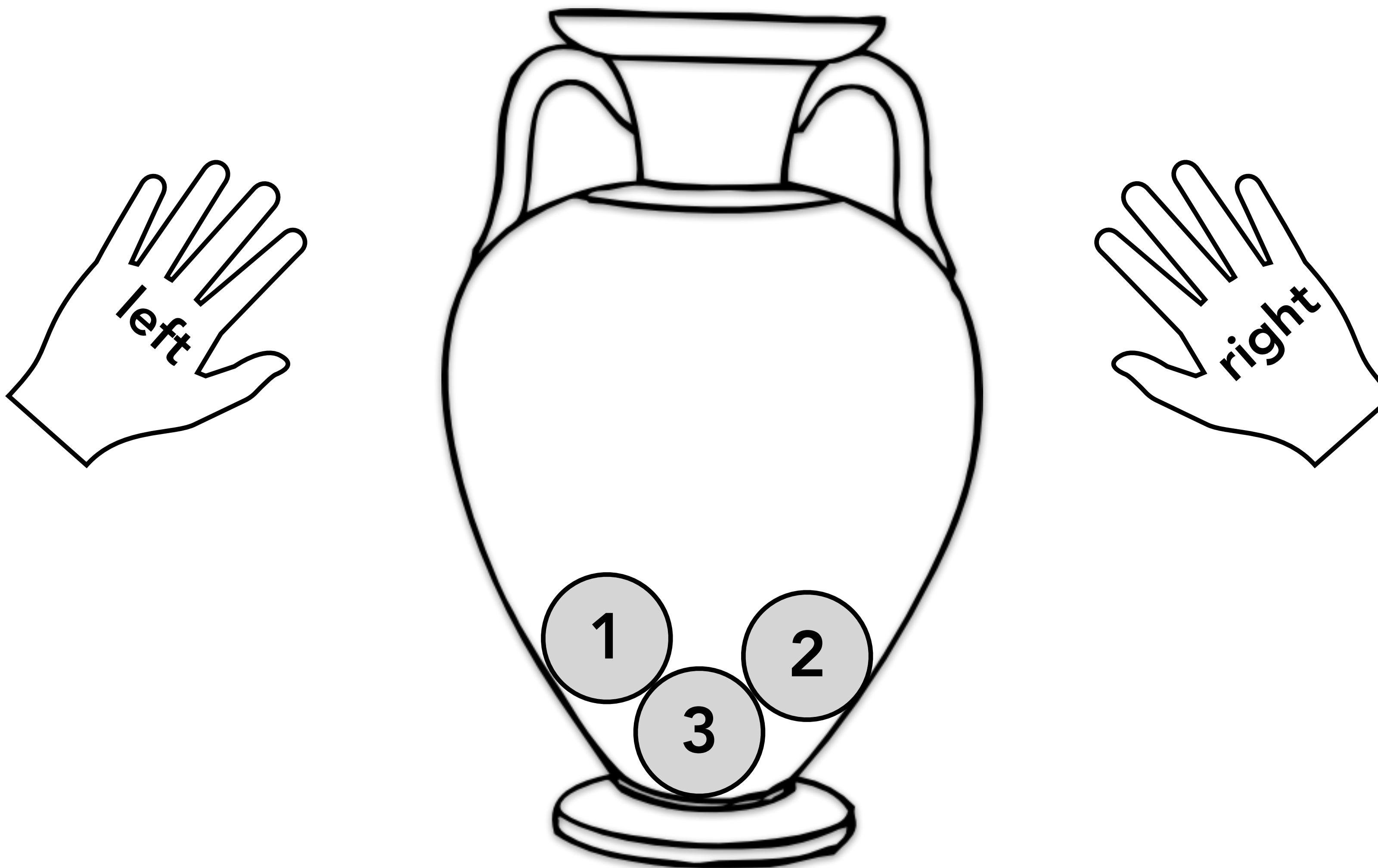
Subjective/Bayesian interpretation

Probabilities = **subjective degrees of belief**

- applies to events which may only happen once
- "**What's the probability that humans will land on Mars someday?**"
- probabilities are not a property of the world, but of a person's beliefs about the world
- at the heart of Bayesian data analysis

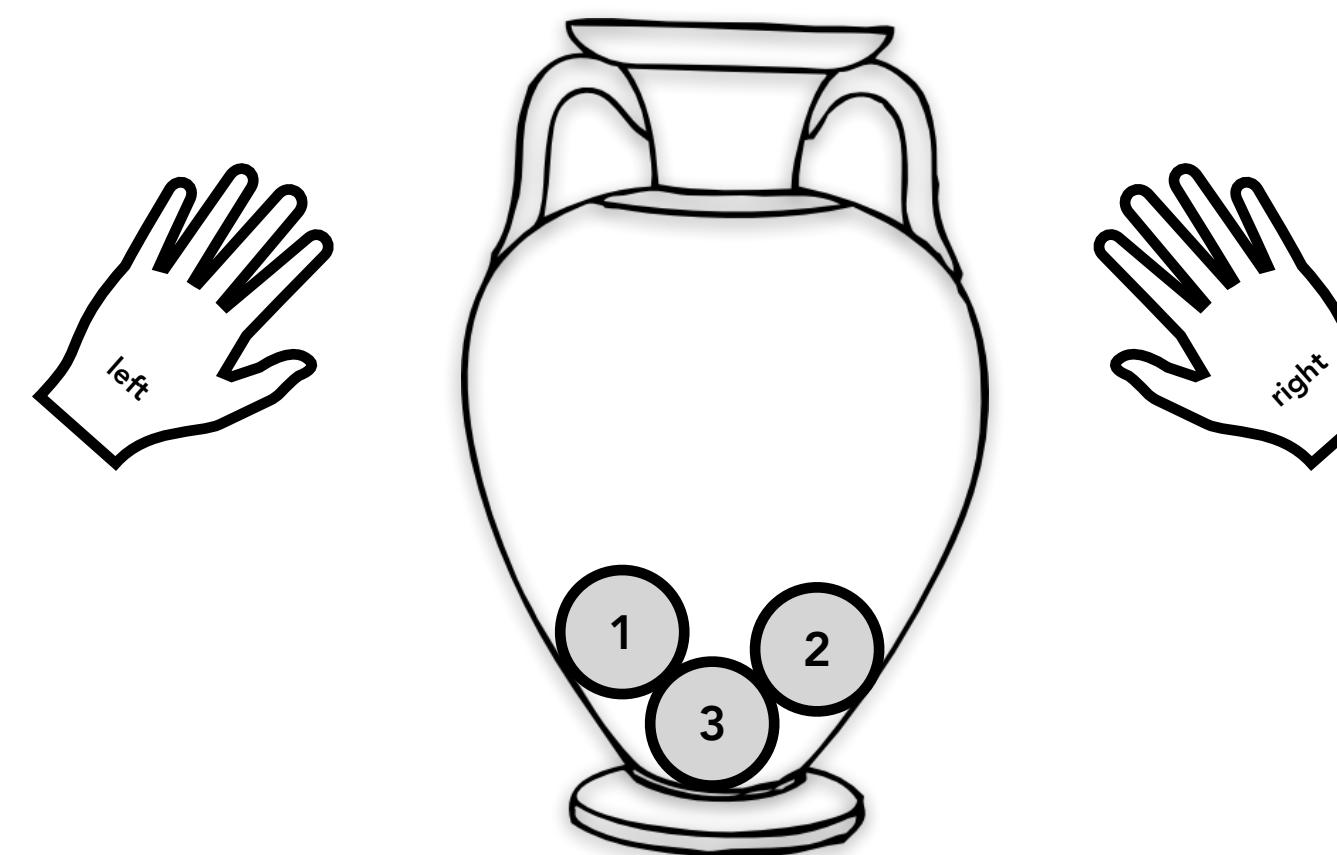
Counting possibilities

no stats class without urns!

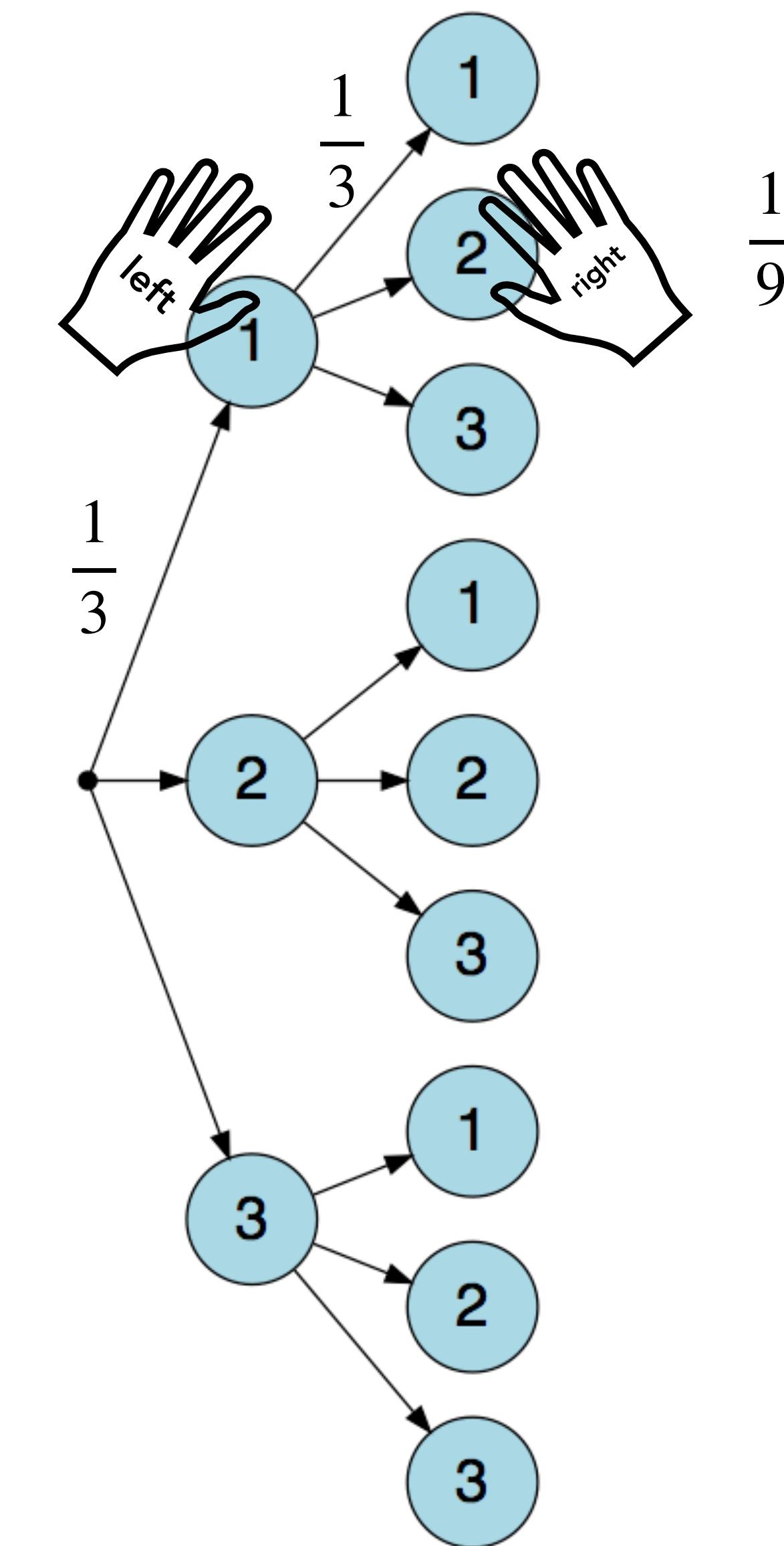


Sampling with replacement

$$p(\text{left} = 1, \text{right} = 2) = ?$$

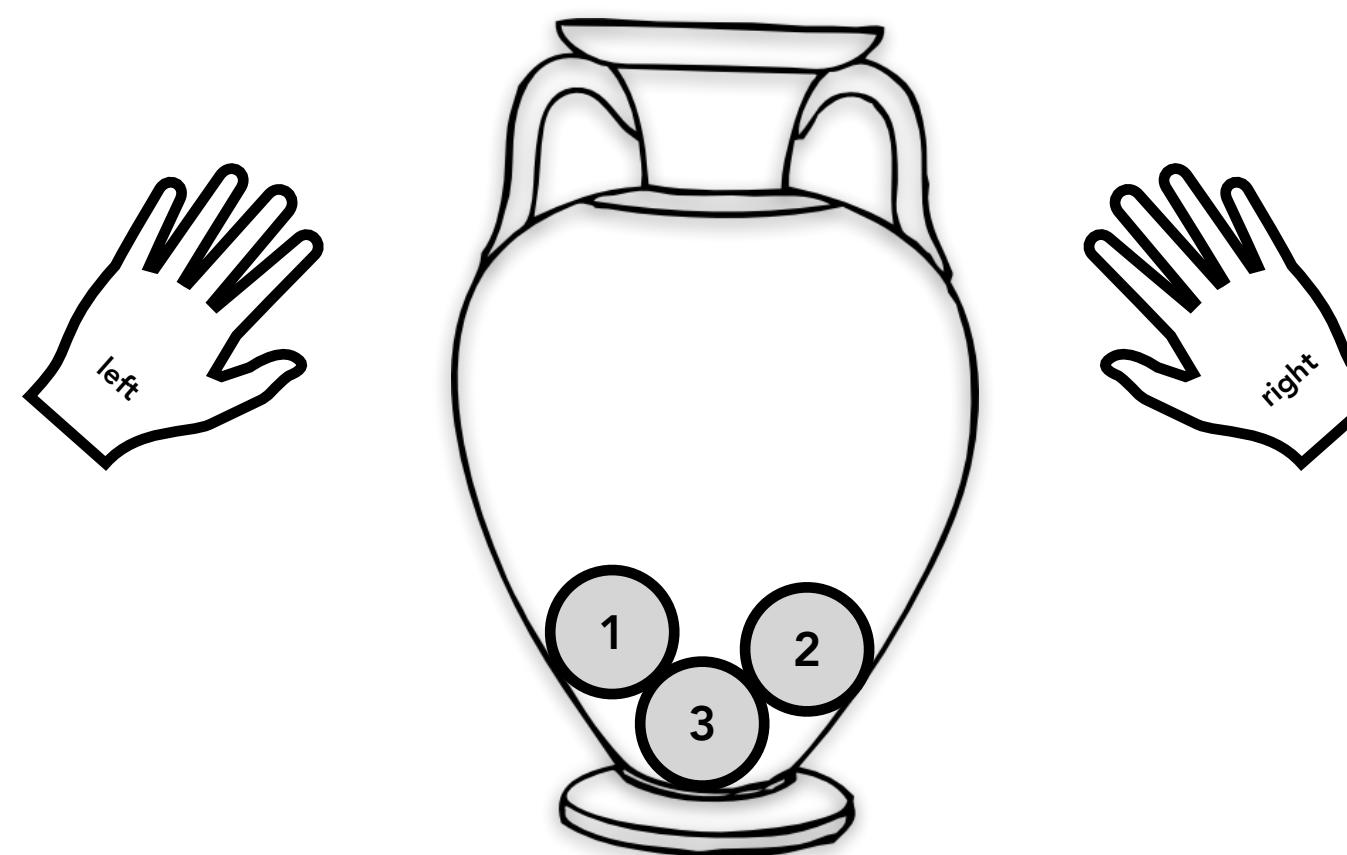


What is the probability that I first draw the 1 with my left hand, and then, after putting the 1 back into the urn again, draw the 2 with my right hand?

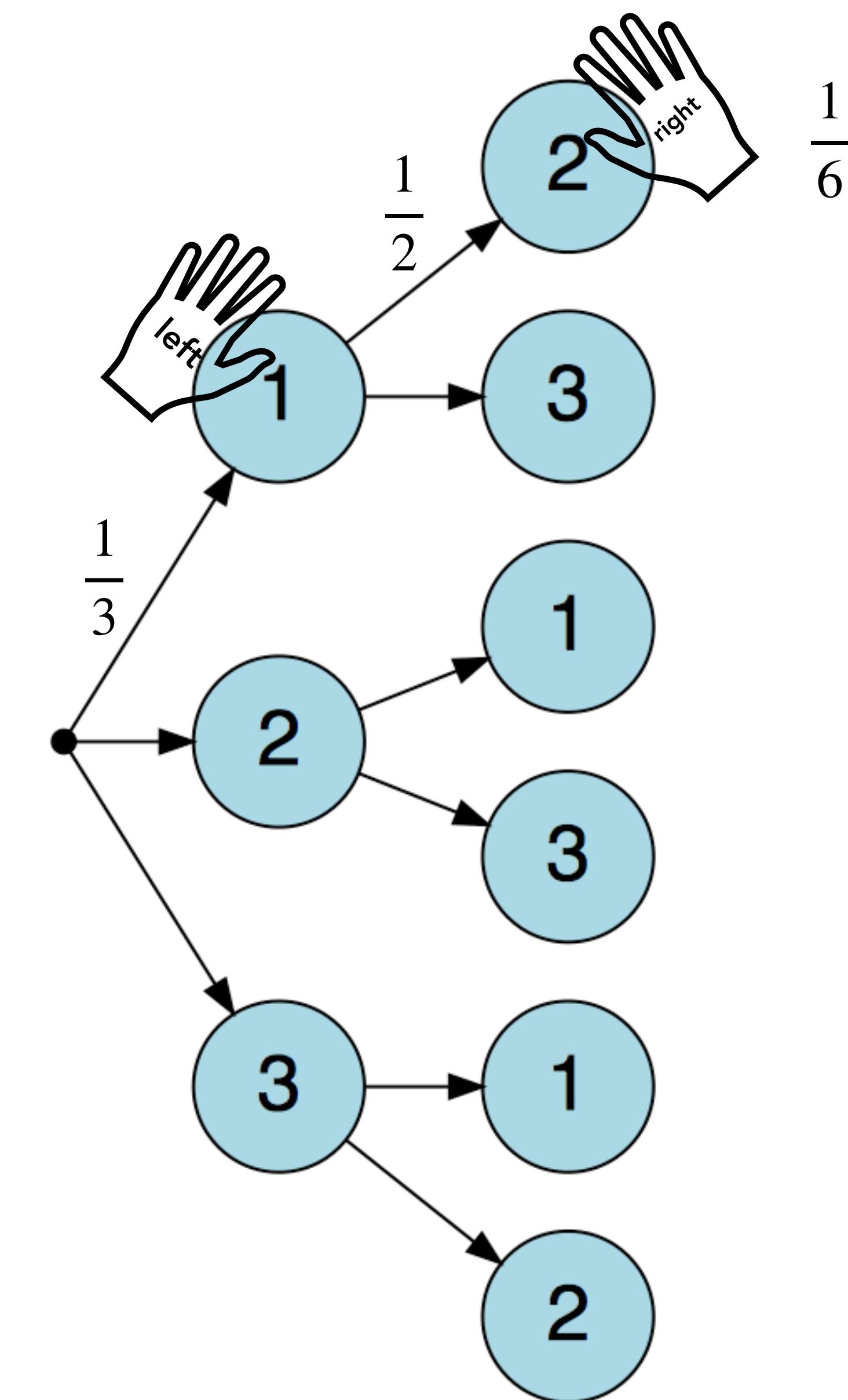


Sampling without replacement

$$p(\text{left} = 1, \text{right} = 2) = ?$$



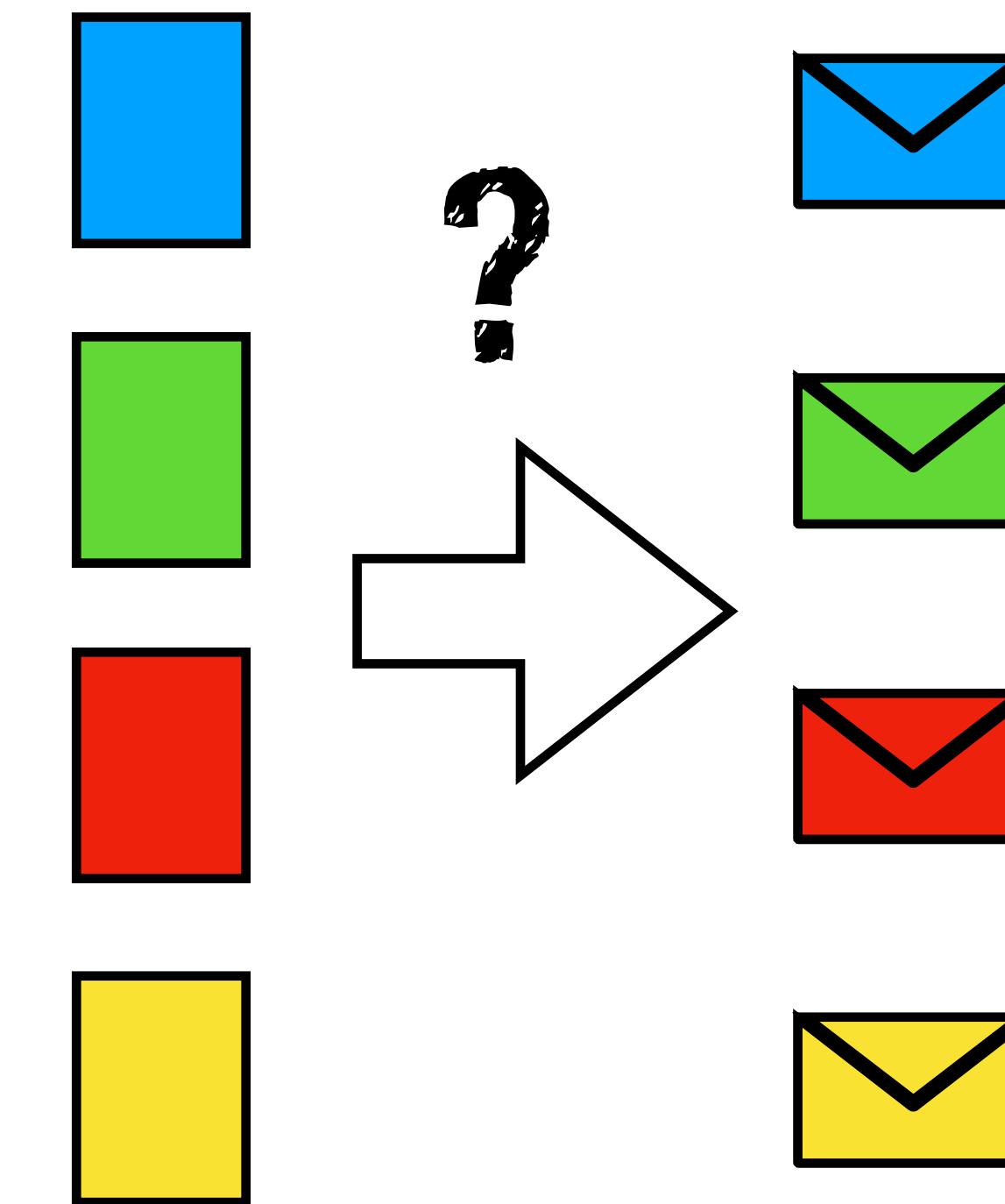
What is the probability that I first draw the 1 with my left hand, and then, without putting the 1 back into the urn, draw the 2 with my right hand?



Random secretary



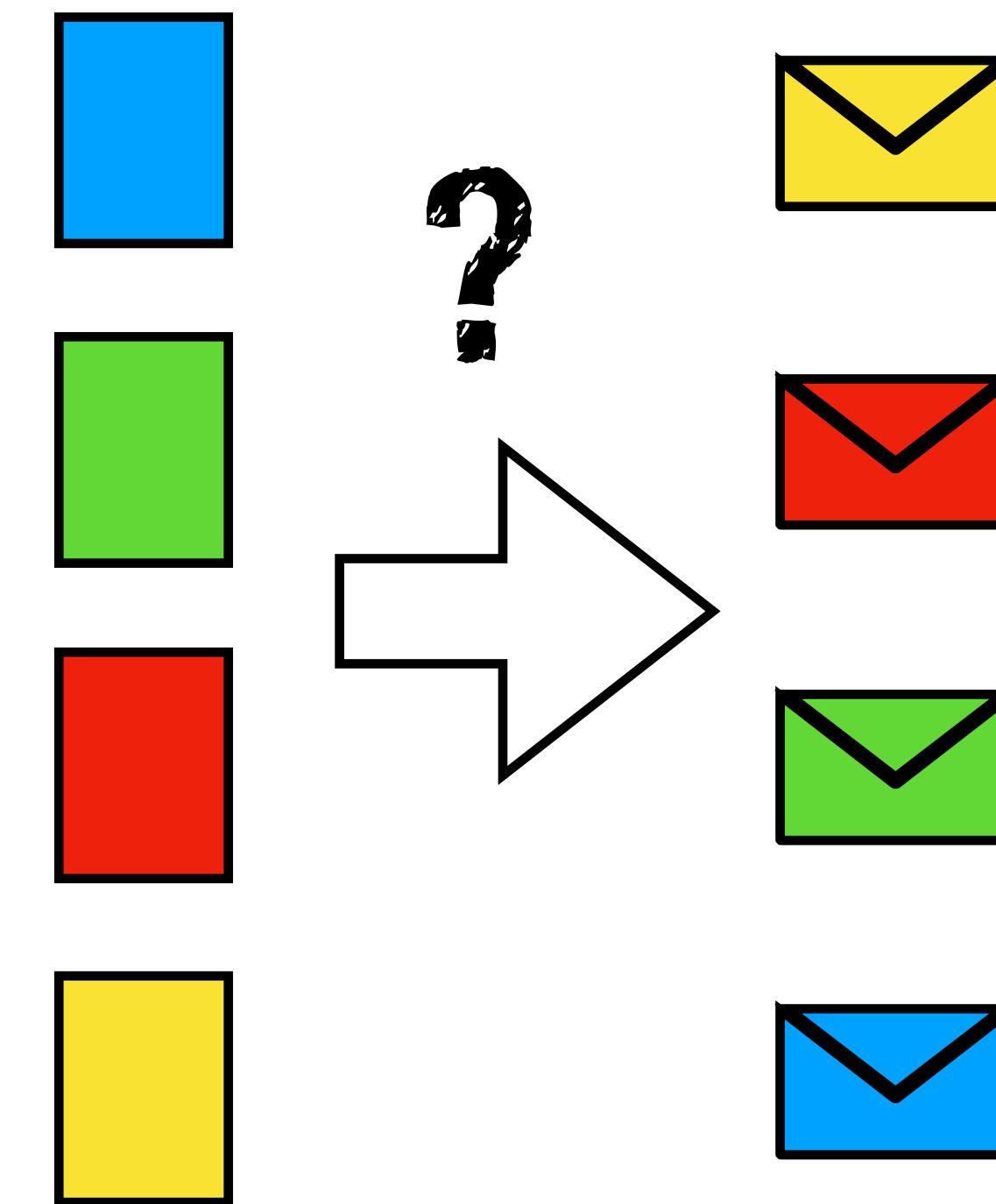
A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?



Random secretary



A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?

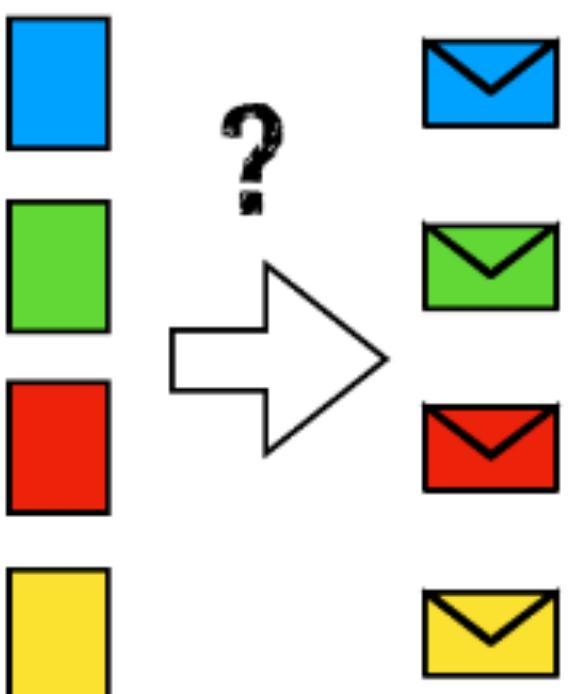


What is the probability that exactly three letters will go into the right envelope?

Random secretary



A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?



0% 25% 50% 75% 100%

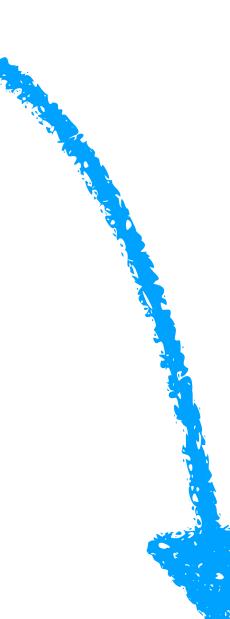
Random secretary

```
1 df.letters = permutations(x = 1:4, k = 4) %>%
2   as_tibble(.name_repair = ~ str_c("person_", 1:4)) %>%
3   mutate(n_correct = (person_1 == 1) +
4         (person_2 == 2) +
5         (person_3 == 3) +
6         (person_4 == 4))
7
8 df.letters %>%
9   summarize(prob_3_correct = sum(n_correct == 3) / n())
```

person_1	person_2	person_3	person_4
1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4

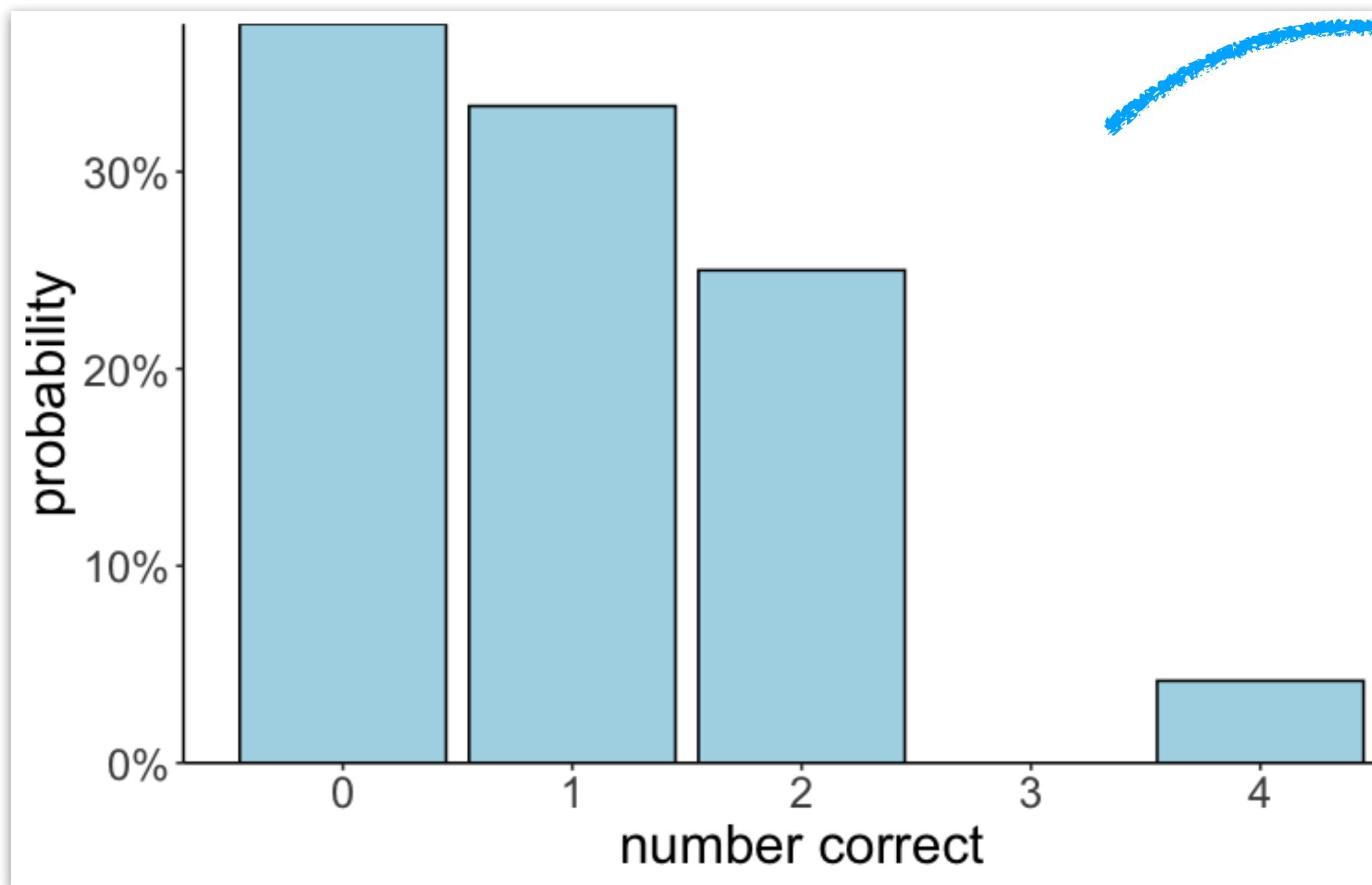
• 24 rows total

prob_3_correct
0



Random secretary

```
1 ggplot(data = df.letters,
2         mapping = aes(x = n_correct)) +
3   geom_bar(aes(y = stat(count)/sum(count)),
4           color = "black",
5           fill = "lightblue") +
6   scale_y_continuous(labels = scales::percent,
7                       expand = c(0, 0)) +
8   labs(x = "number correct",
9        y = "probability")
```



probability of getting
0, 1, 2, 3, or 4
envelopes to the
correct person

Naive definition of probability

$$P_{\text{naive}}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of outcomes}}$$

if all outcomes are equally likely!

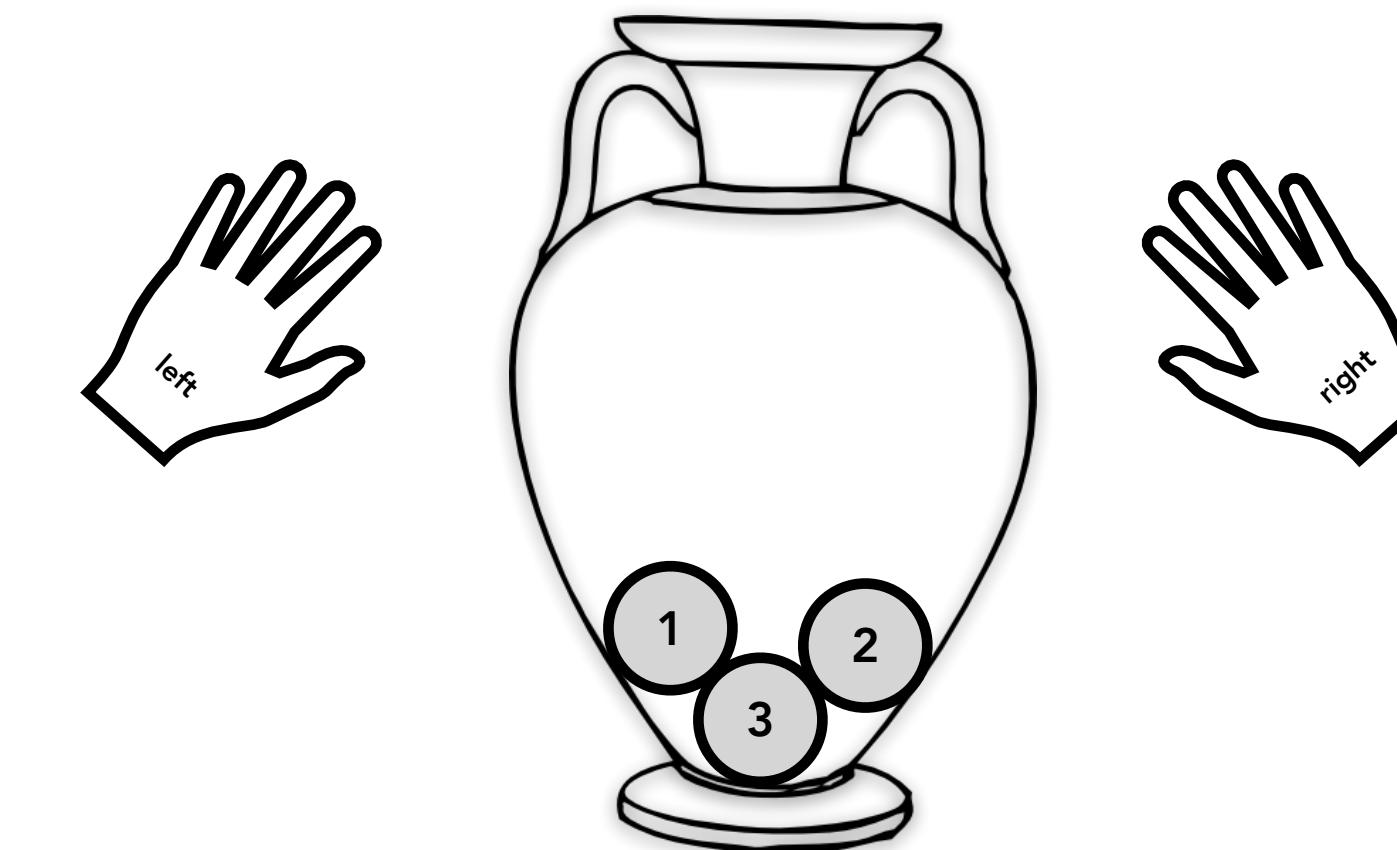
Definitions

Experiment: Activity that produces or observes an outcome.

Drawing 2 marbles from the urn with replacement, and noting the order.

Sample Space: Set of possible outcomes for an experiment.

$$\Omega = \{(1, 1), (1, 2), \dots, (2, 1), \dots, (3, 3)\}$$



Event: Subset of the sample space.

$$(1, 1)$$

Definitions

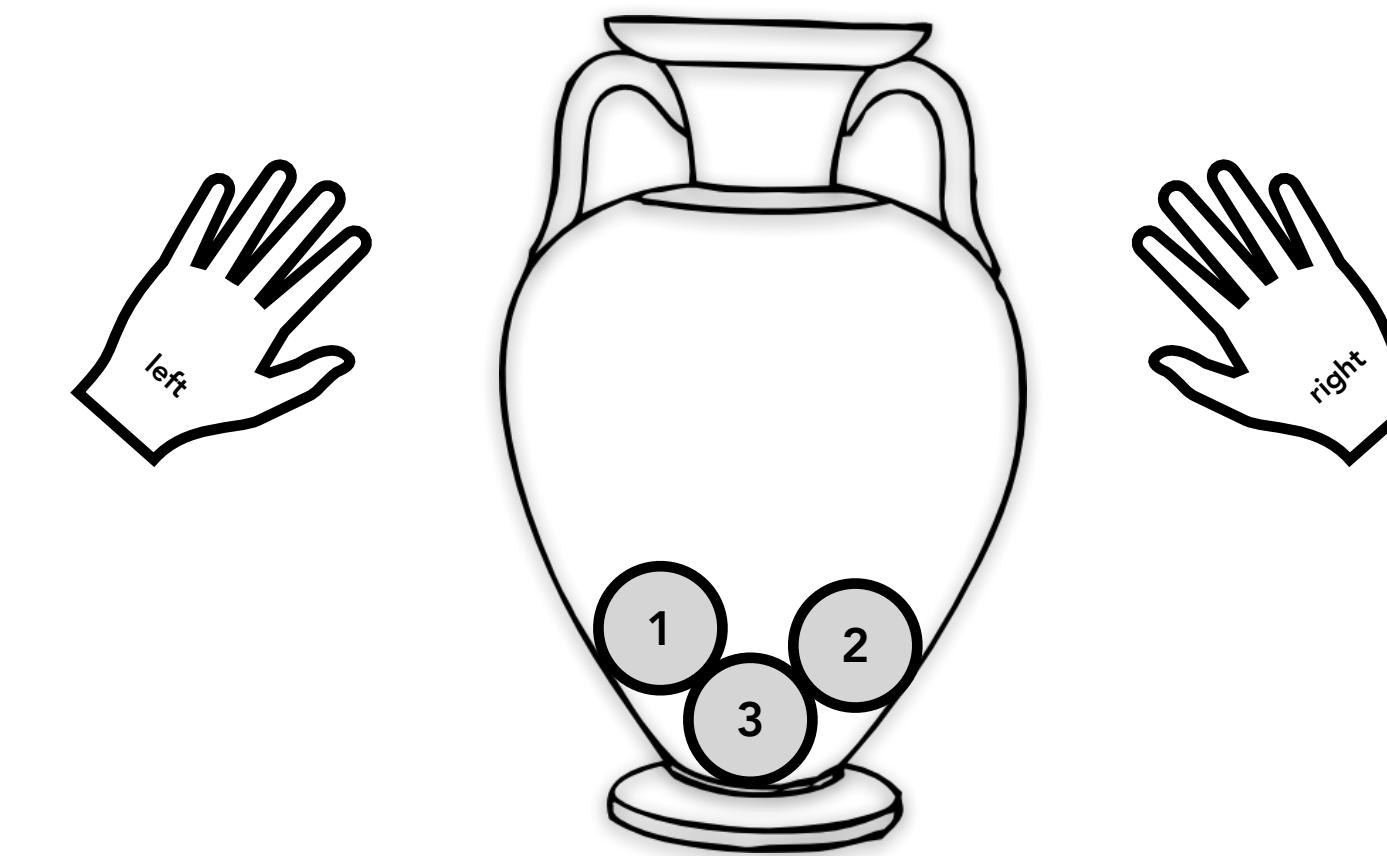
If $P(X_i)$ is the probability of event X_i

1. Probability cannot be negative.

$$P(X_i) \geq 0$$

2. Total probability of all outcomes in the sample space is 1.

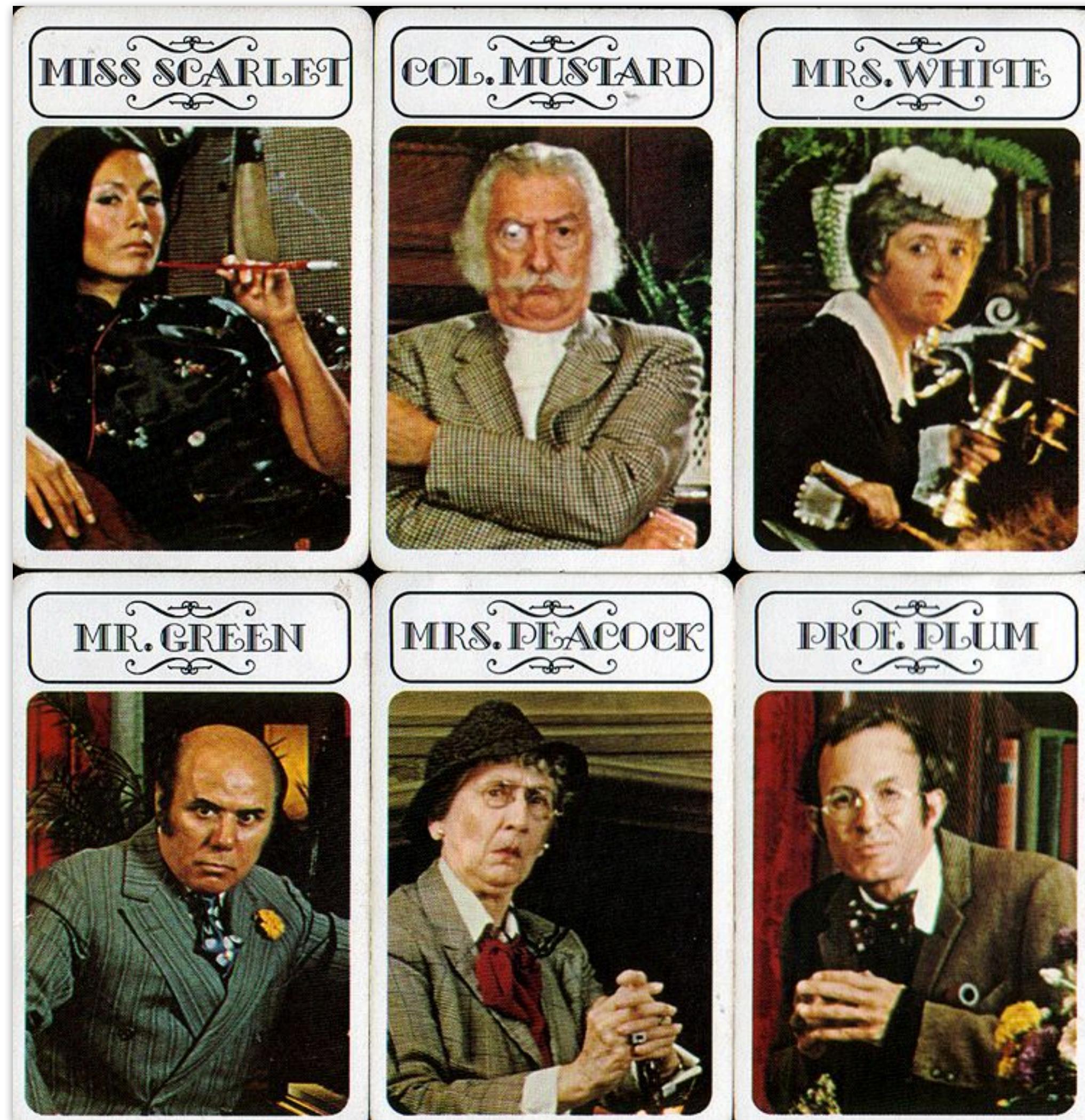
$$\sum_{i=1}^N P(X_i) = P(X_1) + P(X_2) + \dots + P(X_N) = 1$$



Clue guide to probability

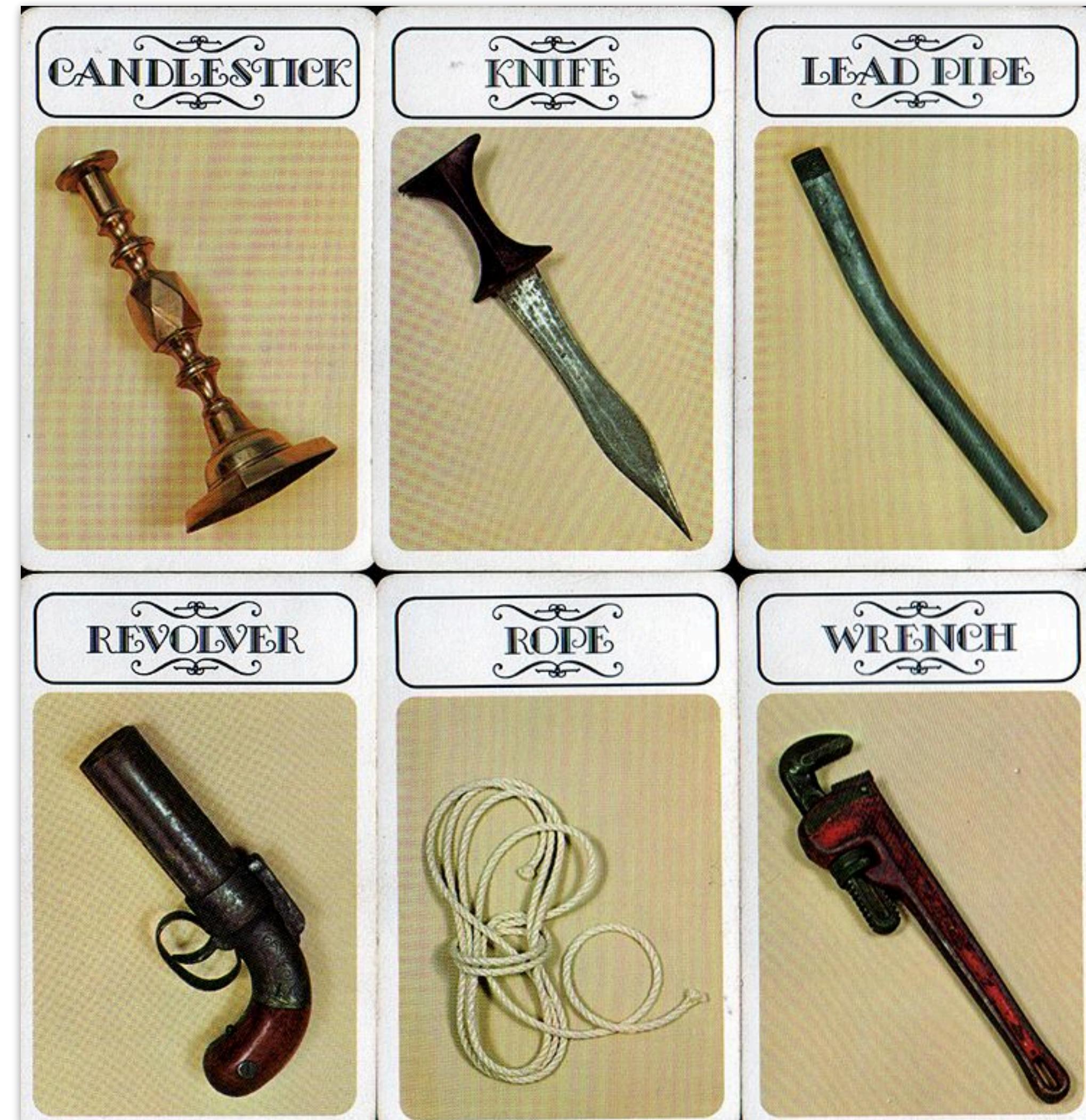
Clue guide to probability

Who killed Mr Boddy?



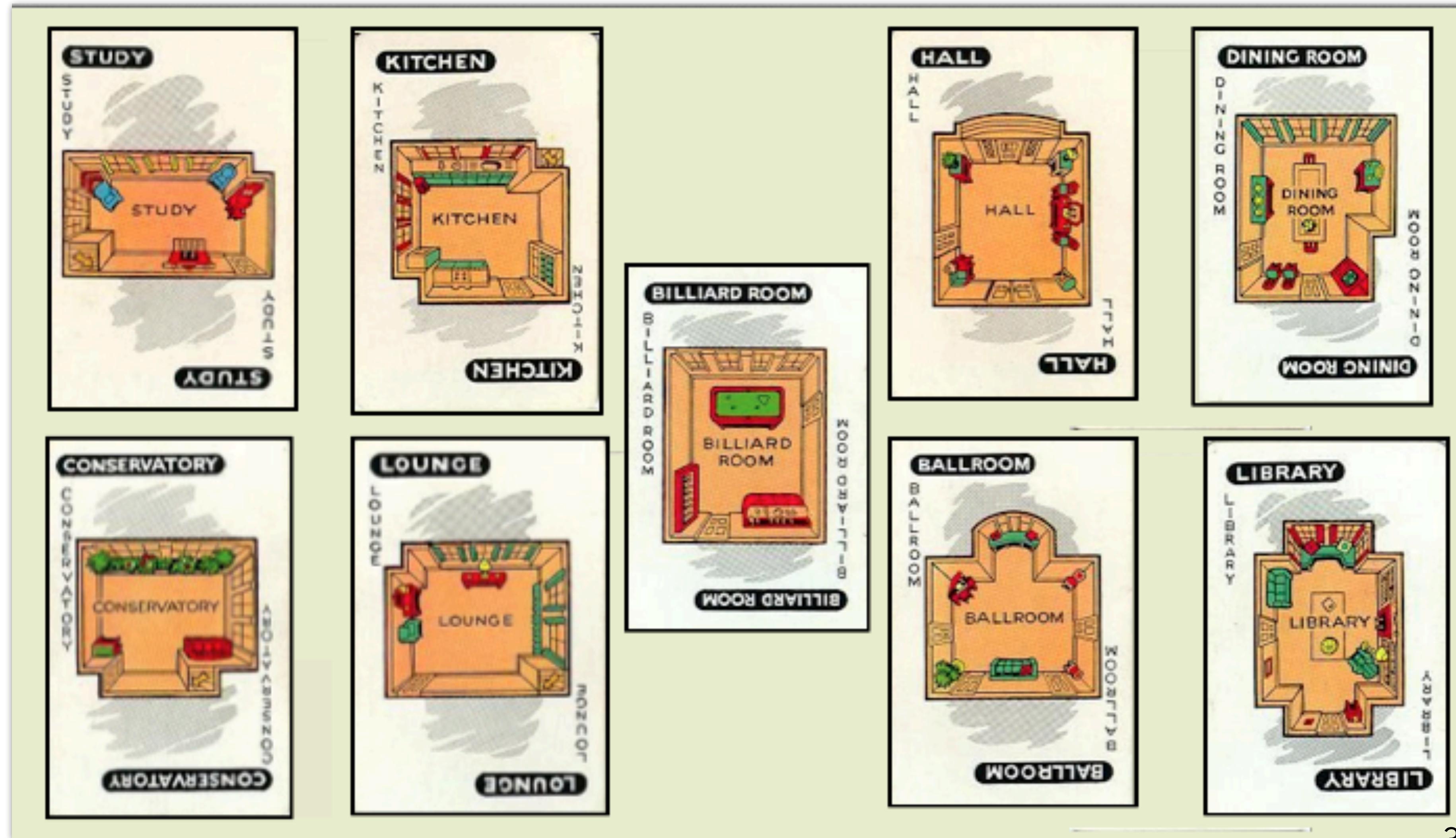
Clue guide to probability

Who killed Mr Boddy, **with what?**

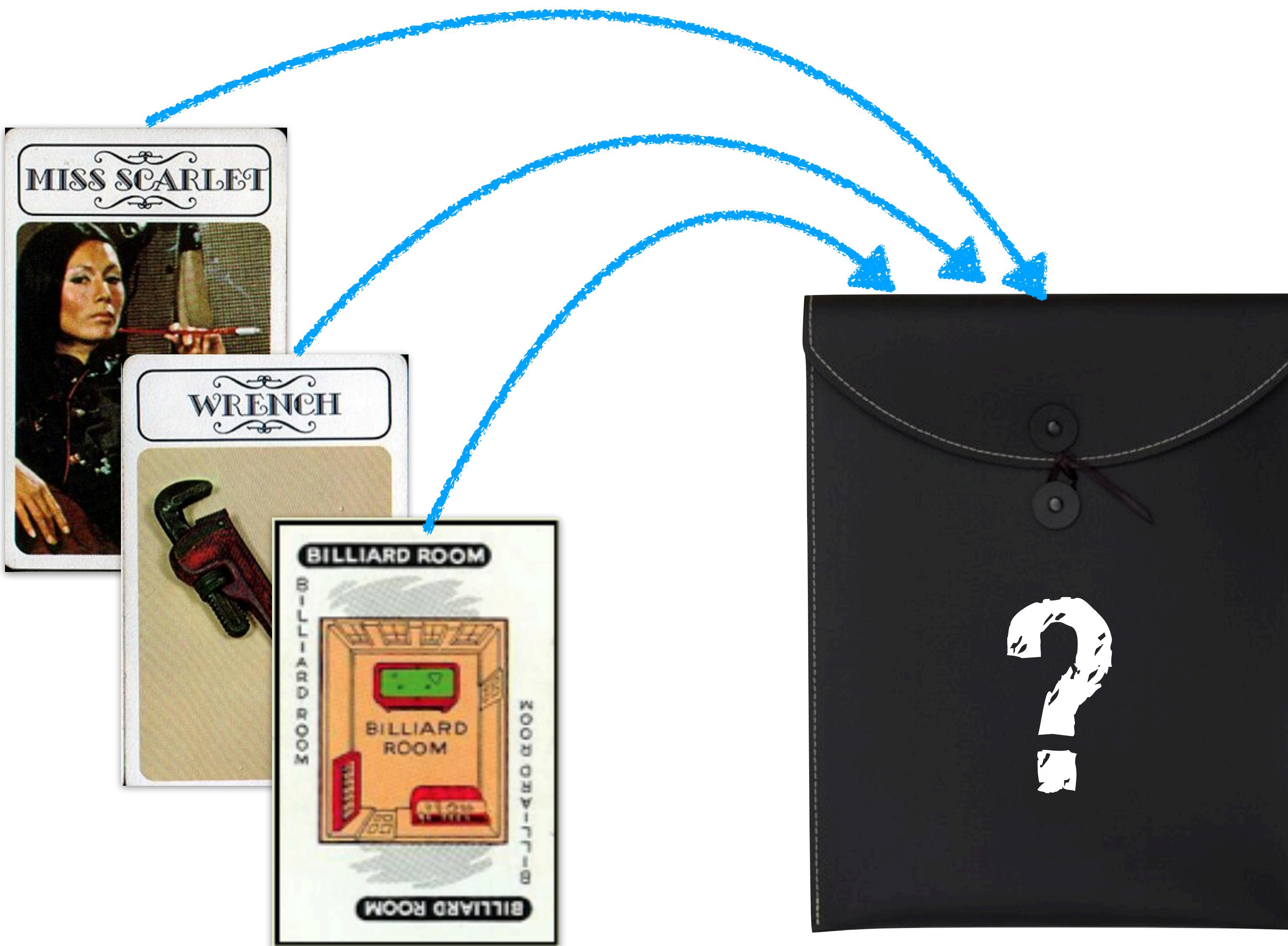


Clue guide to probability

Who killed Mr Boddy, with what, **and where?**



Clue guide to probability



Clue guide to probability

```
1 who = c("ms_scarlet", "col_mustard", "mrs_white",
2       "mr_green", "mrs_peacock", "prof_plum")
3 what = c("candlestick", "knife", "lead_pipe",
4        "revolver", "rope", "wrench")
5 where = c("study", "kitchen", "conservatory",
6           "lounge", "billiard_room", "hall",
7           "dining_room", "ballroom", "library")
8
9 df.clue = expand_grid(who = who,
10                      what = what,
11                      where = where)
```

all combinations

Ω

who	what	where
ms_scarlet	candlestick	study
ms_scarlet	candlestick	kitchen
ms_scarlet	candlestick	conservatory
ms_scarlet	candlestick	lounge
⋮		

`nrow(df.clue) = 324`

Clue guide to probability

Who?

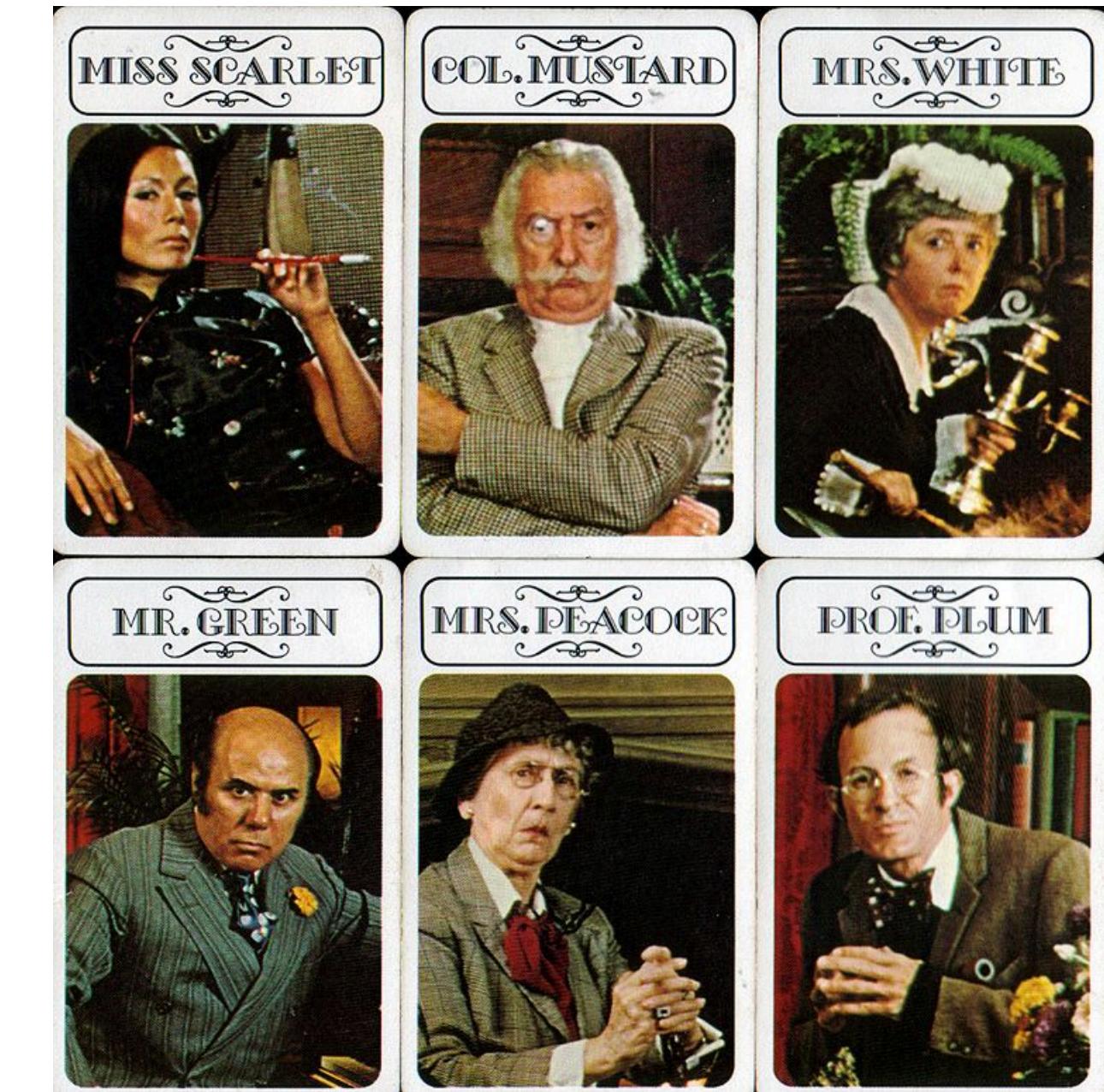
- 6 suspects
- mutually exclusive and exhaustive
- $p(\text{who} = \text{one of the six}) = 1$
- each equally likely a priori
- $p(\text{who} = \text{Prof. Plum}) = \frac{1}{6}$



Clue guide to probability

Who?

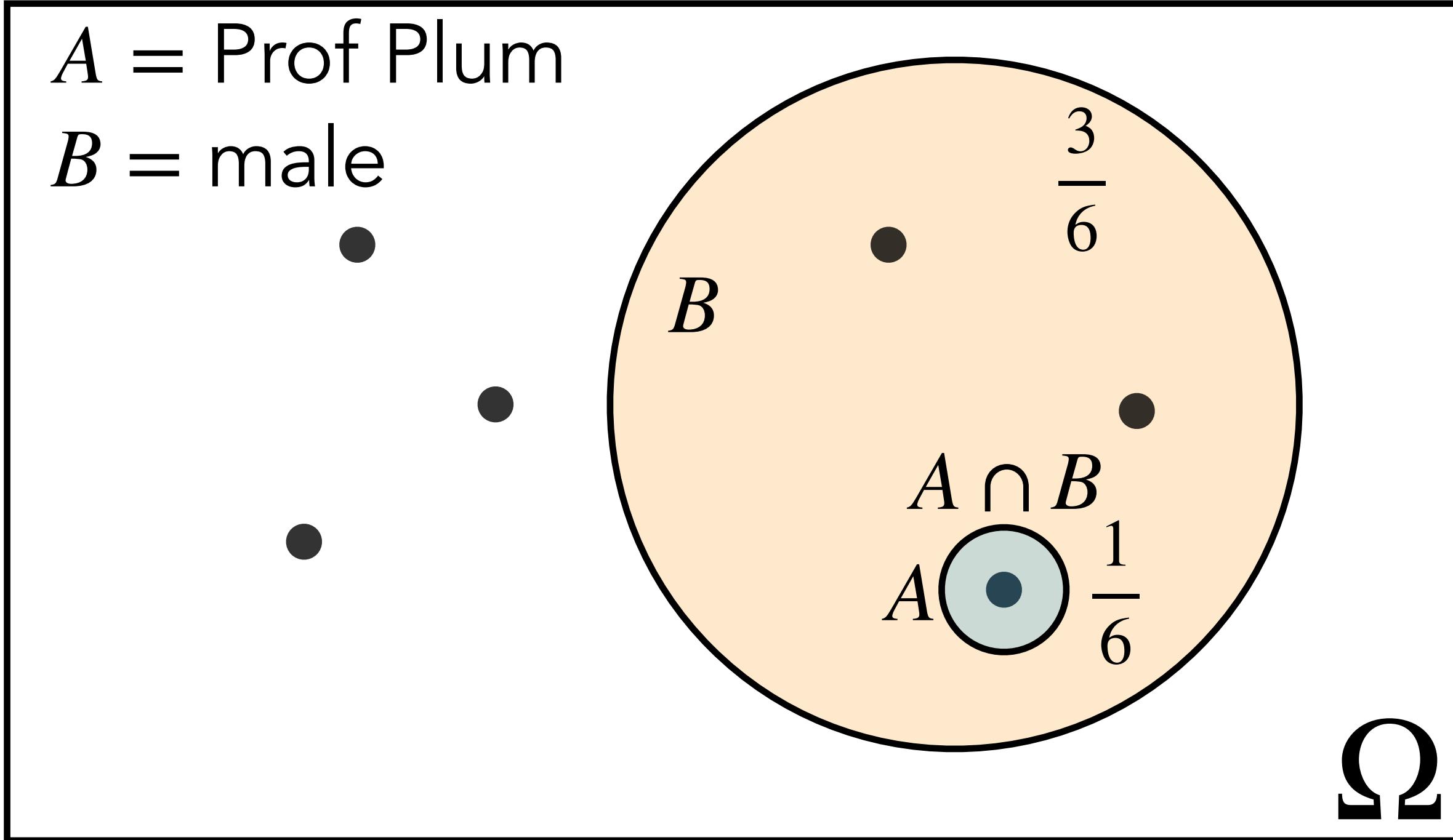
- **conditional probability:**
- $p(A | B)$ (probability of A given B)
- **Definition:** $p(A | B) = \frac{p(A, B)}{p(B)}$



Probability that it was Prof Plum, given that the murderer was male?

$$p(\text{Prof. Plum} | \text{male}) = ?$$

Clue guide to probability



Probability that it was Prof Plum, given that the murderer was male?

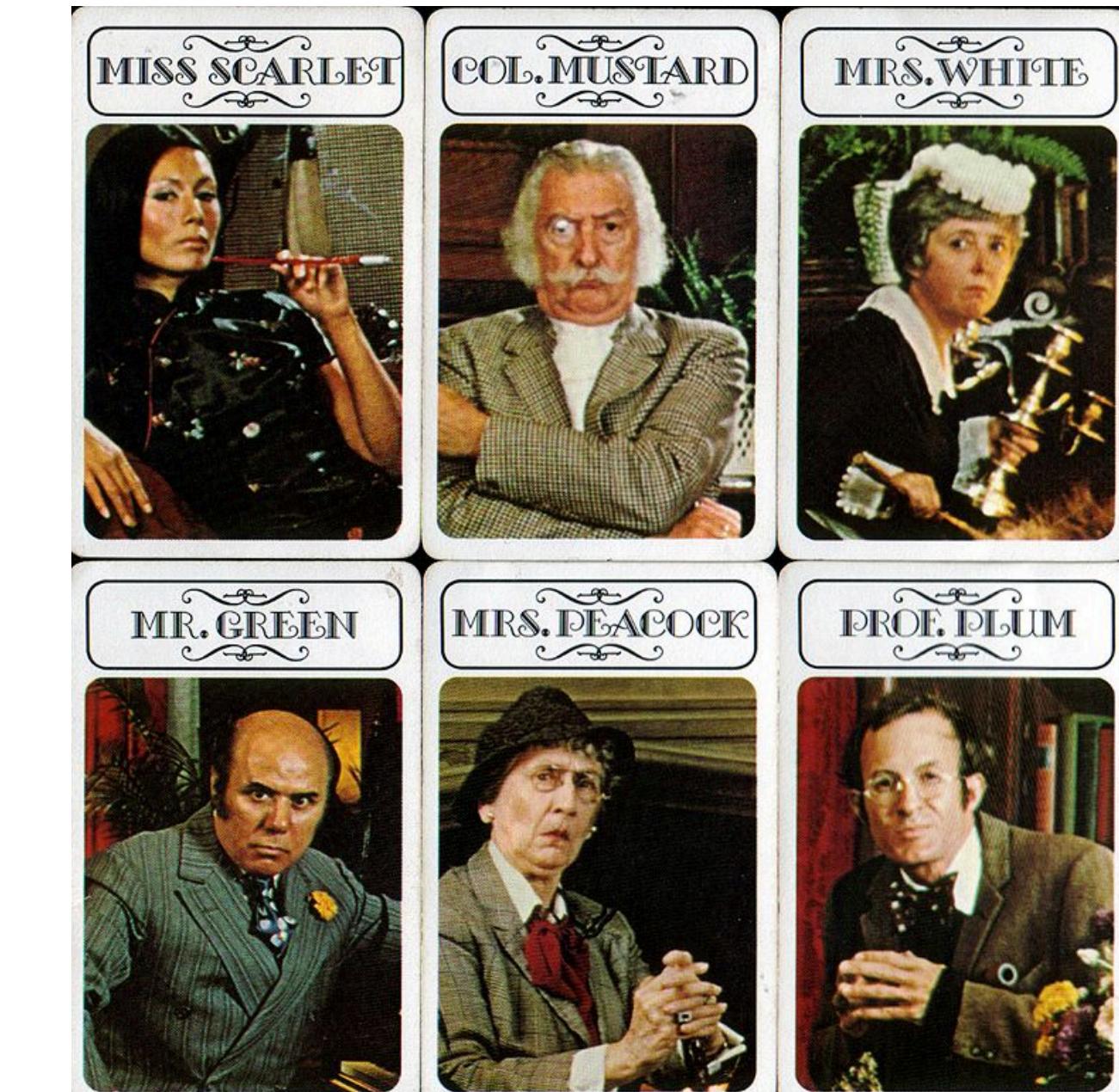
Definition: $p(A | B) = \frac{p(A, B)}{p(B)} = \frac{1}{3}$

$$p(A) = \frac{1}{6} \quad p(A, B) = \frac{1}{6} \quad p(B) = \frac{3}{6}$$

Clue guide to probability

Who?

- **conditional probability:**
- $p(A | B)$ (probability of A given B)
- **Definition:** $p(A | B) = \frac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} | \text{male}) = \frac{1/6}{1/2} = 1/3$



who	gender	
col_mustard	male	spects = df.clue %>% tinct(who) %>%
mr_green	male	ate(gender = ifelse(
prof_plum	male	est = who %in% c("ms_scarlet",
ms_scarlet	female	"mrs_white",
mrs_white	female	"mrs_peacock") ,
mrs_peacock	female	es = "female",
) = "male"))

Clue guide to probability

Who?

- **conditional probability:**
- $p(A | B)$ (probability of A given B)
- **Definition:** $p(A | B) = \frac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} | \text{male}) = \frac{1/6}{1/2} = 1/3$



who	gender
col_mustard	male
mr_green	male
prof_plum	male
ms_scarlet	female
mrs_white	female
mrs_peacock	female

```

1 df.suspects %>%
2   summarize(p_prof_plum_given_male =
3     sum(gender == "male" &
4       == "prof_plum") /
5     sum(gender == "male"))

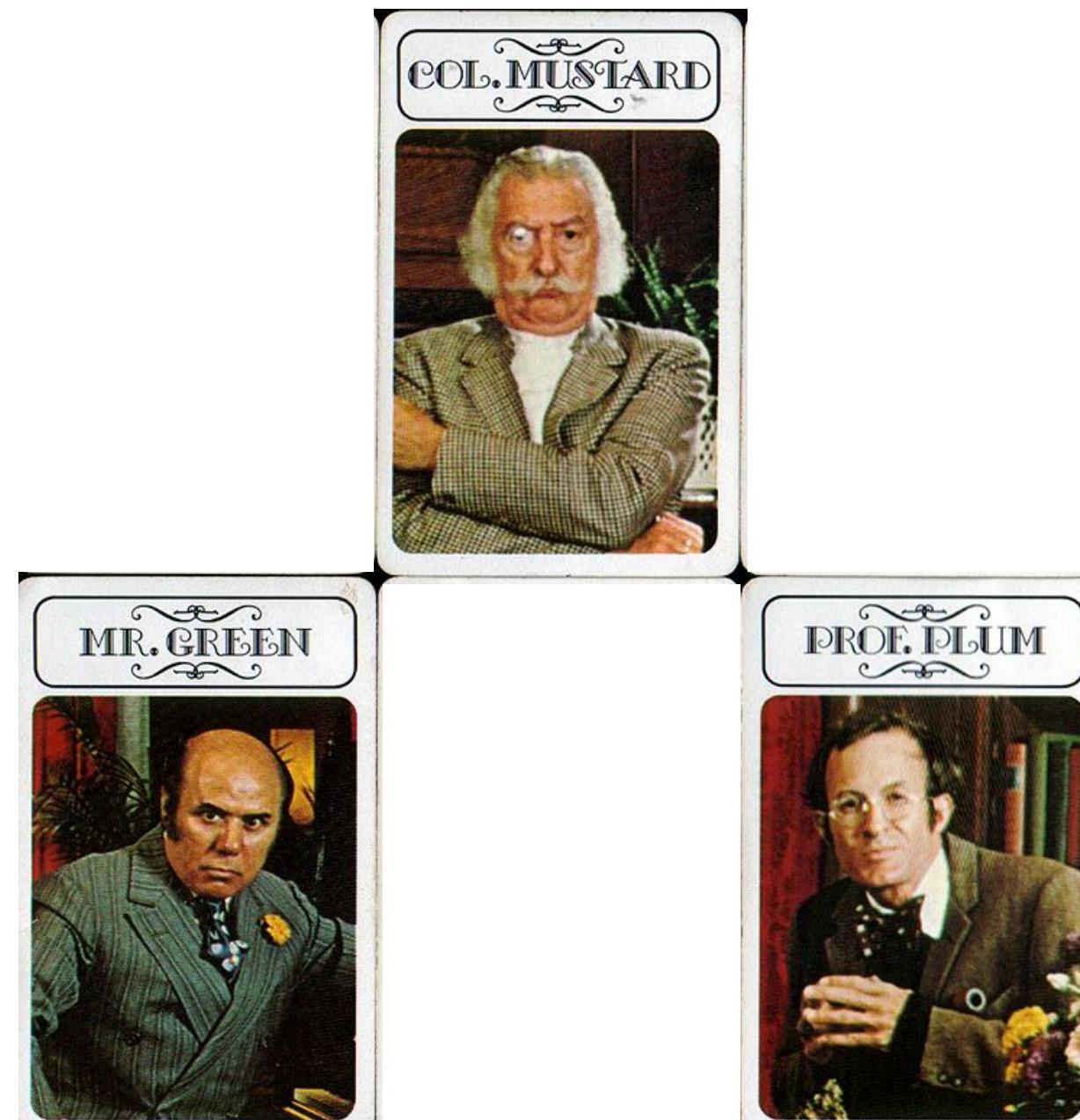
```

use naive definition of probability

Clue guide to probability

Who?

- **conditional probability:**
- $p(A | B)$ (probability of A given B)
- **Definition:** $p(A | B) = \frac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} | \text{male}) = \frac{1/6}{1/2} = 1/3$



who	gender
col_mustard	male
mr_green	male
prof_plum	male

```
1 df.suspects %>%
2   filter(gender == "male") %>%
3   summarize(p_prof_plum_given_male =
4             sum(who == "prof_plum") /
5             n())
```

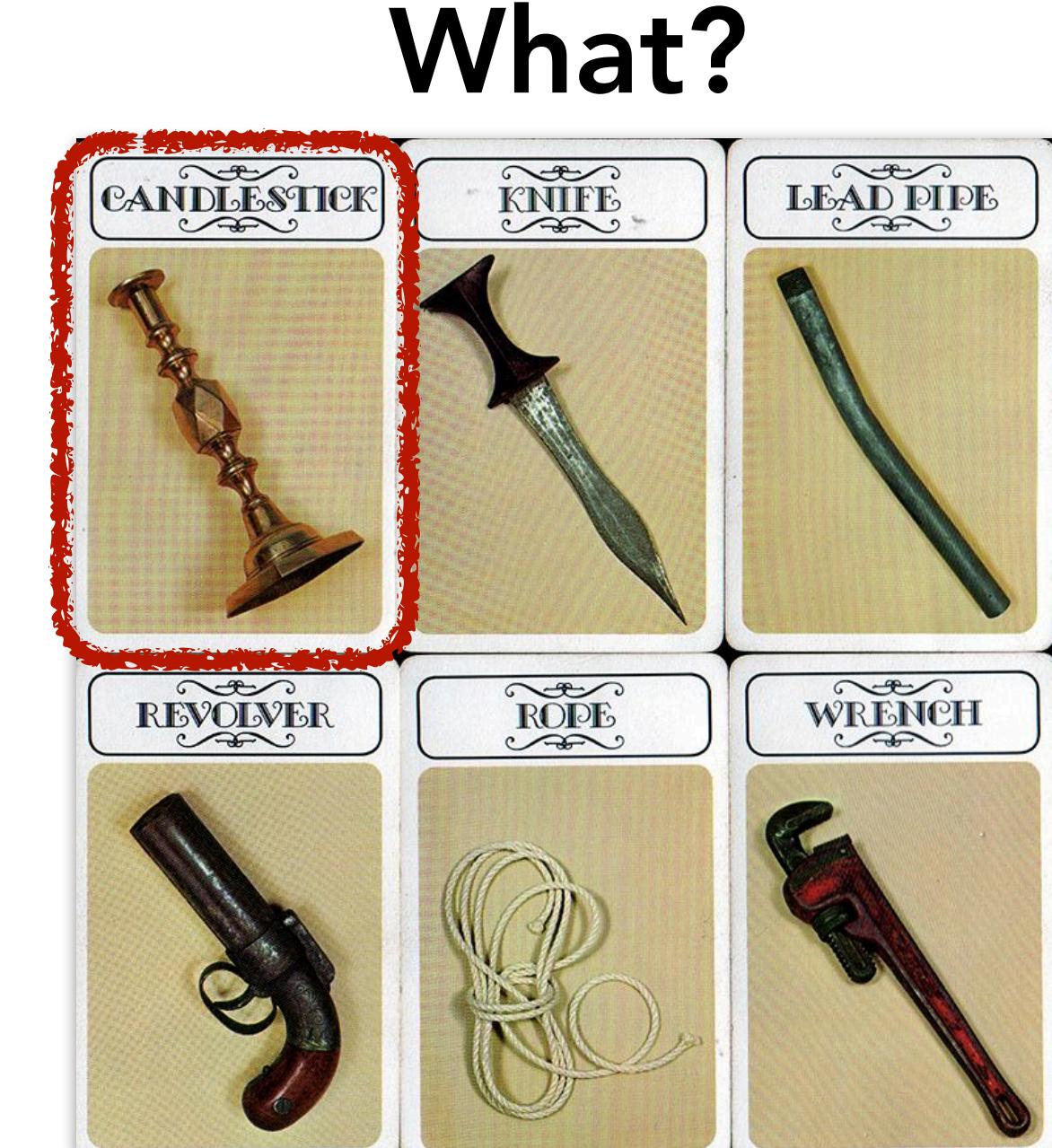
Clue guide to probability

- *independence*:
- A and B are independent if
- **Definition:** $p(A | B) = p(A)$
- (probability of A does not change if you know B)



Who?

- $p(\text{Prof Plum} | \text{candle stick}) = p(\text{Prof Plum})$
- each card (who and what) is drawn from a separate pack of cards



What?

Clue guide to probability

Who?



- ***joint probability:***
- if A and B are independent then
- **Definition:** $p(A, B) = p(A) \cdot p(B)$

What?



- $p(\text{Prof Plum, candle stick}) =$
 $p(\text{Prof Plum}) \cdot p(\text{candle stick}) =$

$$\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Clue guide to probability

- **dependence:**
- **Definition:** $p(A | B) \neq p(A)$
- **Definition:** $p(A, B) = p(A) \cdot p(B | A)$
- if women were more likely than men to use the revolver then
- $p(\text{Mrs. White} | \text{Revolver}) > p(\text{Mrs. White})$

Who?



What?



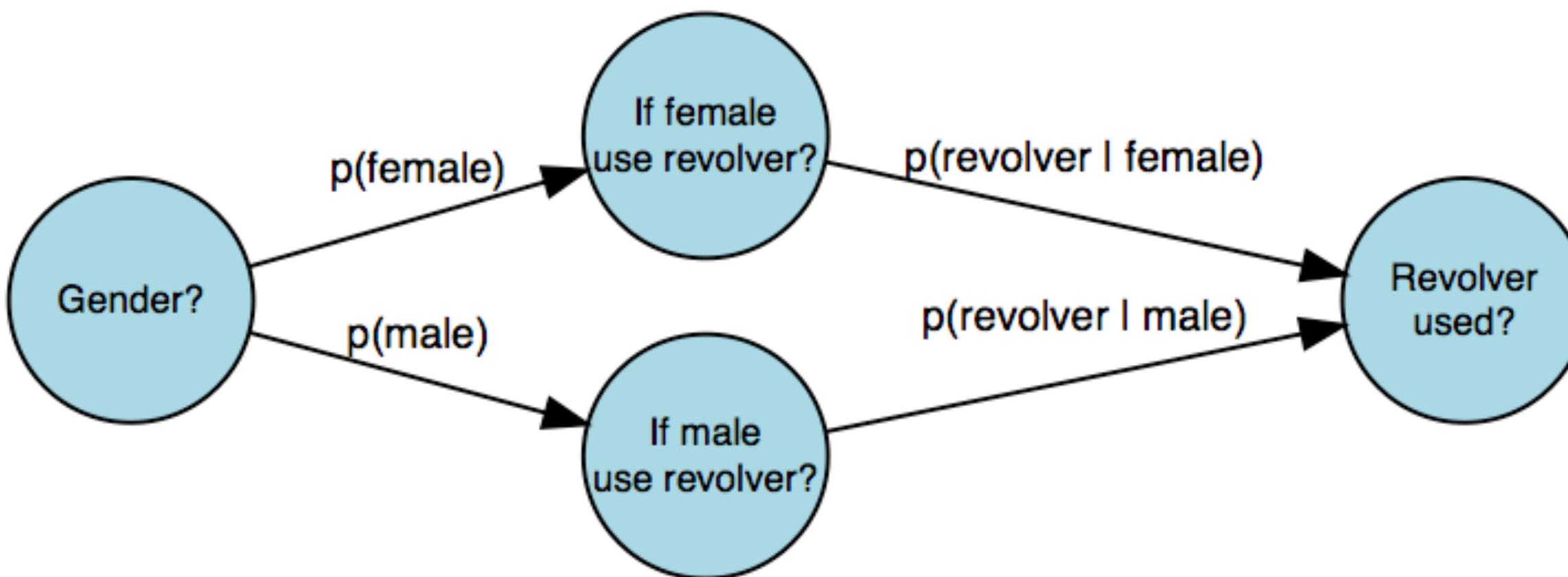
Clue guide to probability

- *law of total probability*
- Definition:

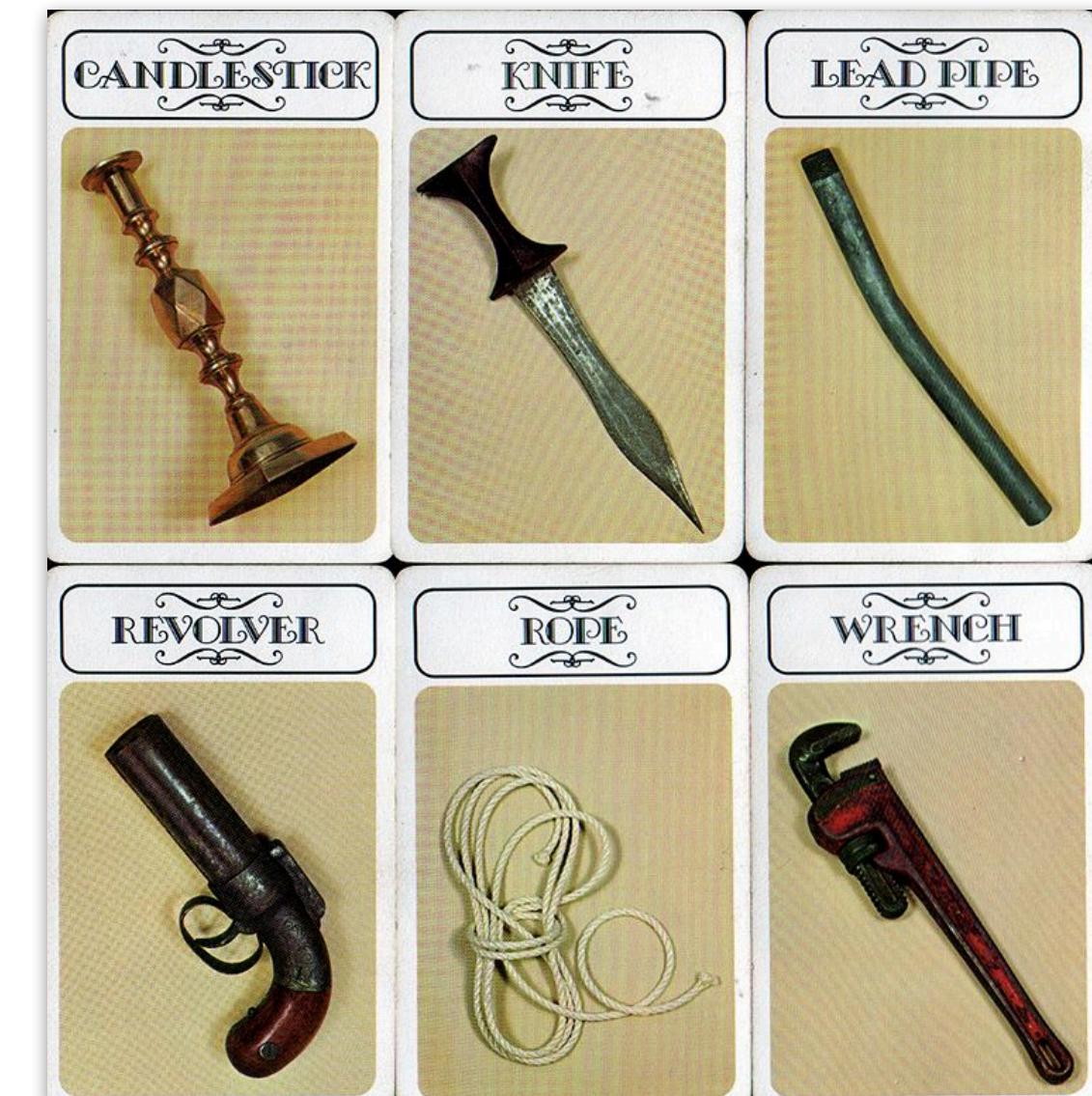
$$p(A) = p(A | B) \cdot p(B) + p(A | \neg B) \cdot p(\neg B)$$

$$p(A) = \sum_{i=1}^n p(A | B_i) \cdot p(B_i)$$

$$p(\text{what} = \text{Revolver}) = ?$$

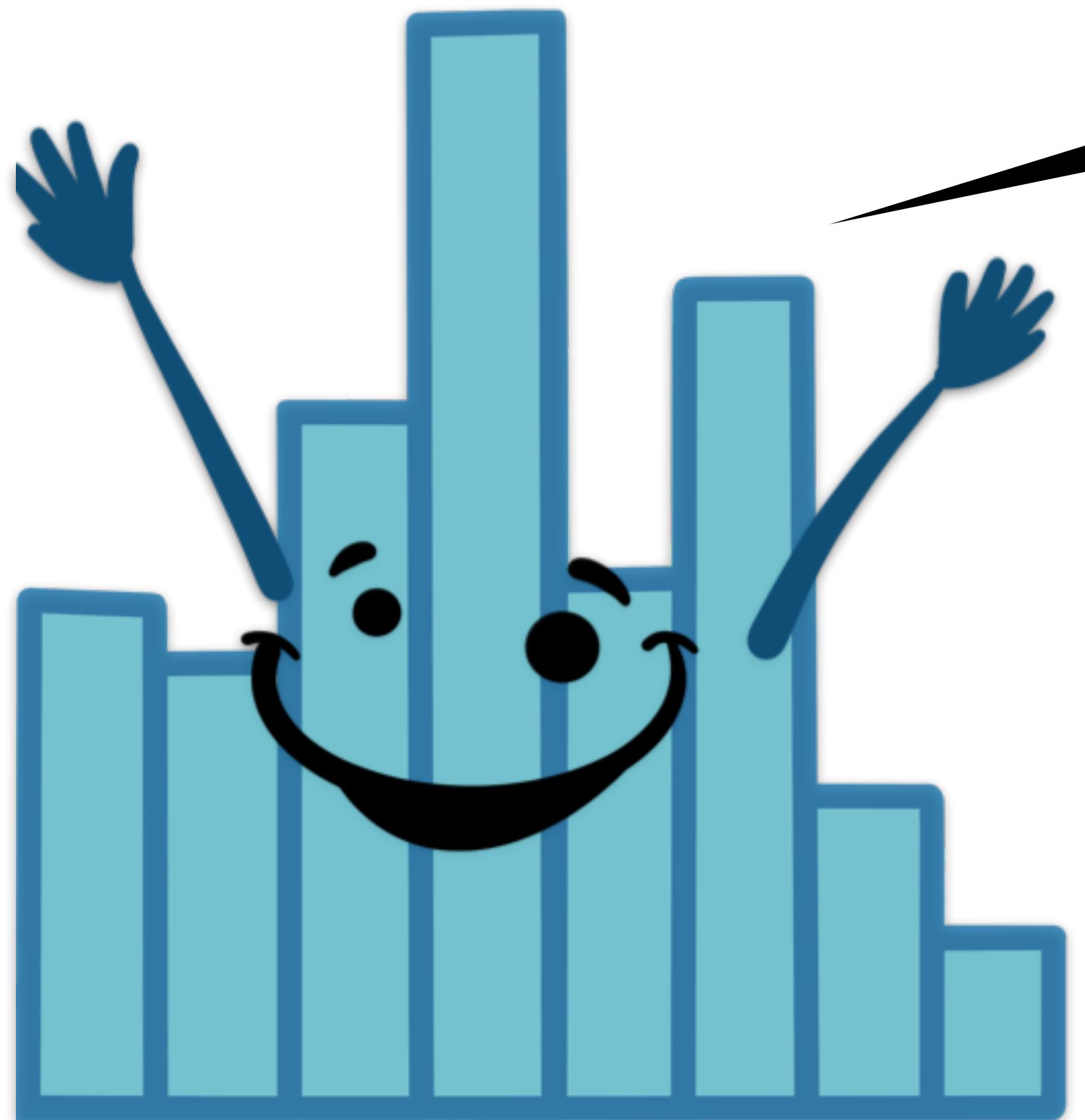


Who?



02:00

stretch break!



Understanding Bayes' rule

Clue guide to probability

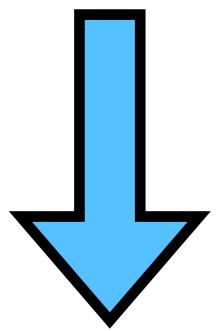
Bayes Theorem in a few steps



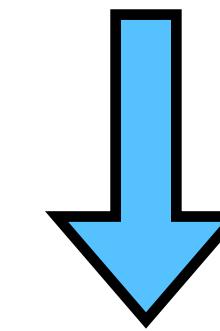
Clue guide to probability

- Bayes' theorem (derivation)

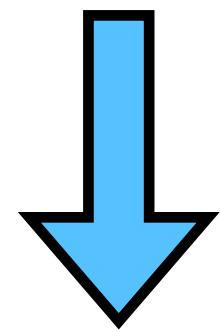
$$p(B | A) = \frac{p(A, B)}{p(A)}$$



$$p(A | B) = \frac{p(A, B)}{p(B)}$$



$$p(A, B) = p(B | A) \cdot p(A) = p(A | B) \cdot p(B)$$



$$p(B | A) = \frac{p(A | B) \cdot p(B)}{p(A)}$$

Clue guide to probability

$$p(B|A) = \frac{p(A|B) \cdot p(B)}{p(A)}$$

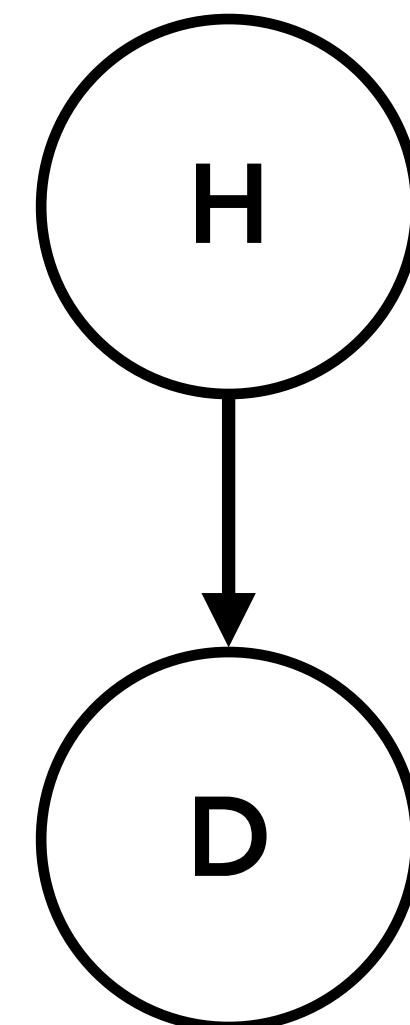
posterior **likelihood** **prior**

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

subjective probability
interpretation

H = Hypothesis

D = Data



formal framework for learning from data

updating our prior belief $p(H)$, to a posterior belief $p(H|D)$
given some data

Clue guide to probability

$$p(H|D) = \frac{\text{likelihood} \cdot \text{prior}}{p(D)}$$

H = Hypothesis
 D = Data

probability of the data?!

law of total probability:

$$p(D) = \sum_{i=1}^n p(D|H_i) \cdot p(H_i)$$

**take into account all the different ways
in which the data could have come about**

Clue guide to probability

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that Fred tests positive.

Suppose that the test is “95% accurate”; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that **$P(T|D) = 0.95$** and **$P(\neg T|\neg D) = 0.95$** .

The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

What's the probability that Fred has conditionitis?

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $P(T|D) = 0.95$ and $P(\neg T|\neg D) = 0.95$. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

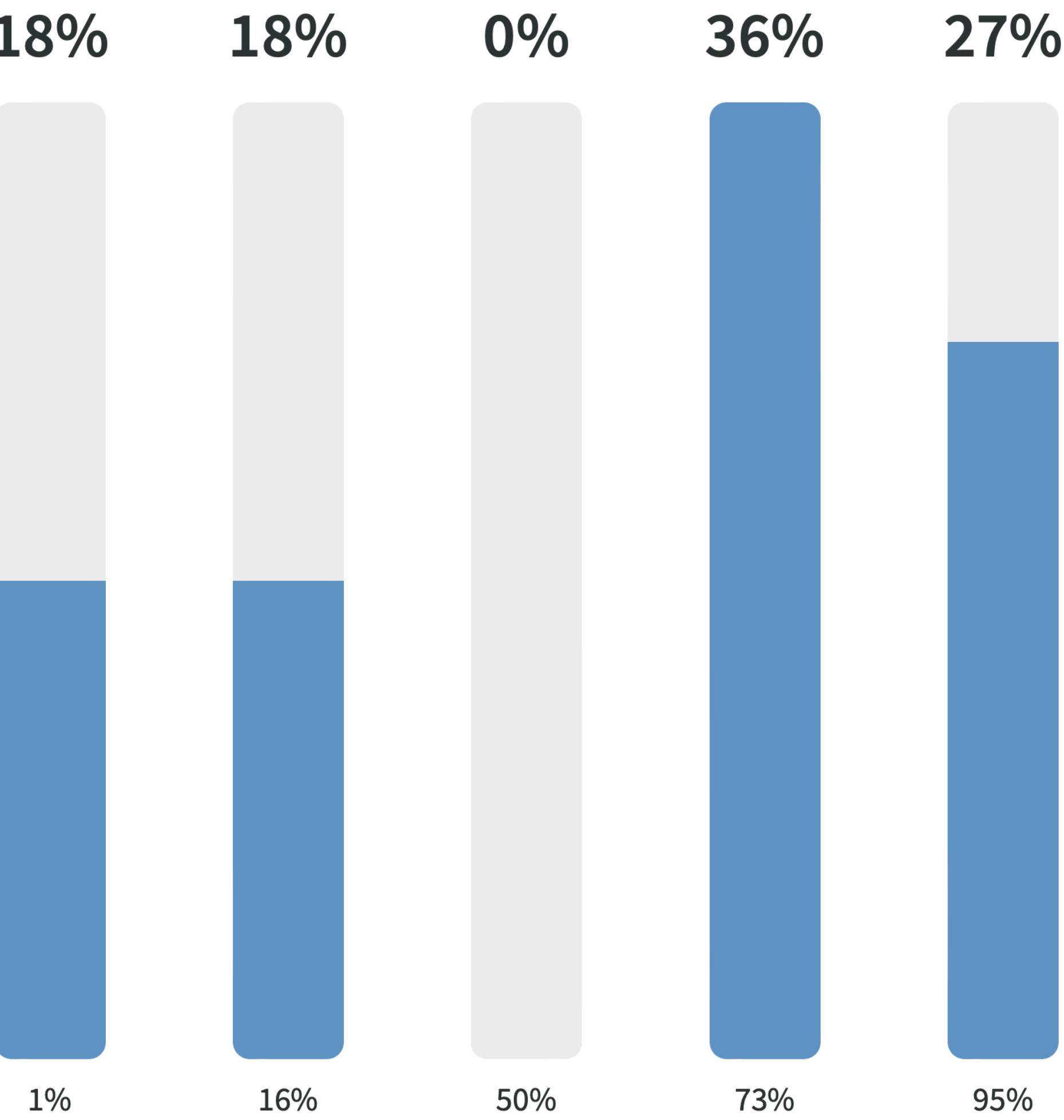
Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

1% 16% 50% 73% 95%

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is “95% accurate”; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $P(T|D) = 0.95$ and $P(\neg T|\neg D) = 0.95$. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.



What's the probability that Fred has conditionitis?

A patient named Fred is tested for a disease called *conditionitis*, a medical condition that afflicts **1% of the population**. The test result is positive, i.e., the test claims that Fred has the disease. Let **D** be the event that Fred has the disease and **T** be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $P(T|D) = 0.95$ and $P(\neg T|\neg D) = 0.95$. The quantity $P(T|D)$ is known as the *sensitivity* (= true positive rate) of the test, and $P(\neg T|\neg D)$ is known as the *specificity* (= true negative rate).

Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

1% 16% 50% 73% 95%

Clue guide to probability

what we know

$$P(D) = 0.01$$

$$P(T|D) = 0.95$$

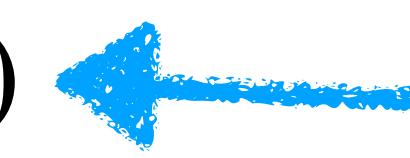
$$P(T|\neg D) = 0.05$$

what we want to know

$$P(D|T) = ?$$

$$p(D|T) = \frac{p(T|D) \cdot p(D)}{p(T)} \quad \text{Bayes' rule}$$

$$p(D|T) = \frac{p(T|D) \cdot p(D)}{p(T|D) \cdot p(D) + p(T|\neg D) \cdot p(\neg D)}$$



law of total
probability

XX

Clue guide to probability

what we know

$$P(D) = 0.01$$

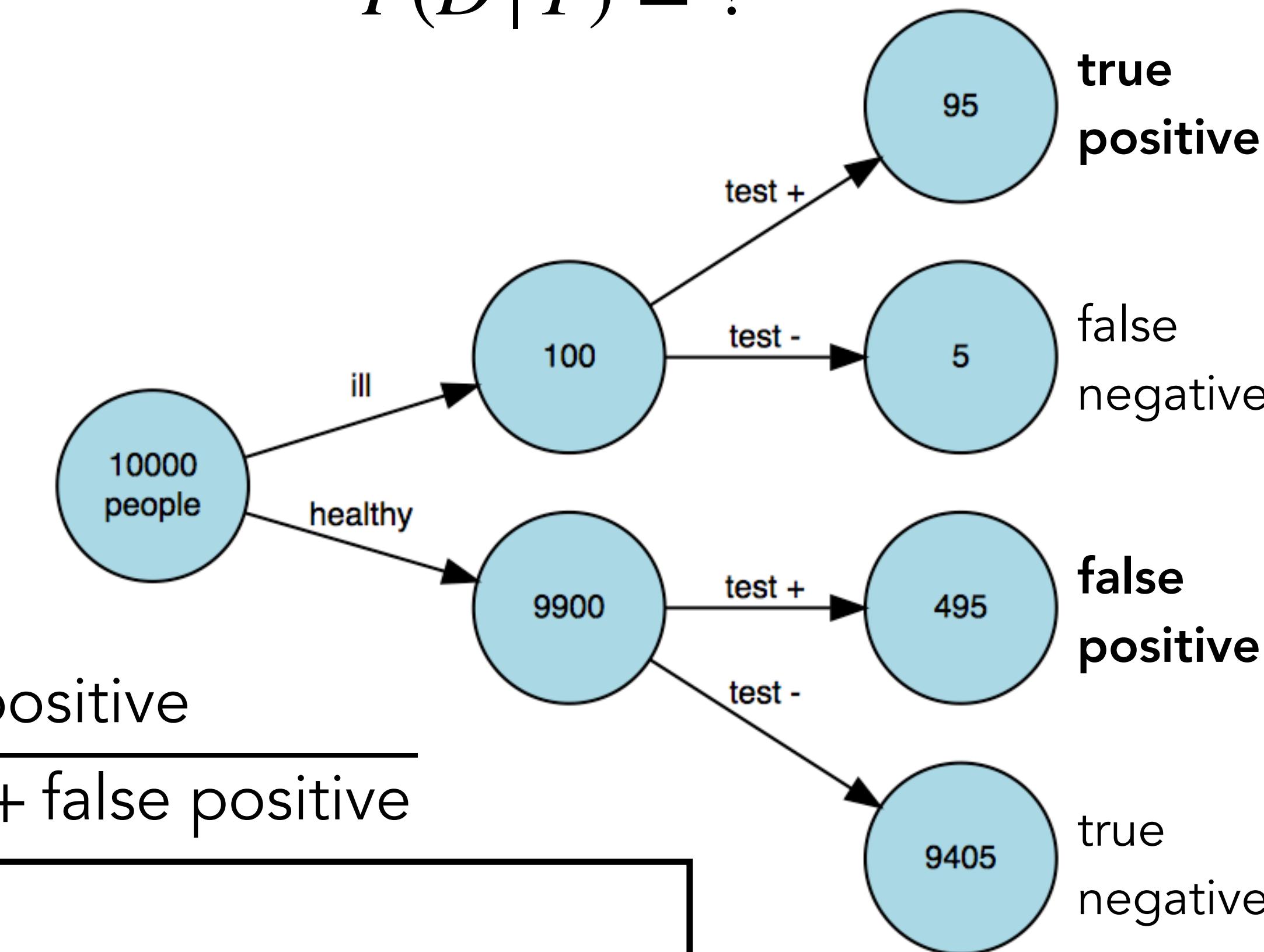
$$P(T|D) = 0.95$$

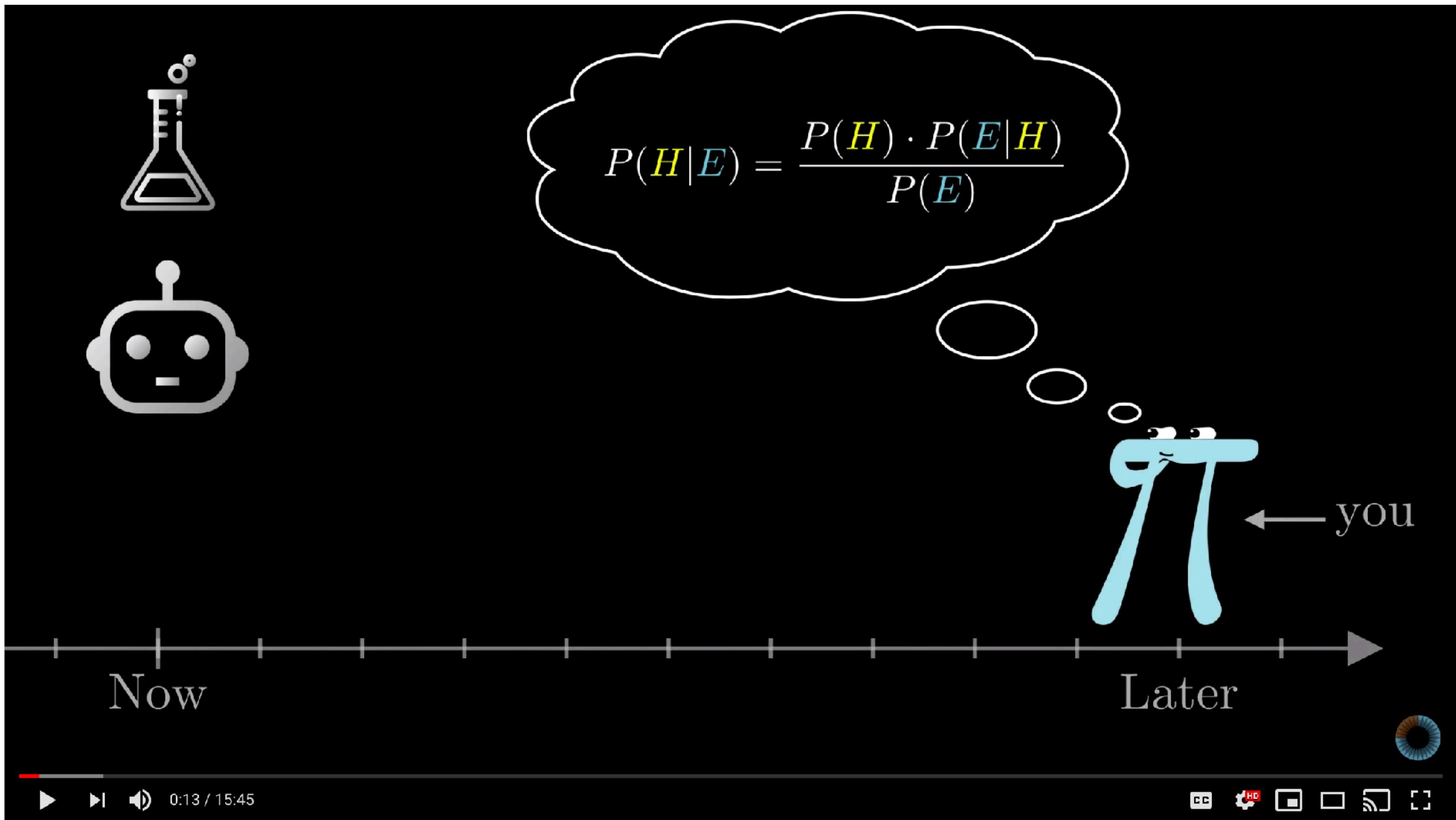
$$P(T|\neg D) = 0.05$$

$$P(D|T) = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

what we want to know

$$P(D|T) = ?$$





Bayes theorem, and making probability intuitive

461,105 views • Dec 22, 2019

1K 26K 228 SHARE SAVE ...

<https://www.youtube.com/watch?v=HZGCoVF3YvM&feature=youtu.be>

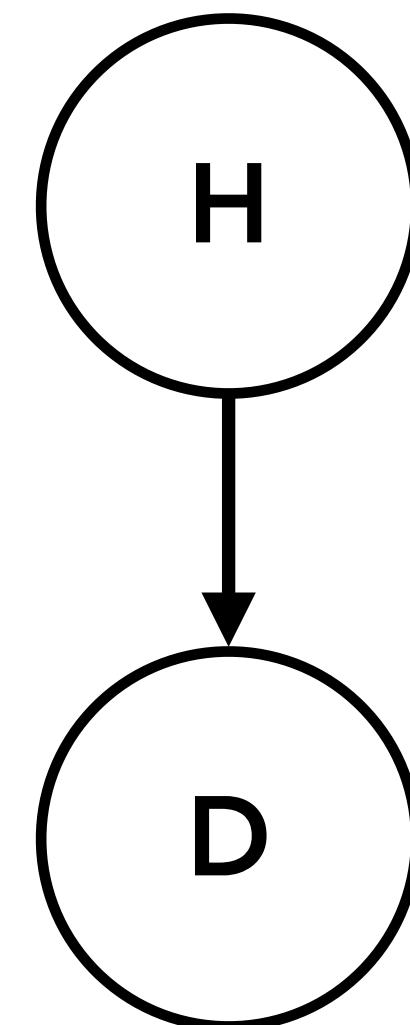
Bayes' theorem in three panels

$$p(H|D) = \frac{\text{likelihood} \quad \text{prior}}{p(D)} \quad \begin{matrix} \text{subjective probability} \\ \text{interpretation} \end{matrix}$$

posterior **likelihood** **prior**

$$p(H|D) = \frac{p(D|H) \cdot p(H)}{p(D)}$$

H = Hypothesis
 D = Data



Getting Bayes' right matters

Getting Bayes right matters!

The image shows the front cover of a PNAS (Proceedings of the National Academy of Sciences) journal issue. The title of the article is "Officer characteristics and racial disparities in fatal officer-involved shootings". The authors listed are David J. Johnson^{a,b,1}, Trevor Tress^b, Nicole Burkell^b, Carley Taylor^b, and Joseph Cesario^b. The journal is identified as "PNAS" and "PSYCHOLOGICAL AND COGNITIVE SCIENCES". A small logo in the top right corner says "Check for updates". The abstract discusses the lack of databases for fatal officer-involved shootings (FOIS) and how this affects the ability to test for racial disparities. It also mentions the use of an approach that sidesteps the benchmark debate by directly predicting the race of civilians fatally shot rather than comparing the rate at which racial groups are shot to some benchmark. The paper reports three main findings related to officer race, county-level violent crime, and the type of shooting.

Original claim:

Requires Bayes' rule

$$\begin{aligned} & \Pr(\text{shot}|\text{minority civilian, white officer}, X) \\ & - \Pr(\text{shot}|\text{minority civilian, minority officer}, X) \\ & \quad \Pr(\text{min. civ. } |\text{shot, white off.}, X) \\ & \quad \times \Pr(\text{shot}|\text{white off.}, X) \\ & = \frac{\Pr(\text{min. civ. } |\text{shot, min. off.}, X)}{\Pr(\text{minority civilian}|\text{white officer}, X)} \\ & \quad \times \Pr(\text{shot}|\text{min. off.}, X) \\ & - \frac{\Pr(\text{min. civ. } |\text{shot, min. off.}, X)}{\Pr(\text{minority civilian}|\text{minority officer}, X)}. \end{aligned} \quad [2]$$

Claim:

"White officers are not more likely to shoot minority civilians than non-White officers"

$$\begin{aligned} & \Pr(\text{shot}|\text{minority civilian, white officer}, X) \\ & - \Pr(\text{shot}|\text{minority civilian, minority officer}, X) \leq 0, \end{aligned} \quad [1]$$

What the statistic says:

"whether a person fatally shot was more likely to be Black (or Hispanic) than White"

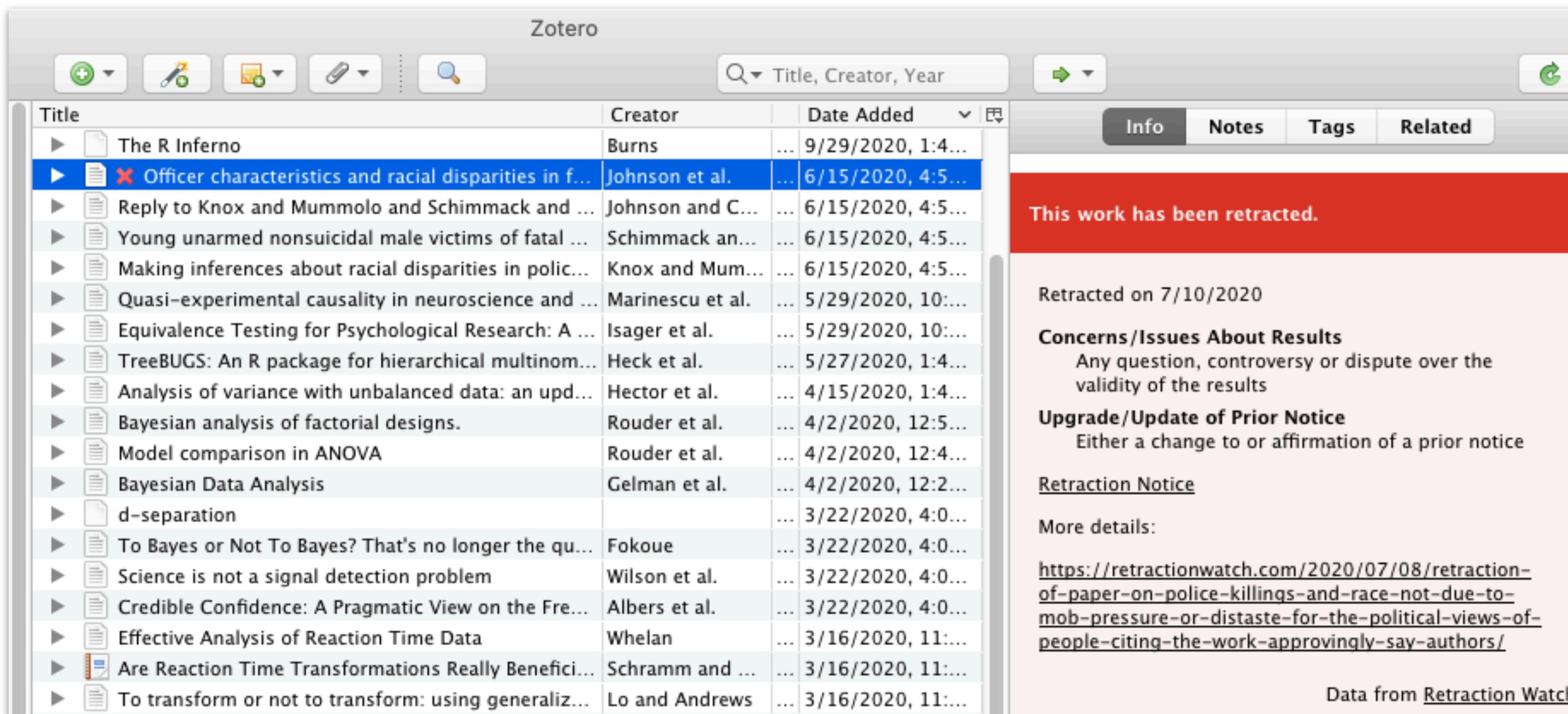
authors didn't have the relevant data to support their claim!

paper was retracted

Johnson, D. J., Tress, T., Burkell, N., Taylor, C., & Cesario, J. (2019). Officer characteristics and racial disparities in fatal officer-involved shootings. *Proceedings of the National Academy of Sciences*, 116(32), 15877–15882.

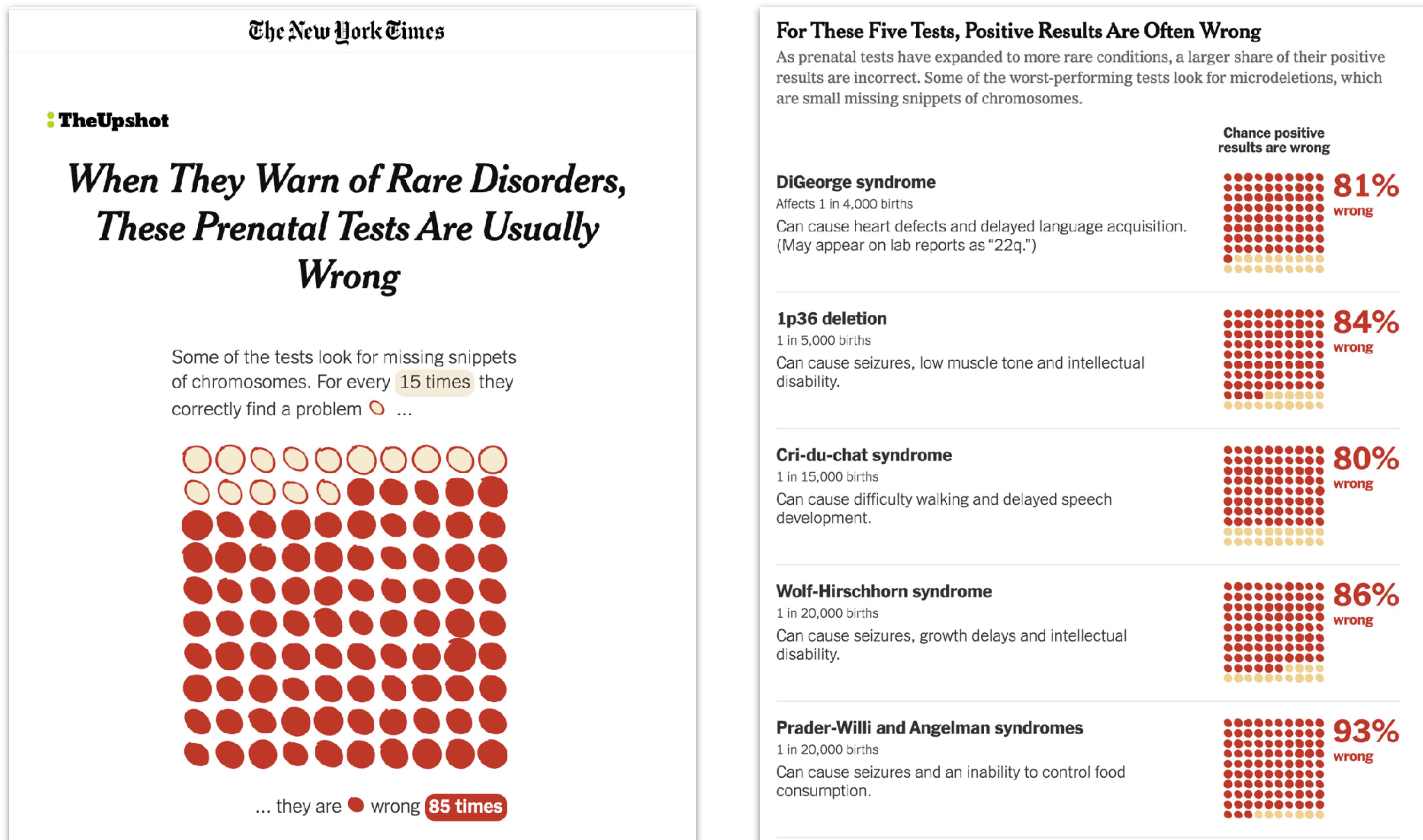
Knox, D., & Mummolo, J. (2020). Making inferences about racial disparities in police violence. *Proceedings of the National Academy of Sciences*, 117(3), 1261–1262.

Getting Bayes right matters!



Tip: Use Zotero as a reference manager!

Getting Bayes right matters!



Getting Bayes right matters!

sensitivity: $p(T|D) = 0.999$

T = positive test result

specificity: $p(\neg T|\neg D) = 0.999$

$\neg T$ = negative test result

prior: $p(D) = 0.0001$

D = disease

$\neg D$ = no disease

data: T (positive test result)

81% wrong

$$\text{posterior: } p(D|T) = \frac{p(T|D) \cdot p(D)}{p(T|D) \cdot p(D) + p(T|\neg D) \cdot p(\neg D)}$$

$$= \frac{0.999 \cdot 0.0001}{0.999 \cdot 0.0001 + 0.001 \cdot 0.9999} \approx 0.09$$

Getting Bayes right matters!

Most people who are in the hospital
being treated for Covid are vaccinated.

Getting Bayes right matters!

likelihood: $p(H|V) = 0.2$

H = hospitalized

$p(H|\neg V) = 0.5$

$\neg H$ = not hospitalized

prior: $p(V) = 0.8$

V = vaccinated

$\neg V$ = no vaccinated

data: H (the person is in the hospital)

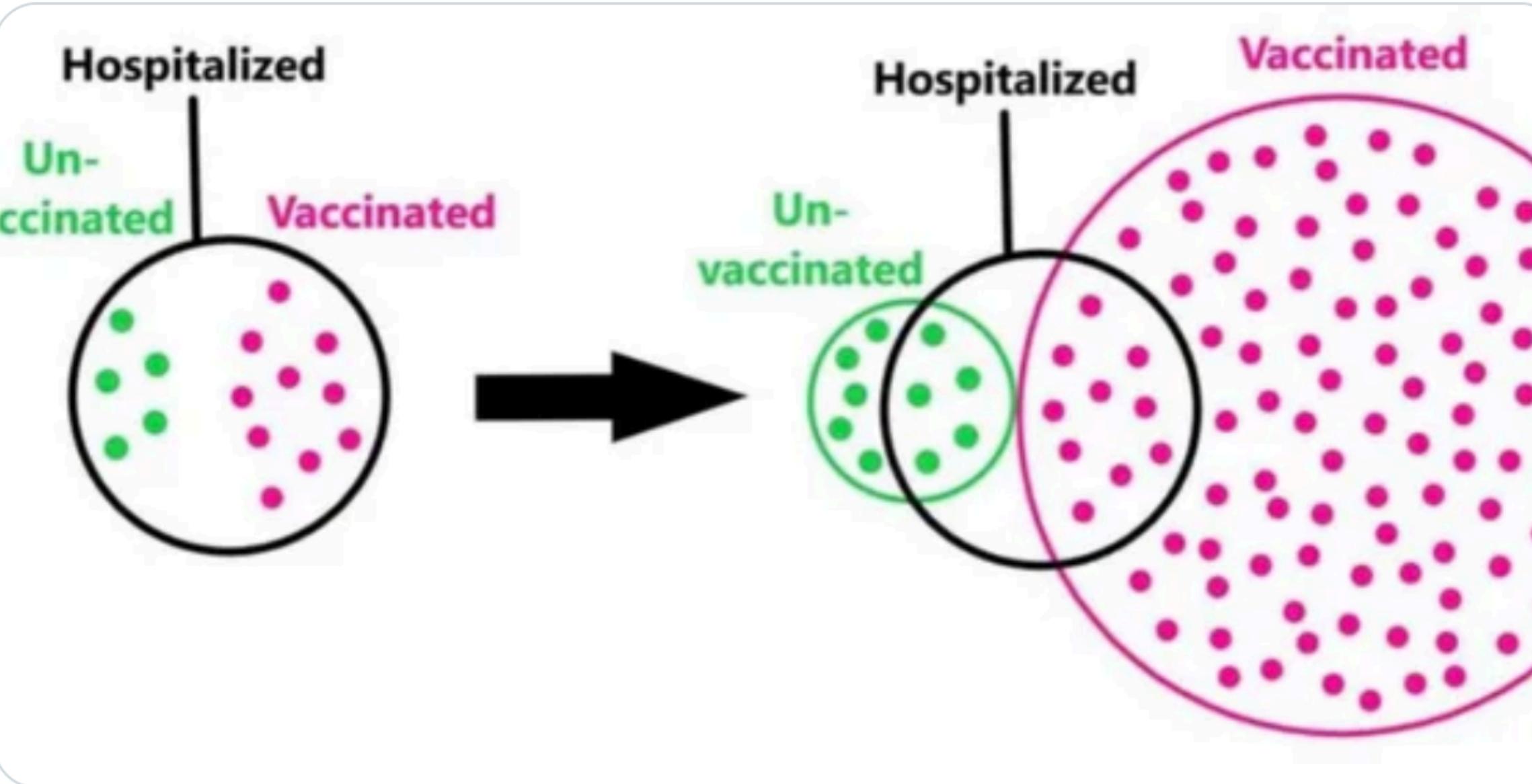
$$\text{posterior: } p(V|H) = \frac{p(H|V) \cdot p(V)}{p(H|V) \cdot p(V) + p(H|\neg V) \cdot p(\neg V)}$$
$$= \frac{0.2 \cdot 0.8}{0.2 \cdot 0.8 + 0.5 \cdot 0.2} \approx 0.62$$

62% of the hospitalized
people are vaccinated

Bayes' rule matters

 **Nick Brown**
@sTeamTraen ...

Stolen from Reddit. May be of some use.



The diagram consists of two circles. The left circle is divided into three regions: a top section labeled "Hospitalized" (grey), a bottom-left section labeled "Un-vaccinated" (green), and a bottom-right section labeled "Vaccinated" (pink). It contains several green and pink dots. A large black arrow points to the right, leading to a second circle. This second circle is also divided into three regions: a top section labeled "Hospitalized" (grey), a bottom-left section labeled "Un-vaccinated" (green), and a bottom-right section labeled "Vaccinated" (pink). The "Vaccinated" section is much larger than the others and contains many more pink dots. The "Un-vaccinated" and "Hospitalized" sections are much smaller and contain fewer dots.

1:22 PM · Nov 20, 2021 · Twitter Web App

565 Retweets **54** Quote Tweets **2,752** Likes

Building a Bayesis



Rolling the dice



Four sided



Six sided

both dice are equally likely to be picked
 $p(\triangle^4) = p(\text{dice}) = 0.5$

both dice are equal sided
(uniform probability over the different numbers)

Which die do you think was rolled?

$$4 \quad p(\triangle^4 | \text{data}) = ?$$

$$4, 2, 1 \quad p(\triangle^4 | \text{data}) = 0.77$$

$$4, 2, 1, 3, 1 \quad p(\triangle^4 | \text{data}) = 0.88$$

$$4, 2, 1, 3, 1, 5 \quad p(\triangle^4 | \text{data}) = 0$$

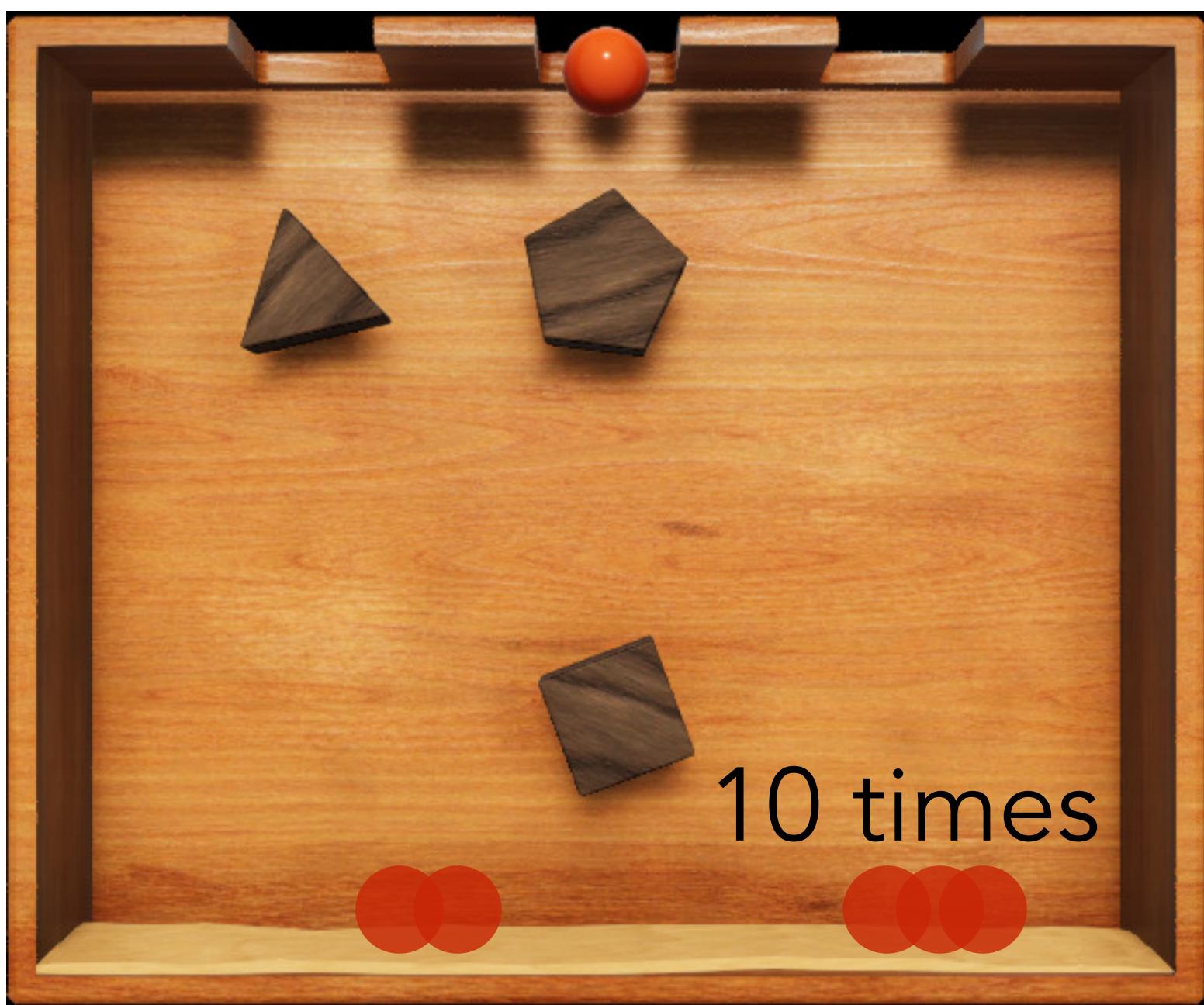
Physical reasoning



Physical reasoning

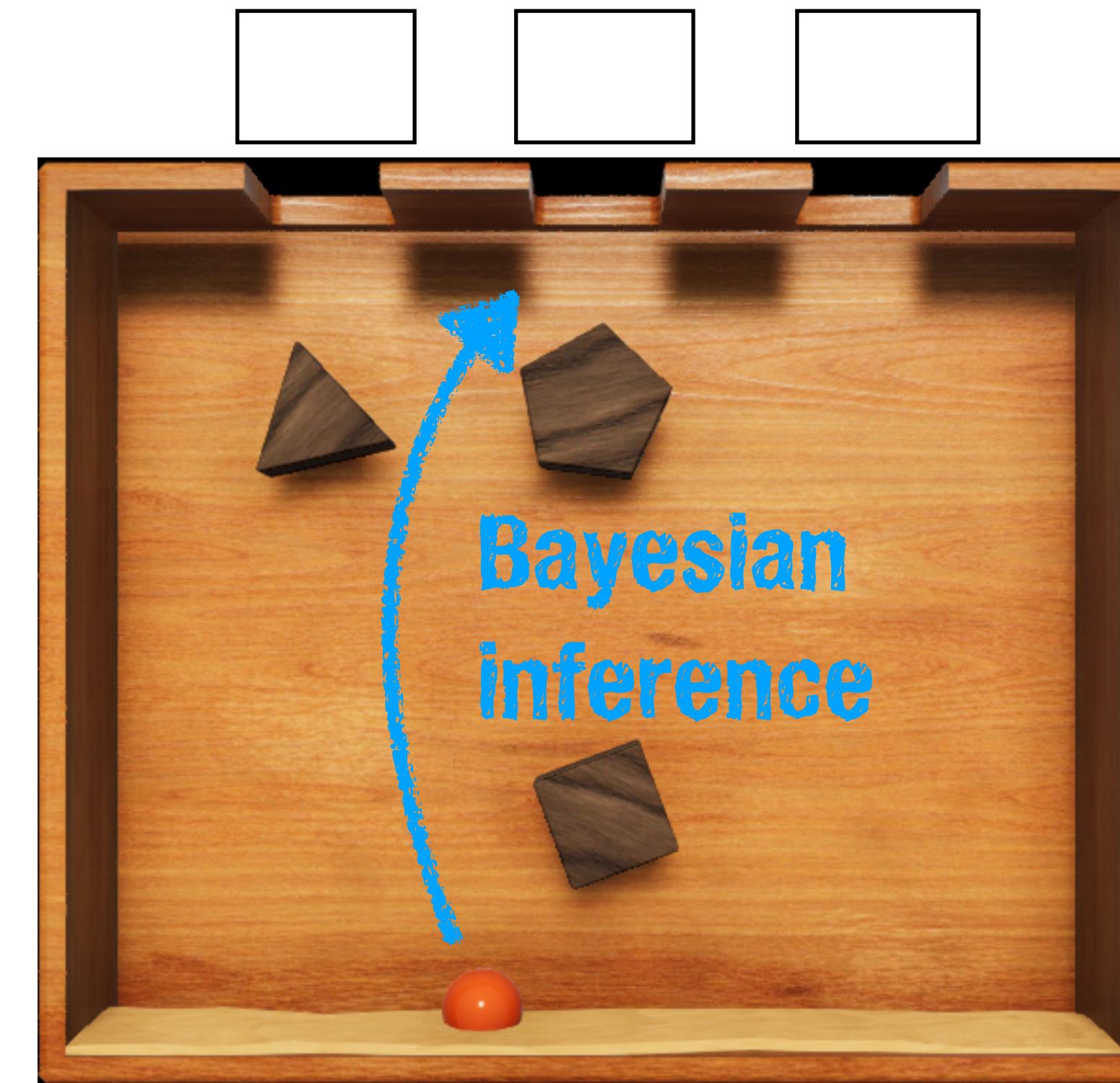


Prediction



Where will the ball land?

Inference



In which hole was the ball dropped?

Outline

- Introduction to probability / Recap
 - Motivation
 - Counting possibilities
 - **Clue** guide to probability
 - Understanding Bayes' Rule
 - Getting Bayes' right matters!
 - Building a Bayesis

I want more!

Chapter 9 Introduction to probability

[God] has afforded us only the twilight ... of Probability.

— John Locke

Up to this point in the book, we've discussed some of the key ideas in experimental design, and we've talked a little about how you can summarise a data set. To a lot of people, this is all there is to statistics: it's about calculating averages, collecting all the numbers, drawing pictures, and putting them all in a report somewhere. Kind of like stamp collecting, but with numbers. However, statistics covers much more than that. In fact, descriptive statistics is one of the smallest parts of statistics, and one of the least powerful. The bigger and more useful part of statistics is that it provides that let you make inferences about data.

Once you start thinking about statistics in these terms – that statistics is there to help us draw inferences from data – you start seeing examples of it everywhere. For instance, here's a tiny extract from a newspaper article in the Sydney Morning Herald (30 Oct 2010):

"I have a tough job," the Premier said in response to a poll which found her government is now the most unpopular Labor administration in polling history, with a primary vote of just 23 per cent.

This kind of remark is entirely unremarkable in the papers or in everyday life, but let's have a think about what it entails. A polling company has conducted a survey, usually a pretty big one because they can afford it. I'm too lazy to track down the original survey, so let's just imagine that they called 1000 NSW voters at random, and 230 (23%) of those claimed that they intended to vote for the ALP. For the 2010 Federal election, the Australian Electoral Commission reported 4,810,795 enrolled voters in NSW; so the opinions of the remaining 4,809,795 voters (about 99.98% of voters) remain unknown to us. Even assuming that no-one lied to the polling

INTERACTIVE COURSE Foundations of Probability in R

Start Course Play Intro Video Bookmark

4 hours 13 Videos 54 Exercises 27,296 Participants 4,350 XP

Course Description

Probability is the study of making predictions about random phenomena. In this course, you'll learn about the concepts of random variables, distributions, and conditioning, using the example of coin flips. You'll also gain intuition for how to solve probability problems through random simulation. These principles will help you understand statistical inference and can be applied to draw conclusions from data.

1 The binomial distribution

One of the simplest and most common examples of a random phenomenon is a coin flip: an event that is either "yes" or "no" with some probability. Here you'll learn about the binomial distribution, which describes the behavior of a combination of yes/no trials and how to predict and simulate its behavior.

VIEW CHAPTER DETAILS Continue Chapter

This course is part of these tracks:

- Probability and Distributions with R
- Statistician with R

David Robinson
Principal Data Scientist at Heap

Probability Cheatsheet v2.0

Compiled by William Chen (<http://wchen.org>) and Joe Blitzstein, with contributions from Sebastian Chiu, Yuan Jing, Yuxi He, and Jossy Hwang. Material based on Joe Blitzstein's (stat110.net) lectures (<http://stat110.net>) and Blitzstein/Hwang's Introduction to Probability (<http://probability.courses.csail.mit.edu>). Licensed under CC BY-NC-ND 4.0. Please share comments, suggestions, and errors at https://github.com/wchen/probability_cheatsheet.

Last Updated September 4, 2015

Counting

Multiplication Rule

Let's say we have a compound experiment (an experiment with multiple components). If the 1st component has n_1 possible outcomes, the 2nd component has n_2 possible outcomes, and the r th component has n_r possible outcomes, then overall there are $n_1 n_2 \dots n_r$ possibilities for the whole experiment.

Sampling Table

The sampling table gives the number of possible samples of size k out of a population of size n , under various assumptions about how the sample is collected.

	Order Matters	Not Matter
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$n!$	$\binom{n}{k}$

Naive Definition of Probability

If all outcomes are equally likely, the probability of an event A happening is:

$$P_{\text{naive}}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of outcomes}}$$

Law of Total Probability (LOTP)

Let $B_1, B_2, B_3, \dots, B_n$ be a partition of the sample space (i.e., they are disjoint and their union is the entire sample space).

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$

De Morgan's Laws: A useful identity that can make calculating probabilities of unions easier by relating them to intersections, and vice versa. Analogous results hold with more than two sets.

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Thinking Conditionally

Independence

Independent Events: A and B are independent if knowing whether A occurs does not give any information about whether B occurs. More formally, A and B (which have nonzero probability) are independent if and only if one of the following equivalent statements holds:

$$P(A \cap B) = P(A)P(B)$$

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

Conditional Independence: A and B are conditionally independent given C if $P(A \cap B|C) = P(A|C)P(B|C)$. Conditional independence does not imply independence, and independence does not imply conditional independence.

Unions, Intersections, and Complements

De Morgan's Laws: A useful identity that can make calculating probabilities of unions easier by relating them to intersections, and vice versa. Analogous results hold with more than two sets.

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Joint, Marginal, and Conditional

Joint Probability: $P(A \cap B)$ – Probability of A and B .

Marginal (Unconditional) Probability: $P(A)$ – Probability of A .

Conditional Probability: $P(A|B) = P(A \cap B)/P(B)$ – Probability of A given that B occurred.

Conditional Probability or Probability: $P(A|B)$ is a probability function for any fixed B . Any theorem that holds for probability also holds for conditional probability.

Probability of an Intersection or Union

Intersections via Conditioning

$$P(A, B) = P(A|B)P(B)$$

$$P(A, B, C) = P(A|B)P(B|C)P(C)$$

Unions via Inclusion-Exclusion

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Simpson's Paradox

It is possible to have $P(A|B, C) < P(A|B')$ and $P(A|B, C') < P(A|B', C')$ yet also $P(A|B) > P(A|B')$.

The PMF satisfies

$$p_X(x) \geq 0 \text{ and } \sum_x p_X(x) = 1$$

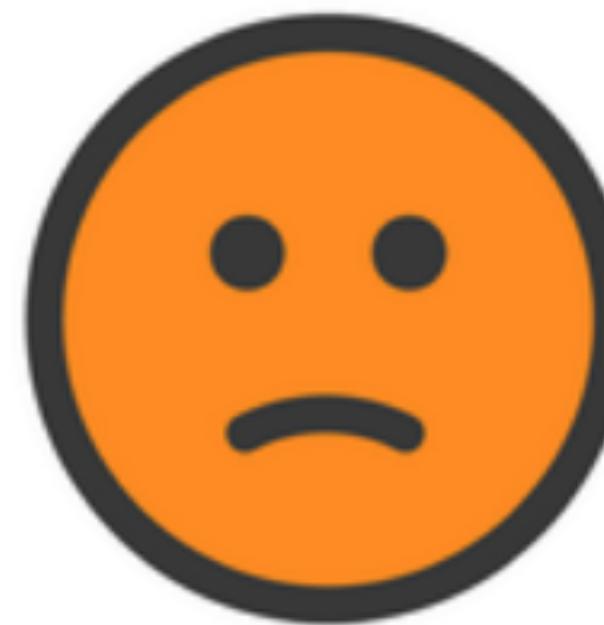
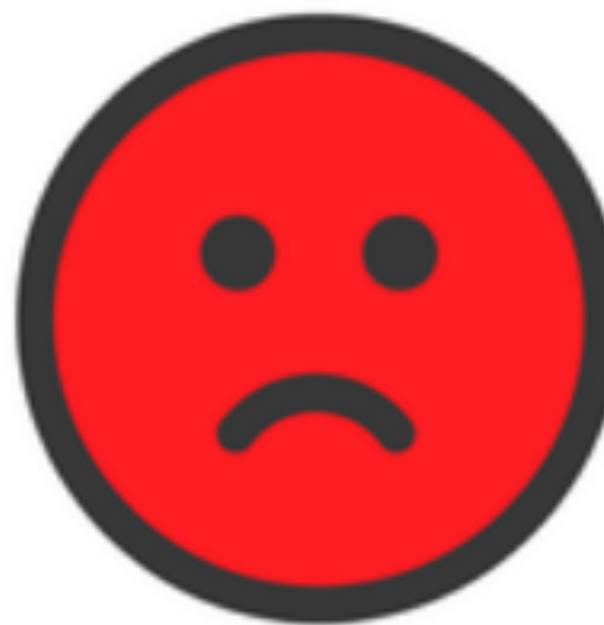
in the figures/
folder for these
materials on
canvas

Feedback

How was the pace of today's class?

much a little just a little much
too too right too too
slow slow fast fast

How happy were you with today's class overall?



What did you like about today's class? What could be improved next time?

Have a nice Martin Luther King, Jr. day!

The time is
always right
to do what
is right.

- Martin Luther King, Jr.

