

Notebook: *Where are we?*

Basic probability

If all outcomes are equally likely – for example, rolling a fair die or flipping a fair coin – then it's easy to compute a probability by counting:

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

How likely are you to get an even number when rolling a die?

$$Even = \{\text{1 die face}, \text{2 die faces}, \text{3 die faces}\}$$

$$All = \{\text{1 die face}, \text{2 die faces}, \text{3 die faces}, \text{4 die faces}, \text{5 die faces}, \text{6 die faces}\}$$

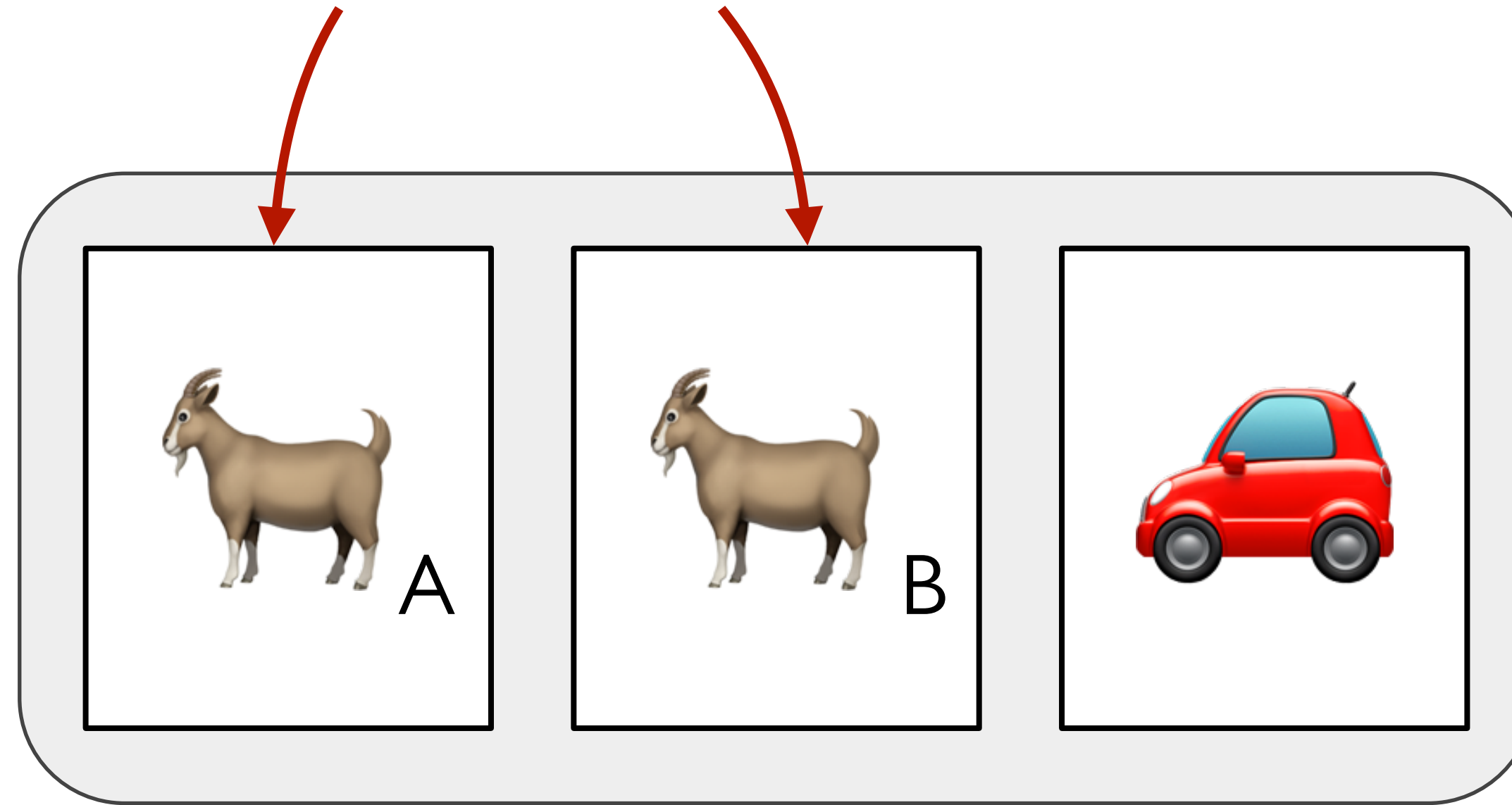
$$|Even| / |All| = 3 / 6 = 50\%$$

*In math, {...} denotes a set,
and |...| denotes the number
of elements.*

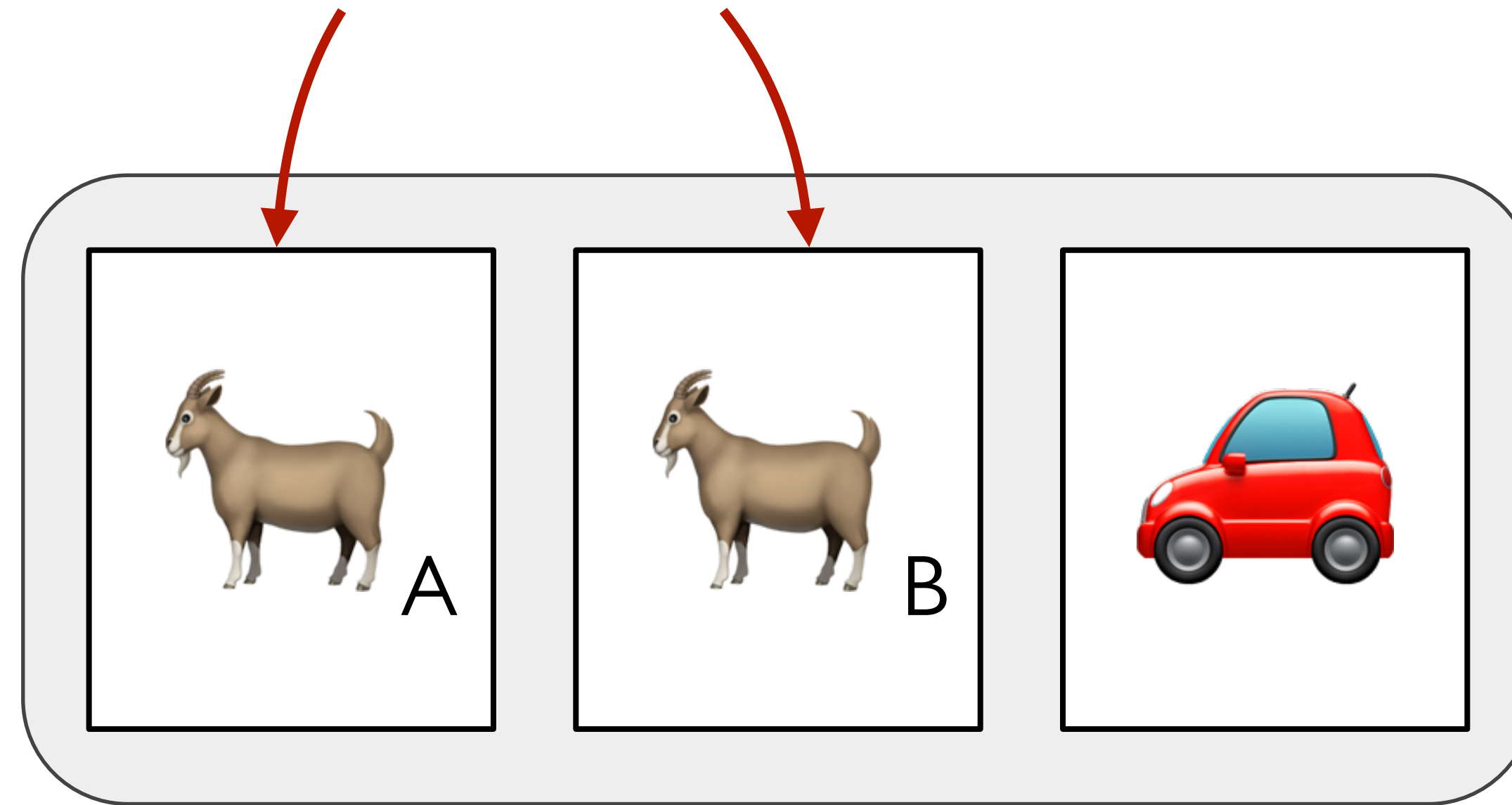
When there are two or more ways an event can happen, the *addition rule* says its probability is the sum of the probabilities for the different ways:

$$\begin{aligned} P(A) = & P(\text{first way } A \text{ can happen}) + \\ & P(\text{second way } A \text{ can happen}) + \\ & \dots \end{aligned}$$

*There are two distinct
goats, Alice (A) and Bob (B)*

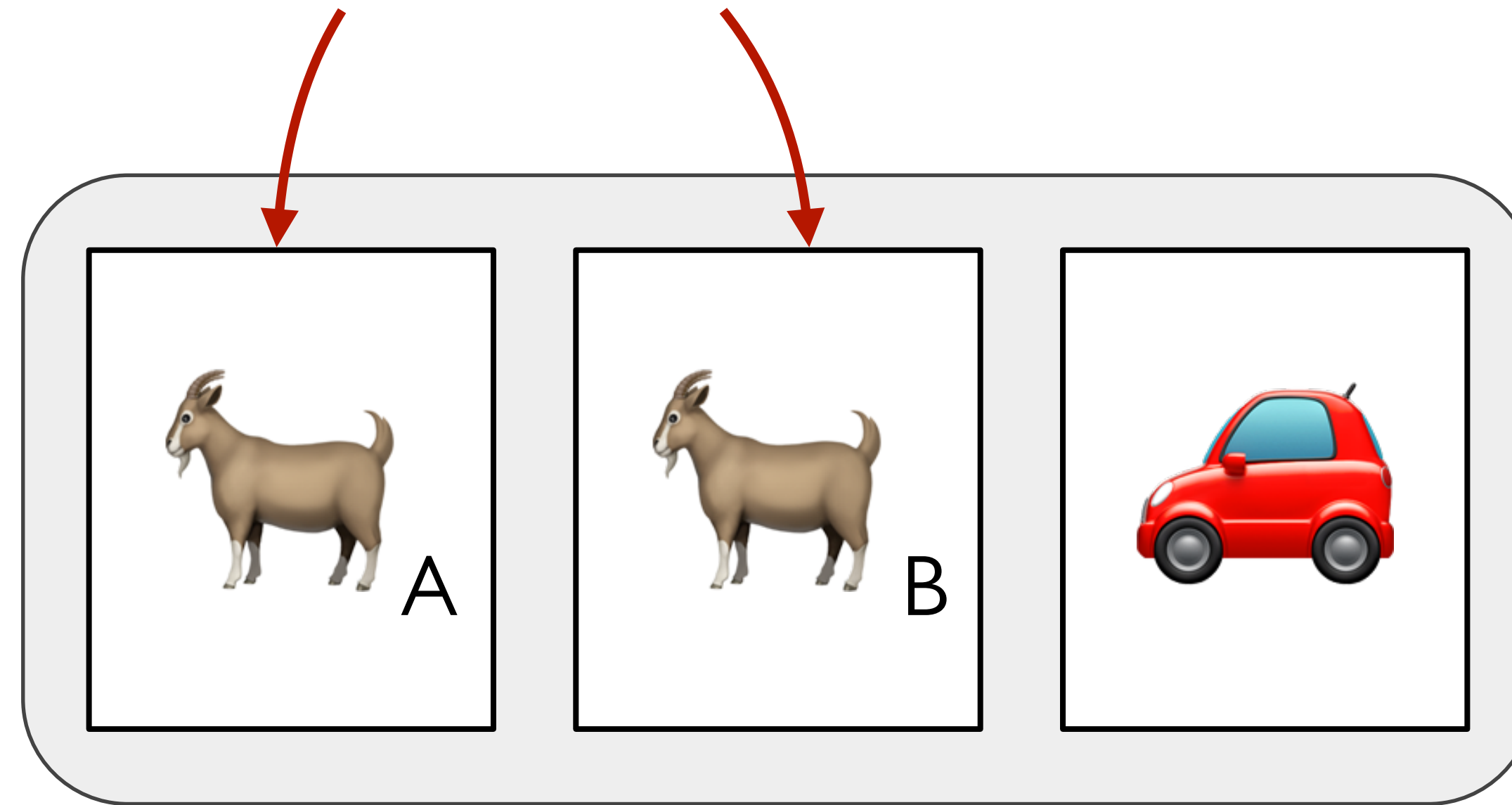


*There are two distinct
goats, Alice (A) and Bob (B)*



$$P(\text{goat}) = P(\text{goat}_A) + P(\text{goat}_B) = 2/3$$

*There are two distinct
goats, Alice (A) and Bob (B)*



$$P(\text{goat}) = P(\text{goat}_A) + P(\text{goat}_B) = 2/3$$

$$P(\text{car}) = 1/3$$

For another example, the probability of rolling a total of six when rolling two dice is:

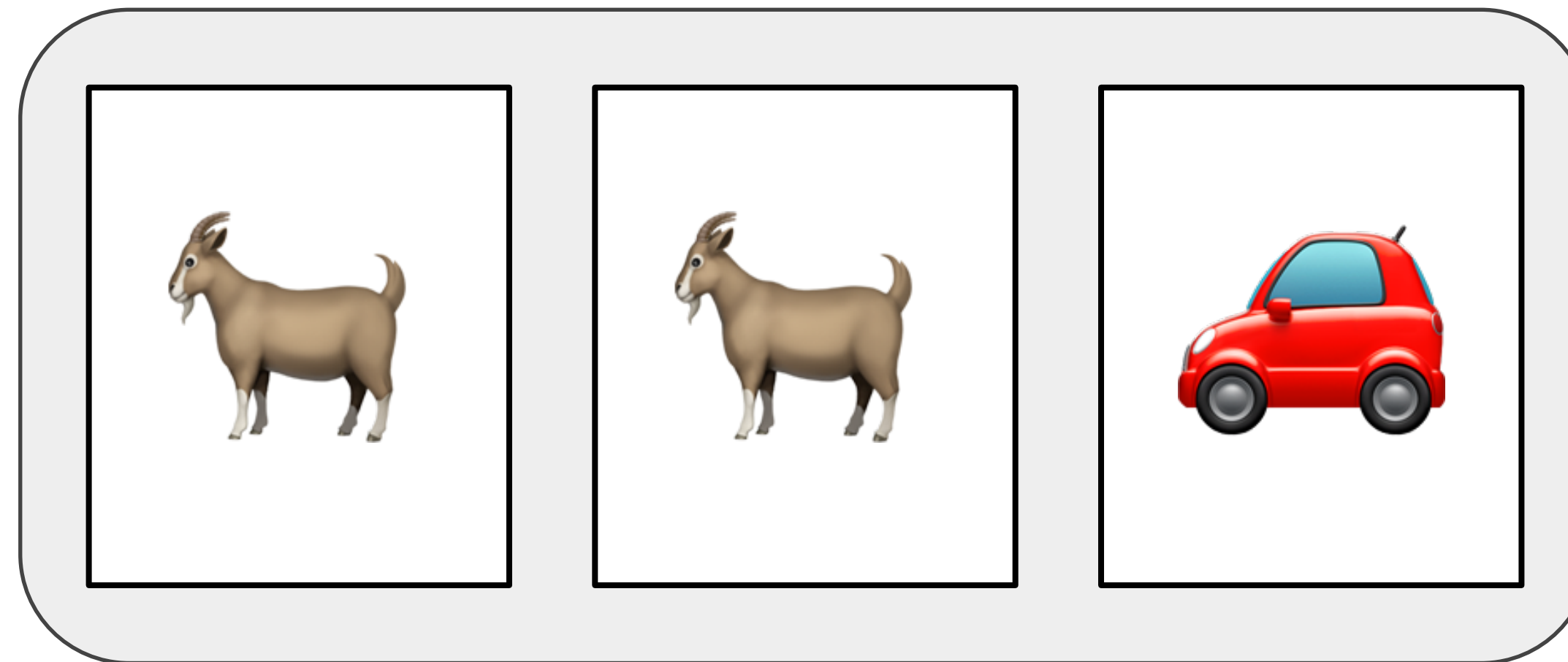
$$\begin{aligned} P(\text{roll six}) = & P(\text{⬢⬢}, \text{⬢}) + \\ & P(\text{⬢}, \text{⬢⬢}) + \\ & P(\text{⬢⬢}, \text{⬢}) + \\ & P(\text{⬢}, \text{⬢⬢}) + \\ & P(\text{⬢⬢}, \text{⬢⬢}) \end{aligned}$$

The chance that event A will *not* occur:

$$P(\text{not } A) = 1 - P(A)$$

So, if $P(\text{☁️🌧️}) = 0.7$ (that is, a 70% chance of rain)

then $P(\text{not } \text{☁️🌧️}) = 1 - 0.7$
 $= 0.3$ (that is, 30%)

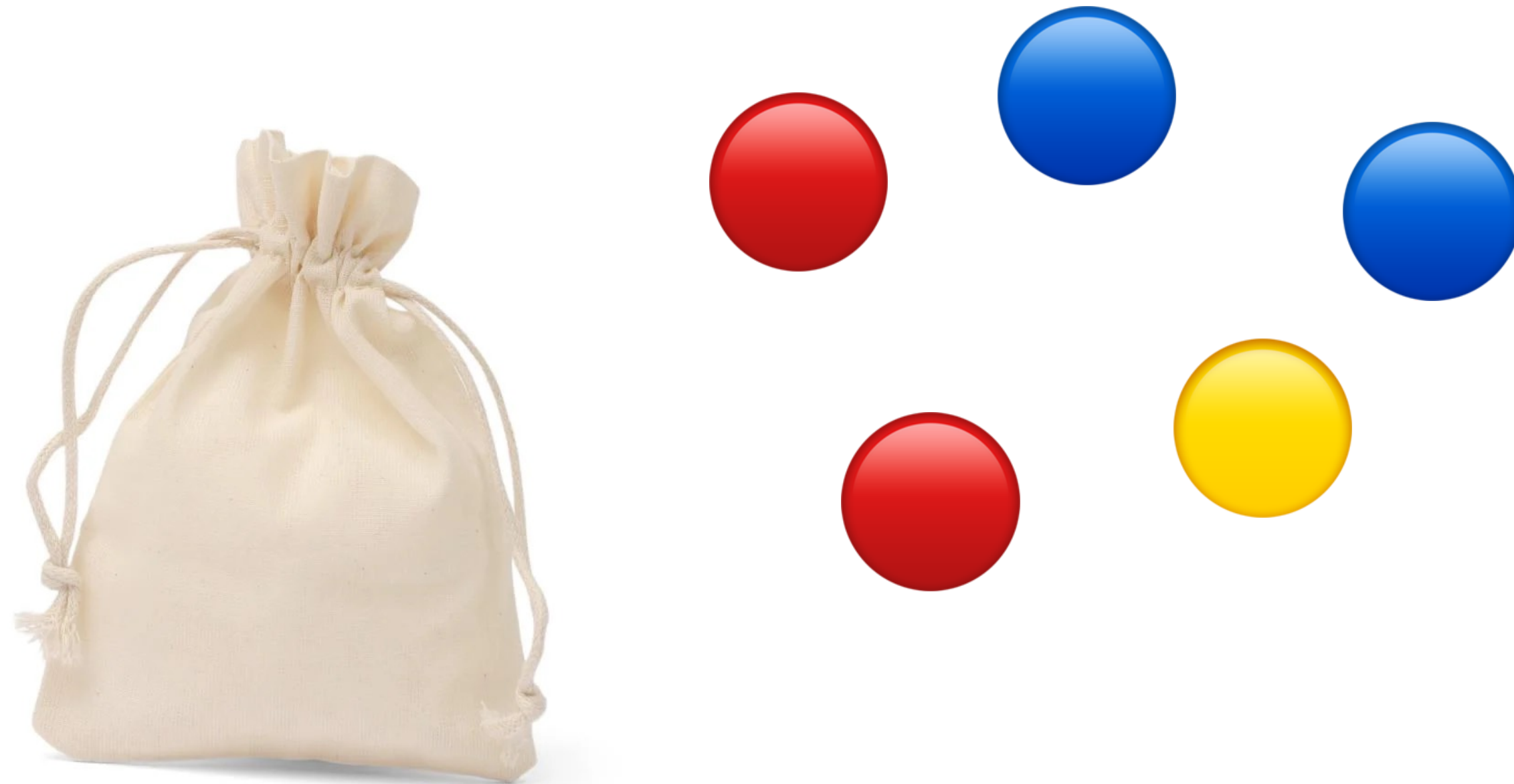


$$\begin{aligned} P(\text{not } \text{goat}) &= 1 - P(\text{goat}) \\ &= 1 - 2/3 \\ &= 1/3 \end{aligned}$$

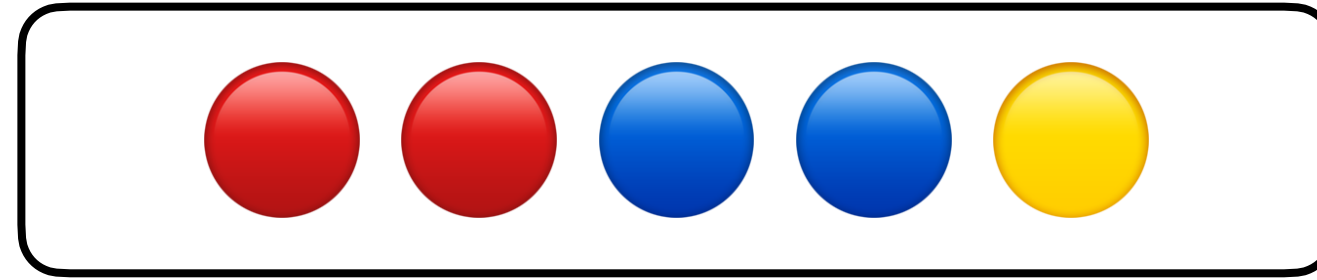
When you want to know the probability of events A and B *both* happening, you use the *multiplication rule*:

$$P(A \text{ and } B) = P(A) \cdot P(B \text{ given that } A \text{ happened})$$

For example, if you are taking colored marbles out of a bag, what is the probability of drawing a yellow marble *then* a blue marble?

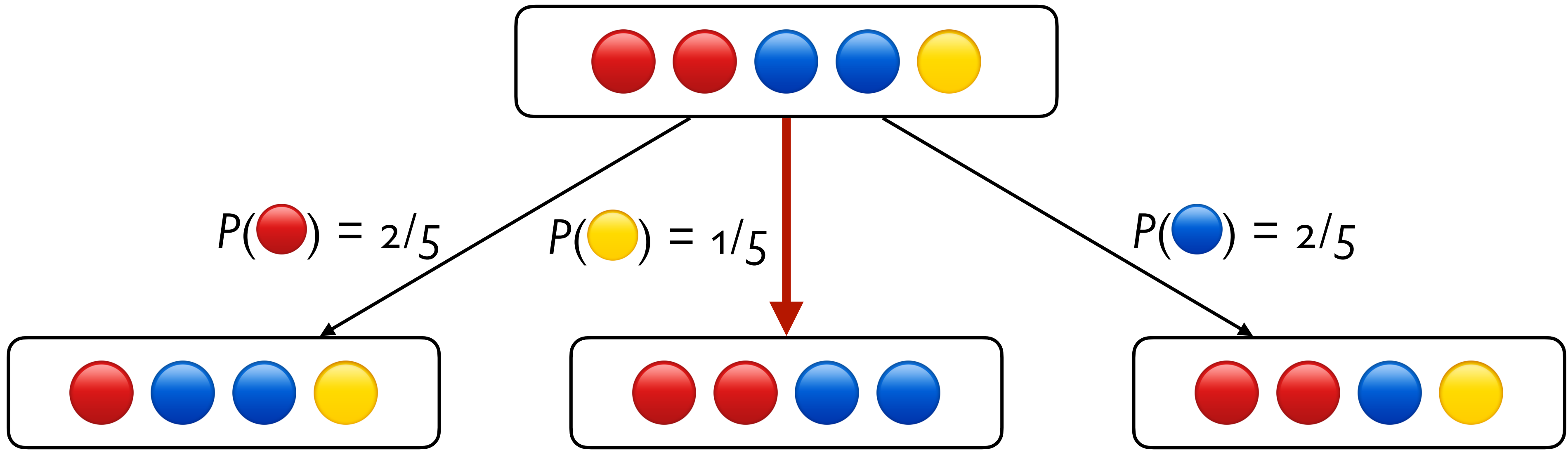


$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = ?$



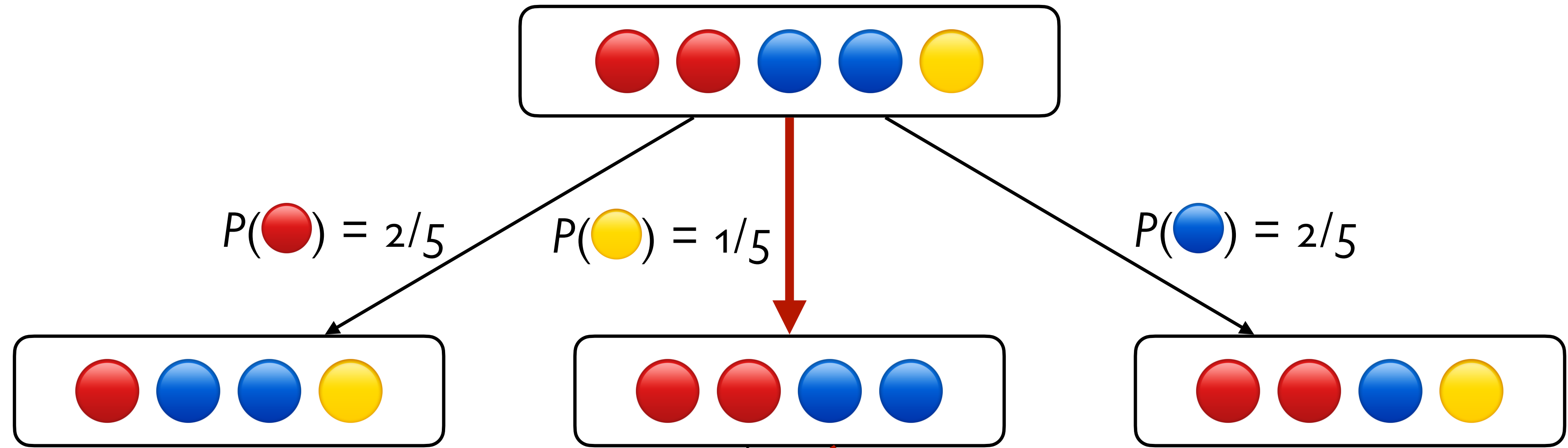
$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = ?$

First draw

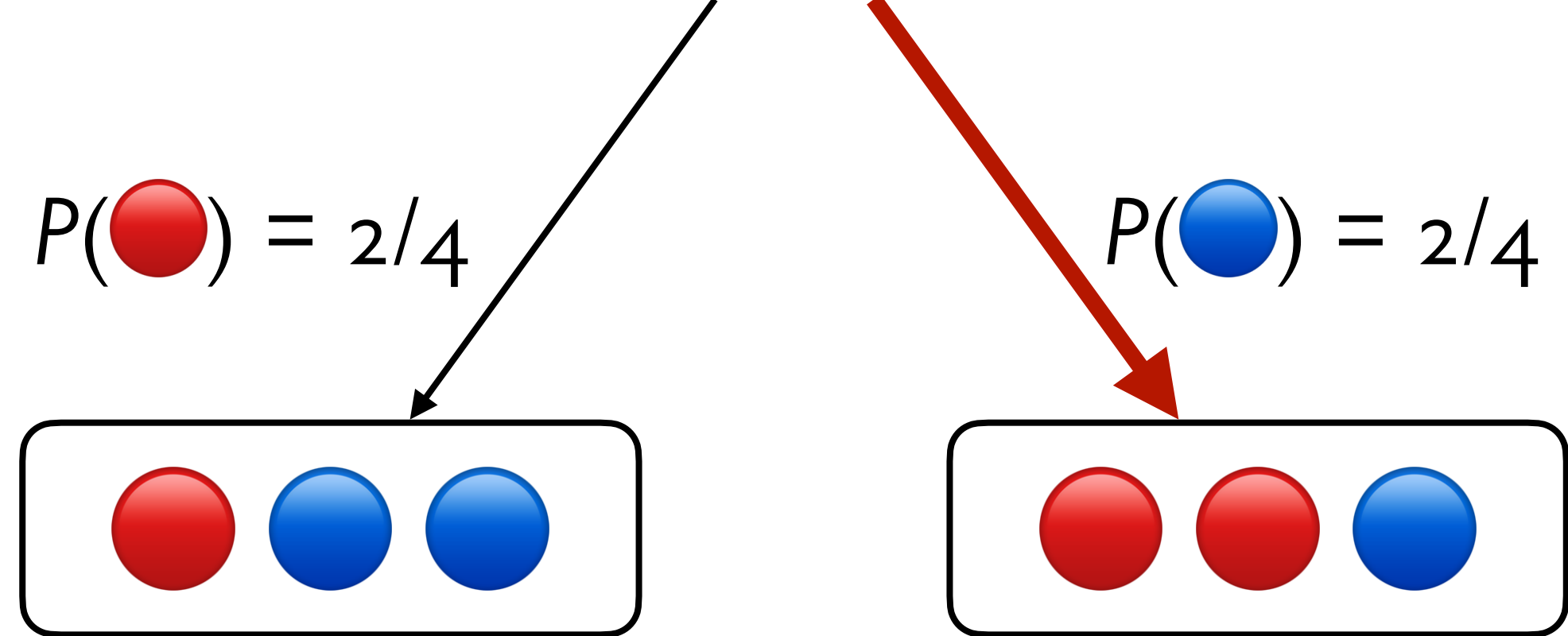


$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = ?$

First draw

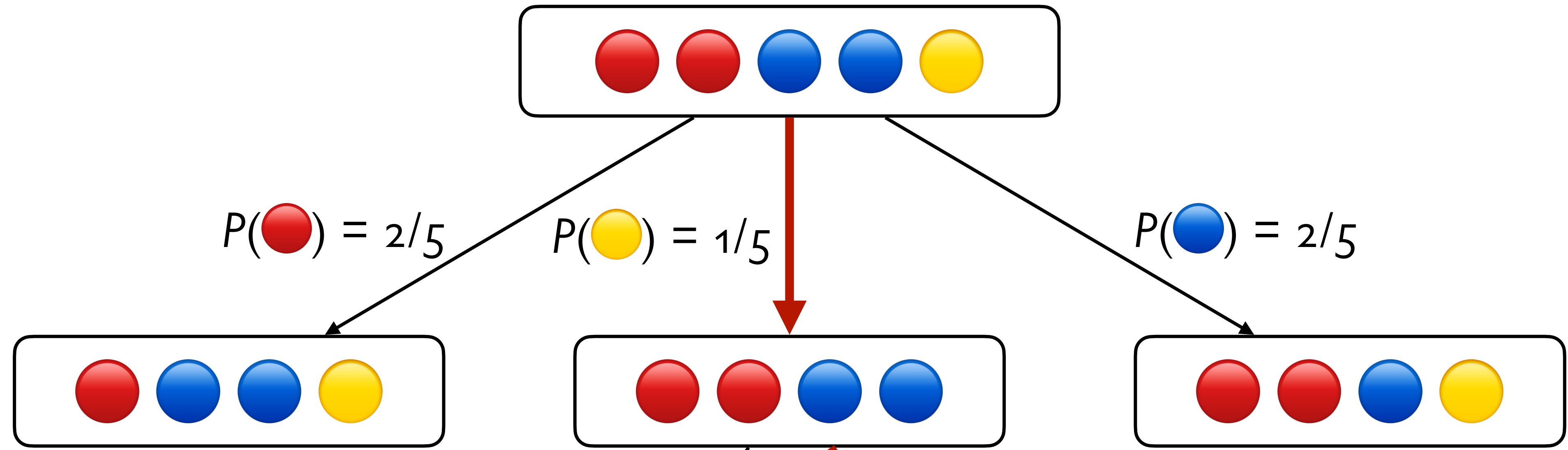


Second draw

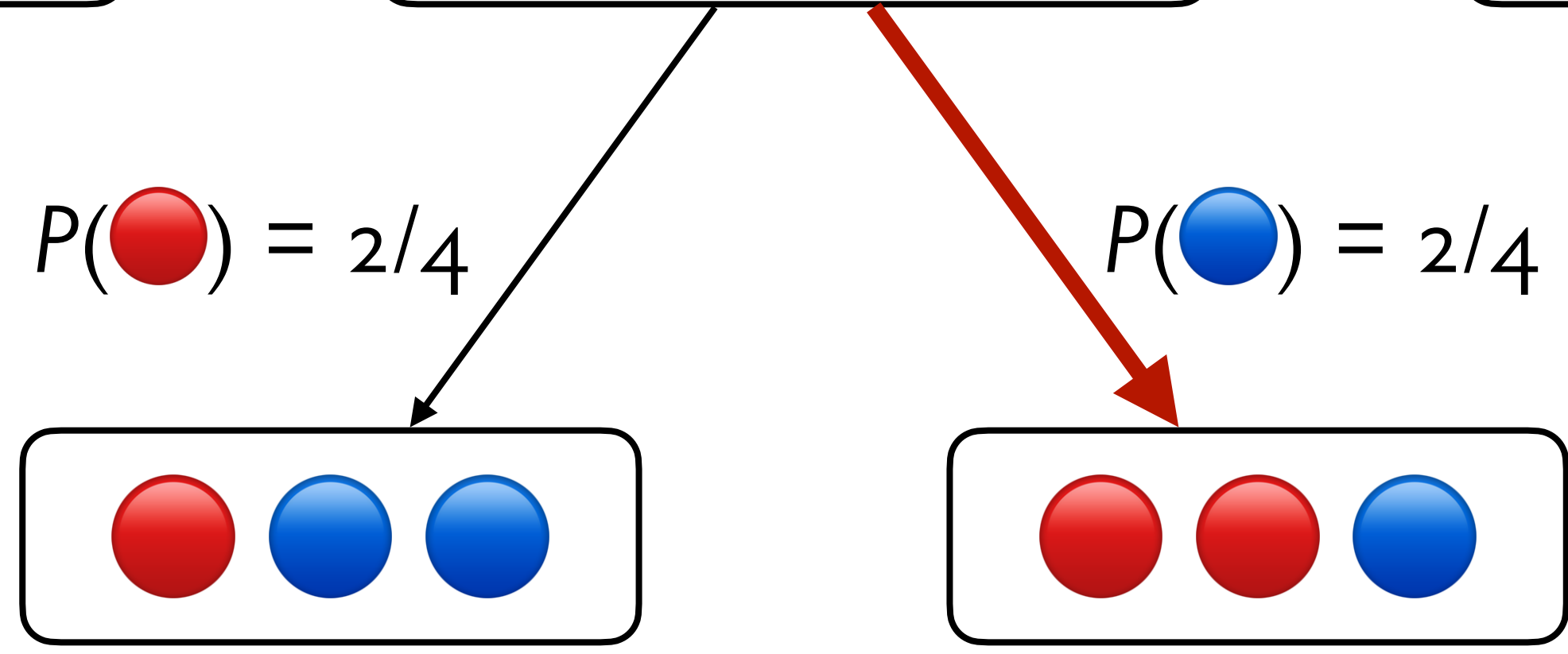


$P(\text{draw a } \text{yellow} \text{ then a } \text{blue}) = \boxed{1/5 \cdot 2/4 = 10\%}$

First draw



Second draw



Practice

Probability the sum of two dice is 12?

Practice

Probability the sum of two dice is 12?

$$P(\text{sum is 12}) = P(\text{roll } \text{⚡}) \cdot P(\text{roll } \text{⚡})$$

Practice

Probability the sum of two dice is 12?

$$\begin{aligned} P(\text{sum is } 12) &= P(\text{roll } \text{⚰}) \cdot P(\text{roll } \text{⚰}) \\ &= 1/6 \cdot 1/6 \end{aligned}$$

Practice

Probability the sum of two dice is 12?

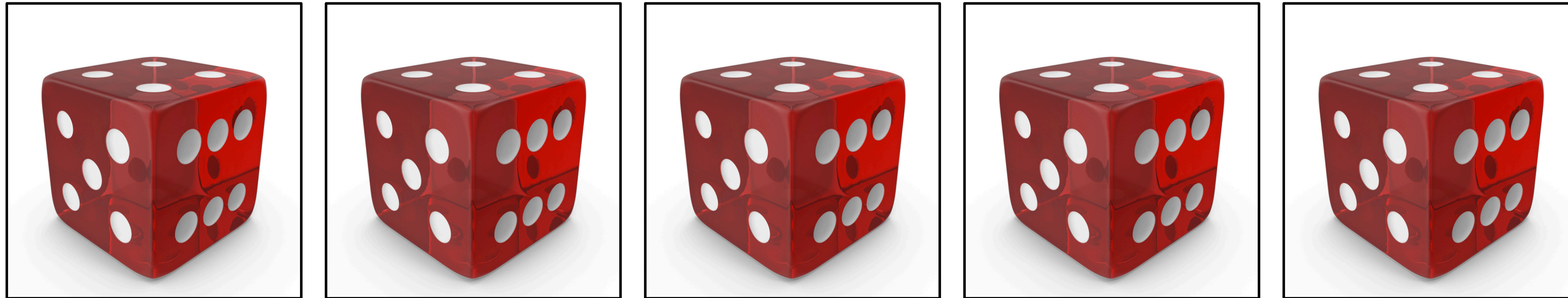
$$\begin{aligned}P(\text{sum is } 12) &= P(\text{roll } \text{⚰}) \cdot P(\text{roll } \text{⚰}) \\&= 1/6 \cdot 1/6 \\&= 1/36\end{aligned}$$

Practice



Probability of all 4s in five rolls?

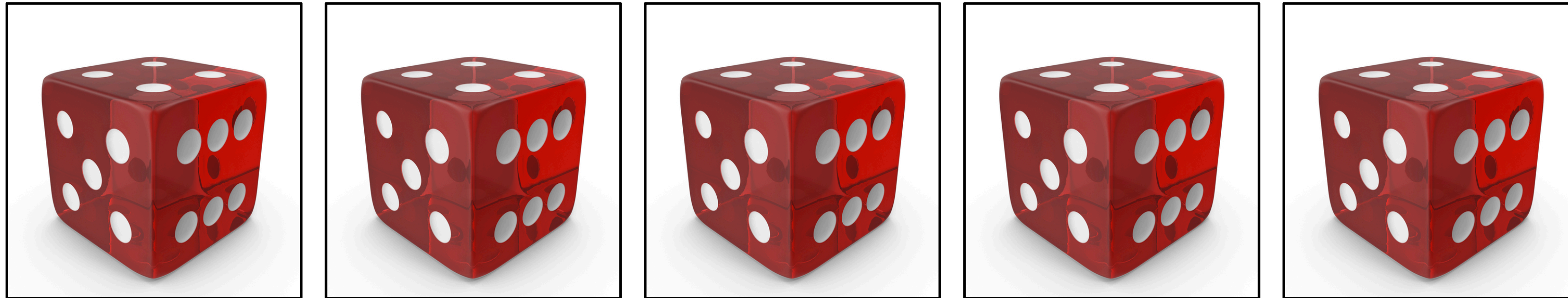
Practice



Probability of all 4s in five rolls?

$$P(\text{five } \text{⚀s}) = P(\text{roll } \text{⚀})^5$$

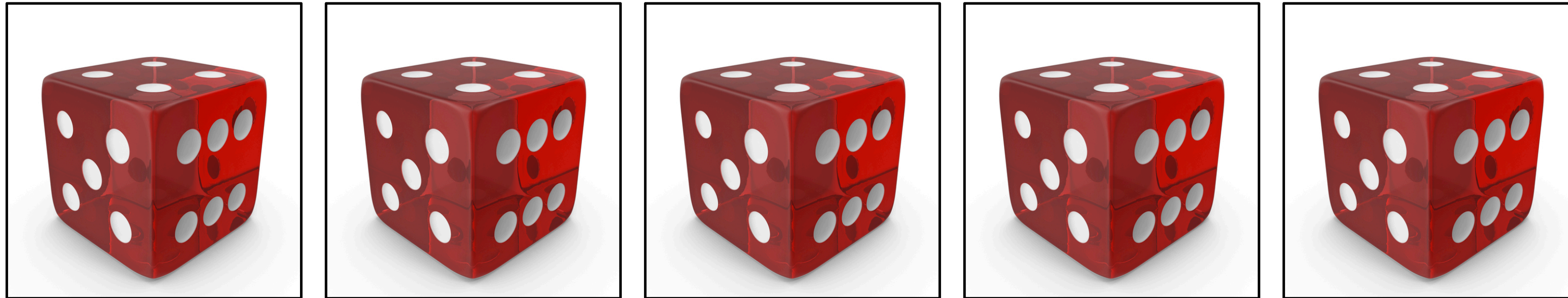
Practice



Probability of all 4s in five rolls?

$$\begin{aligned} P(\text{five } \text{⚡s}) &= P(\text{roll } \text{⚡})^5 \\ &= (1/6)^5 \end{aligned}$$

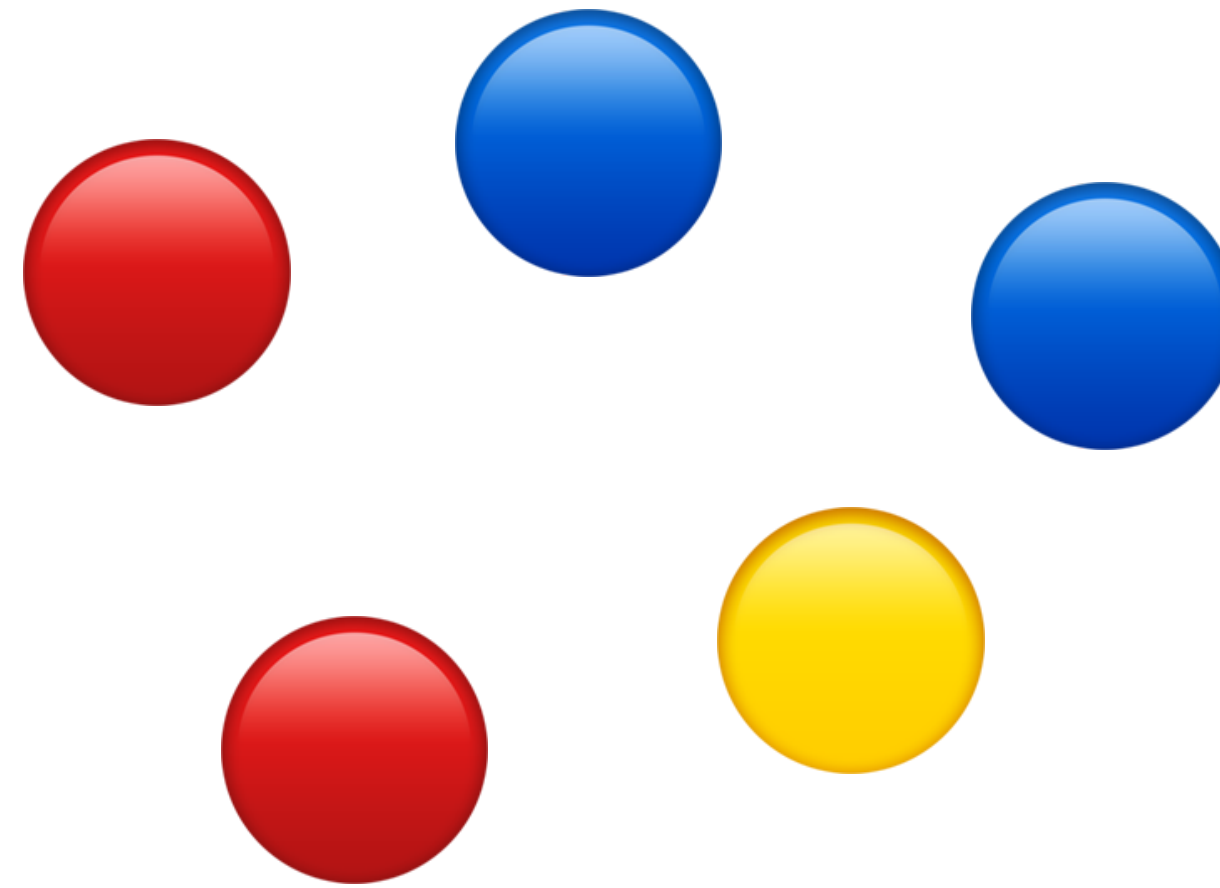
Practice



Probability of all 4s in five rolls?

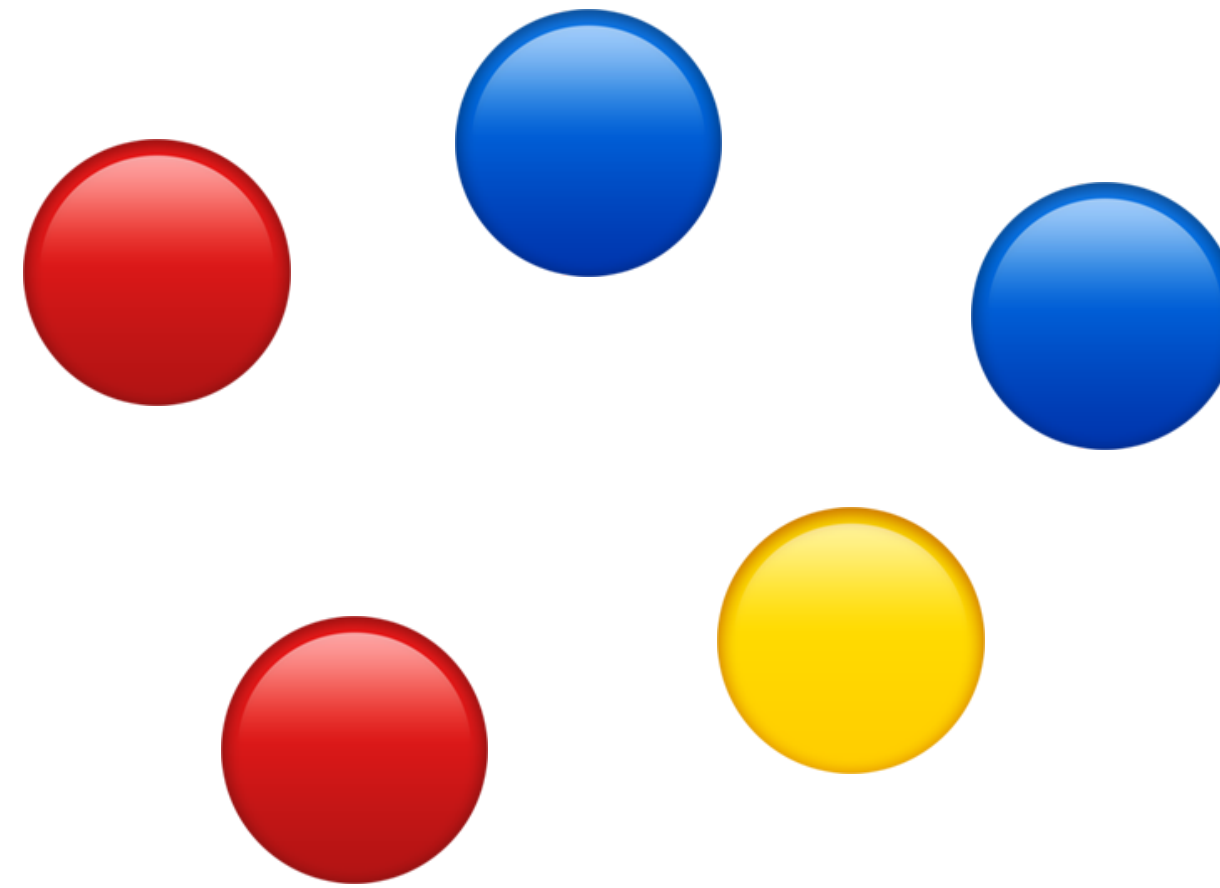
$$\begin{aligned}P(\text{five } \text{⚡}\text{s}) &= P(\text{roll } \text{⚡})^5 \\&= (1/6)^5 \\&= 0.00013\end{aligned}$$

Practice



Probability of drawing two of the same color?

Practice



Probability of drawing two of the same color?

$$\begin{aligned} P(\text{two of same color}) &= P(\text{blue then blue}) + P(\text{red then red}) + P(\text{yellow then yellow}) \\ &= \frac{2}{5} \cdot \frac{1}{4} + \frac{2}{5} \cdot \frac{1}{4} + \frac{1}{5} \cdot 0 \\ &= \frac{2}{20} + \frac{2}{20} + 0 \\ &= \frac{1}{5} \end{aligned}$$

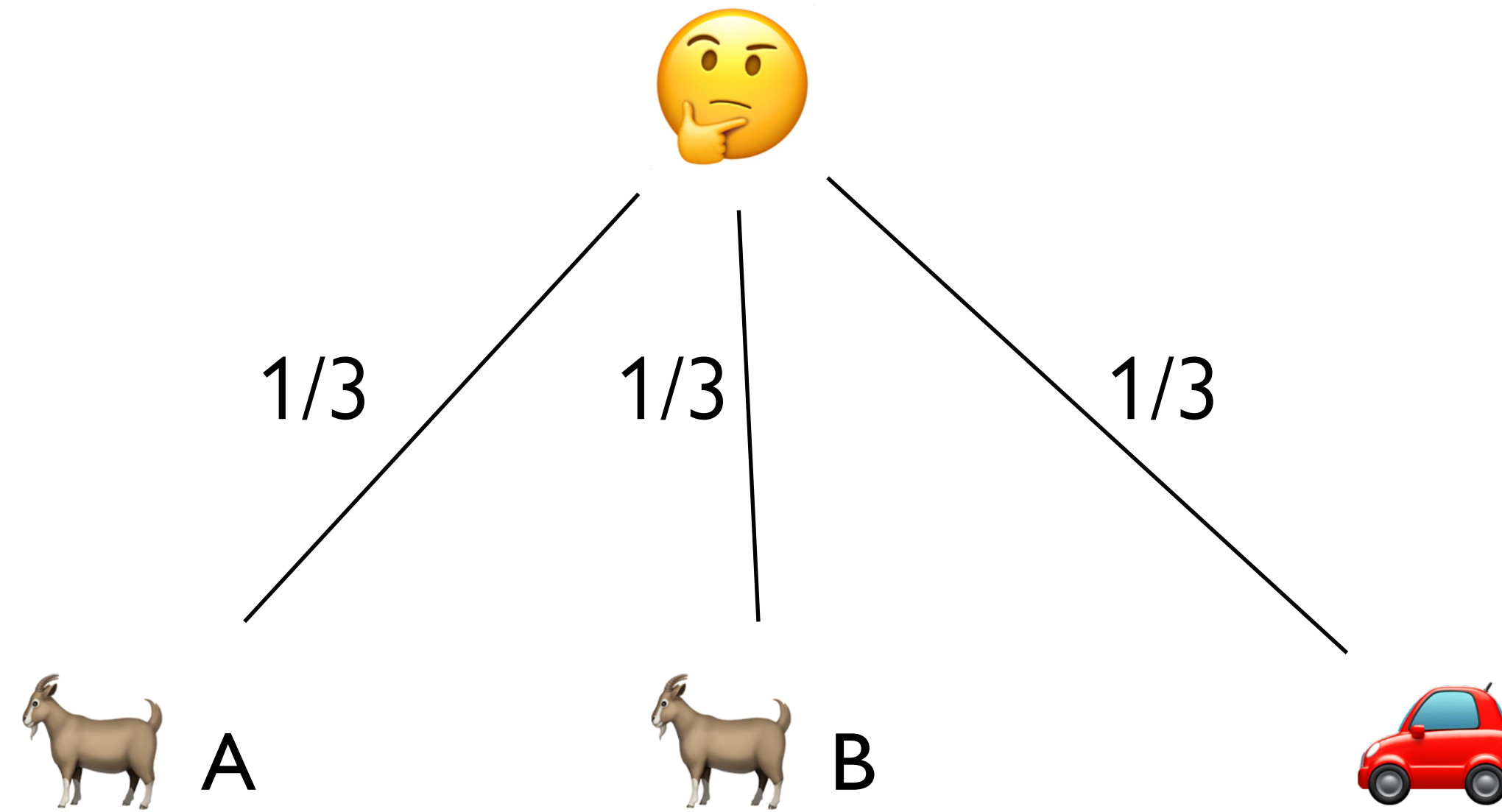
Practice

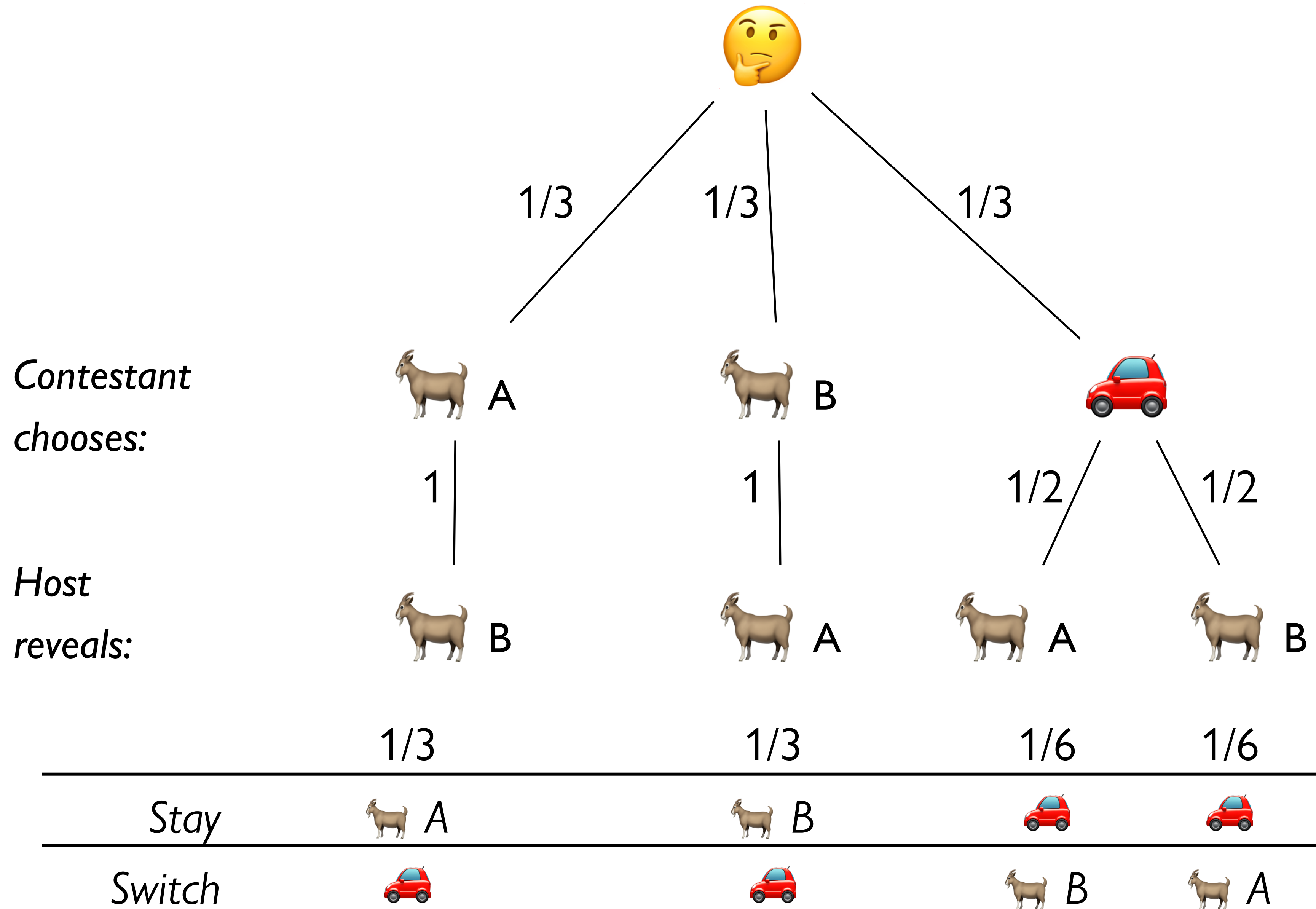
Use the probability rules to answer the Monty Hall Problem:

$$P(\text{stay and win}) = ?$$

$$P(\text{switch and win}) = ?$$

*Contestant
chooses:*





Strategies

Don't switch:

$$P(\text{🚗}) = 1/3 \quad P(\text{🐐}) = 2/3$$

Always switch:

If you picked wrong and switch, you always win.

If you picked right and switch, you always lose

But you were more likely to pick a wrong door initially; switching improves your odds.

$$P(\text{🚗}) = 2/3 \quad P(\text{🐐}) = 1/3$$

Probability and distributions

If you have a random quantity with various possible values, then its *probability (exact) distribution* associates

all the possible values of the quantity
with the probability of each of those values.



Sum of two dice rolls

If you have a random quantity with various possible values, then its *probability (exact) distribution* associates

all the possible values of the quantity
with the probability of each of those values.



Sum of two dice rolls

Random quantity

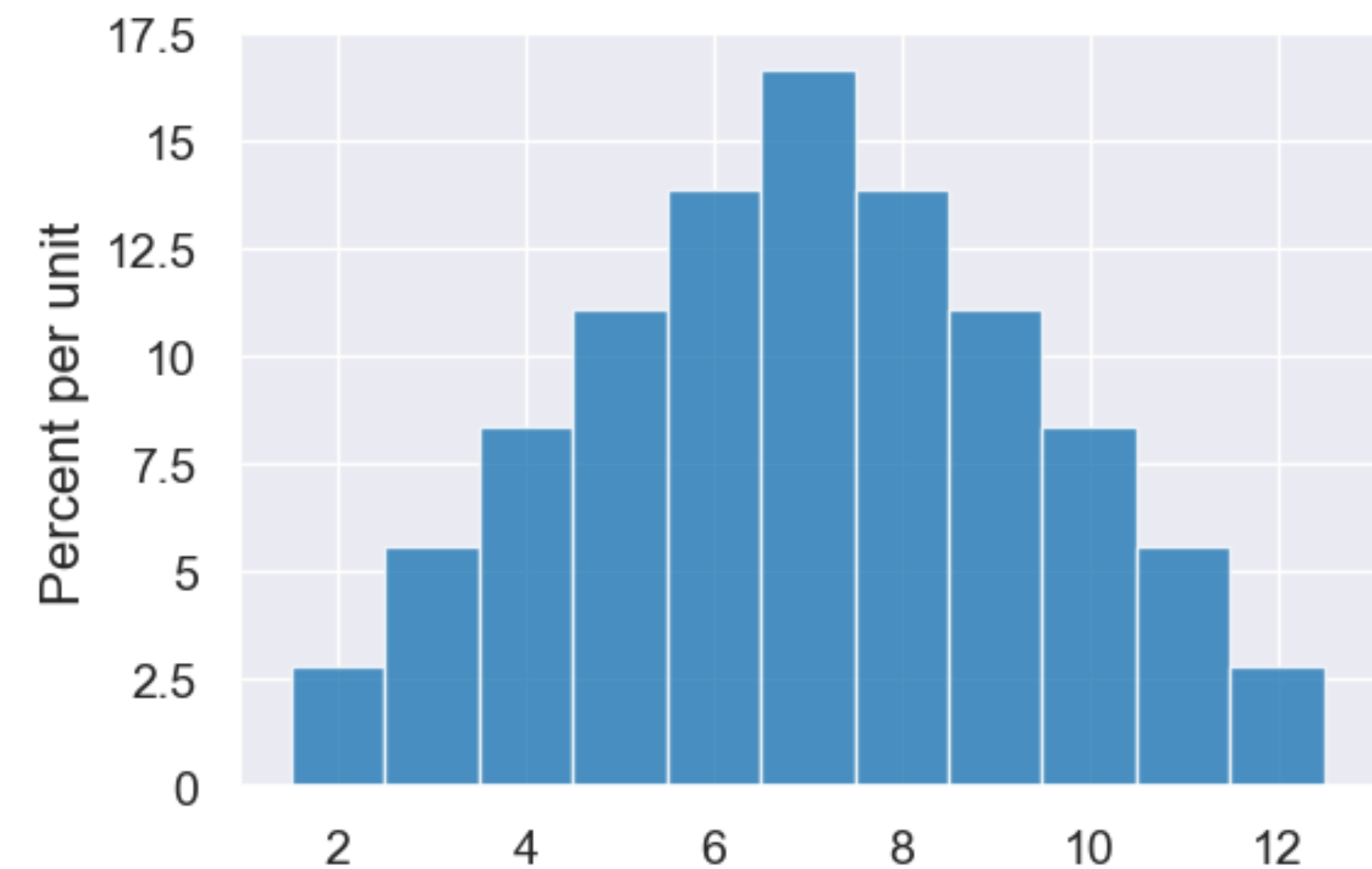
If you have a random quantity with various possible values, then its *probability (exact) distribution* associates

all the possible values of the quantity
with the probability of each of those values.



Sum of two dice rolls

Random quantity

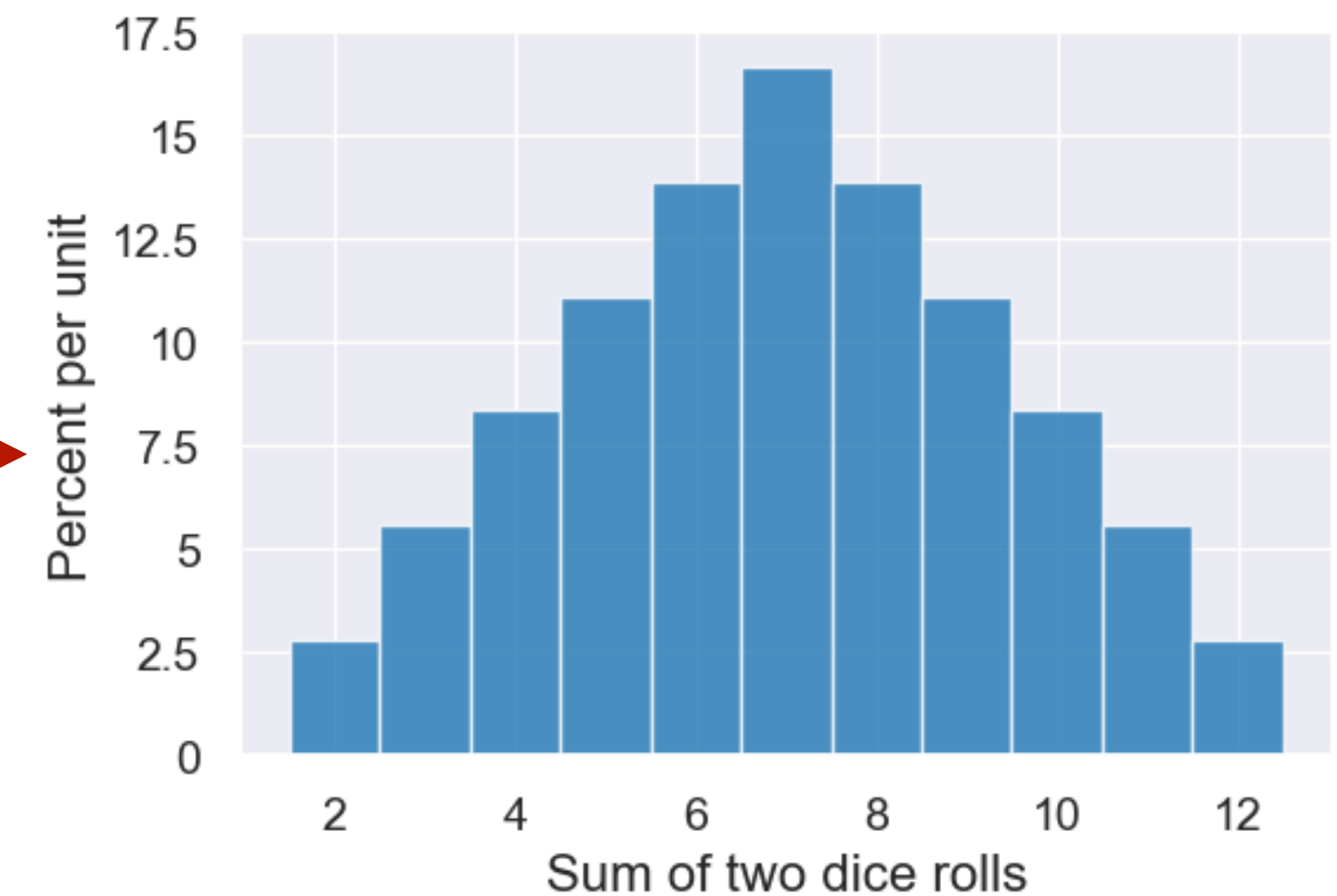


Every possible outcome

If you can do the math, you can work out the probability distribution without ever simulating it:

Employ probability rules

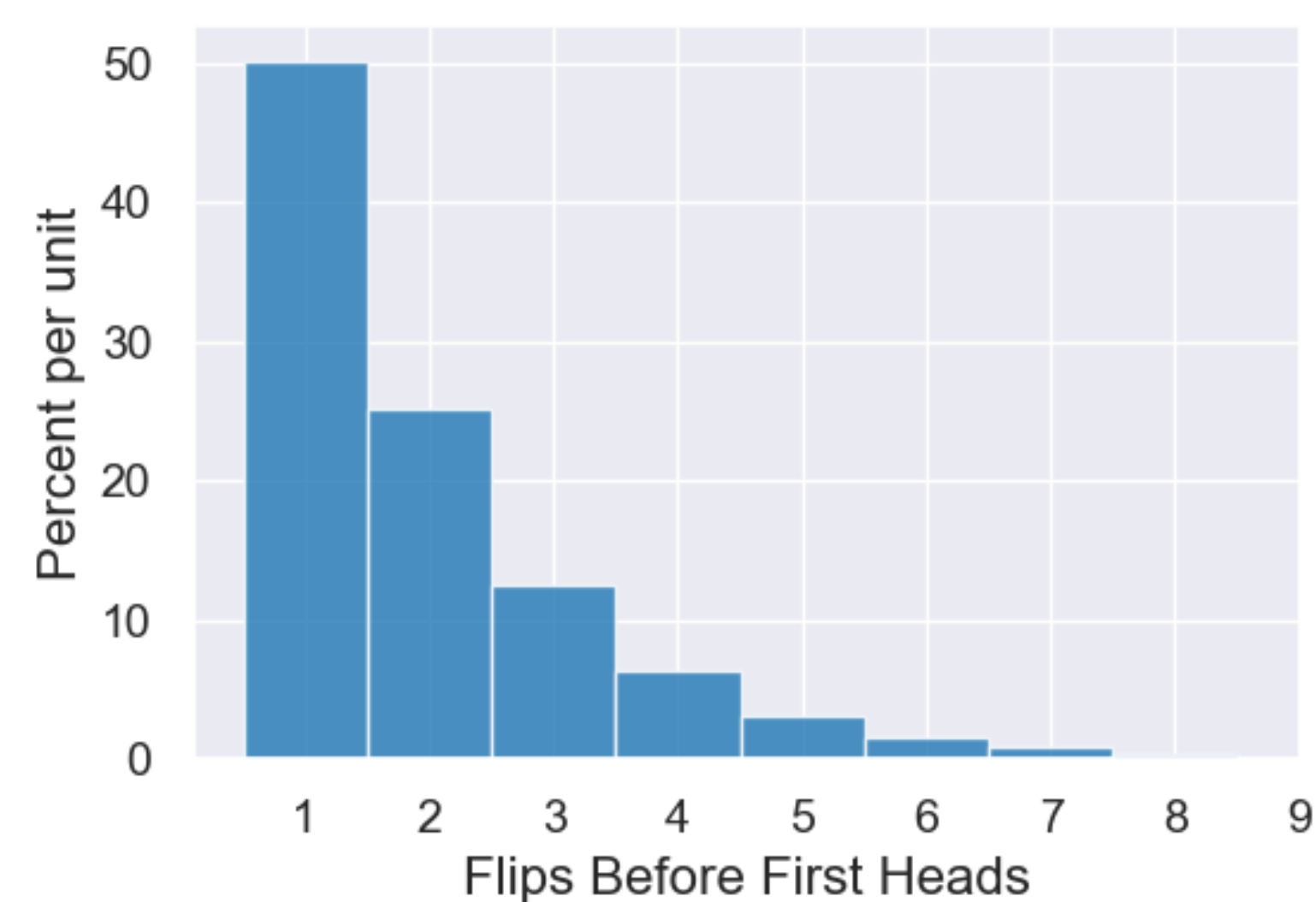
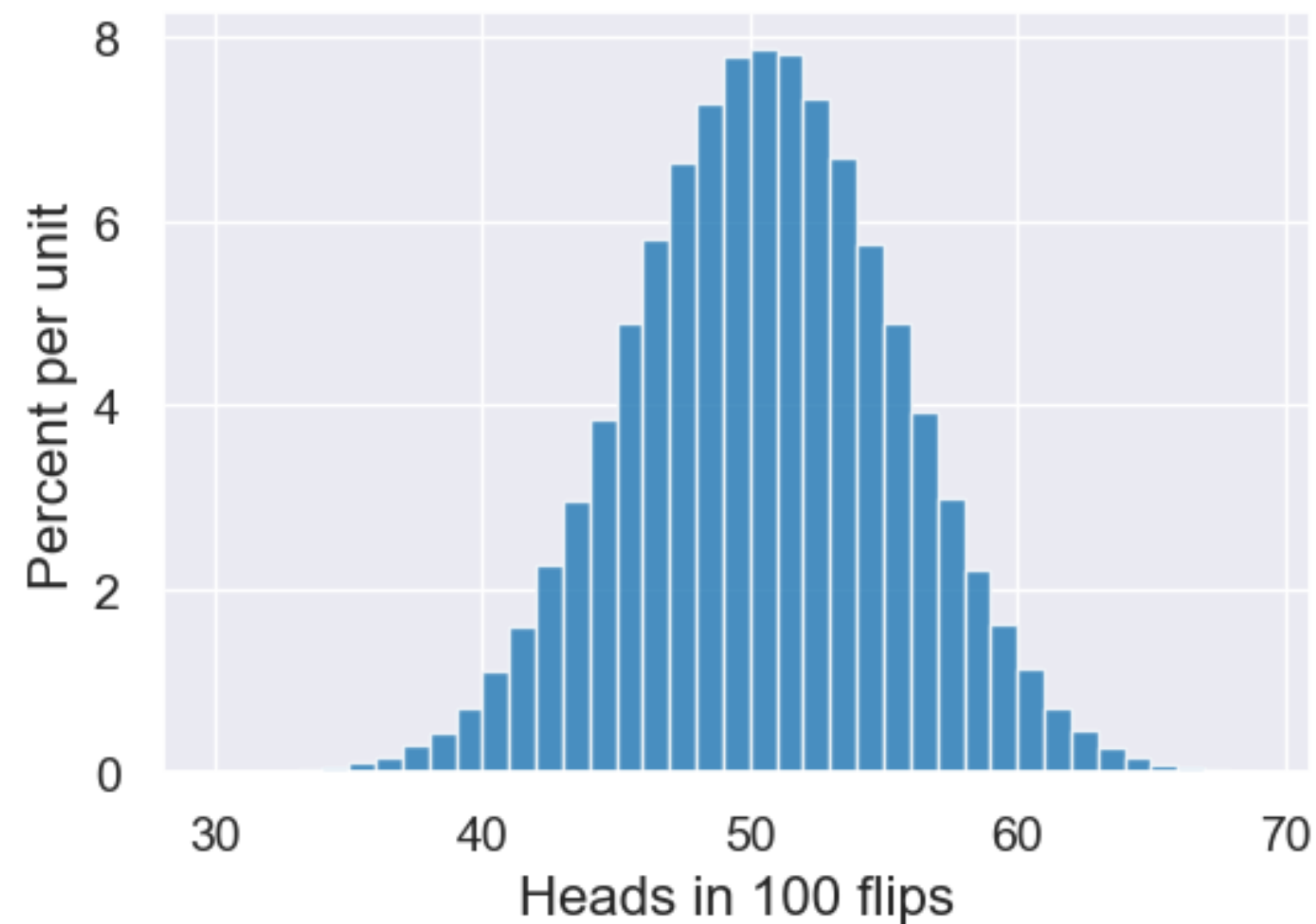
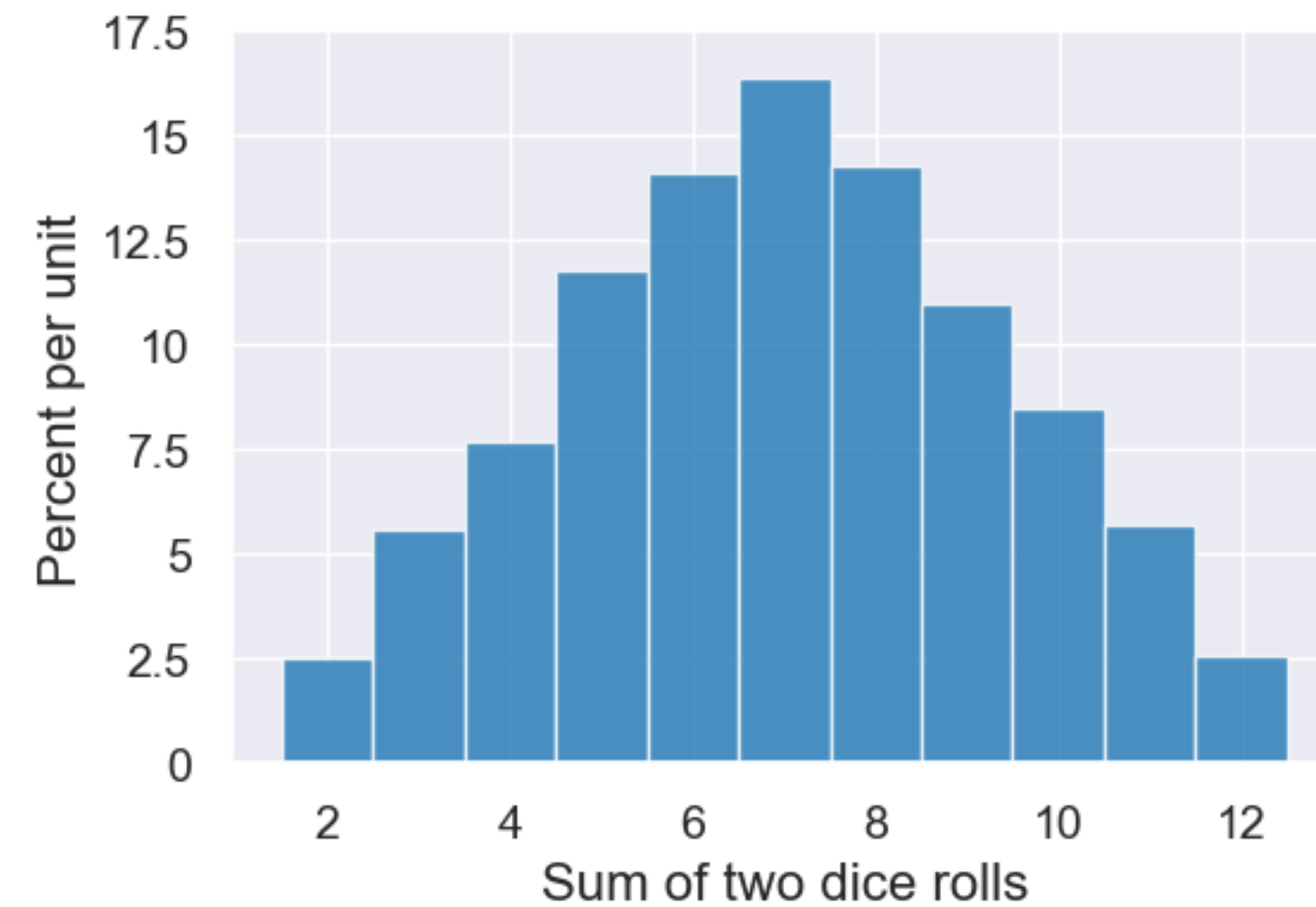
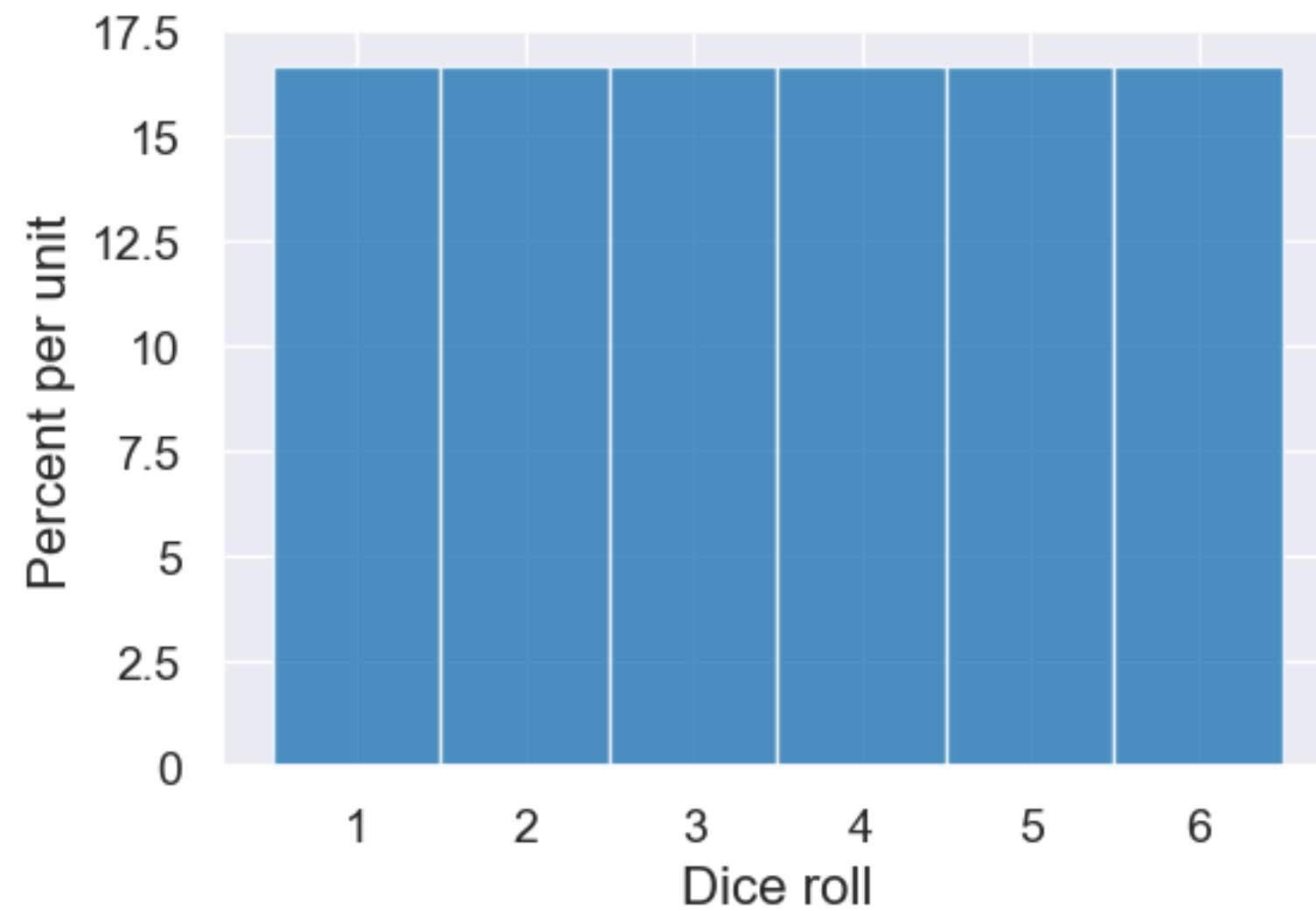
$$\begin{aligned}P(\text{sum is } 2) &= P(\square, \square) \\&= 1/6 \cdot 1/6 \\P(\text{sum is } 3) &= P(\square, \dot{\square}) + \\&\quad P(\dot{\square}, \square) \\&= \dots \\&\dots \\P(\text{sum is } 12) &= \dots\end{aligned}$$



Sum of two dice rolls

but simulating is often easier!

Probability distributions for other random quantities



How to calculate an event's probability?

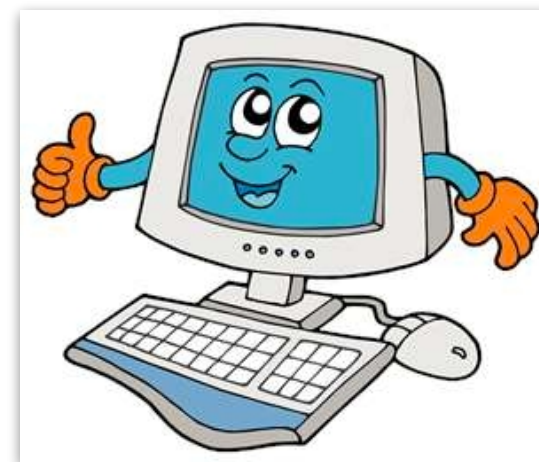
Computers (simulation)

```
N = 1000000 #Roll the dice 1 million times
option_a = np.random.choice(dice, N) + np.random.c
option_b = 2 * np.random.choice(dice, N)

print("Option A Mean: ", np.mean(option_a))
print("Option B Mean: ", np.mean(option_b))

Option A Mean: 7.003198
Option B Mean: 6.99884

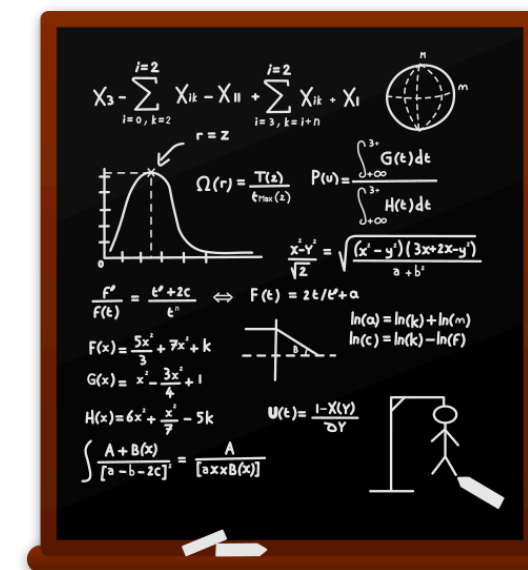
samples = Table().with_columns("Option A", option_
samples.hist("Option A", bins=np.arange(0,14))
samples.hist("Option B", bins=np.arange(0,14))
```



Rooted in *algorithms*

- ✗ Approximate solutions
- ✓ Often convincing
- ✓ Non-trivial problems potentially captured cleanly with code

Math (analytical)

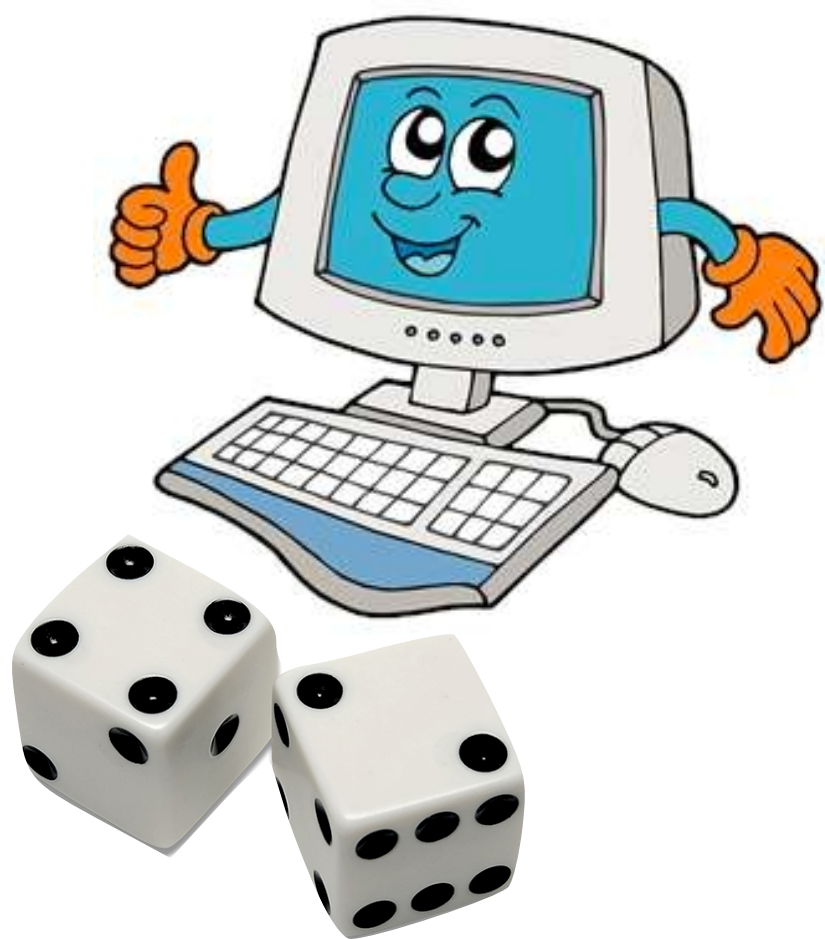


Rooted in *rules (axioms)*

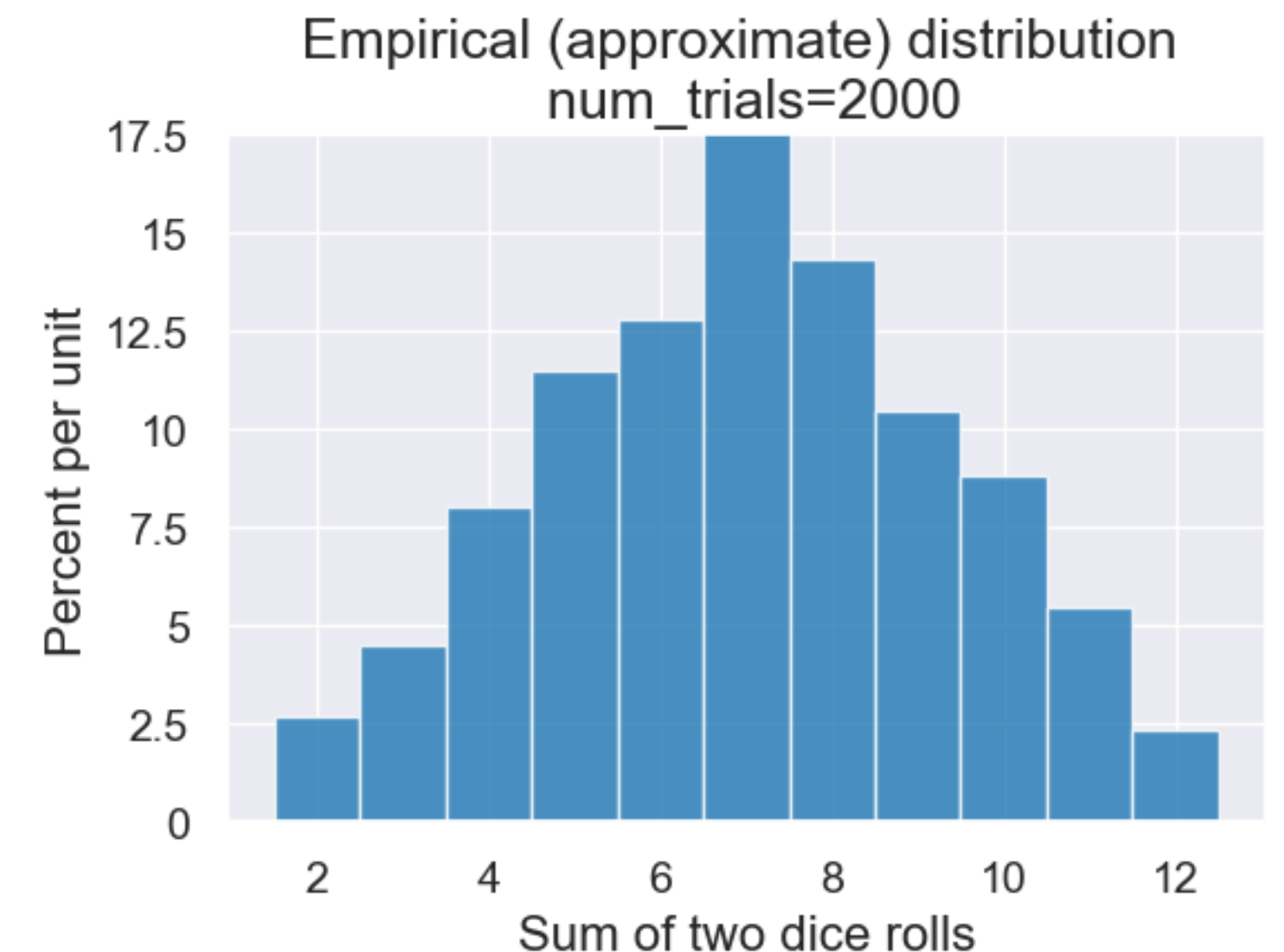
- ✓ Exact solutions
- ✓ Straightforward for simple problems
- ✗ Non-trivial problems potentially difficult to analyze/verify

An *empirical (approximate) distribution* consists of observations, which can be from repetitions of an experiment. It associates

all the unique values you actually *observed* with the proportion of times each value appeared.



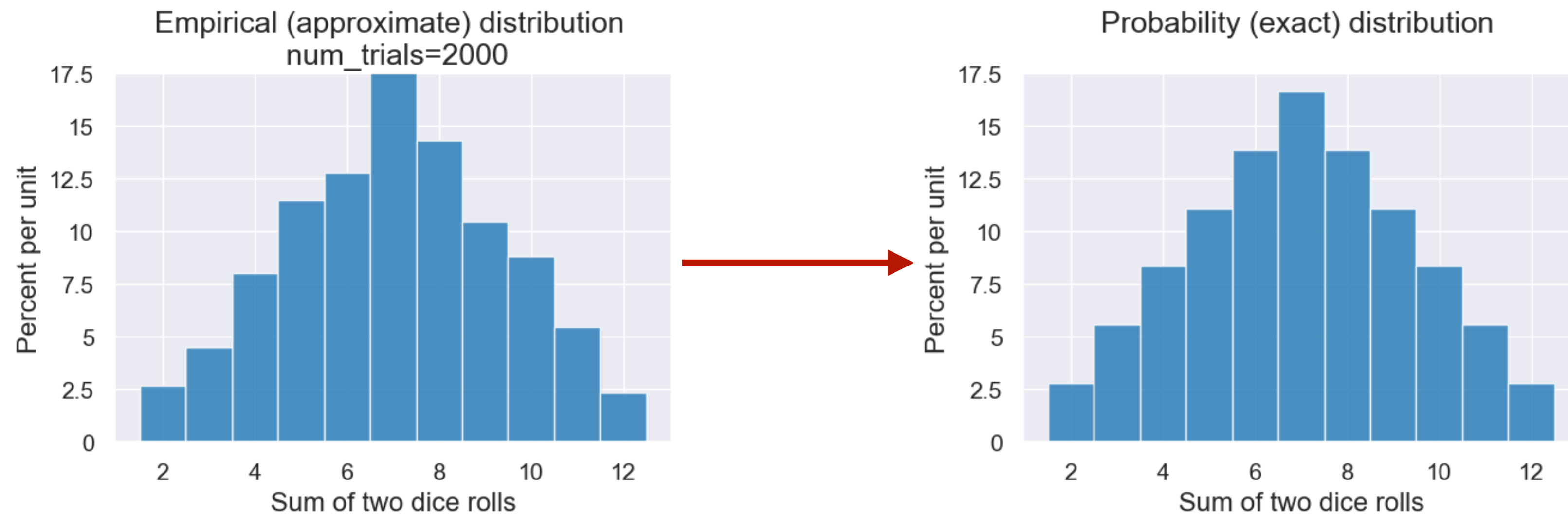
Outcomes: 9, 6, 8, 9, 5, 8, 8, 8, 8,
6, 9, 3, 3, 6, 9, 10, 6, 10, 7, 6, ...



Notebook: *Empirical distribution: Sum of two dice*

The *law of averages* (or *law of large numbers*):

If a chance experiment is repeated many times, independently and under the same conditions, then the proportion of times that an event occurs *gets closer* to the theoretical probability of the event.



Empirical distribution of a sample

If the sample size is large,

then the empirical distribution of a uniform random sample resembles the distribution of the population,

with high probability.

Real-world distributions and sampling

We could only simulate rolling two dice and taking their sum because we knew the true likelihood of each outcome for rolling a single die:

$$P(\text{1}) = 1/6$$

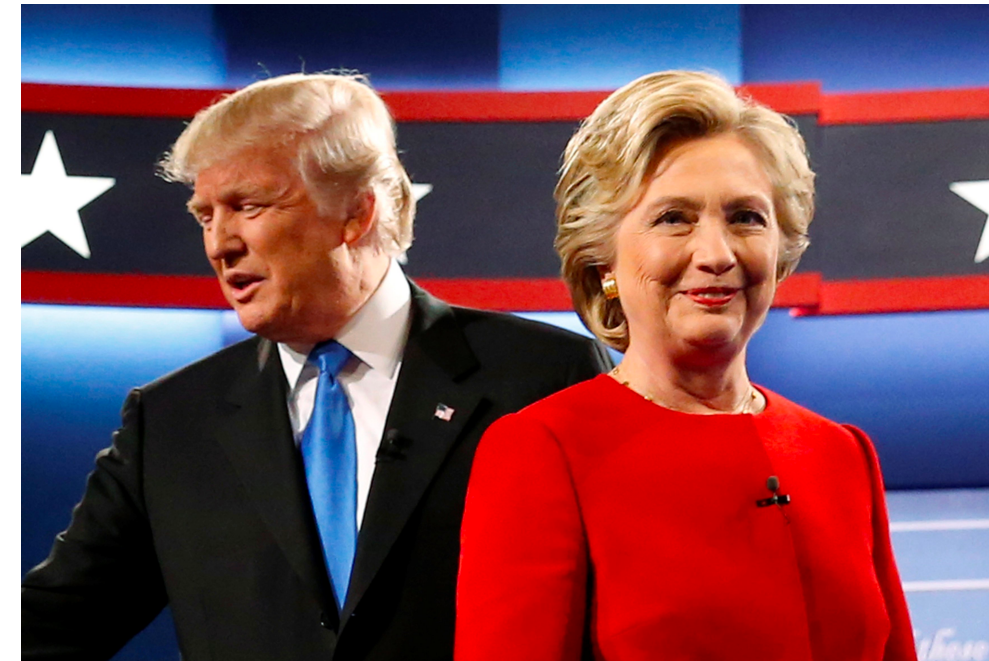
$$P(\text{2}) = 1/6$$

$$P(\text{3}) = 1/6$$

$$P(\text{4}) = 1/6$$

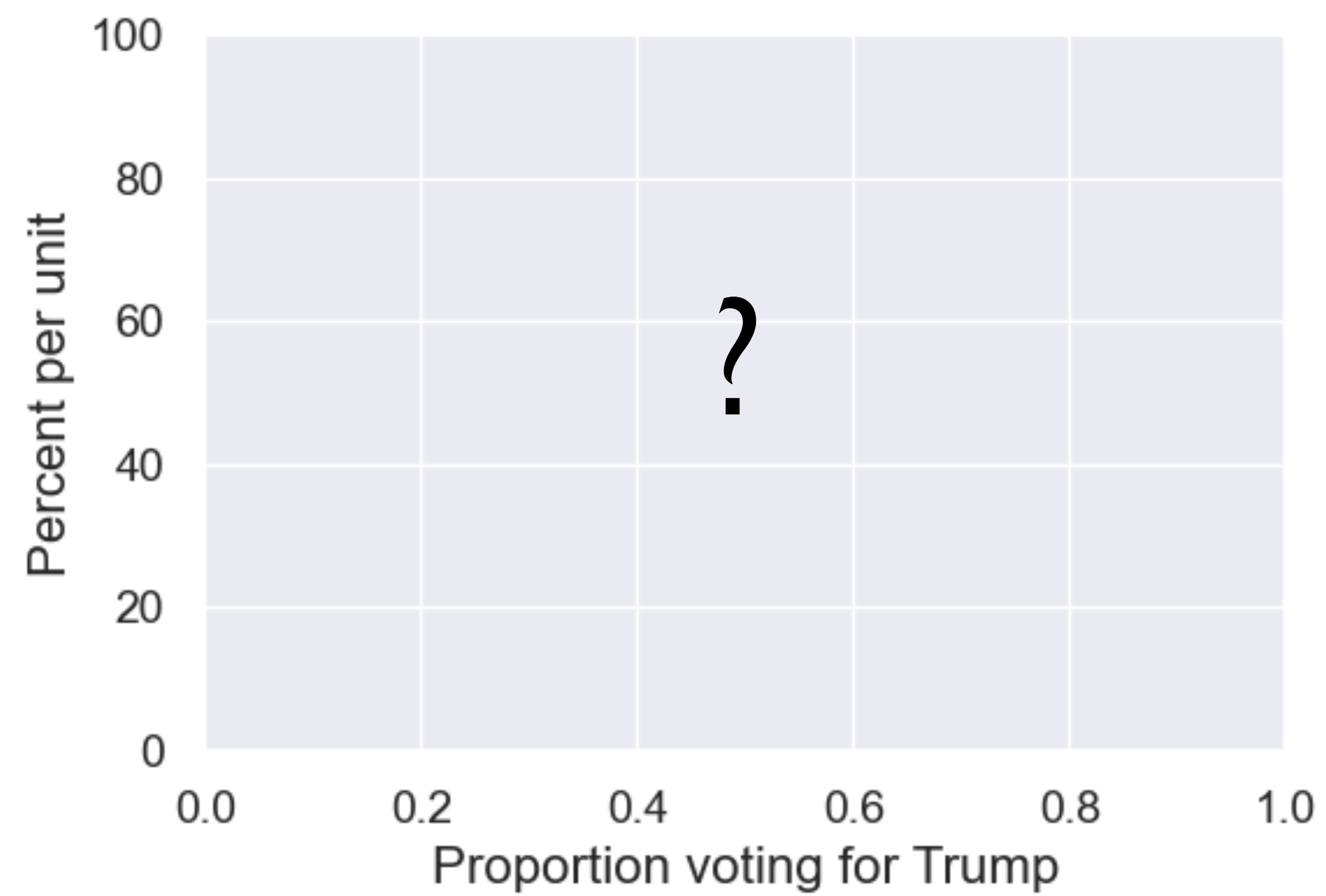
$$P(\text{5}) = 1/6$$

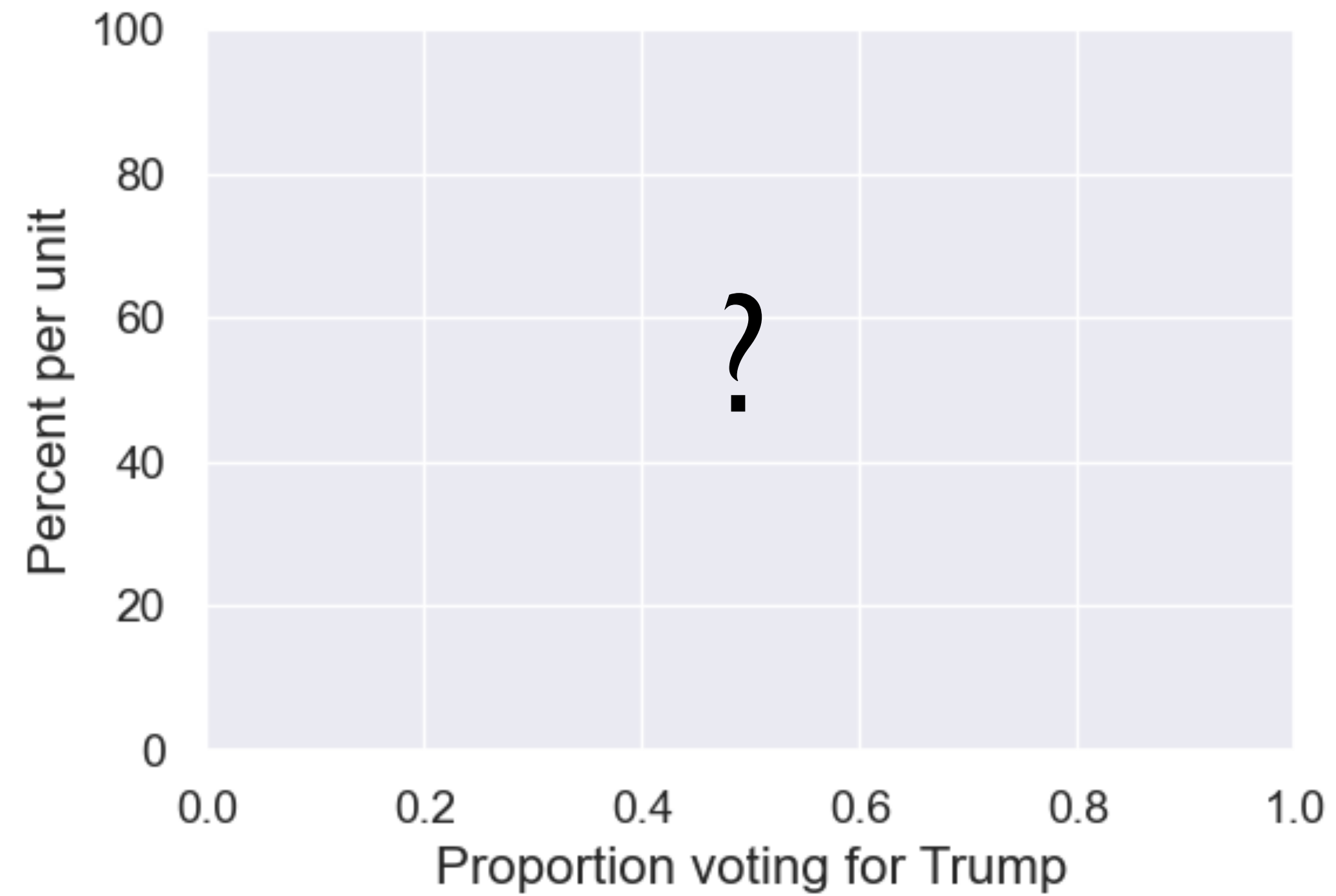
$$P(\text{6}) = 1/6$$



Circa 2016

*What is the probability distribution for a
candidate's chance of getting some percent of votes
in an upcoming election?*





We don't know the true probability of whether each person will vote for a given candidate; we can't compute this distribution the way we did before!



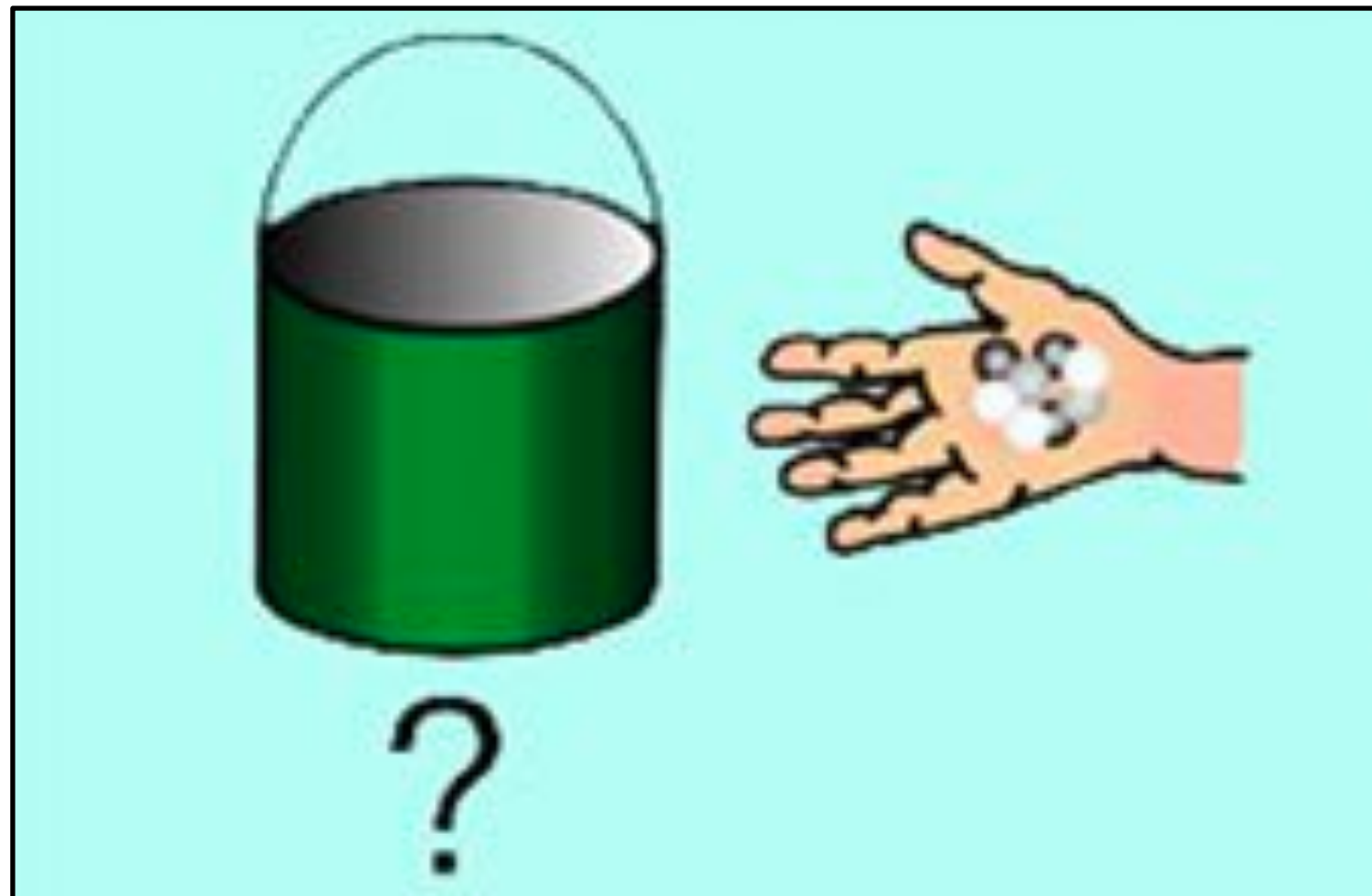
Probability

*Given the information in the
pail, what is in your hand?*



Probability

Given the information in the pail, what is in your hand?



Statistics

Given the information in your hand, what is in the pail?

A *population* is a set of all elements from which a subset called a *sample* will be drawn.

How do we select our sample?

How do we draw meaningful conclusions using a sample?



Statistics

Given the information in your hand, what is in the pail?

Selecting a sample

Not all samples involve chance!

Here's an example of *deterministic sampling*:



Pollster: Call every person with the first name “Bob” and ask him who he’s voting for.

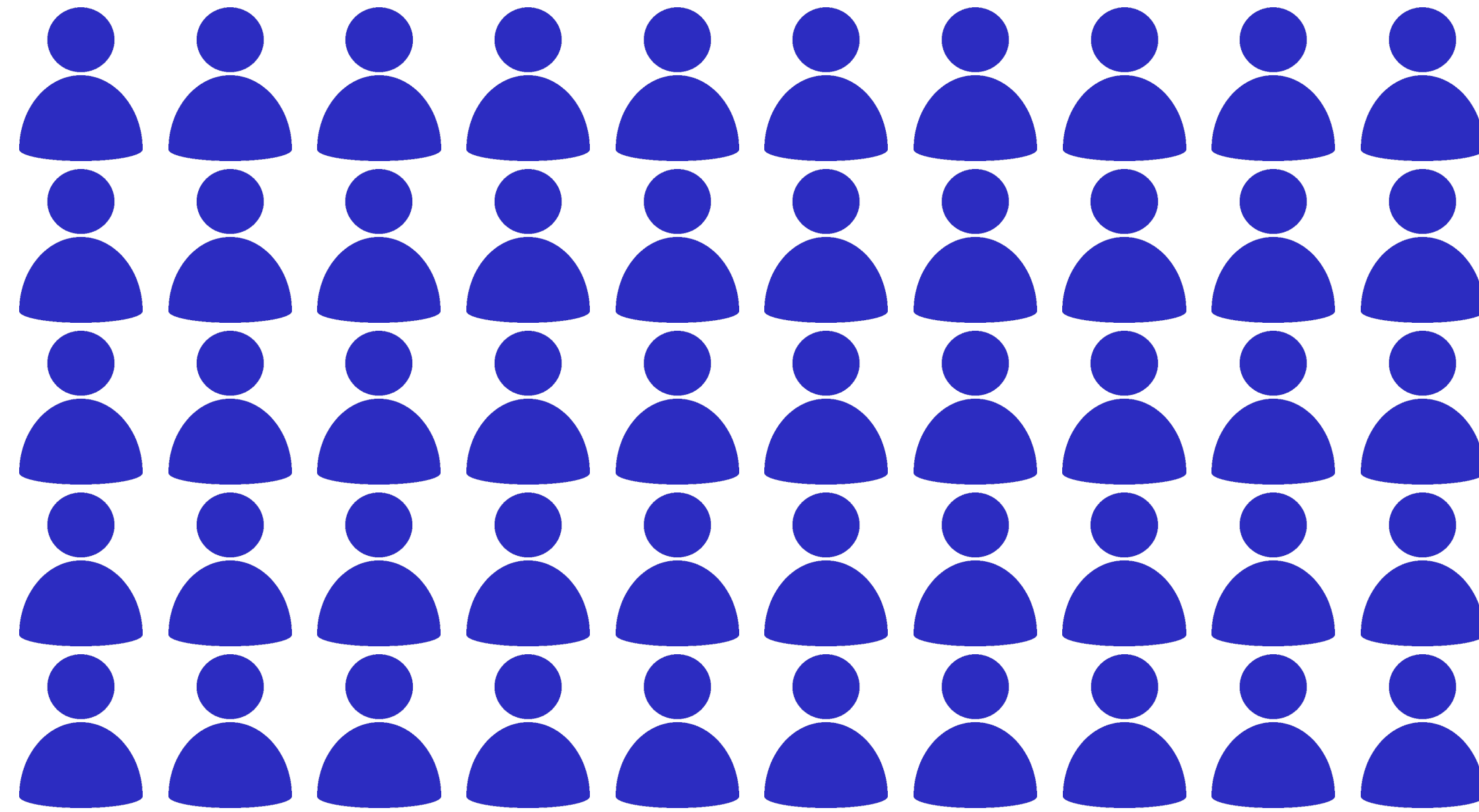
Not all samples involve chance!

Here's an example of *deterministic sampling*:



Pollster: Call every person with the first name "Bob" and ask him who he's voting for.

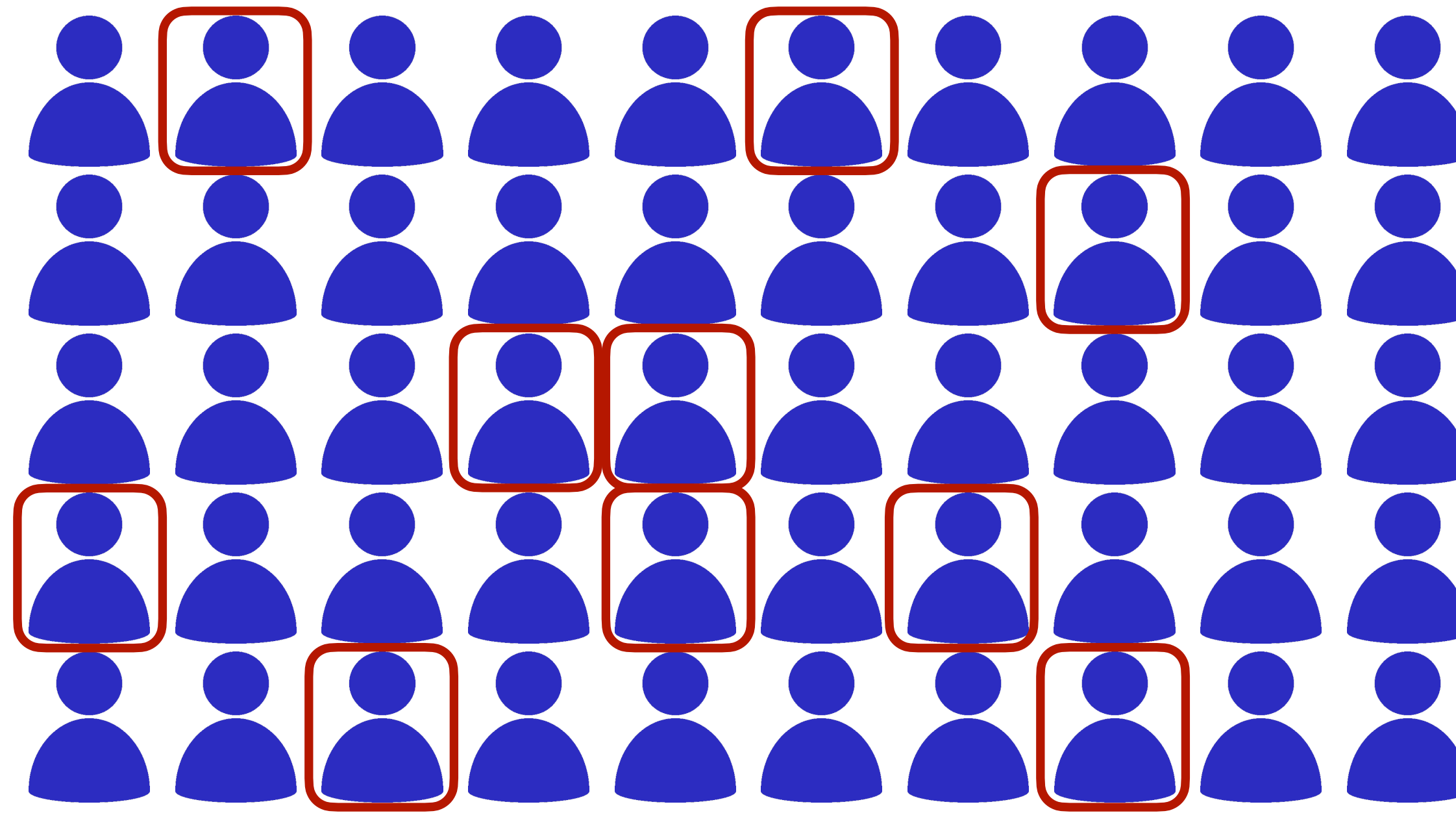
What's wrong with this?



In *random sampling*, each member of the population has some probability of being picked.



*Example:
Random-digit dialing*



In *random sampling*, each member of the population has some probability of being picked.

HOMESEARCH

The New York Times

Alaska Ariz. Colo. Fla. Ga. Ind. Iowa Miss. Mo. N.C. N.H. Nev. Ohio Pa. Wis.

Who Will Be President?

By JOSH KATZ


The estimates on this page are based on pre-election polls. For an estimate including results, The Times is providing live forecasts on election night.

PresidentSenateHouse


Hillary Clinton has an 85% chance to win.

Last updated Tuesday, November 8 at 10:20 PM ET

CHANCE OF WINNING



85%
Hillary Clinton



15%
Donald J. Trump

Forecast historyRecent changesState by stateOther forecastsLikely scenariosExplore paths

The Upshot's elections model suggests that Hillary Clinton is favored to win the presidency, based on [the latest state and national polls](#). A victory by Mr. Trump remains possible: Mrs. Clinton's chance of losing is about the same as the probability that [an N.F.L. kicker misses a 37-yard field goal](#).

For months, we've been updating our estimates with each new poll. Today, it's Election Day, what we've all been waiting for, and there will be no more updates. You can chart different paths to victory below. Here's how our estimates have changed over time:

100%80%60%40%20%0%

Clinton85%

Trump15%

JuneJulyAugustSeptemberOctoberNovember

November 8

"All the News That's Fit to Print"

The New York Times

Late Edition
Today, cloudy, showers midday, high 36. Tonight, strat evening showers, clouds breaking late, colder, low 40. Tomorrow, sunshine, high 56. Weather map appears on Page B1.

VOL. CLXVI ... No. 57,411 ++

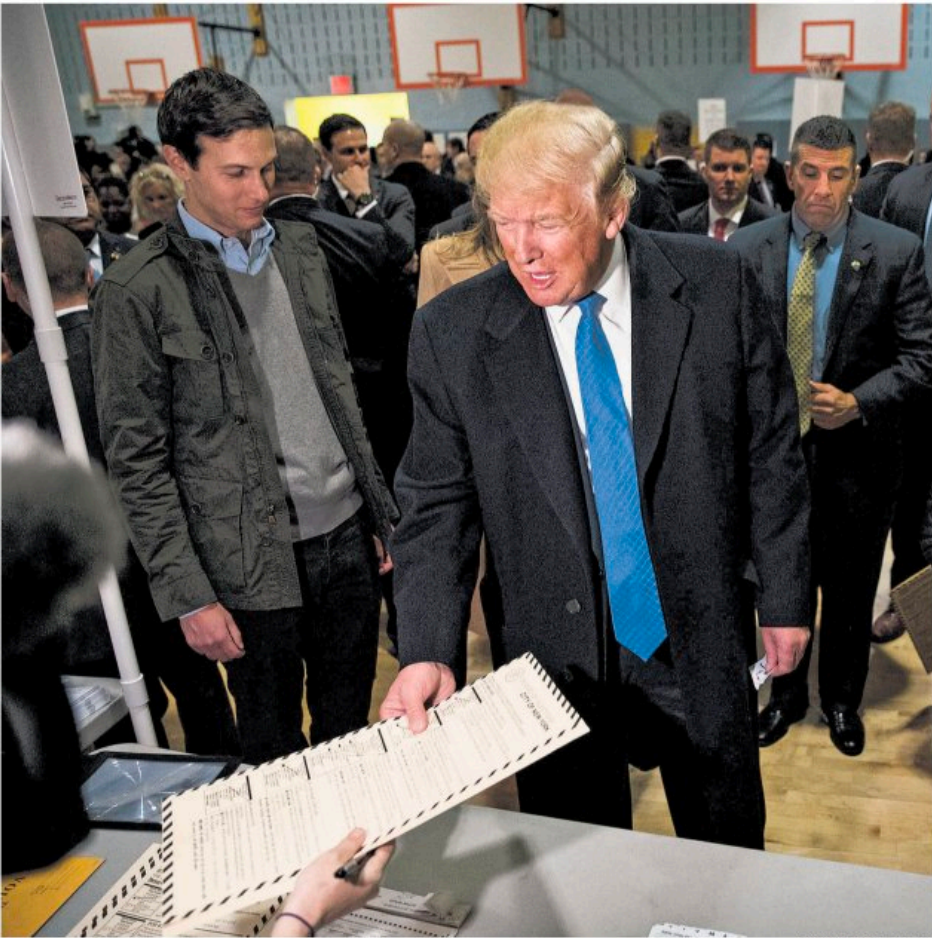
© 2016 The New York Times Company

NEW YORK, WEDNESDAY, NOVEMBER 9, 2016

\$2.50

TRUMP TRIUMPHS

OUTSIDER MOGUL CAPTURES THE PRESIDENCY, STUNNING CLINTON IN BATTLEGROUND STATES



WORKING CLASS SPEAKS

Blue-Collar Whites Give Stinging Rebuke to Democratic Party

By PATRICK HEALY and JONATHAN MARTIN
Donald John Trump was elected the 45th president of the United States on Tuesday in a stunning culmination of an explosive, populist and polarizing campaign that took relentless aim at the institutions and long-held ideals of American democracy. The surprise outcome, defying late polls that showed Hillary Clinton with a modest but persistent edge, threatened convictions throughout the country and the world, where skeptics had watched with alarm as Mr. Trump's unvarnished overtures to disillusioned voters took hold. The triumph for Mr. Trump, 70, a real estate developer-turned-reality television star with no government experience, was a powerful rejection of the establishment forces that had assembled against him, from the world of business to government, and the consensus they had forged on everything from trade to immigration. The results amounted to a repudiation, not only of Mrs. Clinton, but of President Obama, whose legacy is suddenly imperiled. And it was a decisive demonstration of power by a largely overlooked coalition of mostly blue-collar white and working-class voters who felt that the promise of the United States had slipped their grasp amid decades of globalization and multiculturalism. In Mr. Trump, a brash-married Manhattanite who lives in a marble-wrapped three-story penthouse apartment on Fifth Avenue, they found an improbable champion. Mr. Trump's strong showing helped Republicans regain control of the Senate. Only one Republican-controlled seat, in Illinois, fell to Democrats early in the evening. And Senator Richard Burr of North Carolina, a Republican, easily won re-election in a race that had been among the country's most competitive. A handful of other Republican incumbents facing difficult races were running better than expected.

Continued in Election 2016, Page 5

AMBRIDGE JOURNAL

A Blue-Collar Town in Decline And in Despair Turns to Trump

By TRIP GARRIEL
AMBRIDGE, Pa. — As Donald J. Trump's surprisingly strong showing played out on a television above Fred's Diner bar, the men who by day carry pipes, hang dry wall and drive locomotives watched the returns with mounting satisfaction. "He's killing it — that's our next president," said John Gagnio, 50, who had affixed an "I voted" sticker to the blue uniform shirt he wears in a betting plant. "We need a change. We've got to get rid

NEWS ANALYSIS

Around the World, Uncertainty And Fear That 'All Bets Are Off'

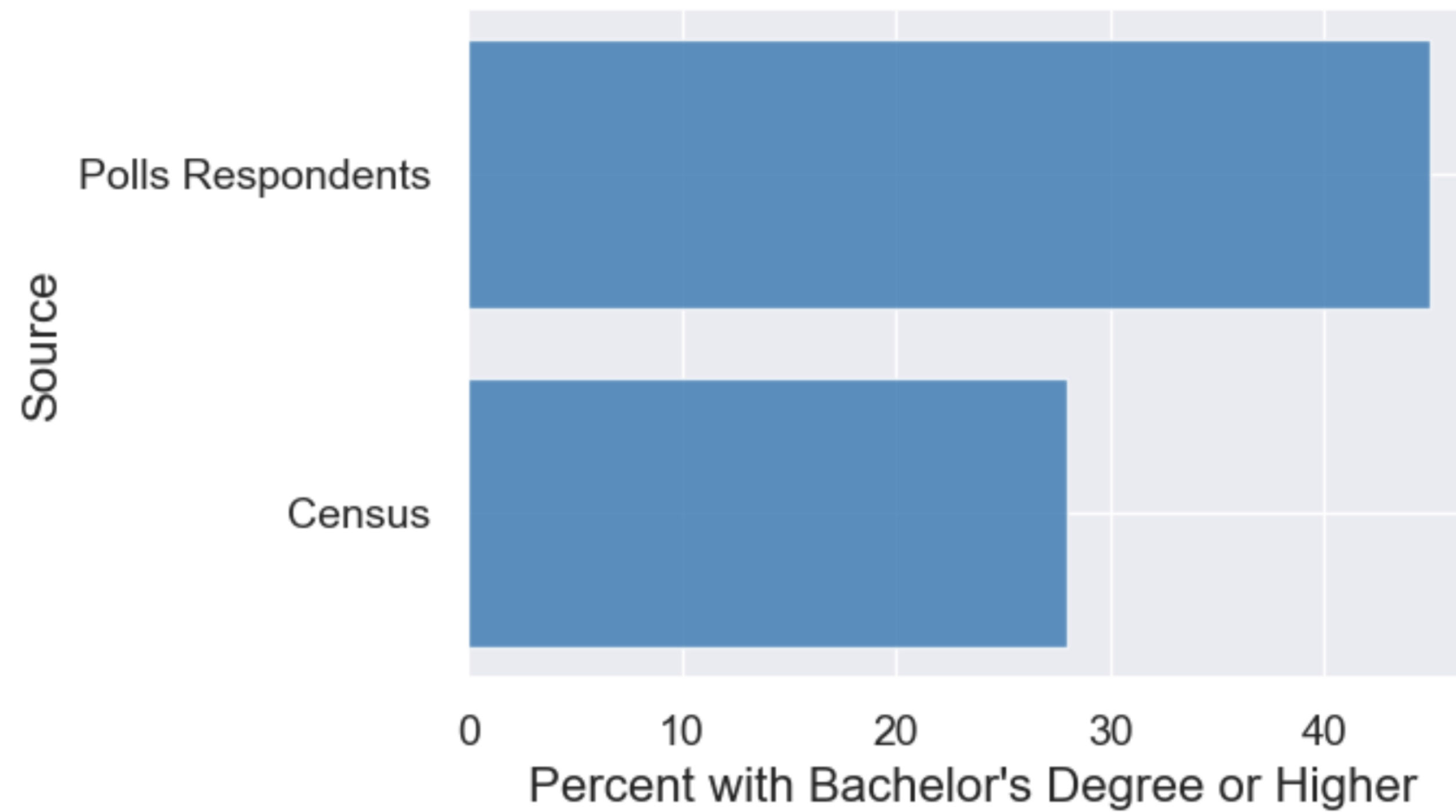
By PETER BAKER
JERUSALEM — Donald J. Trump's stunning election victory on Tuesday night ripped away beyond the nation's borders, sparking an international order that prevailed for decades and raising profound questions about America's place in the world. For the first time since before World War II, Americans chose a president who promised to reverse the internationalism practiced by predecessors of both

MAN IN THE NEWS

Clarion of White Populist Rage Who Vowed 'I Am Your Voice'

By ALEXANDER BURNS
Donald John Trump defied the skeptics who said he would never win, and the political veterans who scoffed at his slapdash campaign. He attacked the norms of American politics, singling out groups for derision on the basis of race and religion and attacking the legitimacy of the political process. He ignored conventions of common decency, employing casual vulgarity and raising personal hostility on his political oppo-

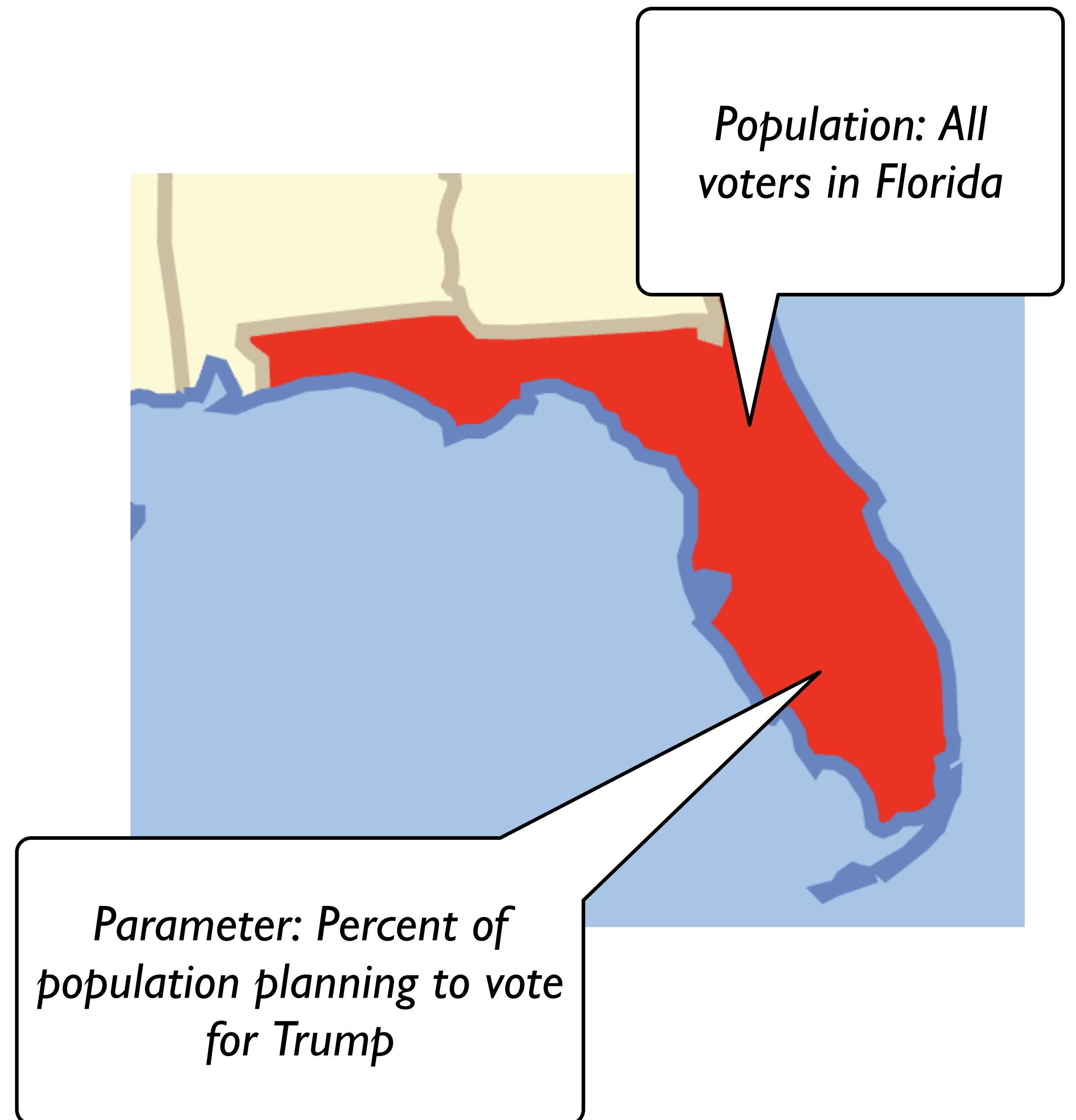
New York Times,
Tuesday November 8, 2016



After the 2016 election, [analysis](#) showed individuals with higher education were *overrepresented* in polling samples.

Terminology

Parameter: A fixed number associated with a *population*

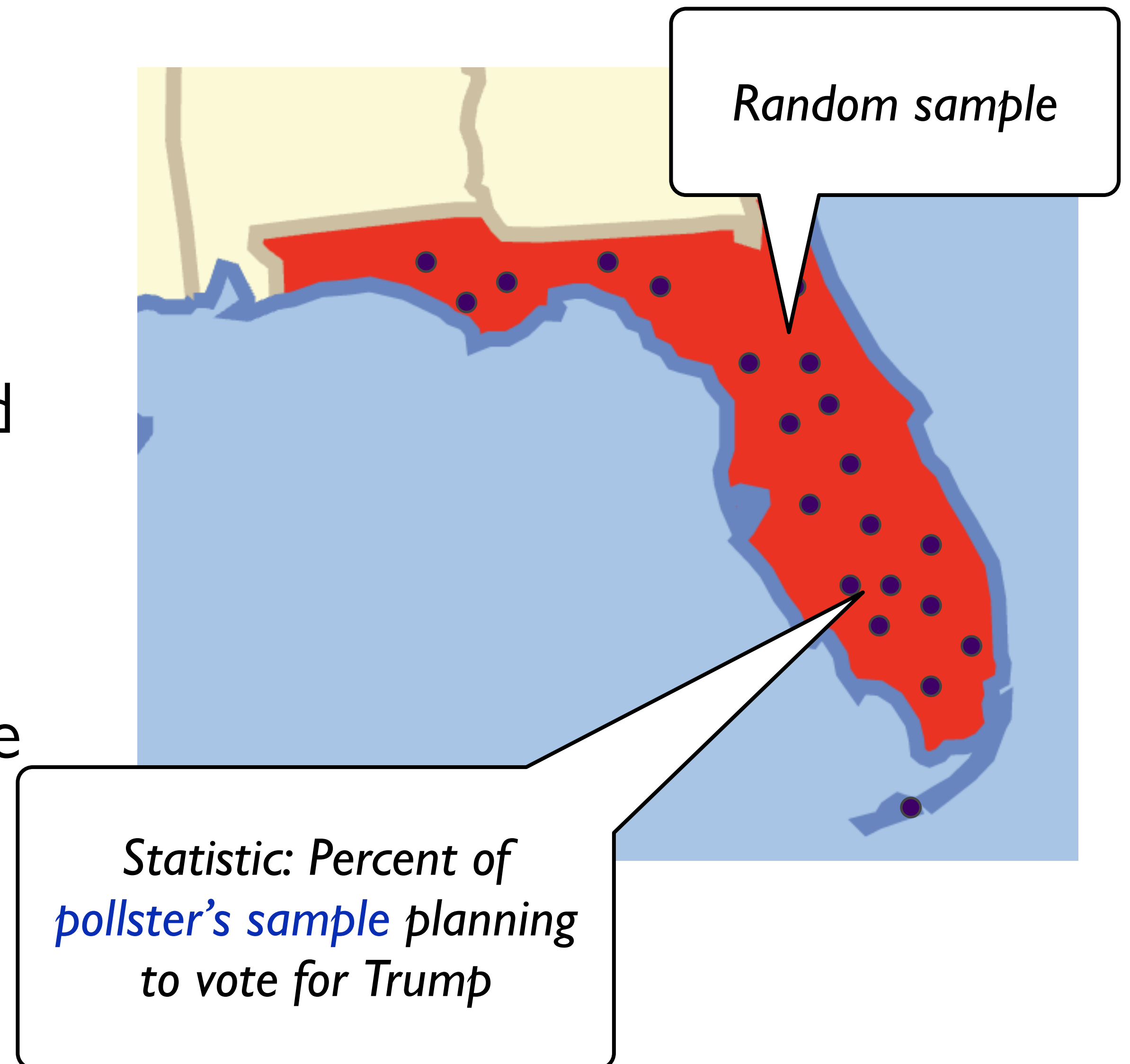


Terminology

Parameter: A fixed number associated with a **population**

Statistic: Any number computed using the data in a sample, e.g., the mean or median.

Statistical inference: Estimate the value of the parameter with statistics

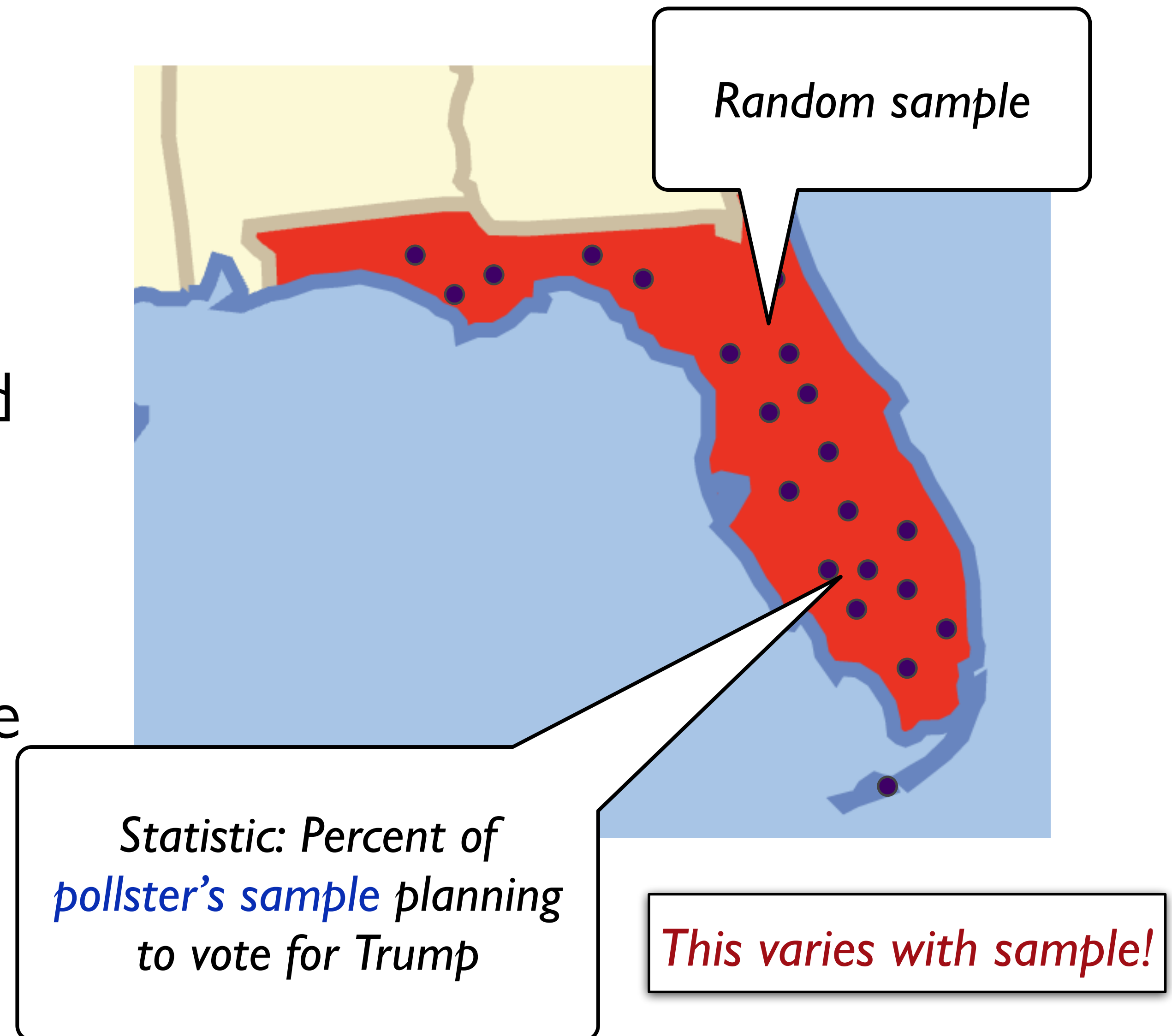


Terminology

Parameter: A fixed number associated with a **population**

Statistic: Any number computed using the data in a sample, e.g., the mean or median.

Statistical inference: Estimate the value of the parameter with statistics



Notebook: *Random sampling: Florida votes in 2016*

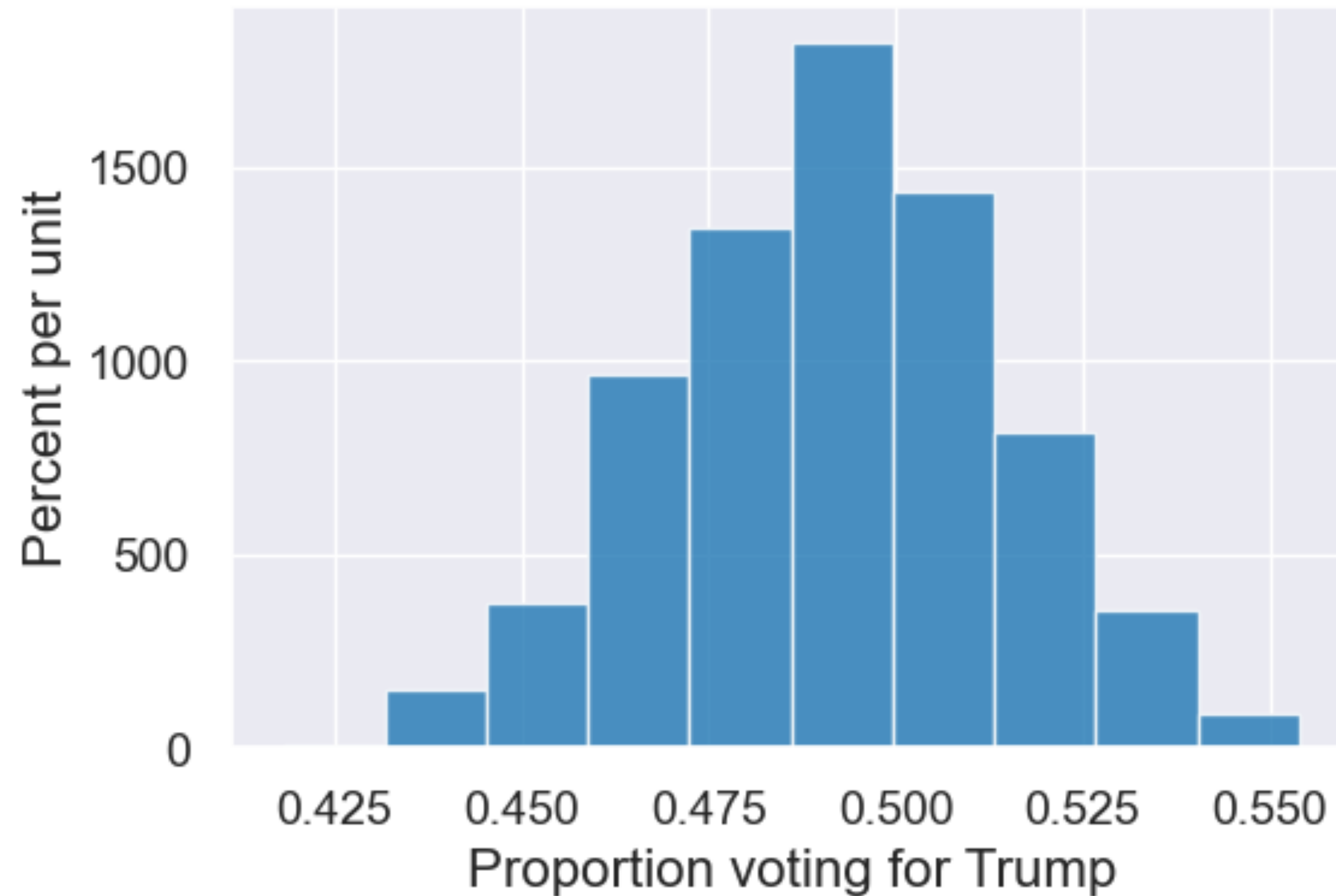
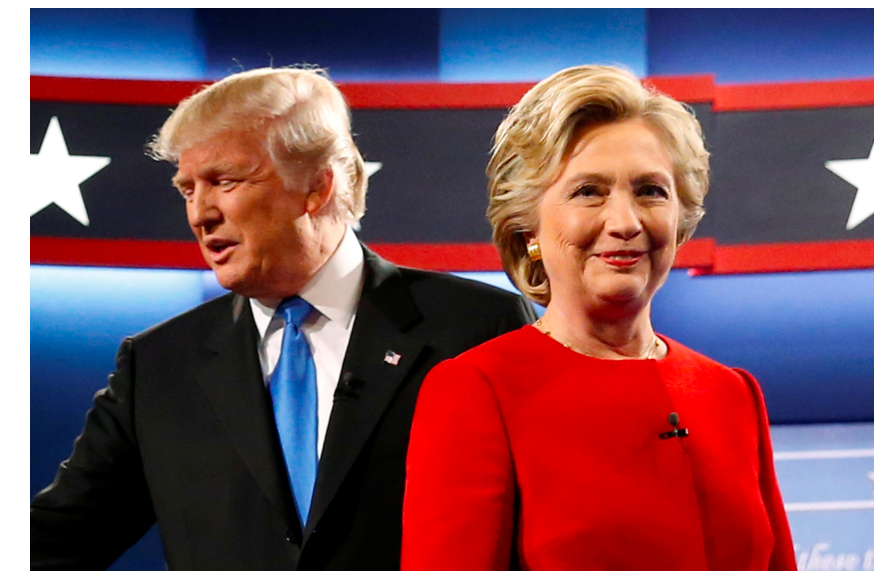
Notebook: *General sampling function*

The *sampling distribution* or probability distribution of the statistic consists of all possible values of the statistic and their corresponding probabilities.

This can be hard to calculate!

Either you need to do the math or you need to generate all possible samples and calculate the statistic based on each sample.

Empirical distribution of a statistic



1. Observe the statistic from repetitions of a (sampling) experiment or simulation.

2. Create a distribution of statistics (i.e., a histogram)

The *empirical distribution of the statistic* is

based on simulated values of the statistic and

consists of

all the observed values of the statistic and

the proportion of times each value appeared.

The empirical distribution is a good approximation to the probability distribution of the statistic – if the number of repetitions in the simulation is large!

A fundamental consideration in using any statistic based on a *random sample* is that

the sample could have come out differently, and

therefore the statistic could have come out differently too!

