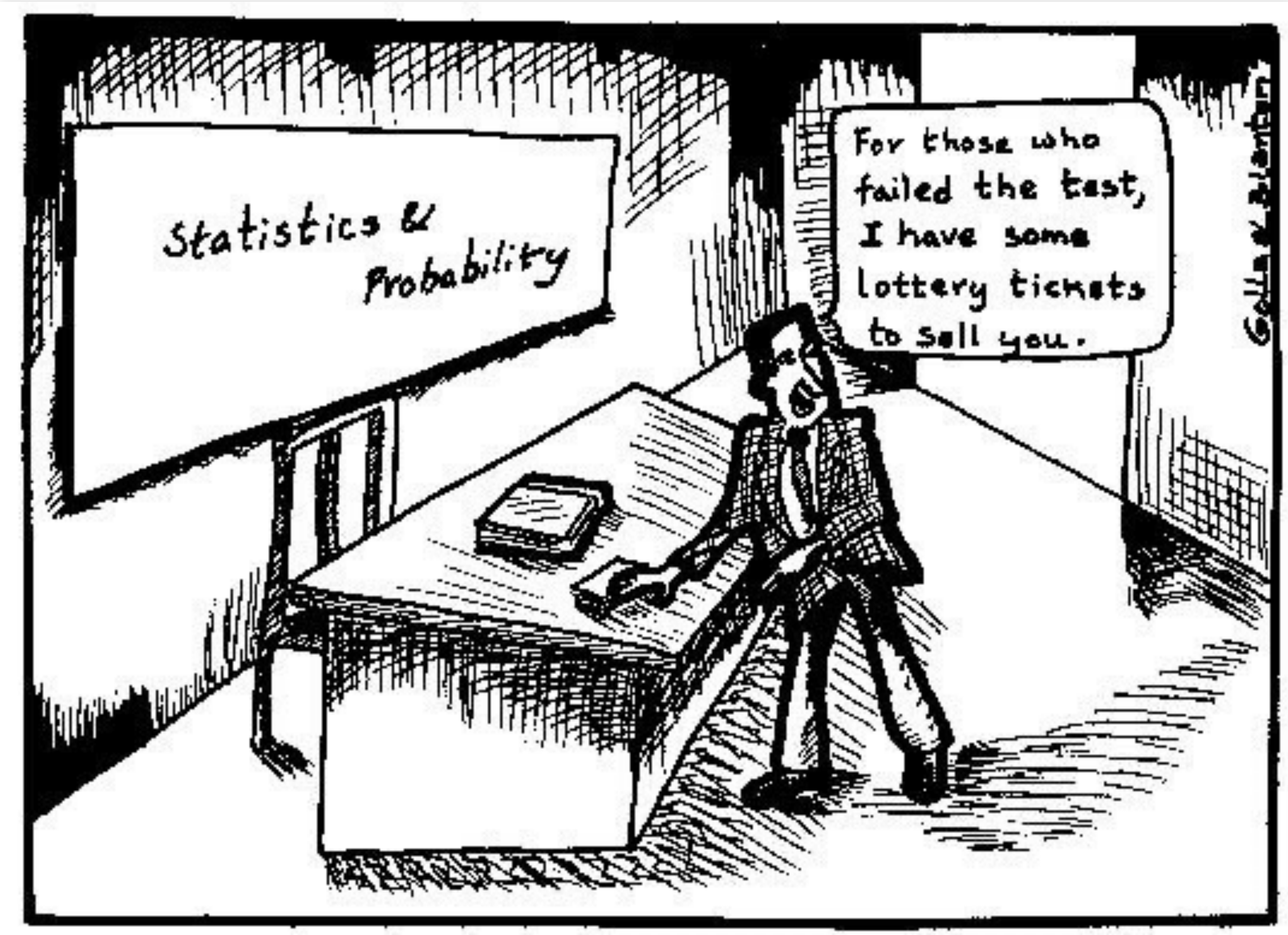
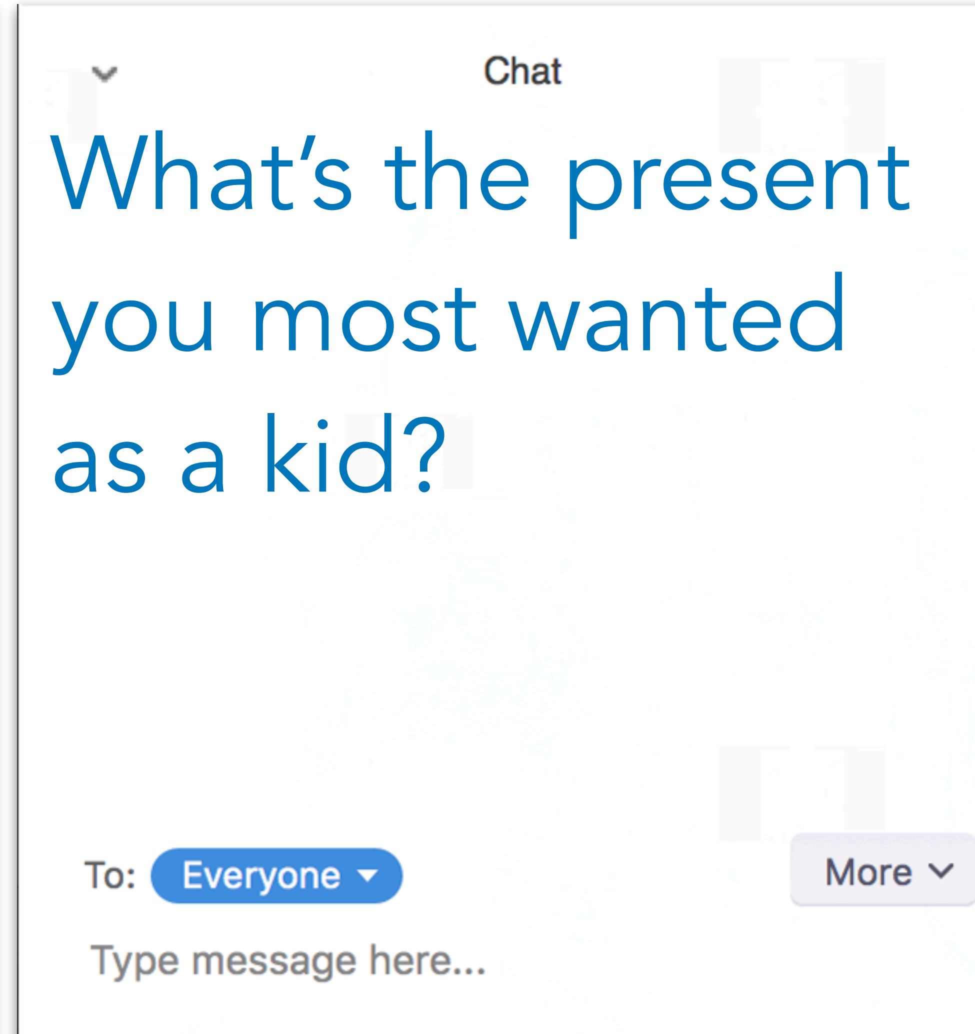
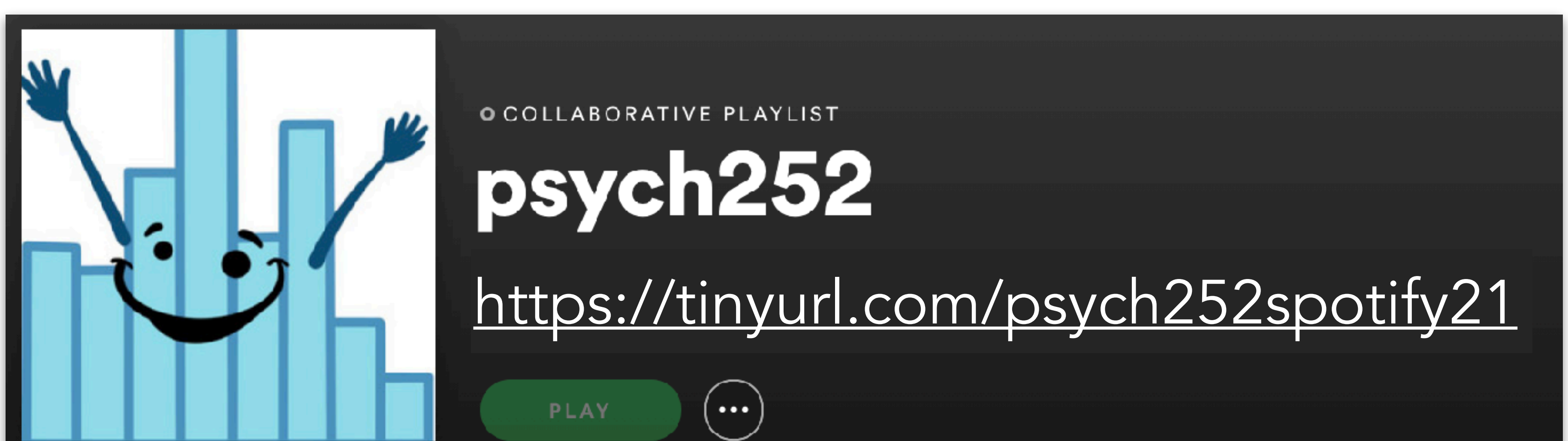
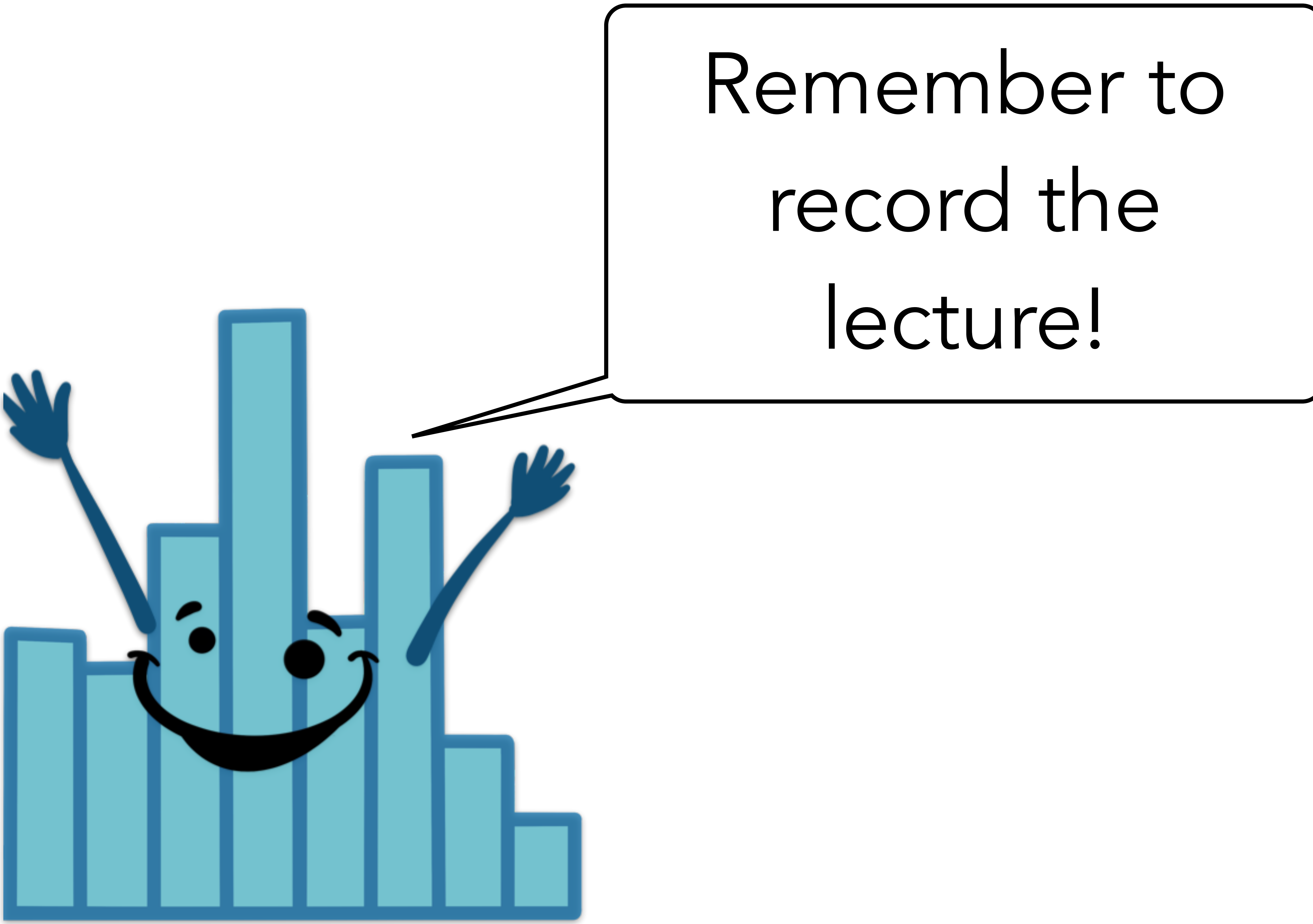
Probability & Causality



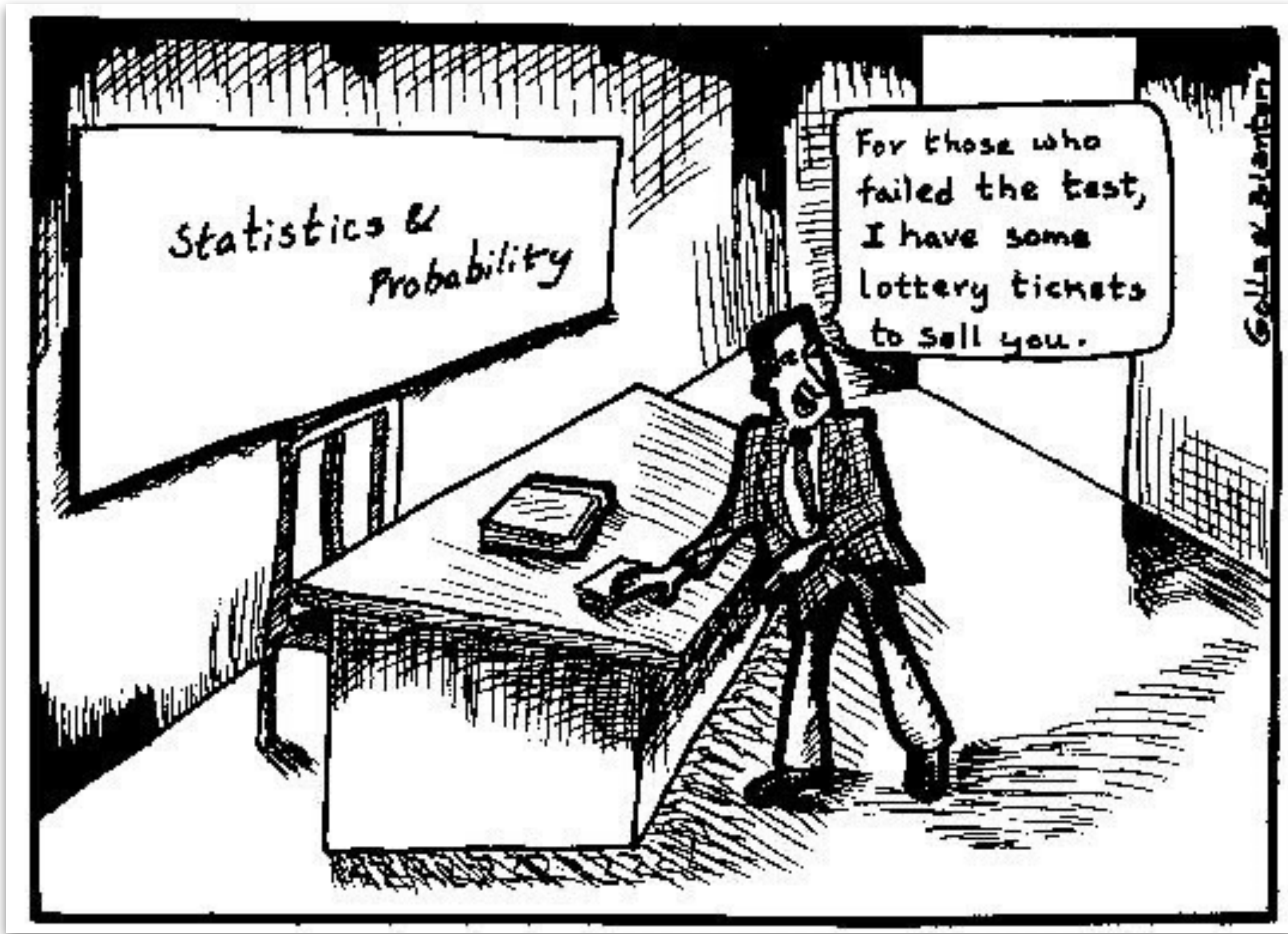




01/25/2021


Remember to record the lecture!

Probability & Causality



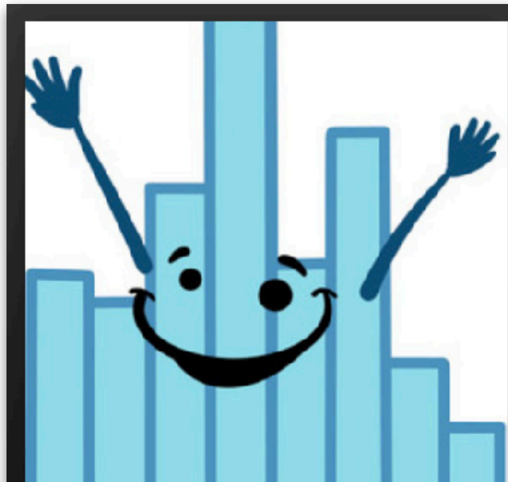
Chat

What's the present you most wanted as a kid?

To: Everyone

More

Type message here...



COLLABORATIVE PLAYLIST

psych252

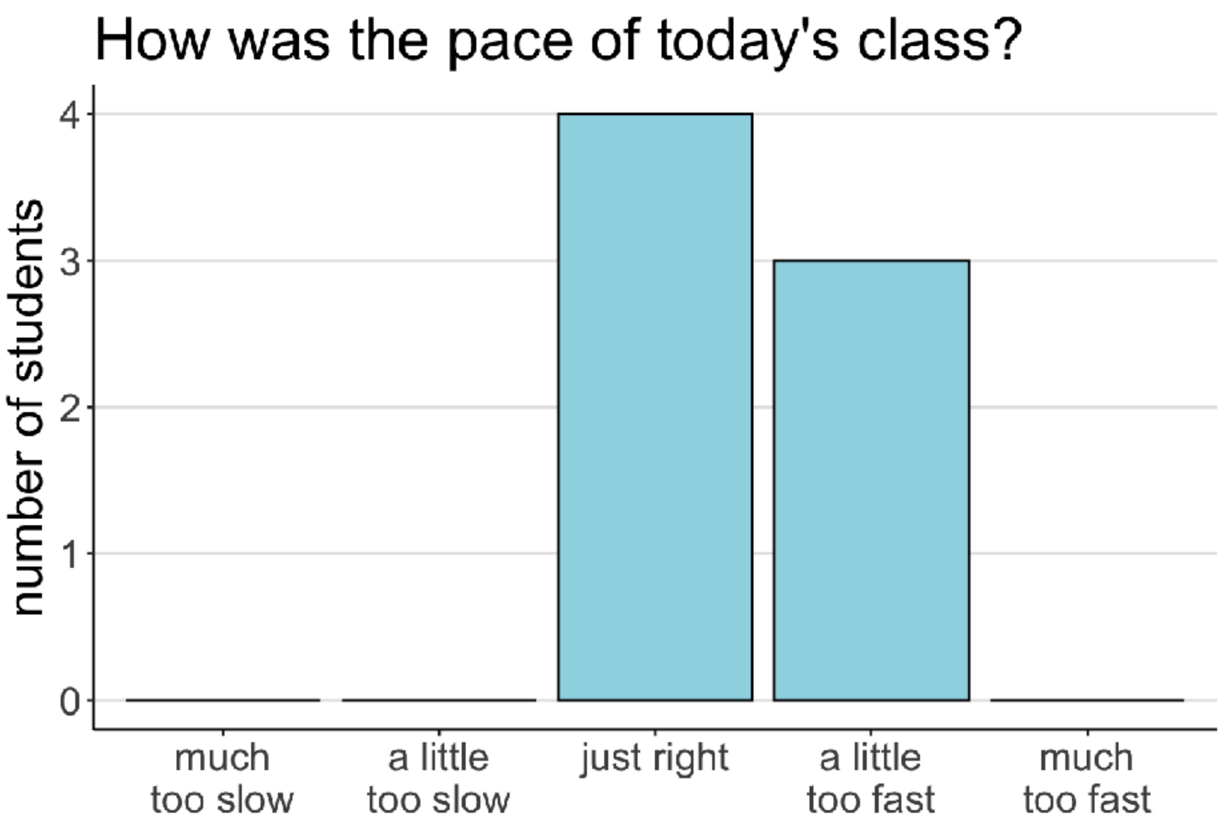
https://tinyurl.com/psych252spotify21

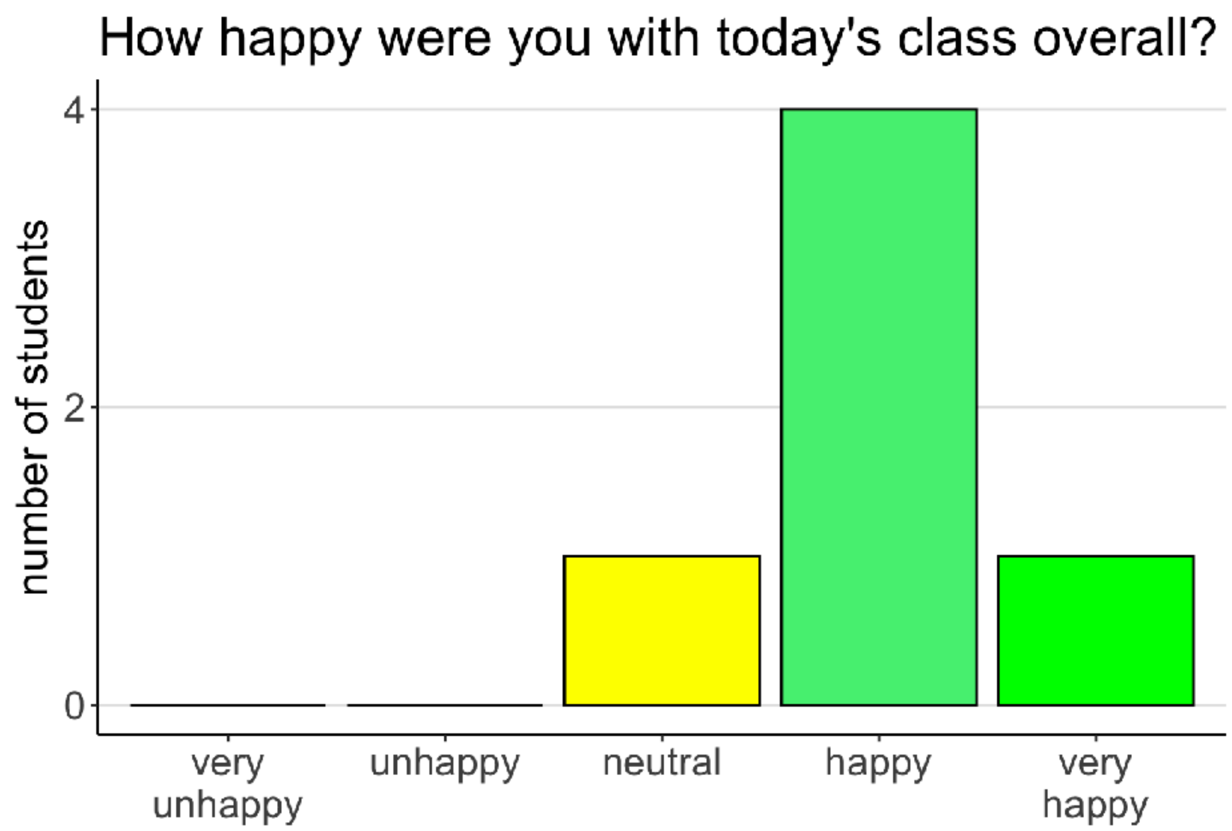
PLAY

01/25/2021

Your feedback

Your feedback





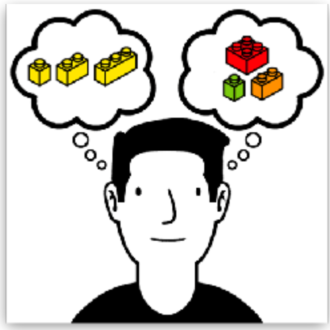
I really liked the pace! The breakout rooms were a little unbalanced so it might be better to ensure equal numbers in all of them for a better learning experience.

Outline

- Introduction to probability / Recap
  - Motivation
  - Counting possibilities
  - Clue guide to probability
- Causal Bayesian Networks
  - seeing (prediction) vs. doing (explanation)
  - correlation is not causation

Motivation

What does statistics have to do with probability?



Theory

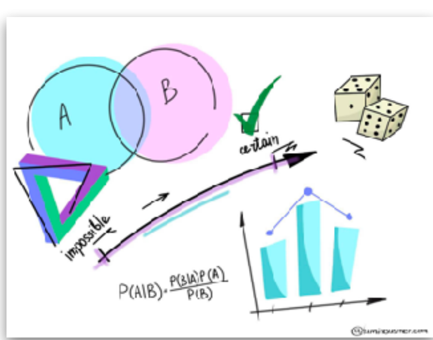
Our goal is to develop theories. In psychology, theories of how the mind works.

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.

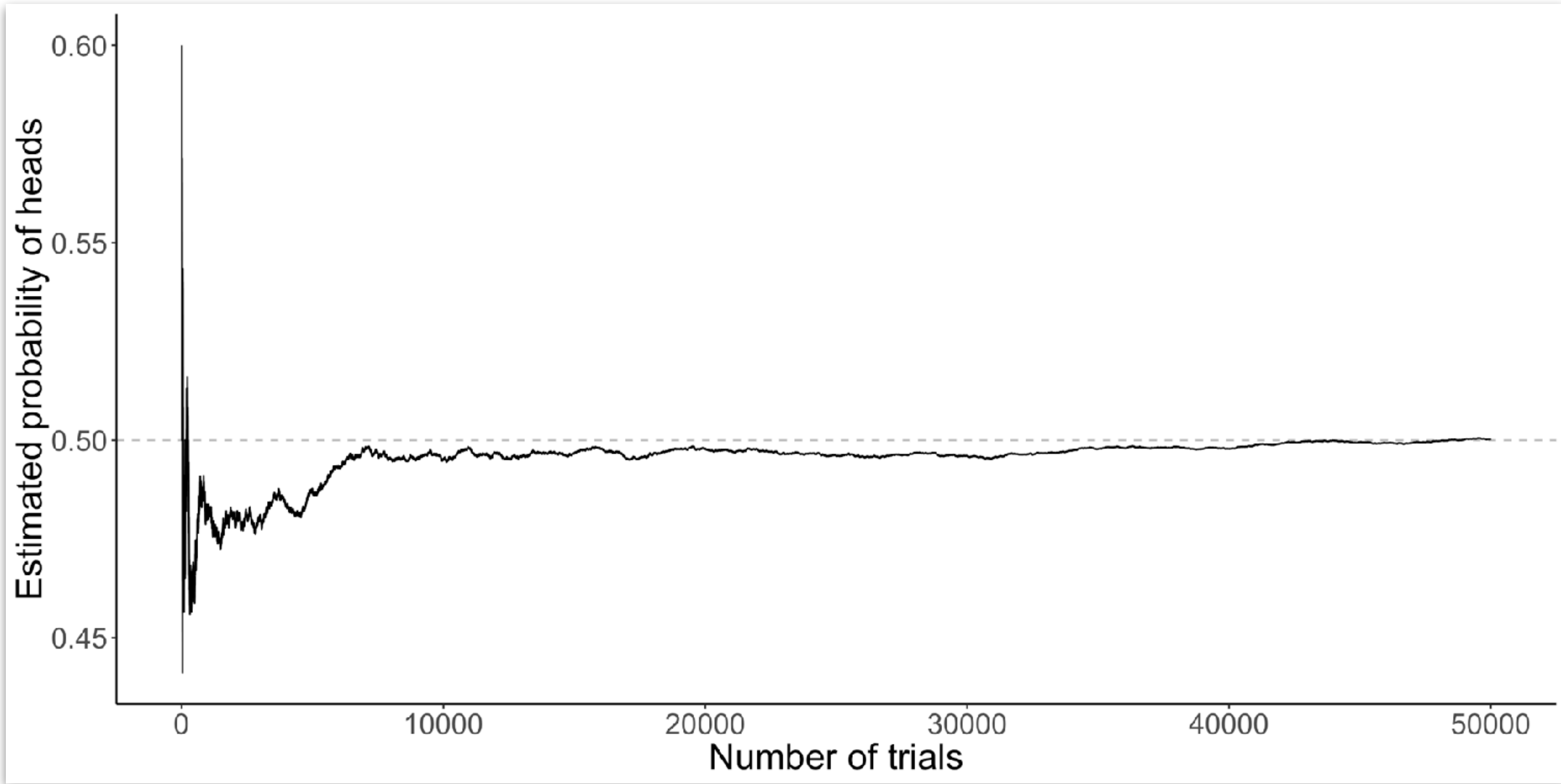


Probability

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

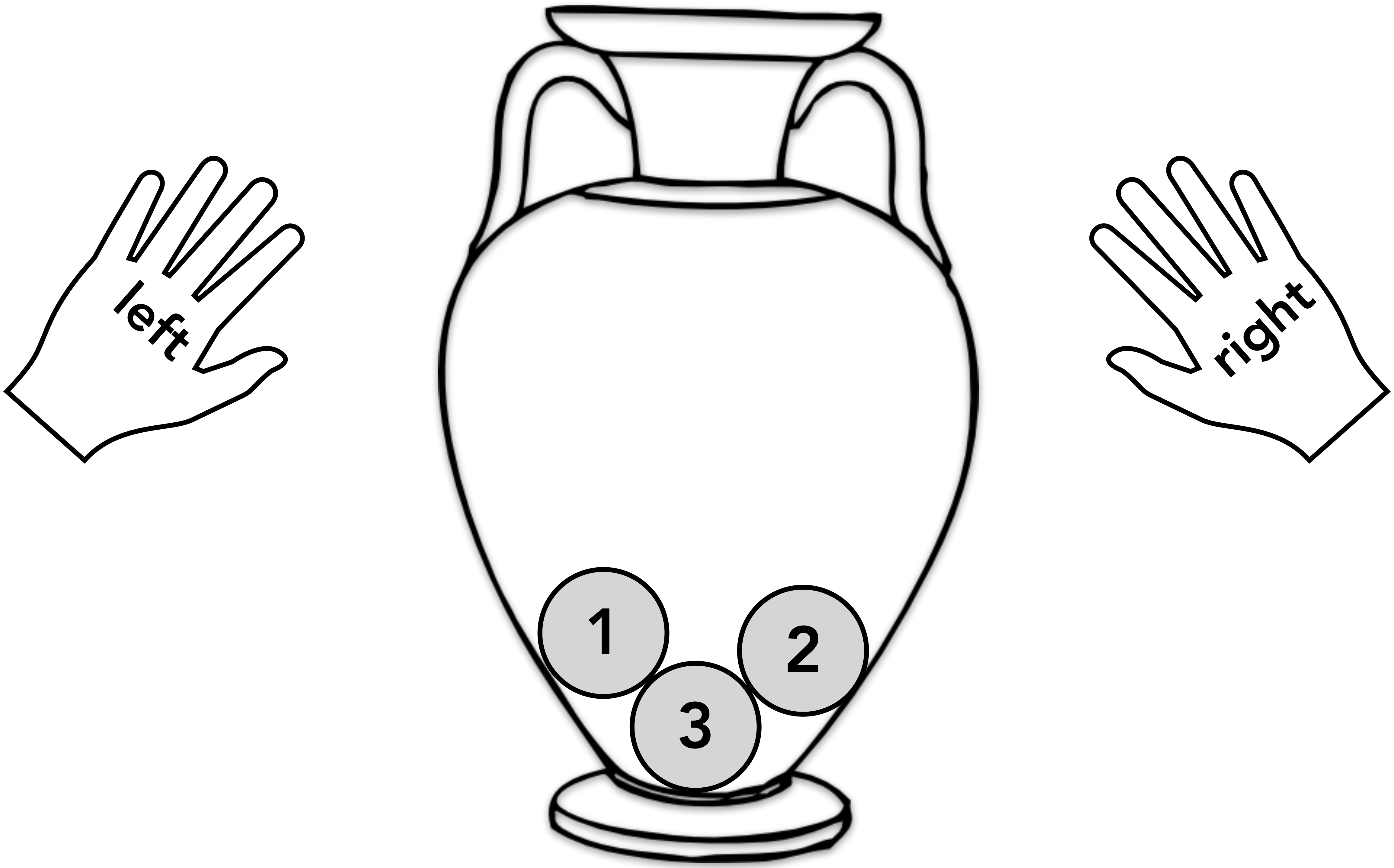
Subjective/Bayesian interpretation

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

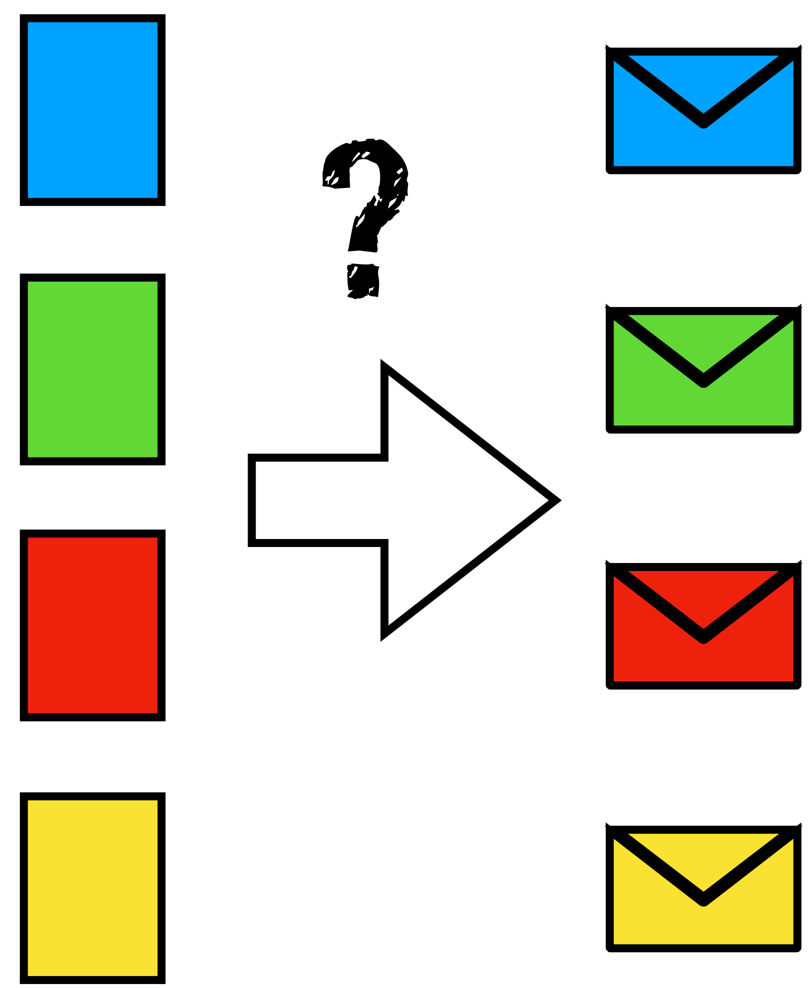
Counting possibilities

no stats class without urns!



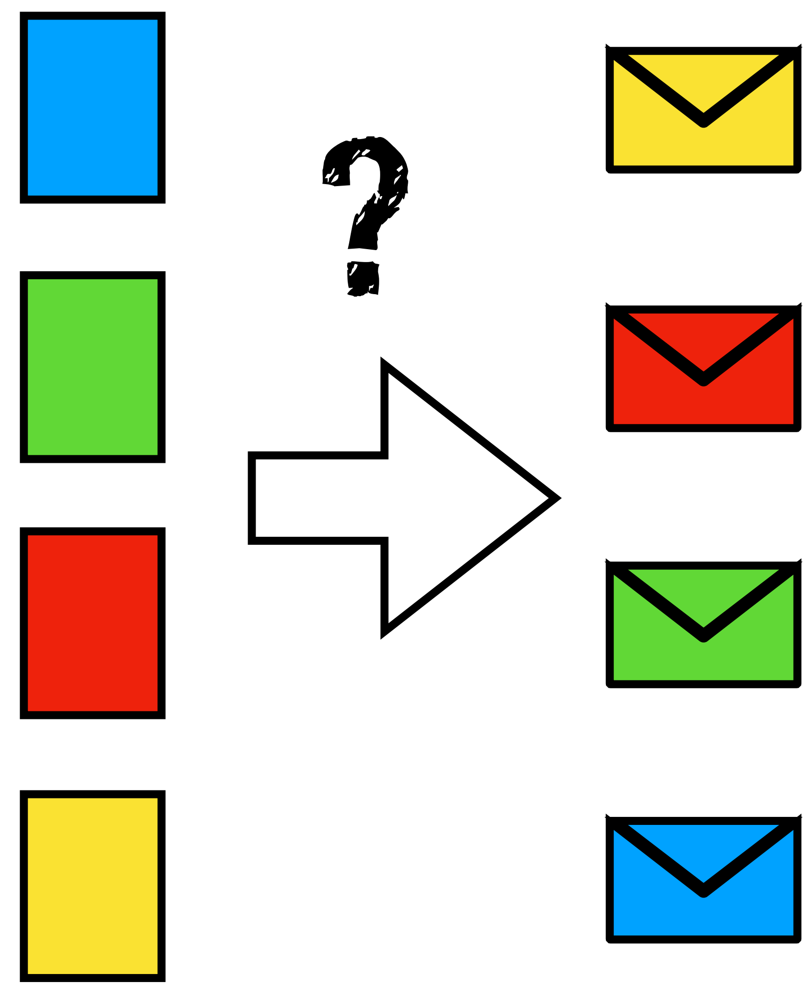
Random secretary

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

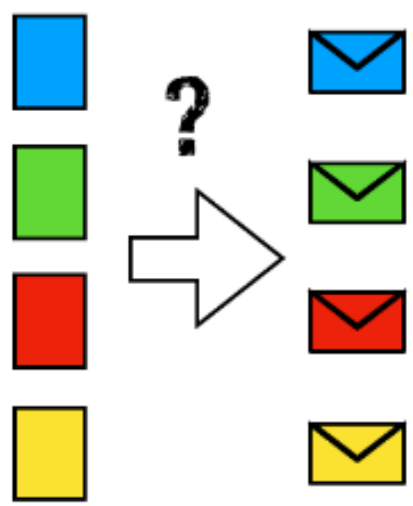
A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?

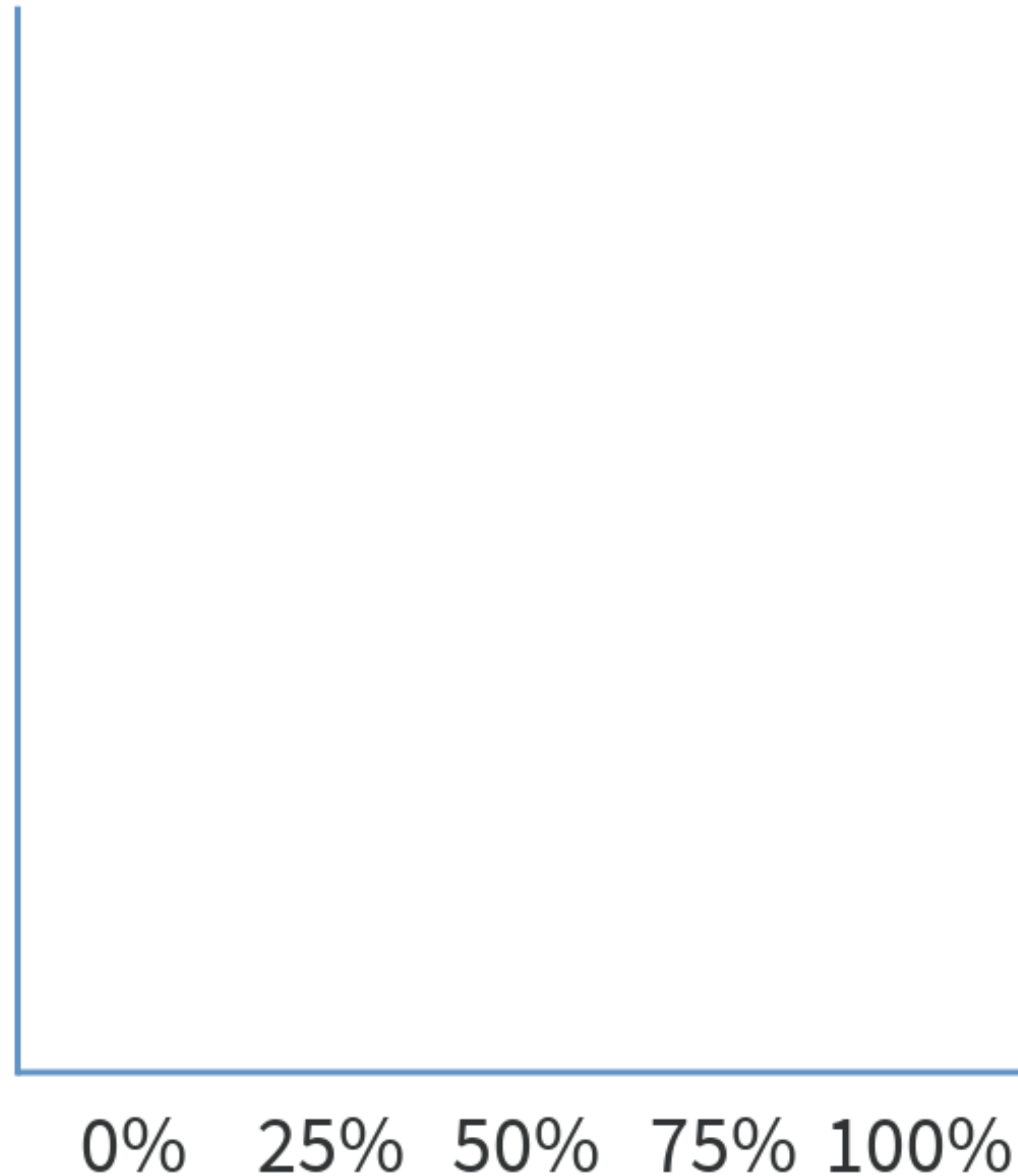


What is the probability that exactly three letters will go into the right envelope?

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?





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!

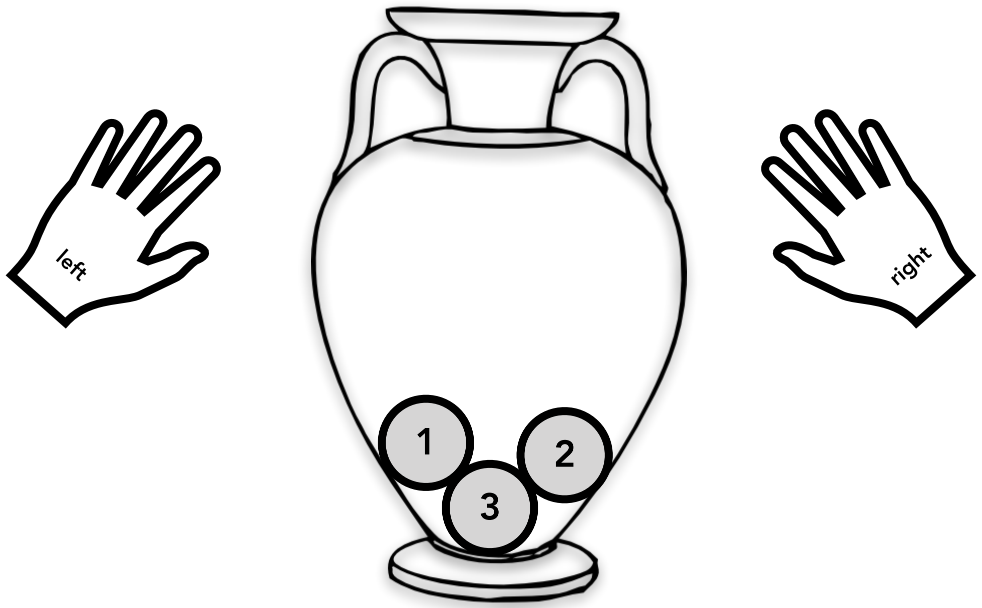
Definitions

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.

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



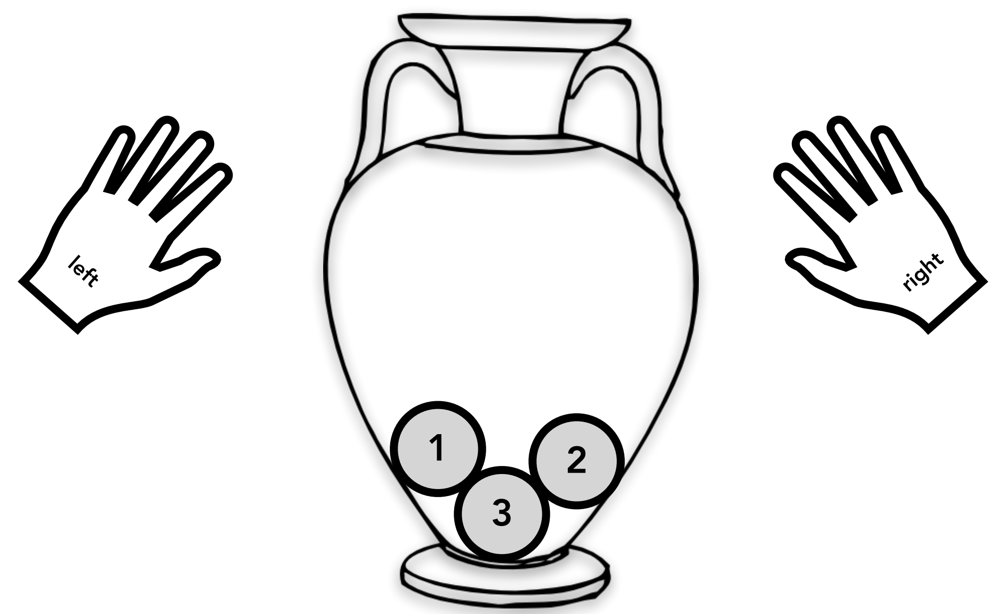
Event: Subset of the sample space. (1, 1)

Definitions

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

1. Probability cannot be negative.

$$P(X_i) \geq 0$$



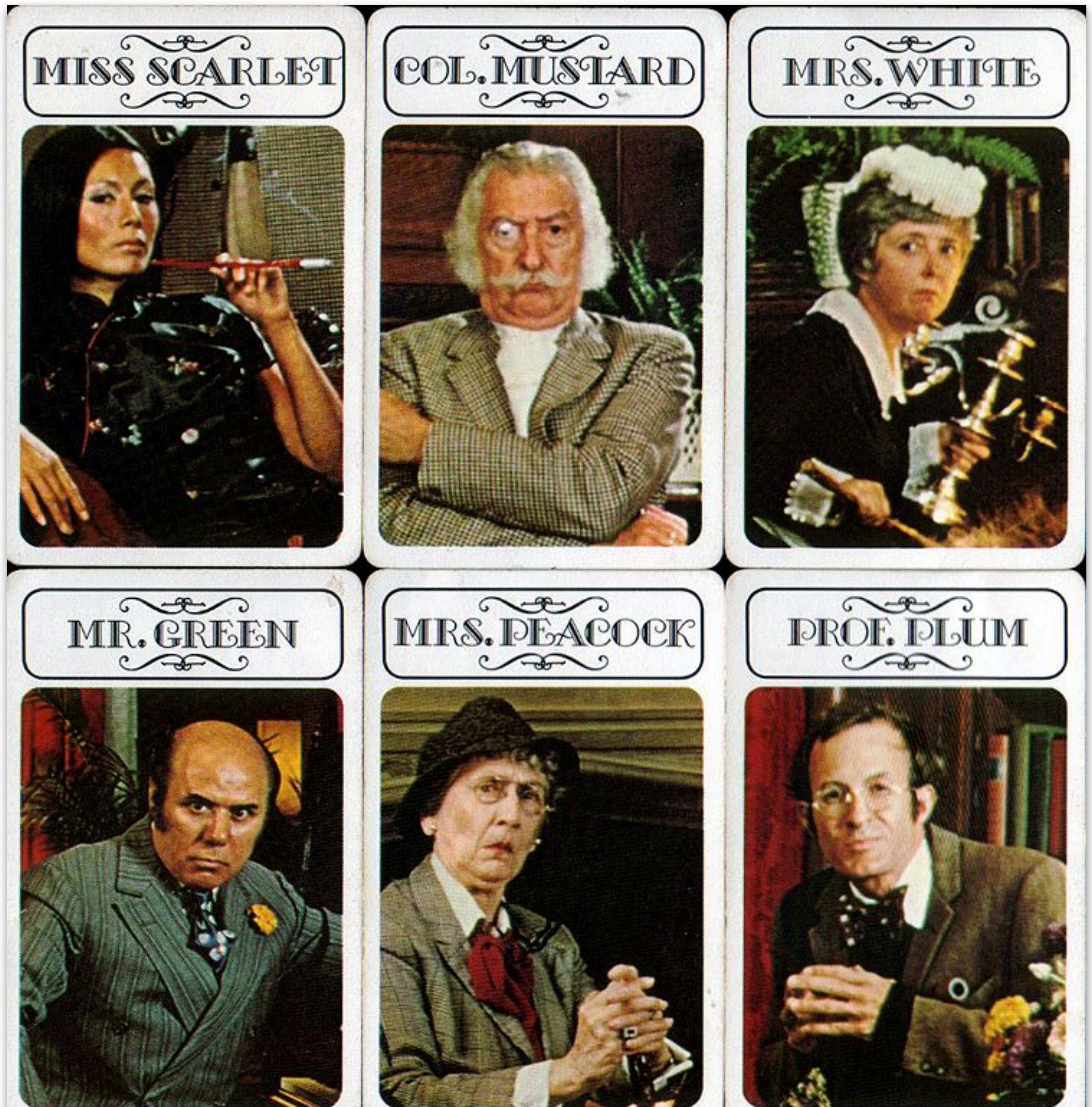
2. Total probability of all outcomes in the sample space is 1.

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

Clue guide to probability

Clue guide to probability

Who killed Mr Boddy?

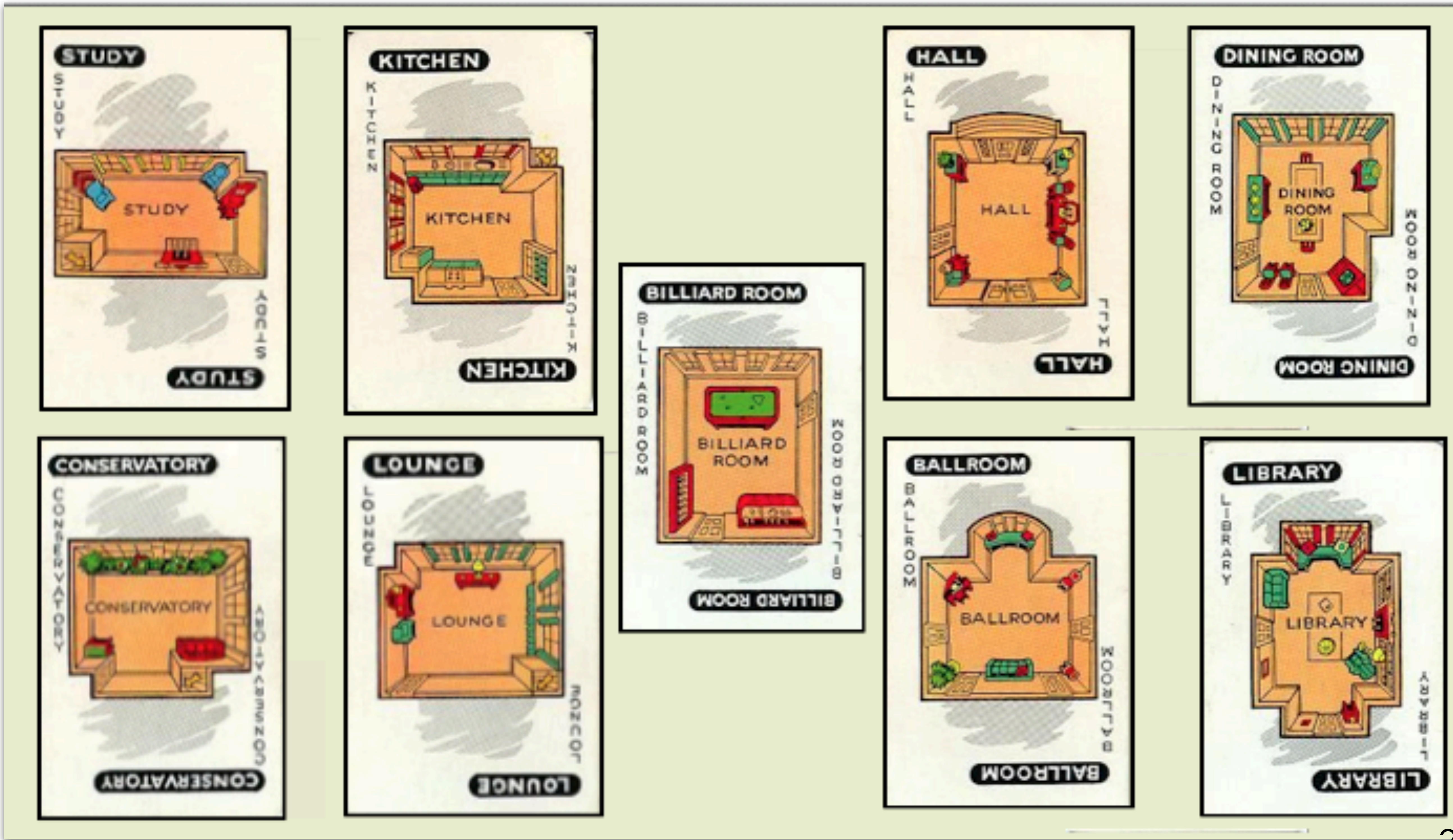


Clue guide to probability

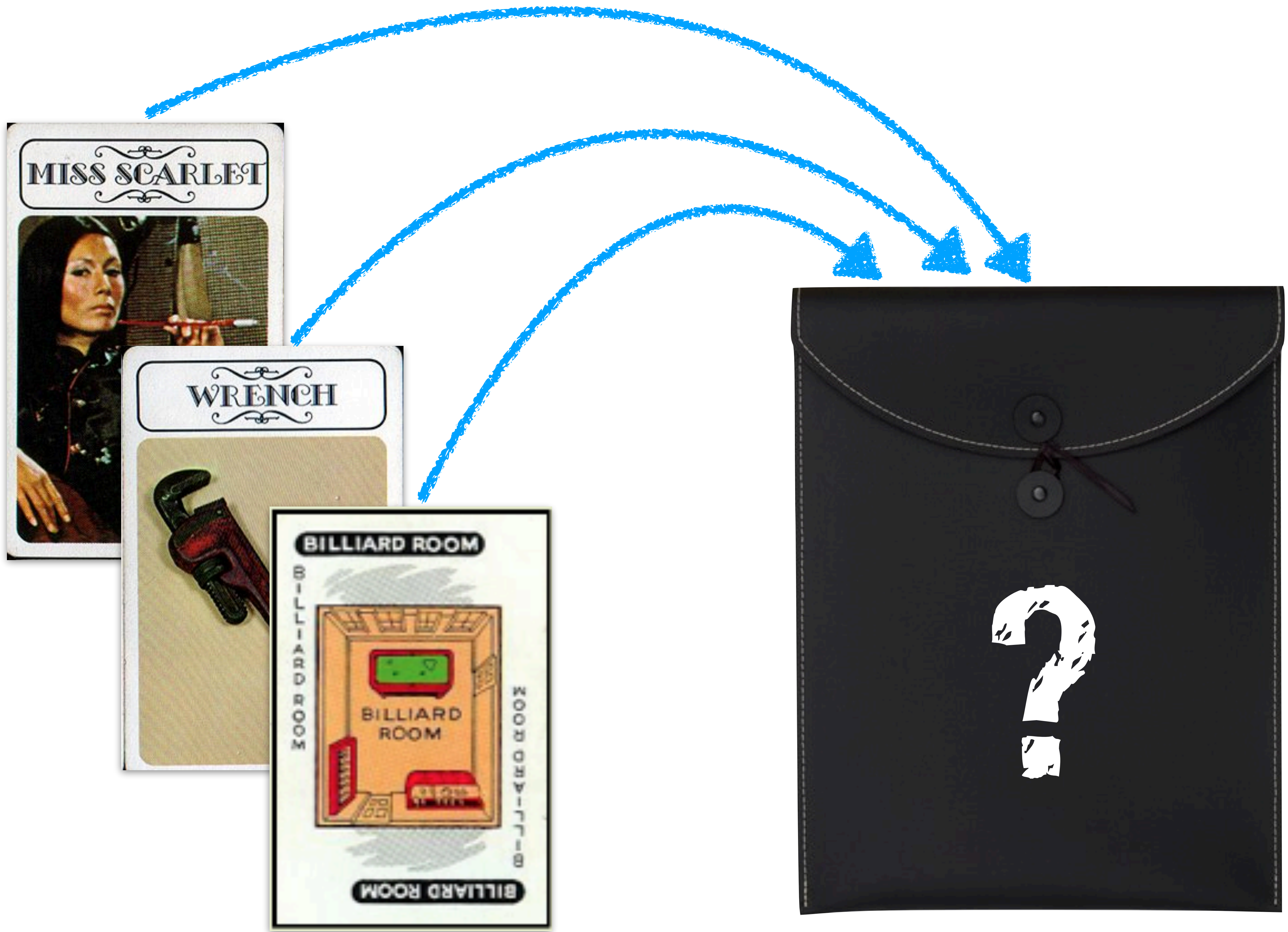
Who killed Mr Boddy, **with what?**

Clue guide to probability

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



Clue guide to probability



Clue guide to probability

```r
who = c("ms_scarlet", "col_mustard", "mrs_white",
        "mr_green", "mrs_peacock", "prof_plum")
what = c("candlestick", "knife", "lead_pipe",
         "revolver", "rope", "wrench")
where = c("study", "kitchen", "conservatory",
          "lounge", "billiard_room", "hall",
          "dining_room", "ballroom", "library")

df.clue = expand_grid(who = who,
                      what = what,
                      where = where)
```

all combinations

Ω

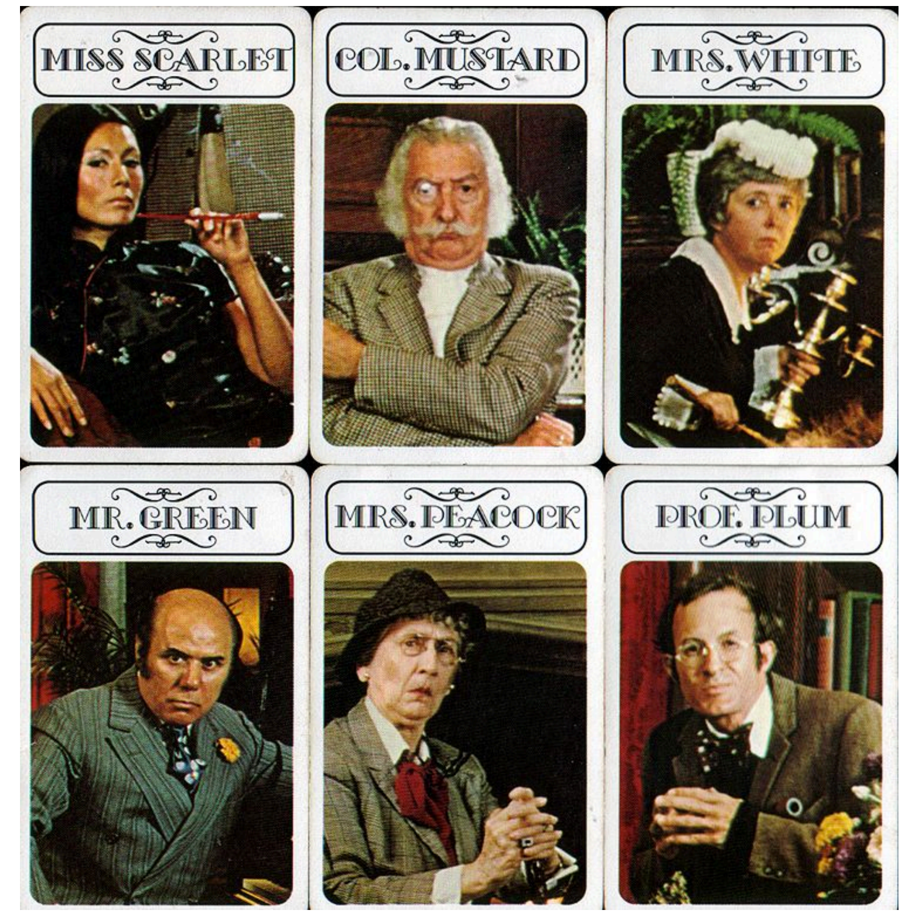
who	what	where
ms_scarlet	candlestick	study
ms_scarlet	candlestick	kitchen
ms_scarlet	candlestick	conservatory
ms_scarlet	candlestick	lounge

⋮

`nrow(df.clue)` = 324

Clue guide to probability

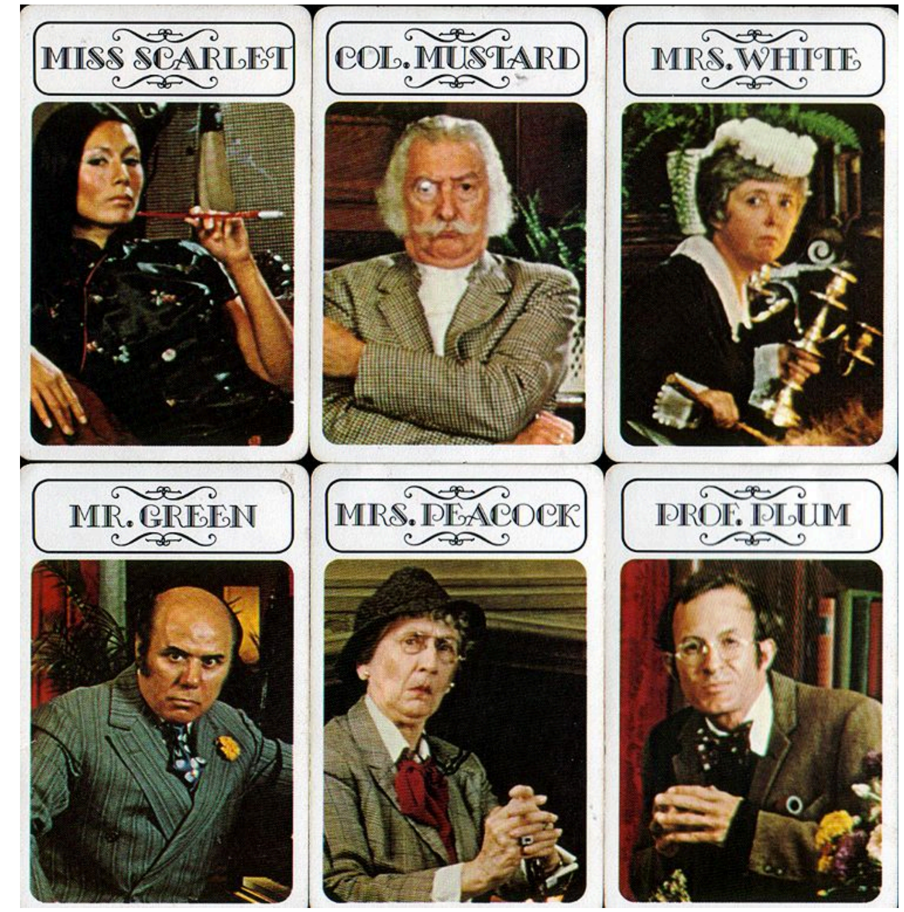
Who?



- 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?

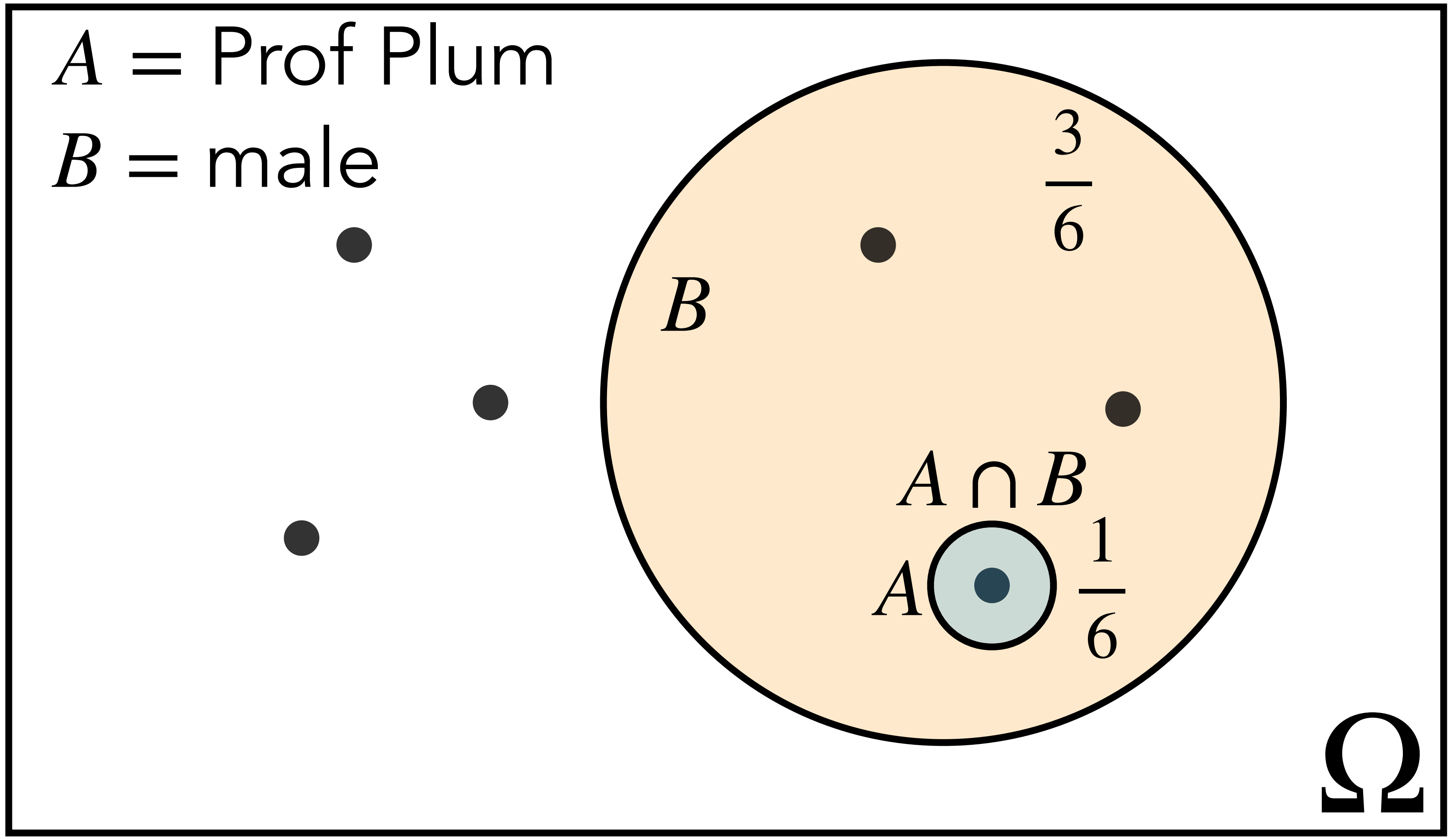


- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A,B)}{p(B)}$

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

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

Clue guide to probability



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

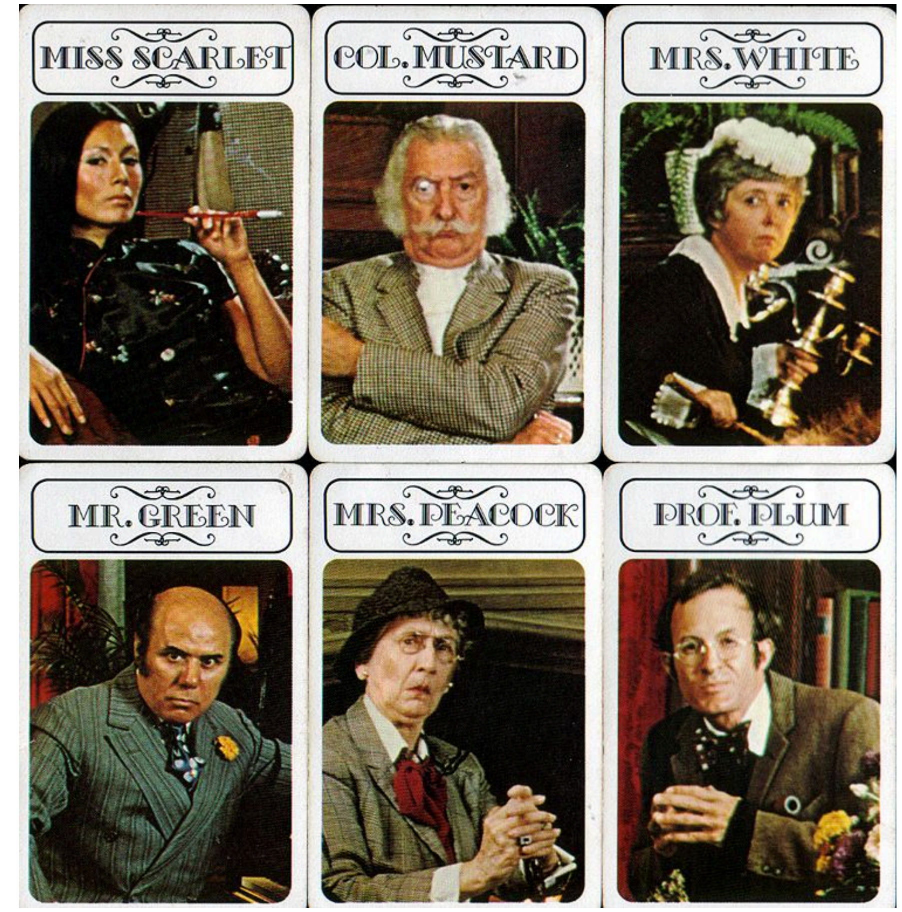
Definition: $p(A \mid B) = \dfrac{p(A,B)}{p(B)} = \dfrac{1}{3}$

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

Clue guide to probability

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

Who?

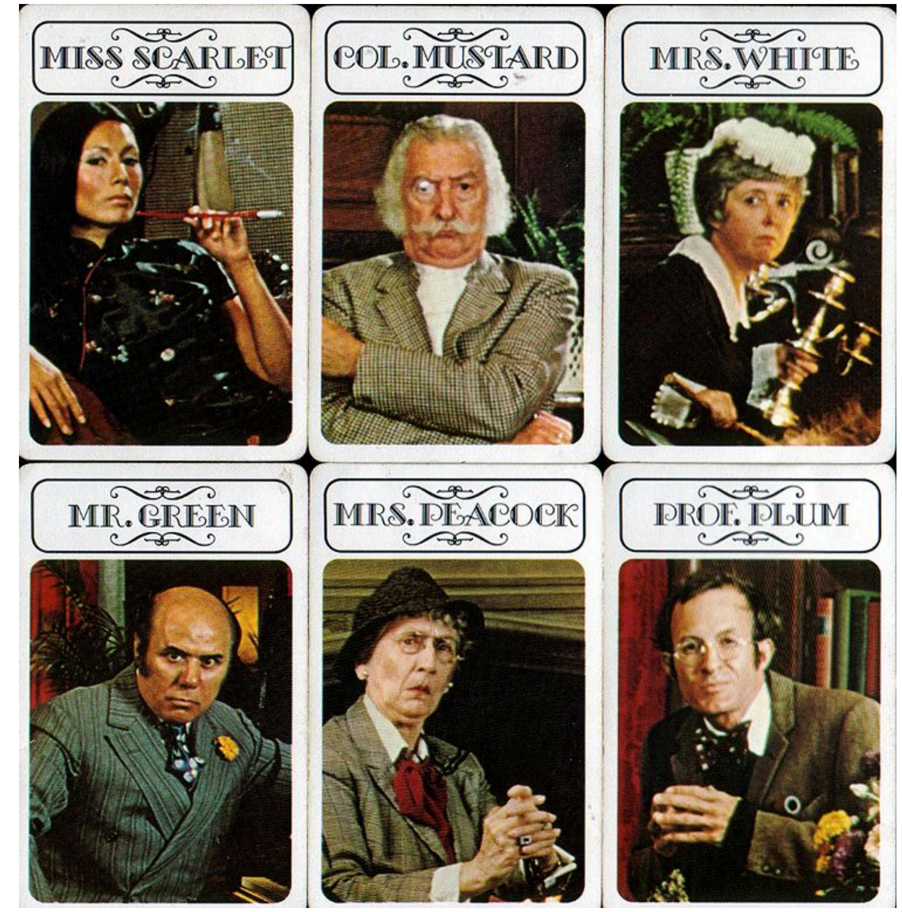


who	gender
col_mustard	male
mr_green	male
prof_plum	male
ms_scarlet	female
mrs_white	female
mrs_peacock	female

```
df.suspects = df.clue %>%
  distinct(who) %>%
  mutate(gender = ifelse(
    test = who %in% c("ms_scarlet",
                      "mrs_white",
                      "mrs_peacock"),
    yes = "female",
    no = "male"))
```

Clue guide to probability

Who?



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

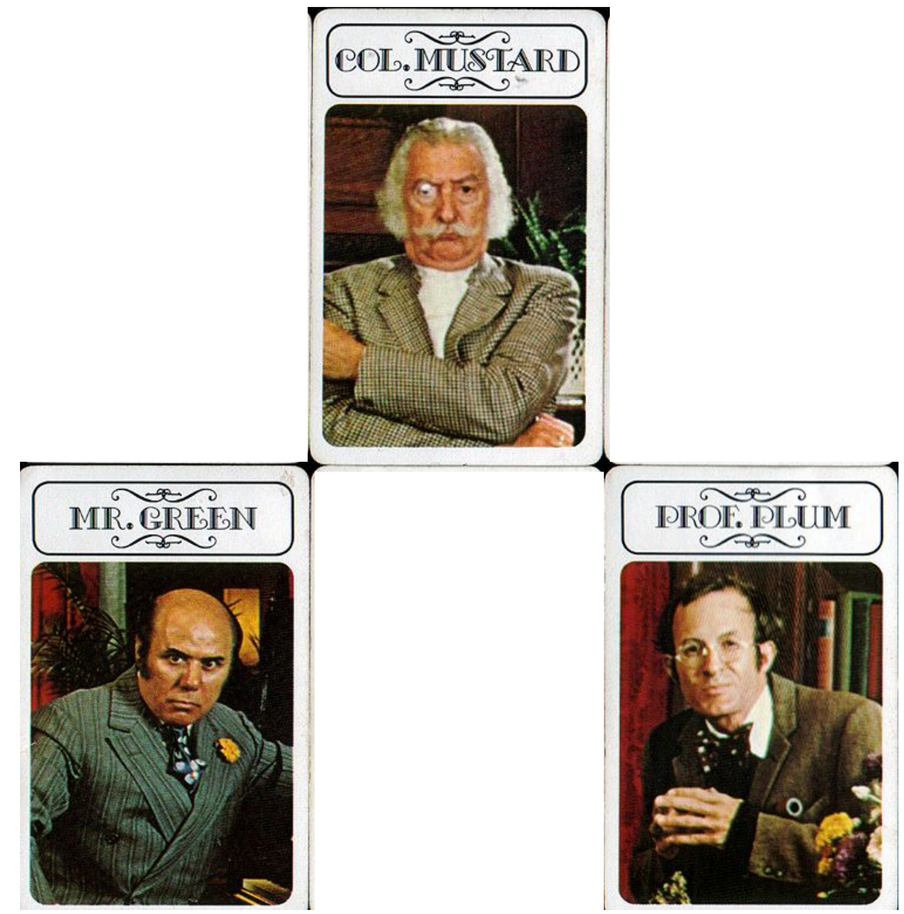
who	gender
col_mustard	male
mr_green	male
prof_plum	male
ms_scarlet	female
mrs_white	female
mrs_peacock	female

```
df.suspects %>%
  summarize(p_prof_plum_given_male =
              sum(gender == "male" &
                  who == "prof_plum") /
              sum(gender == "male"))
```

use naive definition of probability

Clue guide to probability

Who?



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

who	gender
col_mustard	male
mr_green	male
prof_plum	male

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

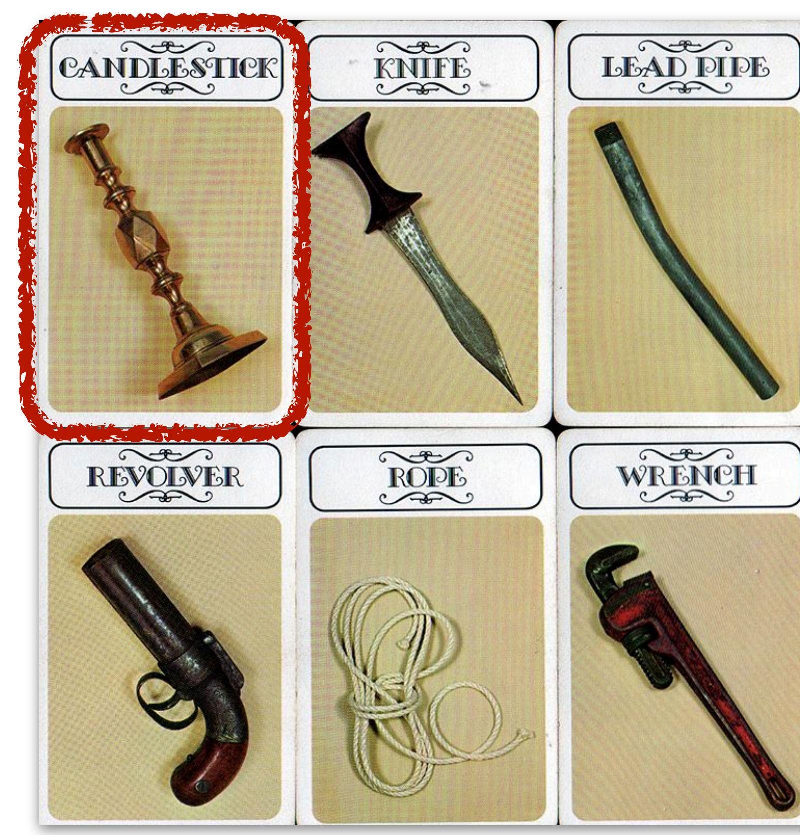
Clue guide to probability

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

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

Who?

What?



Clue guide to probability

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

- $p(\text{Prof Plum}, \text{candle stick}) =$
  $p(\text{Prof Plum}) \cdot p(\text{candle stick}) =$
  $$\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Who?

What?

Clue guide to probability

- *dependence*:
- **Definition**: $p(A \mid B) \neq p(A)$
- **Definition**: $p(A, B) = p(A) \cdot p(B \mid A)$

- if women were more likely than men to use the revolver then
- $p(\text{Mrs. White} \mid \text{Revolver}) > p(\text{Mrs. White})$

Who?

What?

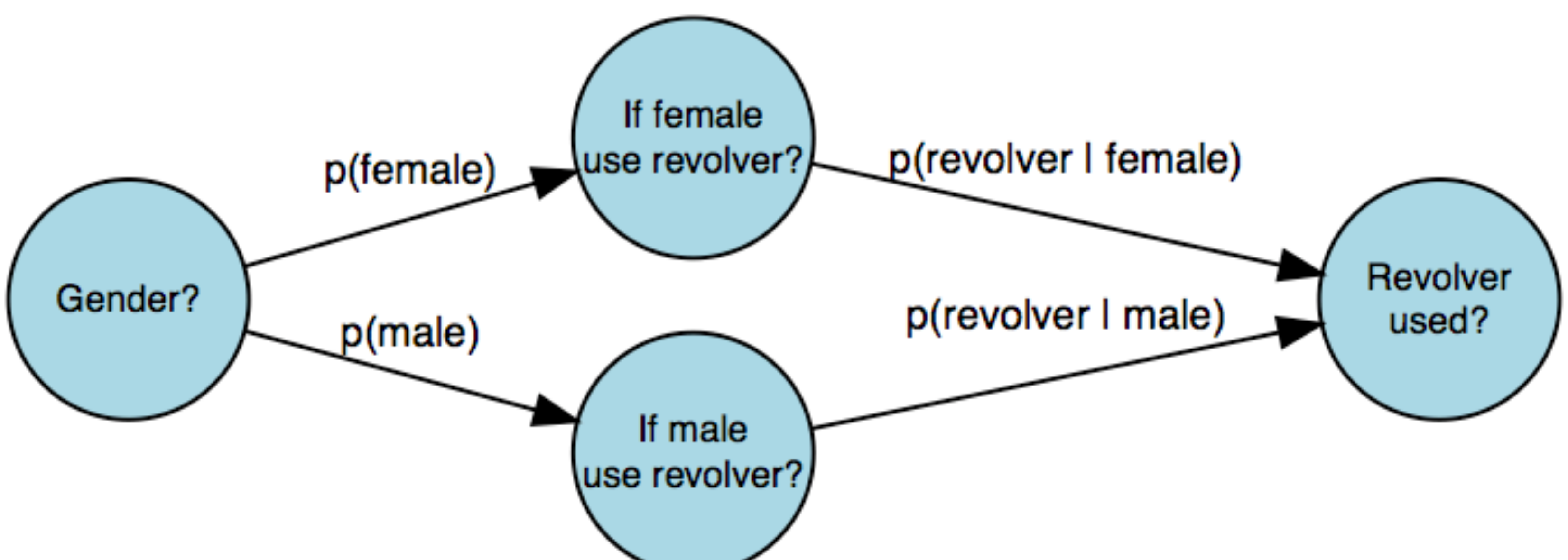
Clue guide to probability

- *law of total probability*
- **Definition**:

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

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

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



Who?

What?

Clue guide to probability

Bayes Theorem in a few steps

Clue guide to probability

- *Bayes' theorem (derivation)*

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

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

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

Clue guide to probability

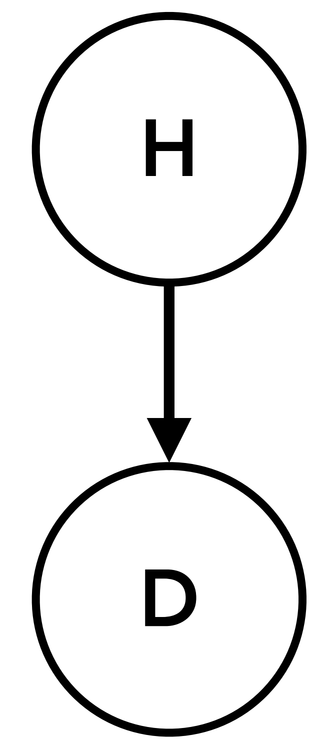
$$p(B\,|\,A) = \frac{p(A\,|\,B)\cdot p(B)}{p(A)}$$

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

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

posterior

likelihood prior

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

probability of the data?!

H = Hypothesis

D = Data

law of total probability:

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

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 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 $\mathbf{P(T|D) = 0.95}$ and $\mathbf{P(\neg T|\neg D) = 0.95}$. 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).

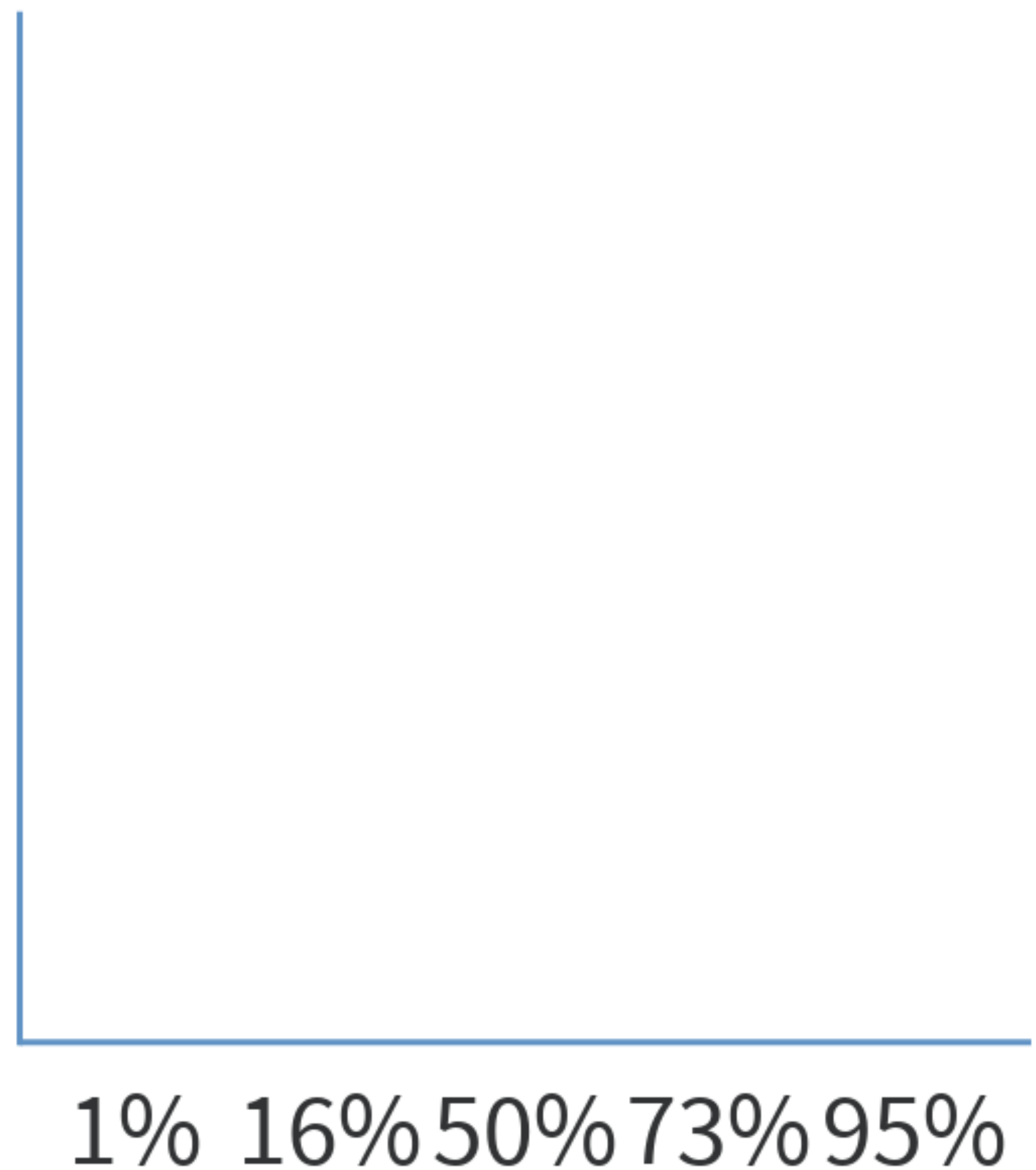
Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

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 $\mathbf{P(T|D) = 0.95}$ and $\mathbf{P(\neg T|\neg D) = 0.95}$. 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.



Breakout rooms

Tasks: 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 $\mathbf{P(T|D) = 0.95}$ and $\mathbf{P(\neg T|\neg D) = 0.95}$. 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.

Size: ~3 people

Time: 5 minutes

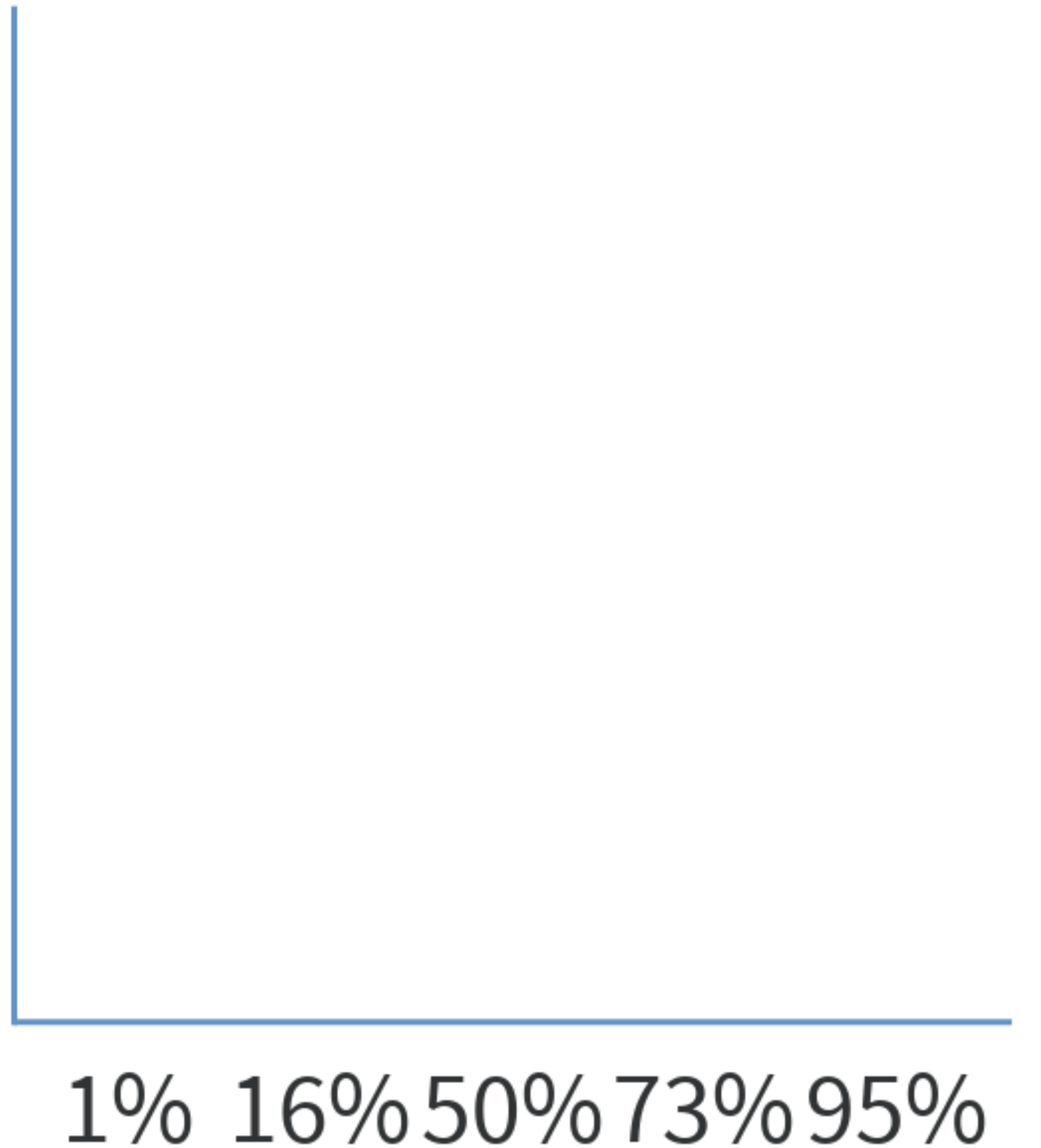
Report: We will vote again.

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).

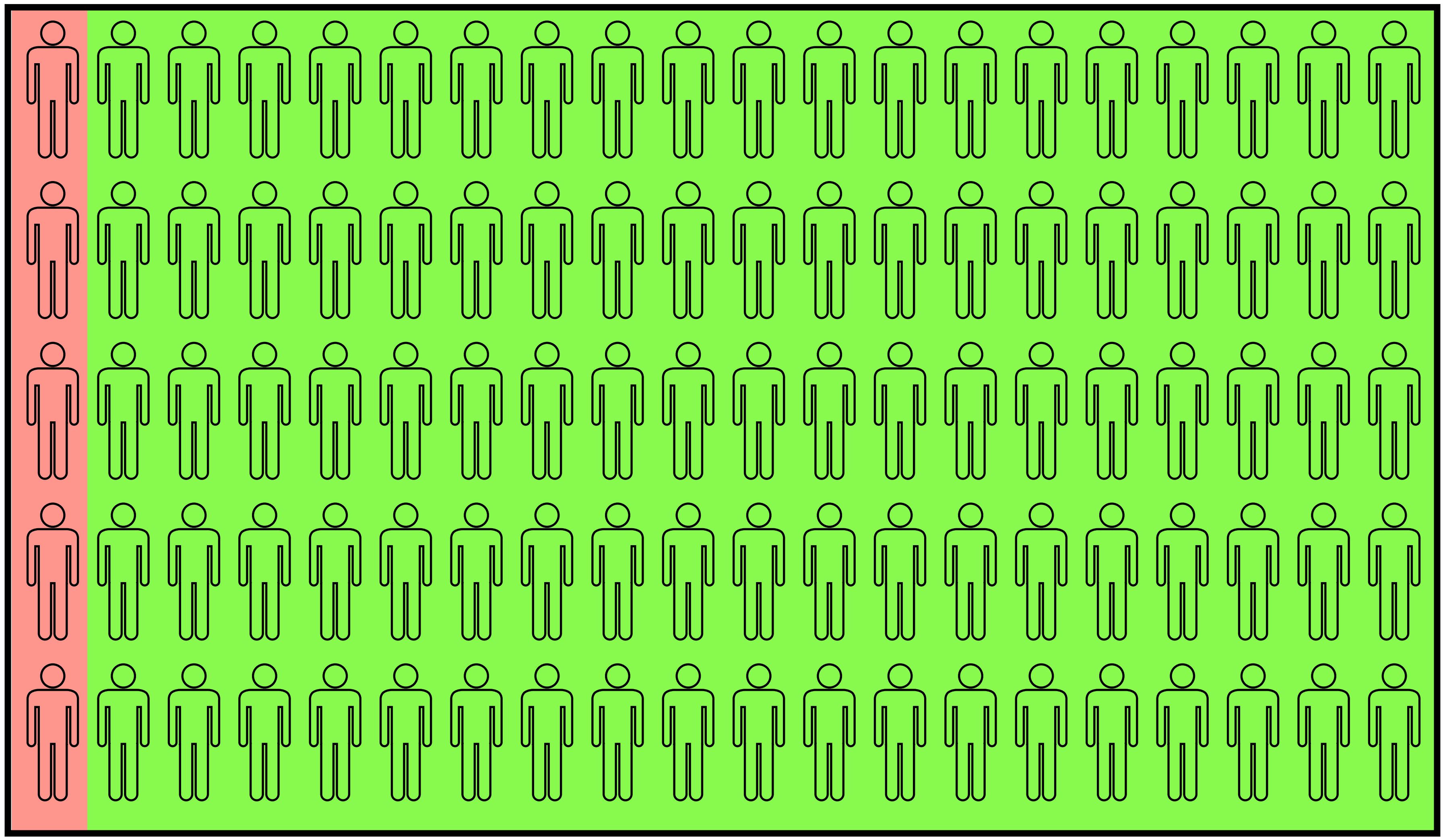
Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.



Clue guide to probability

this example uses different probabilities than the conditionitis one

n = 100



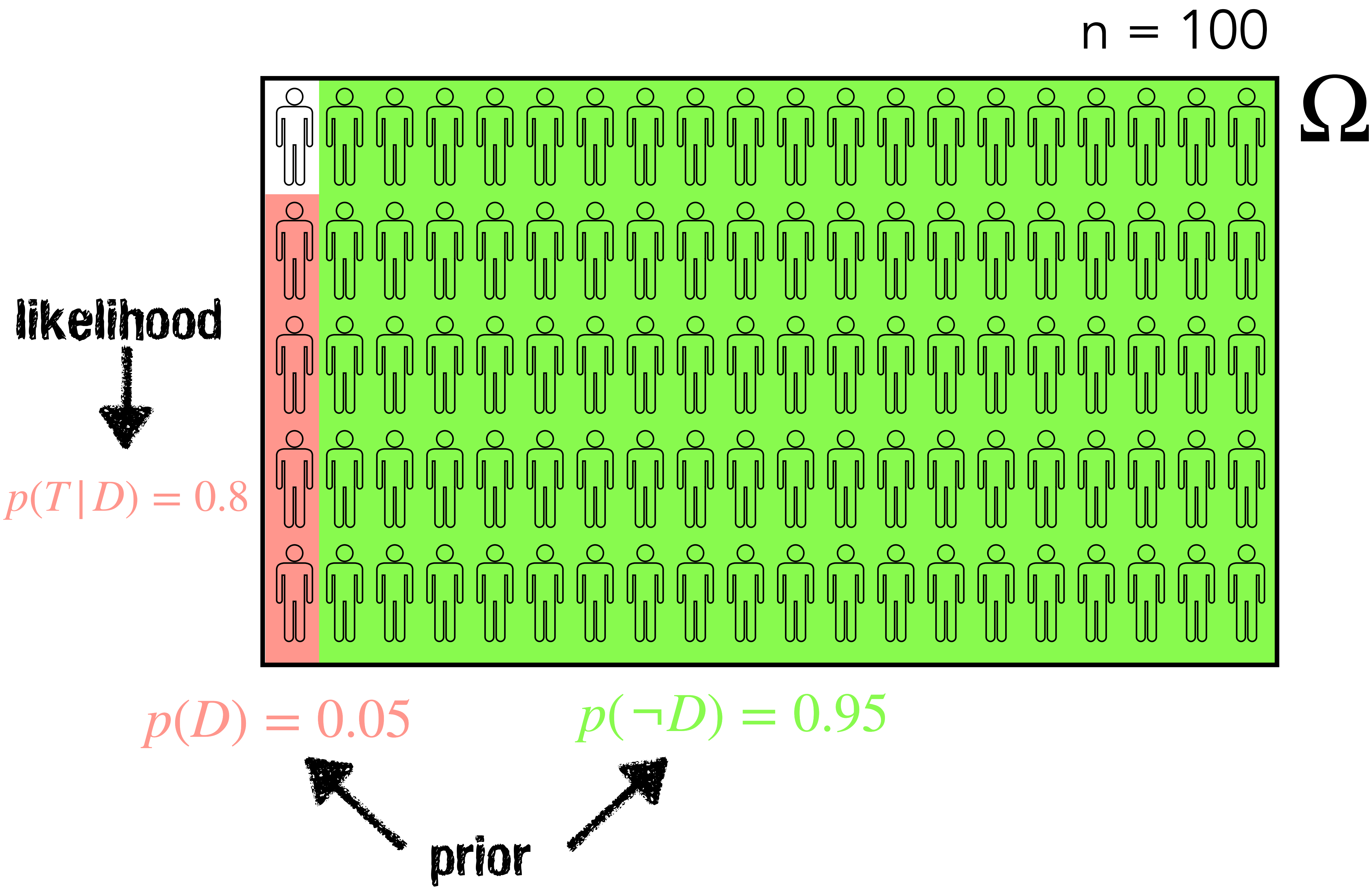
Ω

$p(D) = 0.05$

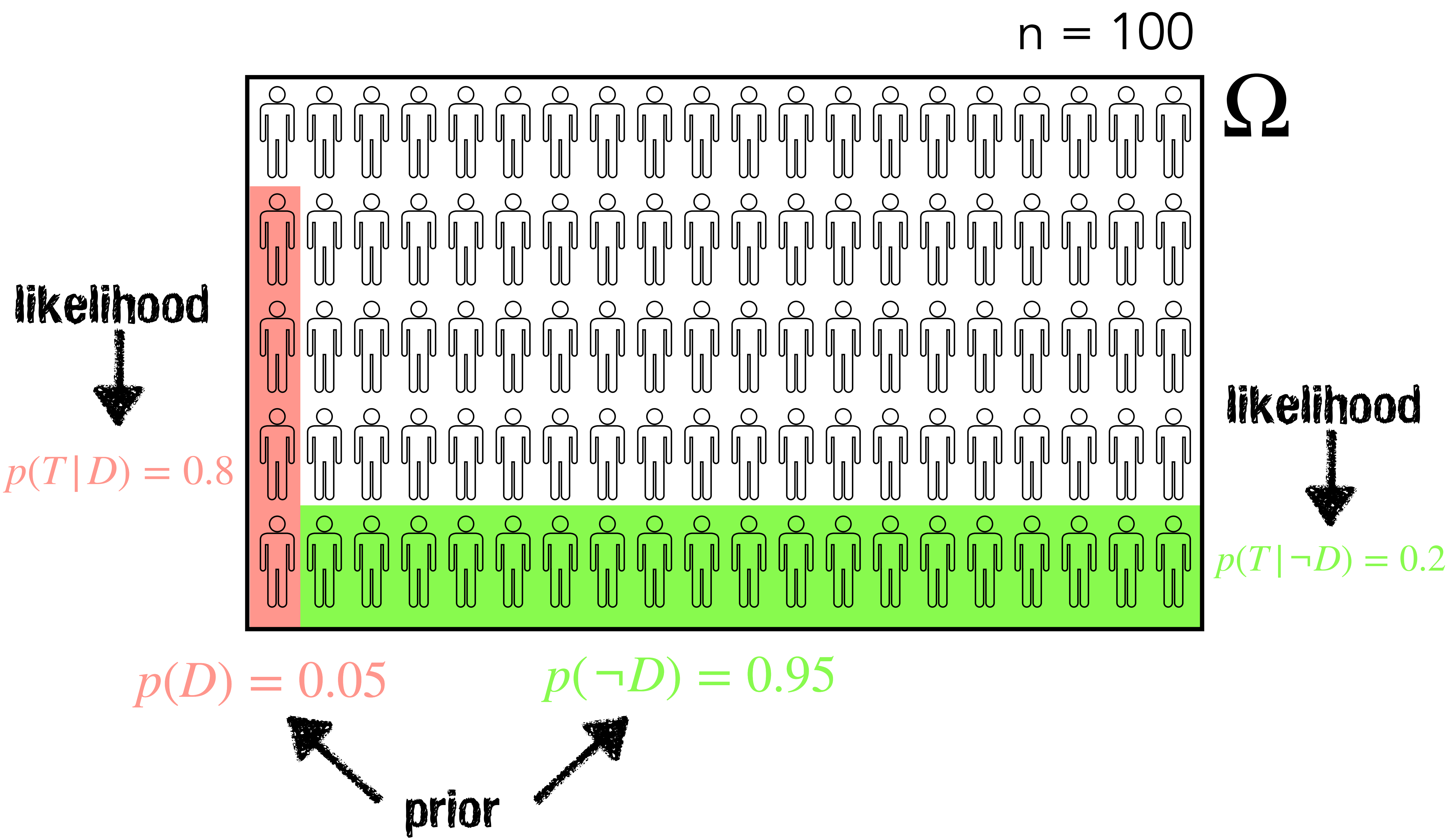
$p(\neg D) = 0.95$

prior

Clue guide to probability



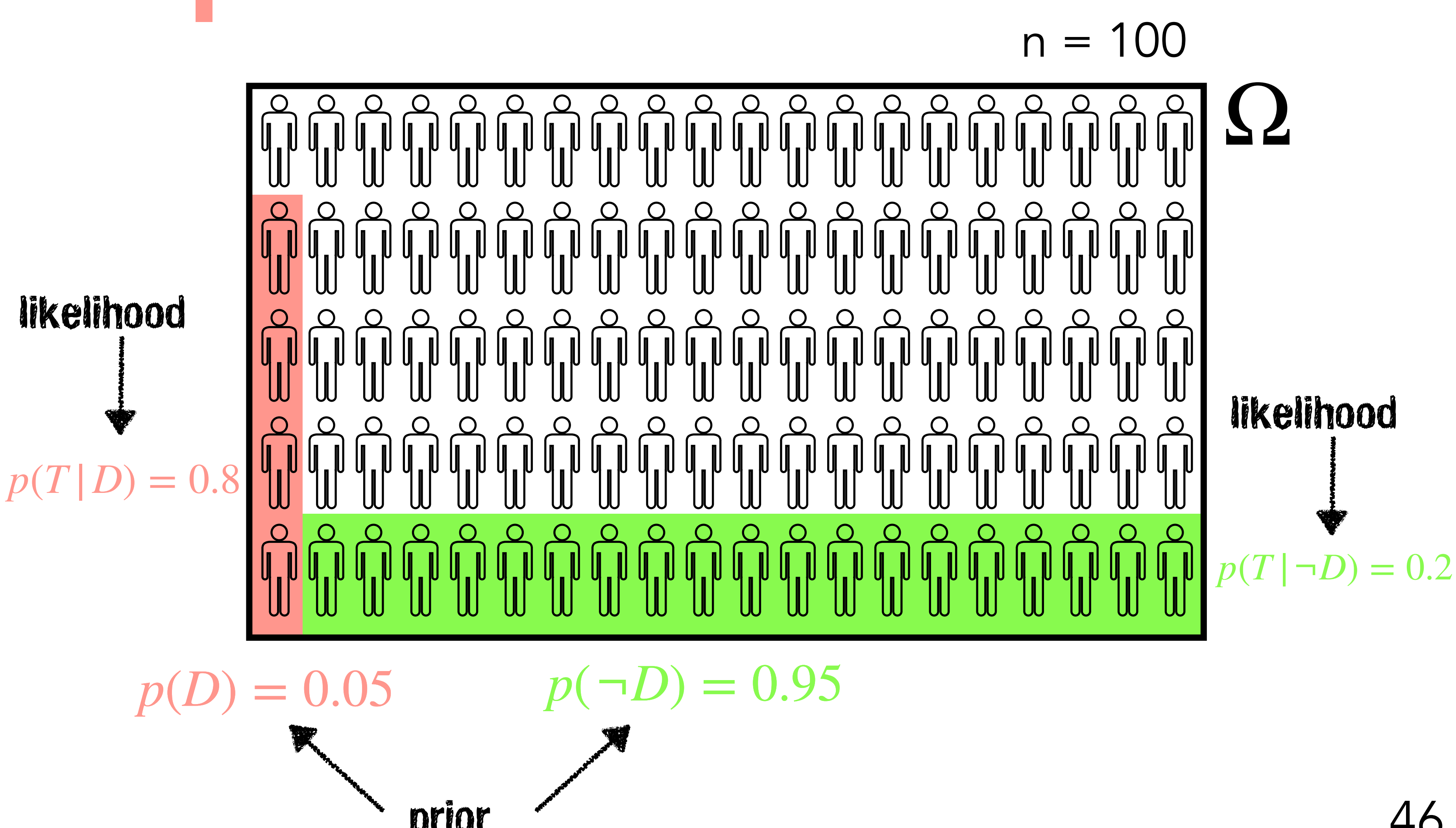
Clue guide to probability

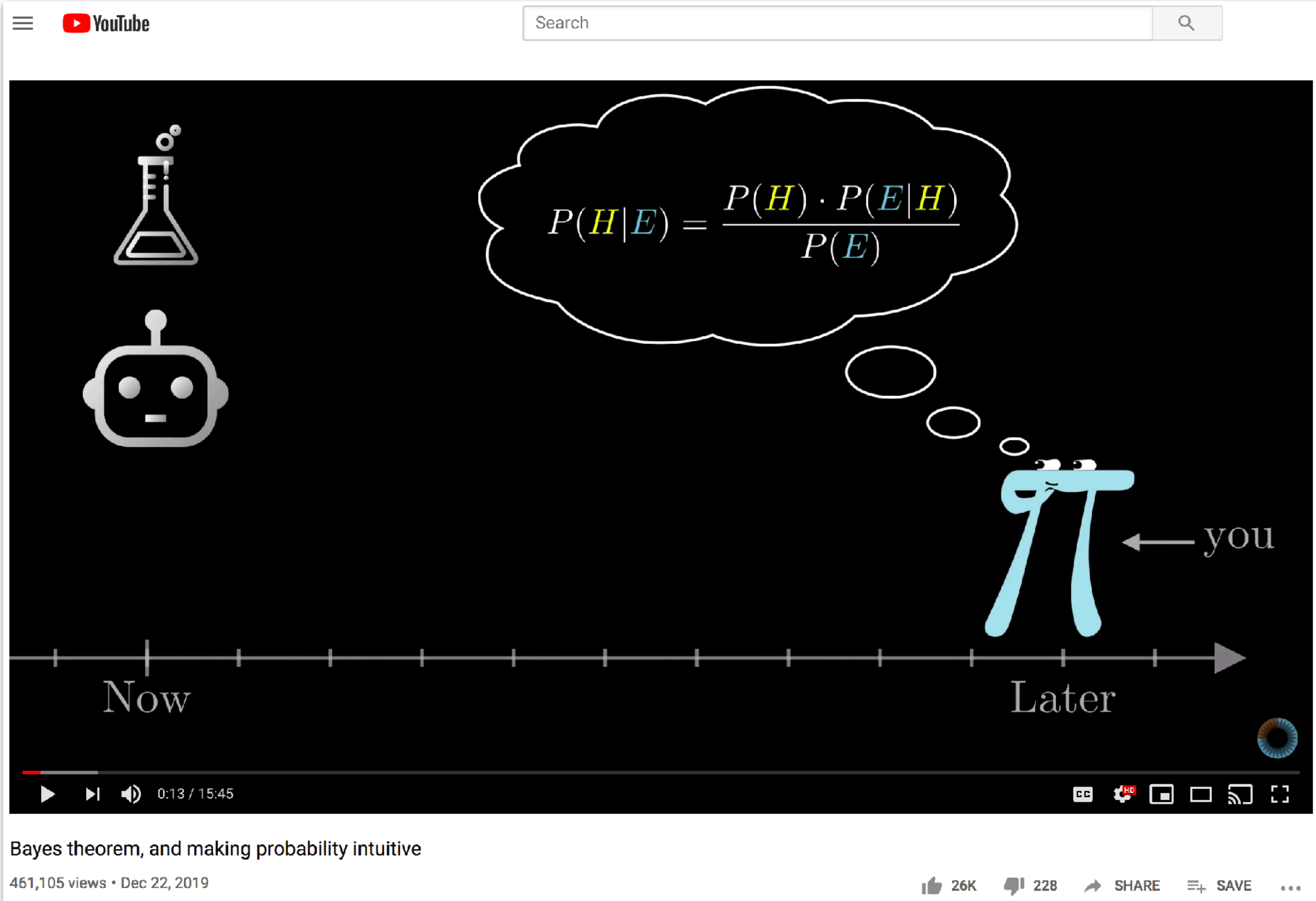


Clue guide to probability

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

$$p(D \mid T) = \frac{4}{4 + 19} = 0.174$$





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

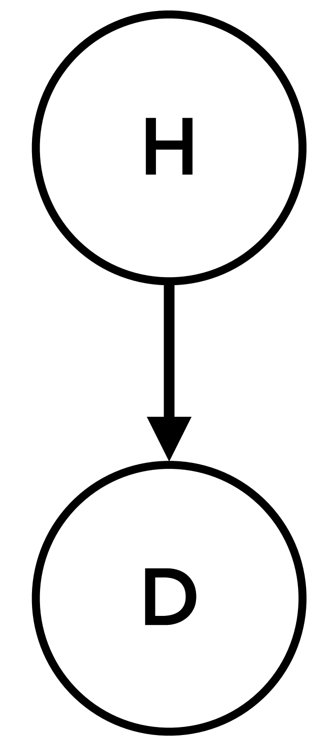
Bayes' theorem in three panels

$$\overset{\text{posterior}}{p(H \mid D)} = \frac{\overset{\text{likelihood}}{p(D \mid H)} \cdot \overset{\text{prior}}{p(H)}}{p(D)}$$

subjective probability interpretation

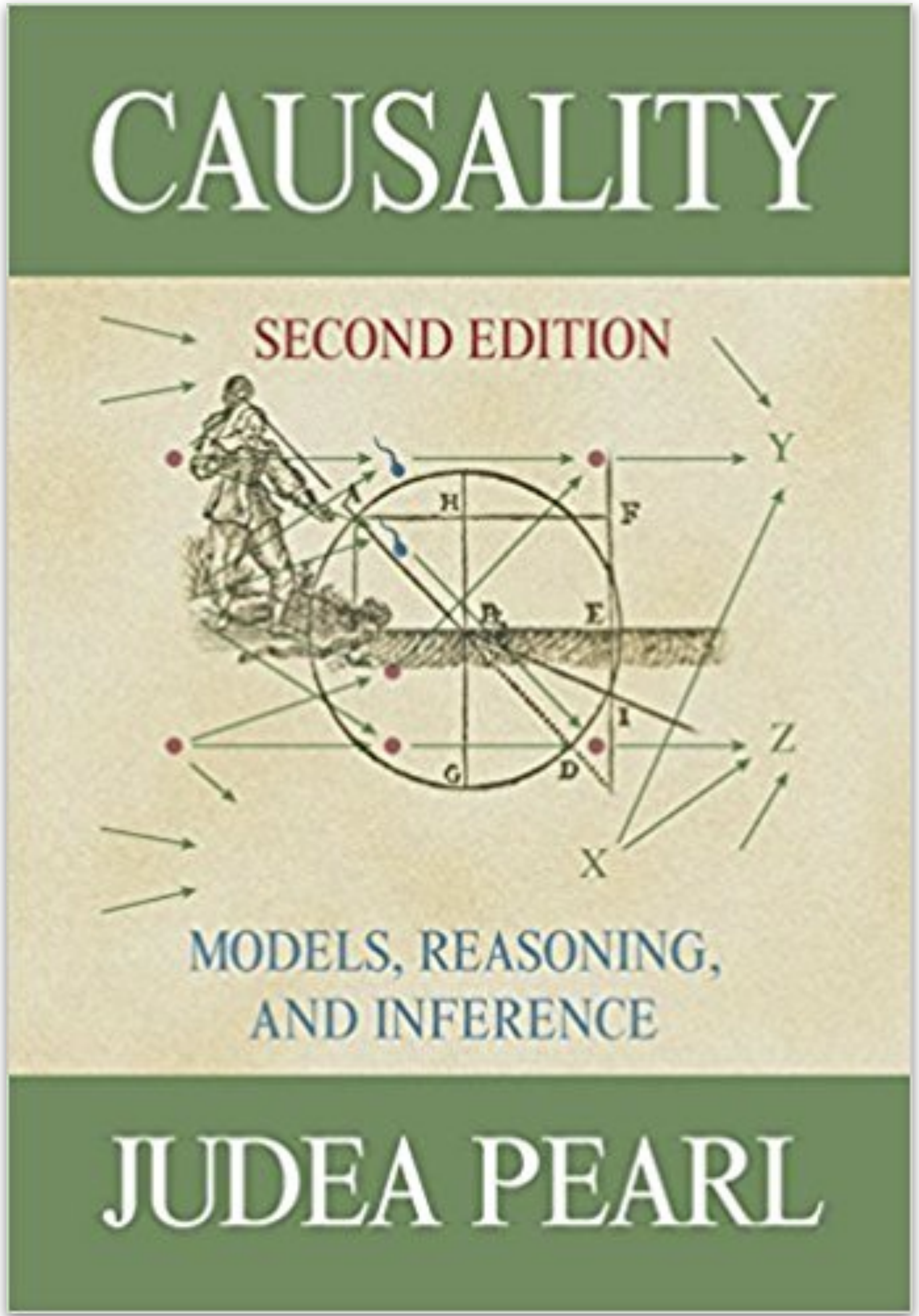
H = Hypothesis

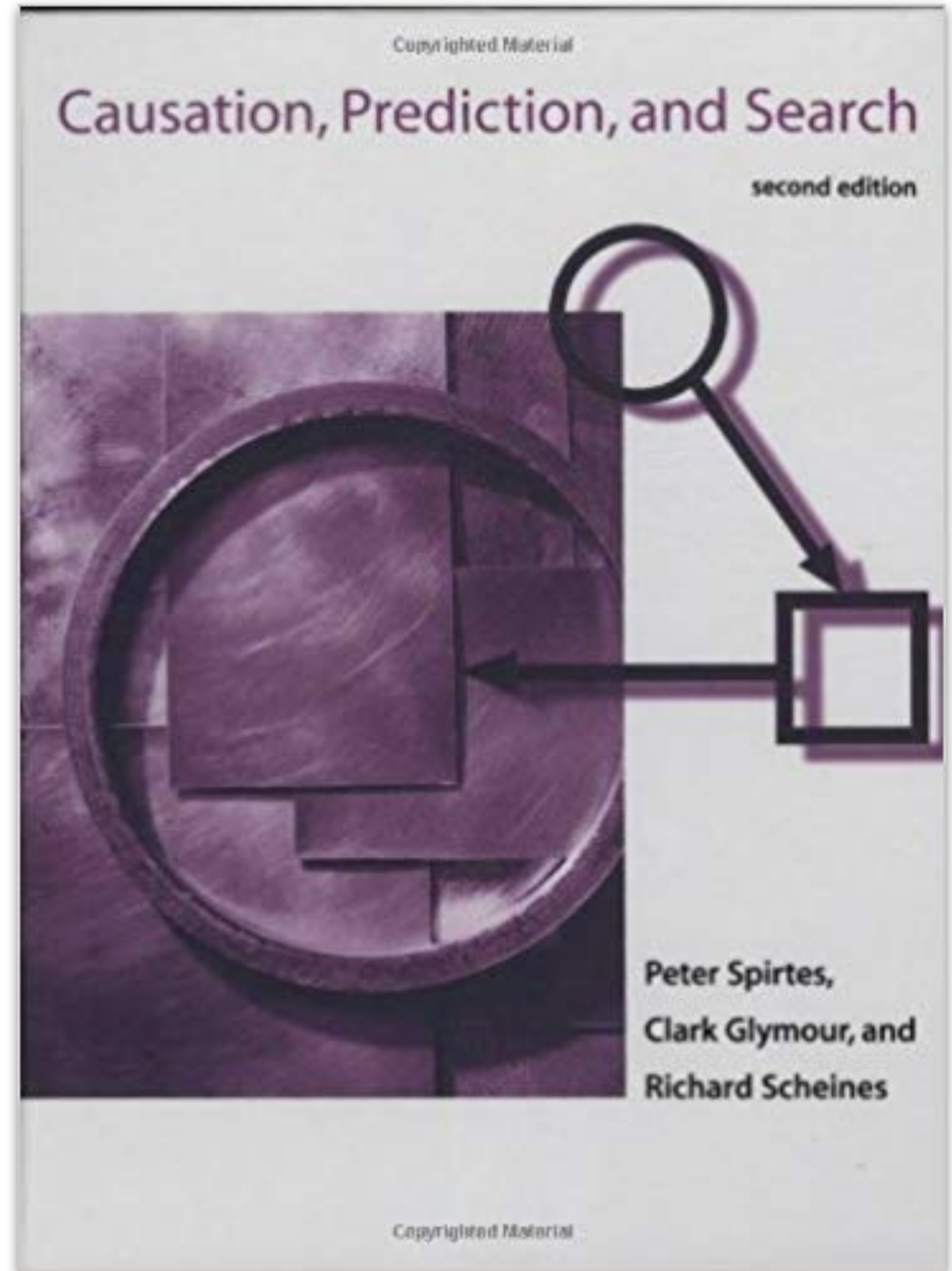
D = Data



https://www.tjmahr.com/bayes-theorem-in-three-panels/

Causal Bayesian Networks



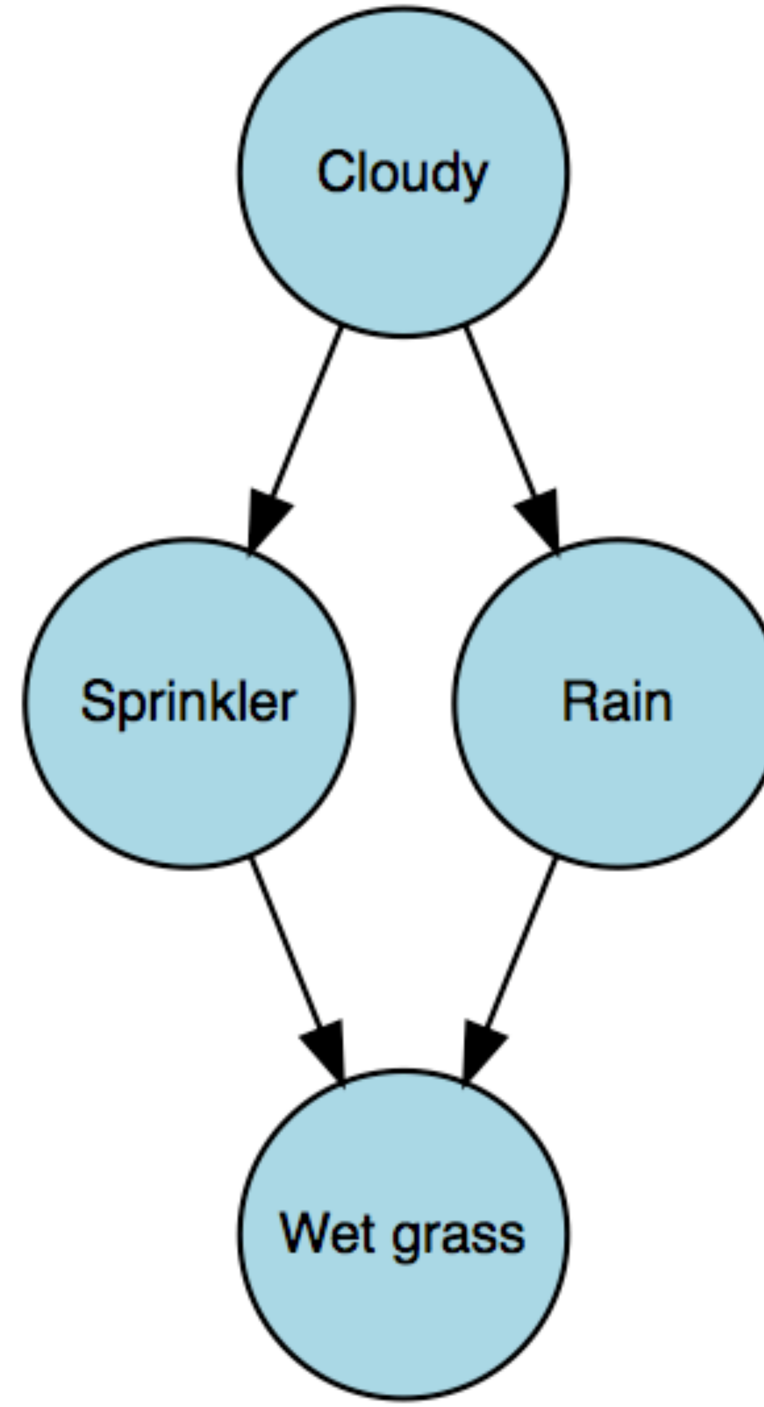


Pearl, J. (2000). Causality: Models, reasoning and inference. Cambridge, England: Cambridge University Press.

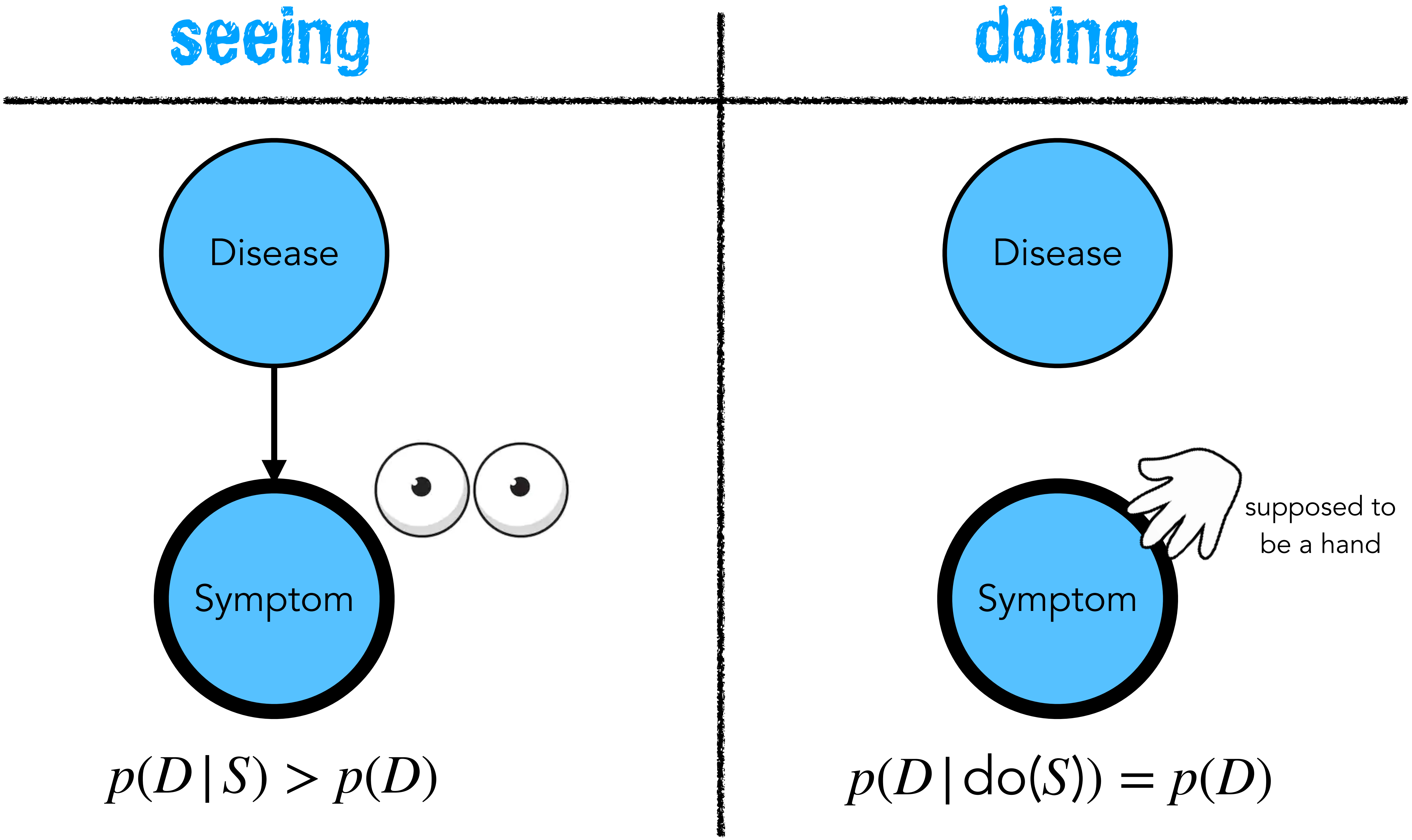
Spirtes, P., Glymour, C. N., & Scheines, R. (2000). Causation, prediction, and search. The MIT Press.

Representation

- **nodes** represent variables of interest
- **links** represent causal relationships between variables
- **conditional probability tables** parameterize the model



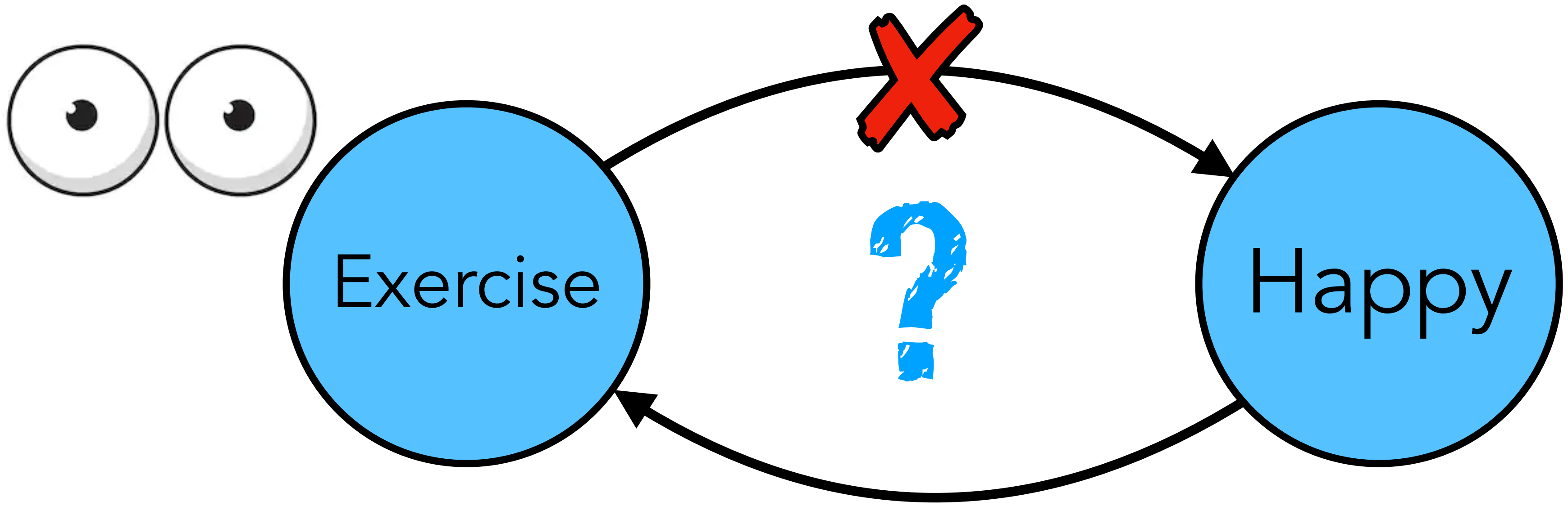
Observation vs. Intervention



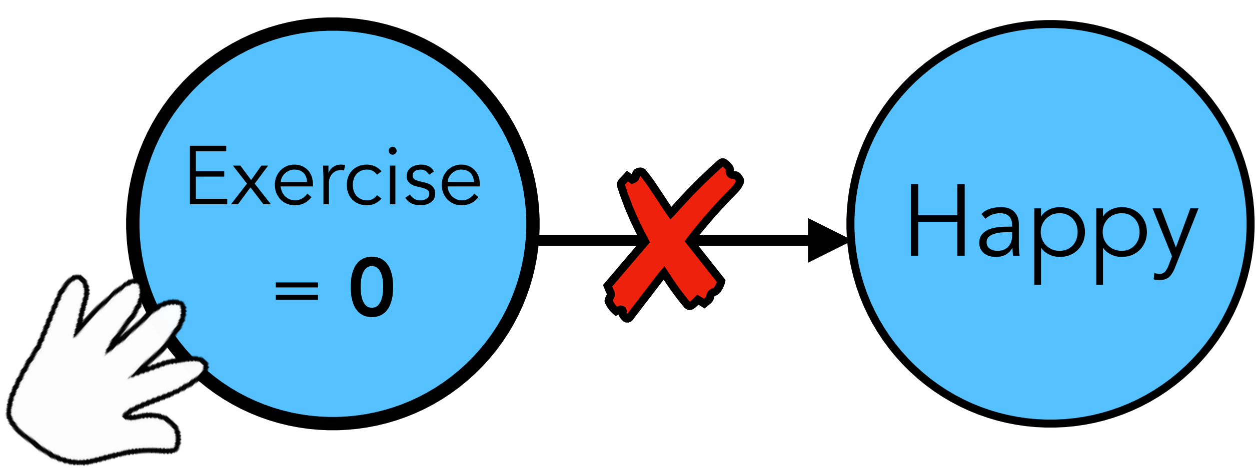
$$p(D \mid S) > p(D)$$

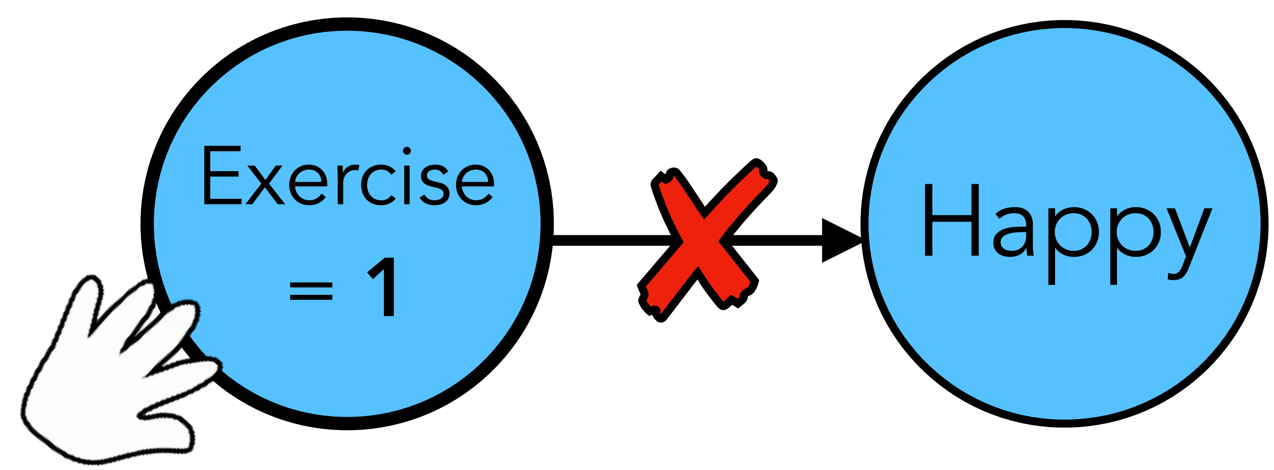
$$p(D \mid \mathrm{do}(S)) = p(D)$$

Inferring causal structure through intervention



Experiment 1

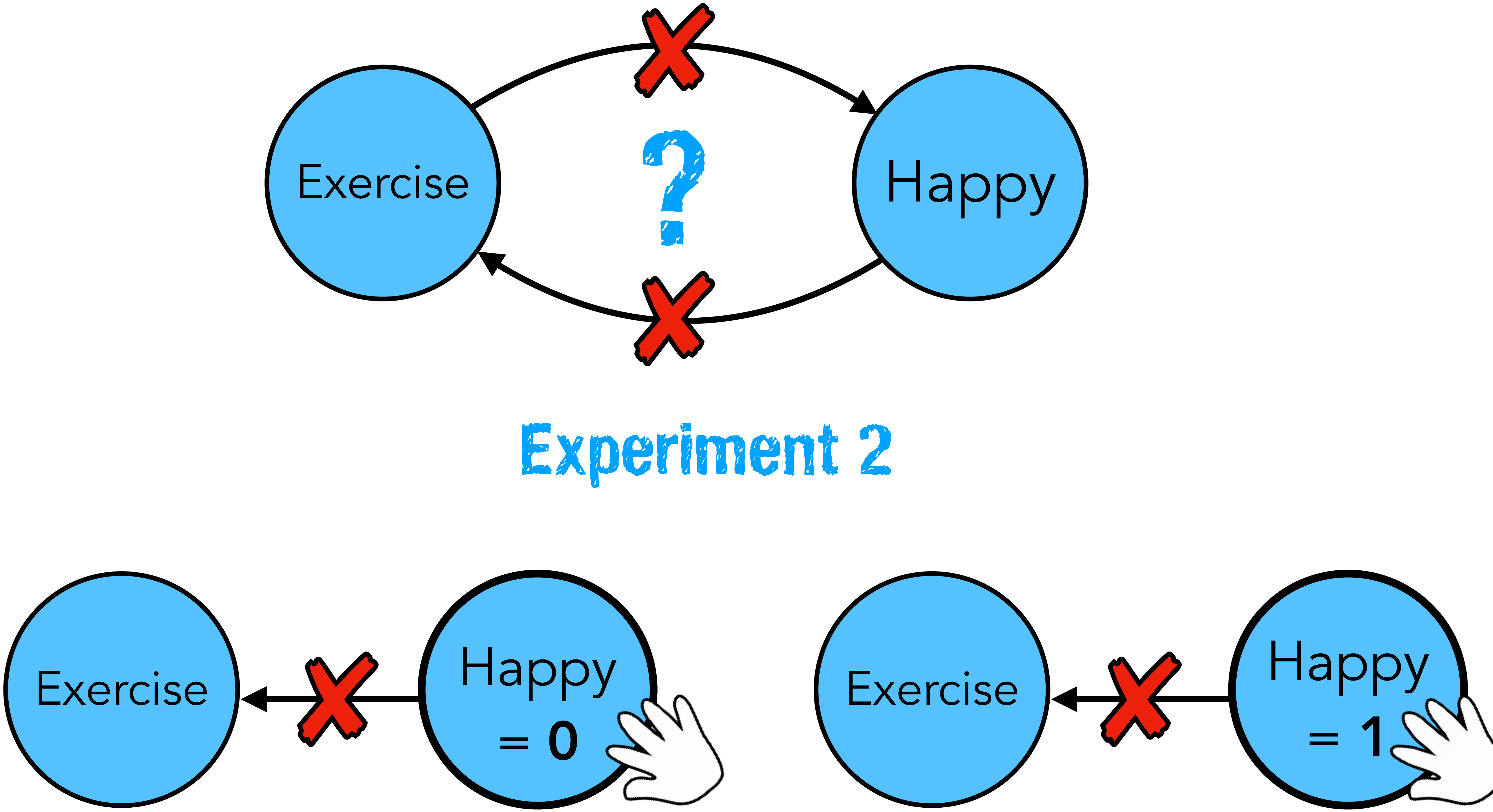




$p(\mathrm{Happy} \,|\, \mathrm{do}(\mathrm{Exercise} = 0)) = 0.3$

$p(\mathrm{Happy} \,|\, \mathrm{do}(\mathrm{Exercise} = 1)) = 0.3$

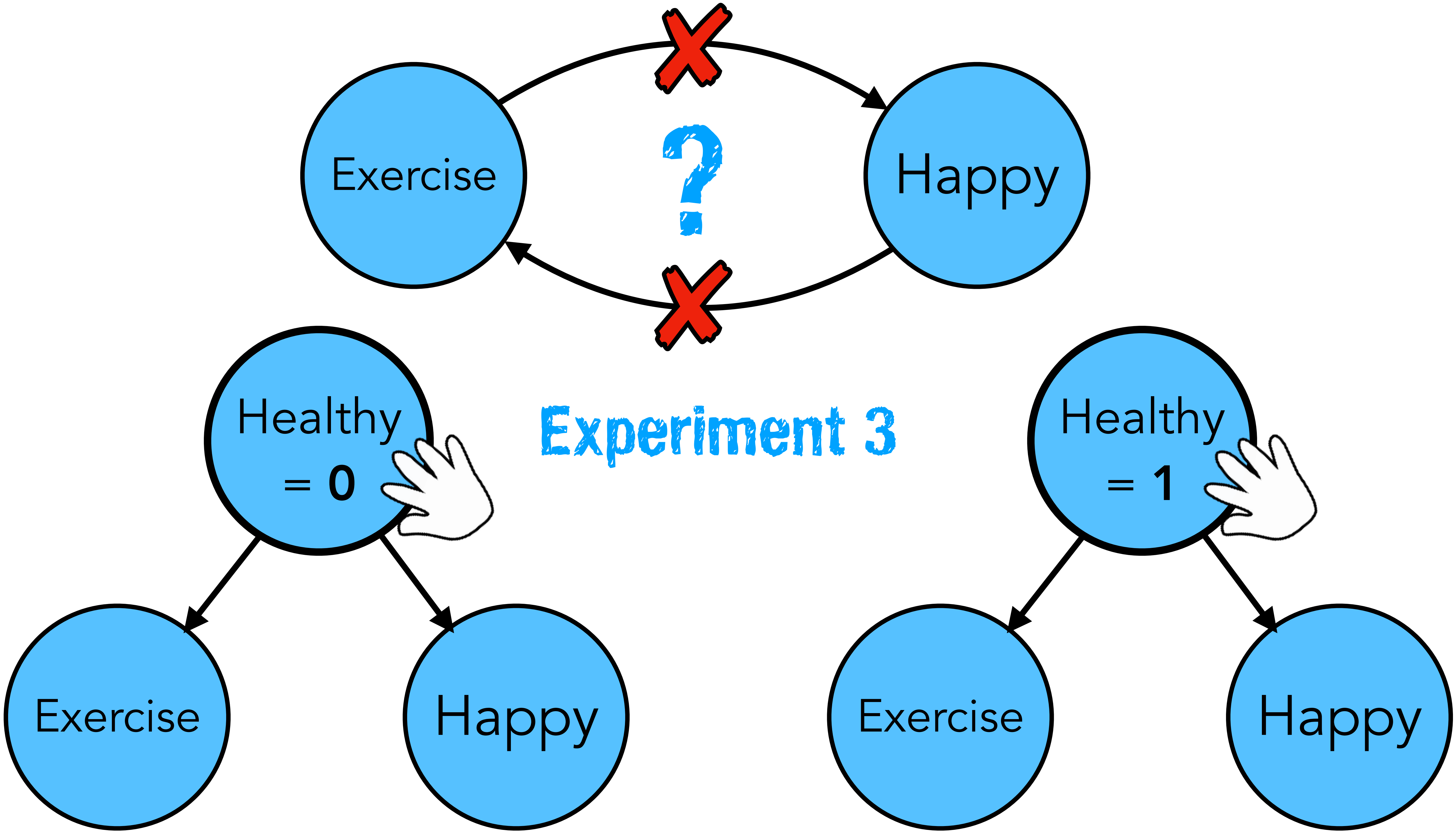
Inferring causal structure through intervention



p(Exercise | do(Happy = 0)) = 0.1

p(Exercise | do(Happy = 1)) = 0.1

Inferring causal structure through intervention



$p(\text{Happy} \,|\, \text{do(Healthy} = 0)) = 0.05$

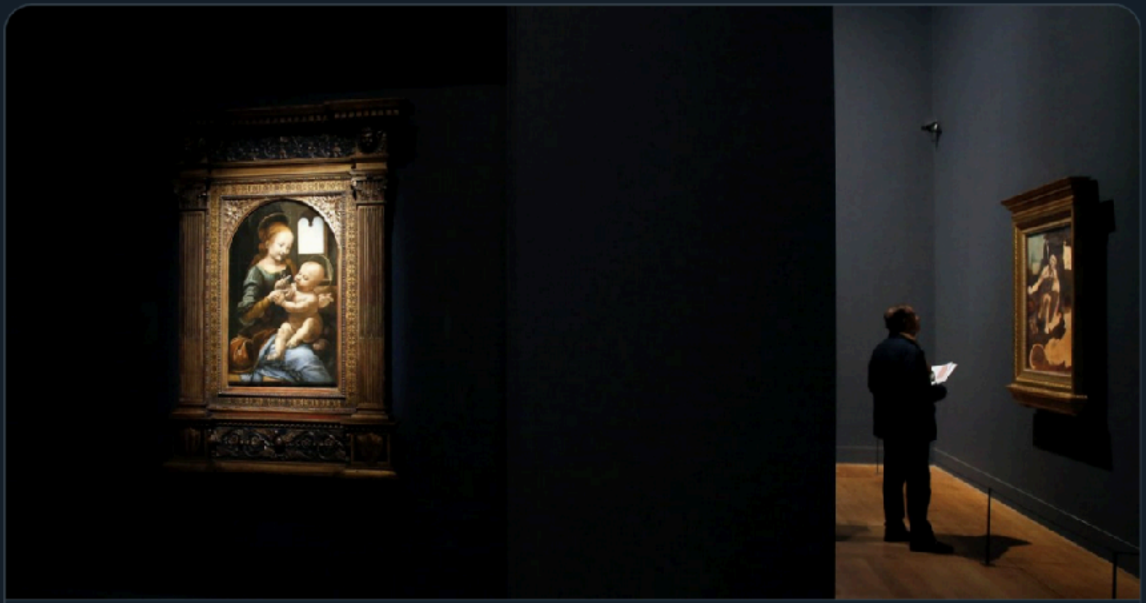
$p(\text{Exercise} \,|\, \text{do(Healthy} = 0)) = 0.1$

$p(\text{Happy} \,|\, \text{do(Healthy} = 0)) = 0.5$

$p(\text{Exercise} \,|\, \text{do(Healthy} = 0)) = 0.75$

NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



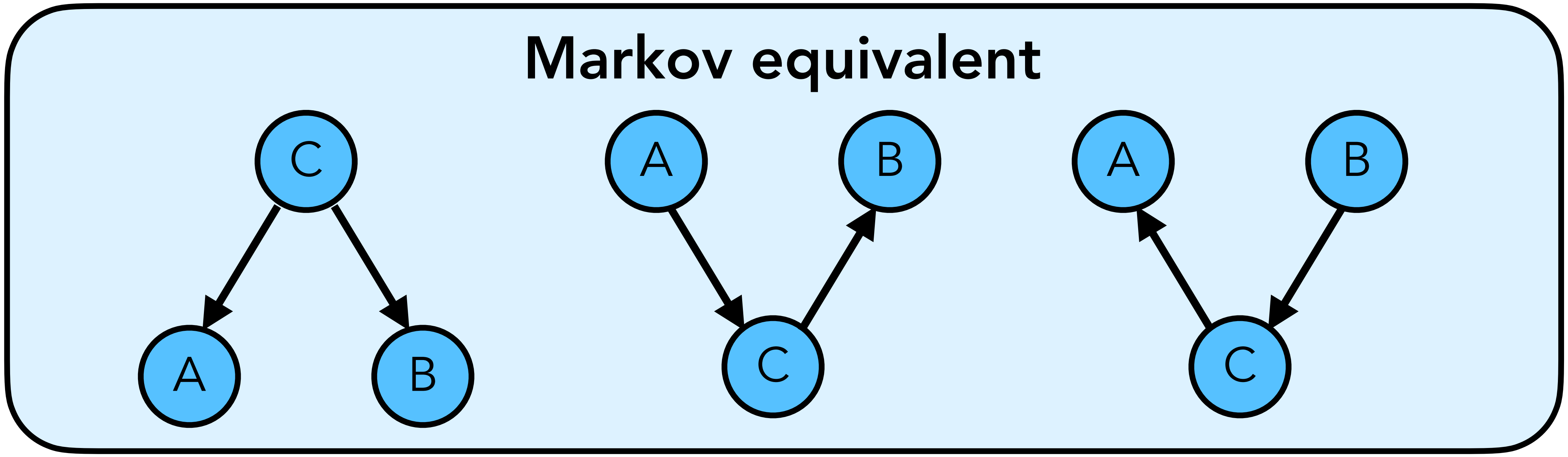
Another Benefit to Going to Museums? You May Live Longer
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

9:19 AM · Dec 22, 2019 · SocialFlow

336 Retweets **1.3K** Likes

Important take home message

- correlation is not causation
- correlation (= probabilistic dependence) suggests that there is *some* causal relationship
- but we don't know which one it is



- **causal interventions** / experiments can reveal the underlying causal structure

Outline

- Introduction to probability / Recap
  - Motivation
  - Counting possibilities
  - Clue guide to probability
- Causal Bayesian Networks
  - seeing (prediction) vs. doing (explanation)
  - correlation is not causation

Feedback

How was the pace of today's class?

much too slow

a little too slow

just right

a little too fast

much too fast

How happy were you with today's class overall?

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

Thank you!