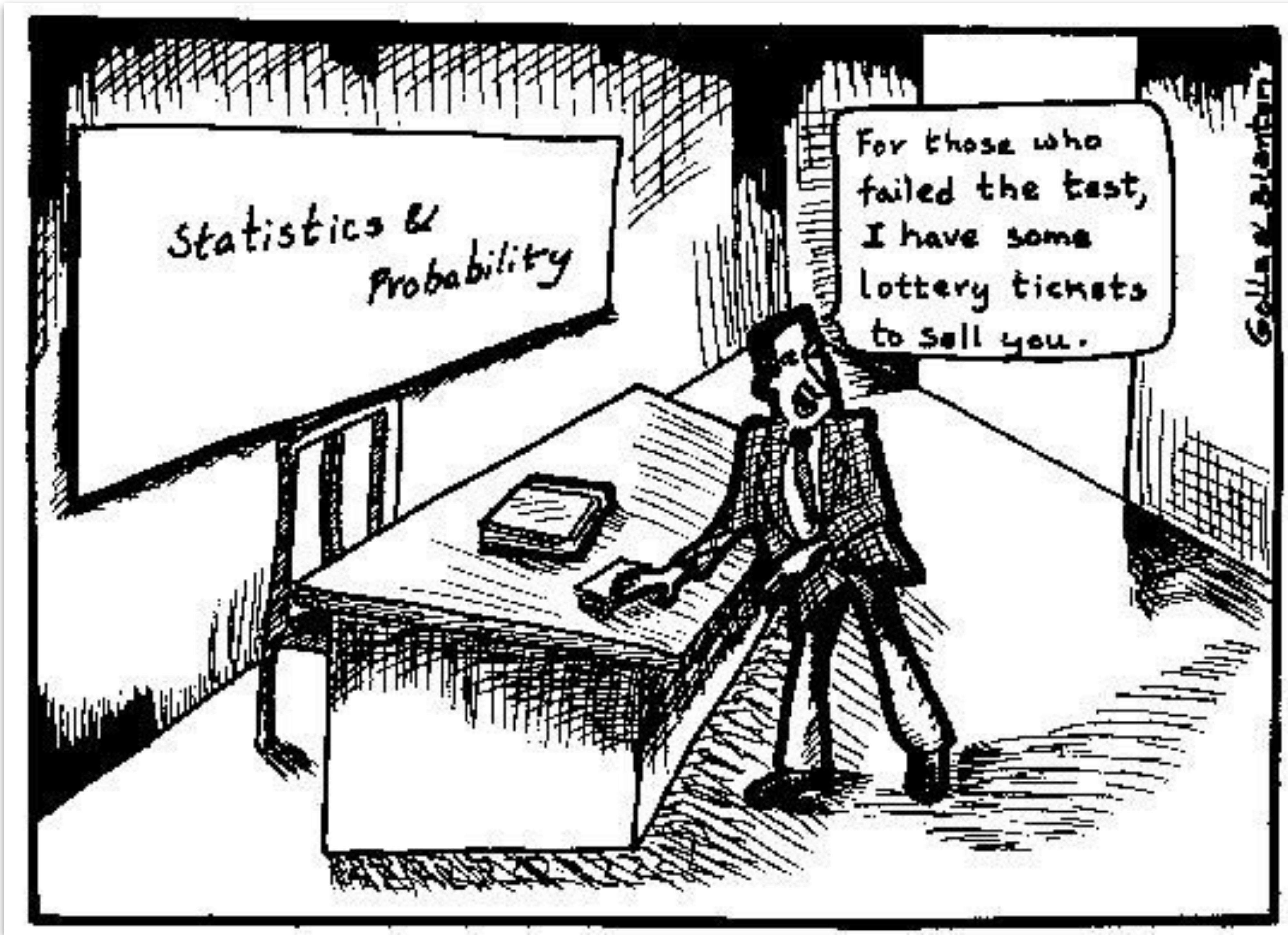
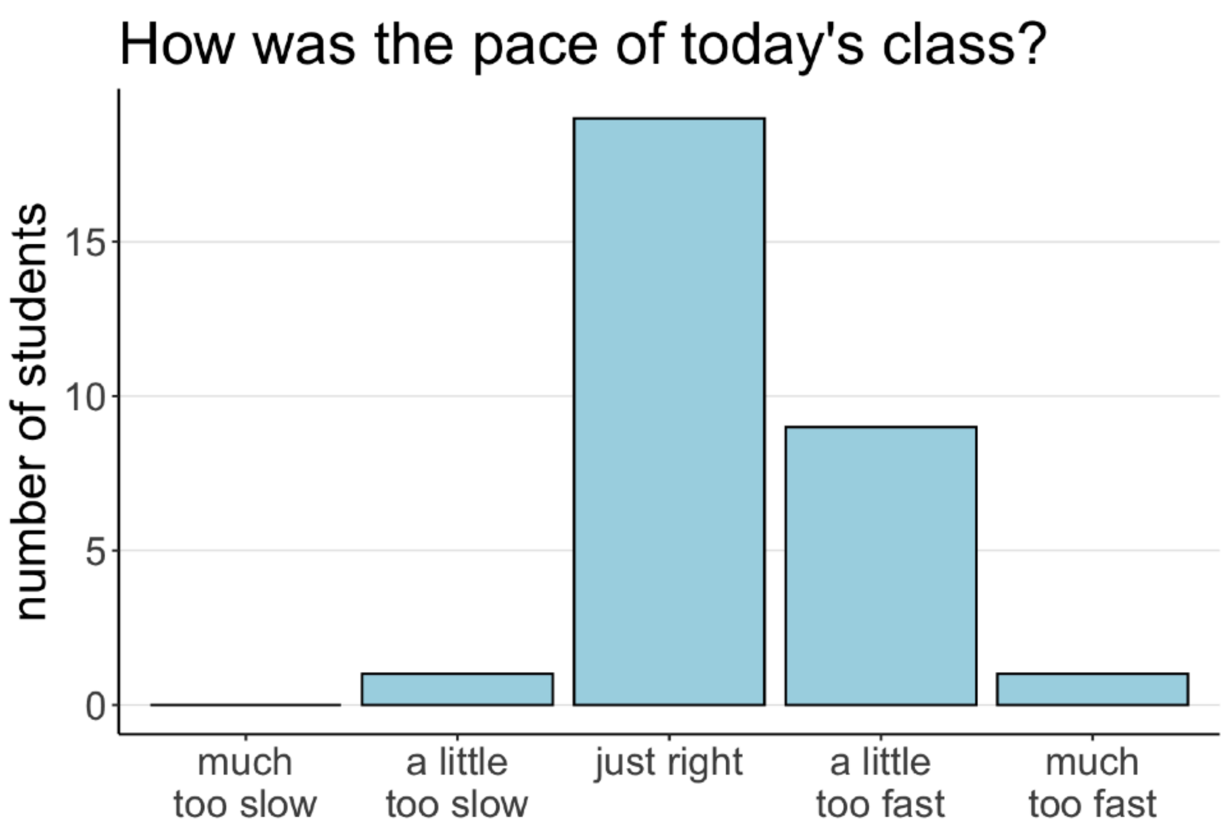
Probability & Causality

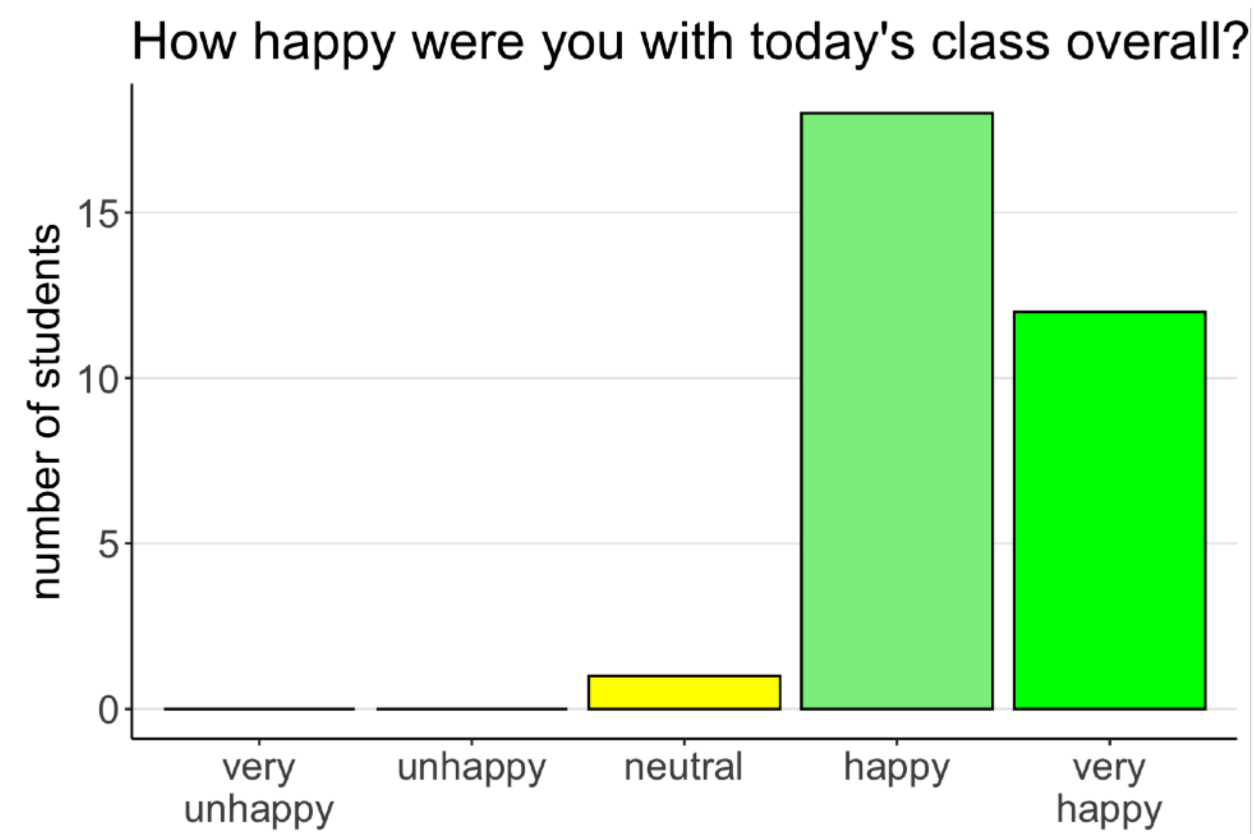


01/17/2020

Your feedback

Your feedback





Your feedback

For merging data frames, just to be clear, you didn't mention the function "merge" because you think it's not a good function (should I learn something else?) or you just didn't have time today?

`left_join()` is the `tidyverse` version of `merge()`

Your feedback

Today's material was extremely helpful and enlightening! Overall good pace as well. Where can we find solutions to the in-class practice problems?

we always post the solutions online after class

Your feedback

I appreciate the applied lecture and like all of the practice we had. Although, I would like additional practice in these kinds of lectures because it moved very quickly for me!

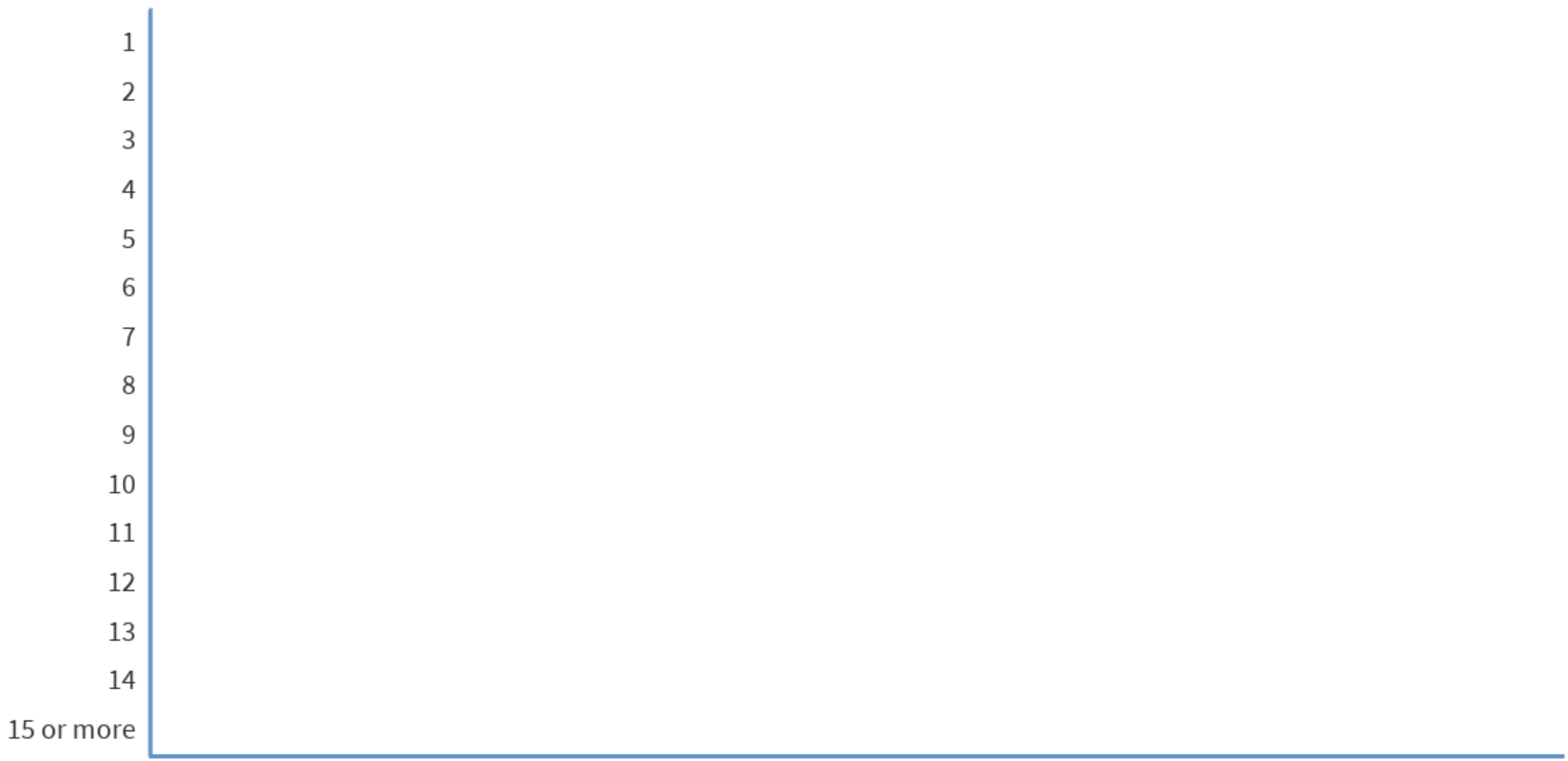
most of the practice will have to happen at home ...

Your feedback

Very helpful & necessary functions - saves life

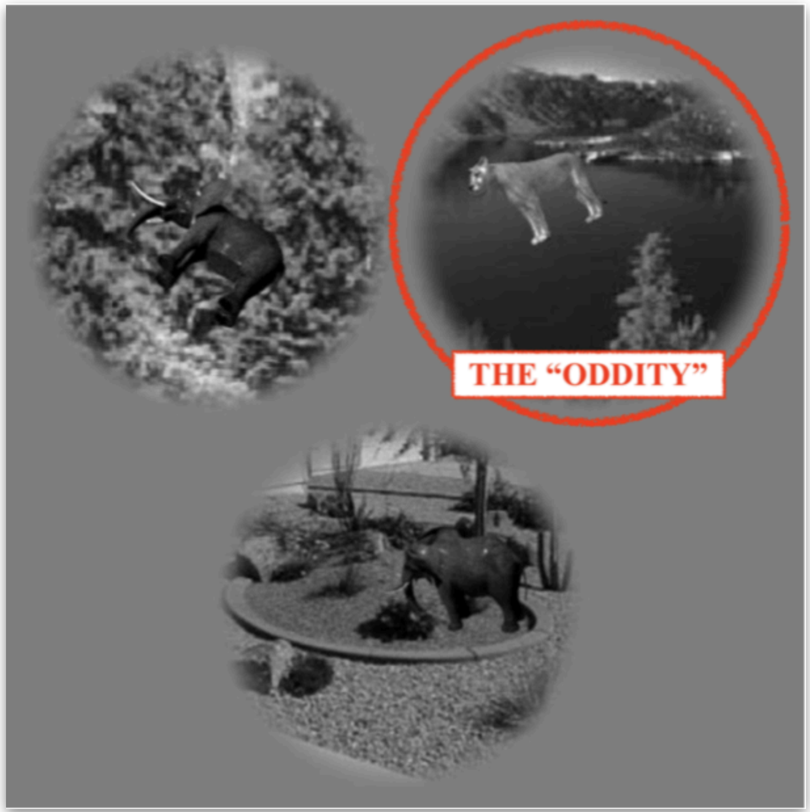
Logistics

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



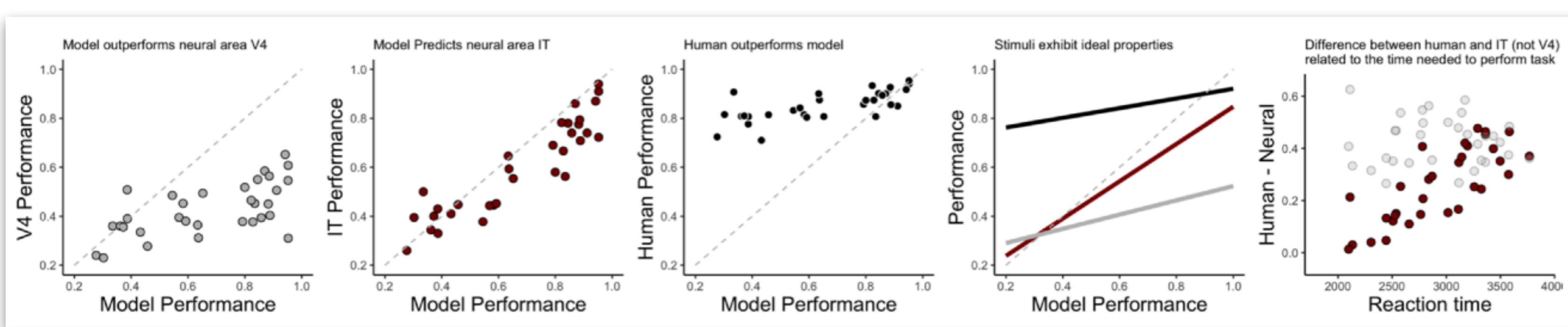
Homework 2

- will be available early this evening
- we'll send out an announcement!
- is due on **Thursday 23rd, at 8pm**



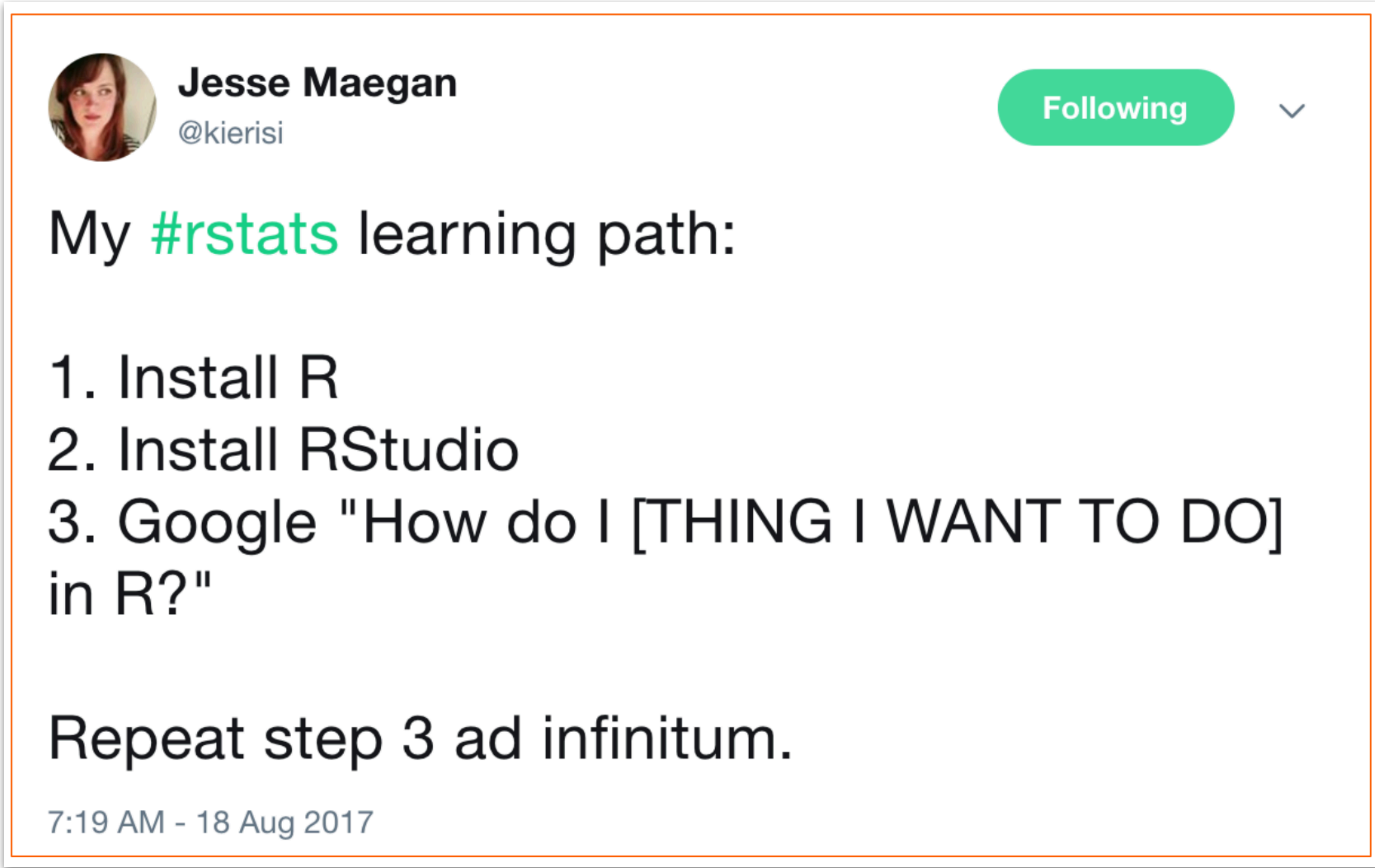
cool experiment

messy dataset



neat plot!

Piazza


Jesse Maegan
@kierisi
Following
My #rstats learning path:
1. Install R
2. Install RStudio
3. Google "How do I [THING I WANT TO DO] in R?"
Repeat step 3 ad infinitum.
7:19 AM - 18 Aug 2017

Piazza

Problem with knitting file

Actions

I'm getting the following error when I try to knit the file. I have already installed TinyTex using the code that was originally on the Homework. When I try to go to the website listed on the error and download LaTeX. I downloaded it, and then I got the 2nd error below.

1st error message:

```
<code>

── Conflicts ──────────────────────────────────────────────────────────────────────
────────────────────────────────────────────────────────── tidyverse_conflicts() ──
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
Quitting from lines 31-33 (1_visualization_homework.Rmd)
Error in eval(expr, envir, enclos) :
  object 'hiphop_cand_lyrics' not found
Calls: <Anonymous> ... handle -> withCallingHandlers -> withVisible -> eval -> eval
Execution halted

No LaTeX installation detected (LaTeX is required to create PDF output). You should
install a LaTeX distribution for your platform: https://www.latex-project.org/get/

  If you are not sure, you may install TinyTeX in R: tinytex::install_tinytex()

  Otherwise consider MiKTeX on Windows - http://miktex.org

  MacTeX on macOS - https://tug.org/mactex/
  (NOTE: Download with Safari rather than Chrome _strongly_ recommended)

  Linux: Use system package manager</code>
```

ideally as a reprex!

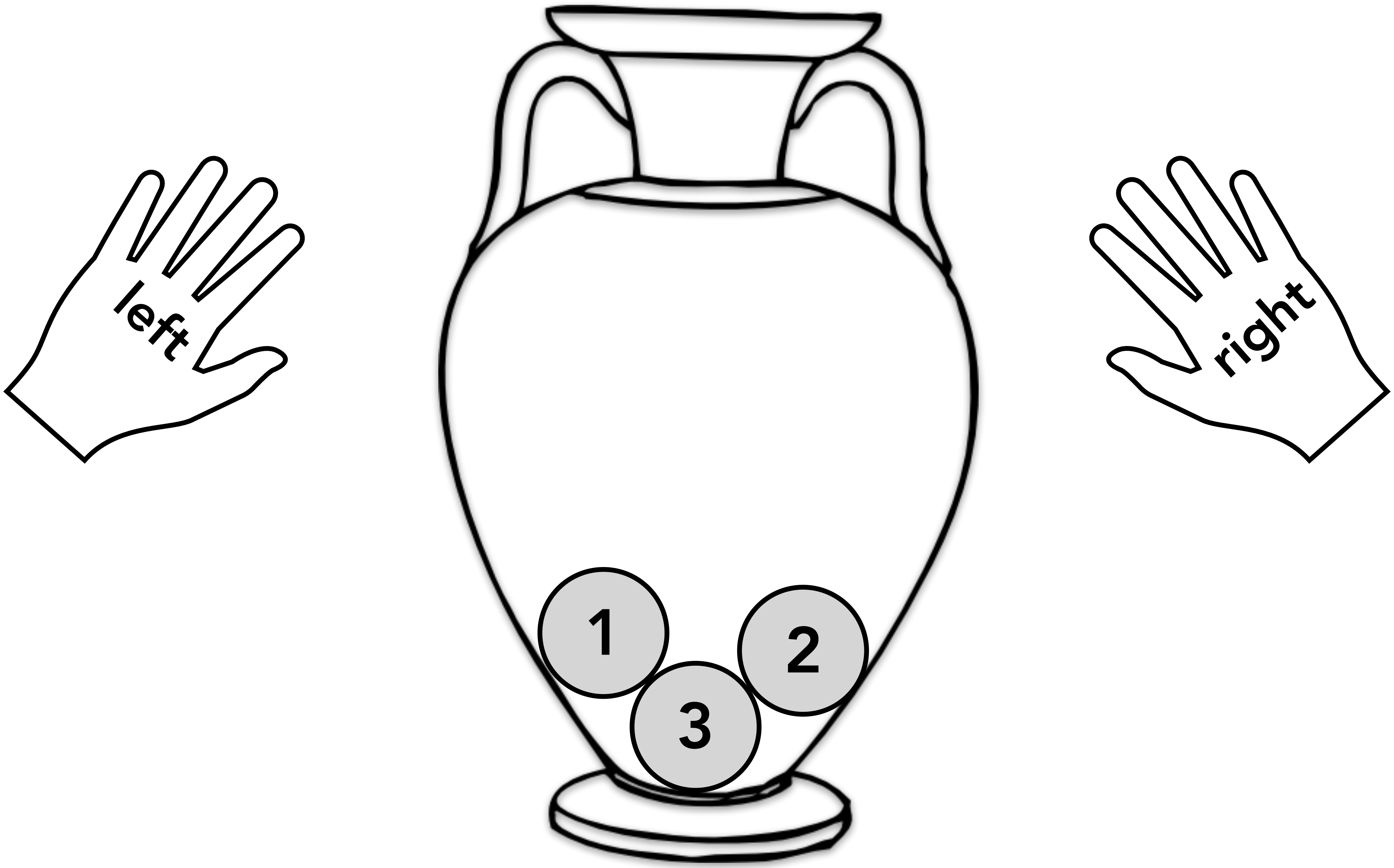
make sure to always post the code!

Outline

- Introduction to probability / Recap
  - Counting possibilities
  - Interpretation of probability
  - Clue guide to probability
- Bayesian Networks
  - representation
  - inference
  - (un-)conditional (in-)dependence
- Causal Bayes nets

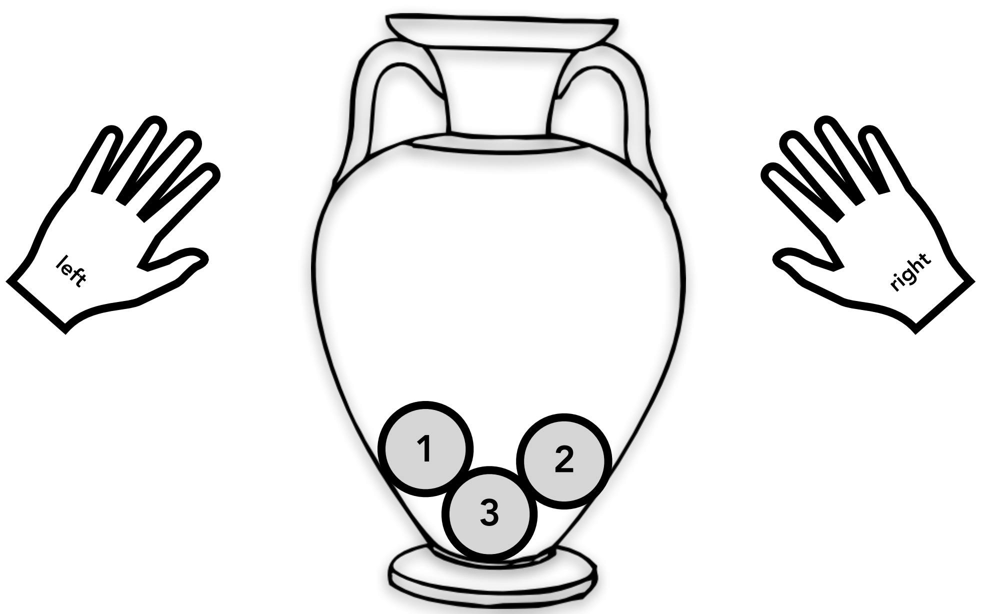
Counting possibilities

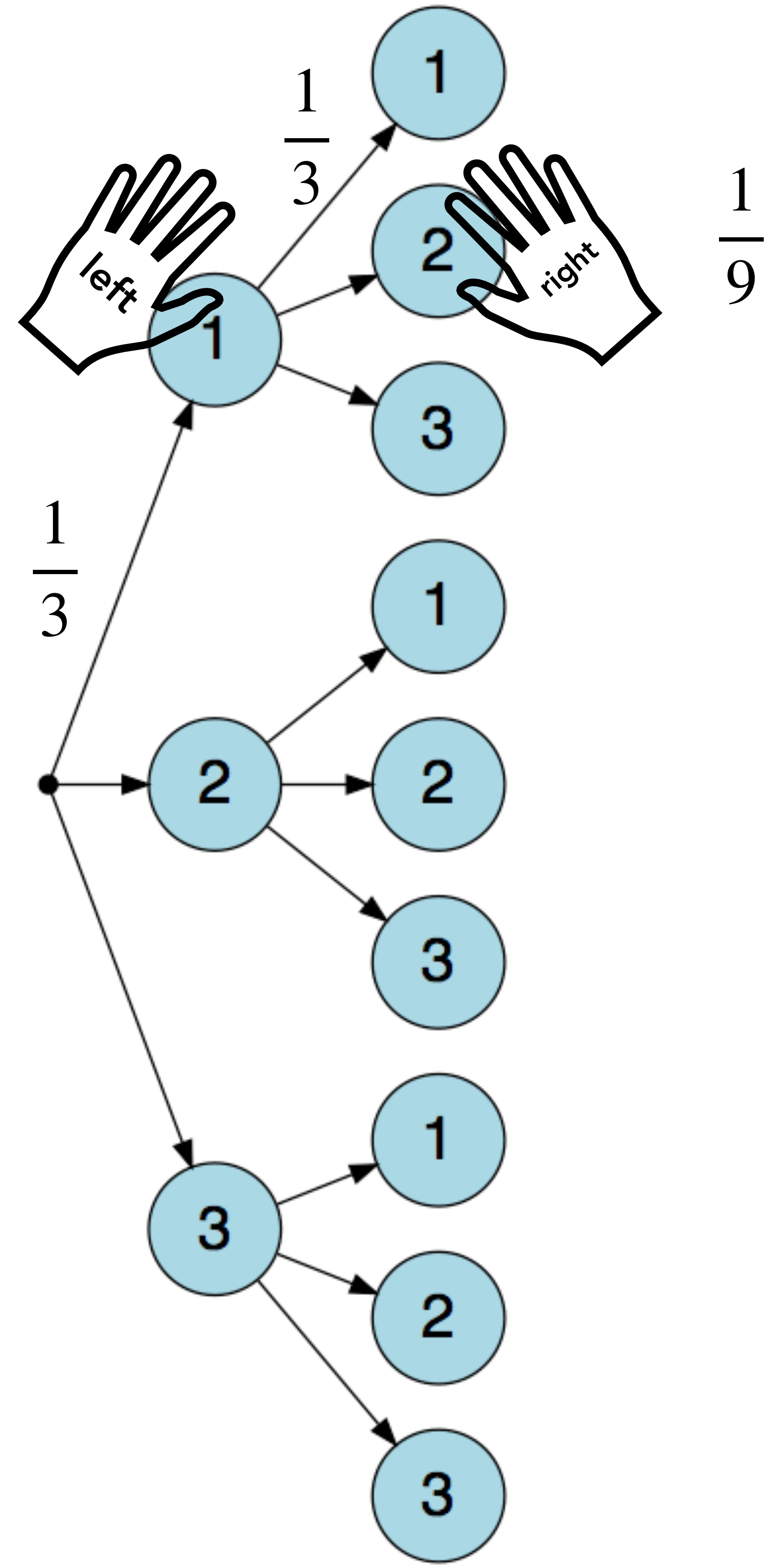
no stats class without urns!



Sampling **with** replacement

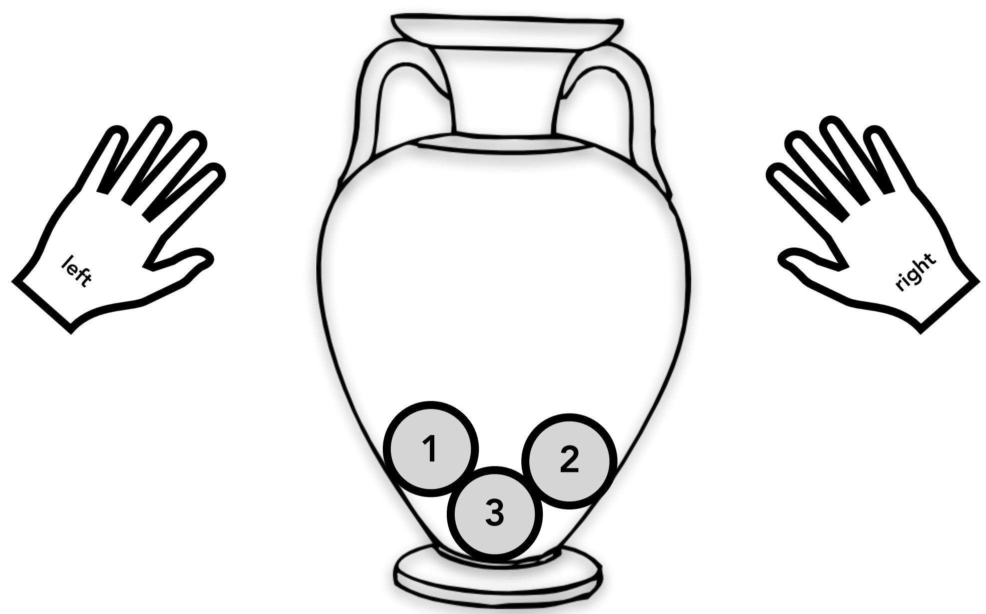
$p(\text{left} = 1, \text{right} = 2) = ?$

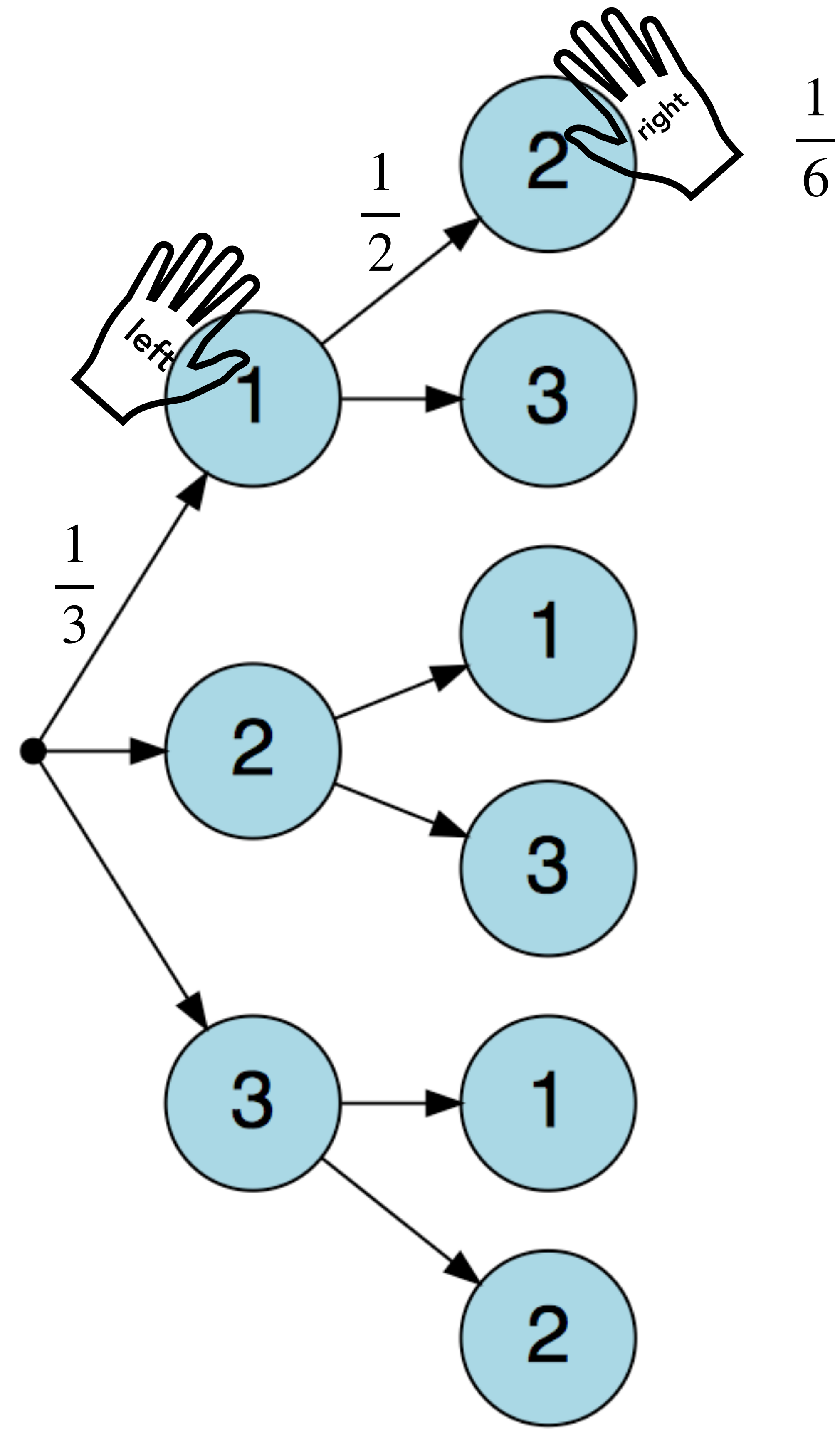




Sampling **without** replacement

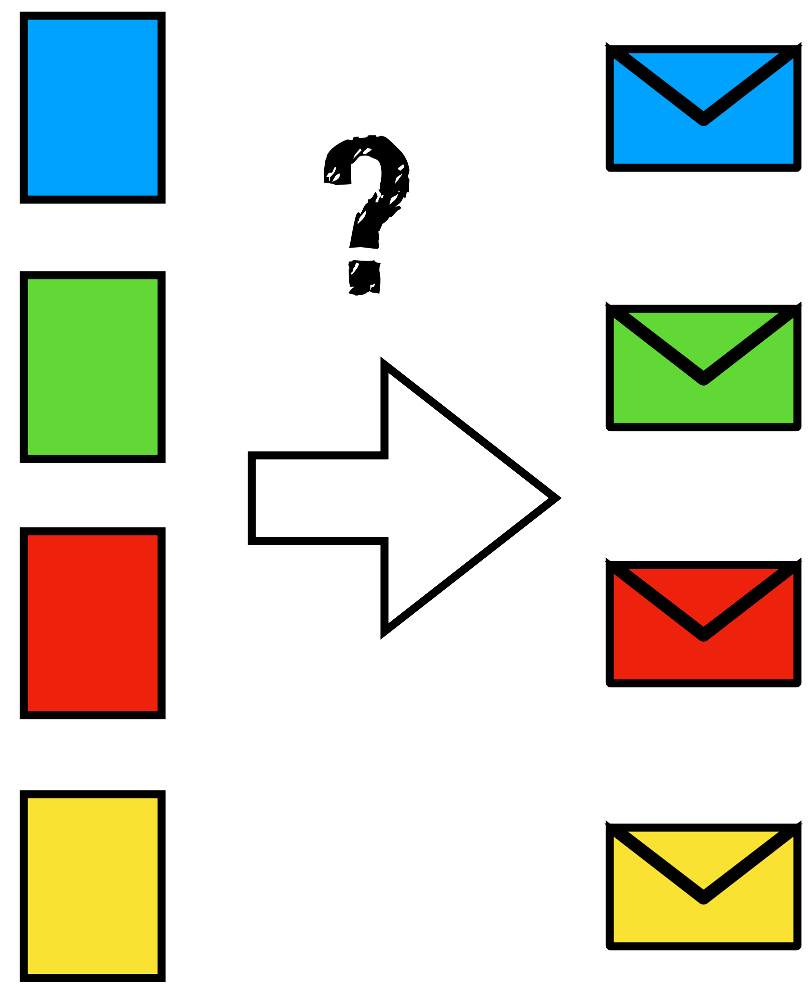
$$p(\text{left} = 1, \text{right} = 2) = ?$$





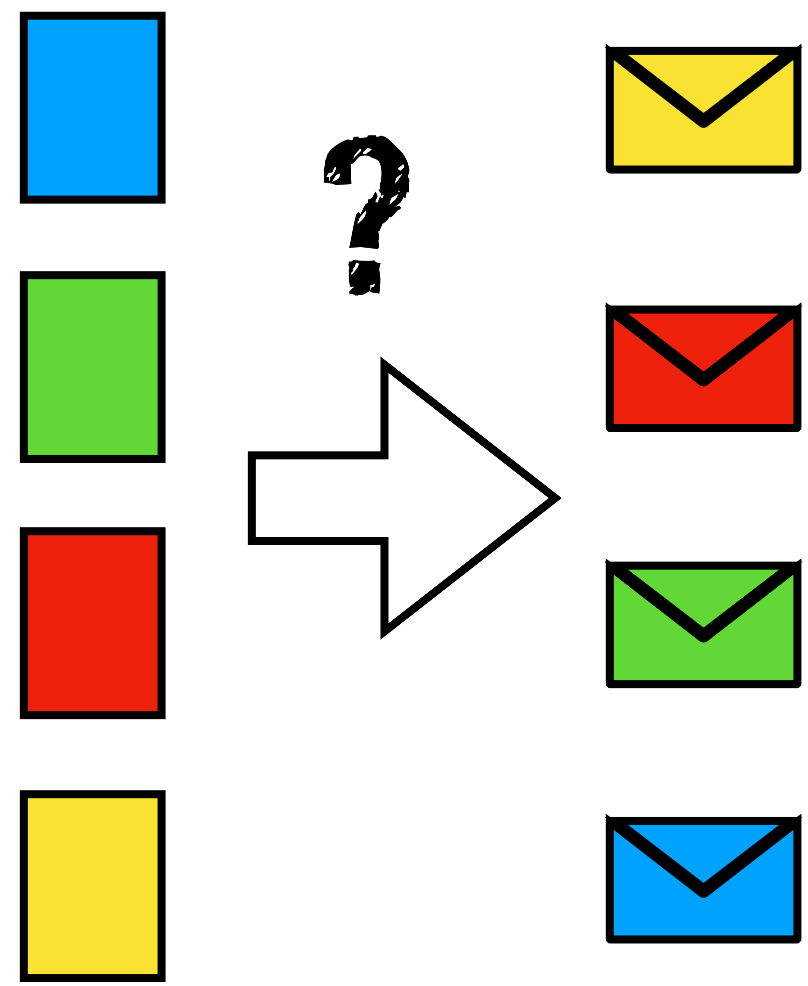
Random secretary

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

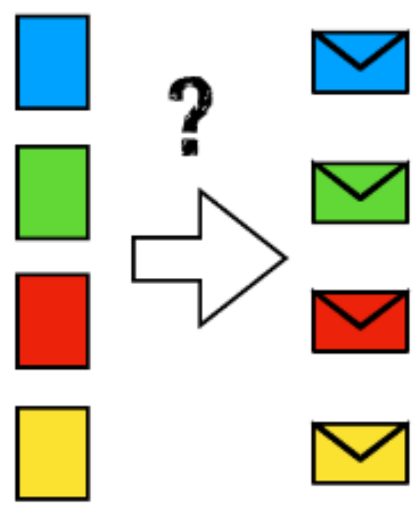
A secretary types four letters to four people and addresses the four envelopes. If he inserts the letters at random, each in a different envelope, what is the probability that exactly three letters will go into the right envelope?

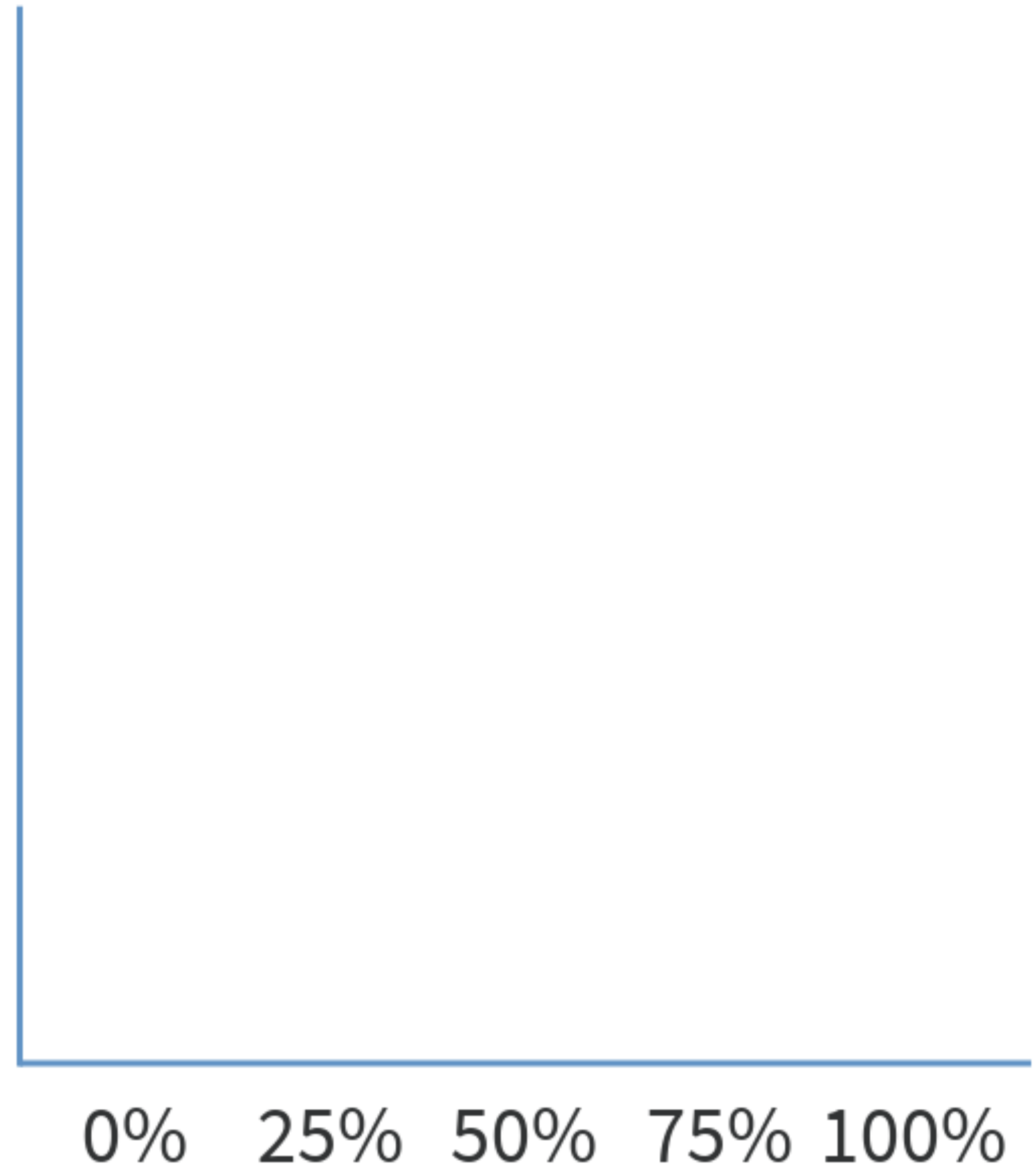


What is the probability that exactly three letters will go into the right envelope?

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?

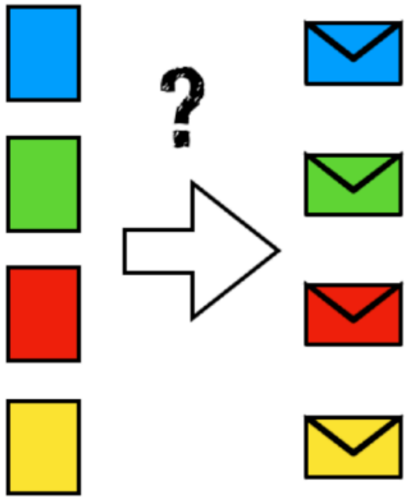


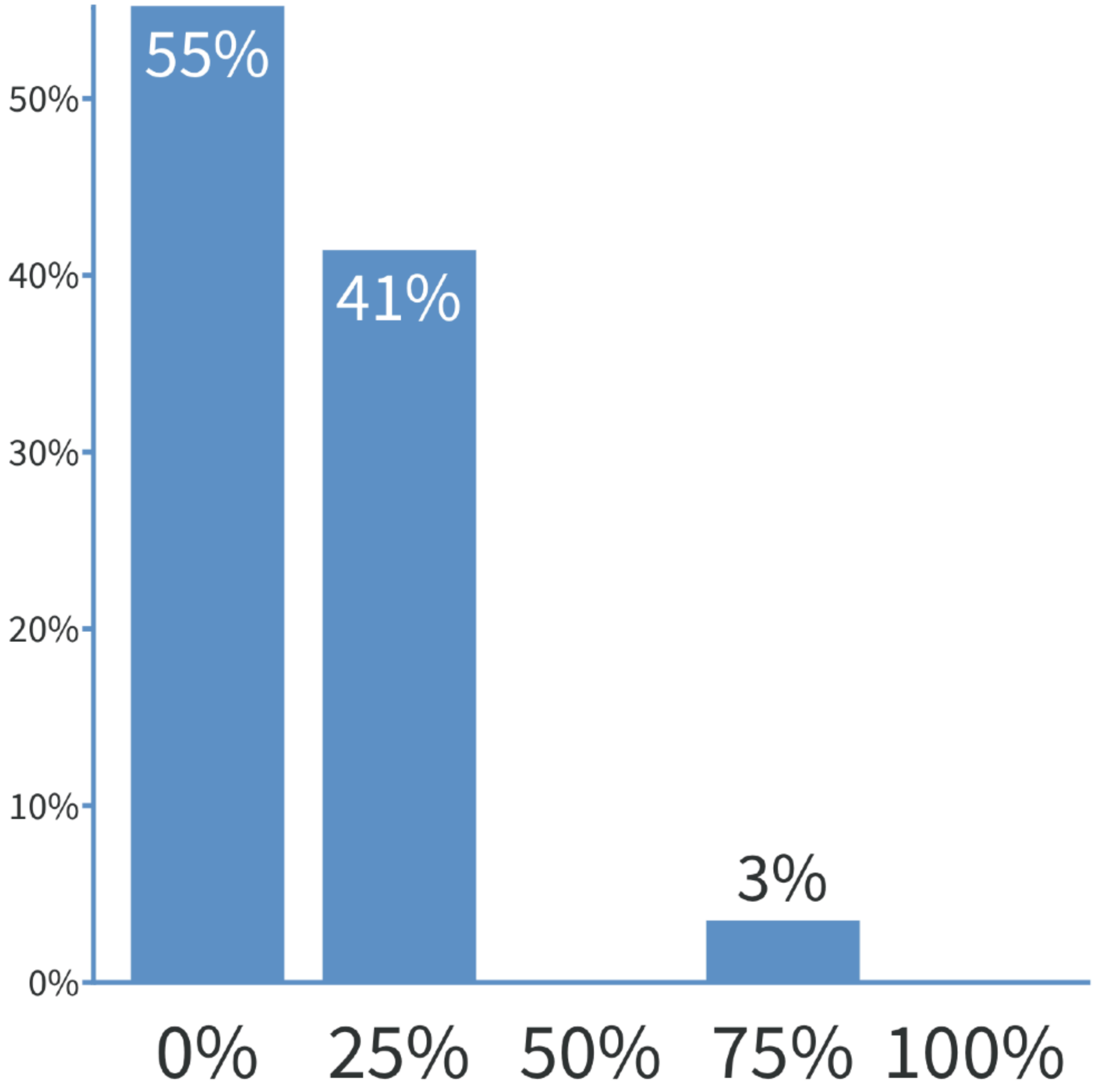


What is the probability that exactly three letters will go into the right envelope?

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?





Random secretary

```
library("arrangements")
df.letters = permutations(x = 1:4, n = 4) %>%
  as_tibble() %>%
  set_names(str_c("person_", 1:4))
```

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

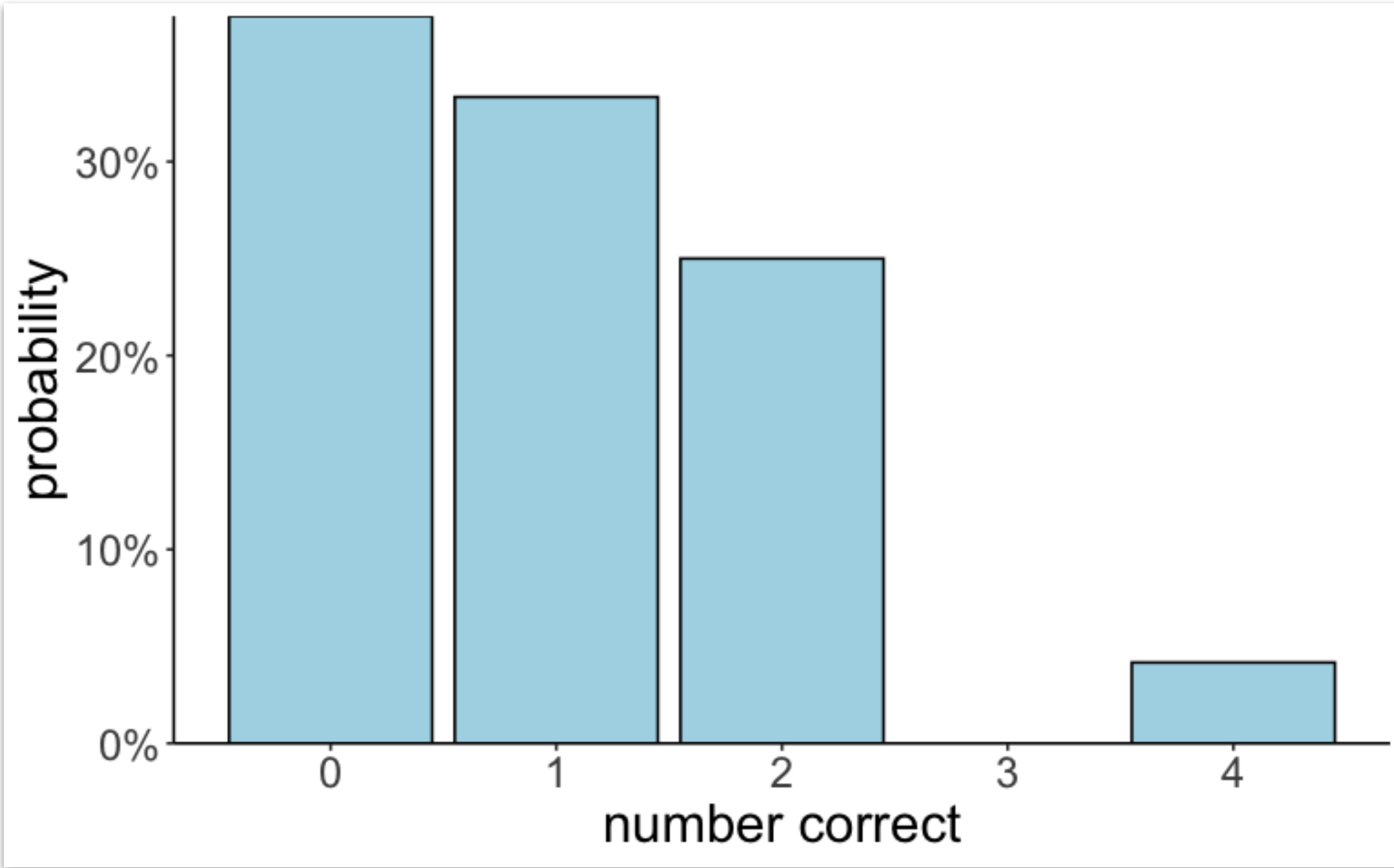
Random secretary

```r
library("arrangements")
df.letters = permutations(x = 1:4, n = 4) %>%
  as_tibble() %>%
  set_names(str_c("person_", 1:4)) %>%
  mutate(n_correct = (person_1 == 1) +
                     (person_2 == 2) +
                     (person_3 == 3) +
                     (person_4 == 4))
```

person_1	person_2	person_3	person_4	n_correct
1	2	3	4	4
1	2	4	3	2
1	3	2	4	2
1	3	4	2	1
1	4	2	3	1
1	4	3	2	2
2	1	3	4	2
2	1	4	3	0
2	3	1	4	1

prob_3_correct
0

Random secretary



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!

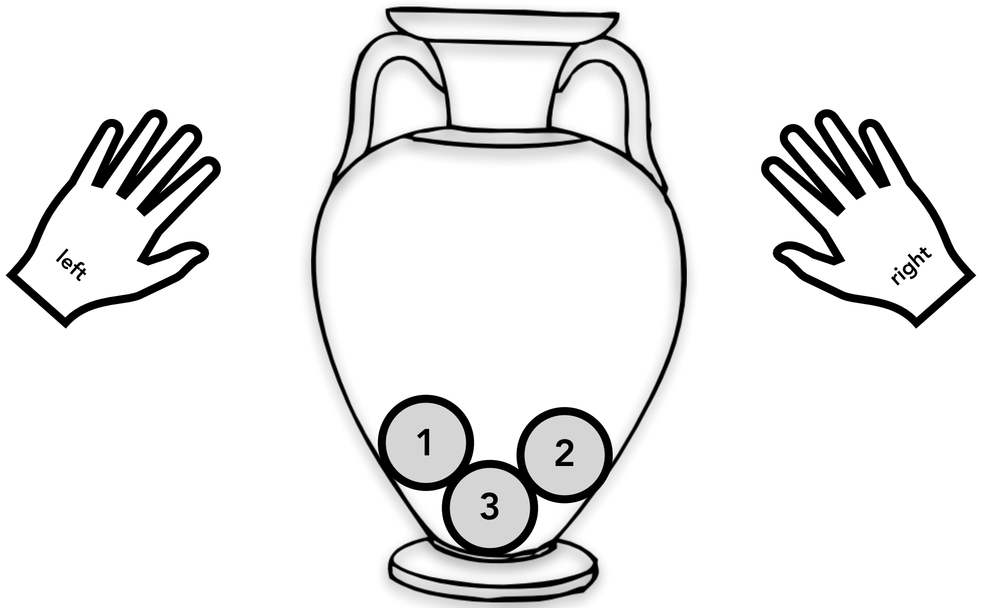
Definitions

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.

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



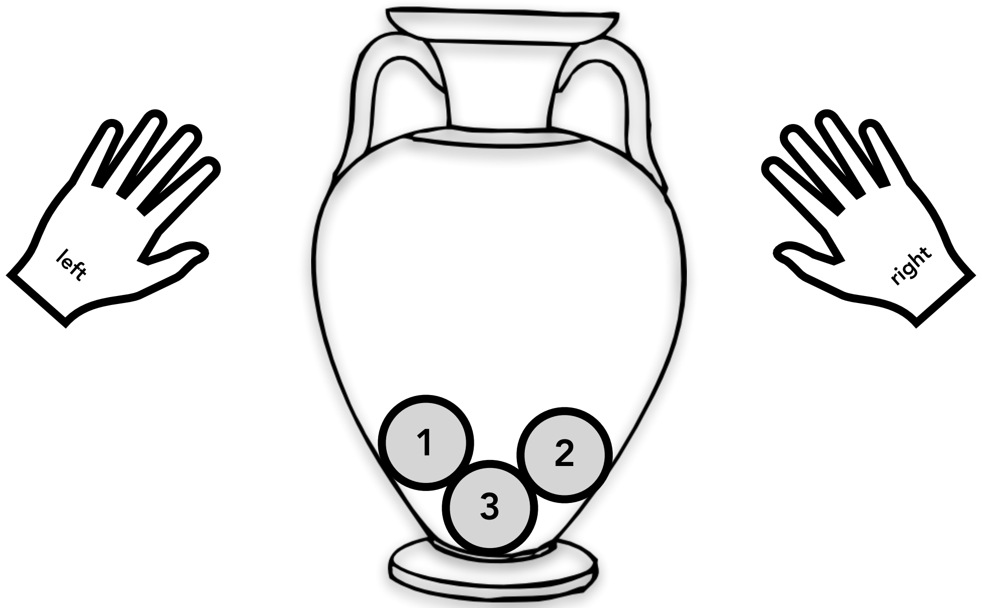
Event: Subset of the sample space. (1, 1)

Definitions

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

1. Probability cannot be negative.

$$P(X_i) \geq 0$$



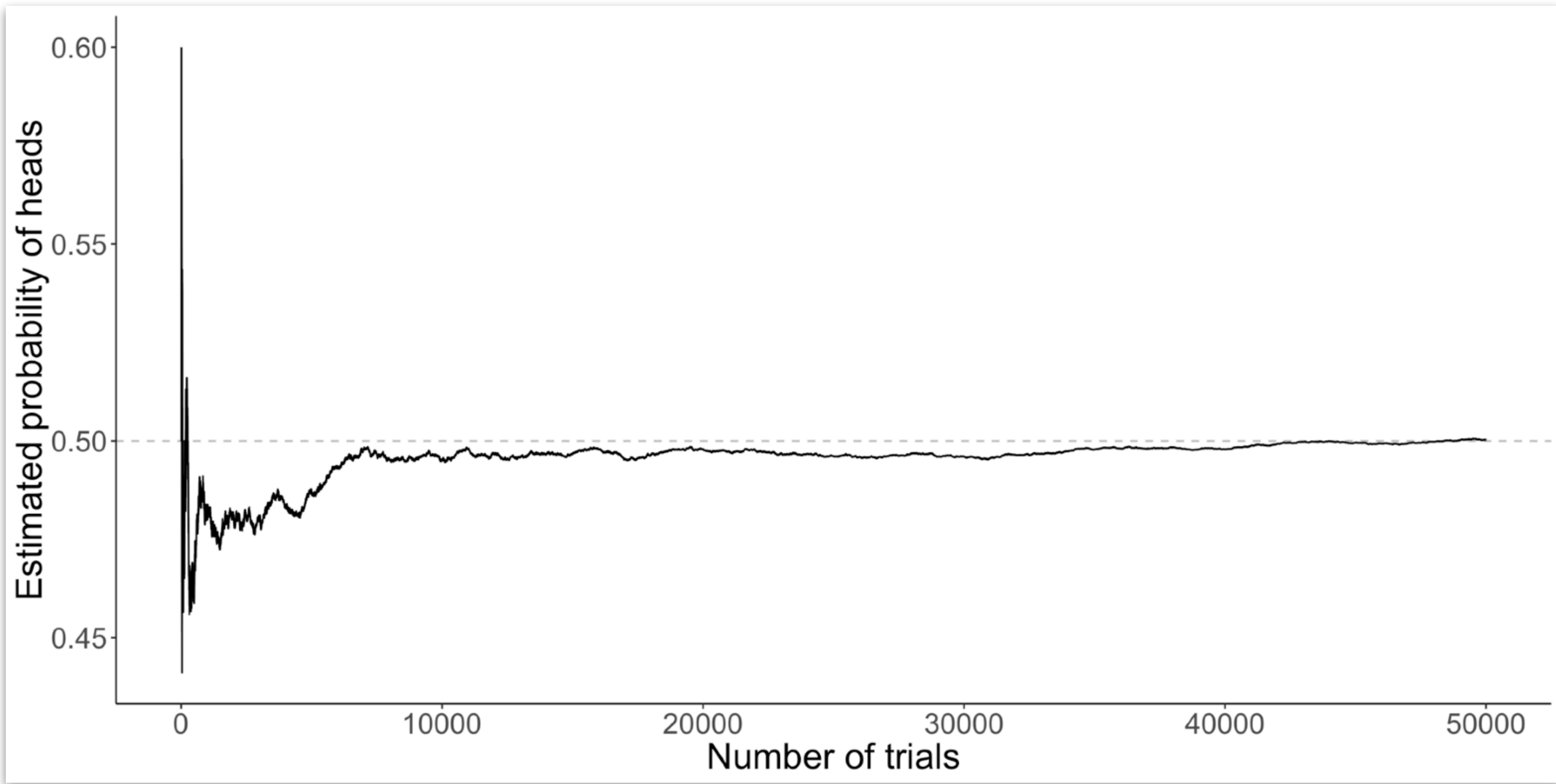
2. Total probability of all outcomes in the sample space is 1.

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

Interpretations of probability

Frequentist interpretation

Probabilities = **long-range frequencies**



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

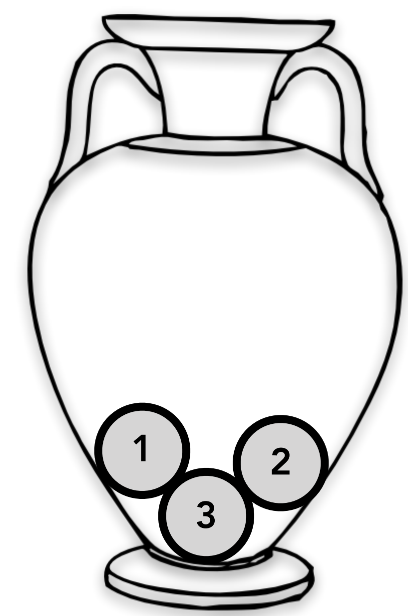
Subjective interpretation

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

Classical interpretation

Probabilities = **computed based on our knowledge of the situation**

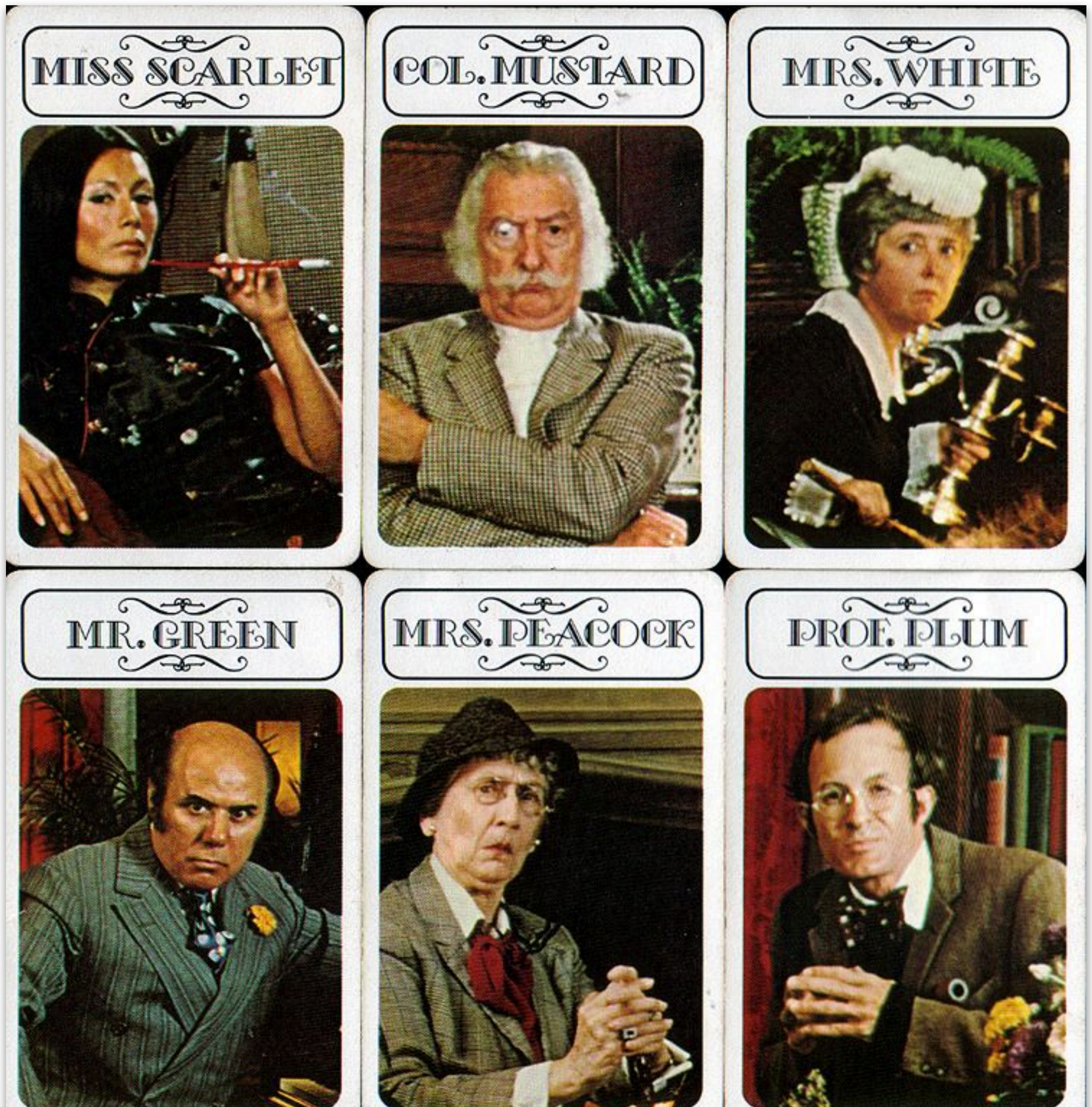


developed by analyzing games of chance

Clue guide to probability

Clue guide to probability

Who killed Mr Boddy?

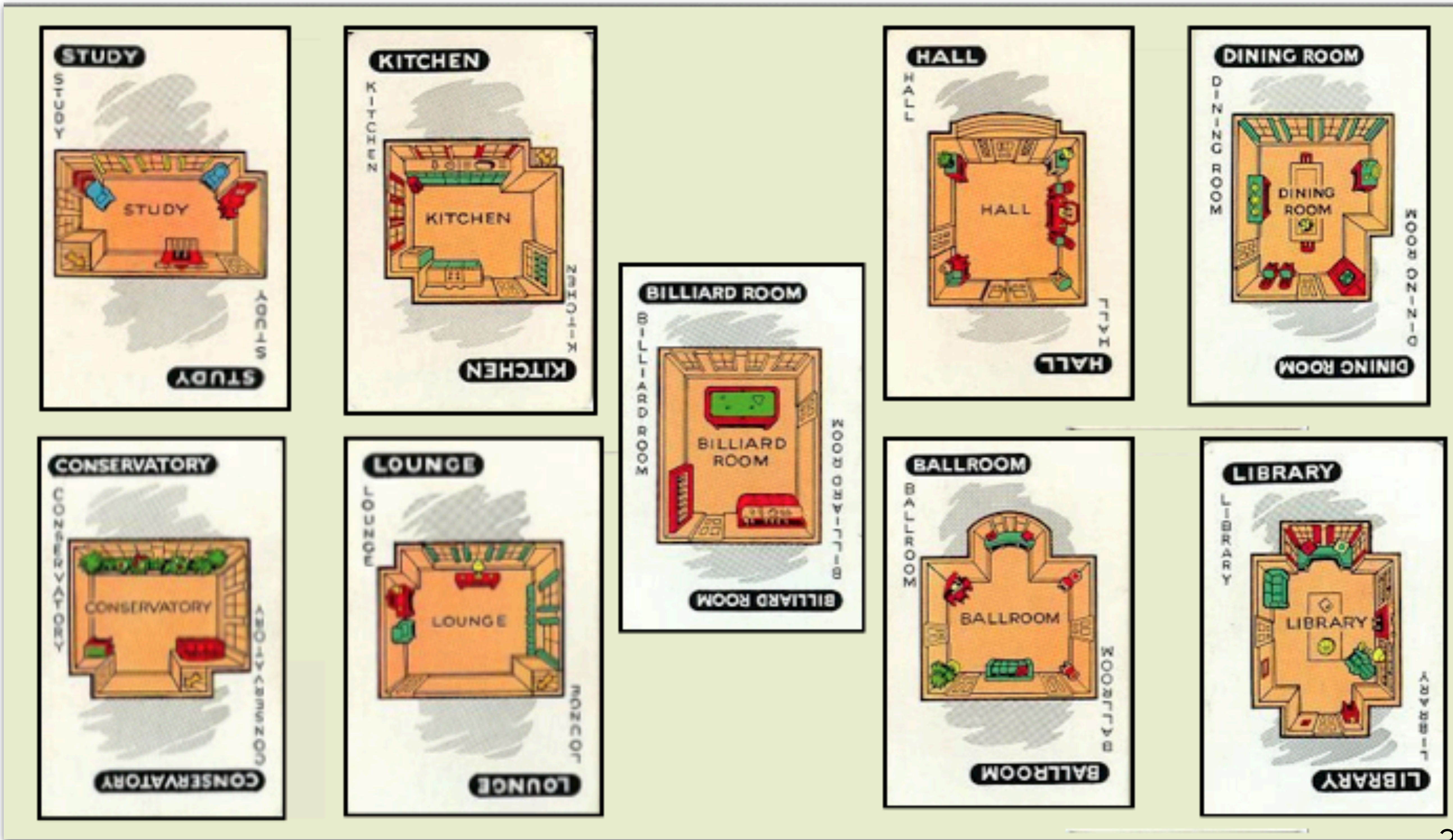


Clue guide to probability

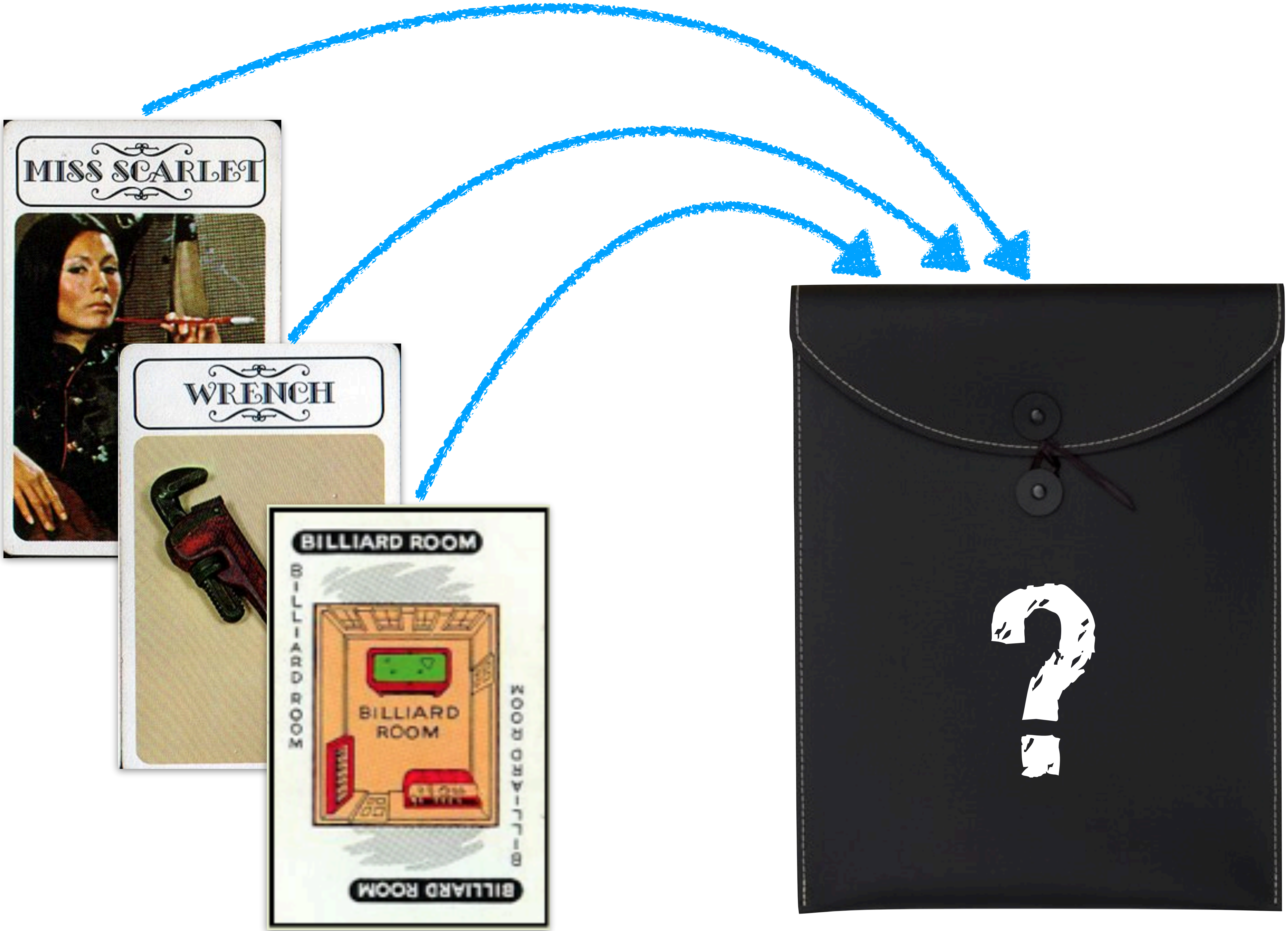
Who killed Mr Boddy, **with what?**

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

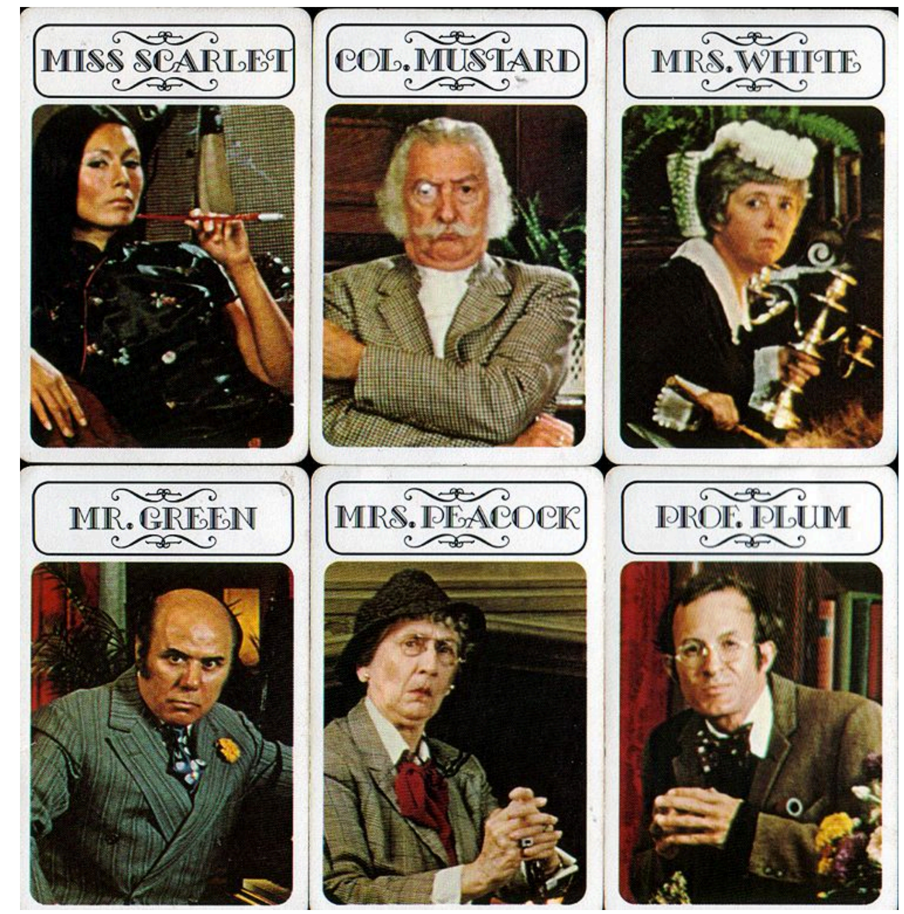
who	what	where
ms_scarlet	candlestick	study
ms_scarlet	candlestick	kitchen
ms_scarlet	candlestick	conservatory
ms_scarlet	candlestick	lounge

⋮

`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
- $p(\text{who} = \text{Prof. Plum}) = \dfrac{1}{6}$

for mutually exclusive events

otherwise

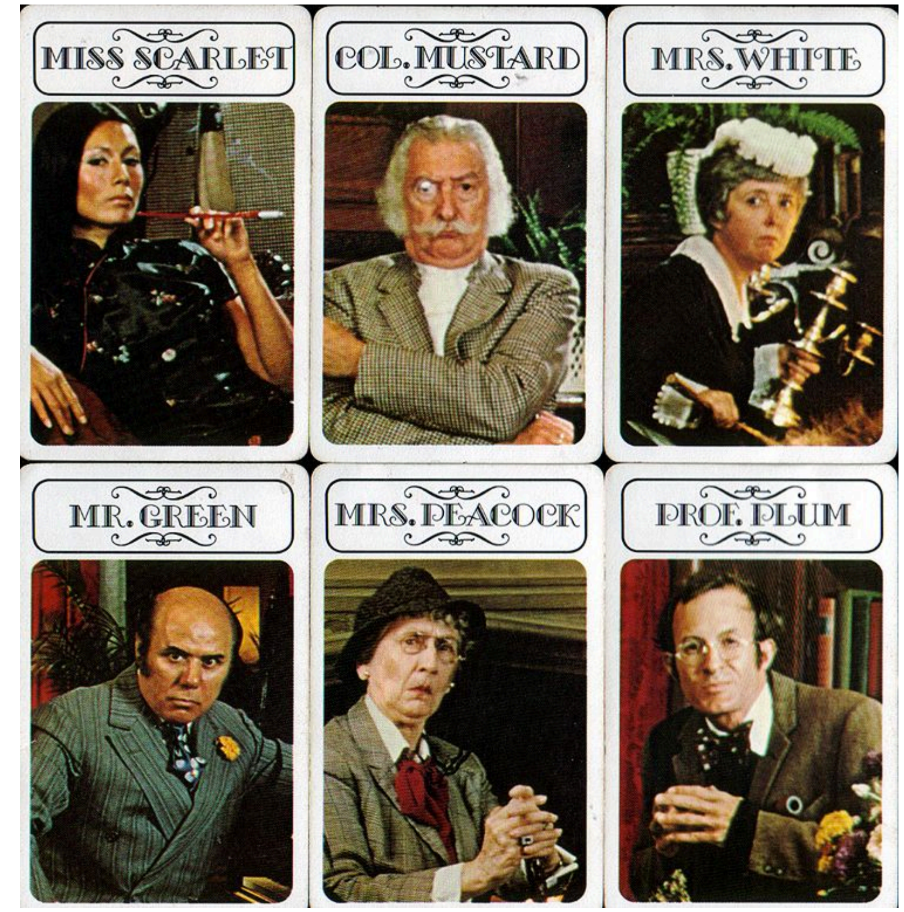
- $p(A \cup B) = p(A) + p(B)$

$$p(A \cup B) = p(A) + p(B) - p(A, B)$$

- $p(\text{who} = \text{Prof. Plum} \cup \text{Mrs. White}) = \dfrac{2}{6}$

Clue guide to probability

Who?

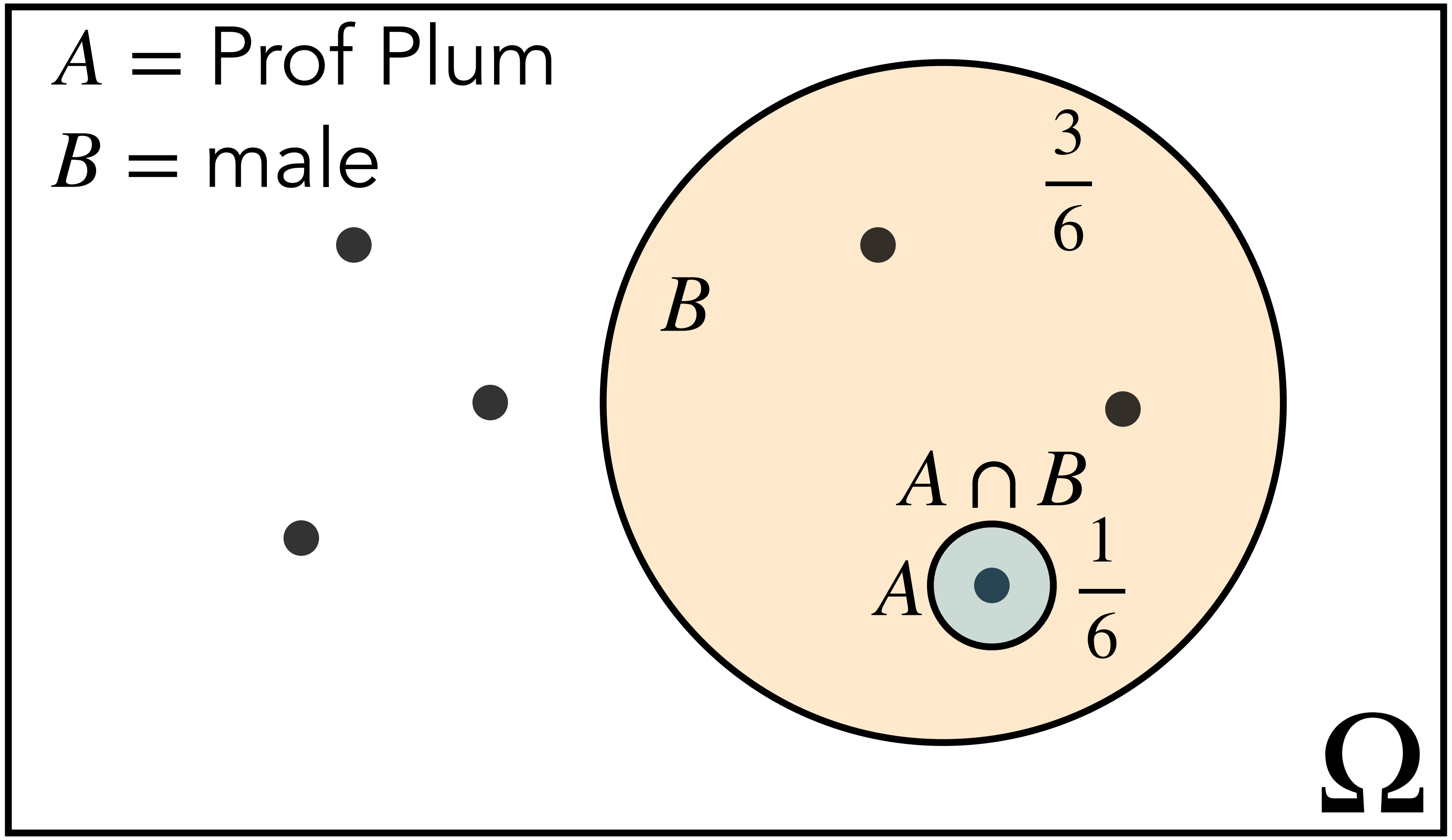


- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A, B)}{p(B)}$

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

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

Clue guide to probability



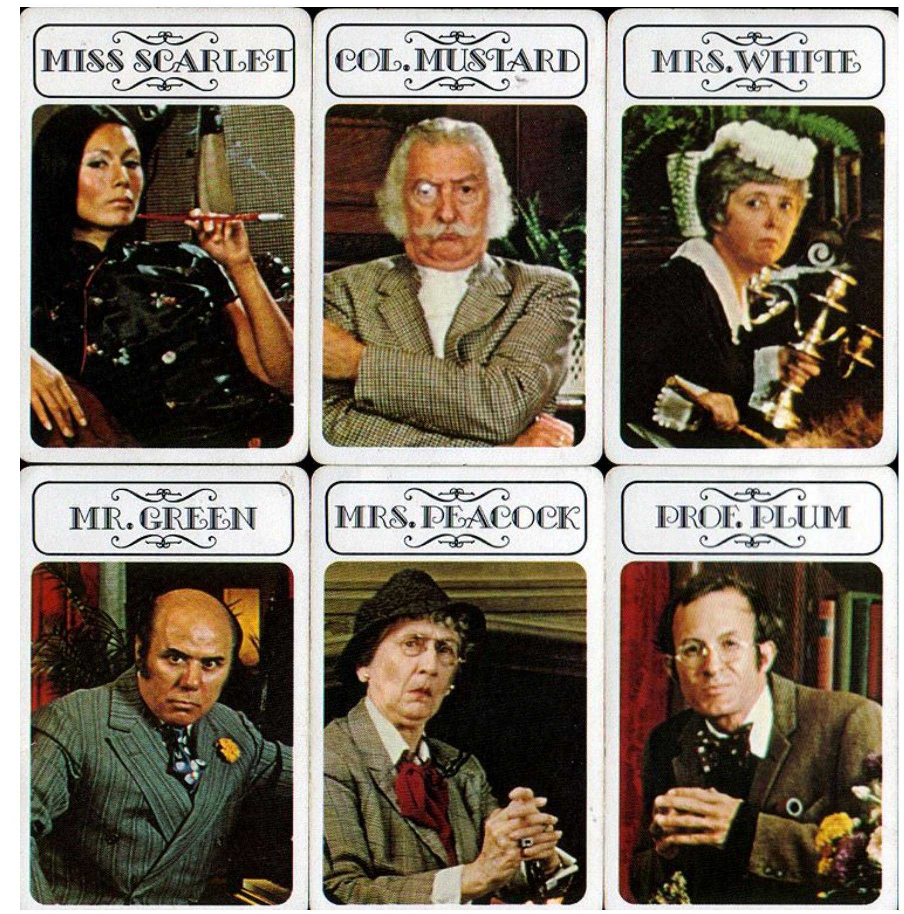
Probability that it was Prof Plum, given that the murderer was male?

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

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

Clue guide to probability

Who?



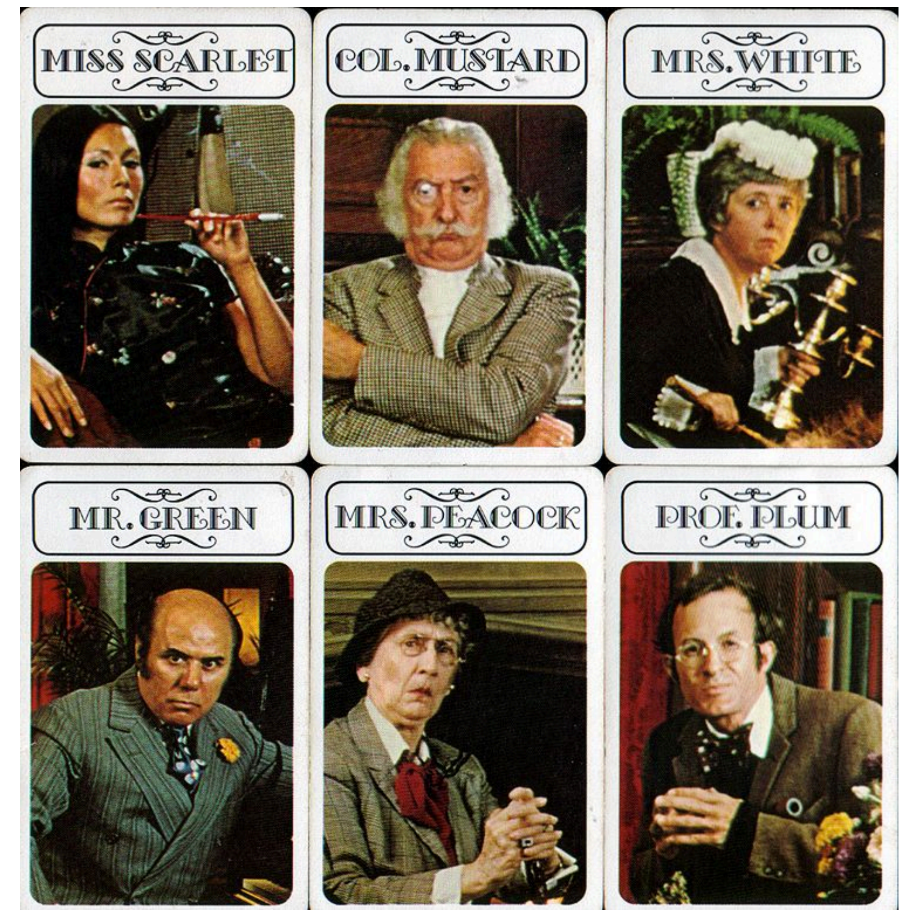
- *conditional probability*:
- $p(A \mid B)$ (probability of A given B)
- **Definition**: $p(A \mid B) = \dfrac{p(A, B)}{p(B)}$
- $p(\text{Prof. Plum} \mid \text{male}) = \dfrac{1/6}{1/2} = 1/3$

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

```r
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**: $p(A\,|\,B) = \dfrac{p(A,B)}{p(B)}$
- $p(\text{Prof. Plum}\,|\,\text{male}) = \dfrac{1/6}{1/2} = 1/3$

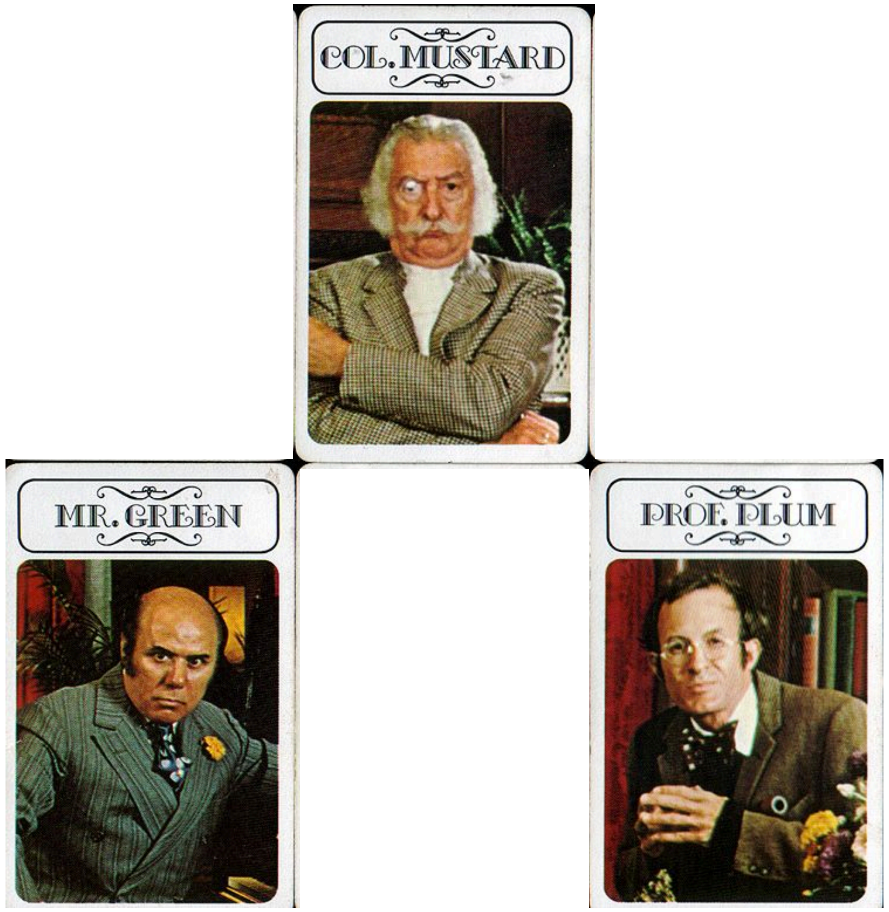
who	gender
col_mustard	male
mr_green	male
prof_plum	male
ms_scarlet	female
mrs_white	female
mrs_peacock	female

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

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)

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

Who?

What?

Clue guide to probability

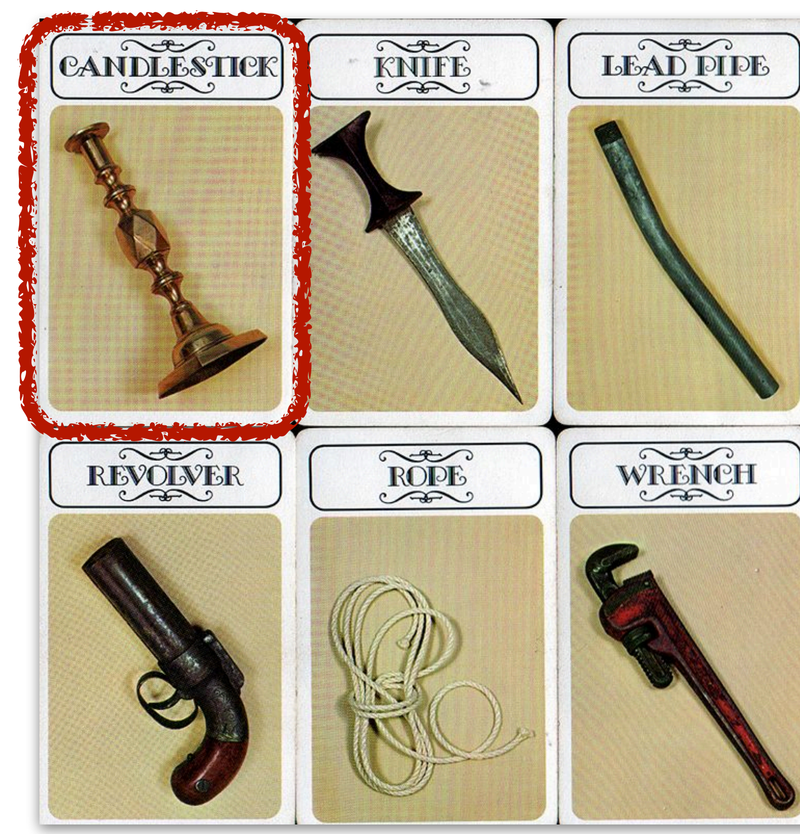
- *joint probability*:
- if A and B are independent then
- **Definition**: $p(A, B) = p(A) \cdot p(B)$

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

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

Who?

What?



Clue guide to probability

- *dependence*:
- **Definition**: $p(A\,|\,B) \neq p(A)$
- **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?

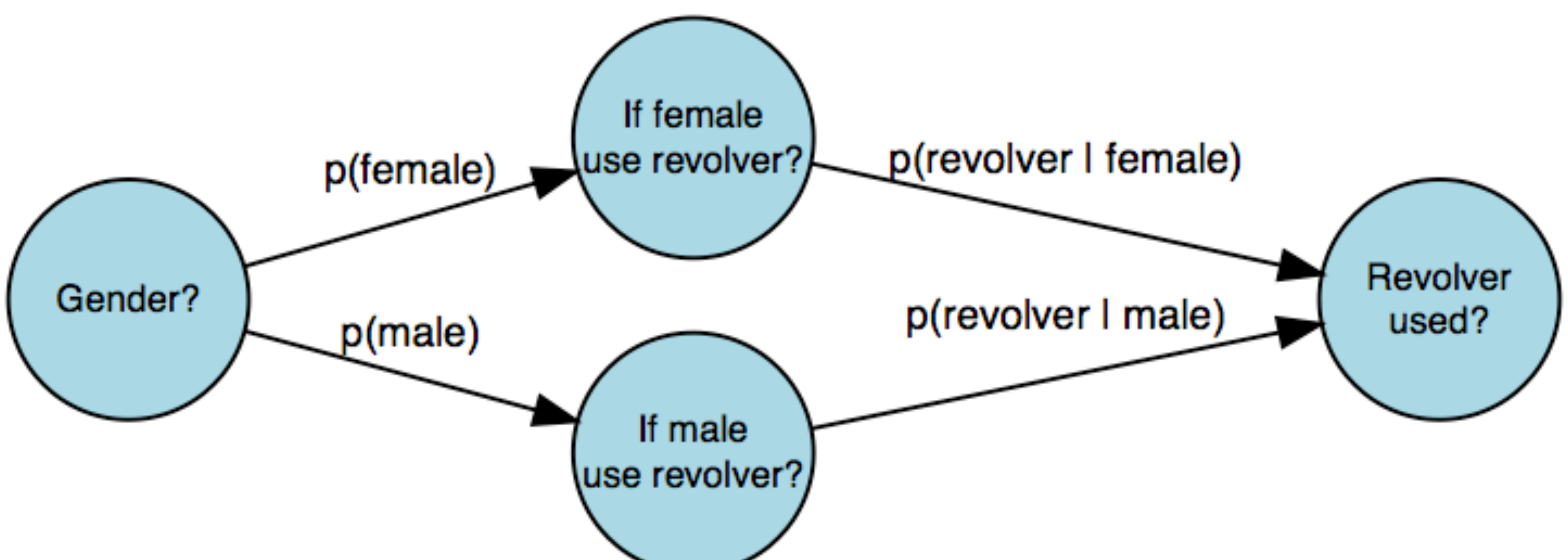
Clue guide to probability

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

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

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

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



Who?

What?

- *Bayes' rule (derivation)*

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

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

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

Clue guide to probability

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

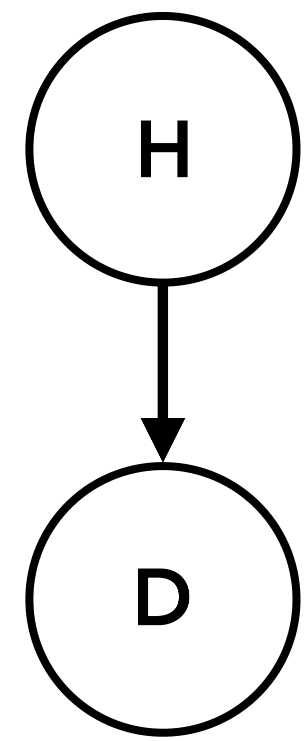
posterior likelihood prior

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

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

$$\overset{\text{posterior}}{p(H \mid D)} = \frac{\overset{\text{likelihood}}{p(D \mid H)} \cdot \overset{\text{prior}}{p(H)}}{\underset{\text{probability of the data?!}}{p(D)}}$$

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).

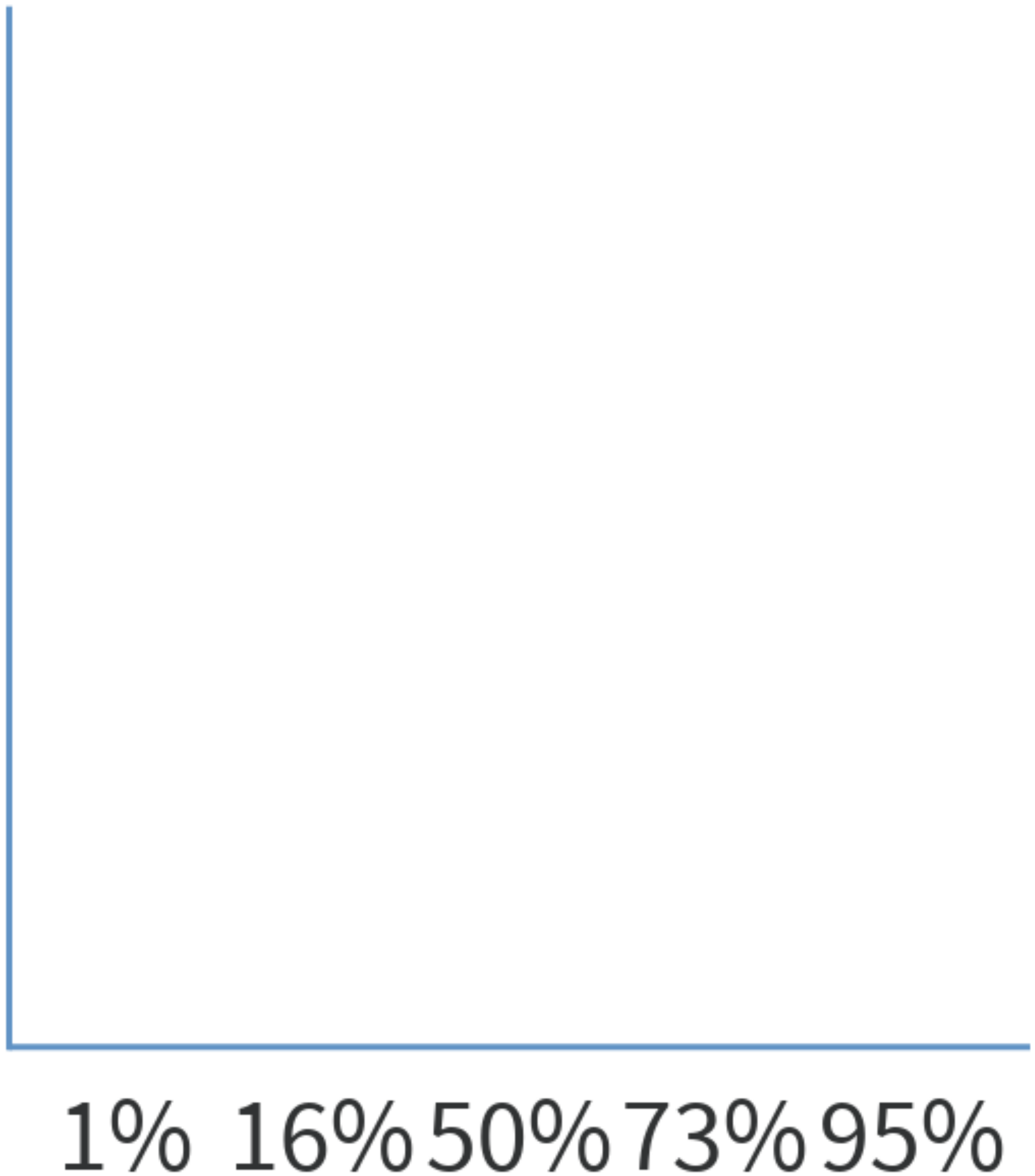
Find the conditional probability that Fred has *conditionitis*, given the evidence provided by the positive test result.

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

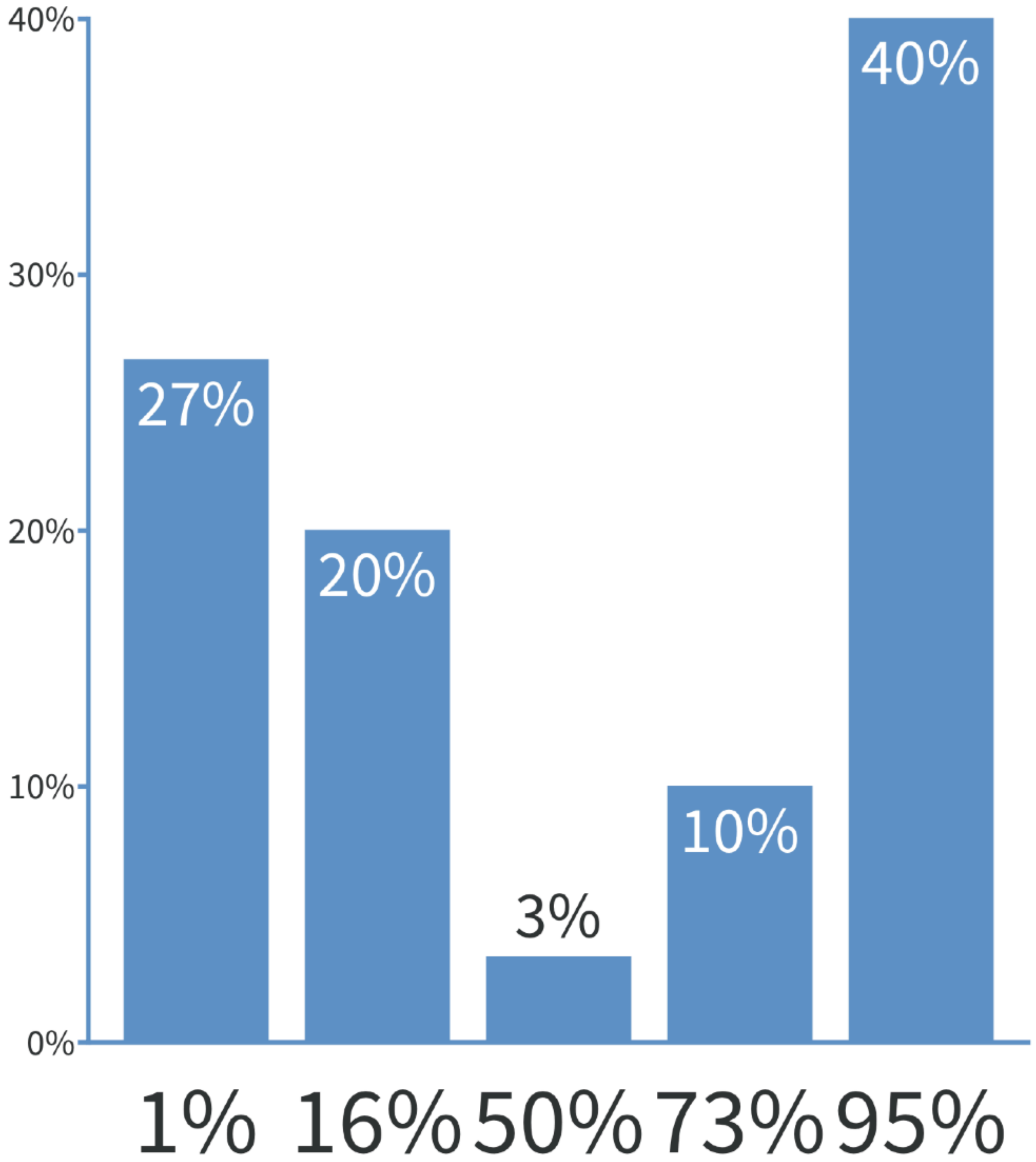


What's the probability that Fred has conditionitis?

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(¬T|¬D) = 0.95**. The quantity P(T|D) is known as the *sensitivity* (= true positive rate) of the test, and P(¬T|¬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

what we know

$P(D) = 0.01$

$P(T \mid D) = 0.95$

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

what we want to know

$P(D \mid T) = ?$

$$p(D \mid T) = \frac{p(T \mid D) \cdot p(D)}{p(T)}$$ Bayes' rule

$$p(D \mid T) = \frac{p(T \mid D) \cdot p(D)}{p(T \mid D) \cdot p(D) + p(T \mid \neg D) \cdot p(\neg D)}$$ law of total probability

$$p(D \mid T) = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.16$$

Clue guide to probability

what we know

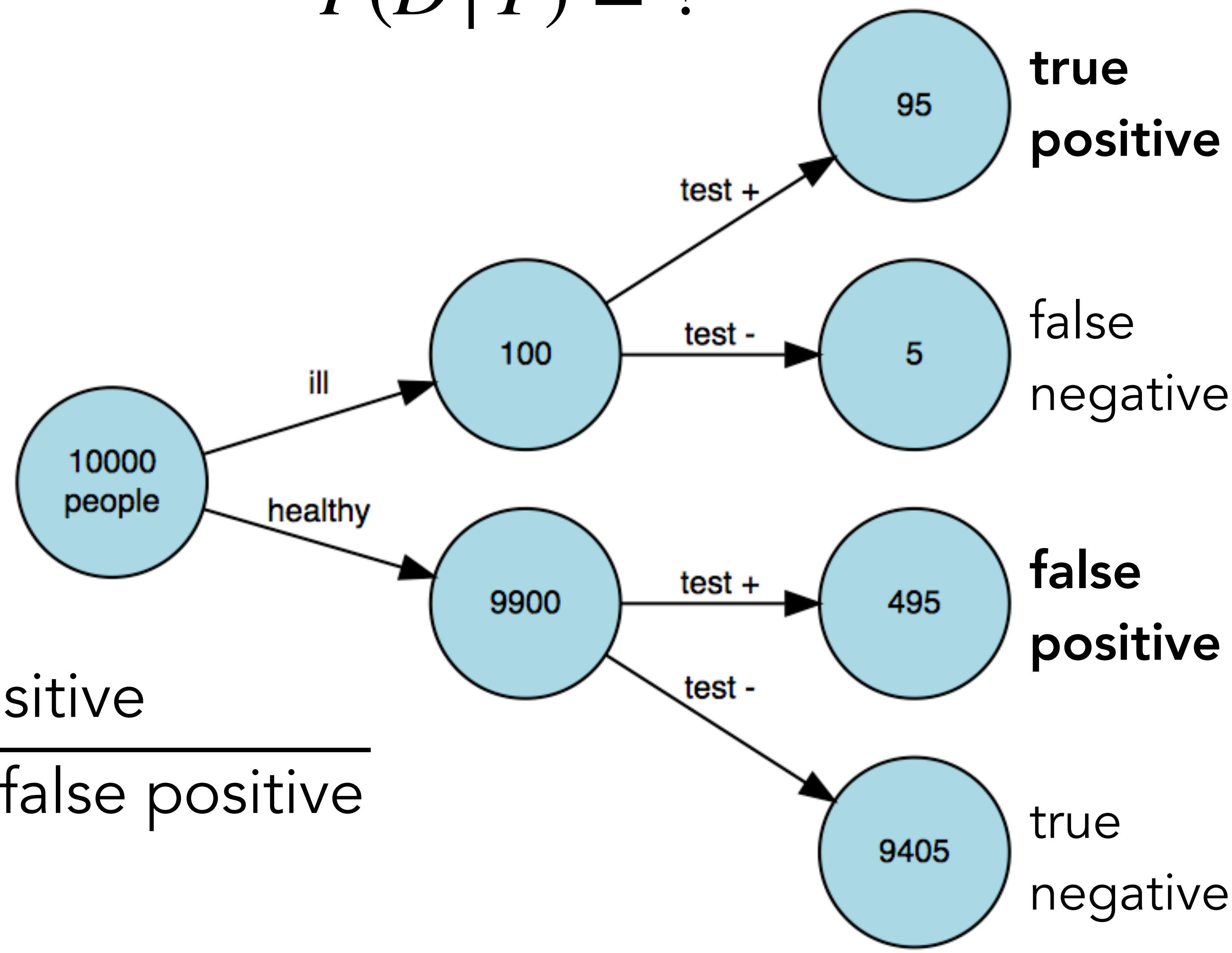
$$P(D) = 0.01$$

$$P(T \mid D) = 0.95$$

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

what we want to know

$$P(D \mid T) = ?$$

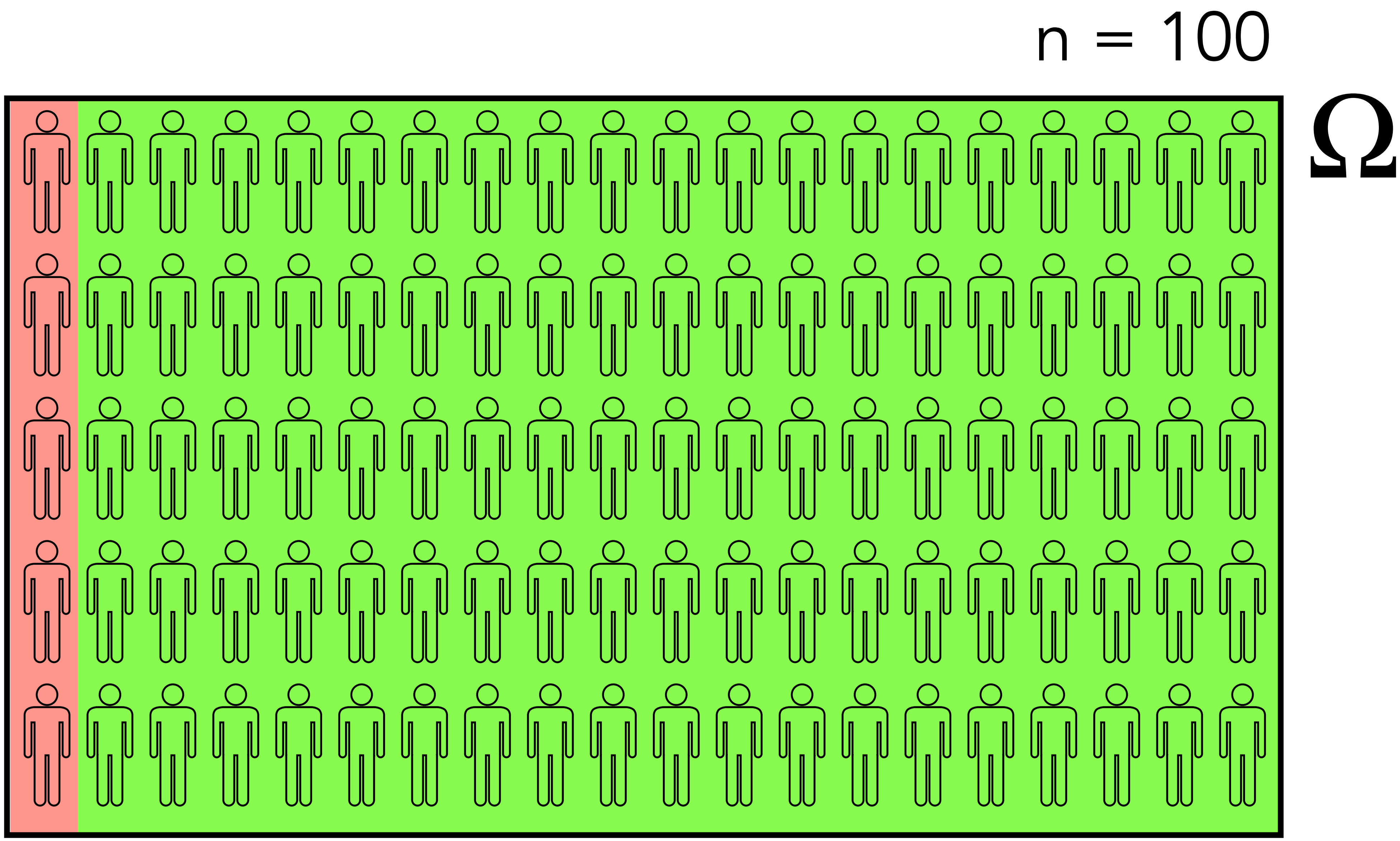


$$P(D \mid T) = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$= \frac{95}{95 + 495}$$

$$\approx 0.16$$

Clue guide to probability

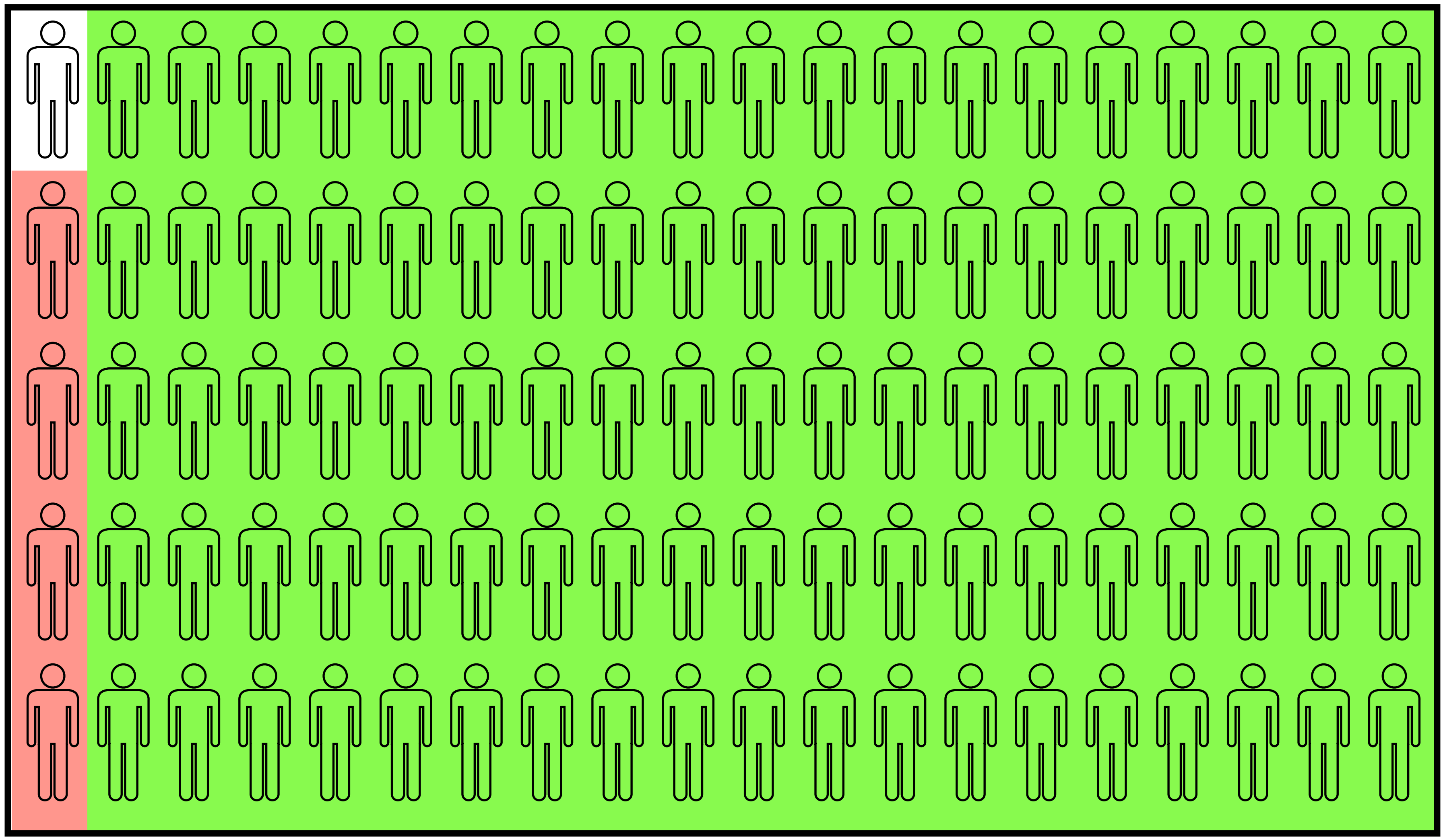


$p(D) = 0.05$ $\quad$ $p(\neg D) = 0.95$

prior

Clue guide to probability

n = 100



Ω

likelihood

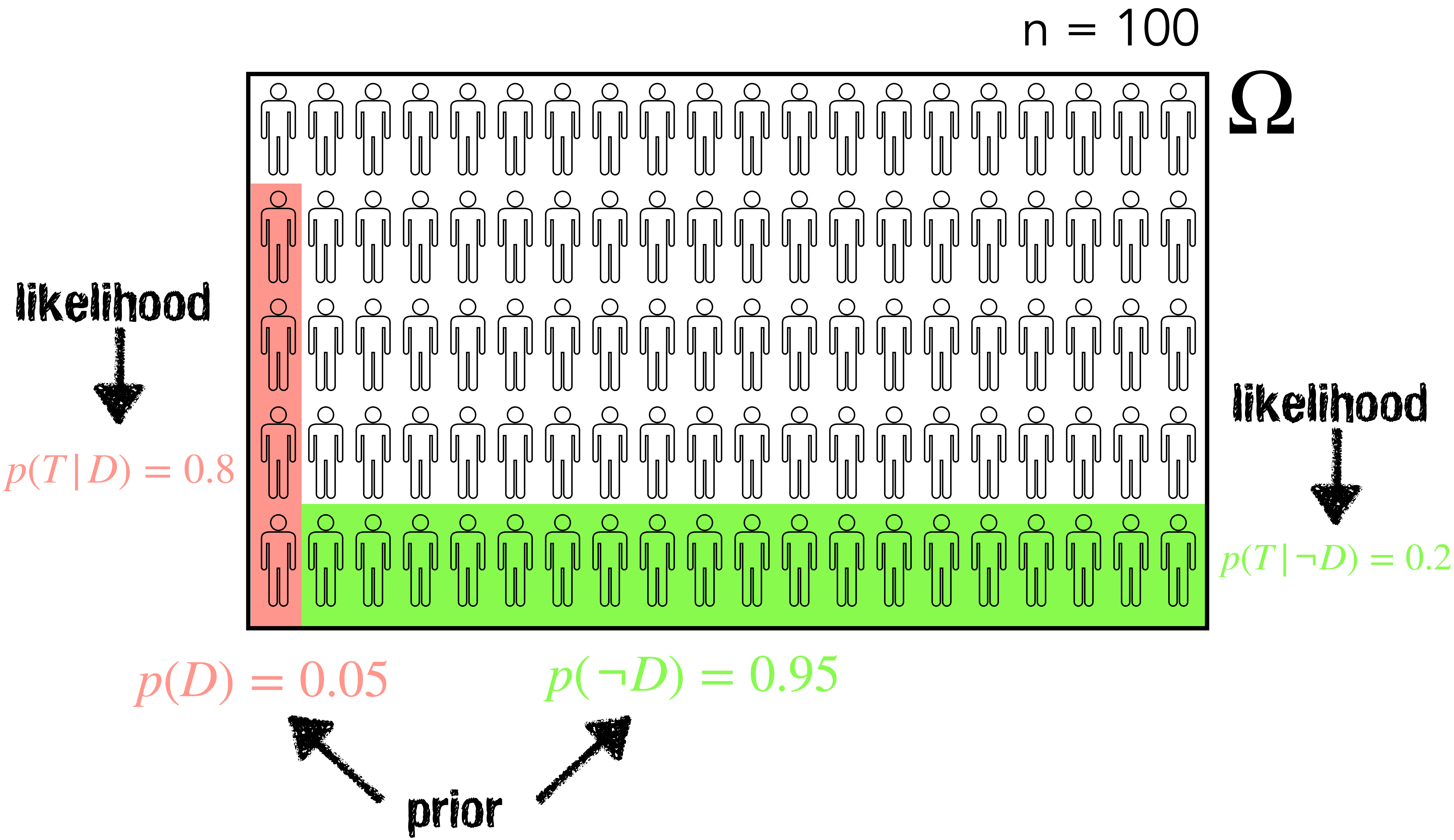
$p(T \mid D) = 0.8$

$p(D) = 0.05$

$p(\neg D) = 0.95$

prior

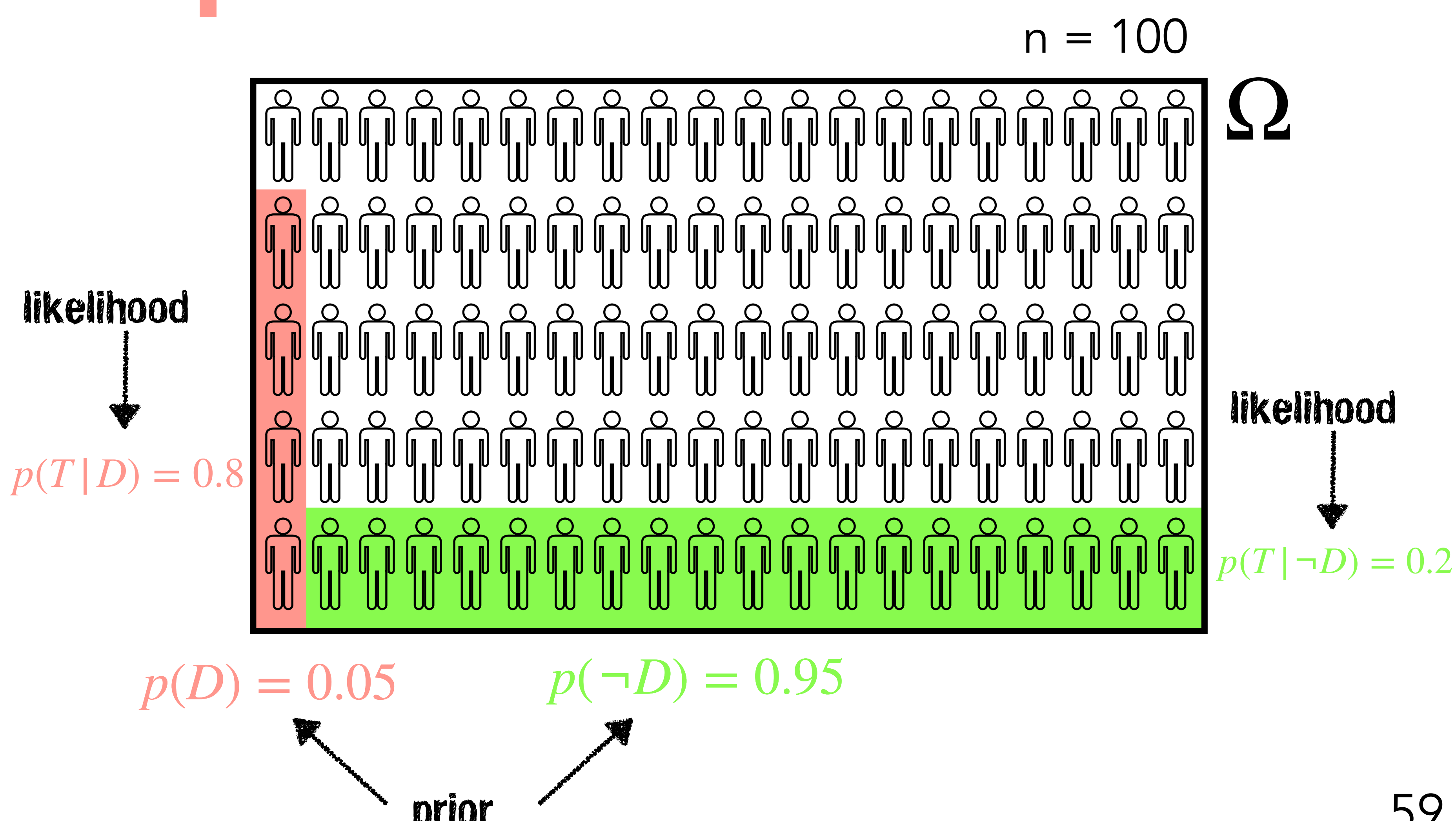
Clue guide to probability

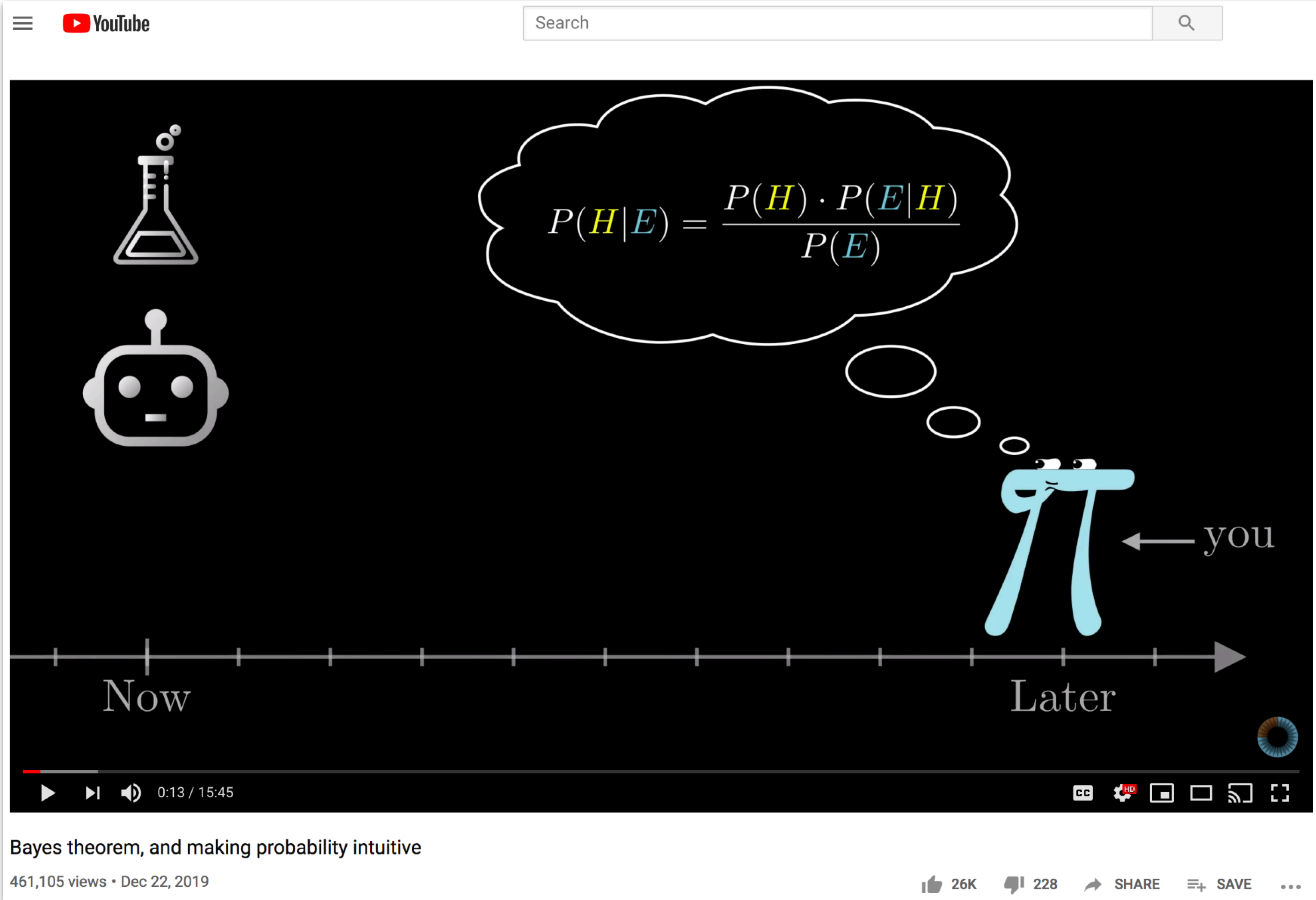


Clue guide to probability

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

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



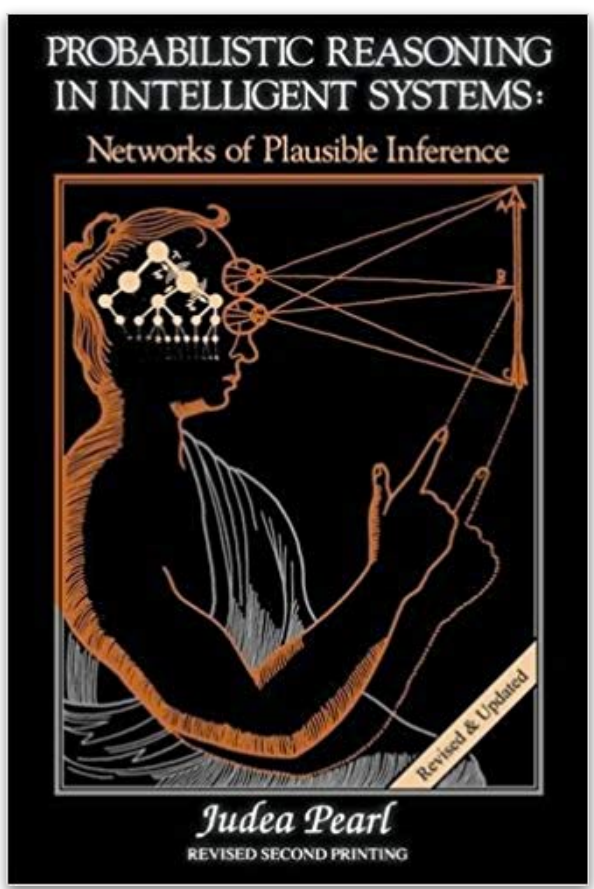


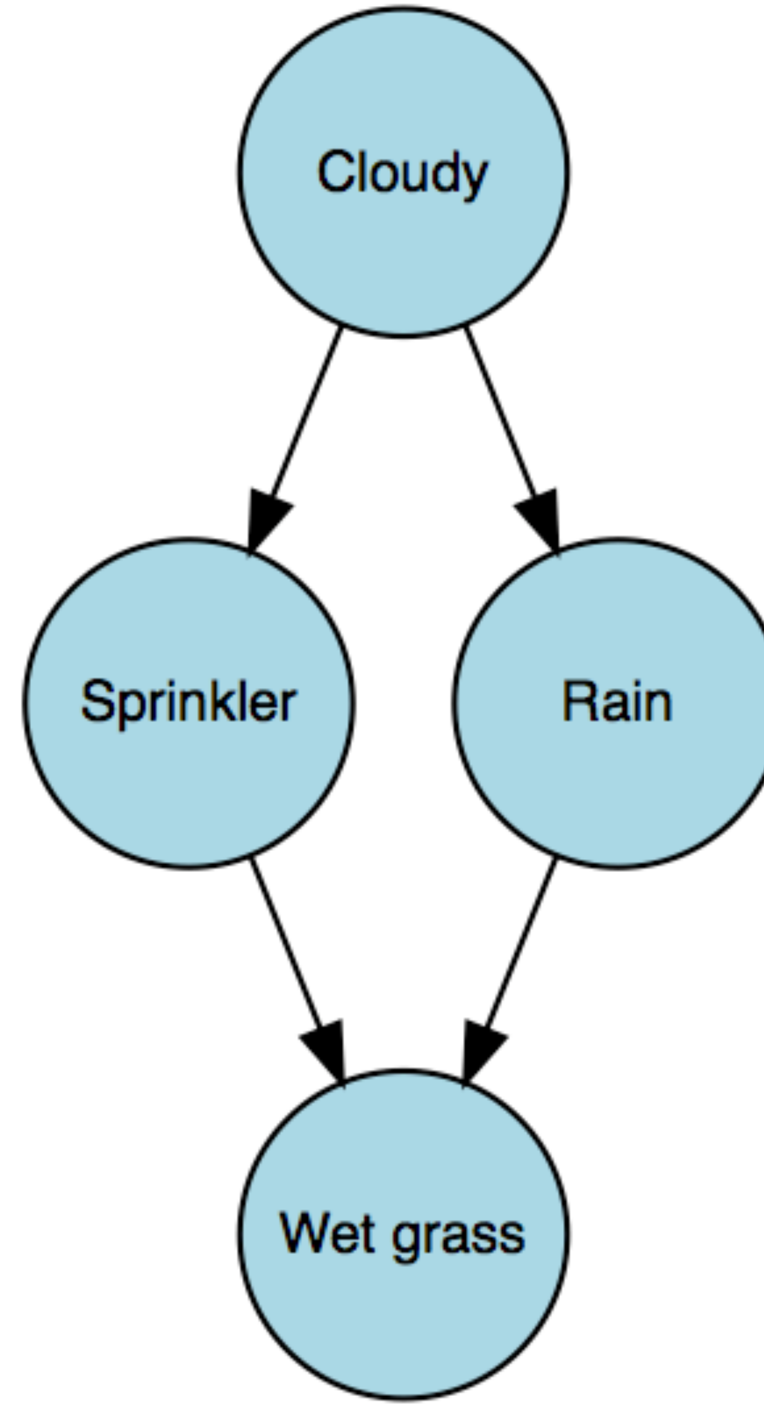
https://www.youtube.com/watch?v=HZGCoVF3YvM&feature=youtu.be

Bayesian Networks

Representation

- **nodes** represent variables of interest
- **links** represent direct dependencies between variables
- **conditional probability tables** parameterize the model



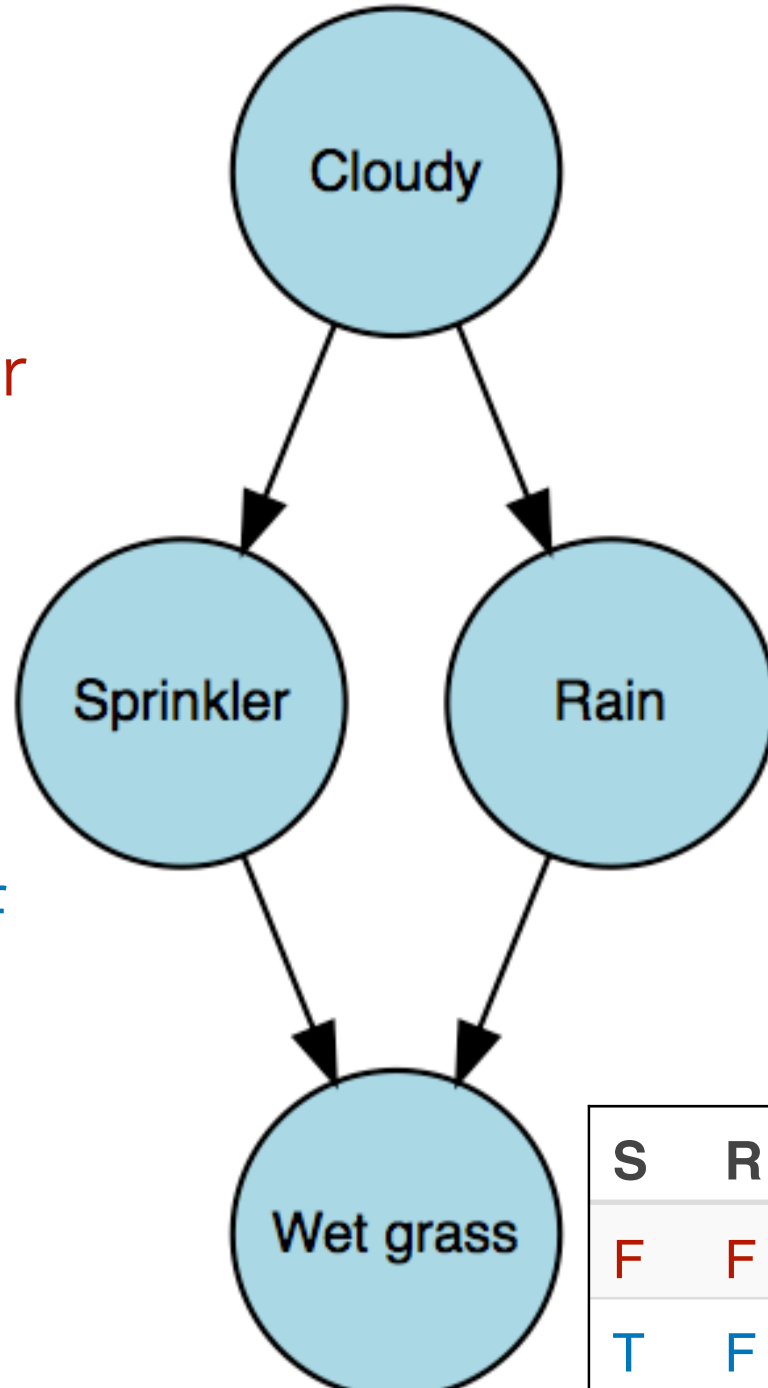


Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Francisco, CA: Morgan Kaufmann.

Representation

p(C)
0.5

equally likely to be cloudy or not



equally likely to be on or off when it's not cloudy

C	p(S)
F	0.5
T	0.1

normally gets turned off when its cloudy

it doesn't rain when it's not cloudy

C	p(R)
F	0
T	0.3

it sometimes rains when it's cloudy

S	R	p(W)
F	F	0.1
T	F	0.90
F	T	0.90
T	T	0.99

wet grass unlikely when neither

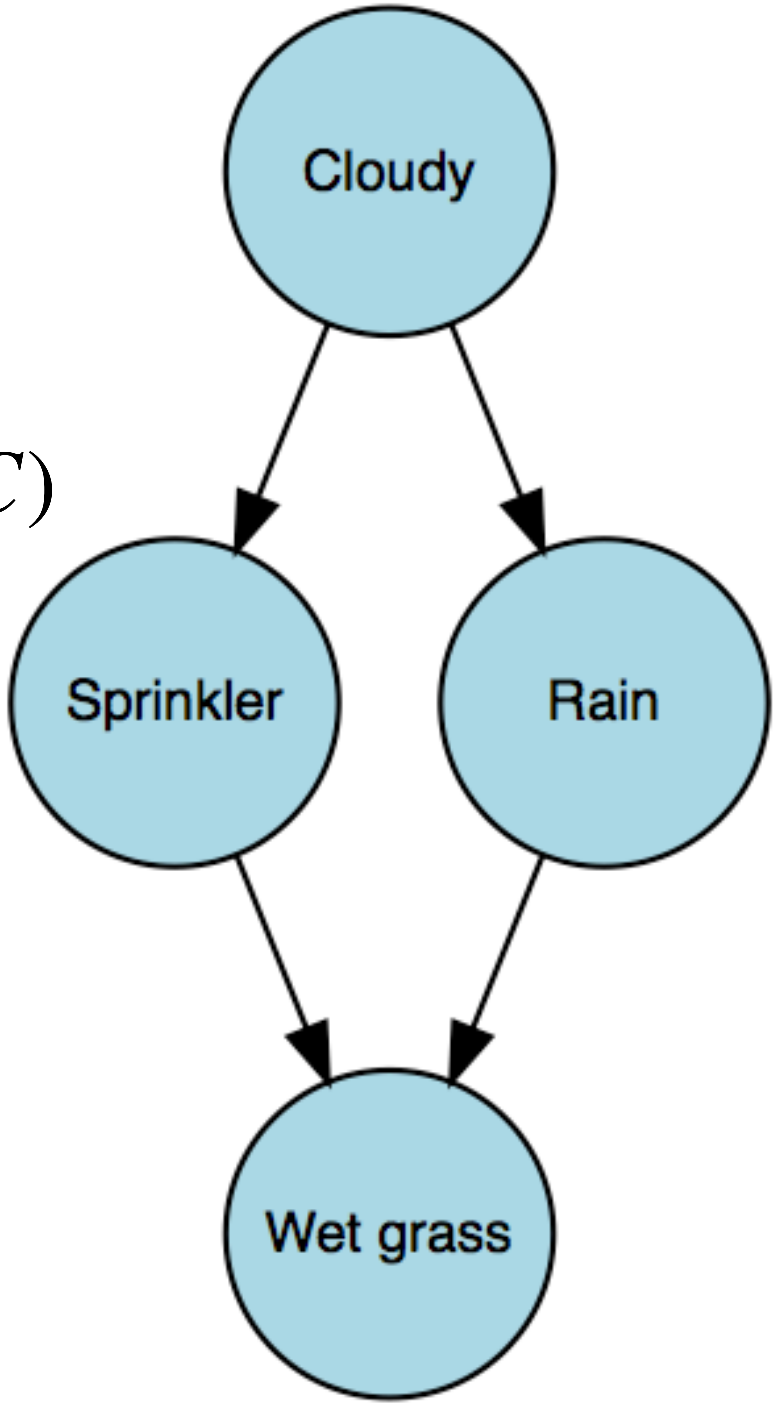
likely wet when either is true

Compact representation of the joint probability distribution

chain rule

$$p(C,S,R,W) = p(W|CSR)\cdot p(CSR)$$
$$= p(W|CSR)\cdot p(R|CS)\cdot p(CS)$$
$$\vdots$$
$$= p(W|CSR)\cdot p(R|CS)\cdot p(S|C)\cdot p(C)$$

number of parameters 8 + 4 + 2 + 1 = 15



considering independence

$$p(C,S,R,W) = p(W|SR)\cdot p(S|C)\cdot p(R|C)\cdot p(C)$$

number of parameters 4 + 2 + 2 + 1 = 9

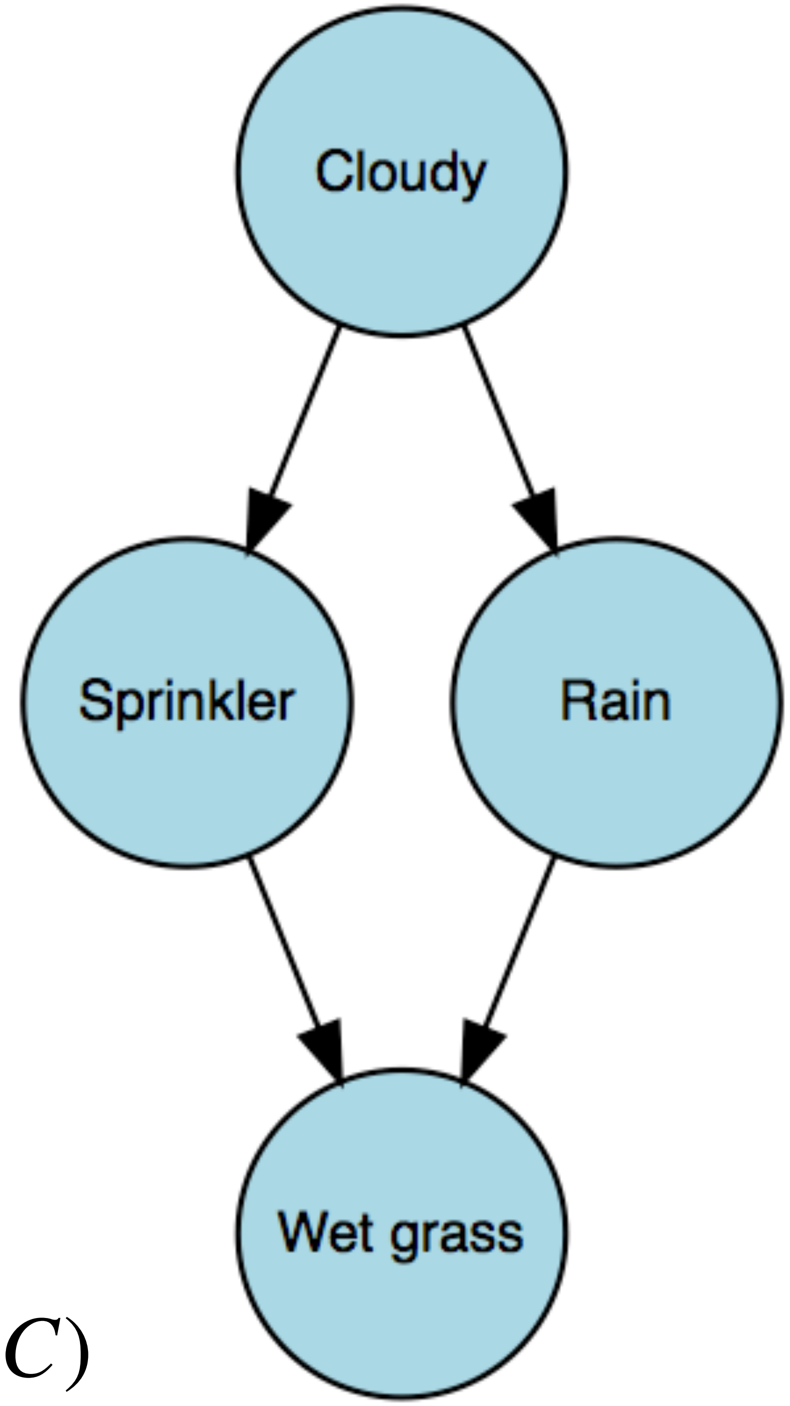
Inference by conditioning

p(C)
0.5

- supports *predictive* reasoning:

$$p(W|C) = ?$$

probability of wet grass given that it's cloudy?



$$p(W|C) = \frac{p(W,C)}{p(C)}$$

$$p(C) = 0.5$$

$$p(W,C) = \sum_{s=0}^{1}\sum_{r=0}^{1} p(C,S,R,W)$$

$$= \sum_{s=0}^{1}\sum_{r=0}^{1} p(W|SR) \cdot p(S|C) \cdot p(R|C) \cdot p(C)$$

Inference by conditioning

- supports *predictive* reasoning:

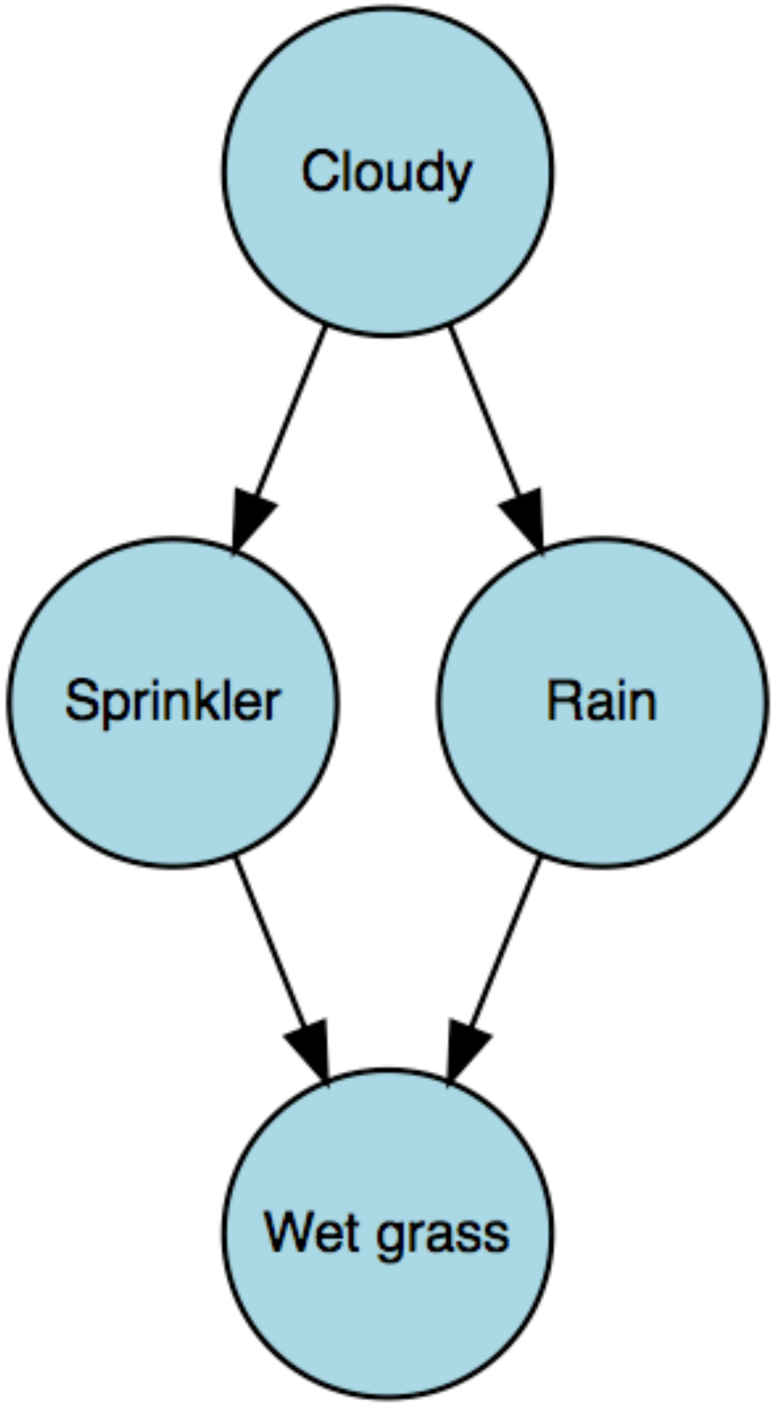
$$p(W \mid C) = ?$$

probability of wet grass
given that it's cloudy?

- and *diagnostic* reasoning:

$$p(R \mid W) = ?$$

probability that it rained
given that the grass is wet?



Patterns of inference: **Common cause**

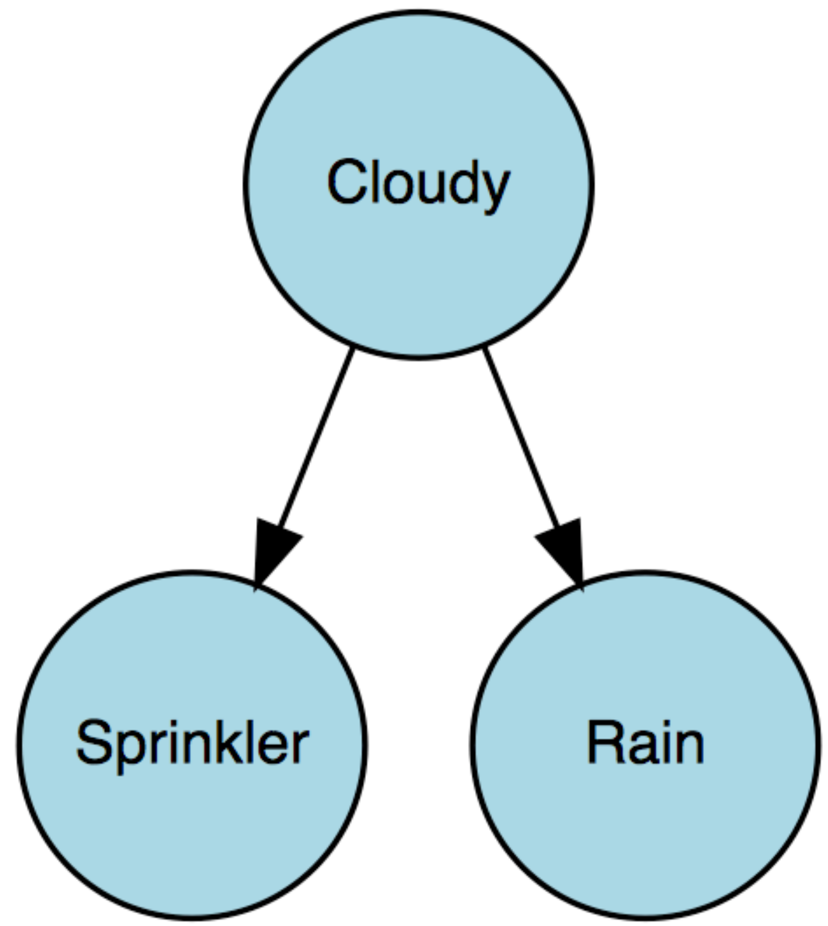
$$p(S \mid R) = p(S)$$

or

$$p(S \mid R) \neq p(S)$$

?

p(C)
0.5



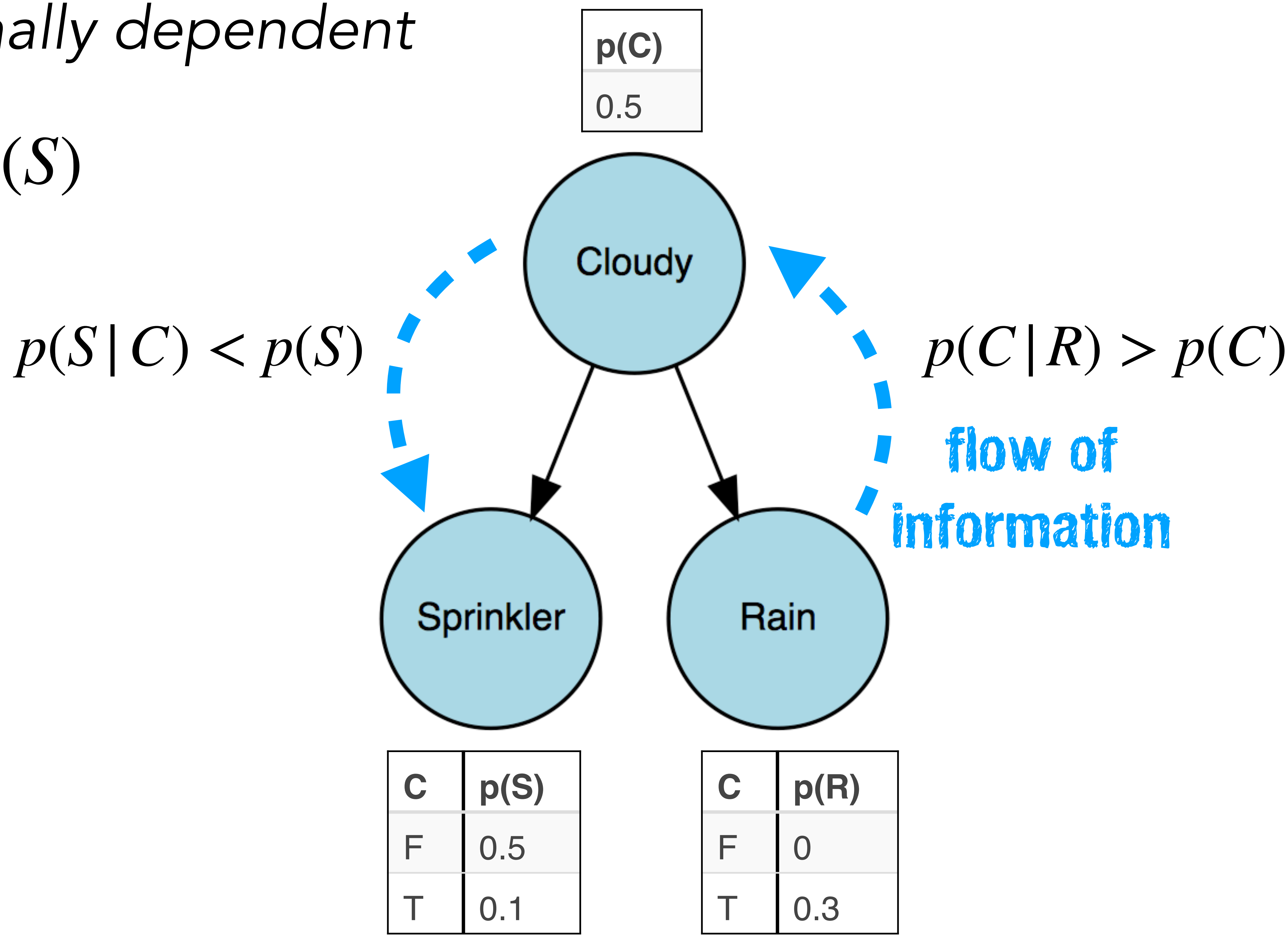
C	p(S)
F	0.5
T	0.1

C	p(R)
F	0
T	0.3

Patterns of inference: **Common cause**

- effects of a common cause are *unconditionally dependent*

$$p(S \mid R) \neq p(S)$$

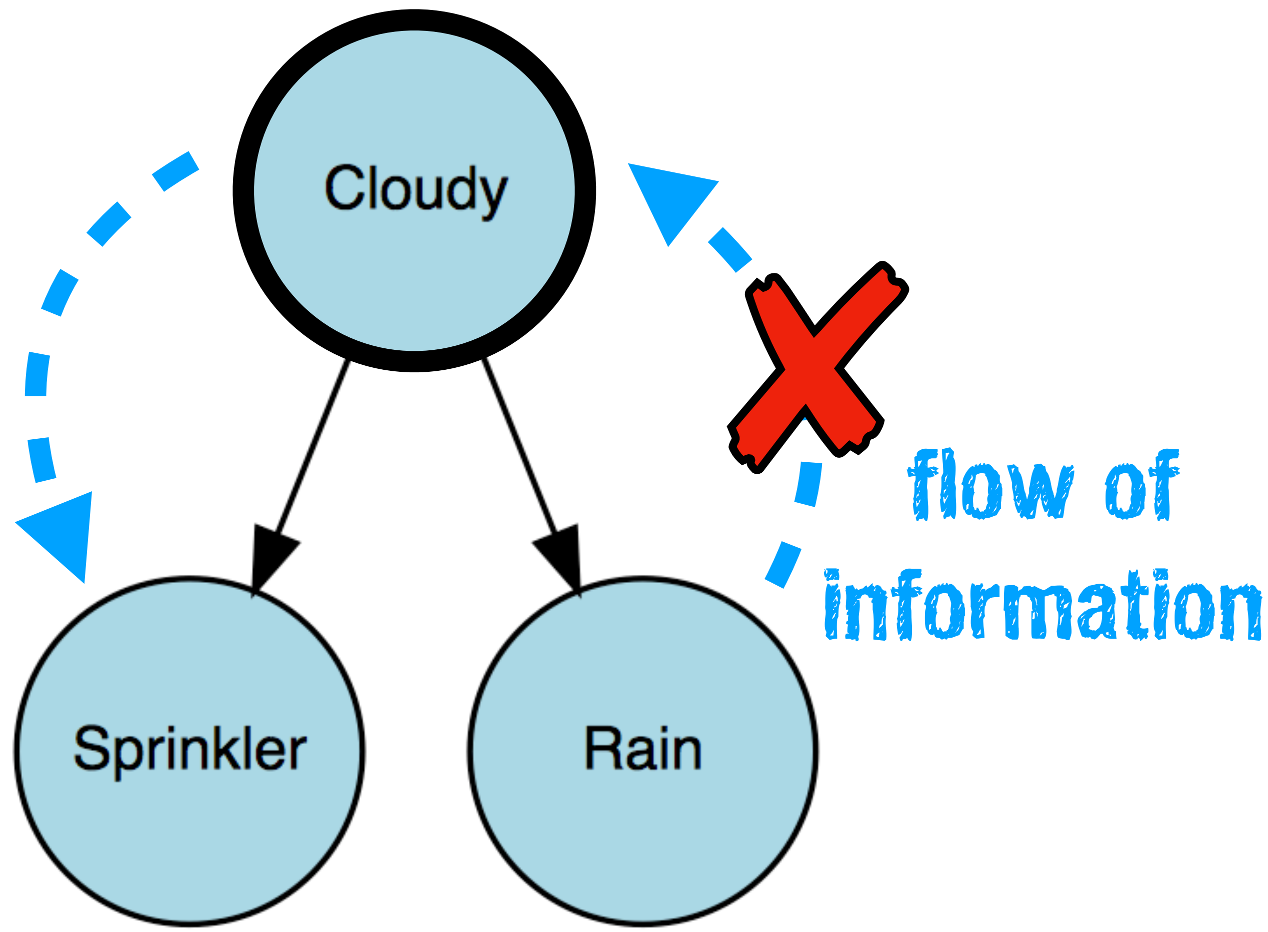


Patterns of inference: **Common cause**

- effects of a common cause are *conditionally independent given the cause*

$$p(S \mid R, C) = p(S \mid C)$$

p(C)
1



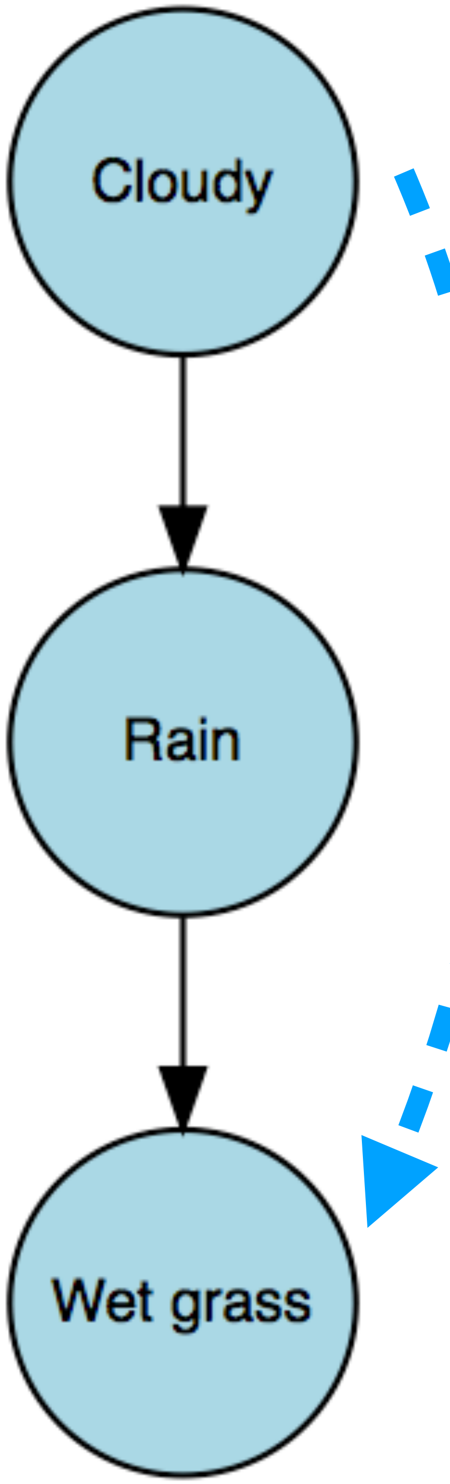
C	p(S)
F	0.5
T	0.1

C	p(R)
F	0
T	0.3

Patterns of inference: **Causal chain**

- cause and effect in a causal chain are *unconditionally dependent*

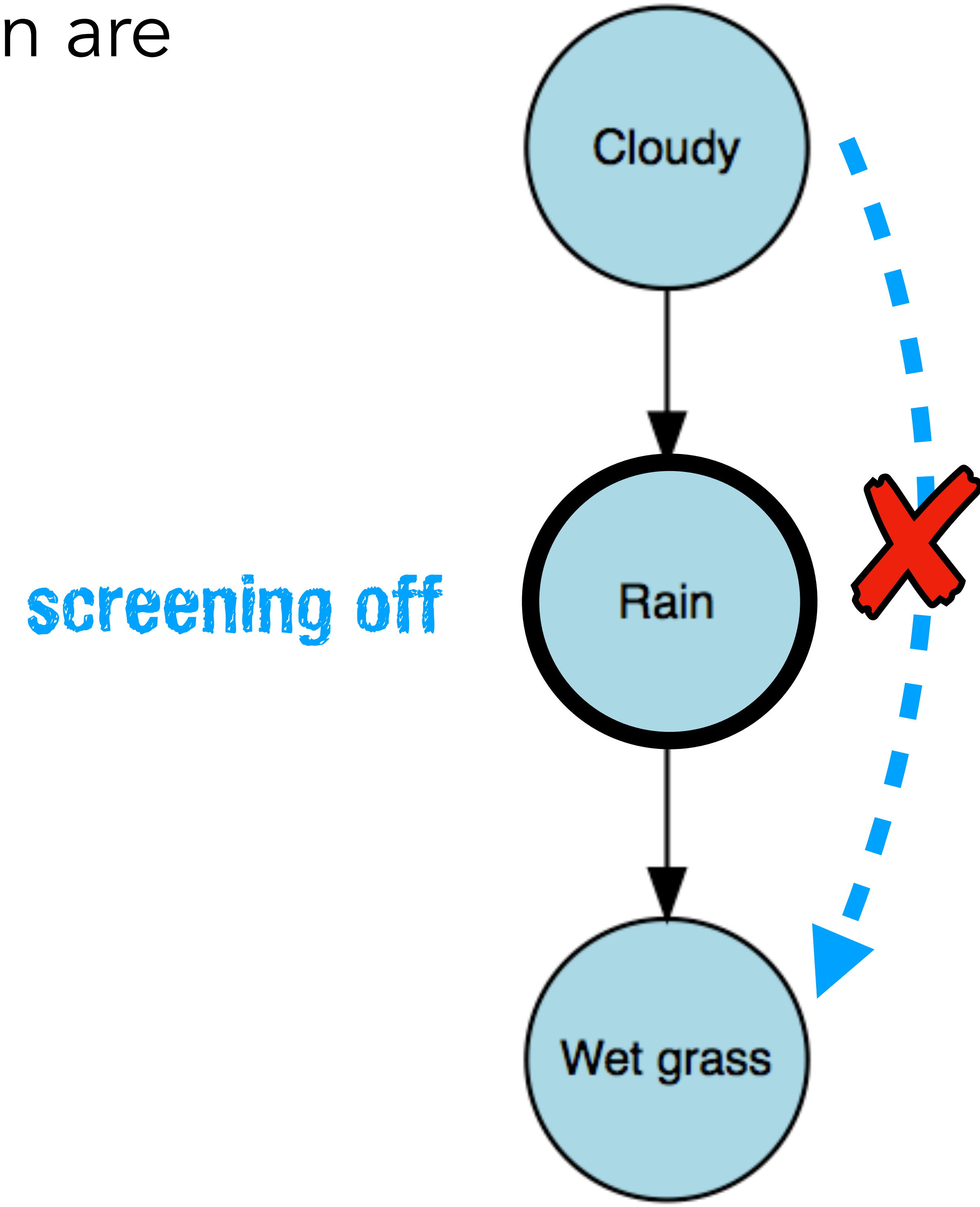
$$p(W \mid C) \neq p(W)$$



Patterns of inference: **Causal chain**

- cause and effect in a causal chain are *conditionally independent*

$$p(W \mid C, R) = p(W \mid R)$$

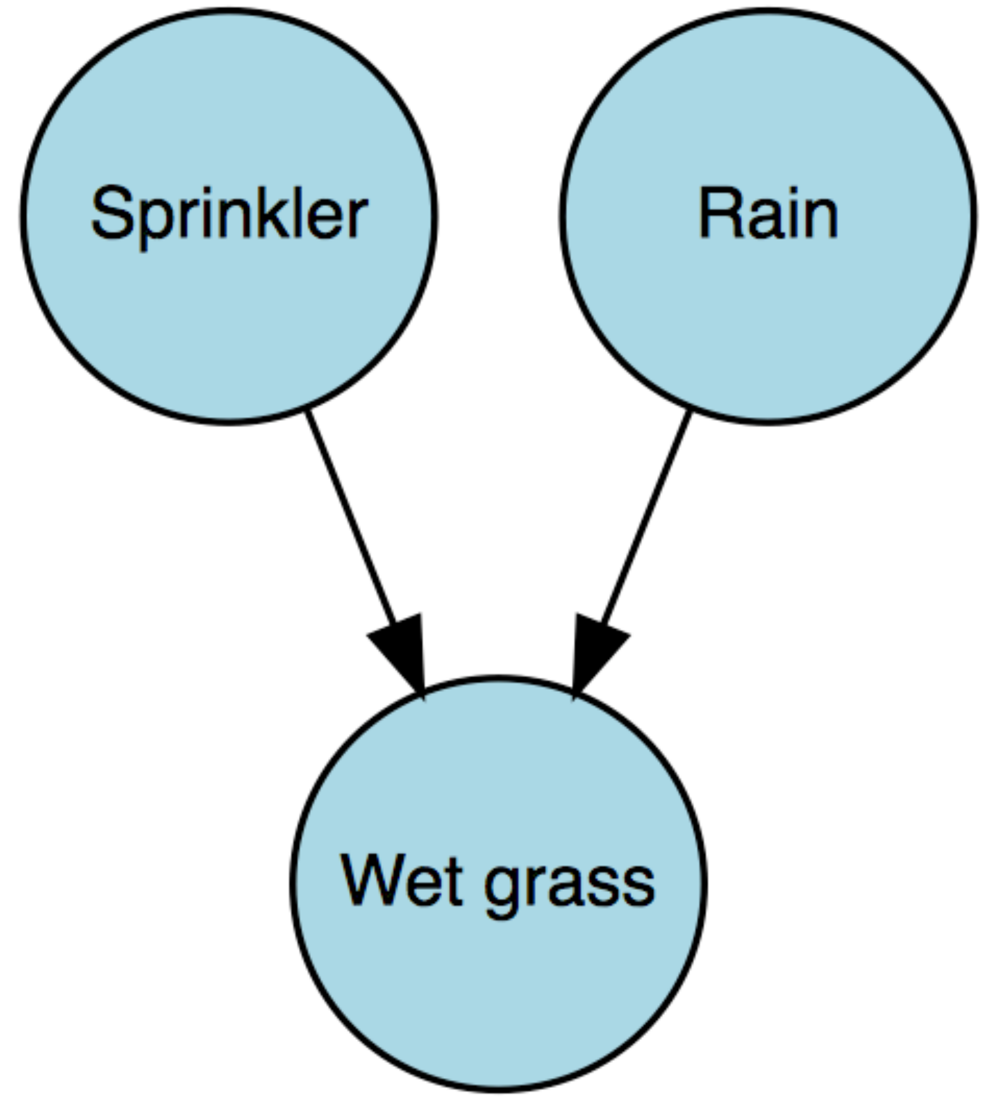


Patterns of inference: **Common effect**

- two causes of a common effect are *unconditionally independent*

$$p(S \mid R) = p(S)$$

(e.g. Sprinklers are set by a timer)

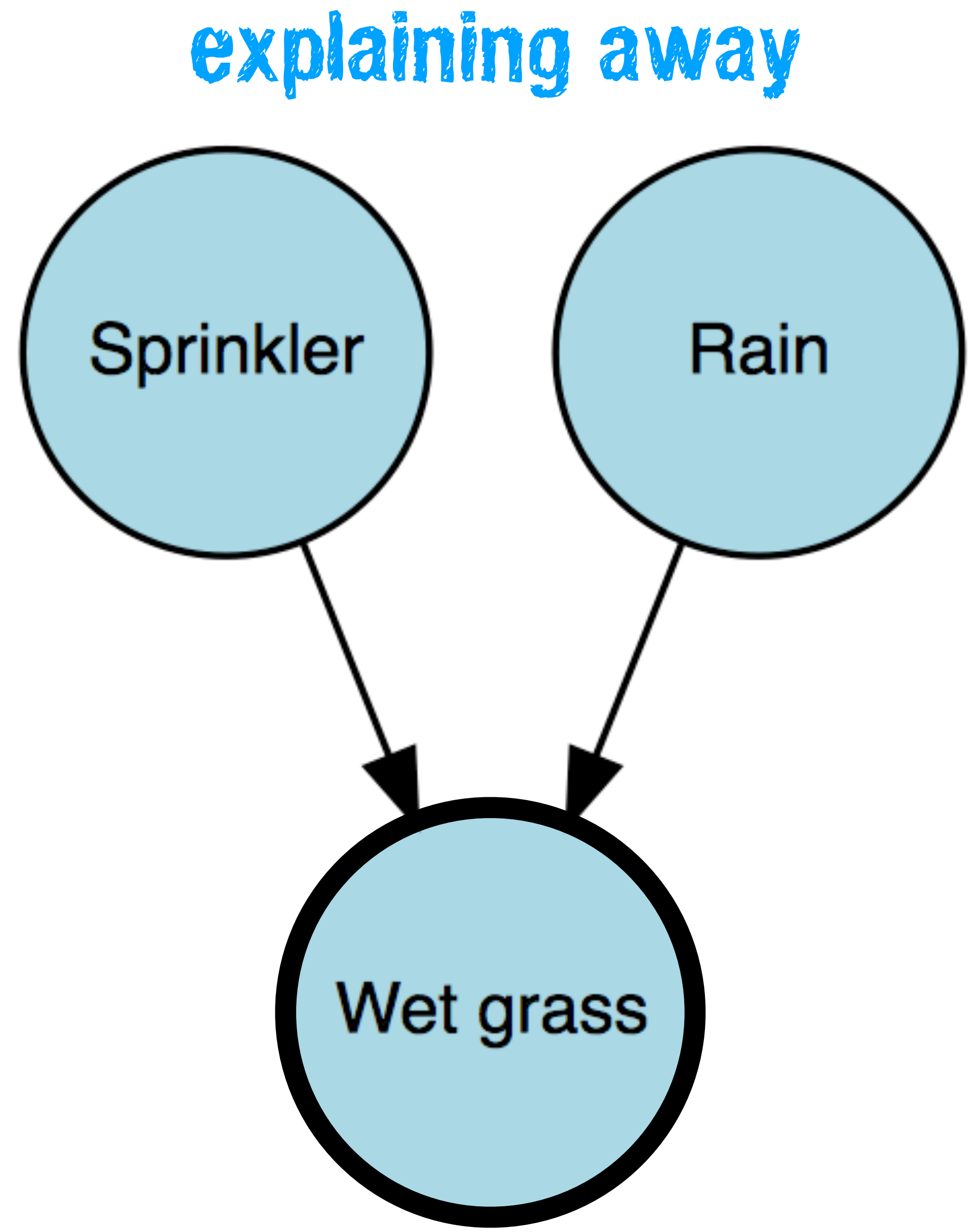


Patterns of inference: **Common effect**

- two causes of a common effect are *conditionally dependent given the effect*

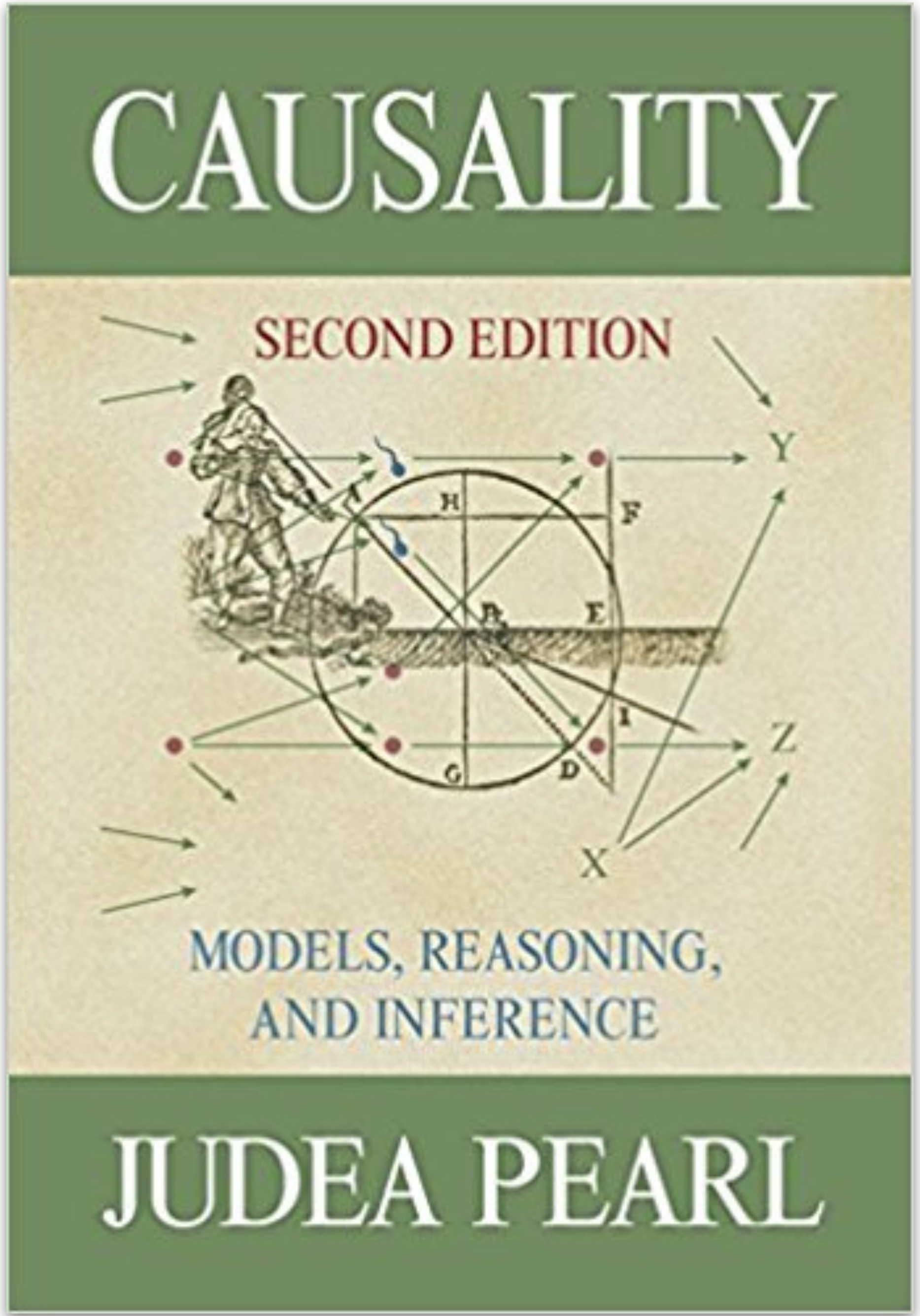
$$p(S \mid R, W) \neq p(S \mid W)$$

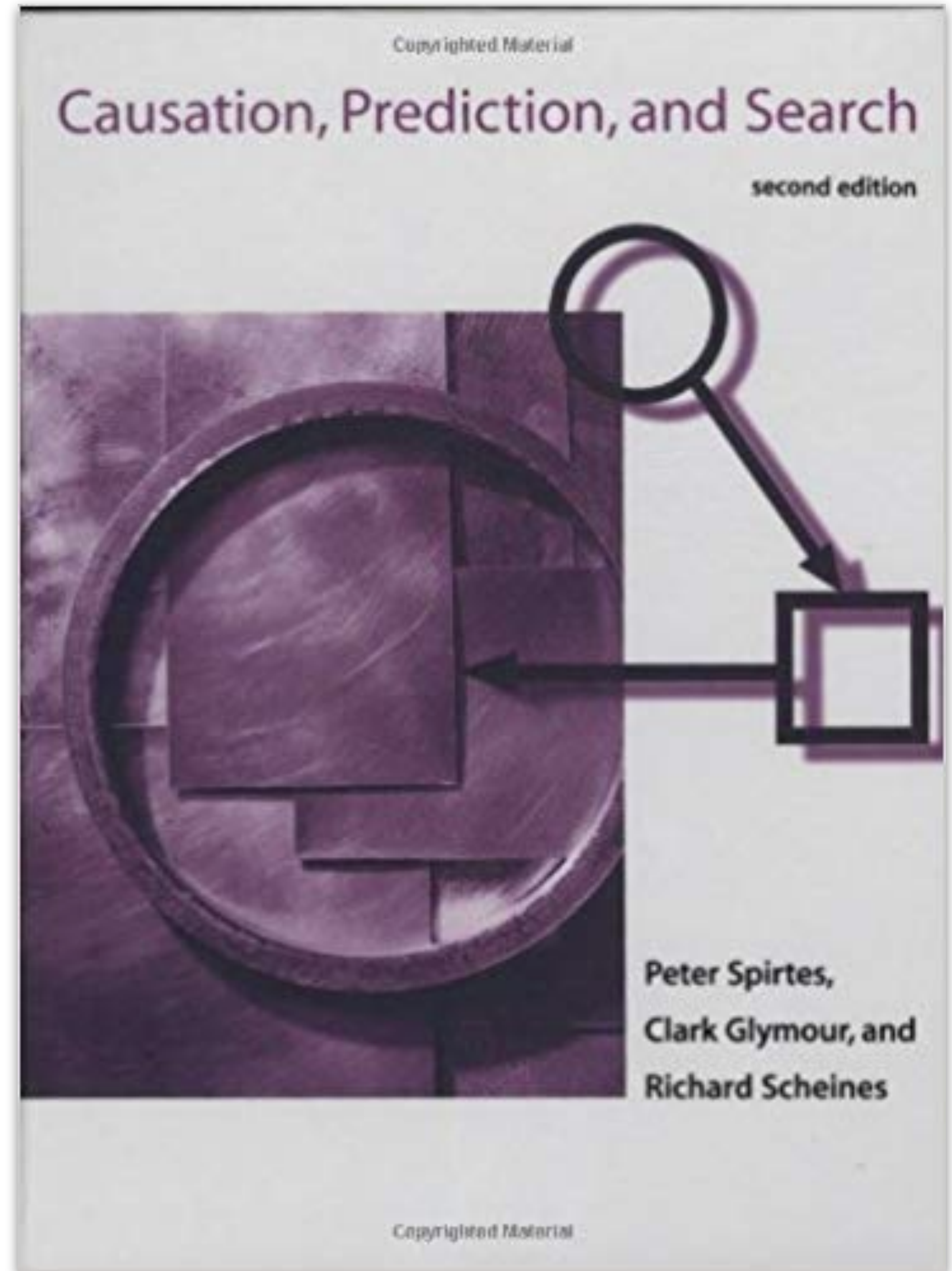
- intuitively: both causes compete to explain the effect



Note: The pattern of inference depends on the structural form which captures how `Sprinkler` and `Rain` jointly affect `Wet grass`. Explaining away holds for the commonly used noisy-or integration function.

Causal Bayesian Networks



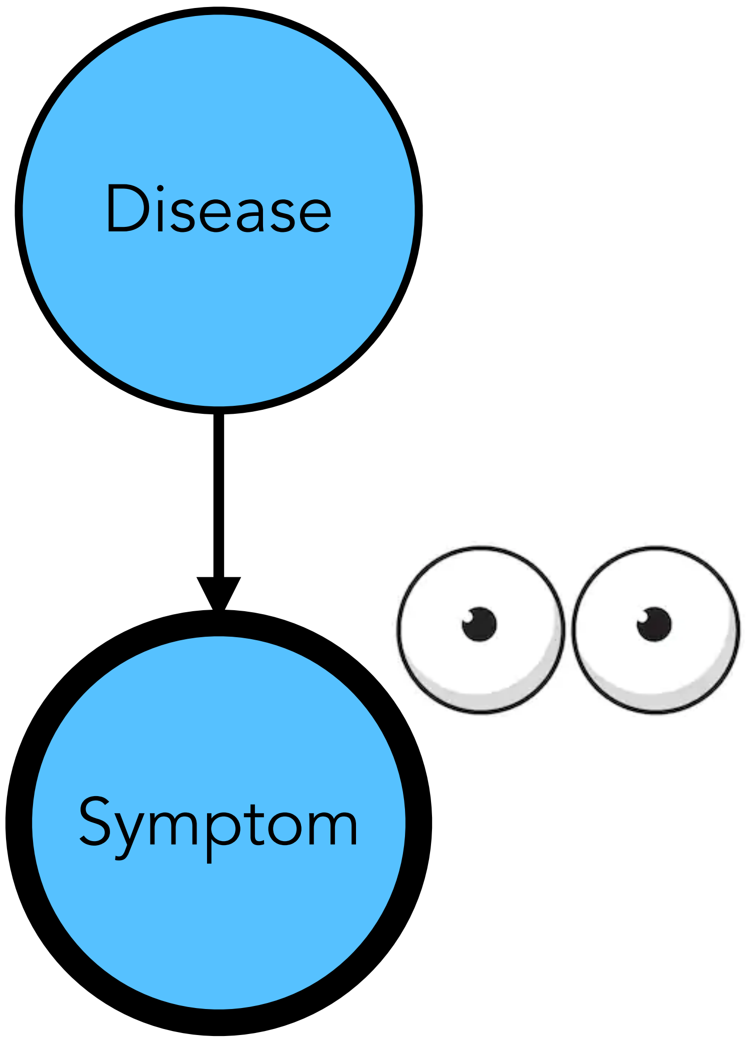


Pearl, J. (2000). Causality: Models, reasoning and inference. Cambridge, England: Cambridge University Press.

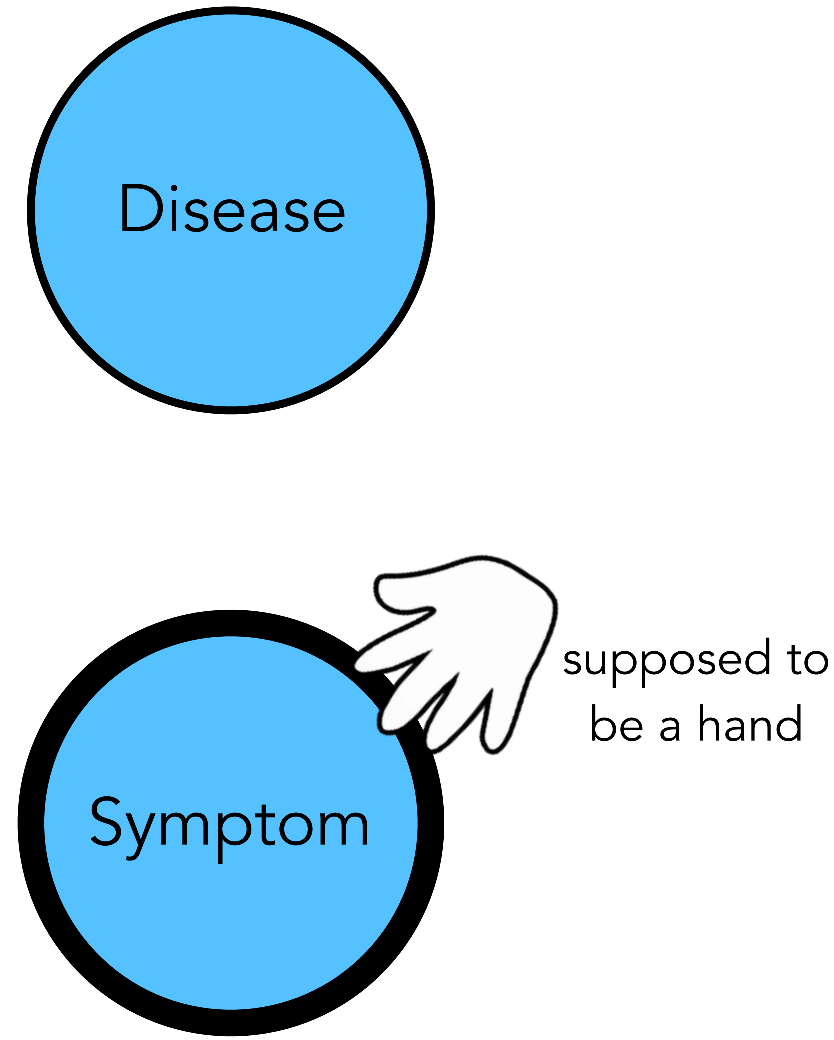
Spirtes, P., Glymour, C. N., & Scheines, R. (2000). Causation, prediction, and search. The MIT Press.

Observation vs. Intervention

seeing	doing

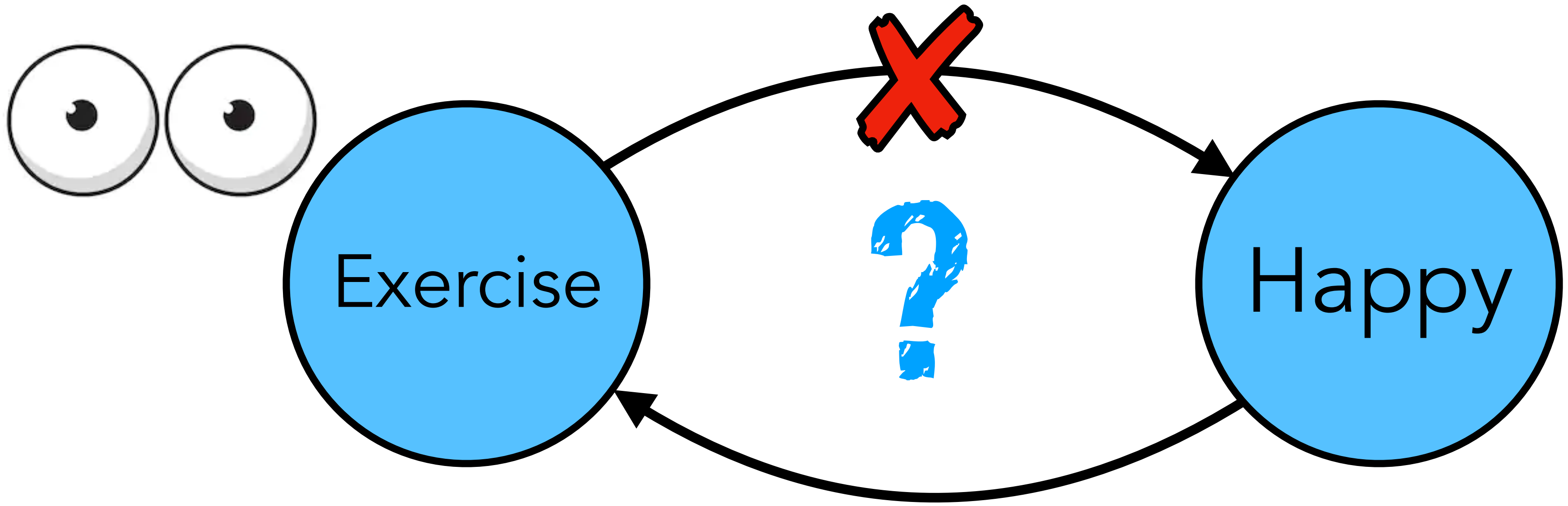


$$p(D \mid S) > p(D)$$

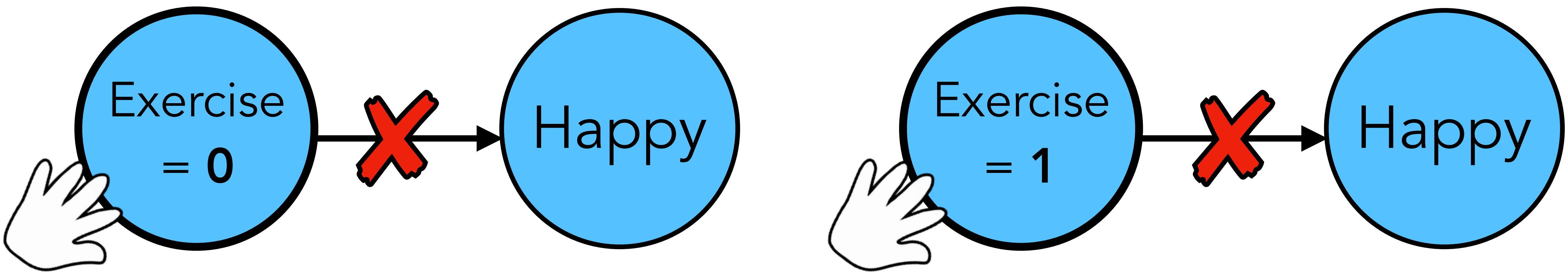


$$p(D \mid \mathrm{do}(S)) = p(D)$$

Inferring causal structure through intervention

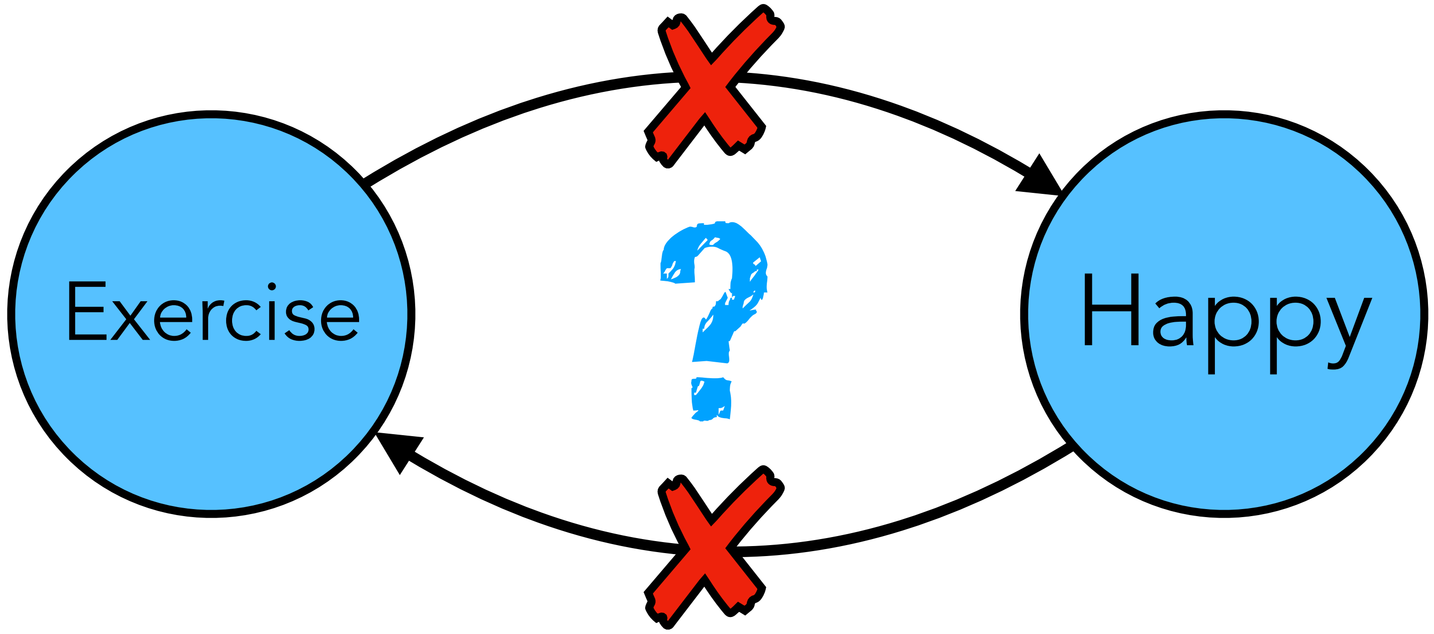


Experiment 1

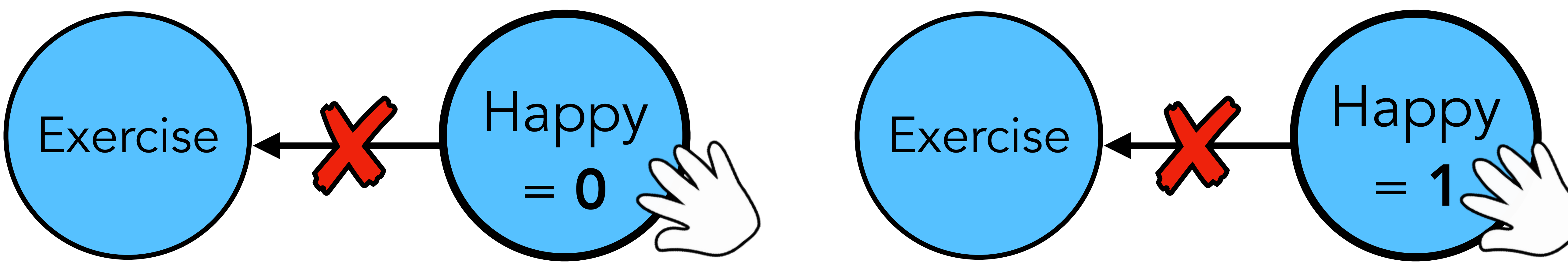


$p(\text{Happy} \mid \text{do}(\text{Exercise} = 0)) = 0.3$ $p(\text{Happy} \mid \text{do}(\text{Exercise} = 1)) = 0.3$

Inferring causal structure through intervention

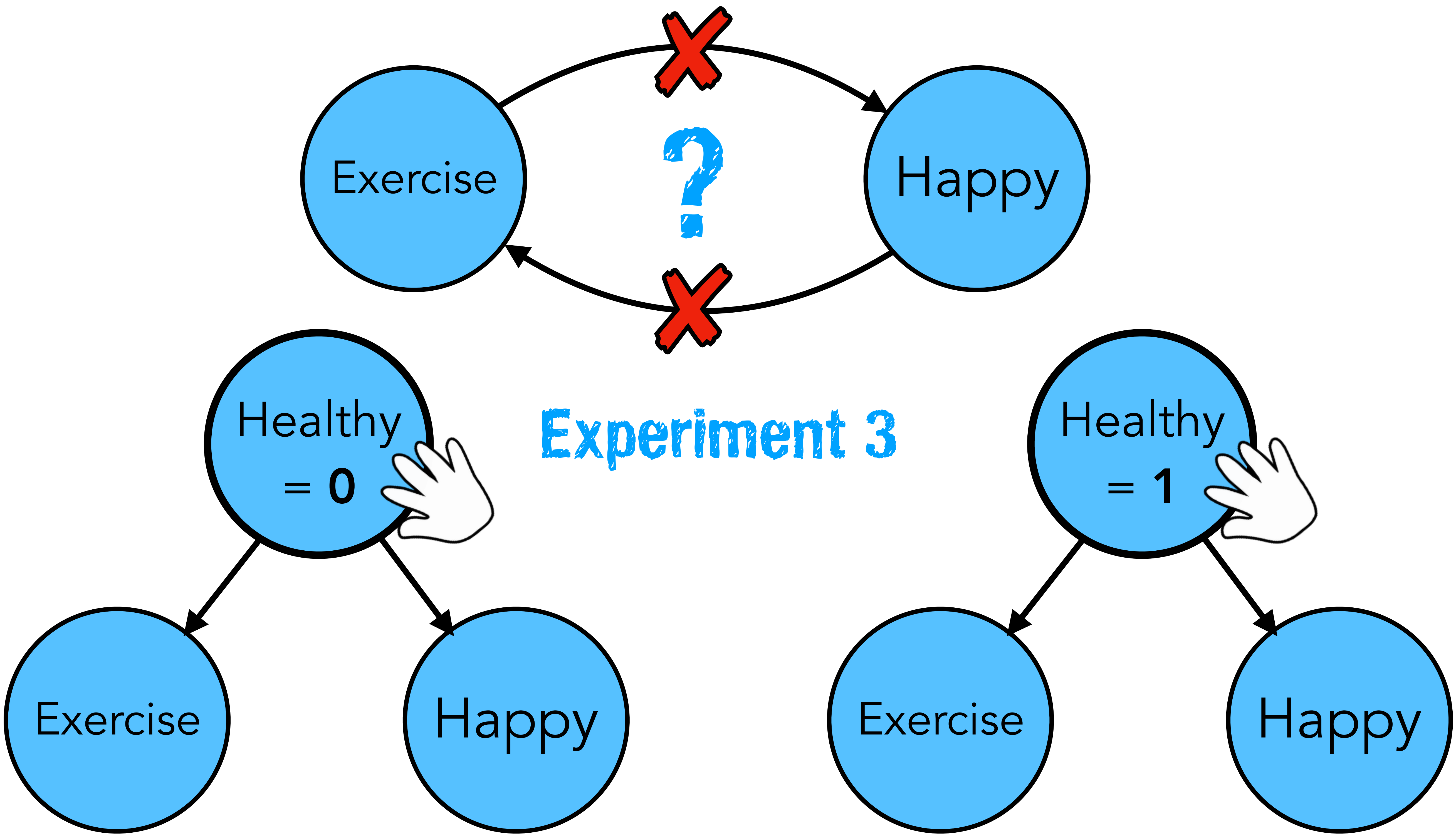


Experiment 2



$p(\text{Exercise}\,|\,\text{do}(\text{Happy} = 0)) = 0.1$ $p(\text{Exercise}\,|\,\text{do}(\text{Happy} = 1)) = 0.1$

Inferring causal structure through intervention



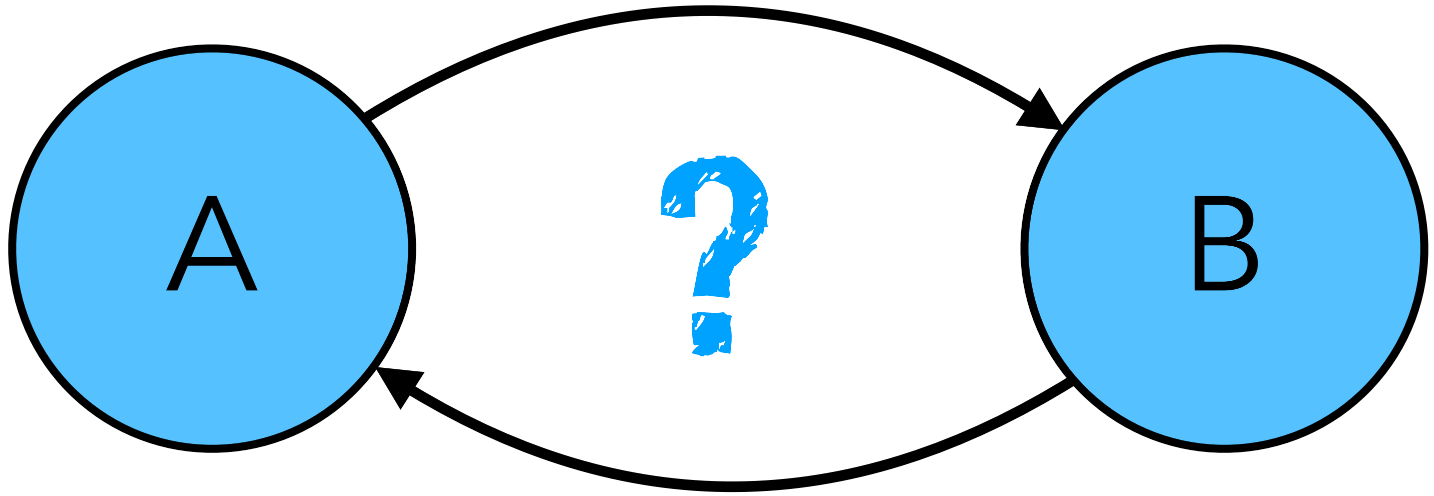
$p(\text{Happy} \mid \text{do}(\text{Healthy} = 0)) = 0.05$

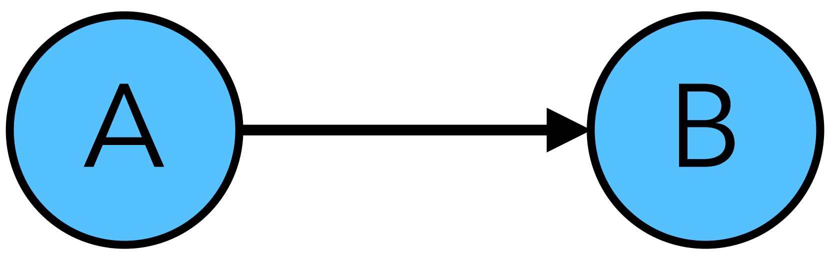
$p(\text{Exercise} \mid \text{do}(\text{Healthy} = 0)) = 0.1$

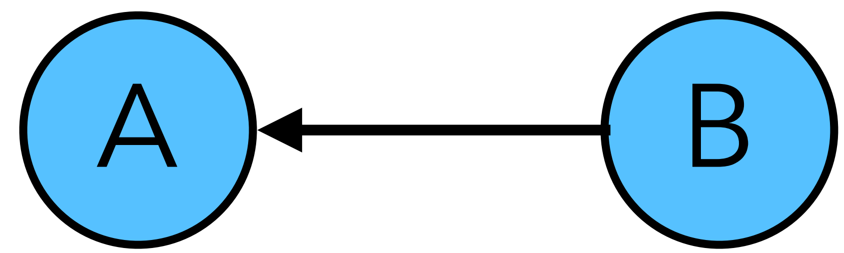
$p(\text{Happy} \mid \text{do}(\text{Healthy} = 0)) = 0.5$

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

Inferring causal structure through intervention

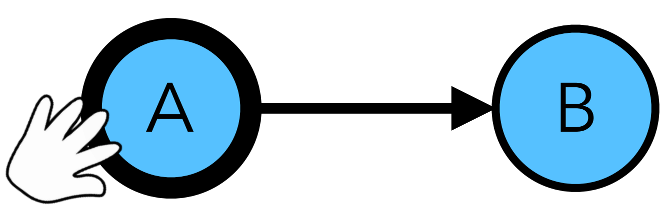






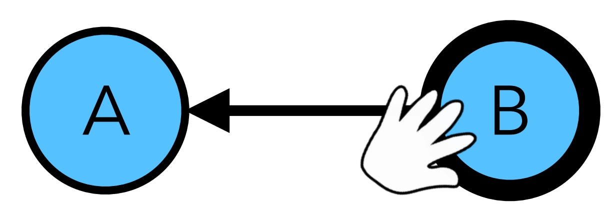
$$p(B \,|\, \mathrm{do}(A)) = p(B \,|\, A)$$

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



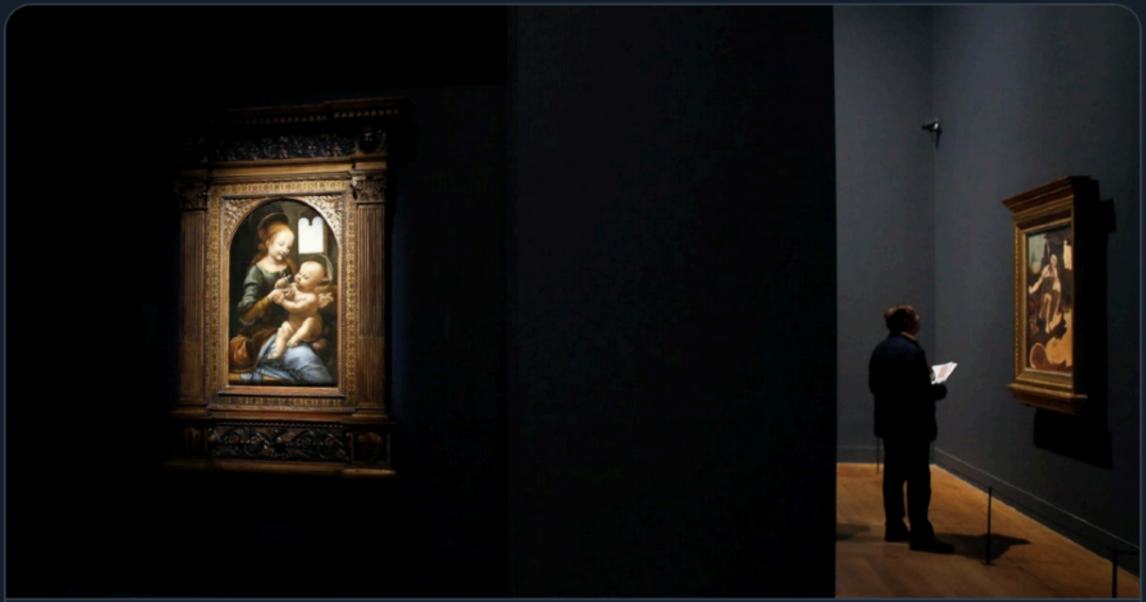
$$p(A \,|\, \mathrm{do}(B)) = p(A)$$

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



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.



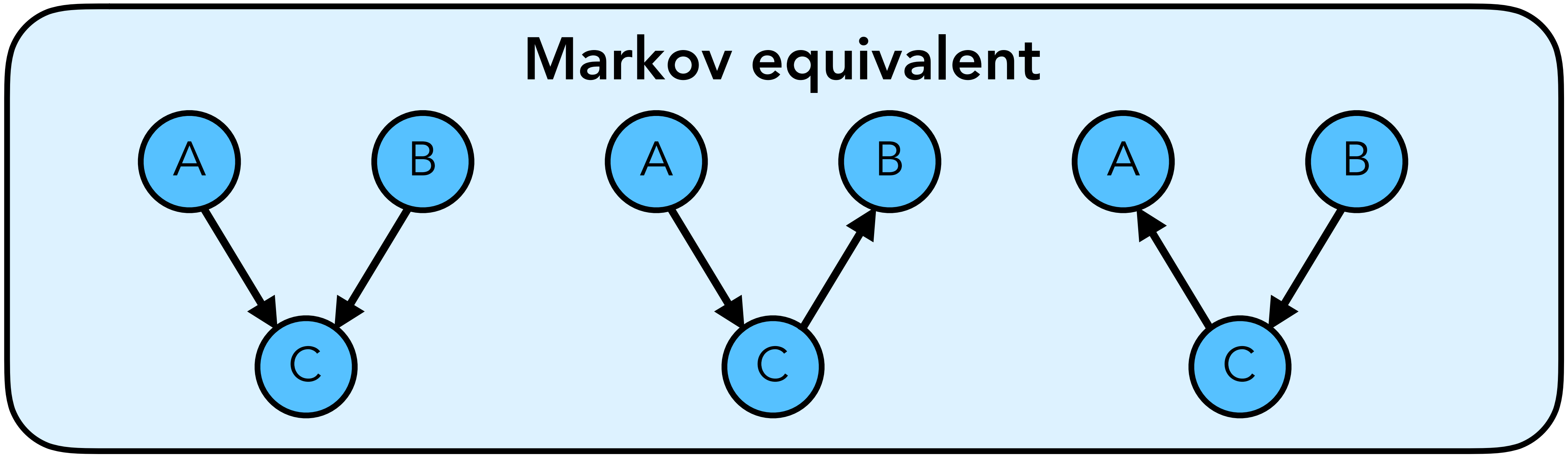
Another Benefit to Going to Museums? You May Live Longer
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

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

Summary

- Introduction to probability / Recap
  - Counting possibilities
  - Interpretation of probability
  - Clue guide to probability
- Bayesian Networks
  - representation
  - inference
  - (un-)conditional (in-)dependence
- Causal Bayes nets

Have a nice Martin Luther King, Jr. day!

The time is always right to do what is right.

- Martin Luther King, Jr.