Materials: http://huppenkothen.org/bayesian-statistics-tutorial/

Bayesian Statistics

Daniela Huppenkothen

DIRAC, University of Washington

Tiana_Athriel

dhuppenkothen

dhuppenk@uw.edu

All models are wrong, but some are useful

— George Box

Nature is complex ...

… and so is our data collection!

Quick poll!

Part 1: Probabilities

What snacks will we have for coffee break today?

What snacks will we have for coffee break today?

p()

p()

Basic rule of probability (1)

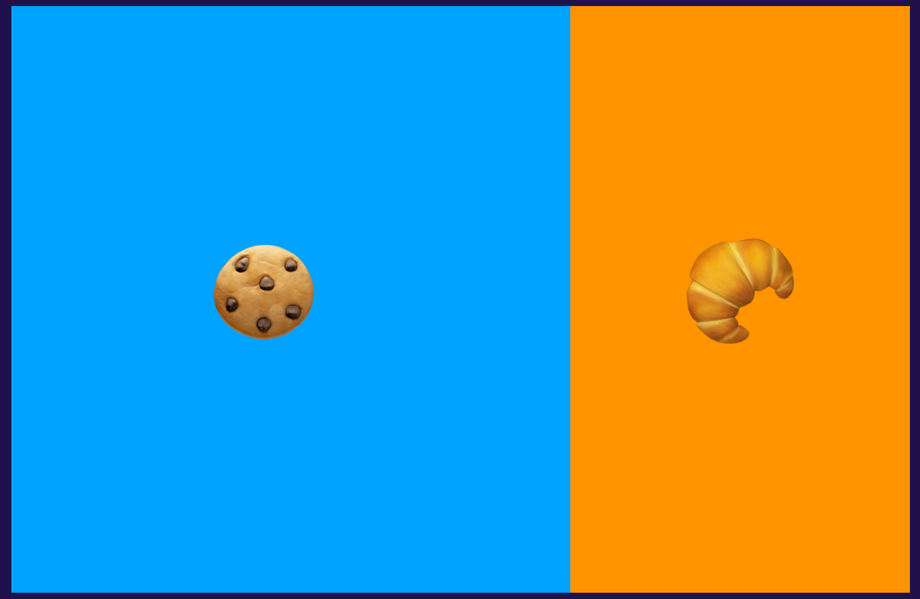
$$0 \leq p(\text{🍪}) \leq 1$$

If p(🍪) = 0.3, what is p(🥐)?

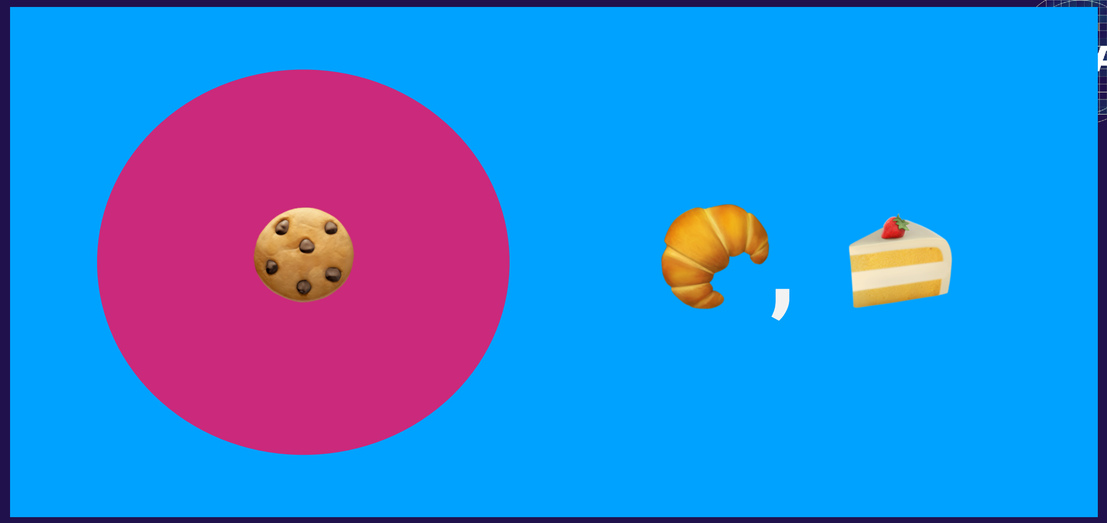
a) 0.3
b) 0.5
c) 0.01
d) 0.7

p(🍪) + p(🥐) = 1

$$\sum_i P_i(x) = 1$$



p(🍪) + p(🥐) + p(🍰) = 1



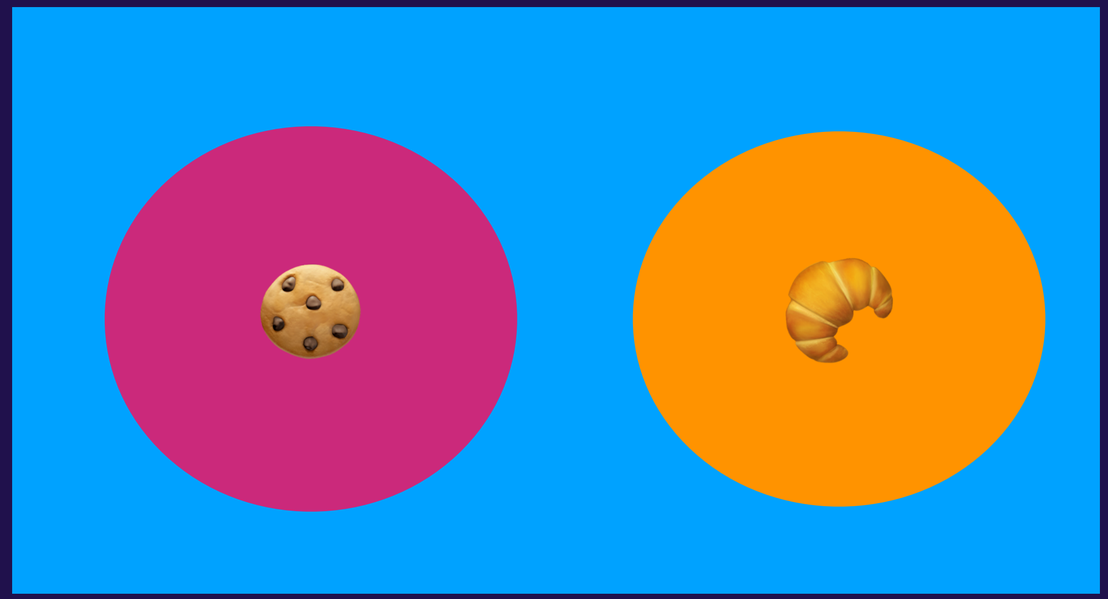
p(🥐) + p(🍰) = p(not 🍪)

“complement”

p(🍪) = 1 - p(not 🍪)

p(🍪 ∩ 🥐) = “p(🍪 and 🥐)”

p(🍪 ∪ 🥐) = “p(🍪 or 🥐)”



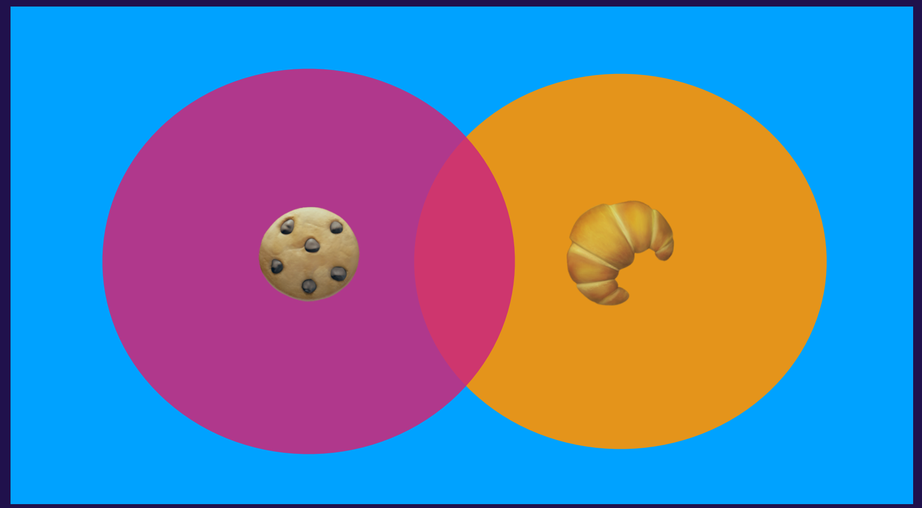
🍪 , 🥐 are independent:

p(🍪 ∩ 🥐) = 0

p(🍪 ∪ 🥐) = p(🍪) + p(🥐)

p(🍪 ∩ 🥐) = “p(🍪 and 🥐)”

p(🍪 ∪ 🥐) = “p(🍪 or 🥐)”



🍪 , 🥐 are **not** independent:

p(🍪 ∩ 🥐) = x

p(🍪 ∪ 🥐) = p(🍪) + p(🥐) - p(🍪 ∩ 🥐)

Let’s add another category

Is there a break with snacks in the morning or the afternoon?

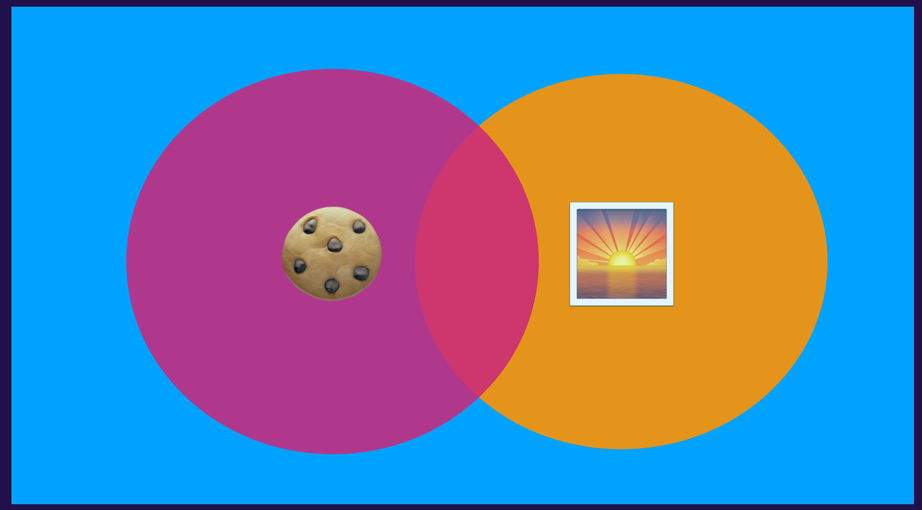
p(🌅) = there are snacks in the morning
p(😴) = there are snacks in the afternoon

Do we get cookies more often in the morning or in the afternoon?

p(🍪 | 🌅) = “probability of 🍪 given 🌅 ”

p(🍪 | 🌅) = p(🍪 ∩ 🌅) / p(🌅)

p(🍪 | 🌅) = p(🍪 , 🌅) / p(🌅)



Which snacks do we get at which time of the day?

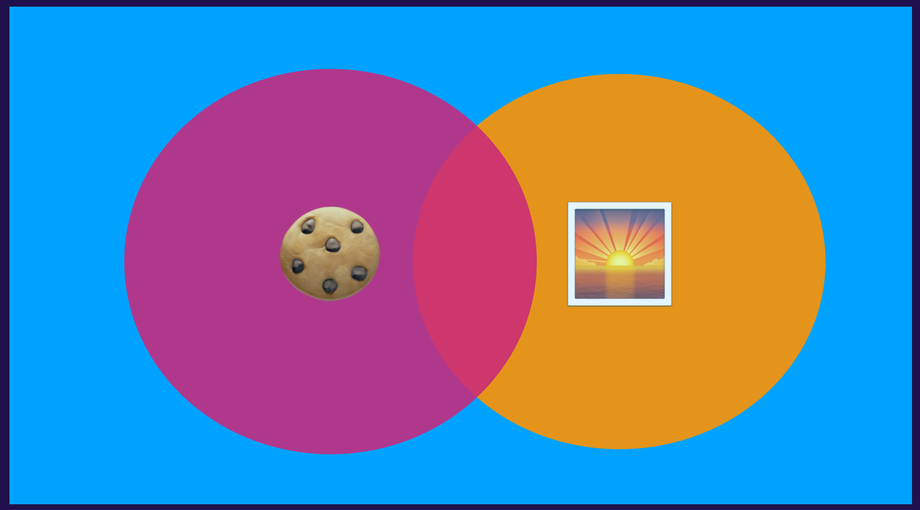
	p(🍪)	p(🥐)	p(🍰)
p(🌅)	0.25	0.06	0.29
p(😴)	0.25	0.14	0.01

- p(🍰) = ?
- p(🌅 | 🥐) = ?
- p(🥐) = ?
- p(🍰 | 🌅) = ?

p(🍪 | 🌅) = p(🍪 , 🌅) / p(🌅)

What is p(🌅 | 🍪) ?

p(🌅 | 🍪) = p(🍪 , 🌅) / p(🍪)

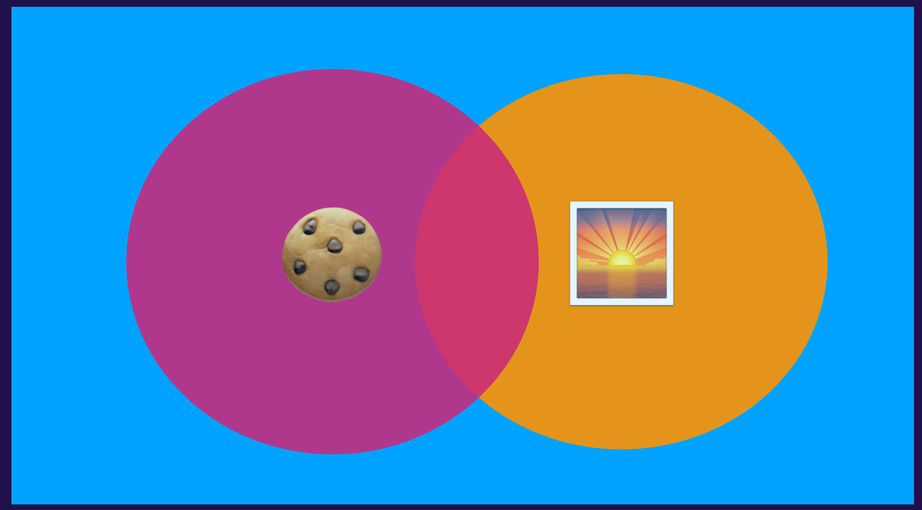


p(🍪 | 🌅) = p(🍪 , 🌅) / p(🌅)

p(🌅 | 🍪) = p(🍪 , 🌅) / p(🍪)

p(🌅 | 🍪) p(🍪) = p(🍪 | 🌅) p(🌅)

Bayes rule



$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Bayes rule

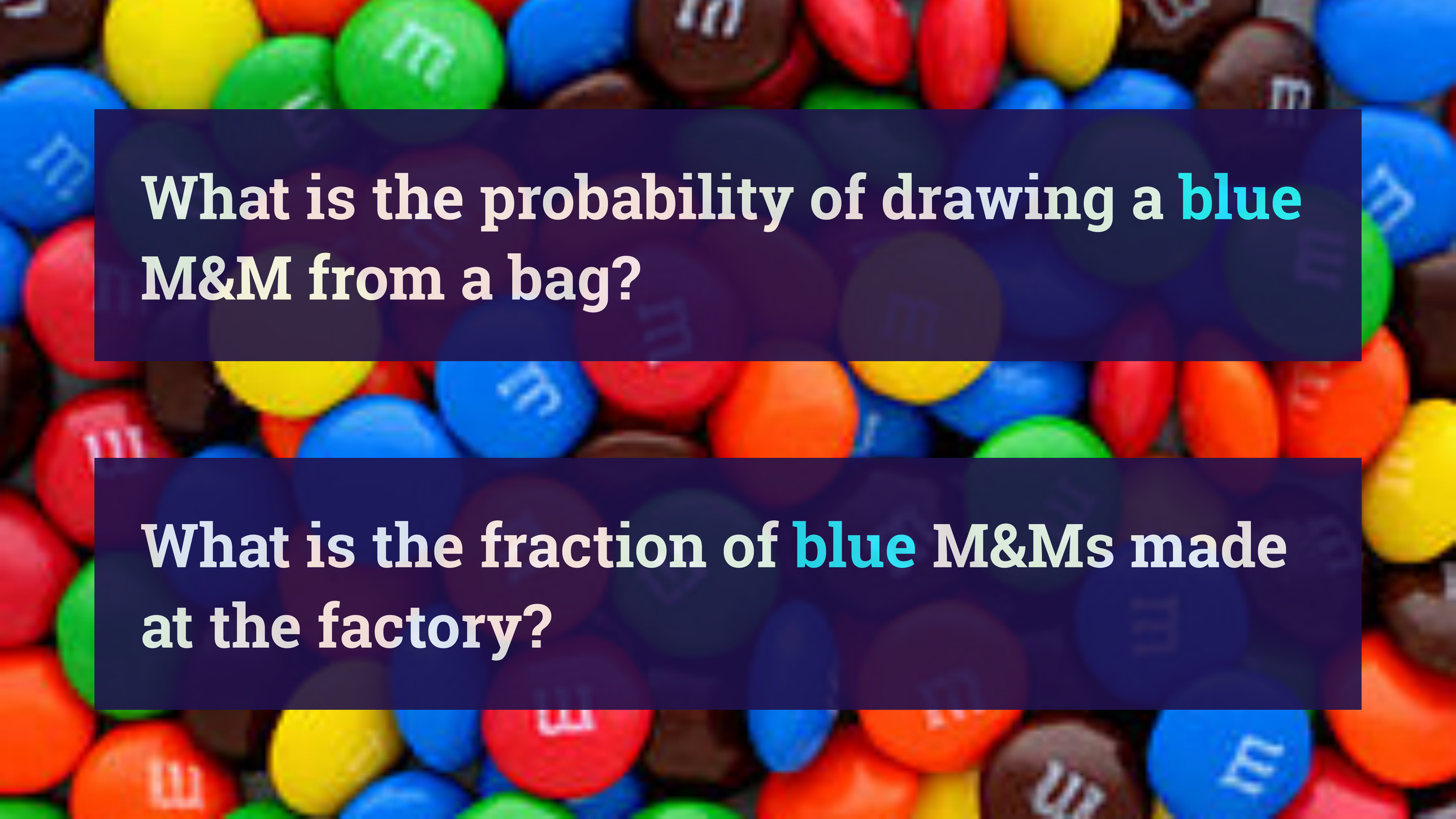
Important: $P(A \mid B) \neq P(B \mid A)$

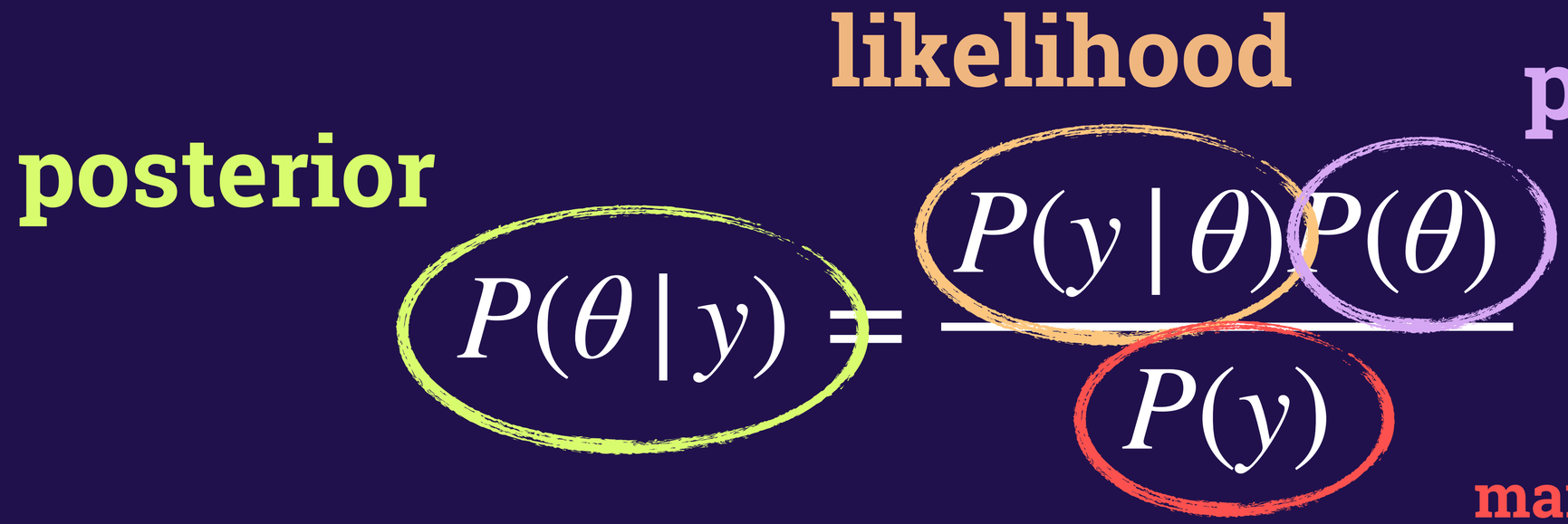
p(☔ | ☁️) ≠ p(☁️ | ☔)

$$P(\theta \mid y) = \frac{P(y \mid \theta) P(\theta)}{P(y)}$$

y = data — black hole X-ray spectrum

θ = model parameters — black body temperature


What is the probability of drawing a blue M&M from a bag?
What is the fraction of blue M&Ms made at the factory?



$$P(\theta \mid y) \propto P(y \mid \theta) P(\theta)$$

posterior

likelihood

prior

$$P(\theta \mid y) \propto P(y \mid \theta) P(\theta)$$

Find the posterior distribution for the percentage of blue m&ms made at the factory

What kind of data are m&ms?

1. nominal
2. continuous

How will you record the data? Table (all colours)

How might we model the probability of drawing a blue m&m?

likelihood

$$P(\theta \mid y) \propto P(y \mid \theta) P(\theta)$$

$y =$ **number of successes (blue m&ms)**

$n - y =$ **number of failures (not-blue m&ms)**

$\theta =$ **probability of drawing a blue m&m**

I have drawn the following sequence of blue (b) and not-blue (t) m&ms:

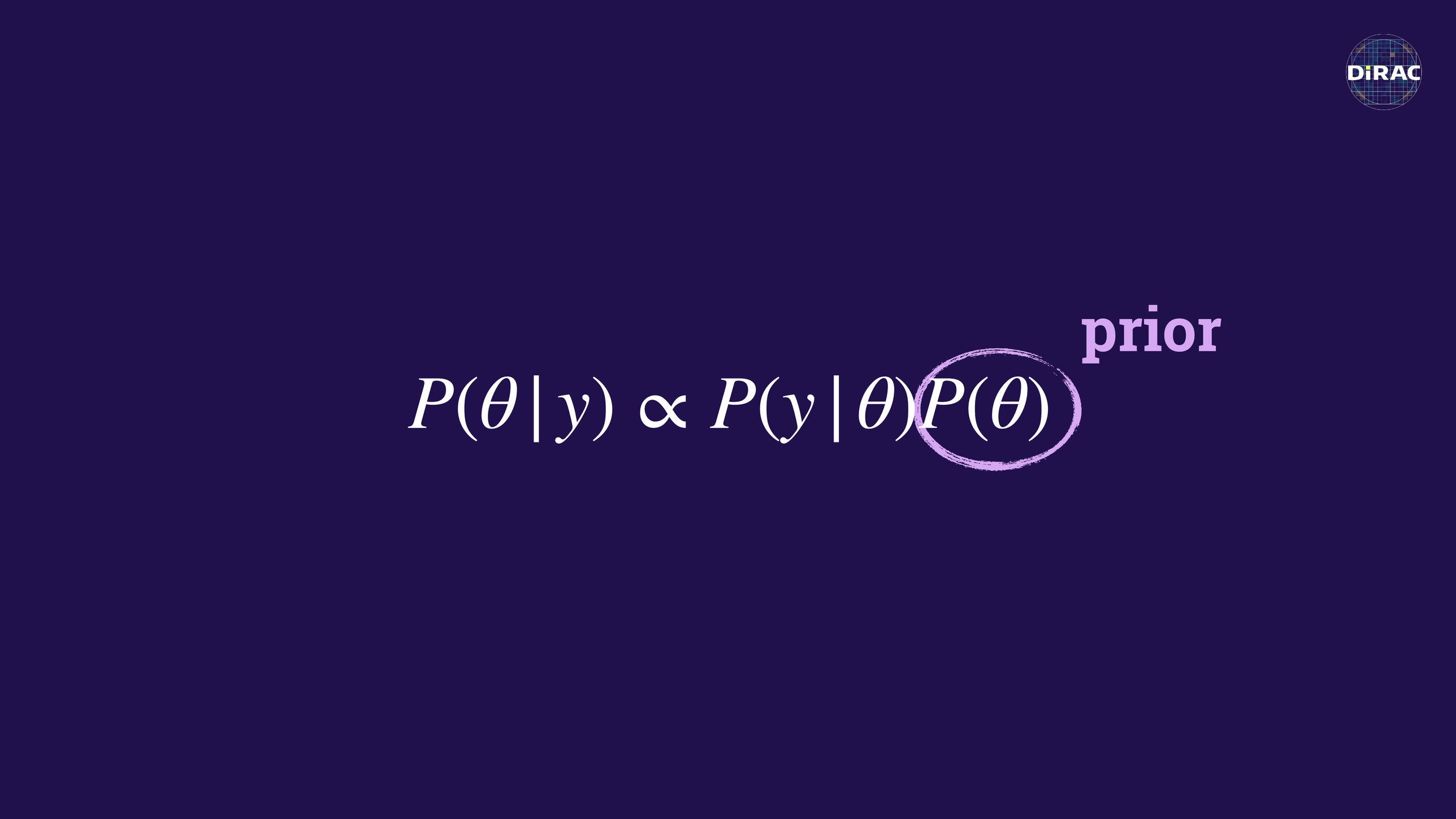
$$S = b\ b\ t\ b\ t\ t\ \ t\ b\ t\ b$$

What's the probability of drawing exactly that sequence?

$$\begin{aligned} p(S) &= \theta\theta(1-\theta)\theta(1-\theta)(1-\theta)(1-\theta)\theta(1-\theta)\theta \\ &= \theta^5(1-\theta)^5 \end{aligned}$$

Binomial distribution

$$p(y \mid \theta) \propto \theta^{y}(1-\theta)^{n-y}$$


prior
$P(\theta|y) \propto P(y|\theta)P(\theta)$

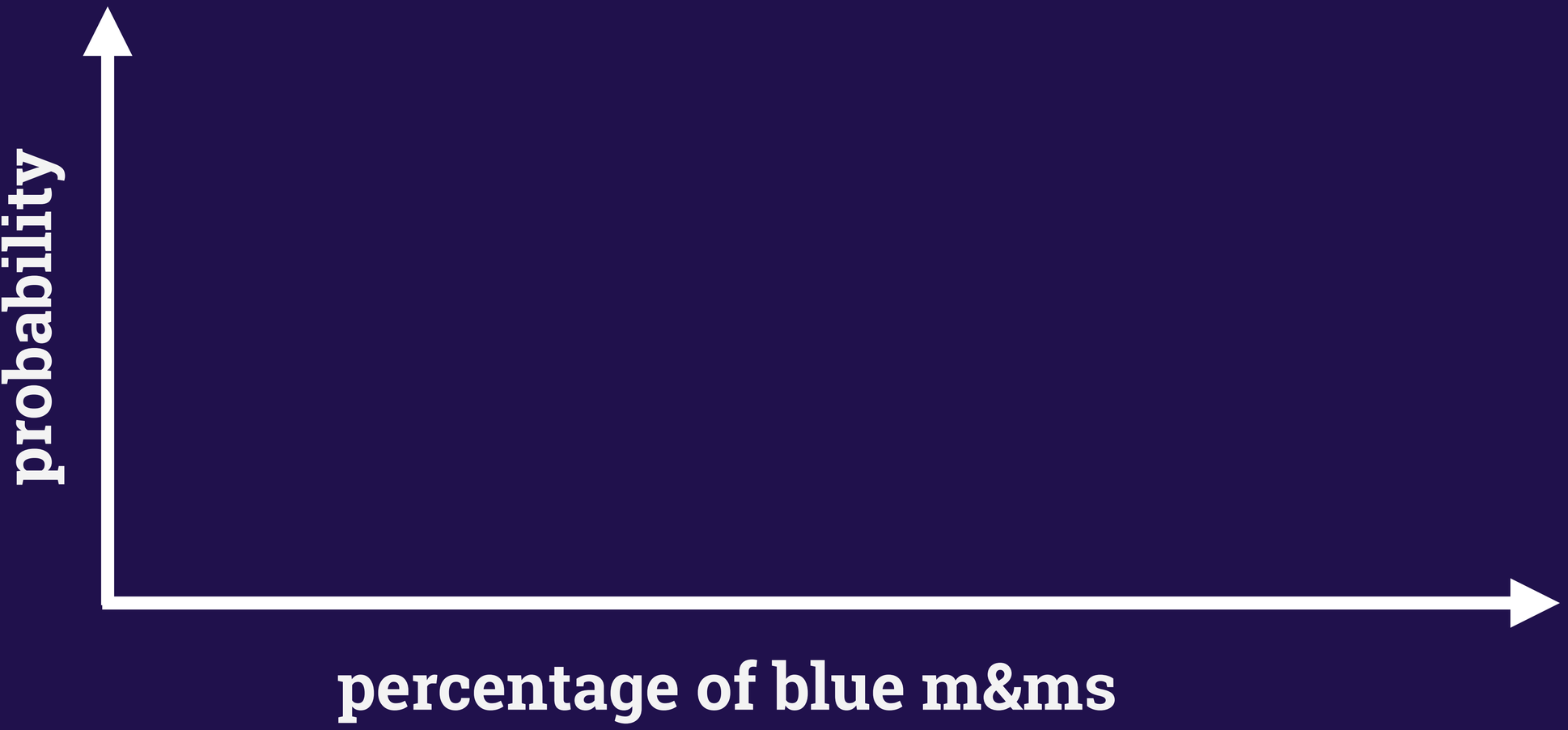
Prior distribution: What do we think we know about the colour distribution of m&ms?

How many different colours of M&Ms are there?

What percentage of blue m&ms are made at the factory?

Do you think every bag of m&ms will have the same percentage of blue m&ms?

Sketch your prior



prior

$$P(\theta \mid y) \propto P(y \mid \theta) P(\theta)$$

beta distribution

$$P(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

“conjugate prior” $\alpha = ?$ $\beta = ?$

Write a function for the beta distribution and try out different values for alpha and beta

Some values to try:

$$\alpha = \beta$$

$$\alpha = \beta = 1$$

$$\alpha, \beta \quad \text{very large}$$

$$\alpha, \beta \quad \text{very small}$$

likelihood

$$p(y) \propto \theta^{y}(1-\theta)^{n-y}$$

prior

$$P(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$P(\theta \mid y) \propto P(y \mid \theta)P(\theta)$$

$$p(\theta) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

Gather some data!

take the first 20 m&ms out of your bag (n=20)

record the counts for all colours

Use your function for the beta distribution to plot both the prior and the posterior in the same plot

- Is the posterior distribution what you expected?
- Compare the posterior distribution to the prior distribution: Is this the result you expected, given six different colours?
- How sensitive is the posterior to the prior?

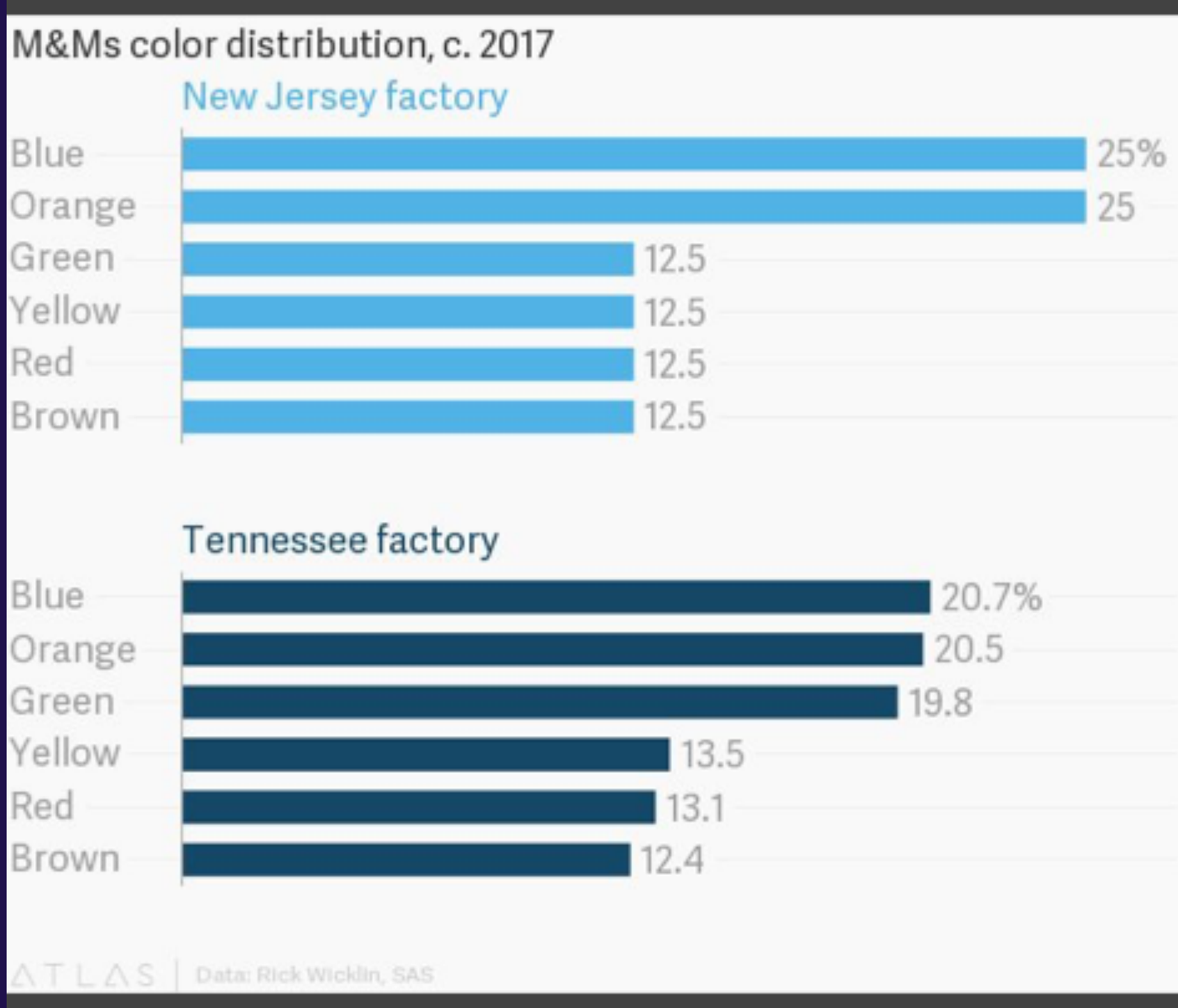
How would you expect the posterior to change with more data?

Let’s pool all the data and find out!

https://forms.gle/zHy8LSbJsB8Pm9XA9

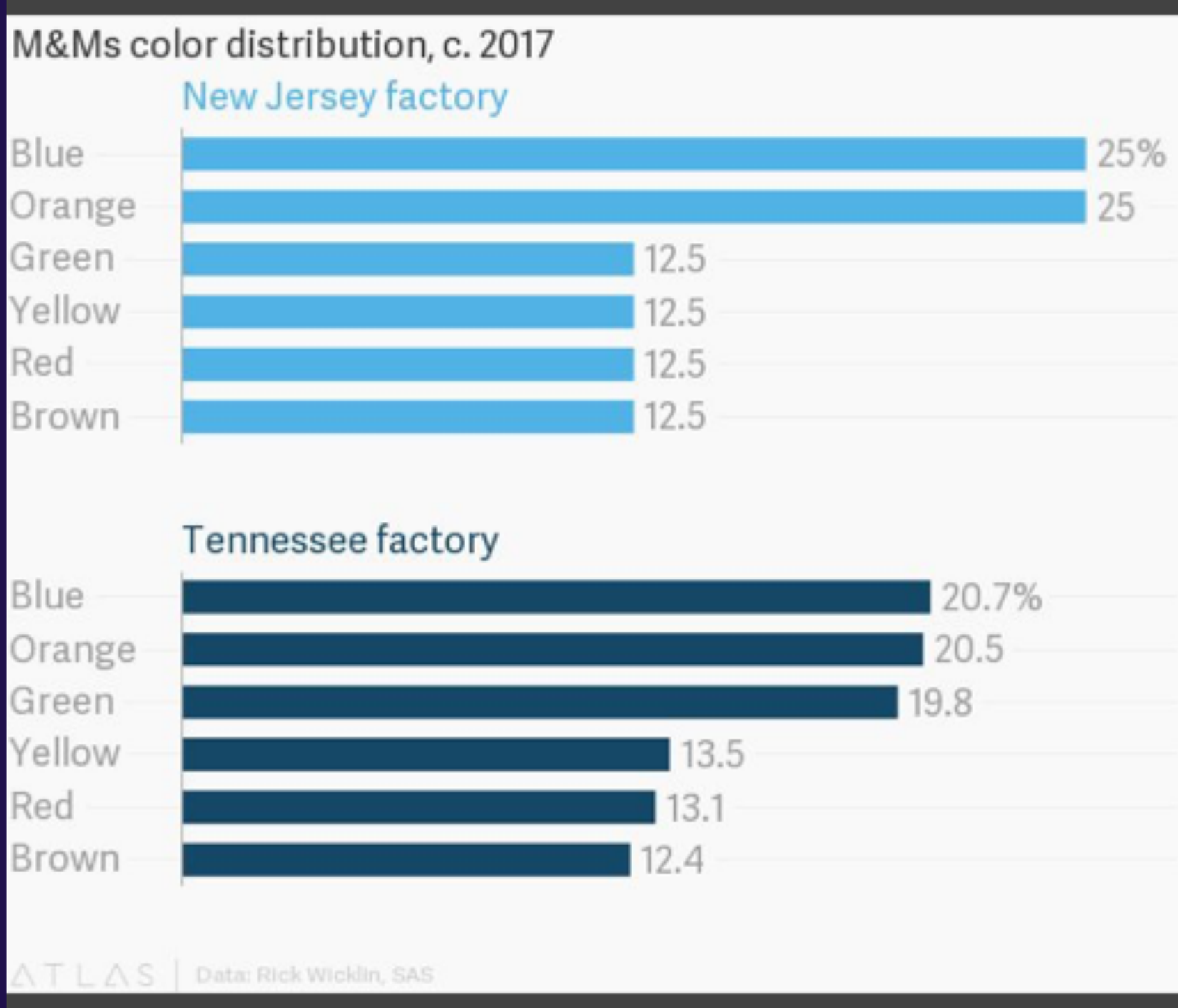
Surprise Twist!

Different factories make different distributions of m&ms!



Surprise Twist!

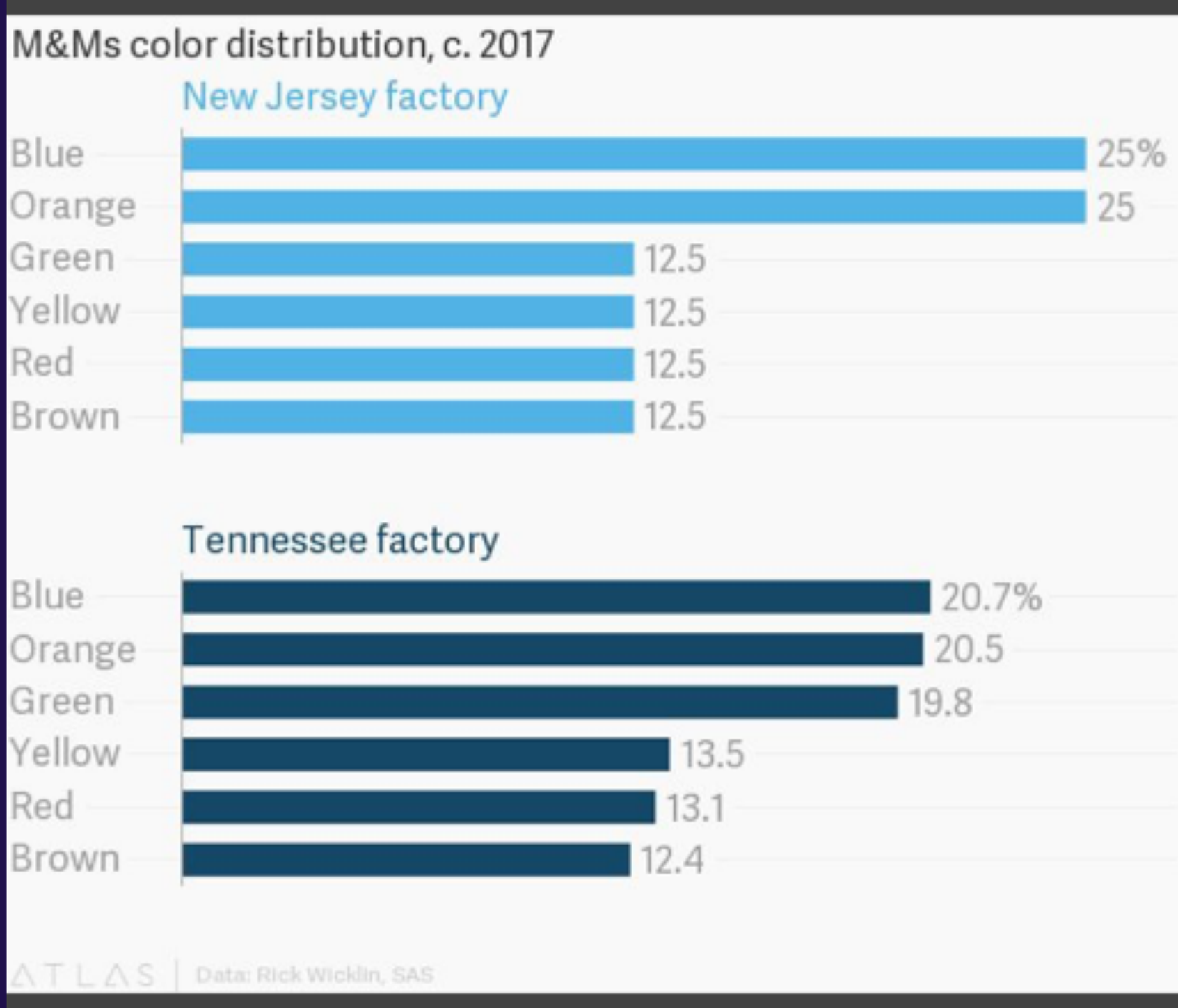
Which factory did your m&ms come from?



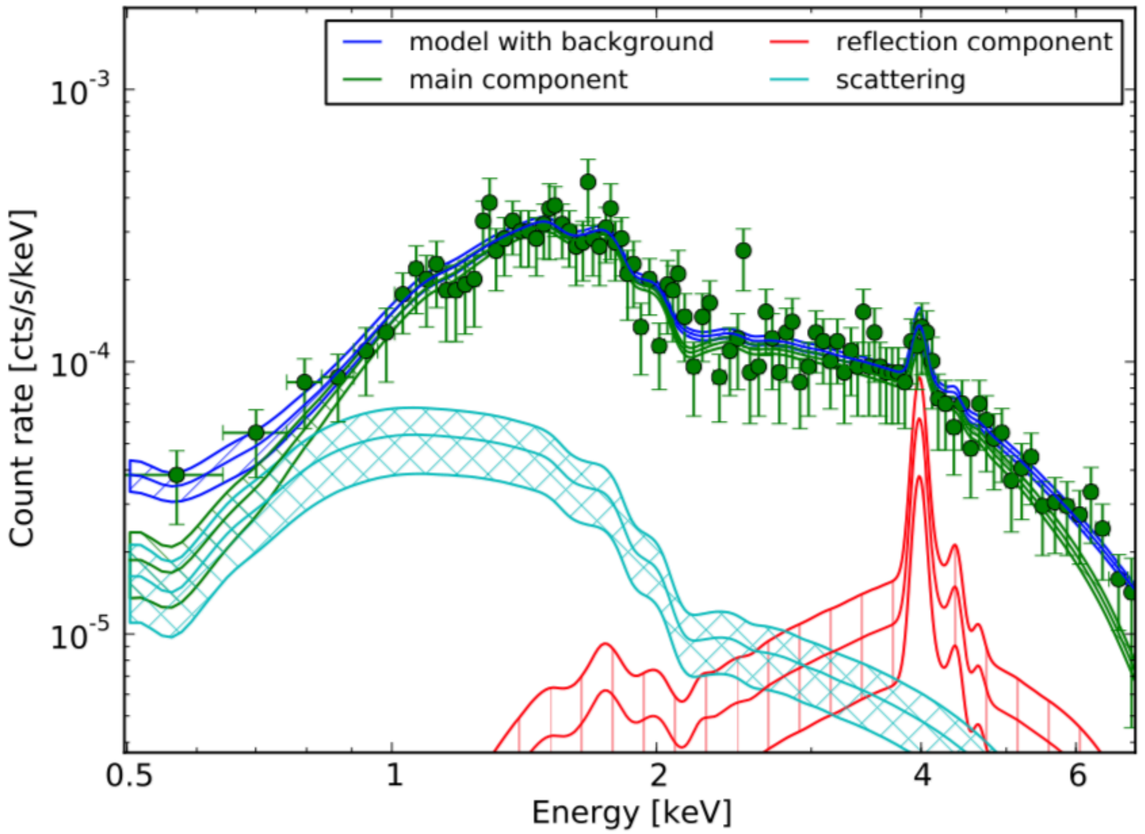
Surprise Twist!

New Jersey = HKL

Tennessee = CLV



What about astronomy?



Data: X-ray spectrum

Model: absorption + scattering + reflection

Likelihood: Poisson

Buchner et al (2014)

Ethical considerations in statistics

data and algorithms are social constructs

... are created by humans

... encode biases

... can be (mis-)used to serve an agenda

Based on the M&Ms exercise, can you think of ways our statistical procedure could be mis-used or be made to be misleading?

Opinion: Is science really facing a reproducibility crisis, and do we need it to?

Daniele Fanelli

PNAS March 13, 2018 115 (11) 2628-2631; first published March 12, 2018
https://doi.org/10.1073/pnas.1708272114

Edited by David B. Allison, Indiana University Bloomington, Bloomington, IN, and accepted by Editorial Board Member Susan T. Fiske November 3, 2017 (received for review June 30, 2017)

Article | Figures & SI | Info & Metrics | PDF

Abstract

Efforts to improve the reproducibility and integrity of science are typically justified by a narrative of crisis, according to which most published results are unreliable due to growing problems with research and publication practices. This article provides an overview of recent evidence suggesting that this narrative is mistaken, and argues that a narrative of epochal changes and empowerment of scientists would be more accurate, inspiring, and compelling.

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Aut

Archive > Volume 519 > Issue 7541 > Research Highlights: Social Selection > Article

NATURE | RESEARCH HIGHLIGHTS: SOCIAL SELECTION

Psychology journal bans *P* values

Test for reliability of results 'too easy to pass', say editors.

Chris Woolston

26 February 2015 | Clarified: 09 March 2015

PDF | Rights & Permissions

A controversial statistical test has finally met its end, at least in one journal. Earlier this month, the editors of *Basic and Applied Social Psychology* (*BASP*) announced that the journal would no longer publish papers containing *P* values because the statistics were too often used to support lower-quality research [1].

Most mistakes I see made in statistical procedures are either **mis-applications** of methods or **mis-interpretation** of the results

Know your assumptions!