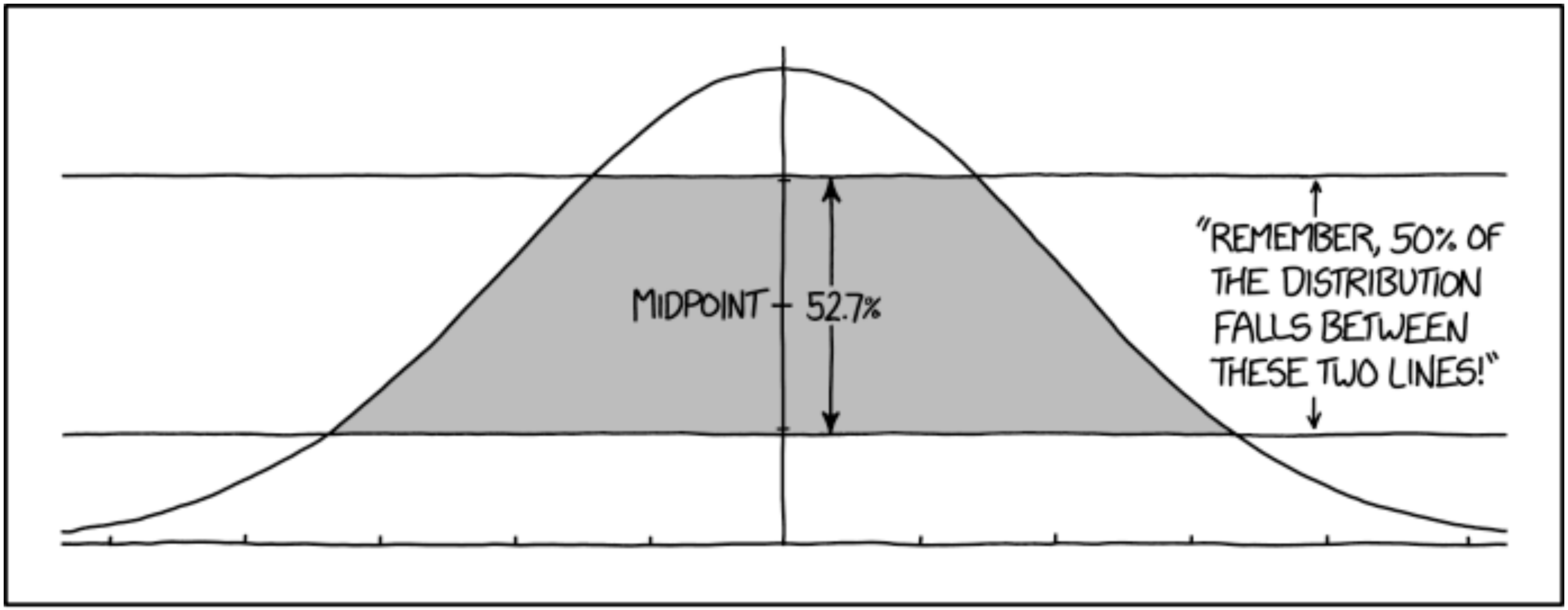
Bayesian data analysis 2



HOW TO ANNOY A STATISTICIAN

03/02/2020

Logistics

Final projects

Project proposals

Project presentations

When/how will you present? *

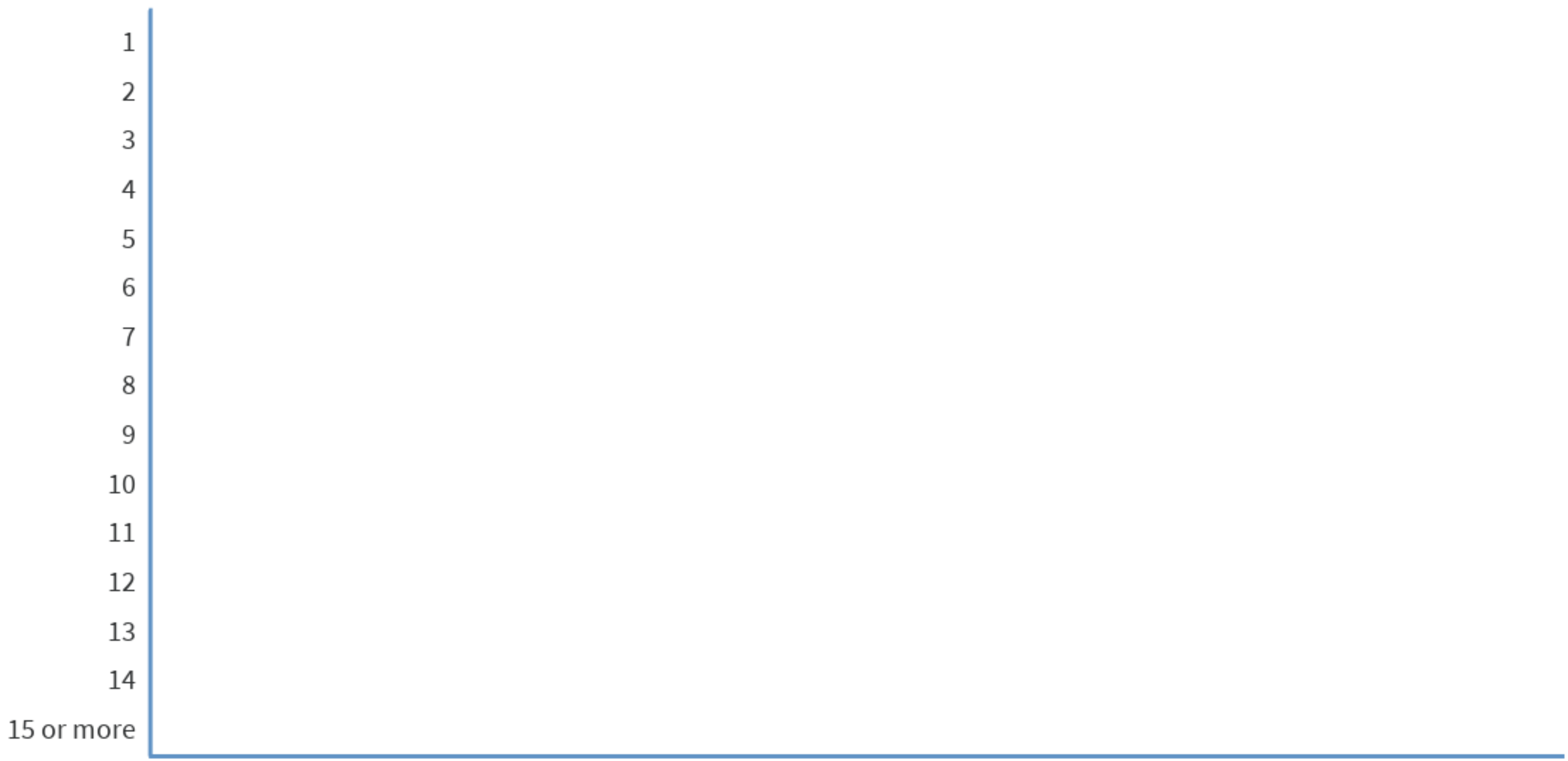
- On March 18th (Final presentations day)
- On March 13th (Final class)
- On March 16th at 3pm (Stats instructors meeting)
- I will record the presentation and submit a video.

survey link

https://tinyurl.com/psych252presentation

Homework 5

How many hours did it take you to complete Homework 5?



RStudio®

time

Plan for today

- Quick Bayes recap
- Ingredients: likelihood, prior, inference
- Doing Bayesian data analysis
- Bayesian models of cognition

Plan for today

- **Quick Bayes recap**
- Ingredients: likelihood, prior, inference
- Doing Bayesian data analysis
- Bayesian models of cognition

Quick Bayes recap

Quick Bayes recap

Frequentist statistics

- generate a sampling distribution of the test statistic assuming H_0
- compare observed value of the test statistic with the sampling distribution
- reject the H_0 if probability of observed value (or more extreme values) is less than α

Bayesian statistics

- directly test hypotheses of interest
- define prior over hypotheses **p(H)**
- compute likelihood of the data for each hypothesis **p(D|H)**
- use Bayes' rule to infer the posterior over hypotheses given the data **p(H|D)**

Quick Bayes recap

Bayesian Recipe

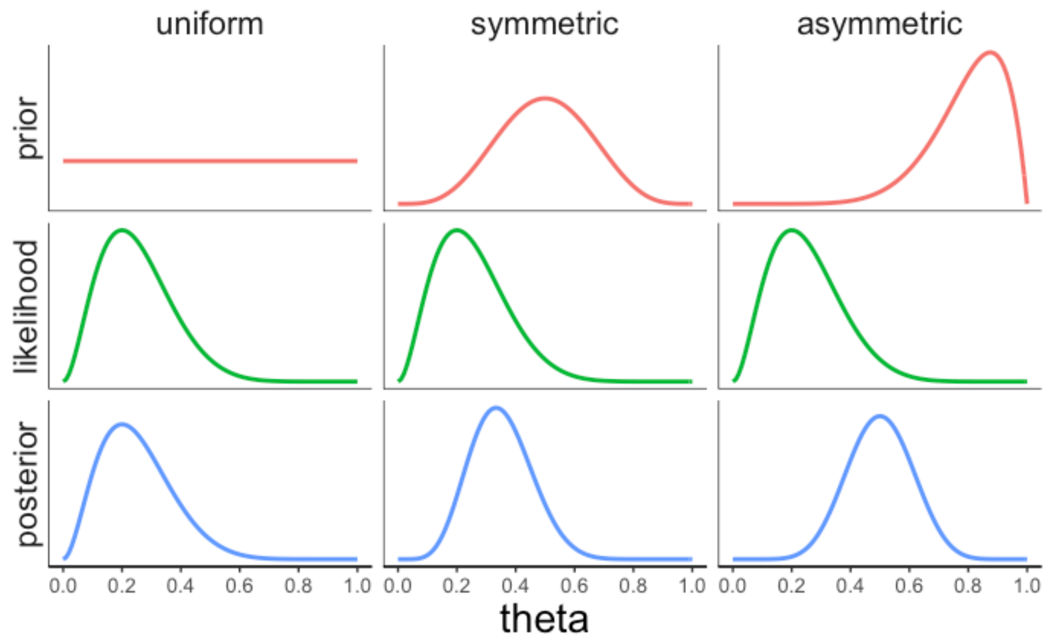
- Hypotheses
- Prior over hypotheses
- Data
- Likelihood of the data given the hypotheses
- Posterior over hypotheses given the data

+ a healthy dose of Bayes' rule

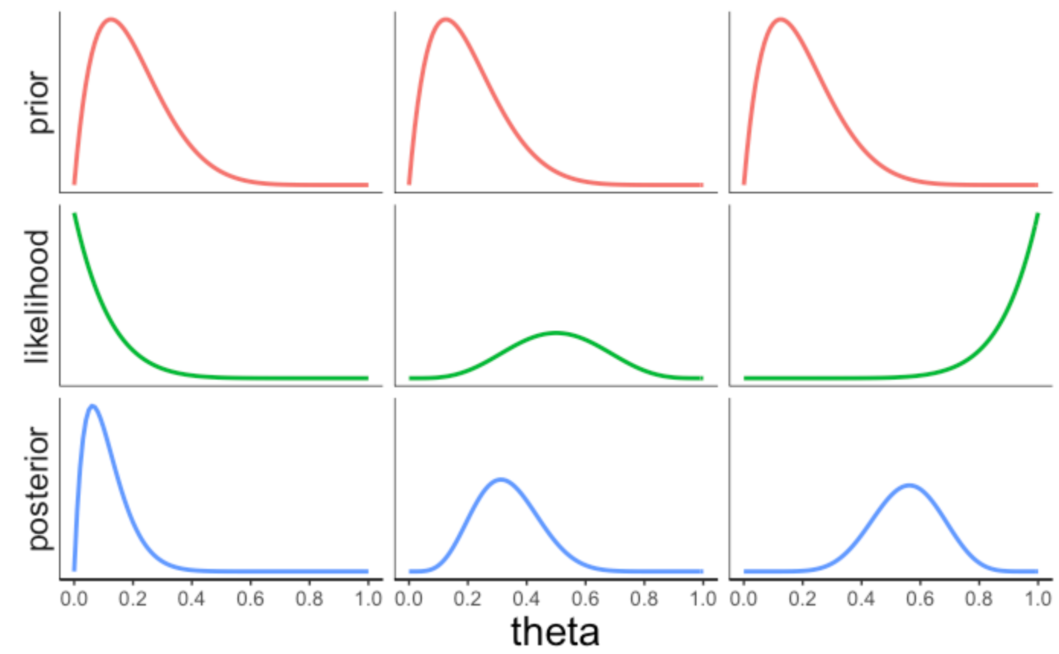
$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{p(D)}$$

Quick Bayes recap

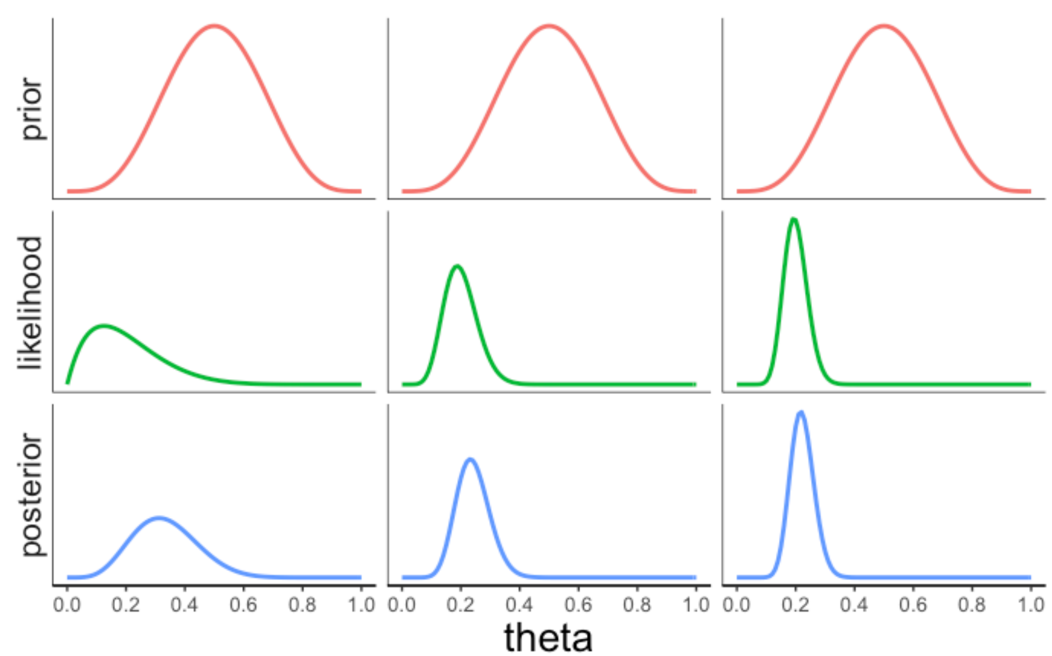
Effect of the **prior**



Effect of the **likelihood**



Effect of the **sample size**



Plan for today

- Quick Bayes recap
- **Ingredients: likelihood, prior, inference**
- Doing Bayesian data analysis
- Bayesian models of cognition

Ingredients: likelihood, prior, inference

Ingredients

Likelihood Prior

Posterior

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

Normalizing constant

Likelihood Prior

Posterior

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

Normalizing constant

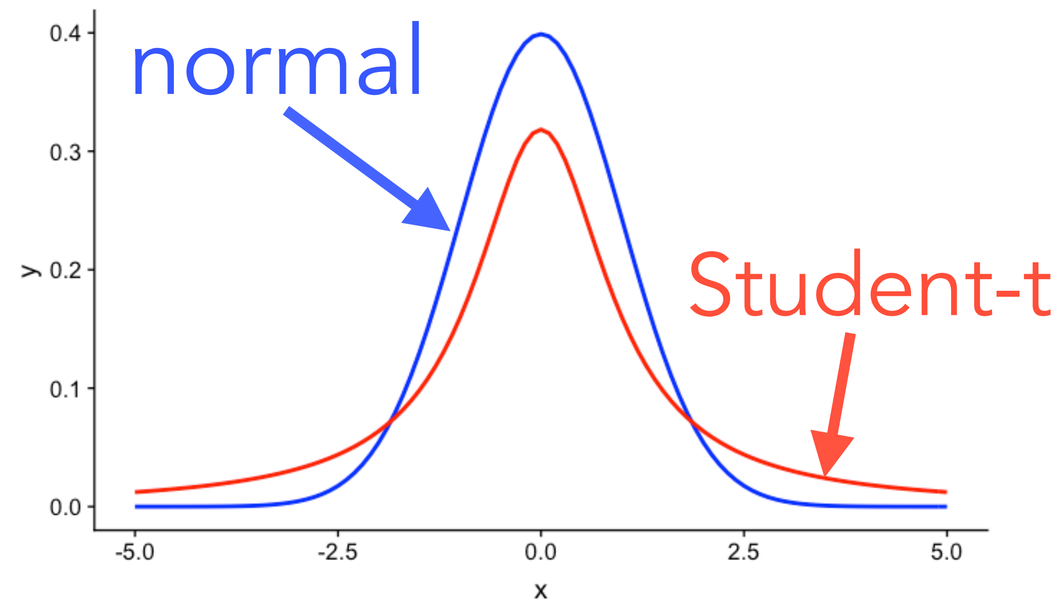
Likelihood

- **What probabilistic model describes best how the data were generated?**
- How to build a (Bayesian) model?
  - What real-life behavior should the model explain?
  - What assumptions can you make about the behavior?
    - What's the nature of your dependent variable (e.g. binary, ordered, continuous)?
  - Does the model re-create the behavior of interest?

Likelihood

- **Bernoulli**:
  - binary data
  - a single trial
- **Binomial**:
  - binary data
  - fixed number of total trials
  - trial outcomes are independent
  - probability of success is the same in each trial
- **Poisson**: count of discrete events
- **Beta-binomial**: like binomial but probability of success may change across trials
- ...

Likelihood



- **Normal**:
  - continuous data
  - unbounded outcomes
  - outcome is the result of a large number of additive factors
- **Student-*t***:
  - same as Normal
  - handles greater variability in the data (distribution has **fat tails**)

Prior

Likelihood Prior

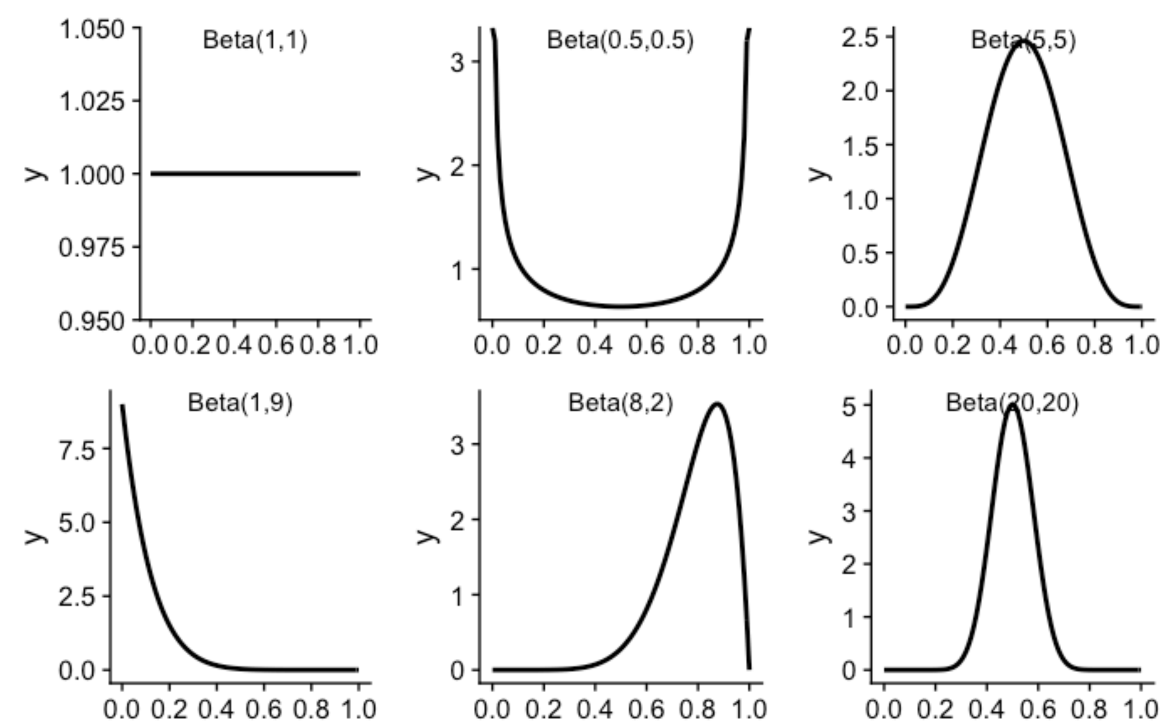
Posterior

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

Normalizing constant

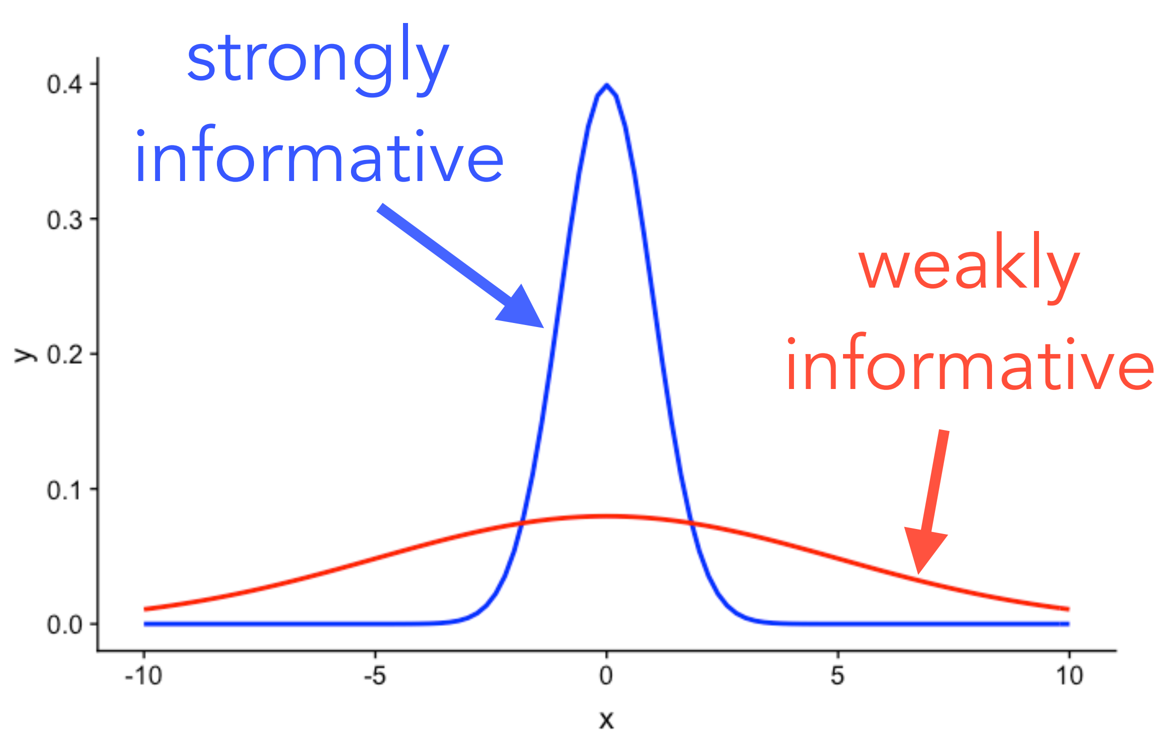
Prior

- **uniform**:
  - continuous or discrete
  - bounded between minimum and maximum
- **beta:**
  - continuous parameters
  - bounded between 0 and 1
  - can model a wide range of priors



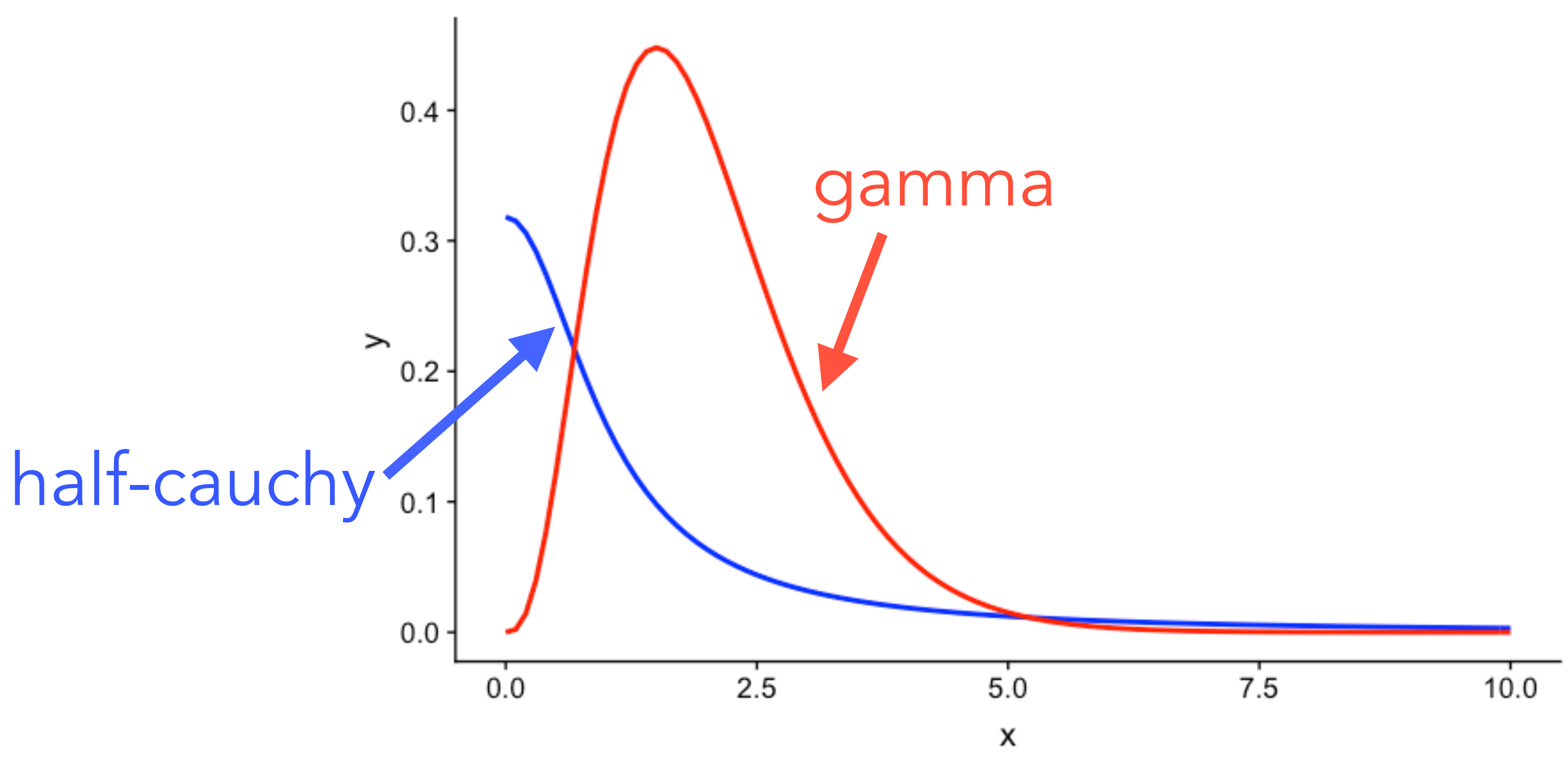
Prior

- **normal**:
  - continuous
  - unbounded outcomes
  - can range from weakly informative to strongly informative



Prior

- What prior should we use for inferring the standard deviation?
  - **uniform** (positive)
    - but: large values might be less plausible a priori than smaller values
  - **cauchy** (truncated)
  - **gamma**



Inference

Likelihood Prior

Posterior

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

Normalizing constant

the devil is in the denominator ...

Doing Bayesian inference

Discrete hypothesis space

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

sum over all possibilities

Continuous hypothesis space

$$p(H \mid D) = \frac{p(D \mid H) \cdot p(H)}{\int_{-\infty}^{\infty} p(D \mid H_i) \cdot p(H_i) dH_i}$$

integral over all possibilities

Discretizing the parameters

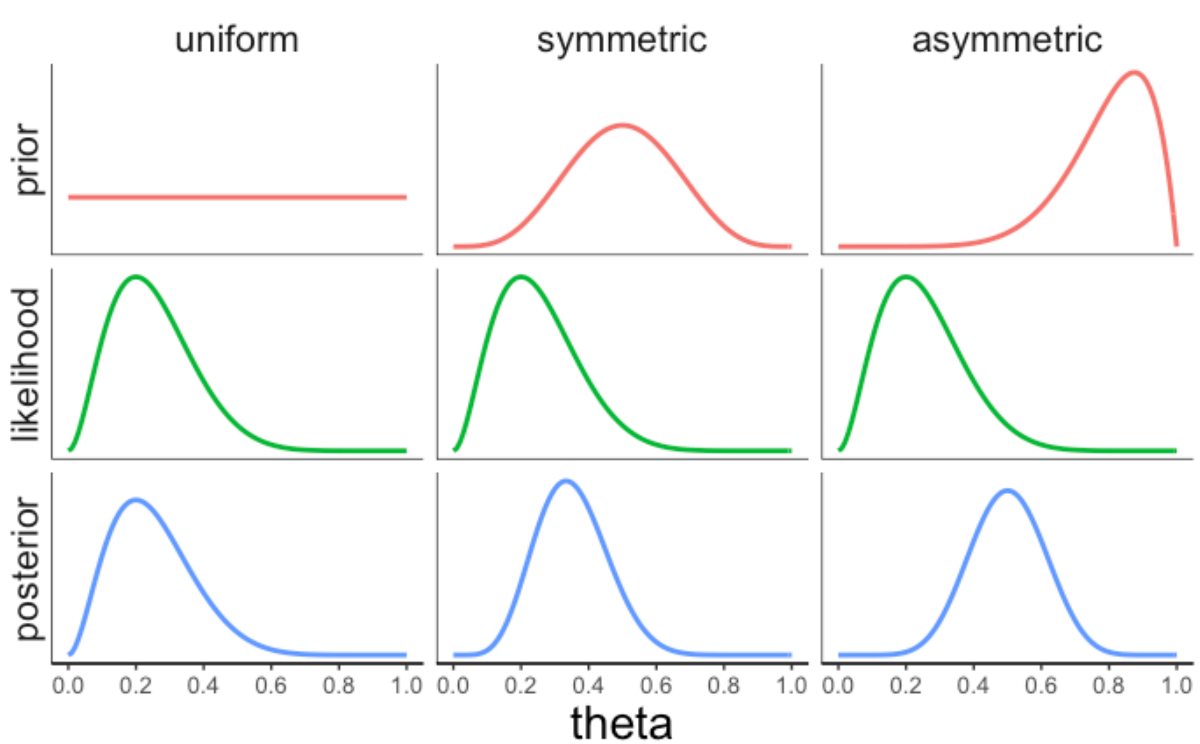
```
# grid
theta = seq(0, 1, 0.01)

# data
data = rep(0:1, c(8, 2))

# calculate posterior
df.prior = tibble(theta = theta,
                  prior_uniform = dbeta(grid, shape1 = 1, shape2 = 1),
                  prior_normal = dbeta(grid, shape1 = 5, shape2 = 5),
                  prior_biased = dbeta(grid, shape1 = 8, shape2 = 2)) %>%
  pivot_longer(cols = -theta,
               names_to = "prior_index",
               values_to = "prior") %>%
  mutate(likelihood = dbinom(sum(data == 1),
                             size = length(data),
                             prob = theta)) %>%
  group_by(prior_index) %>%
  mutate(posterior = likelihood * prior / sum(likelihood * prior)) %>%
  ungroup() %>%
  pivot_longer(cols = -c(theta, prior_index),
               names_to = "index",
               values_to = "value")
```

100 discrete values

for 3 variables, we would already need 1 Mio combinations



The CURSE of dimensionality

Inference via sampling

- we cannot directly calculate the probability of the posterior (because it might have a pretty weird shape)
- **but**: we can draw random samples from the posterior
- we can then use our data wrangling and visualization skills to make inferences based on these samples

Inference via sampling

- imagine that we don't have a **`dnorm`**`()` function in R but we want to compute probabilities
- luckily, we do have the **`rnorm`**`()` function, so we can create random samples

recent work on Bayesian inference has developed efficient algorithms for generating independent samples from the posterior distribution

Inference via sampling

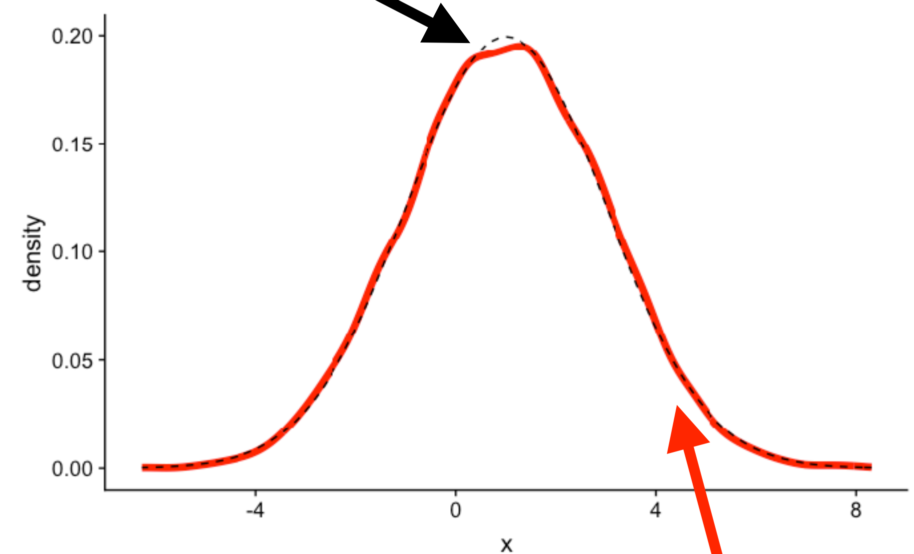
```r
# generate samples
df.samples = tibble(x = rnorm(n = 10000, mean = 1, sd = 2))

# visualize distribution
ggplot(data = df.samples,
       mapping = aes(x = x)) +
  stat_density(geom = "line",
               color = "red",
               size = 2) +
  stat_function(fun = "dnorm",
                args = list(mean = 1, sd = 2),
                color = "black",
                linetype = 2)

# calculate probability based on samples
df.samples %>%
  summarize(prob = sum(x >= 0 & x < 4)/n())

# calculate probability based on theoretical distribution
pnorm(4, mean = 1, sd = 2) - pnorm(0, mean = 1, sd = 2)
```

theoretical distribution



distribution based on samples

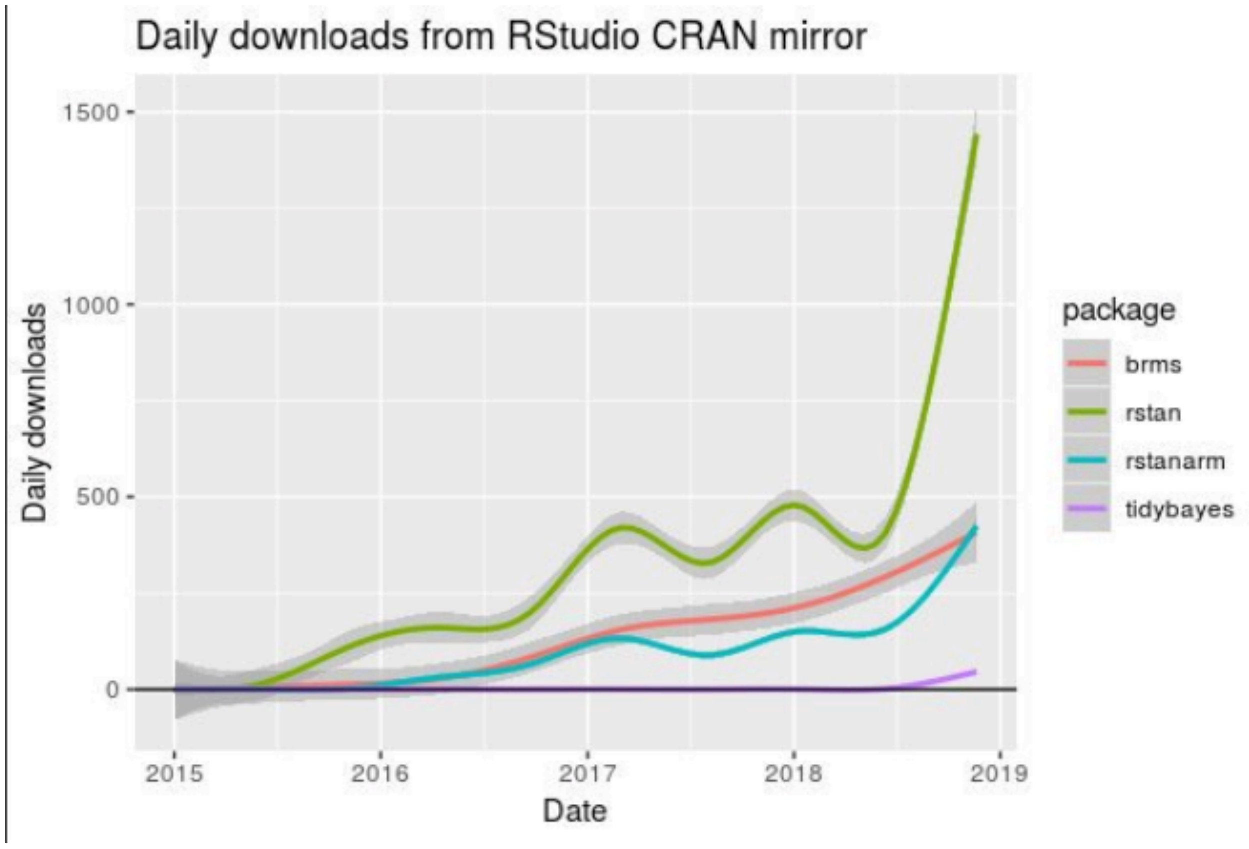
x
0.07
0.10
-0.56
-0.26
1.59
0.37
0.01
3.68
0.87
1.24

⋮

both methods yield ≅63%

Inference via sampling

- Bayesian data analysis is becoming more popular because:
  - computers are getting more powerful
  - inference techniques are getting better
  - software packages become easier to use



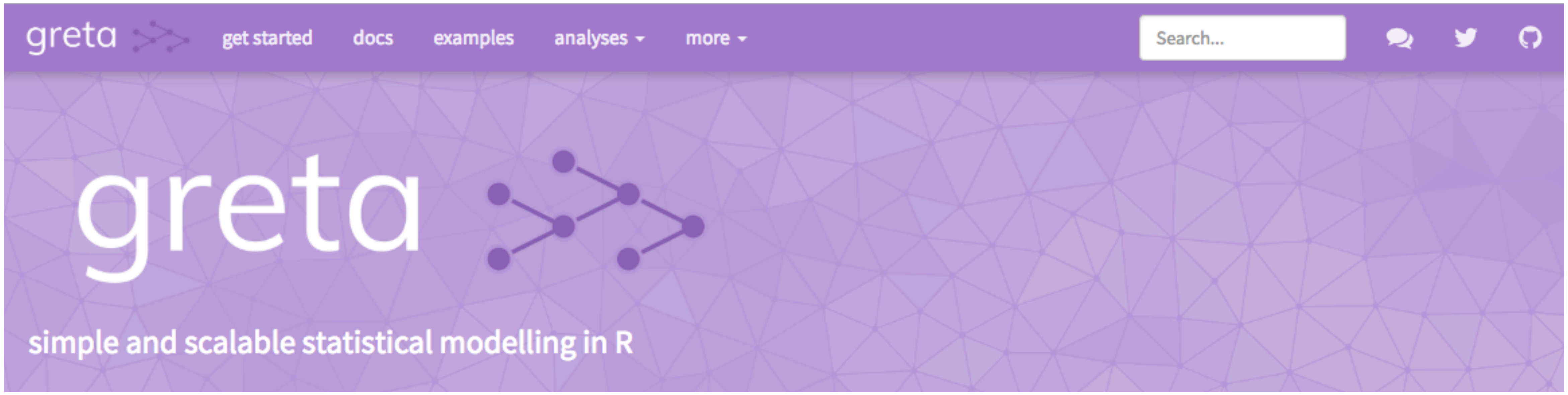
Plan for today

- Quick Bayes recap
- Ingredients: likelihood, prior, inference
- **Doing Bayesian data analysis**
- Bayesian models of cognition

Doing Bayesian data analysis

Software packages

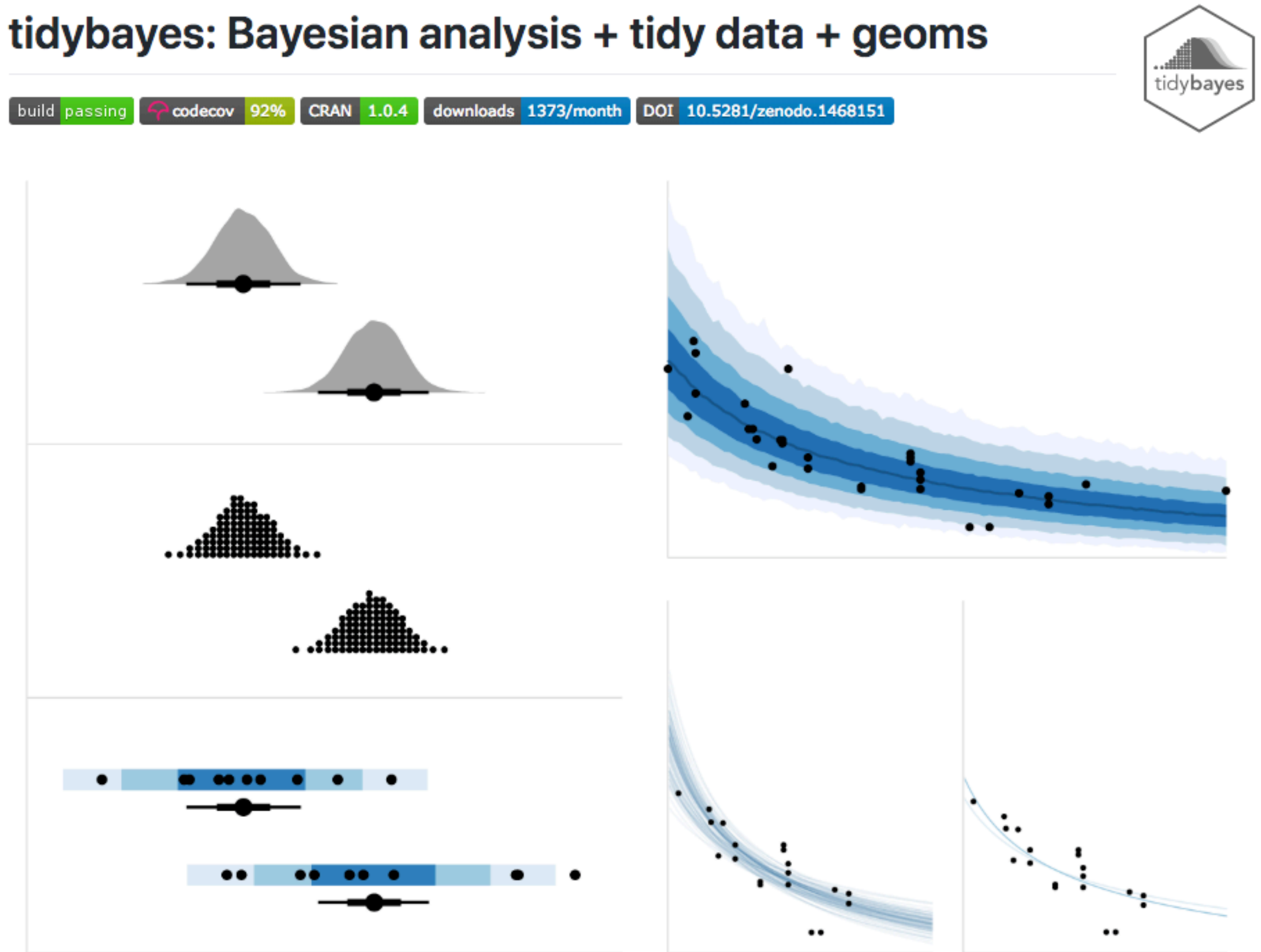
`library("greta")`



- let's us write Bayesian models directly in R with a simple syntax
- uses Tensorflow to implement Hamiltonian Monte Carlo sampling (a fast inference algorithm ...)

Software packages

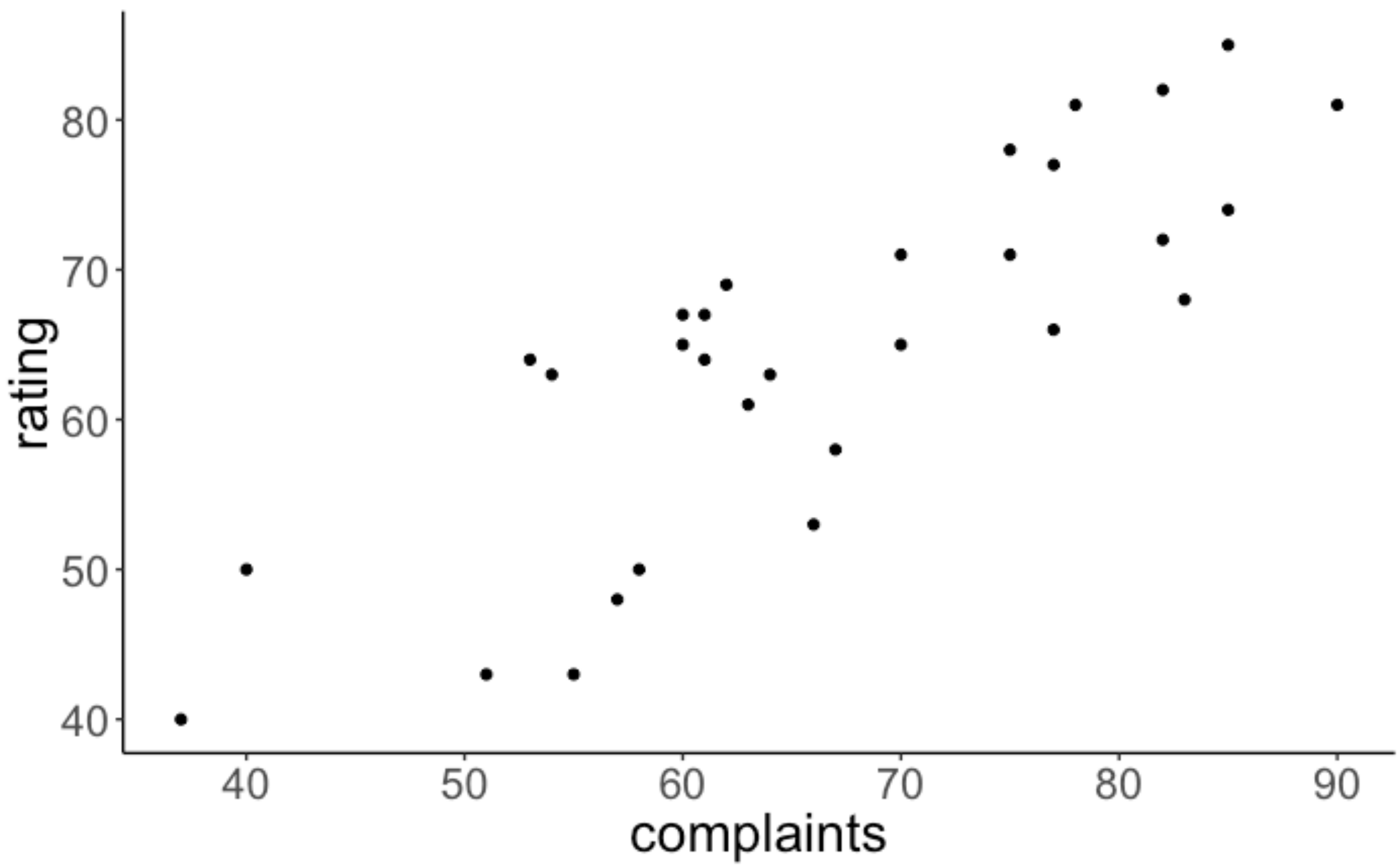
`library("tidybayes")`



- great tool for wrangling and visualizing the results of Bayesian data analysis

Attitude data set

What's the relationship between how well an employee handles complaints and their overall performance rating?



Frequentist analysis

Frequentist analysis

```r
# fit model
fit = lm(formula = rating ~ 1 + complaints,
         data = df.attitude)

# print summary
fit %>% summary()
```

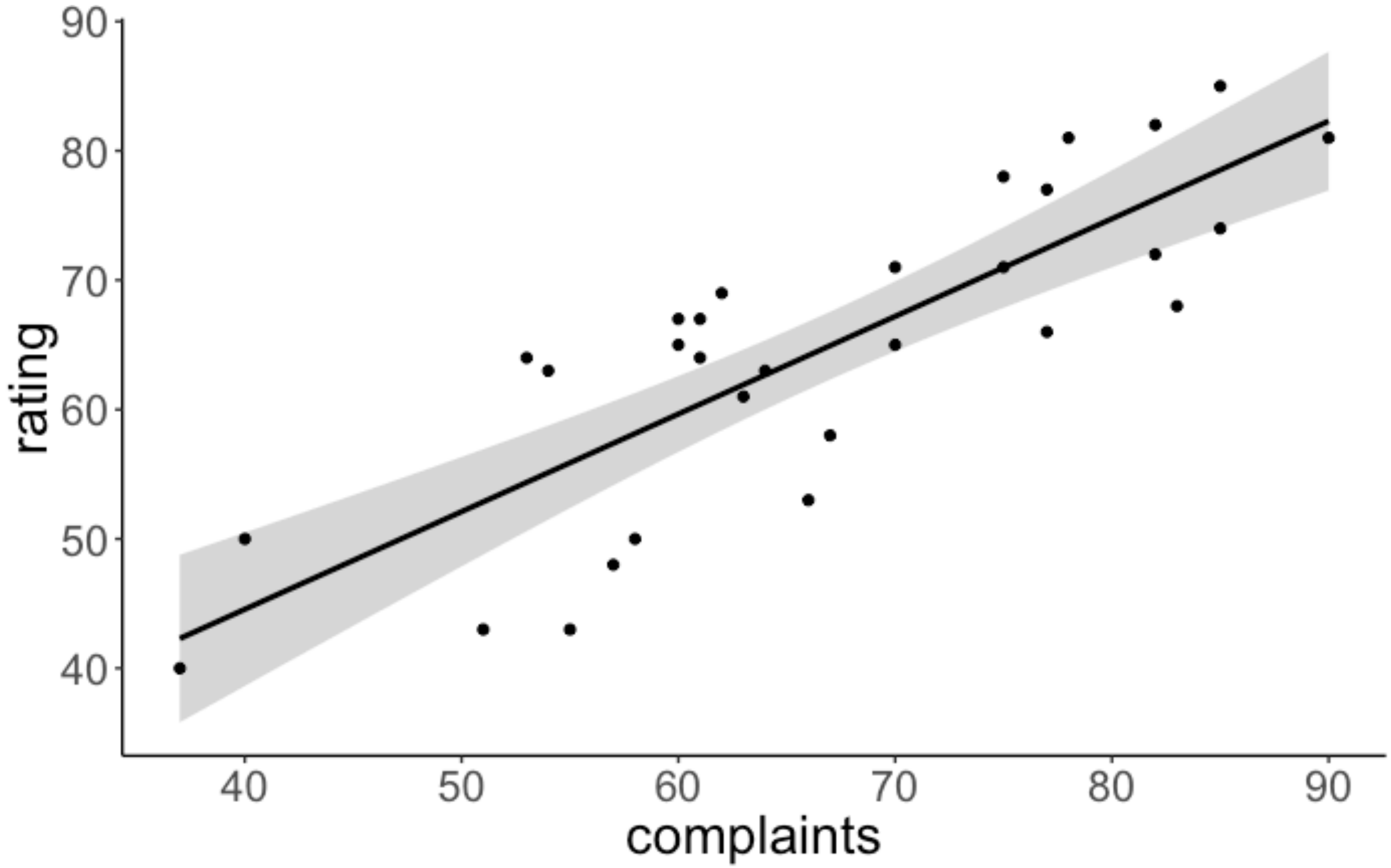
```
Call:
lm(formula = rating ~ 1 + complaints, data = df.attitude)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.8799  -5.9905   0.1783   6.2978   9.6294 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 14.37632    6.61999   2.172   0.0385 *  
complaints   0.75461    0.09753   7.737 1.99e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.993 on 28 degrees of freedom
Multiple R-squared:  0.6813,	Adjusted R-squared:  0.6699 
F-statistic: 59.86 on 1 and 28 DF,  p-value: 1.988e-08
```

Visualize model predictions



Best-fitting regression line with confidence interval

Bayesian analysis

Model specification

```r
library("greta")
library("tidybayes")

# variables & priors
b0 = normal(0, 10)
b1 = normal(0, 10)
sd = cauchy(0, 3, truncation = c(0, Inf))

# linear predictor
mu = b0 + b1 * attitude$complaints

# observation model (likelihood)
distribution(attitude$rating) = normal(mu, sd)

# define the model
m = model(b0, b1, sd)
```

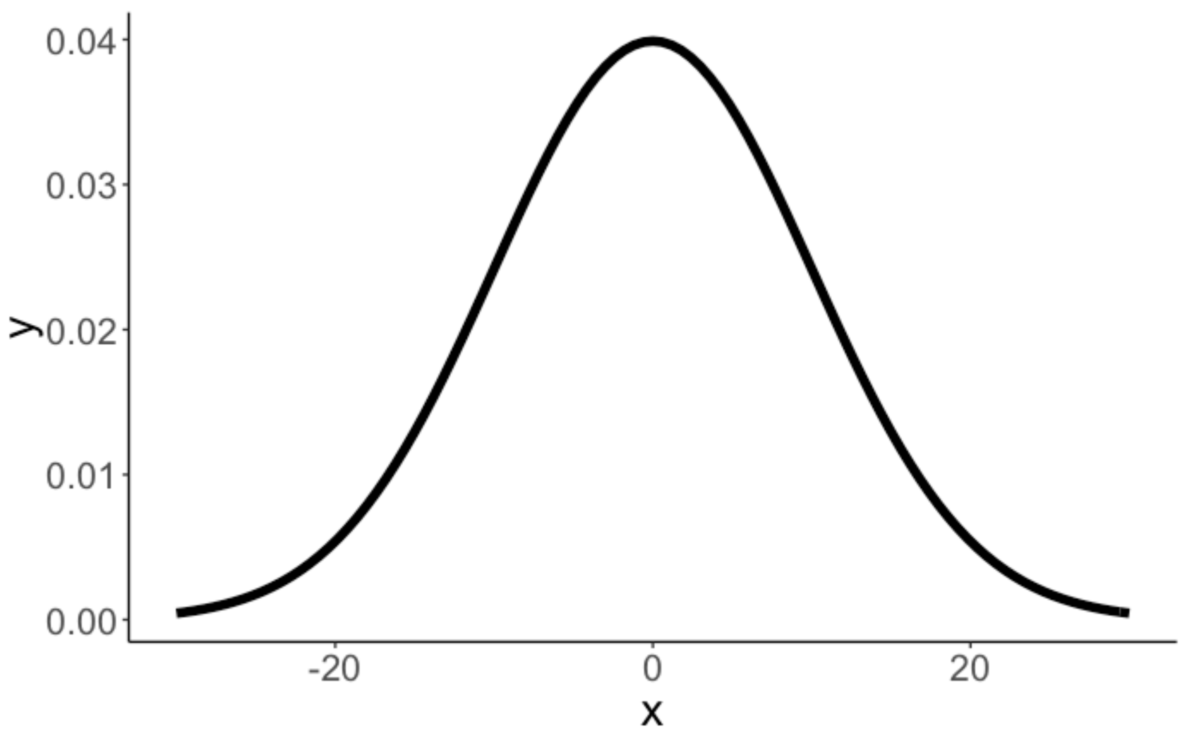
priors

linear combination

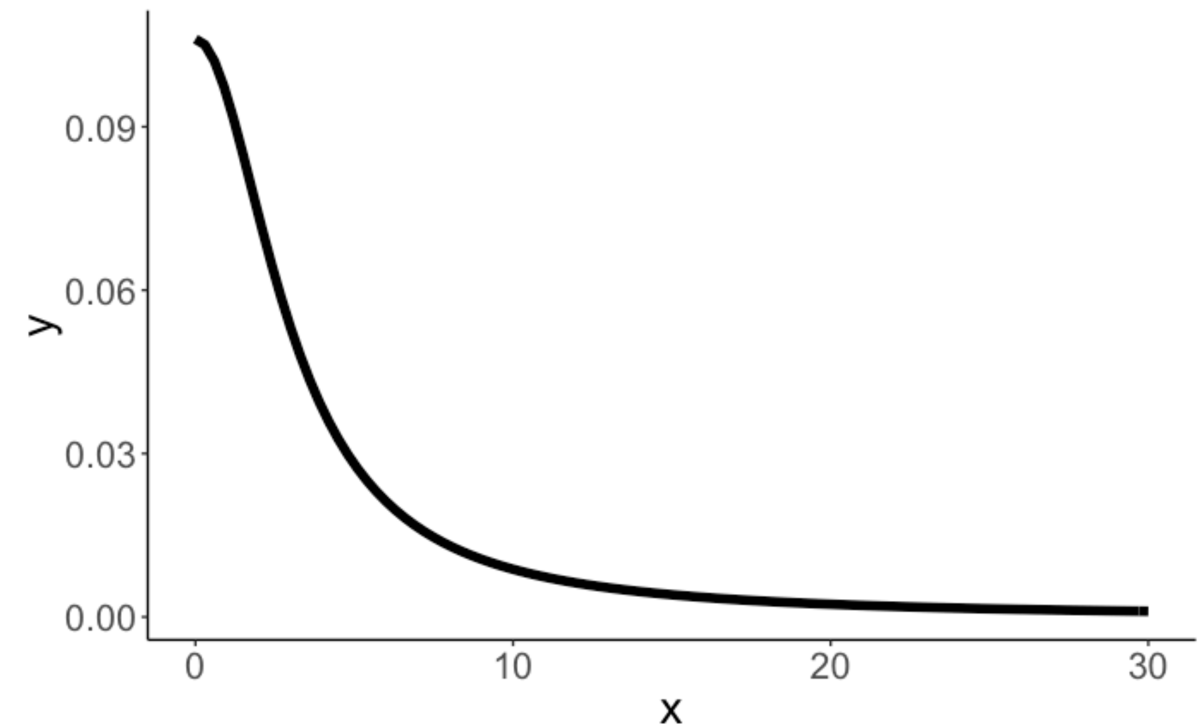
Gaussian likelihood

build the model

Priors



Gaussian prior on intercept and coefficient

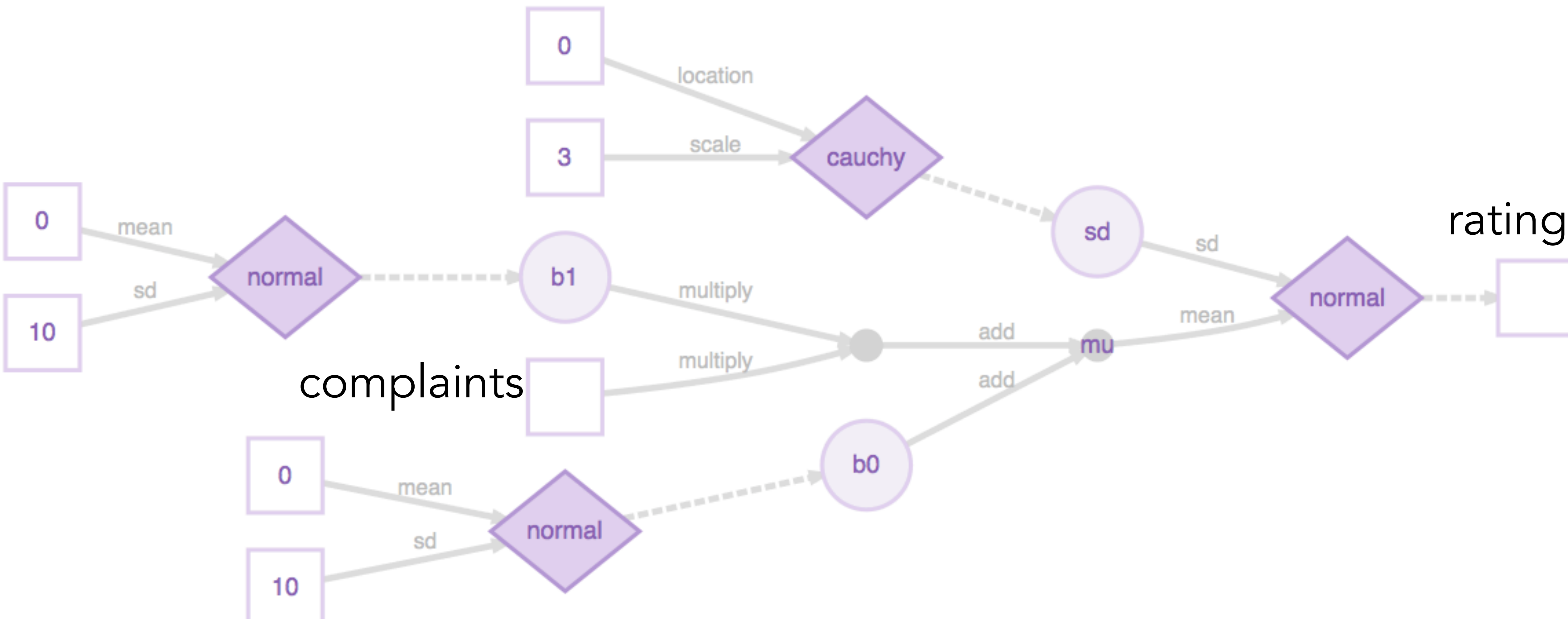


Truncated Cauchy prior on the standard deviation

weakly informative priors (allow for a wide range of possible values)

Graphical representation of the model

```r
# plotting
plot(m)
```



the model describes how the data are generated

Inference via sampling

Markov Chain
Monte Carlo
inference

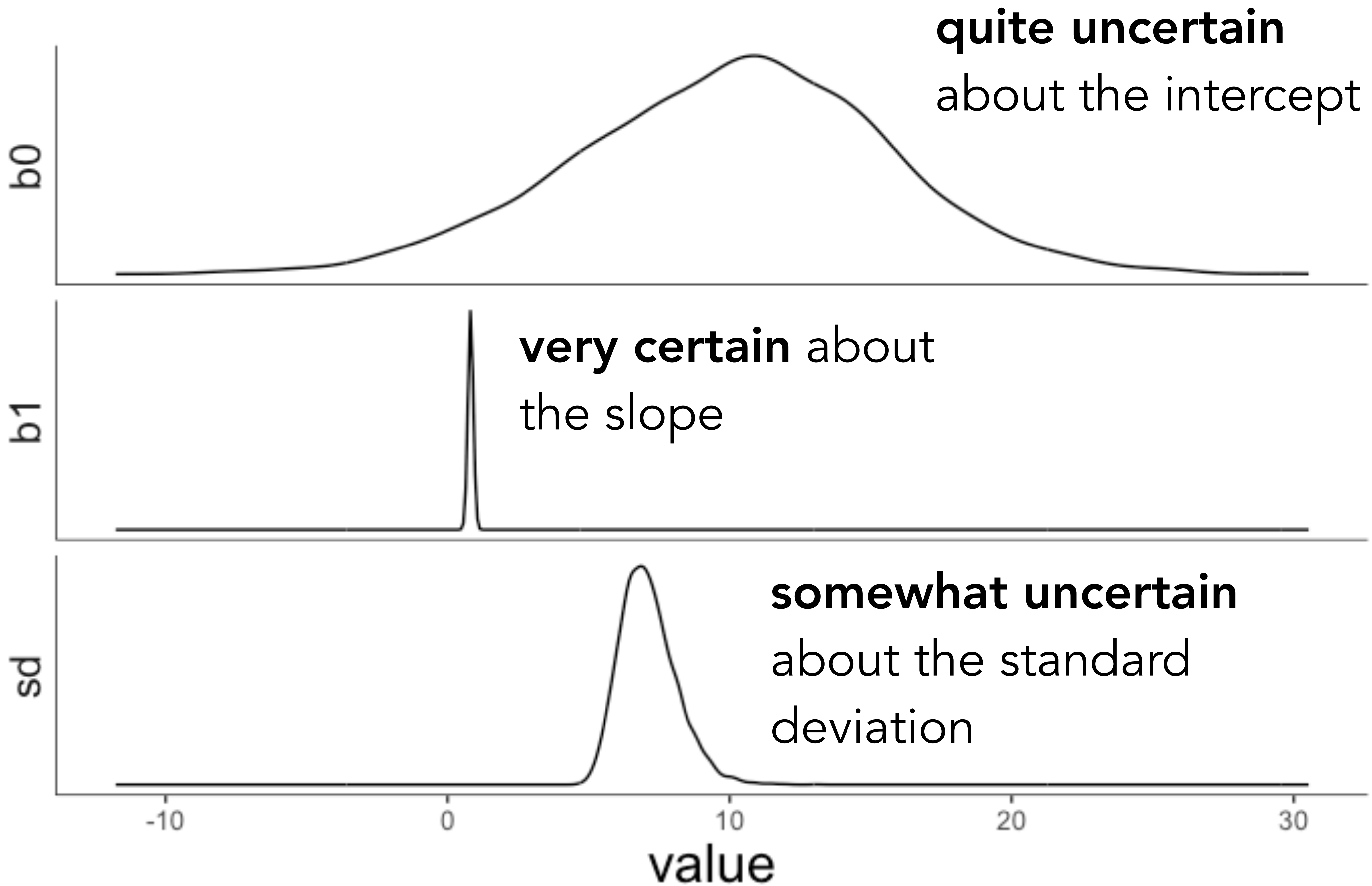
```
# sampling
draws = mcmc(m, n_samples = 1000)

# tidy up the draws
df.draws = tidy_draws(draws) %>%
  clean_names()
```

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

each of these is a solution for explaining the data

Visualize the posterior

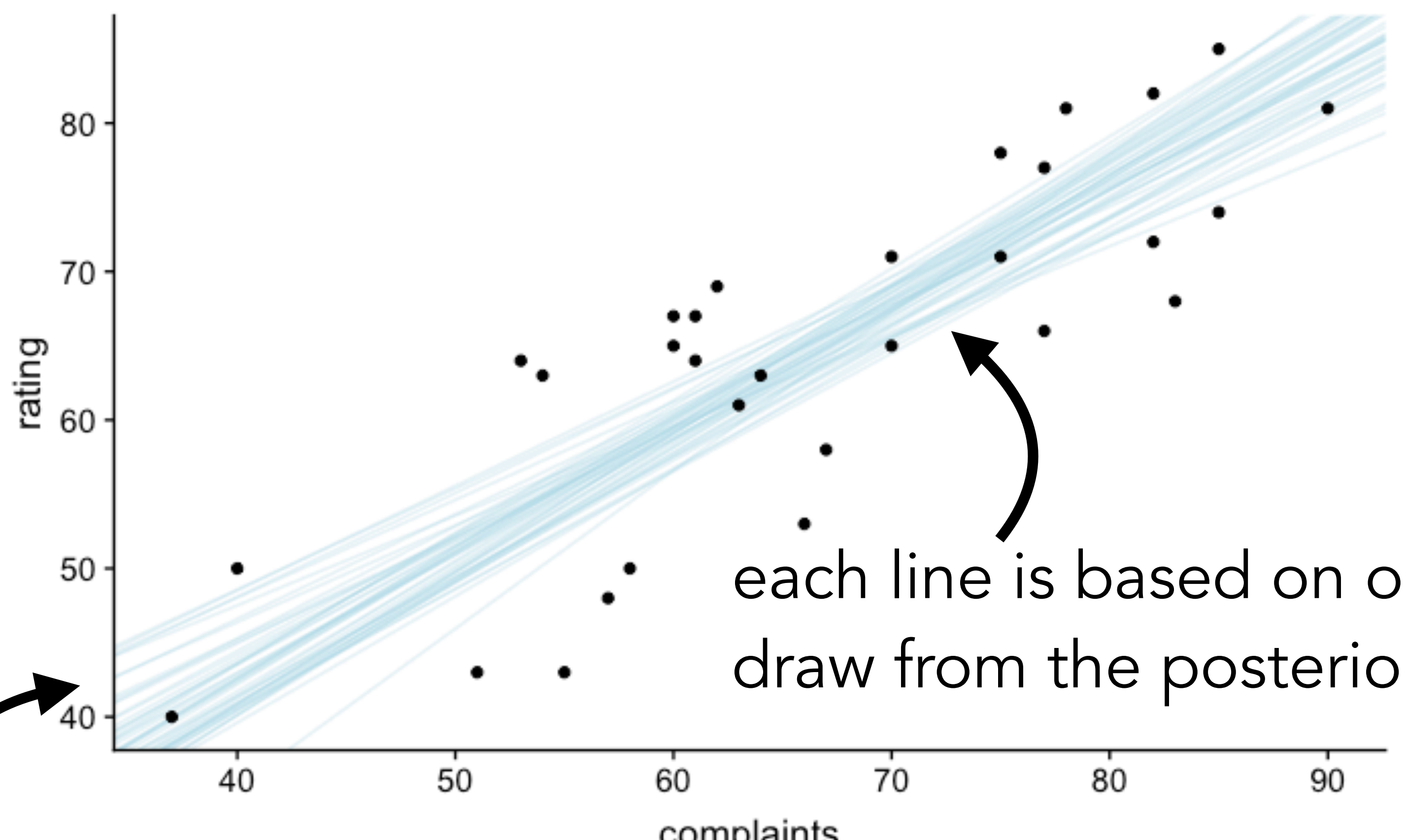


this is the result of a Bayesian data analysis

Posterior distribution over the three parameters in the model

Visualize the model predictions

```r
ggplot(data = df.attitude,
       mapping = aes(x = complaints,
                     y = rating)) +
  geom_abline(data = df.draws %>%
                sample_n(size = 50),
              aes(intercept = b0,
                  slope = b1),
              alpha = 0.3,
              color = "lightblue") +
  geom_point()
```



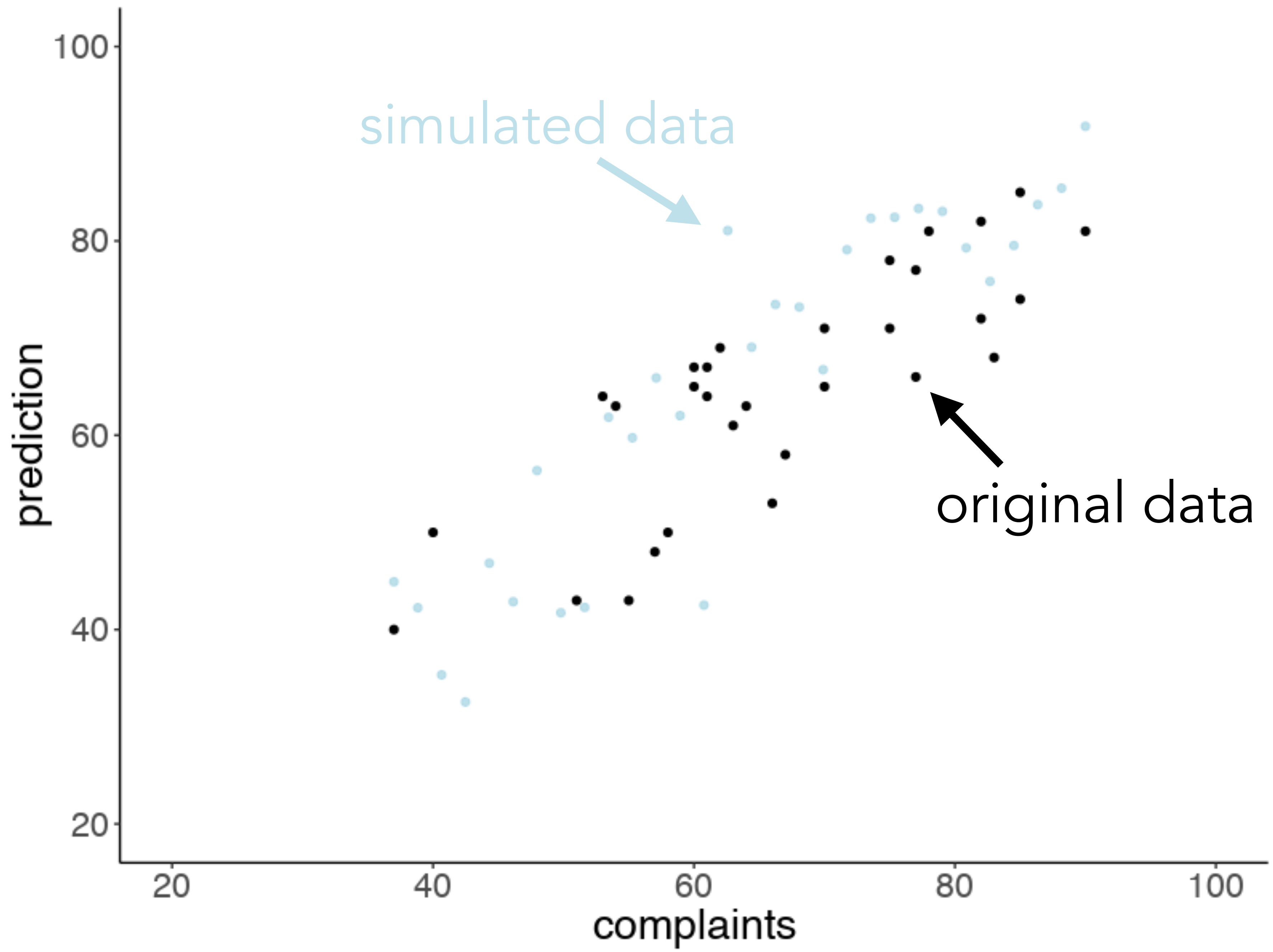
each line is based on one draw from the posterior

explains the "high" uncertainty about the intercept

chain	iteration	draw	b0	b1	sd
1	1	1	6.08	0.87	7.60
1	2	2	1.12	0.95	7.66
1	3	3	-1.83	0.99	7.01
1	4	4	-4.23	1.02	6.64
1	5	5	3.26	0.87	7.96
1	6	6	-1.04	0.98	7.67
1	7	7	-0.83	0.97	10.12
1	8	8	-1.41	0.97	8.02
1	9	9	9.46	0.81	6.30
1	10	10	10.02	0.84	6.57

Posterior predictive check

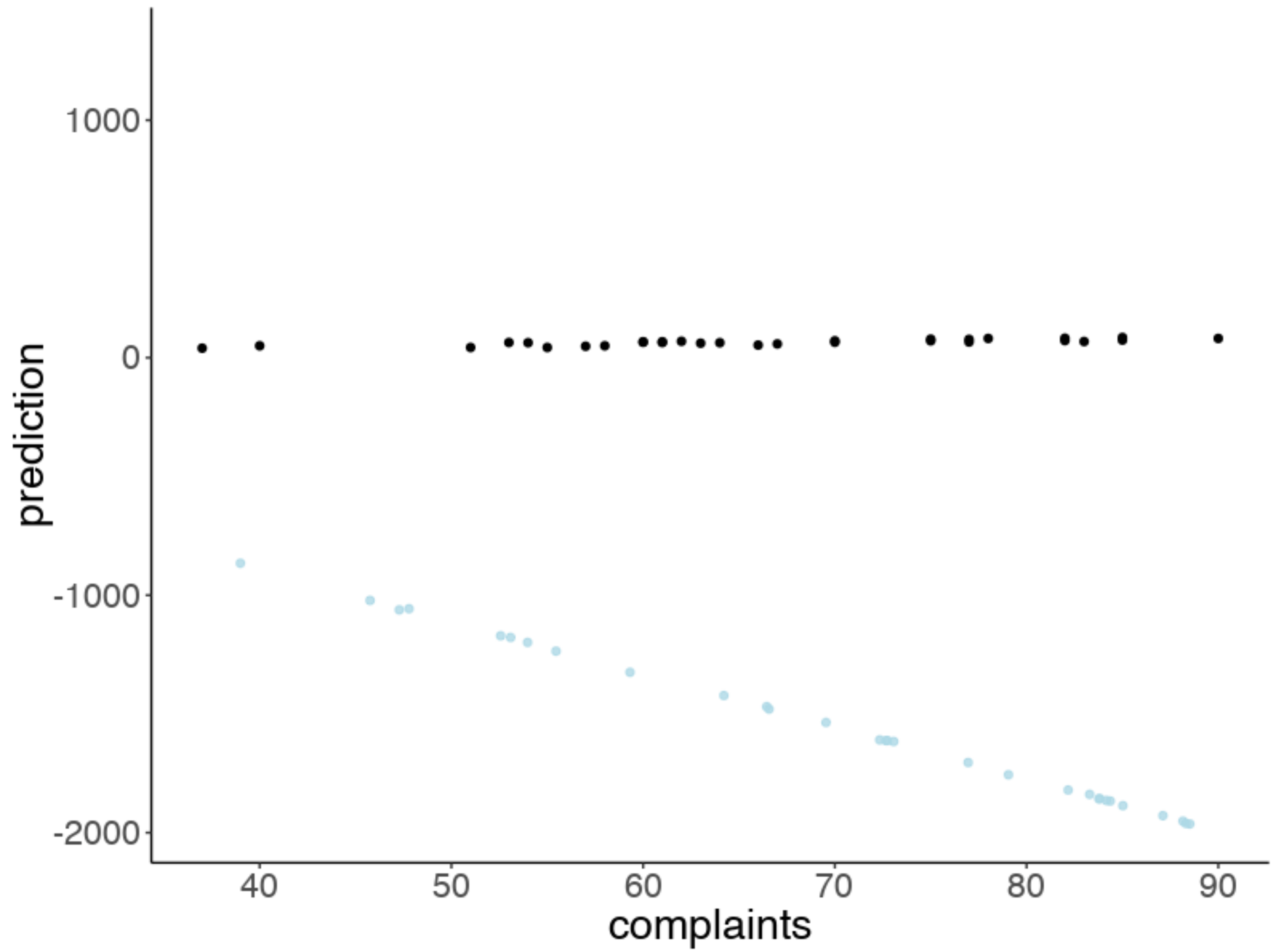
1. sample parameters from the posterior distribution
2. generate data using these parameters (using the likelihood function)



Hypothetical outcome plot

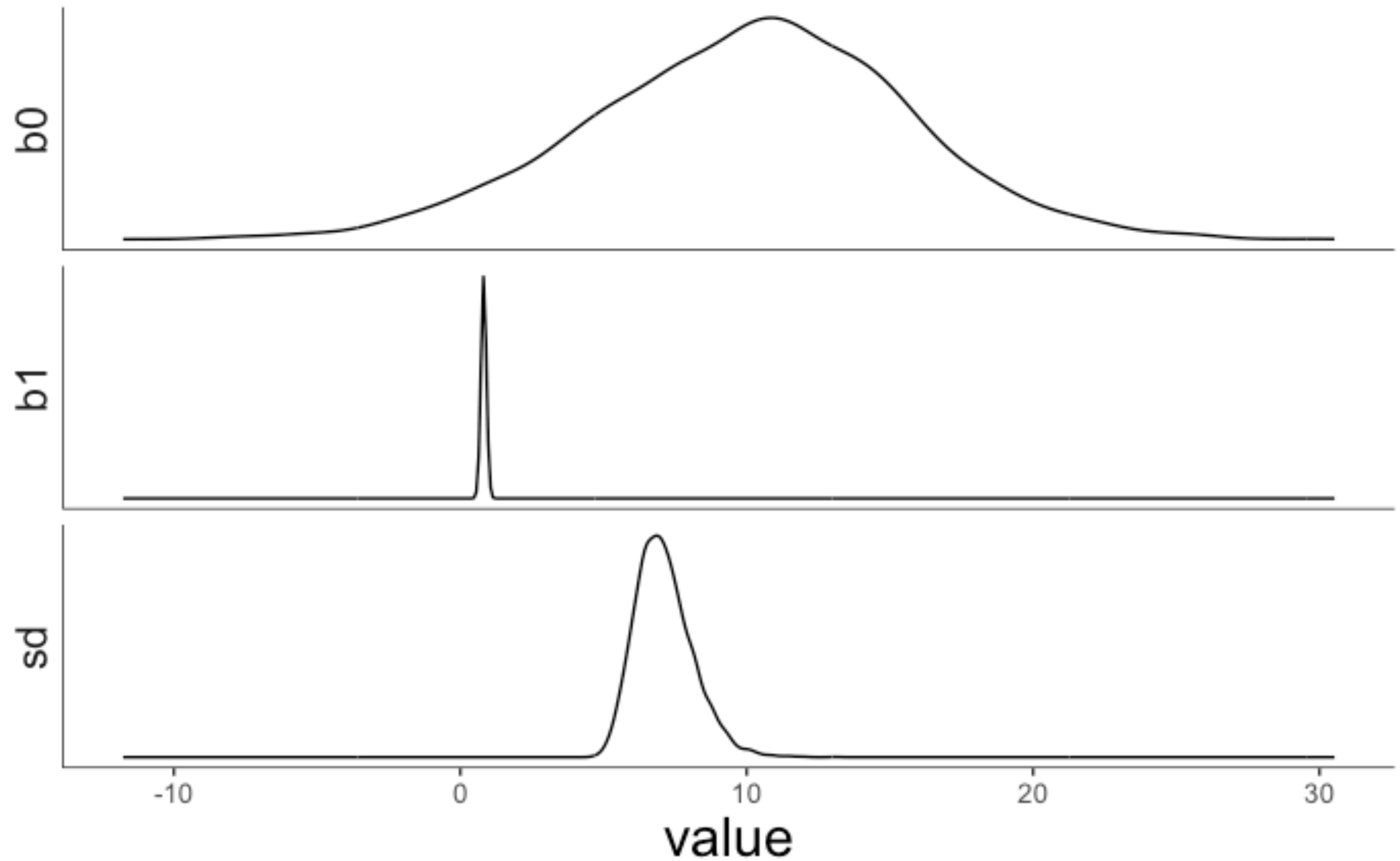
Prior predictive check

1. sample parameters from the **prior distribution**
2. generate data using these parameters (using the likelihood function)



Hypothetical outcome plot

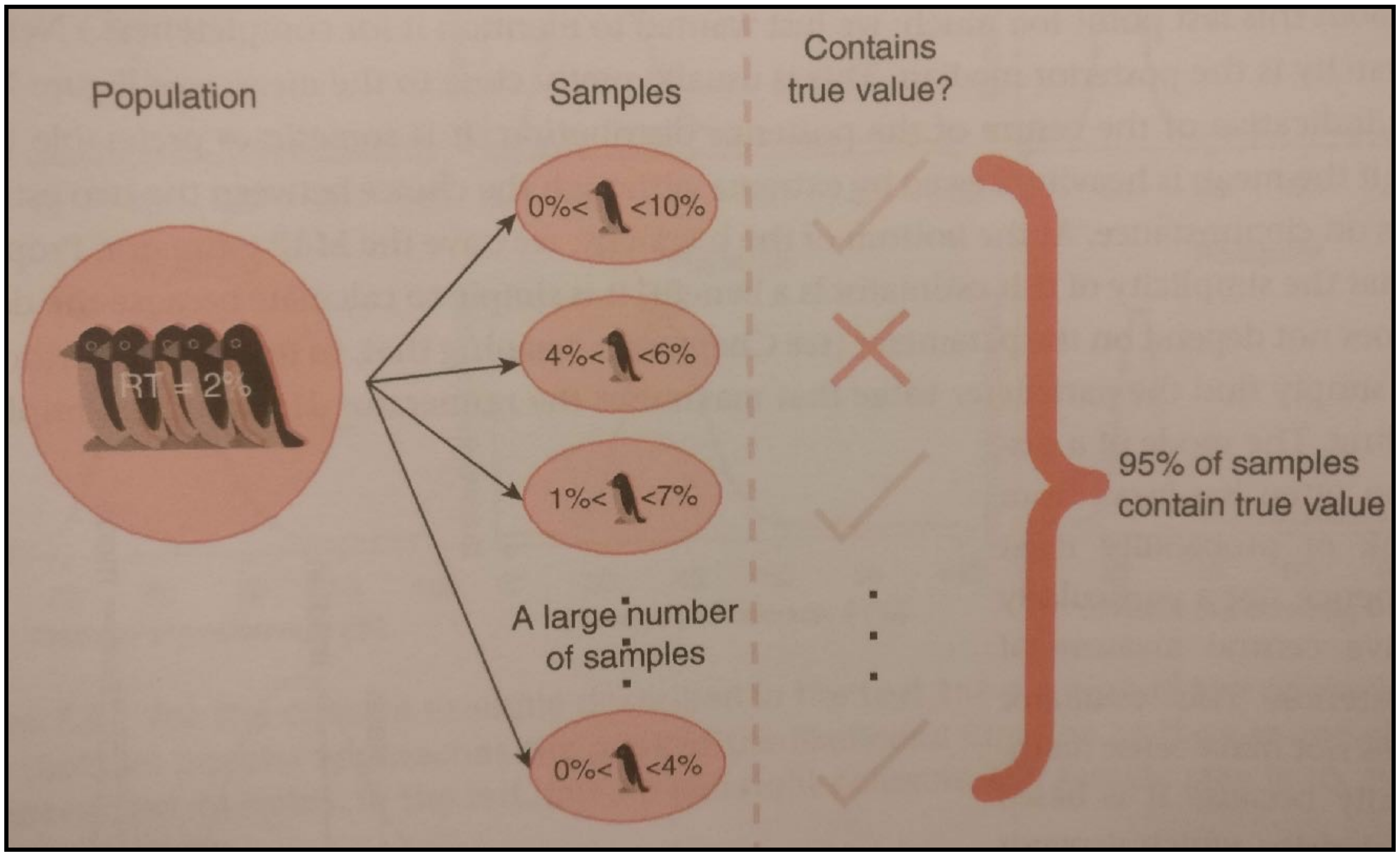
Summarizing results



- Posterior over each parameter is the result of the Bayesian data analysis.
- no p-values
- no confidence intervals

Confidence interval vs. credible interval

"From our research, we concluded that the percentage of penguins with red tails, RT, has a 95% **confidence interval** of 1% < RT < 5%."



For 95% of the (hypothetical) samples, the confidence interval contains the true value.

Confidence interval vs. credible interval

"From our research, we concluded that the percentage of penguins with red tails, RT, has a 95% **credible interval** of 0% < RT < 4%."

Straightforward interpretation

There is a 95% probability that the percentage of penguins with red tails lies in the range of 0% < RT < 4%.

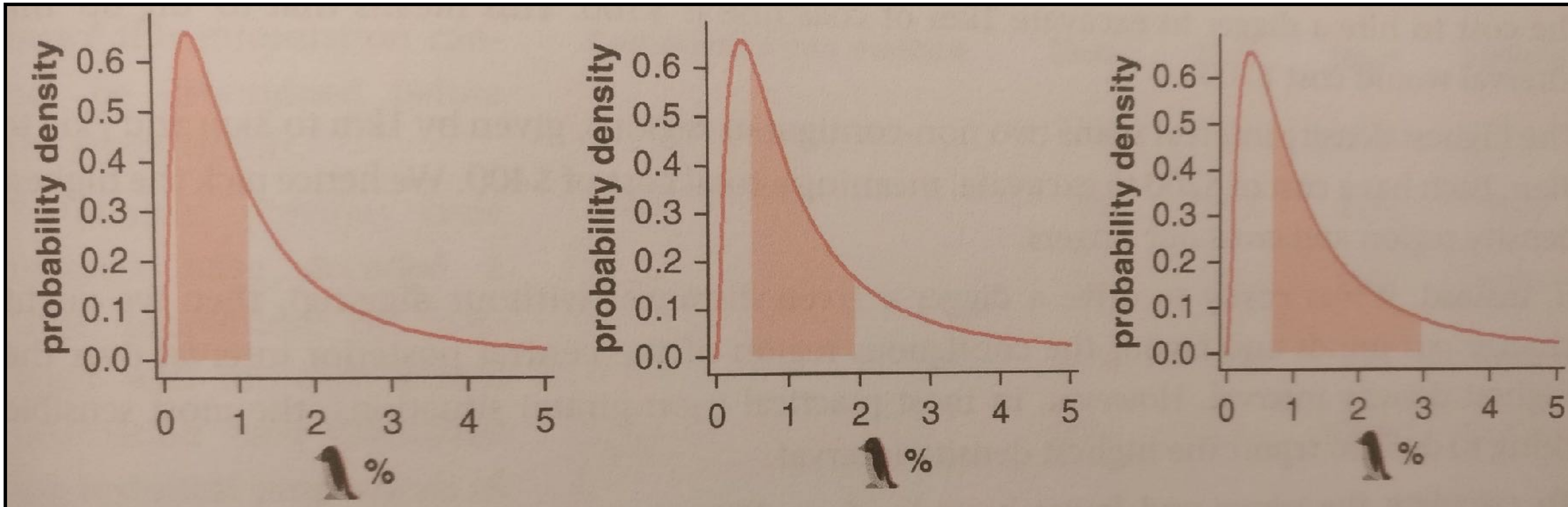
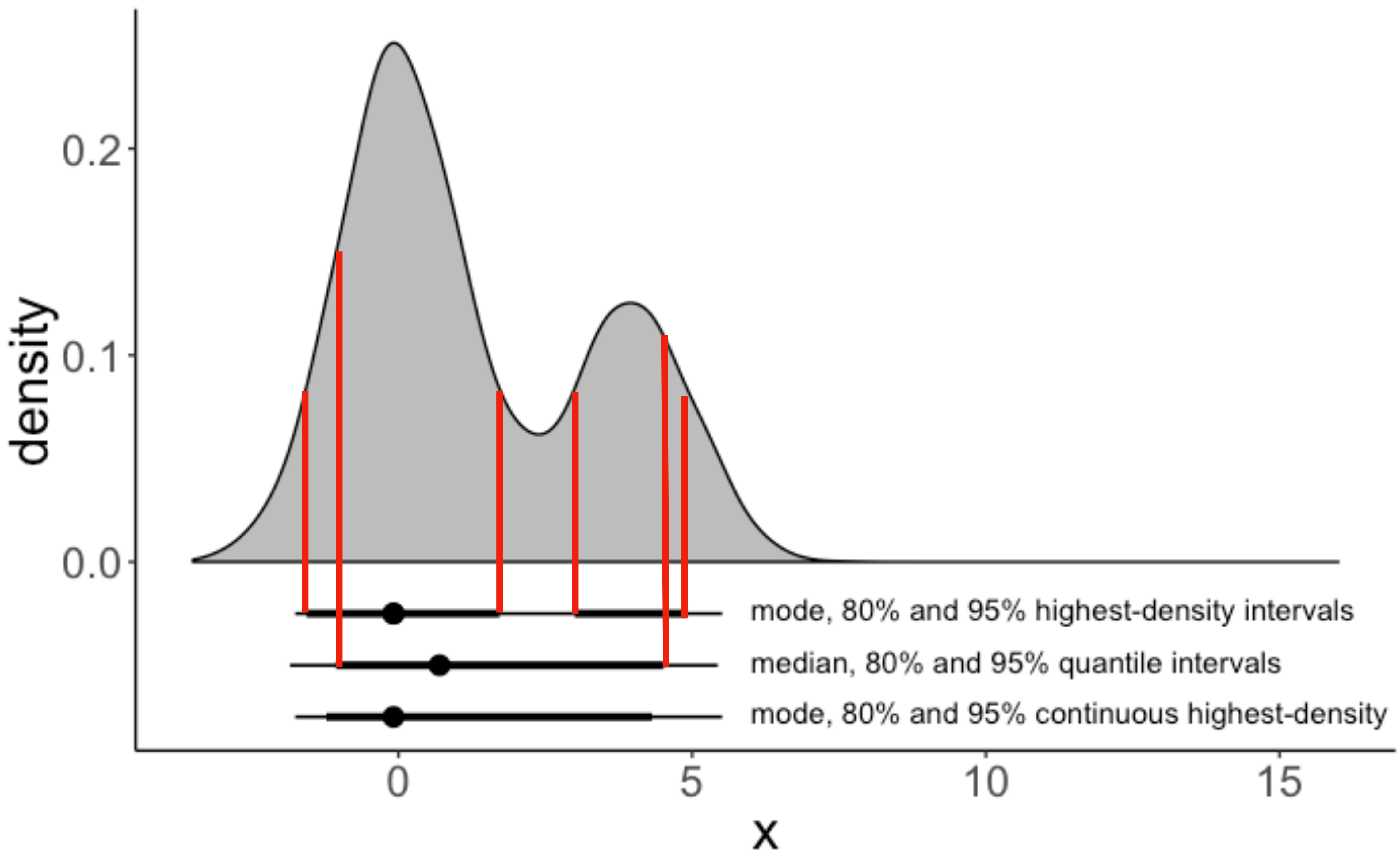


Figure 7.8 Three examples of 50% credible intervals for a parameter representing the proportion of penguins with red tails.

Different kinds of credible intervals

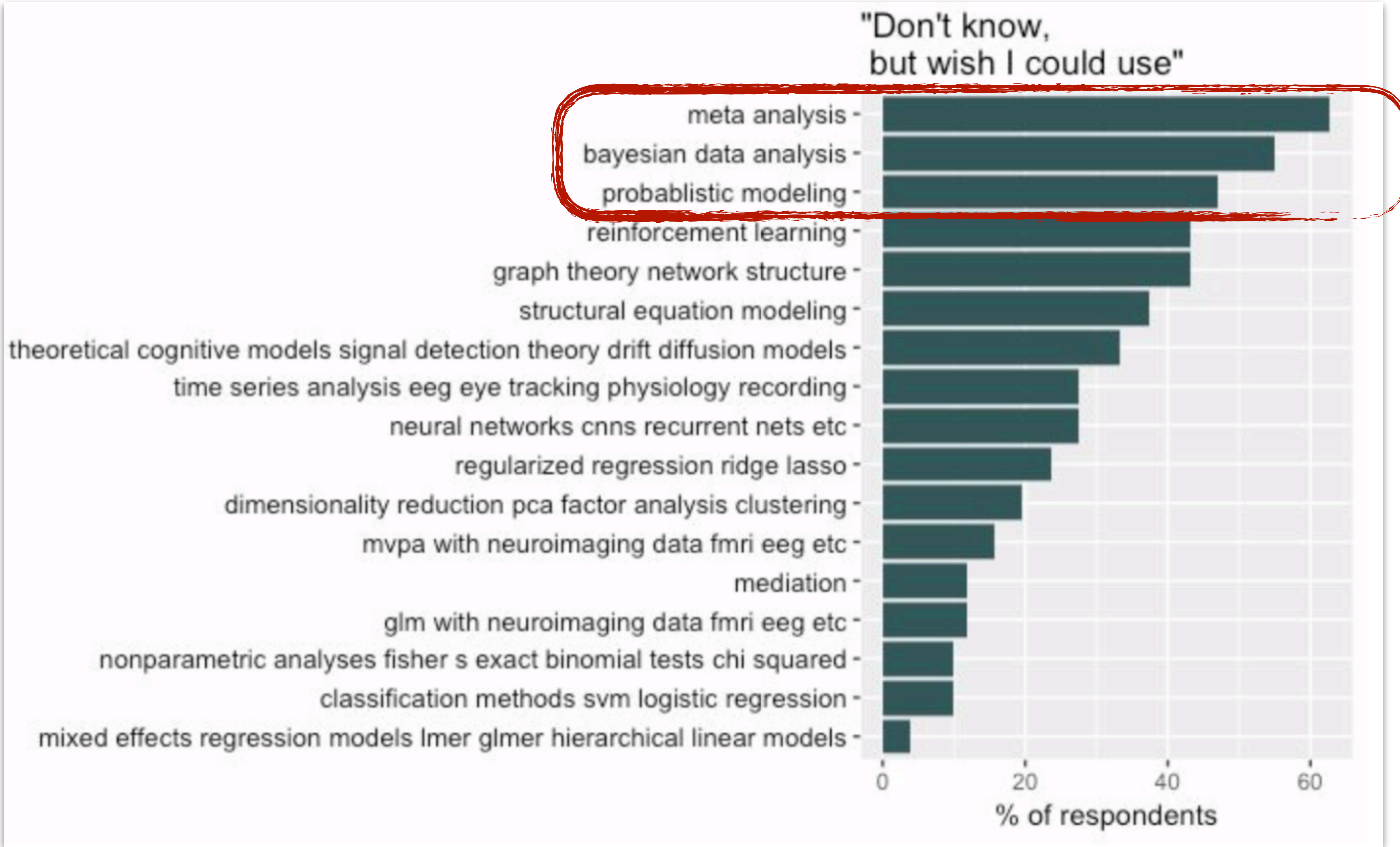


Plan for today

- Quick Bayes recap
- Ingredients: likelihood, prior, inference
- Doing Bayesian data analysis
- **Bayesian models of cognition**

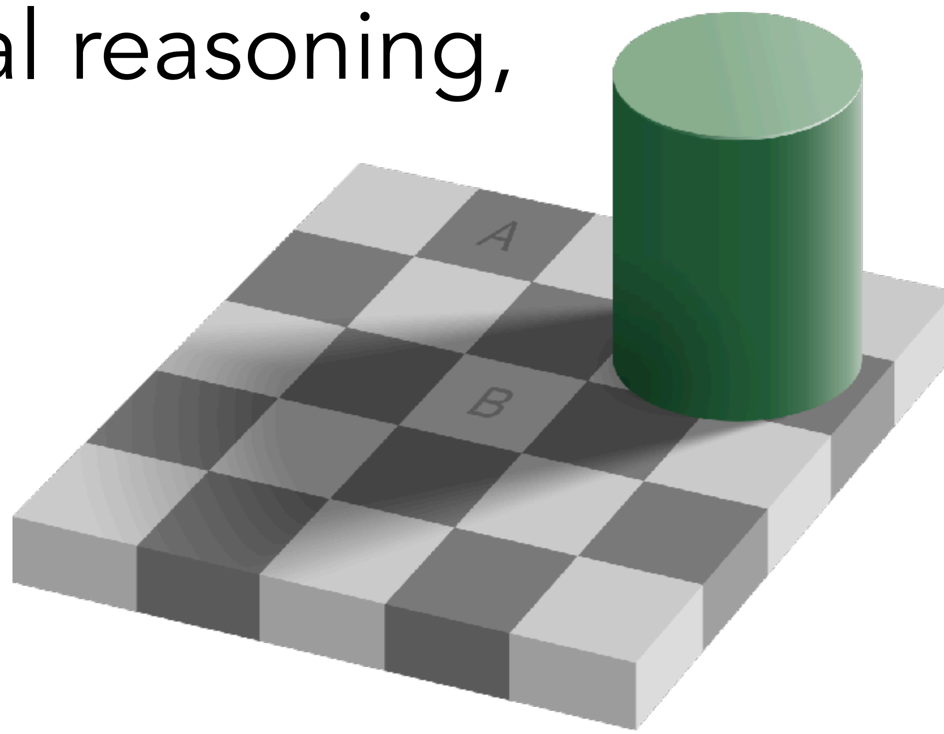
Bayesian models of cognition

Bayesian models of cognition



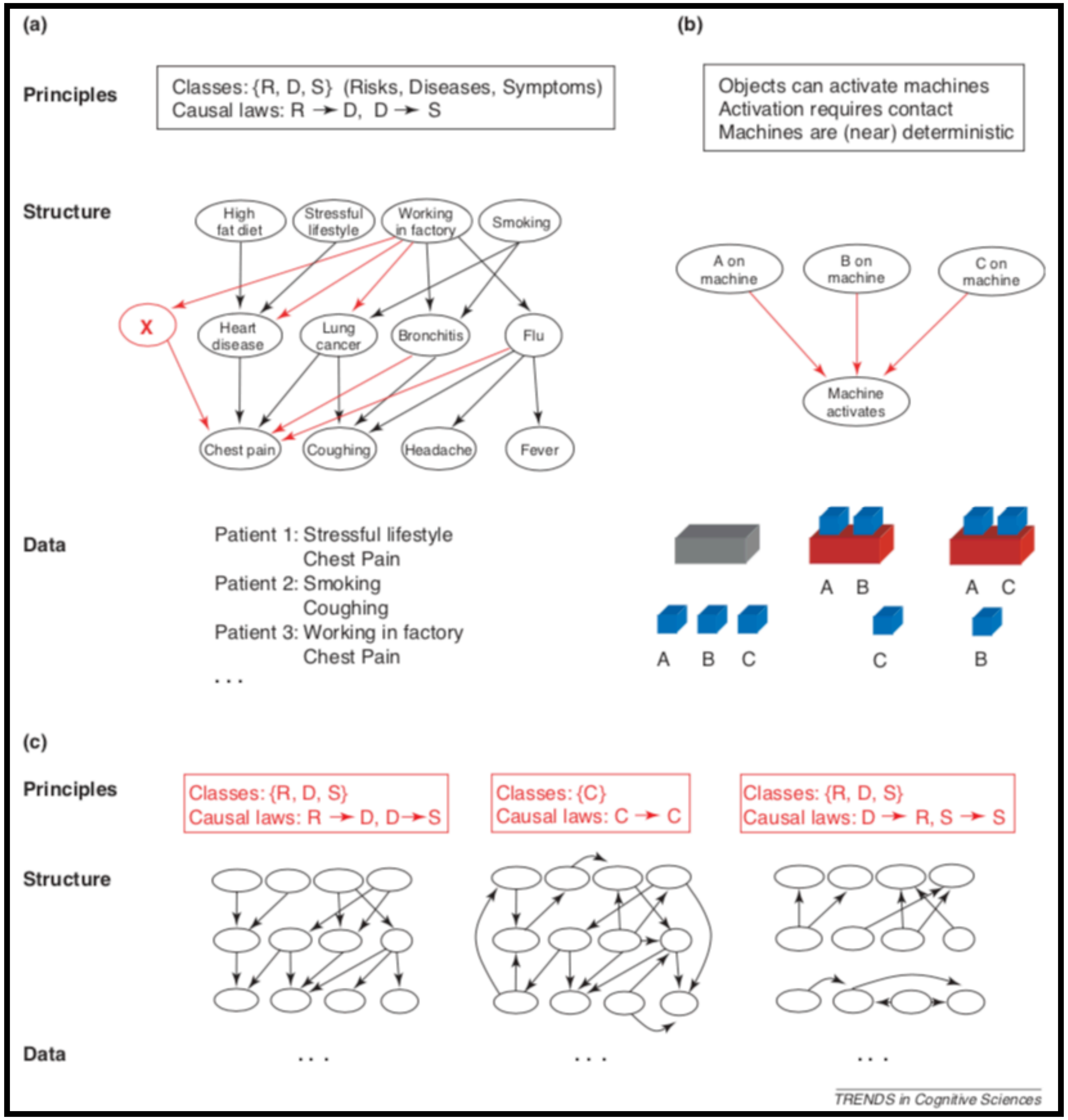
Bayesian models of cognition

- Is the mind doing Bayesian inference?
  - combines prior knowledge with data to draw conclusions
  - many cognitive phenomena can be thought of as inductive problems (e.g. perception, learning, theory of mind, physical reasoning, …)



Yuille & Kersten (2006) Vision as Bayesian inference: Analysis by synthesis? Trends in Cognitive Sciences

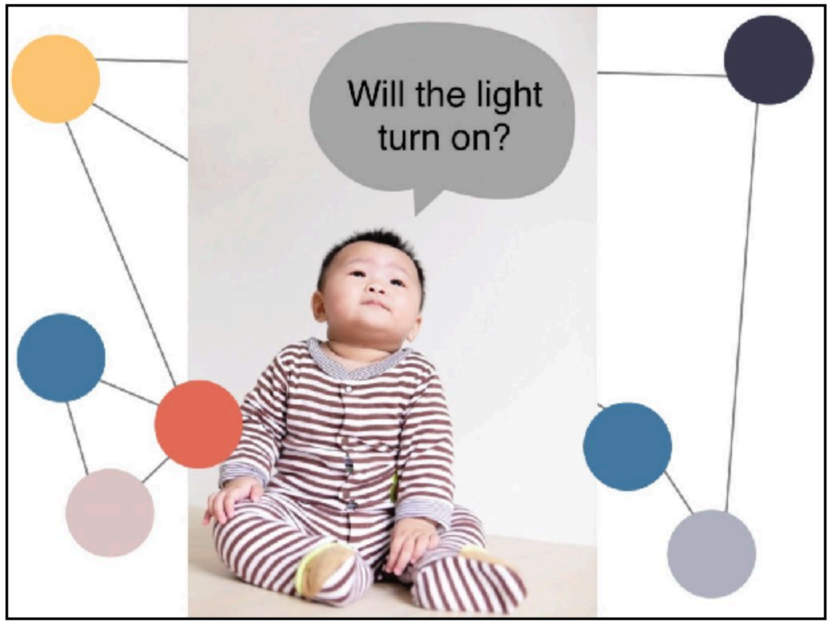
Bayesian models of cognition



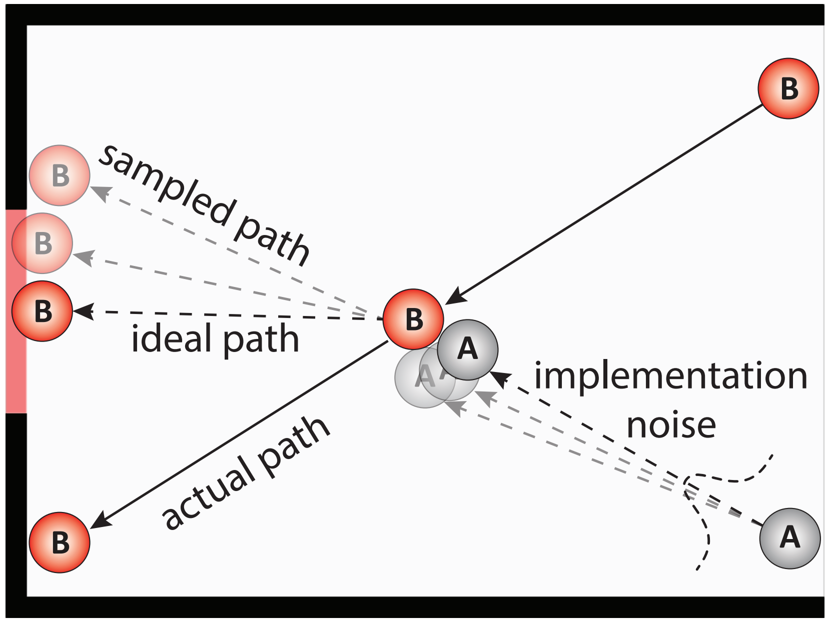
Tenenbaum, Griffiths, & Kemp (2006) Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences

Causality in Cognition

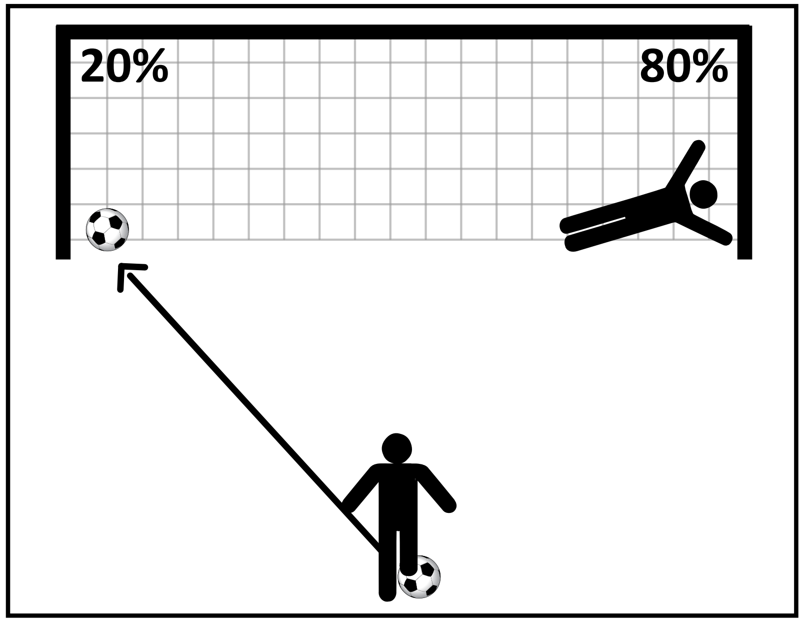
Our lab studies the role of causality in our understanding of the world, and of each other.



learning



reasoning

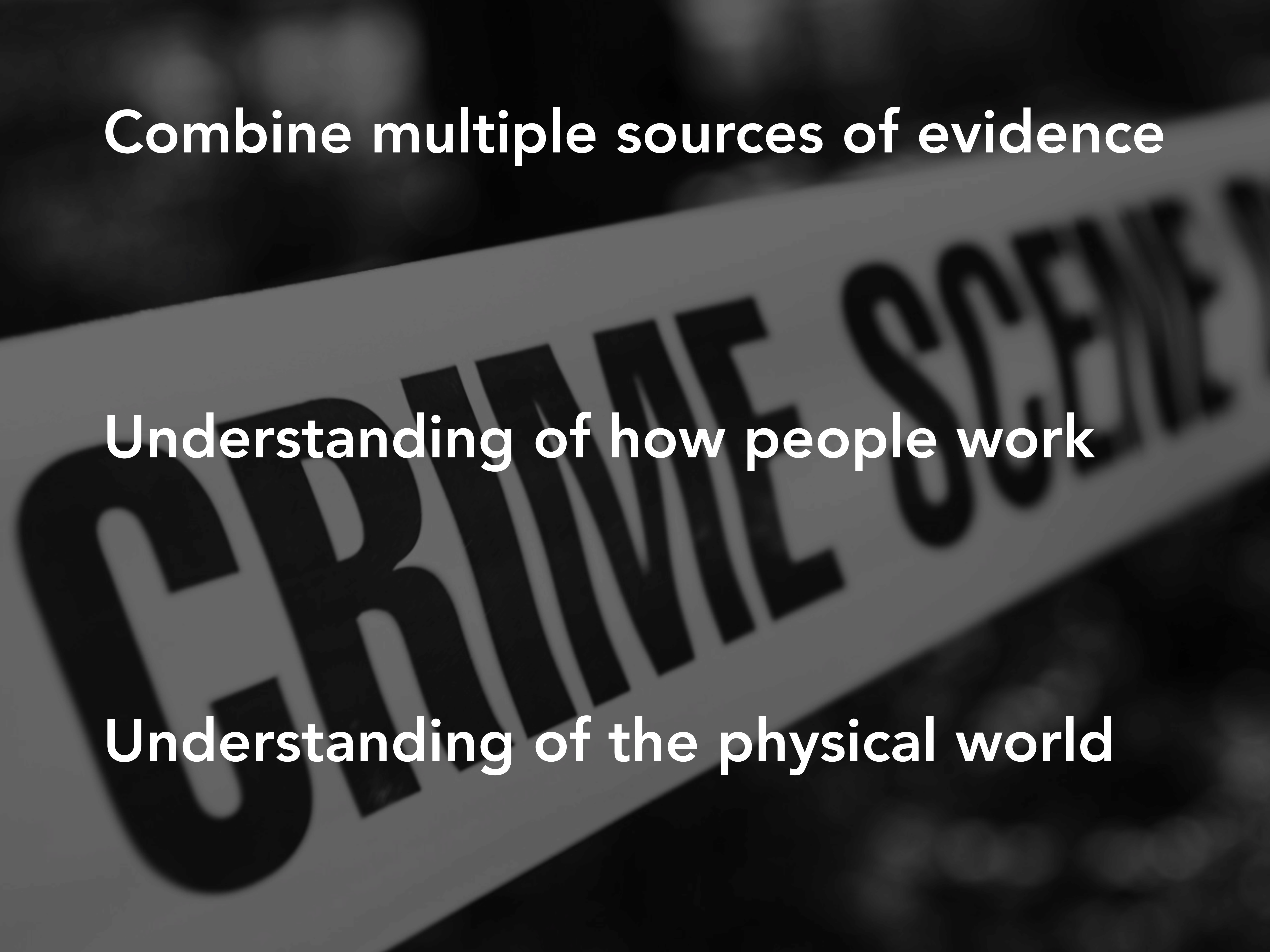


judgment

@tobigerstenberg

http://cicl.stanford.edu/


Coca-Cola
Coca-Cola
Coca-Cola
+
BANG
BANG



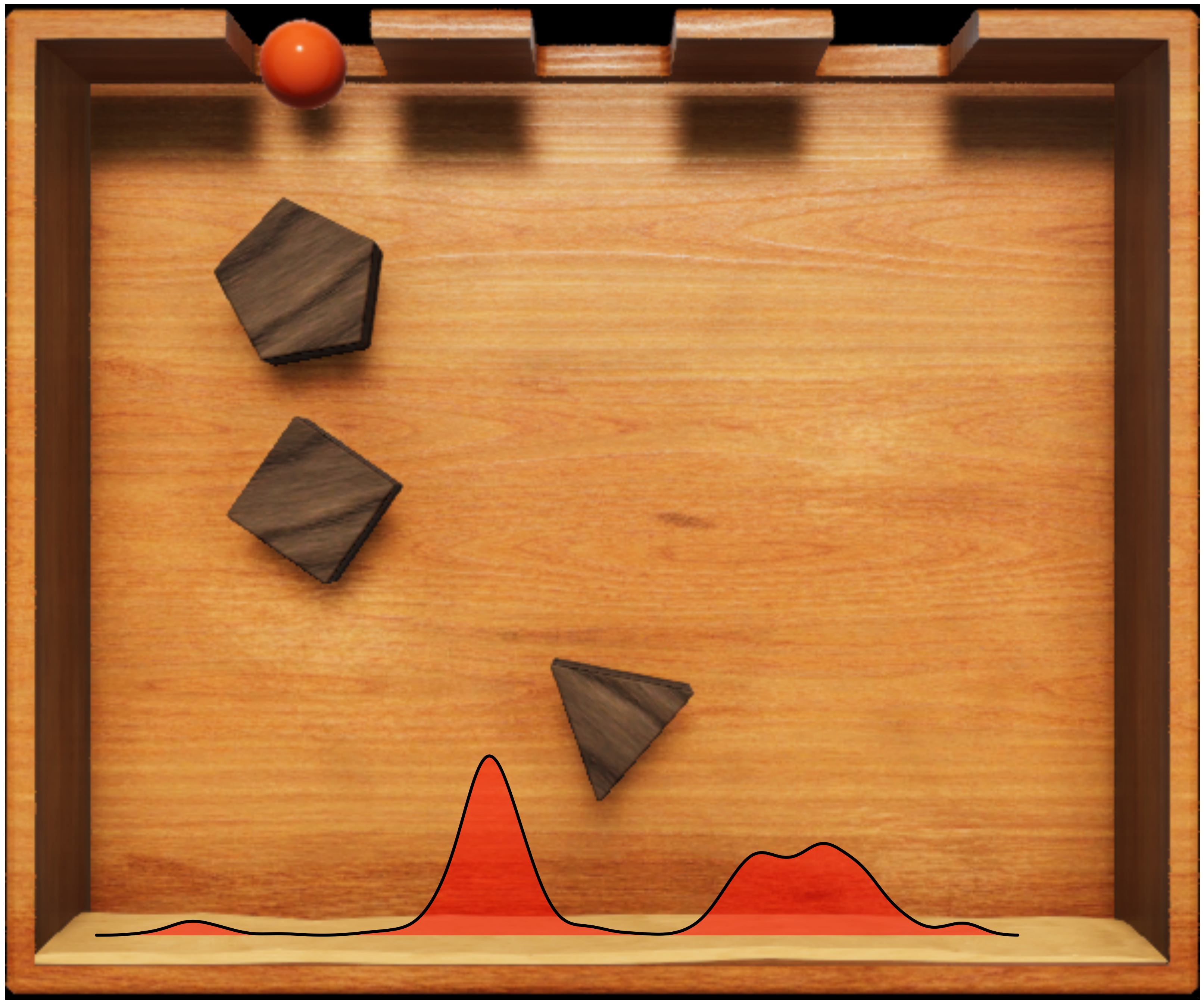
Combine multiple sources of evidence

Understanding of how people work

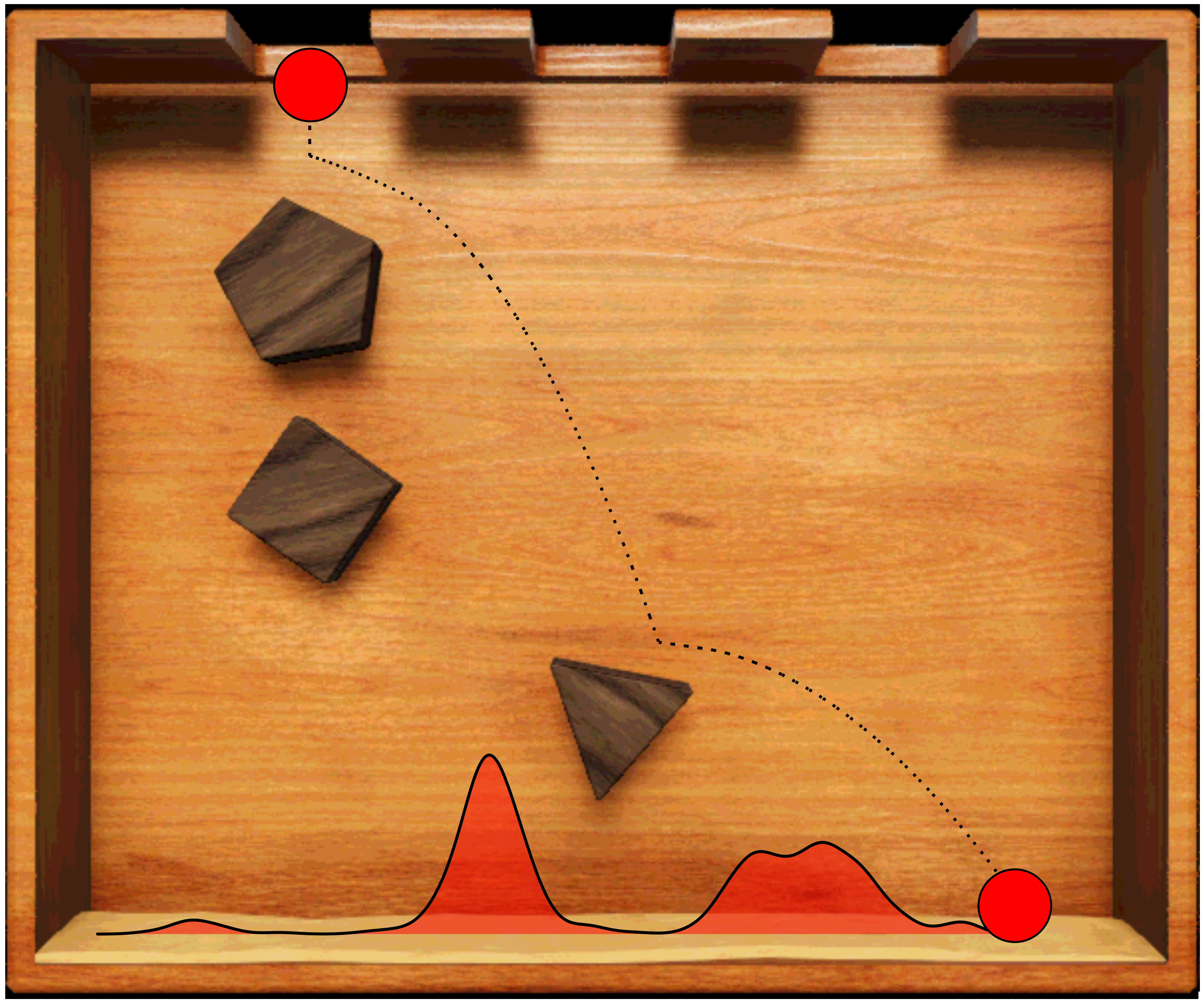
Understanding of the physical world

Prediction: Where will the ball land?

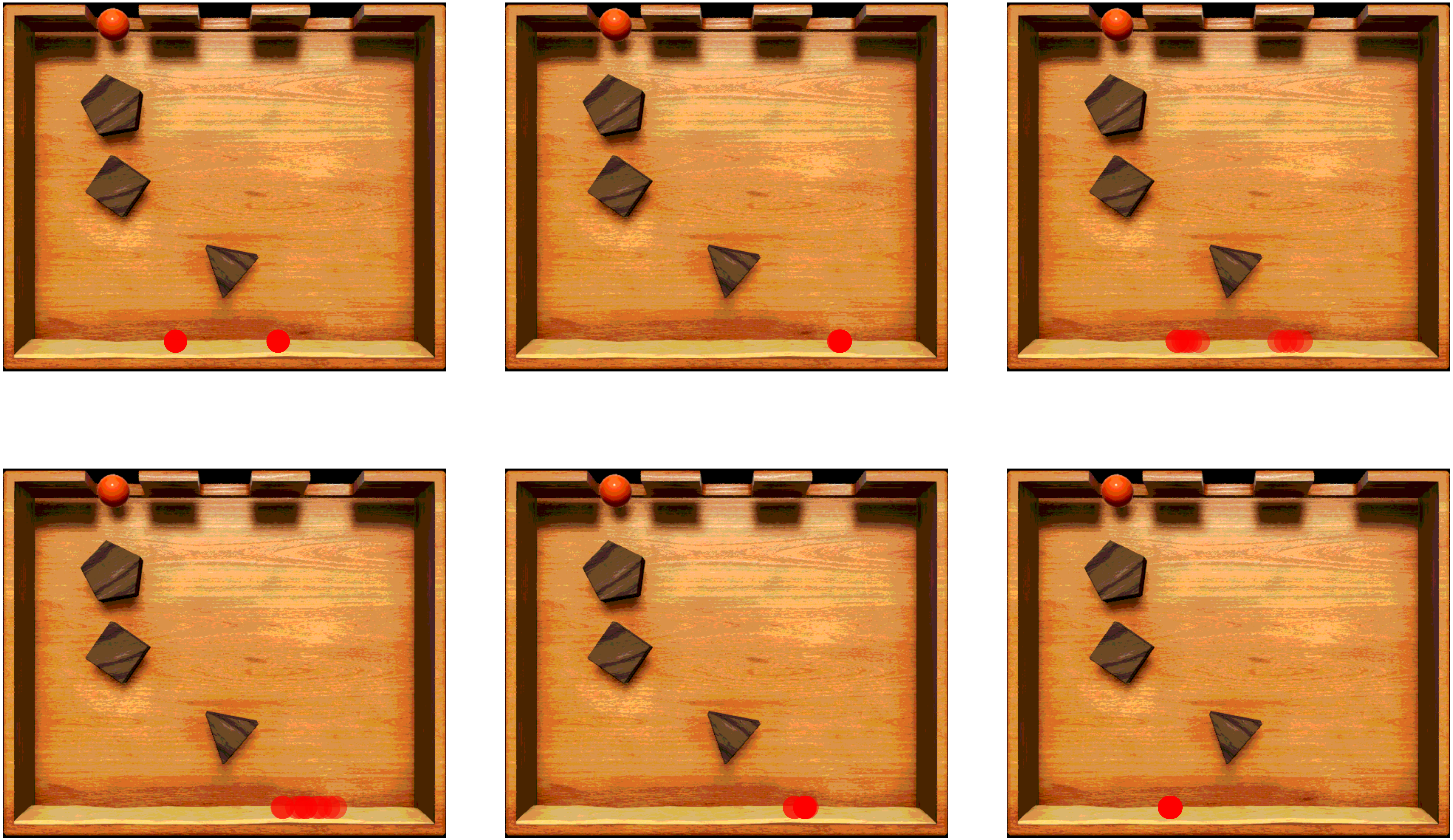
Prediction: Where will the ball land?



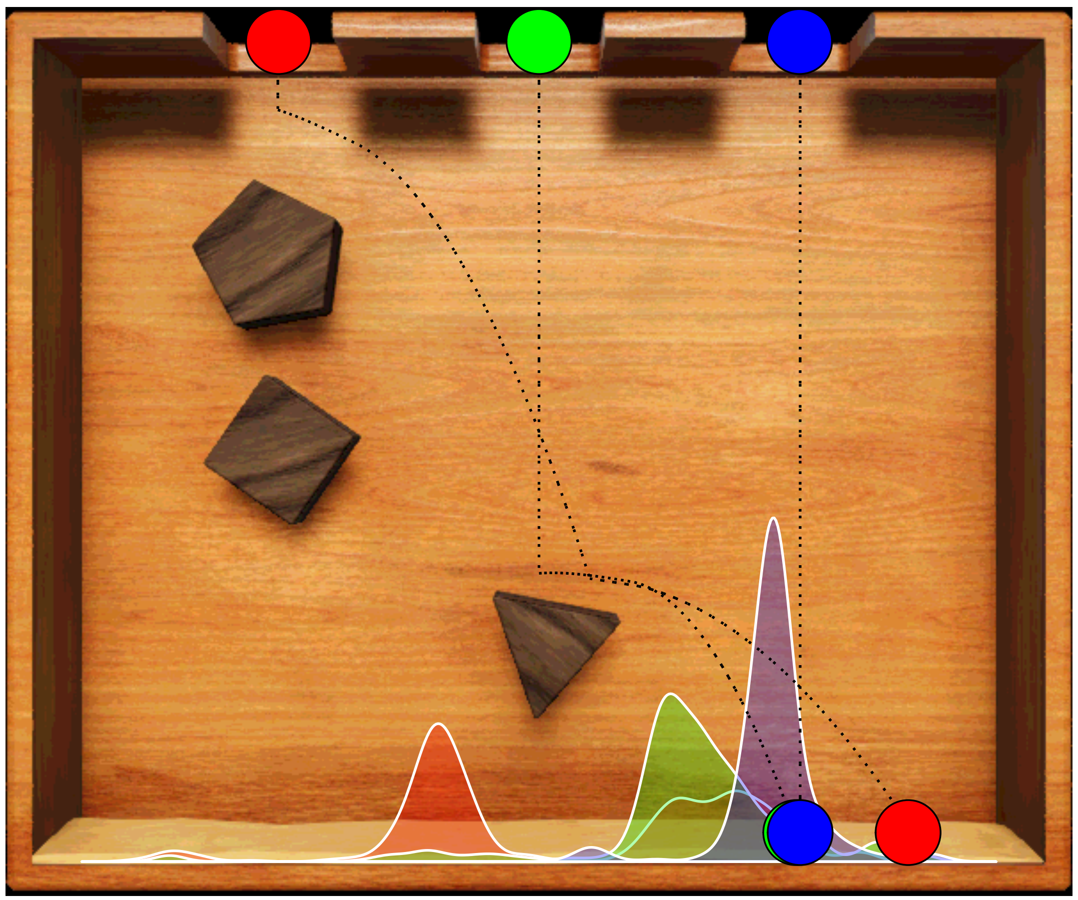
Prediction: Where will the ball land?



Prediction: Where will the ball land?



Prediction: Where will the ball land?

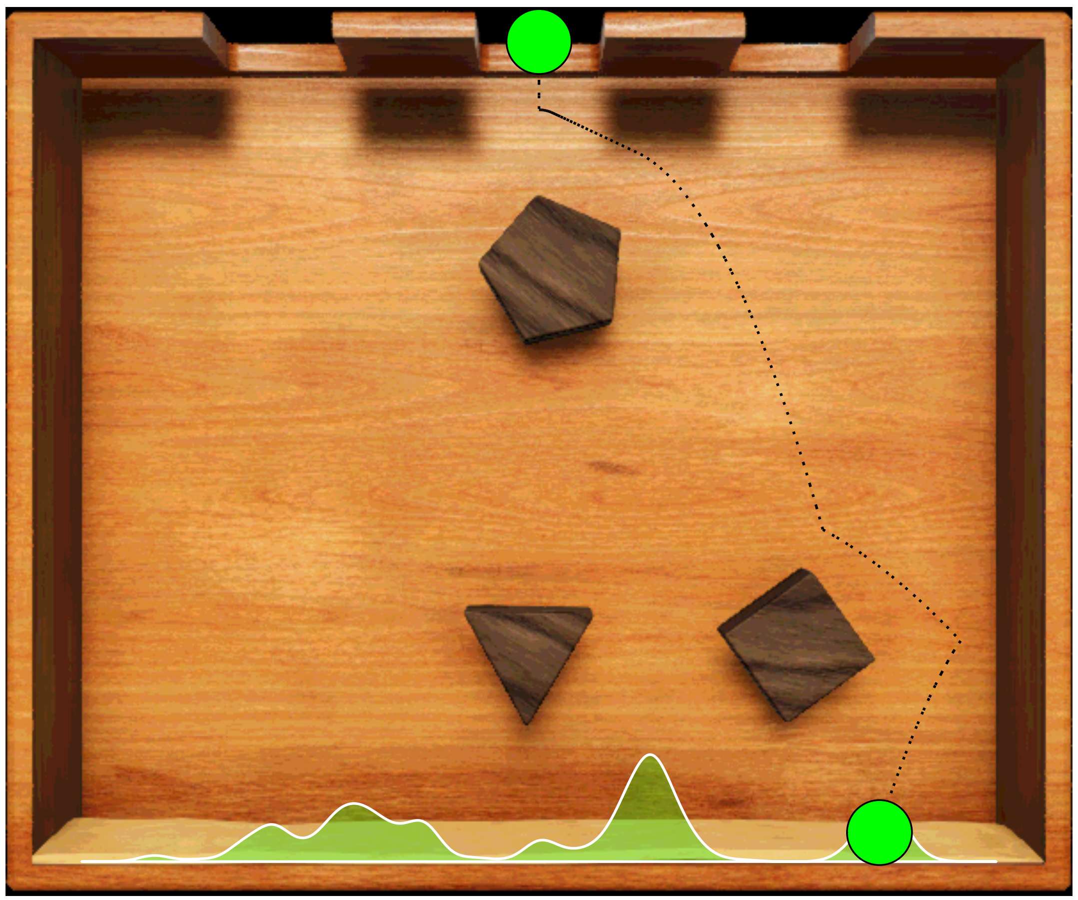


people

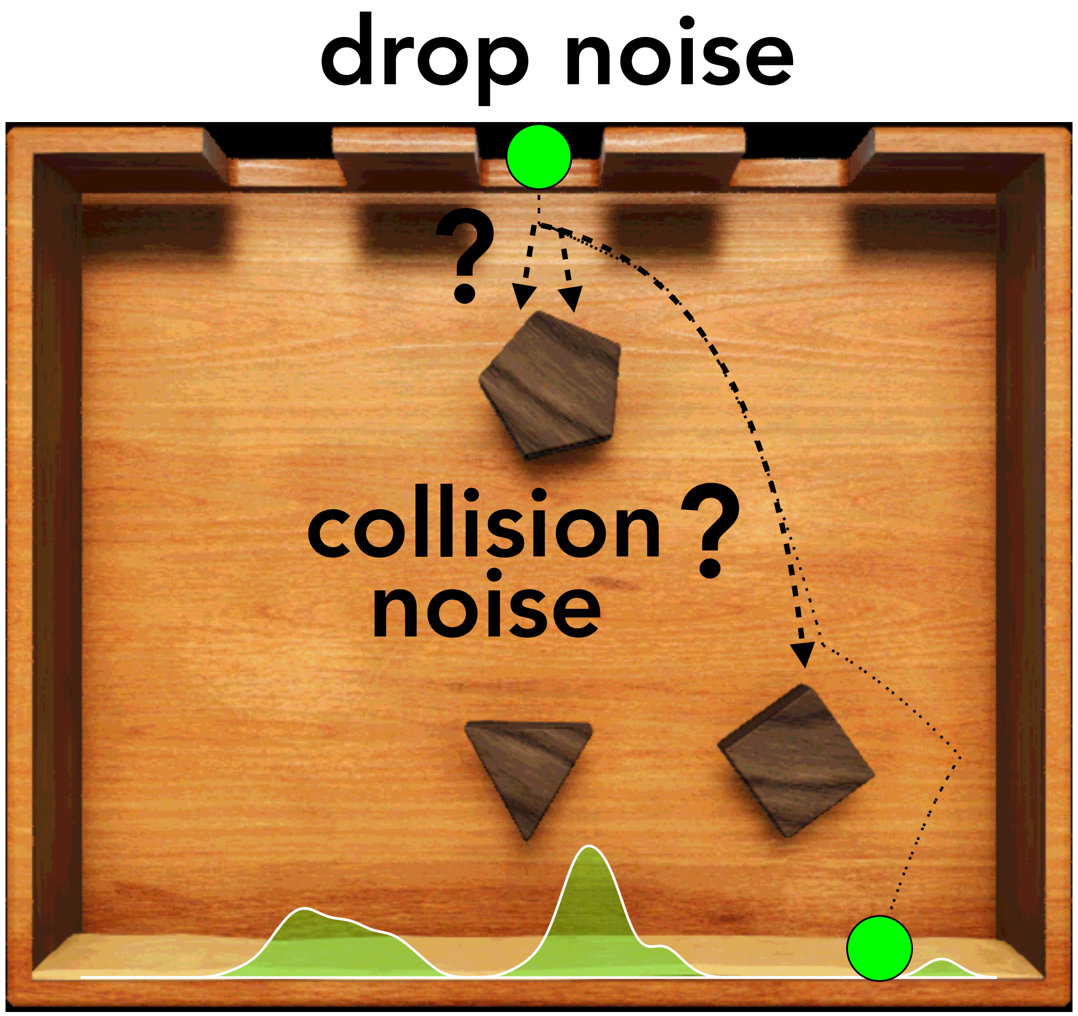
Ullman, Spelke, Battaglia, & Tenenbaum (2017) Mind Games: Game Engines as an Architecture for Intuitive Physics. *Trends in Cognitive Sciences*

Smith & Vul (2013) Sources of uncertainty in intuitive physics. *Topics in Cognitive Science*

Prediction: Where will the ball land?

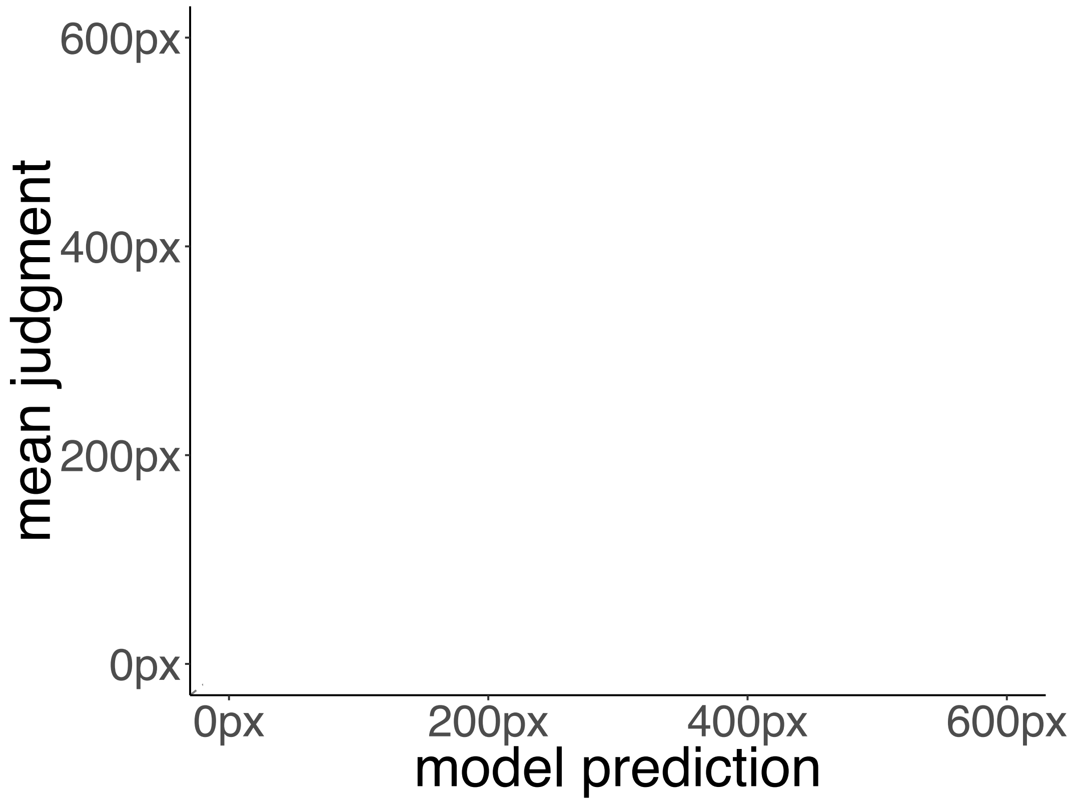


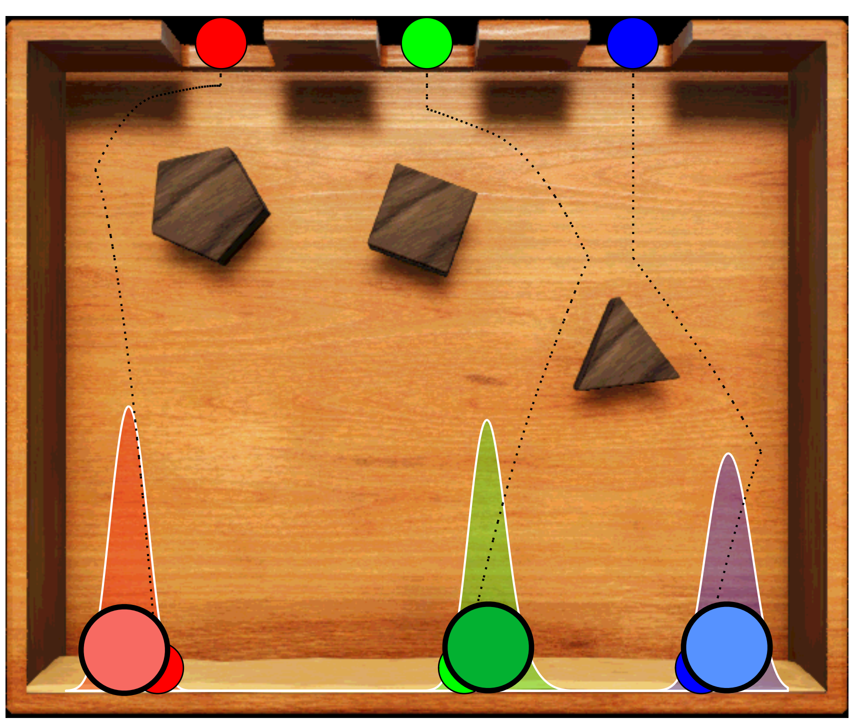
people



model

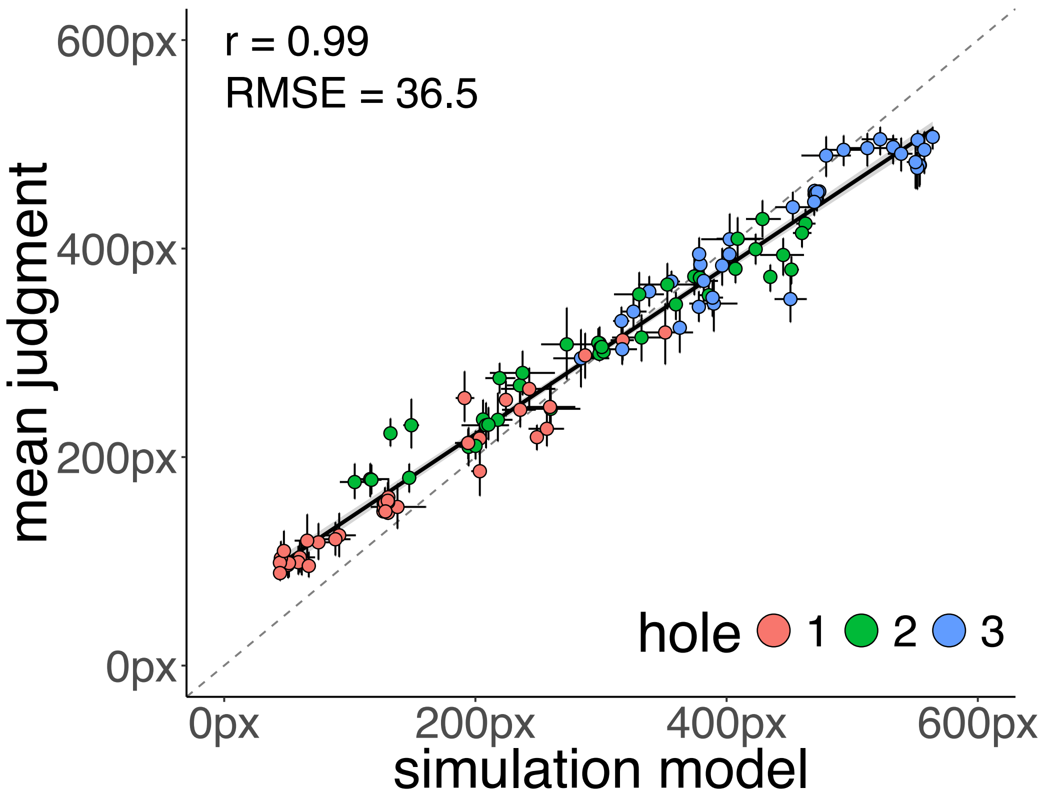
Prediction: Where will the ball land?





mean model prediction

Prediction: Where will the ball land?

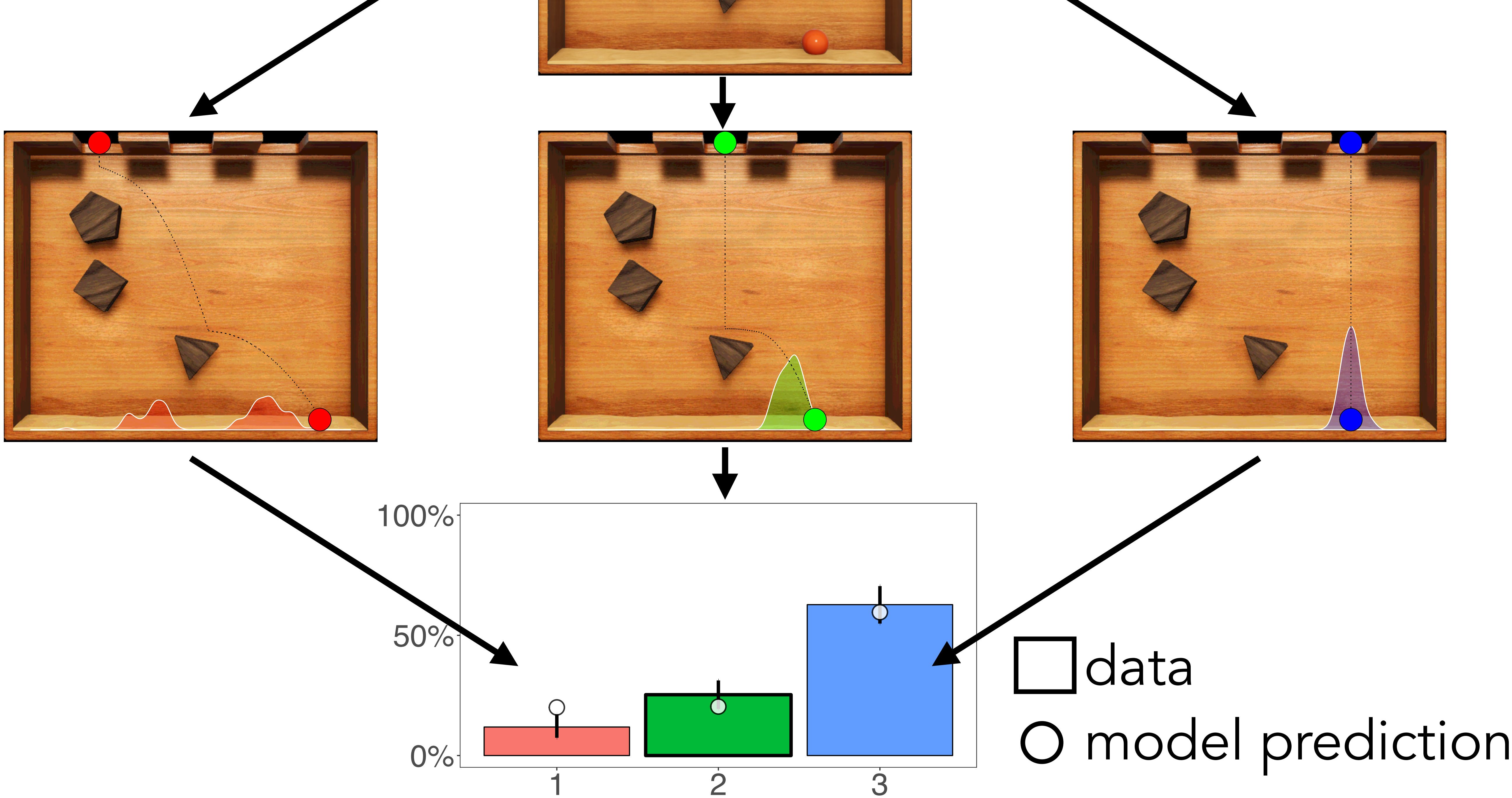


Inference: In which hole was the ball dropped?

Inference: In which hole was the ball dropped?

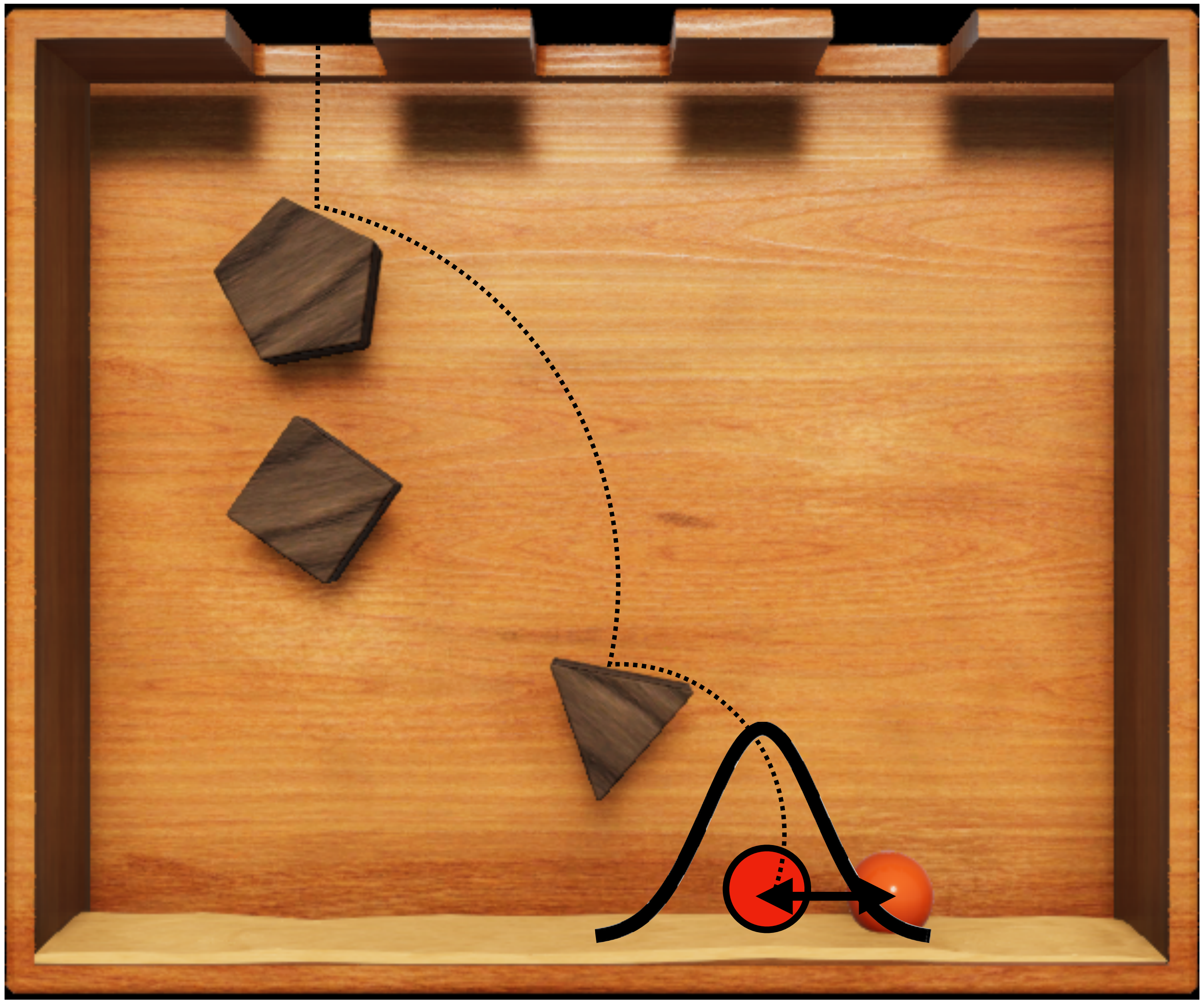
distance between ball's true x position and x position in sample

$$\exp\left(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2}\right)$$



Inference: In which hole was the ball dropped?

Uniform prior over the three different holes

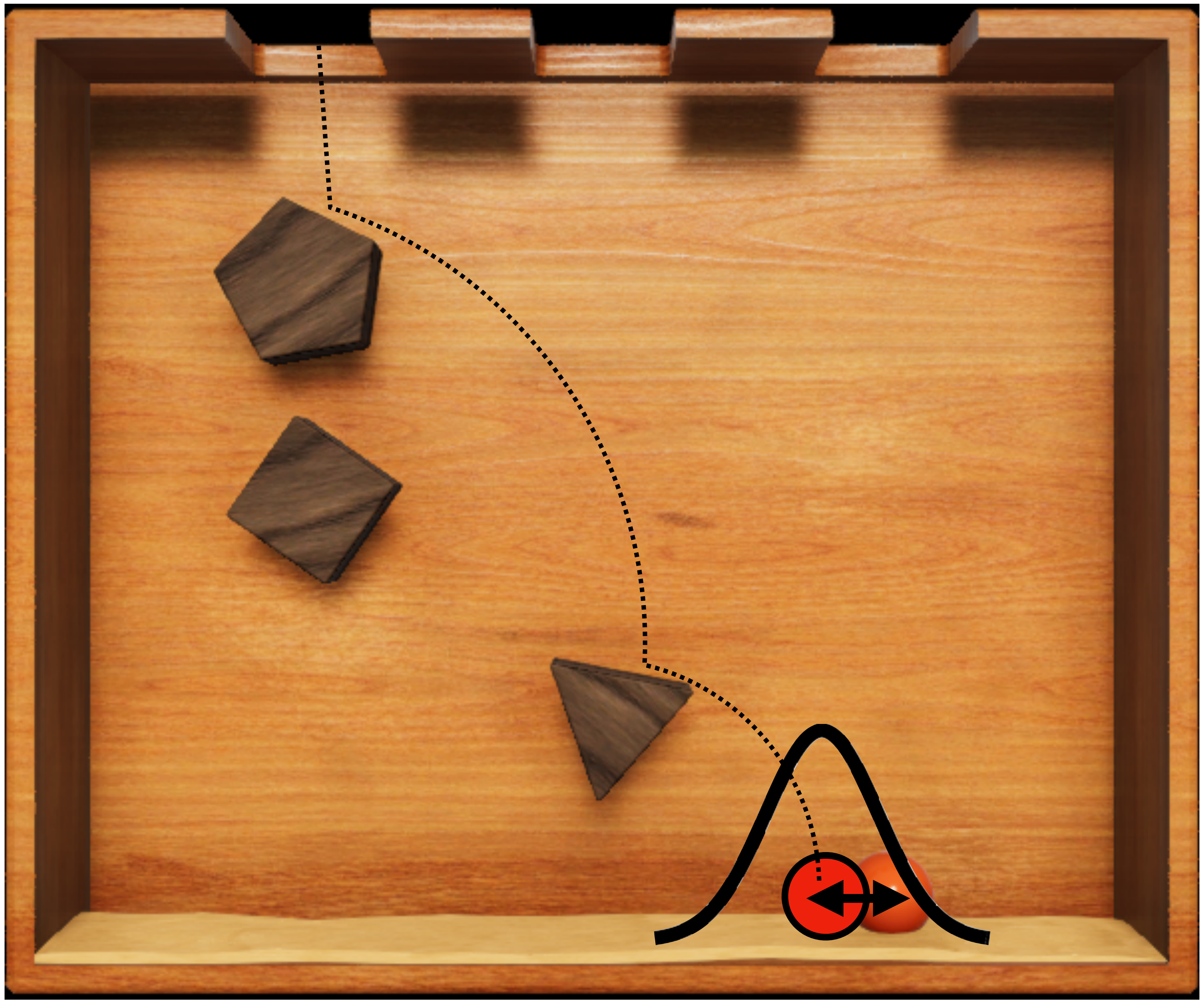


Likelihood of the ball ending up here if it was dropped in hole 1/2/3?

- drop the ball (with noise)
- look at the distance between the simulated ball and where the ball actually ended up
- convert this distance into a likelihood via a Gaussian distribution

Inference: In which hole was the ball dropped?

Uniform prior over the three different holes



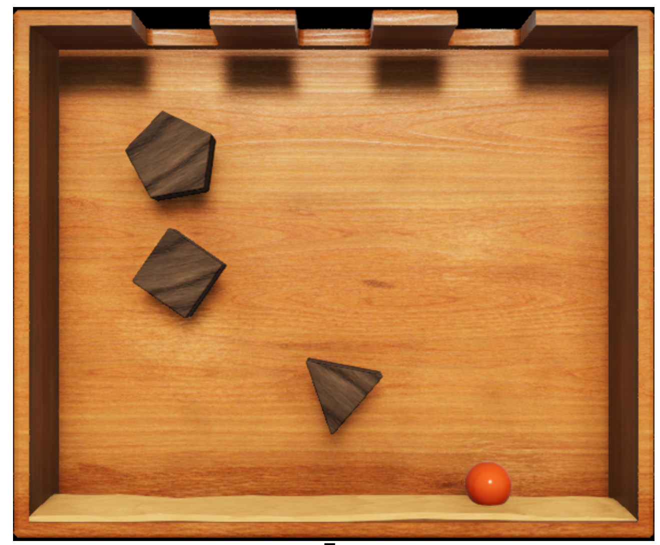
Likelihood of the ball ending up here if it was dropped in hole 1?

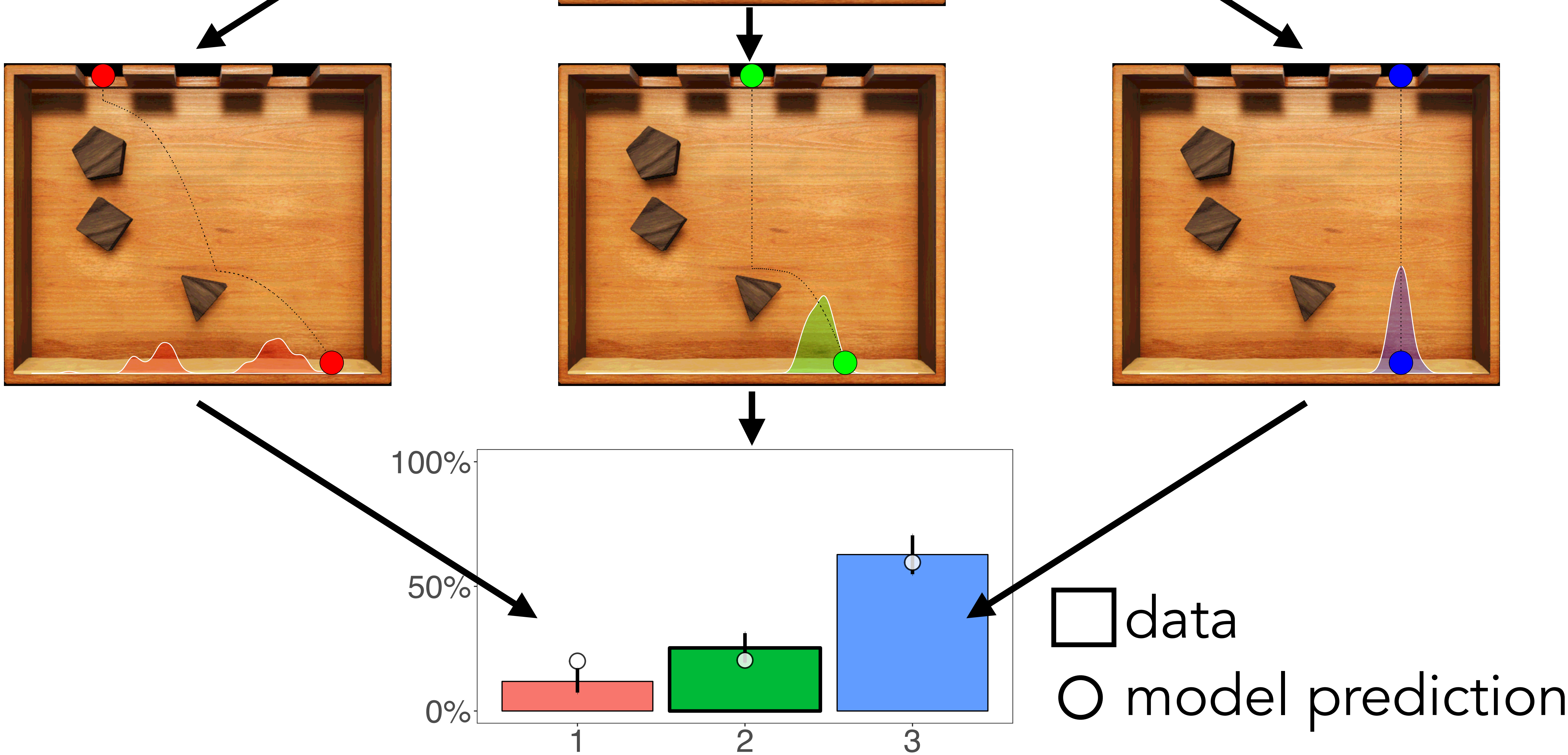
- drop the ball (with noise)
- look at the distance between the simulated ball and where the ball actually ended up
- convert this distance into a likelihood via a Gaussian distribution

Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

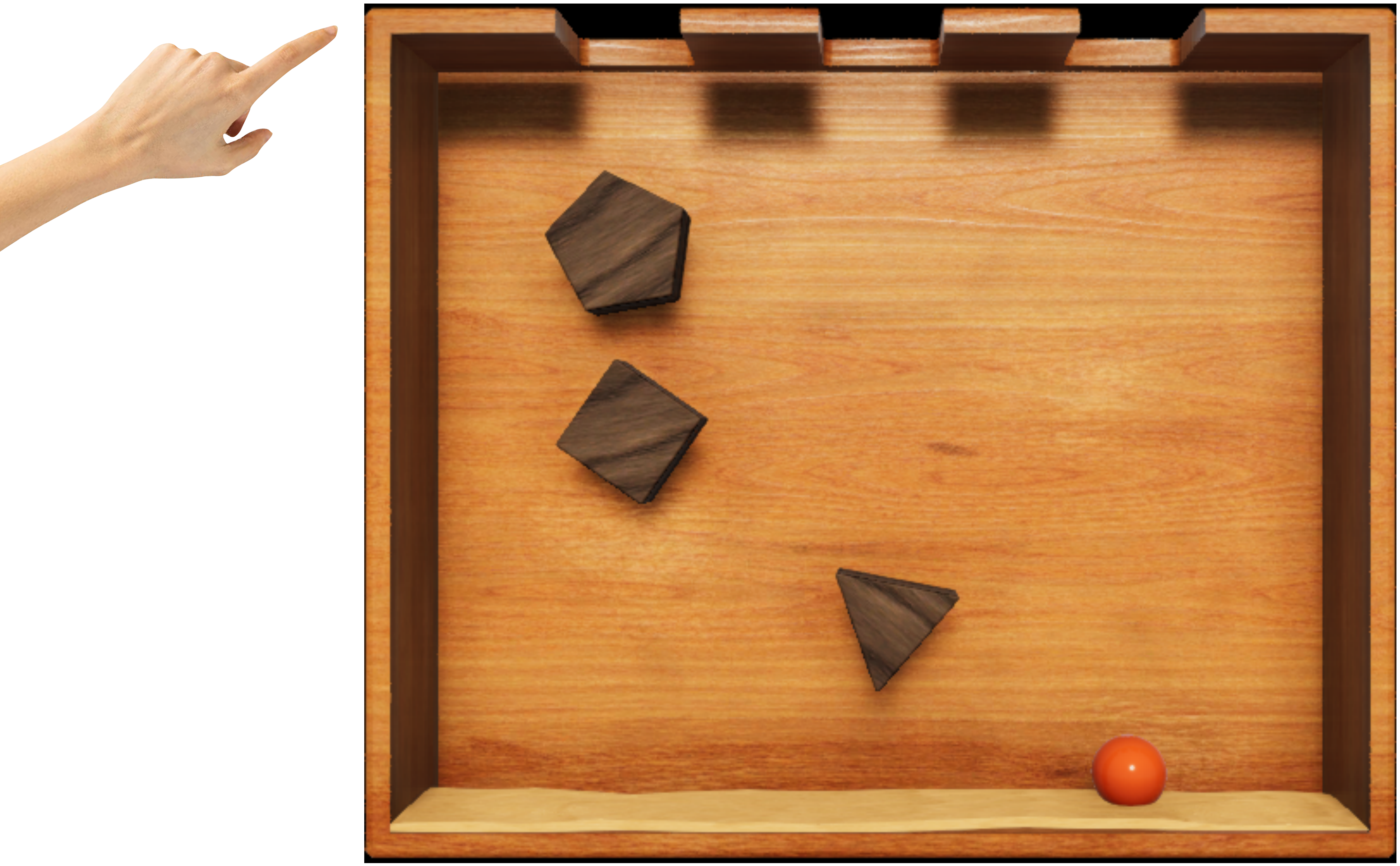
$$\exp(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2})$$





Inference: In which hole was the ball dropped?

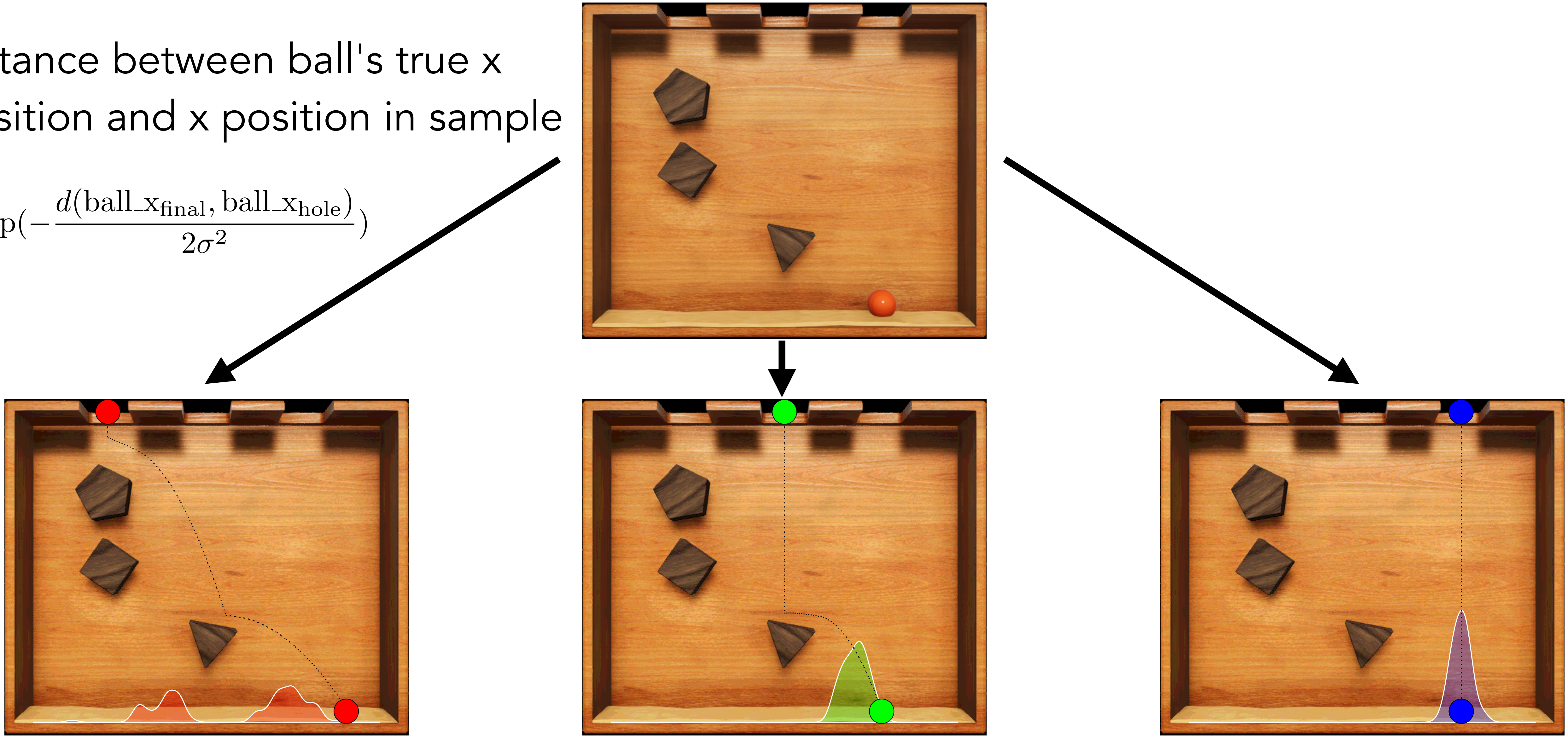
Inference: In which hole was the ball dropped?



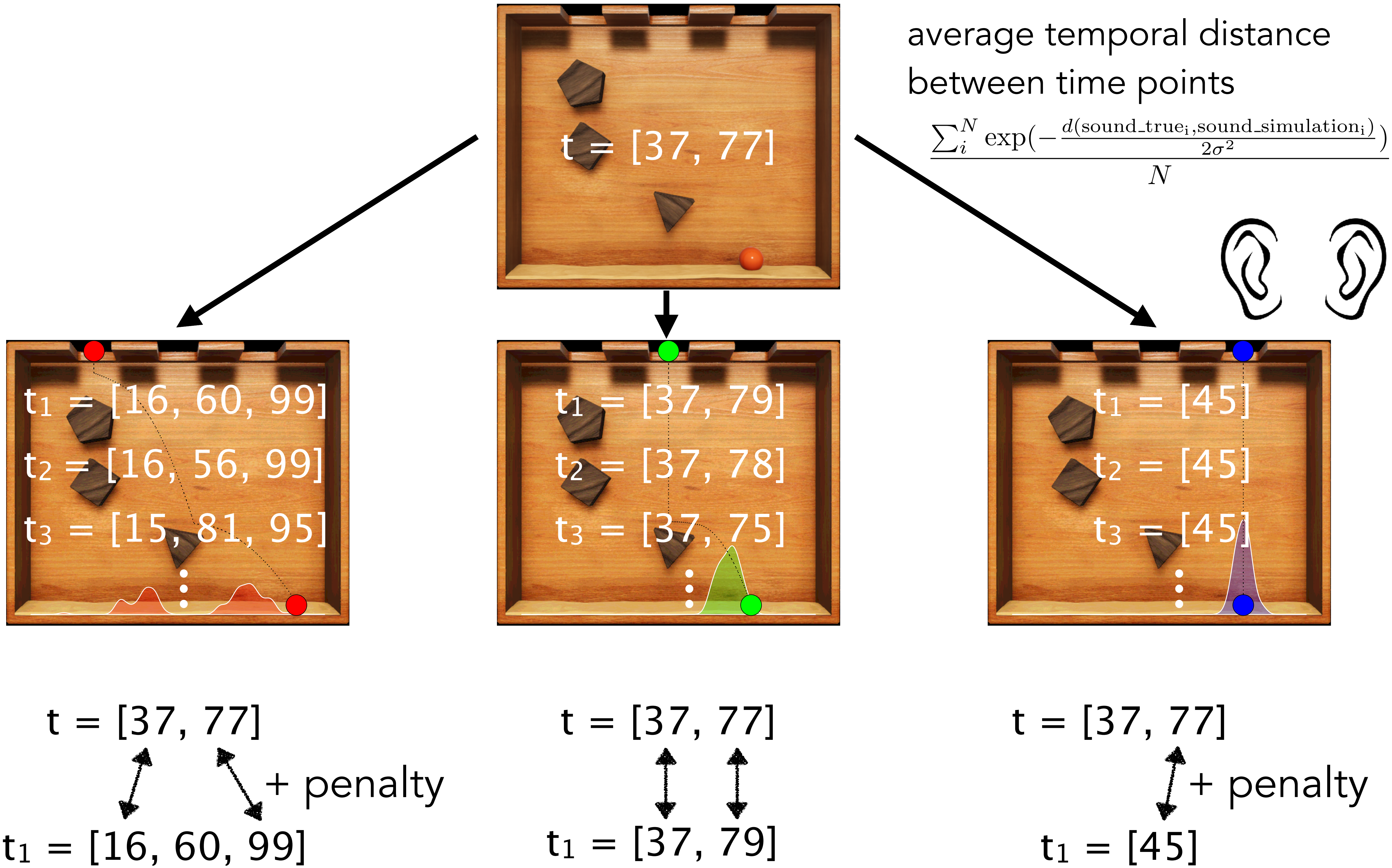
Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

$$\exp(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2})$$

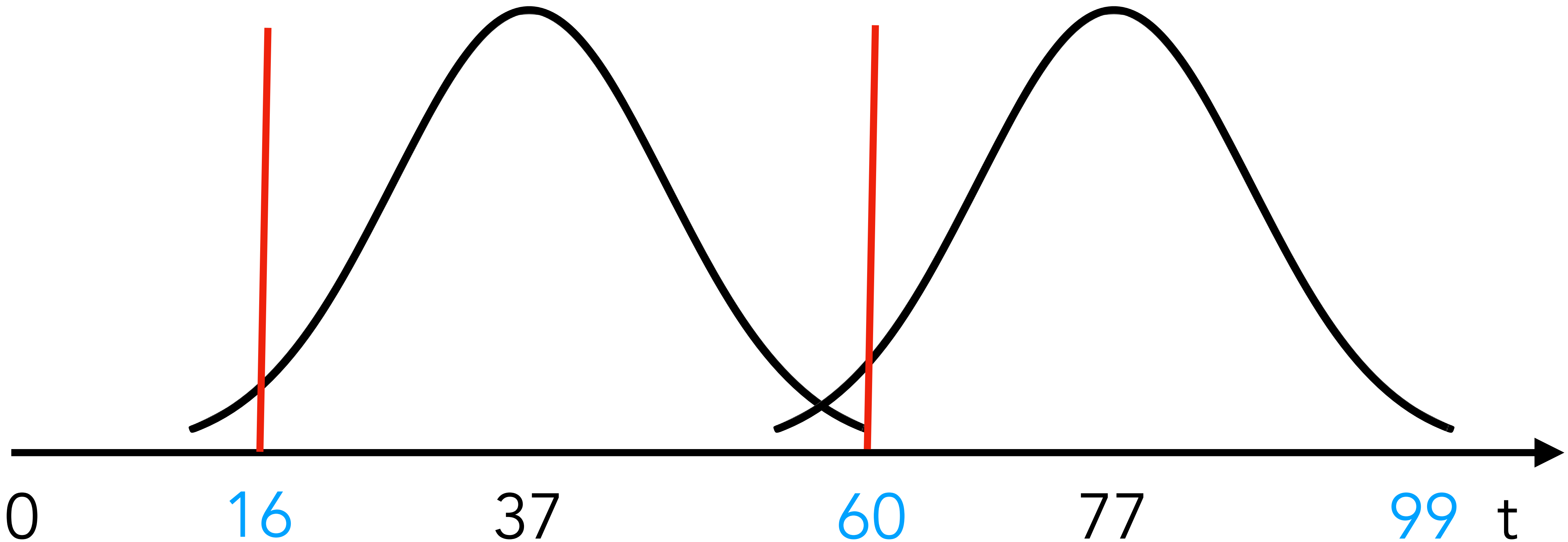


Inference: In which hole was the ball dropped?



Likelihood of a sequence of collision sounds

actual sounds $t = [37, 77]$



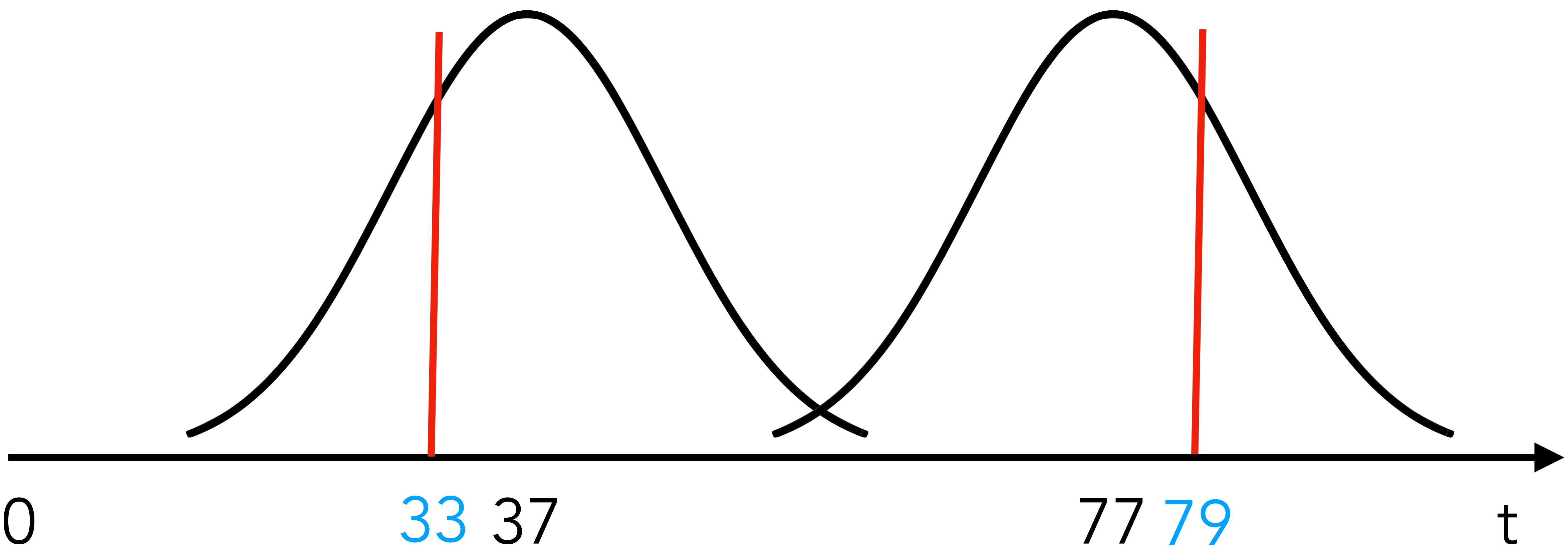
simulated sounds $t_1 = [16, 60, 99]$

fixed penalty

- determine which ones are closest to the actual sounds
- calculate their likelihood
- add a penalty for each additional (or missed) sound

Likelihood of a sequence of collision sounds

actual sounds $t = [37, 77]$



simulated sounds $t_1 = [33, 79]$

- determine which ones are closest to the actual sounds
- calculate their likelihood
- add a penalty for each additional (or missed) sound

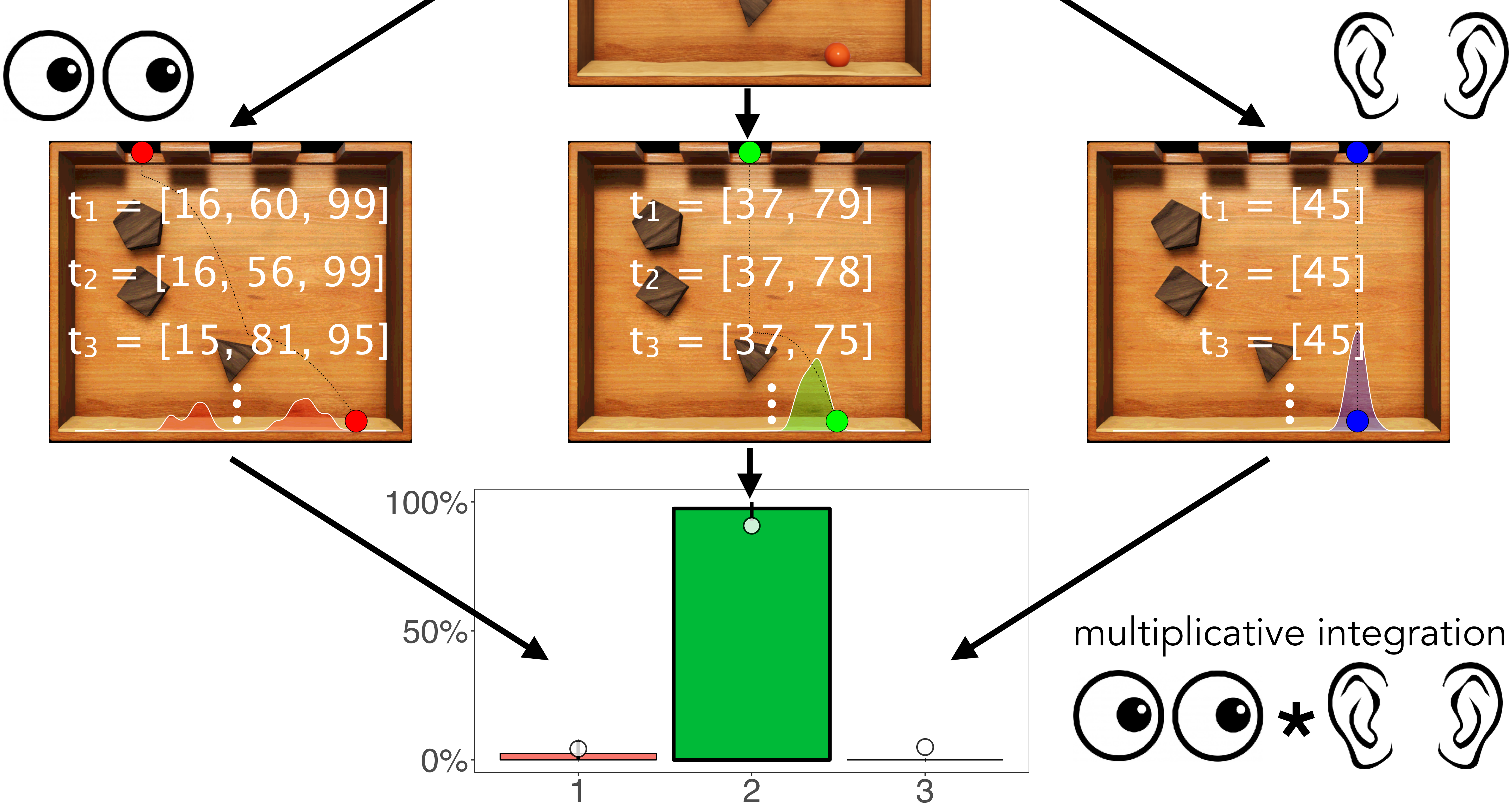
Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

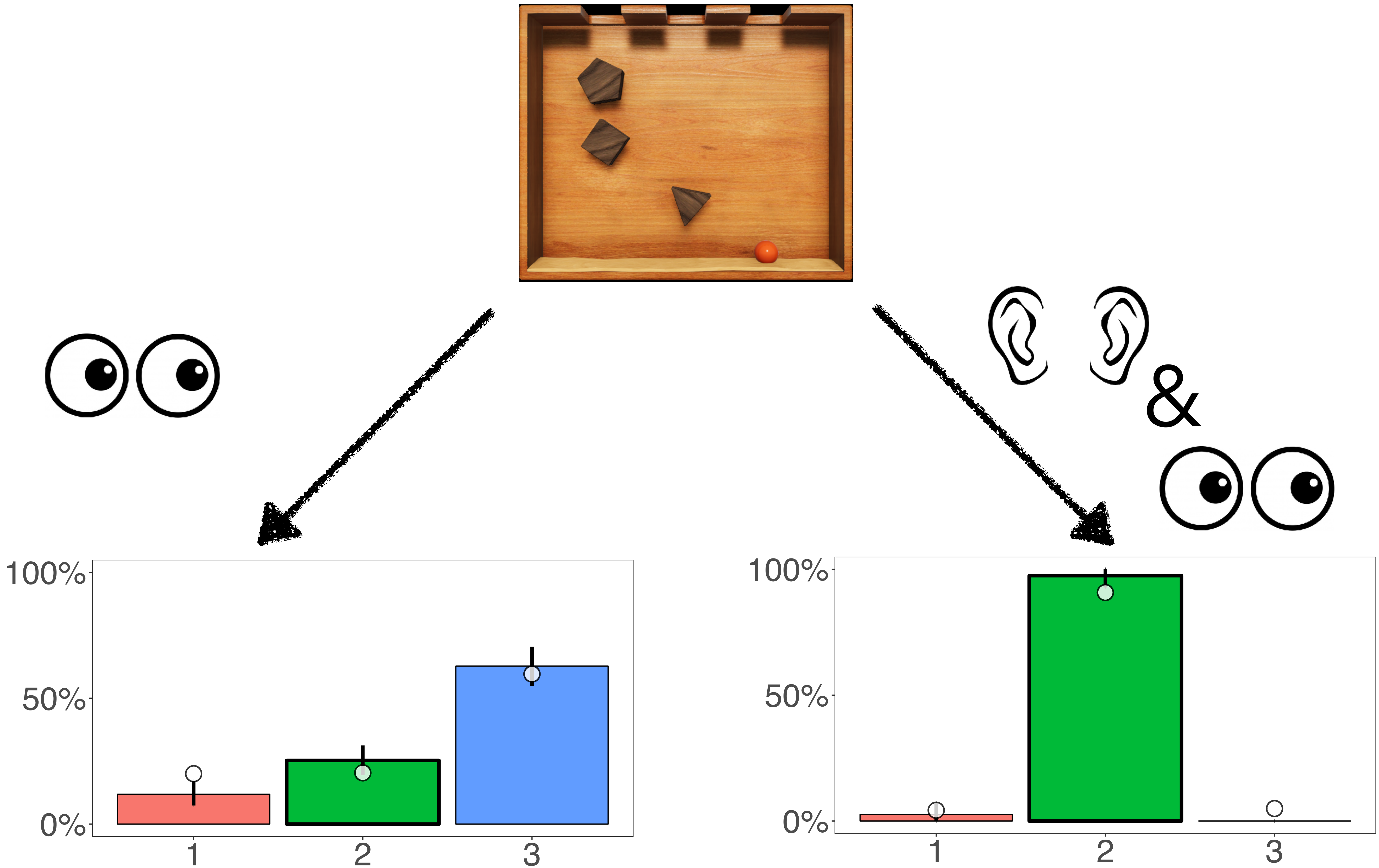
$$\exp(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2})$$

average temporal distance between time points

$$\frac{\sum_i^N \exp(-\frac{d(\text{sound\_true}_{\text{i}}, \text{sound\_simulation}_{\text{i}})}{2\sigma^2})}{N}$$



Inference: In which hole was the ball dropped?



Summary

- Quick Bayes recap
- Ingredients: likelihood, prior, inference
- Doing Bayesian data analysis
- Bayesian models of cognition

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!