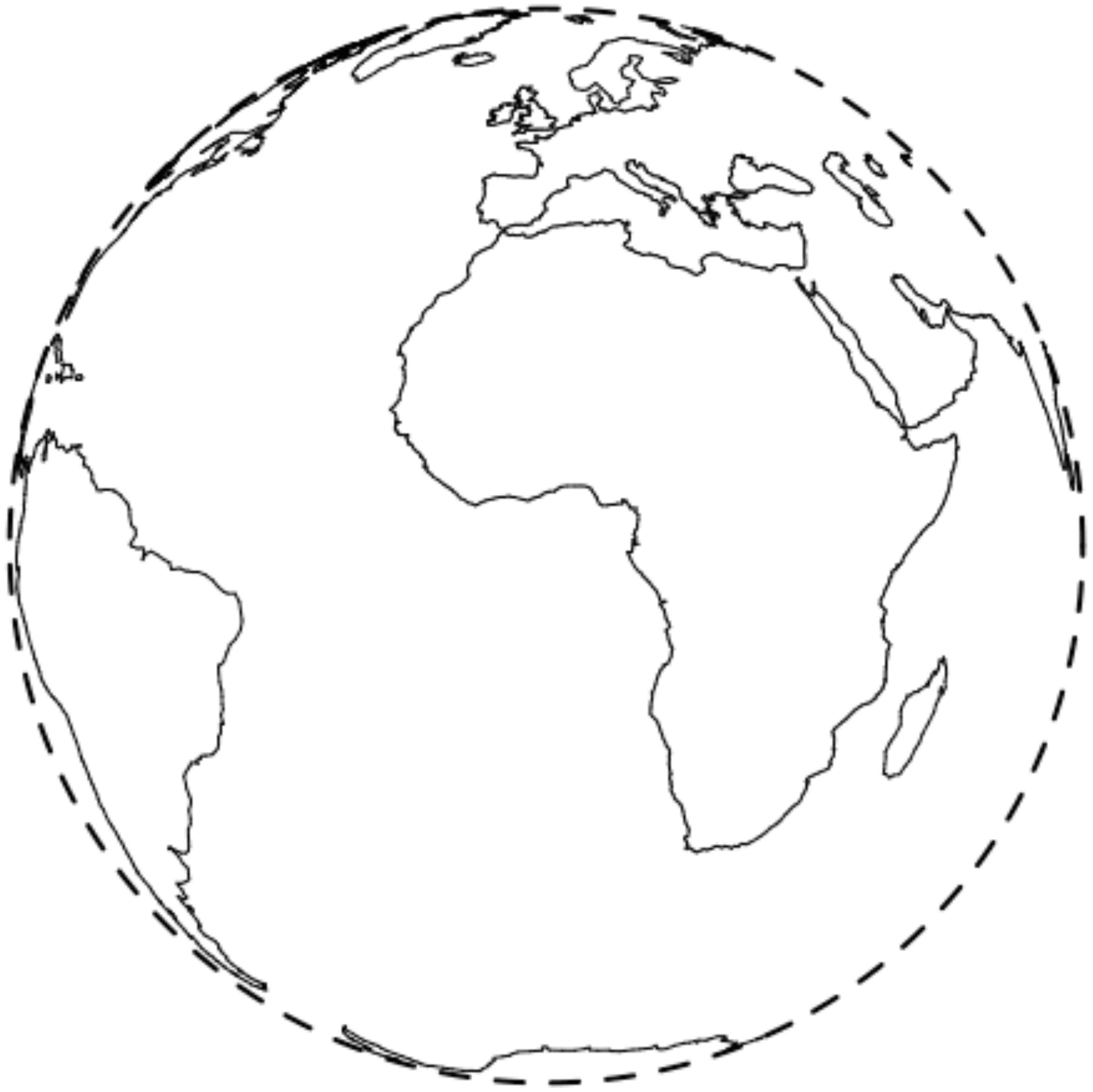
Statistical Rethinking

2022

02: Foundations of Bayesian Inference

What proportion of the surface is covered with water?



How should we use the sample?

How to produce a summary?

How to represent uncertainty?

Bayesian data analysis

For each possible explanation of the data,

Count all the ways data can happen.

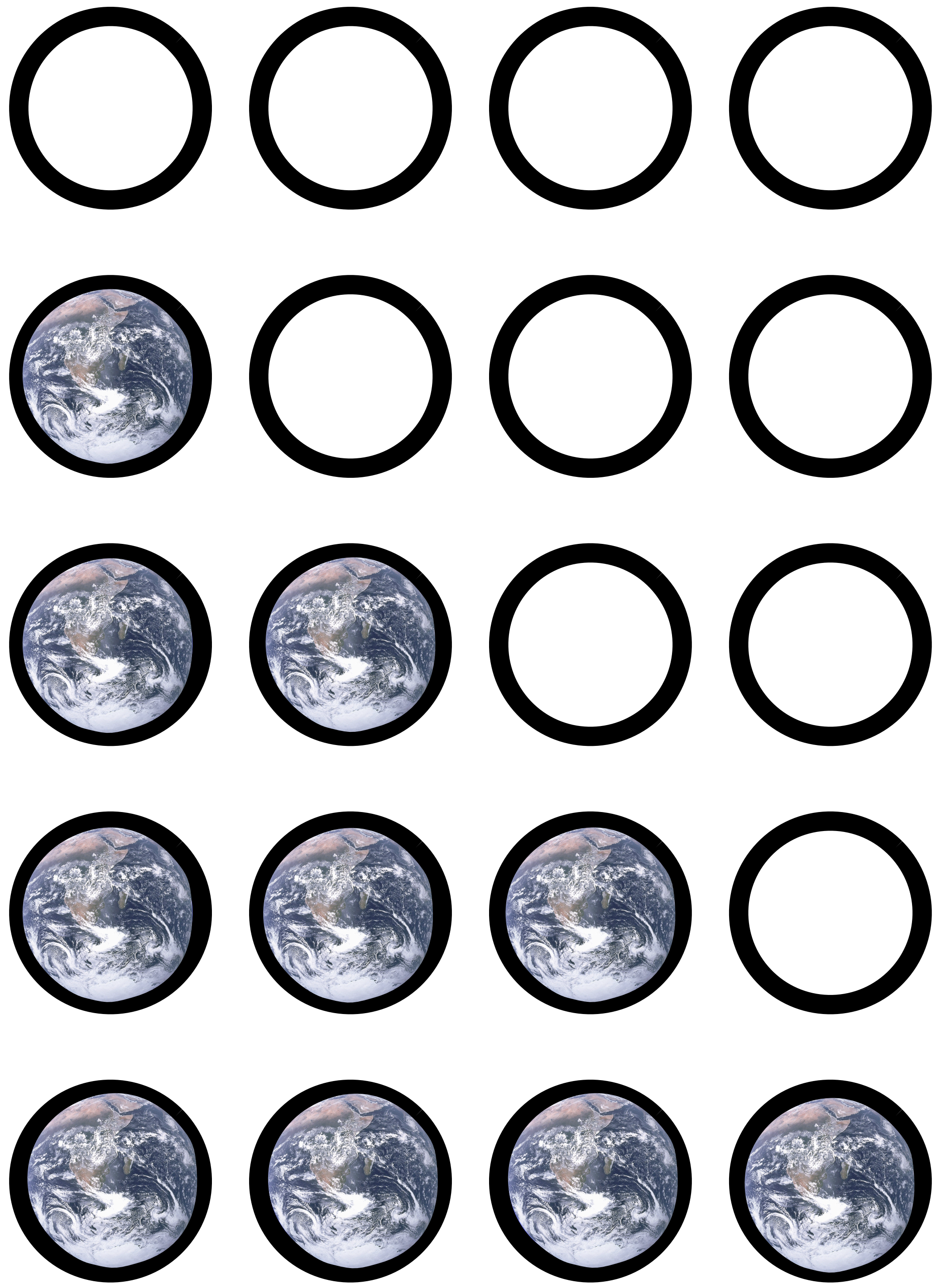
Explanations with more ways to produce the data are more plausible.

Garden of Forking Data

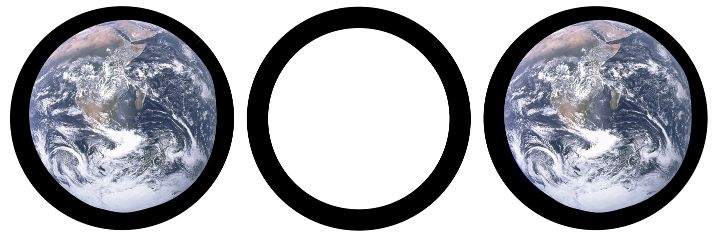
Contains 4 marbles

Possible contents:

(1)
(2)
(3)
(4)
(5)



Observe:



Garden of Forking Data

Contains 4 marbles

Possible contents:

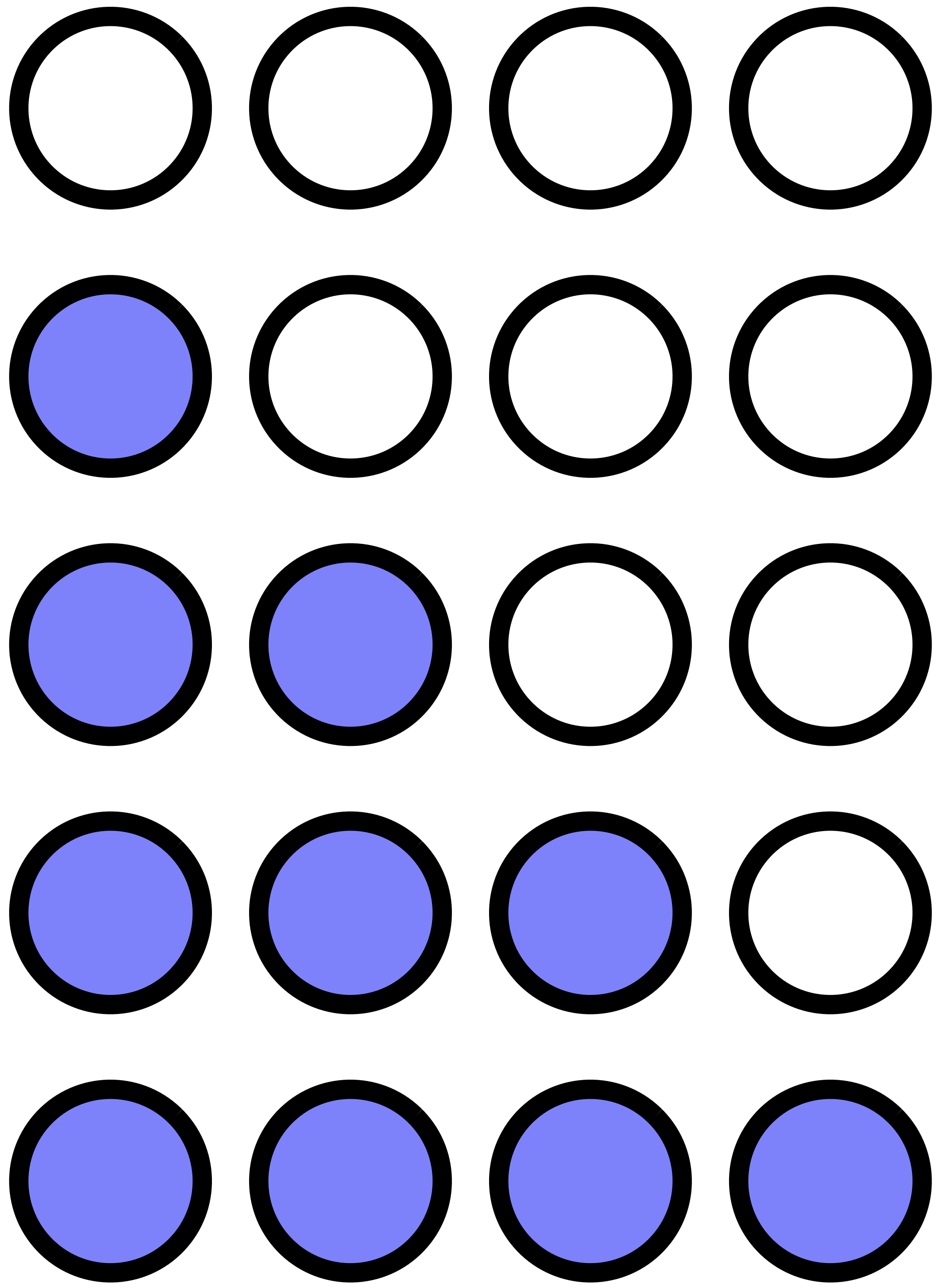
(1)

(2)

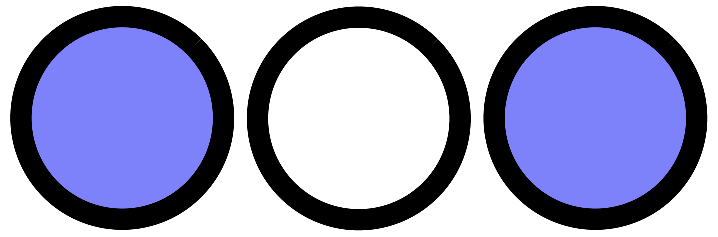
(3)

(4)

(5)



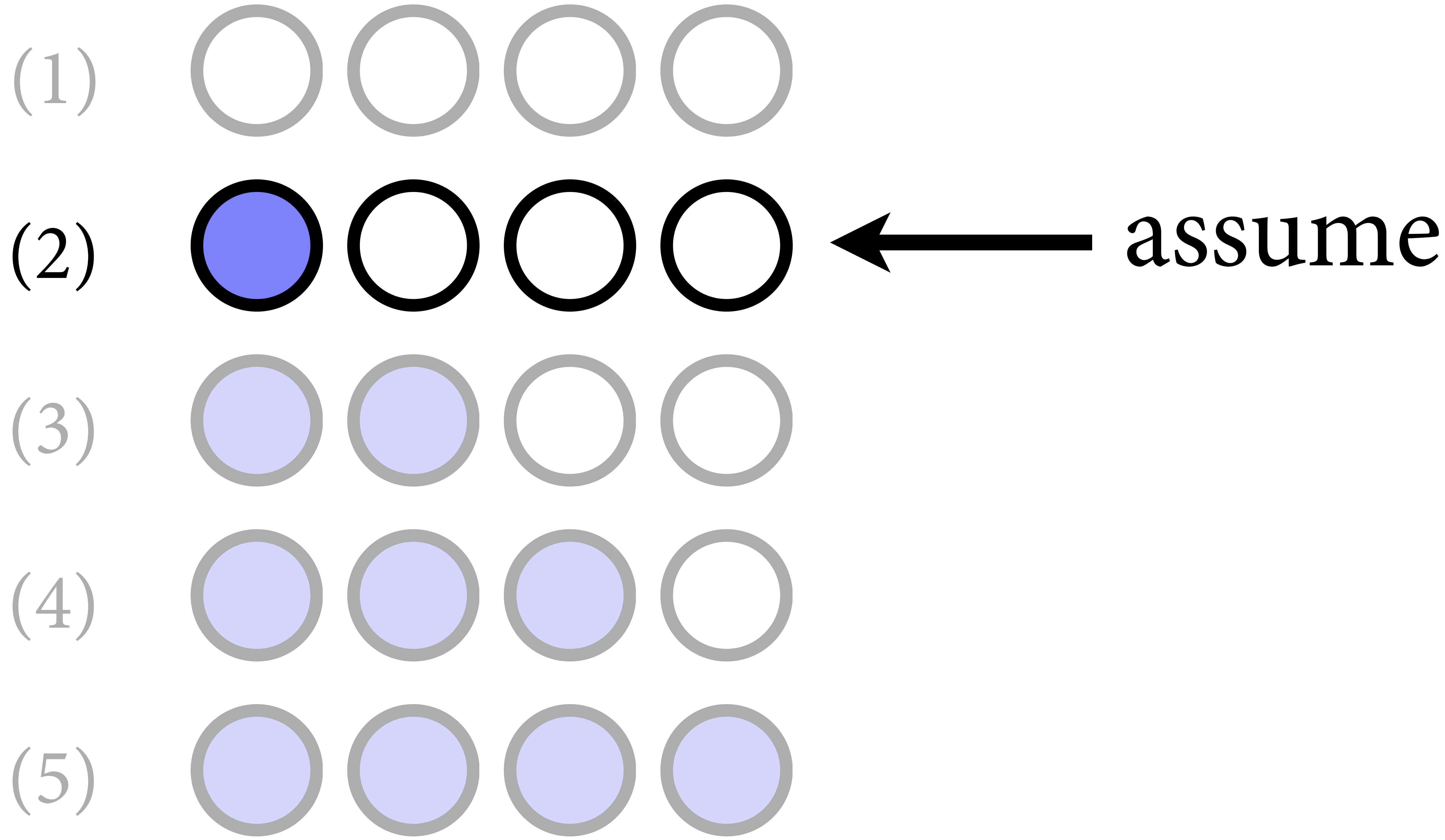
Observe:



Garden of Forking Data

Contains 4 marbles

Possible contents:



How many ways to observe ?

First Possibility

Figure 2.2

Second Possibility

Figure 2.2

Third Possibility

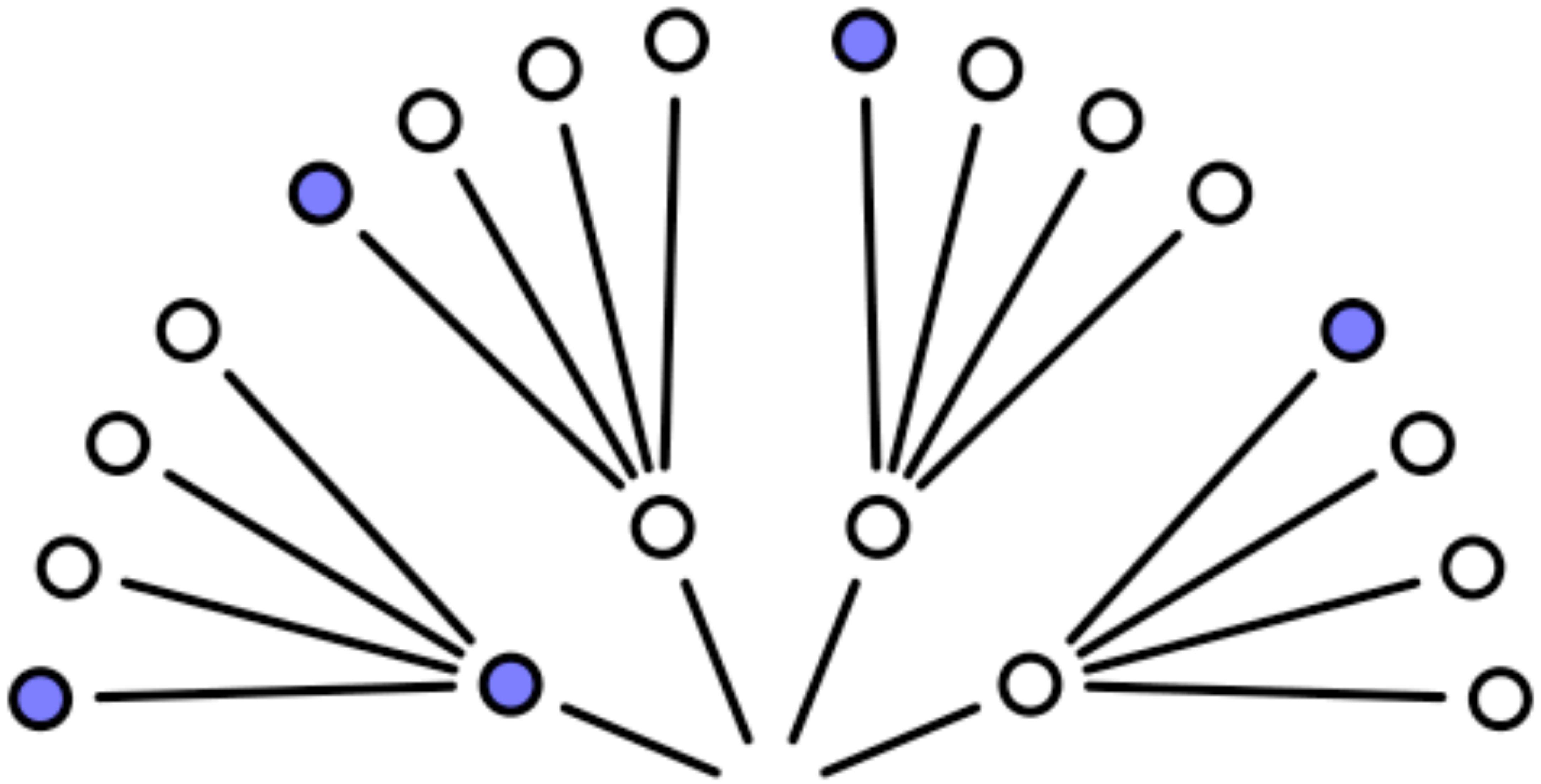


Figure 2.2

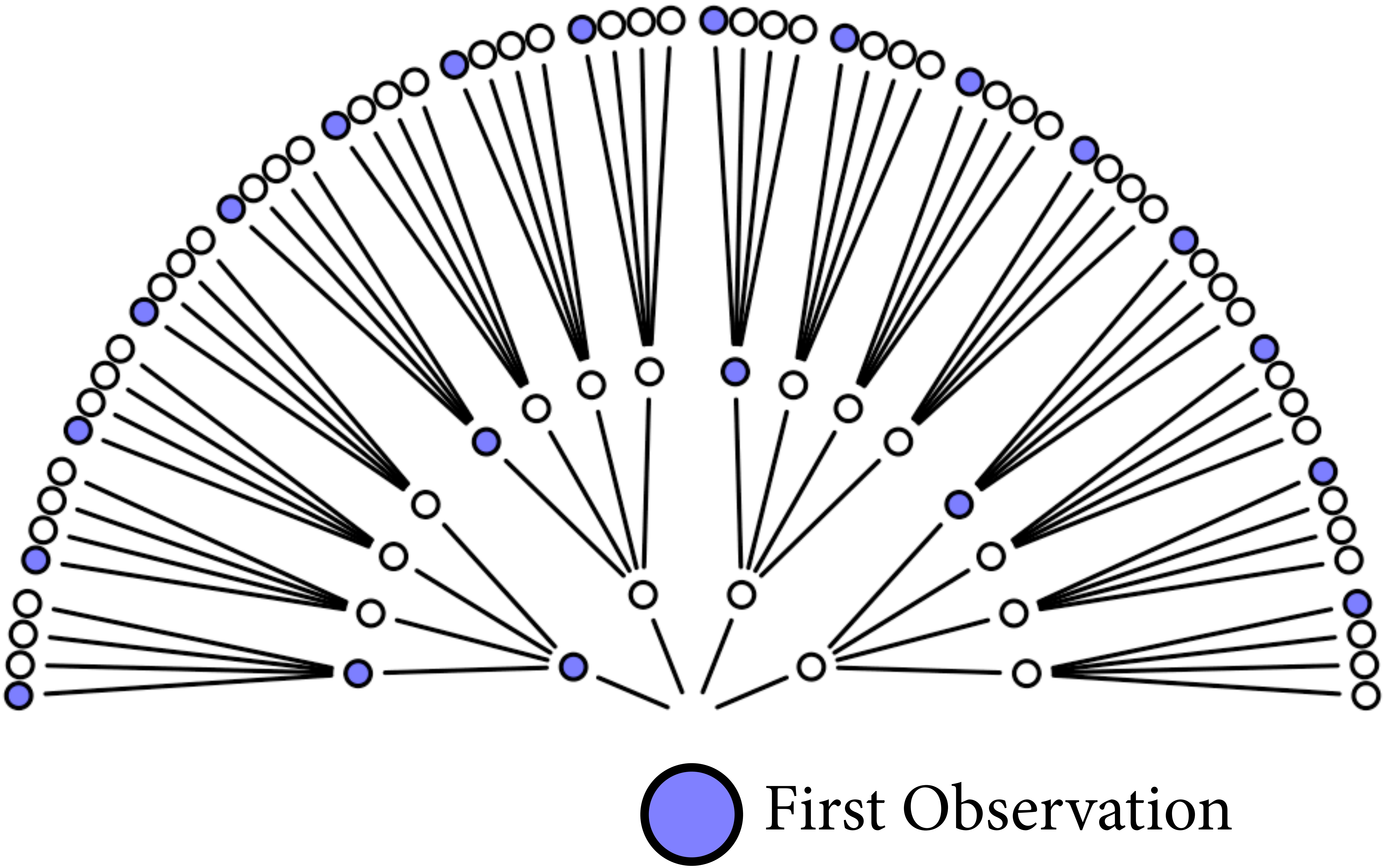


Figure 2.2

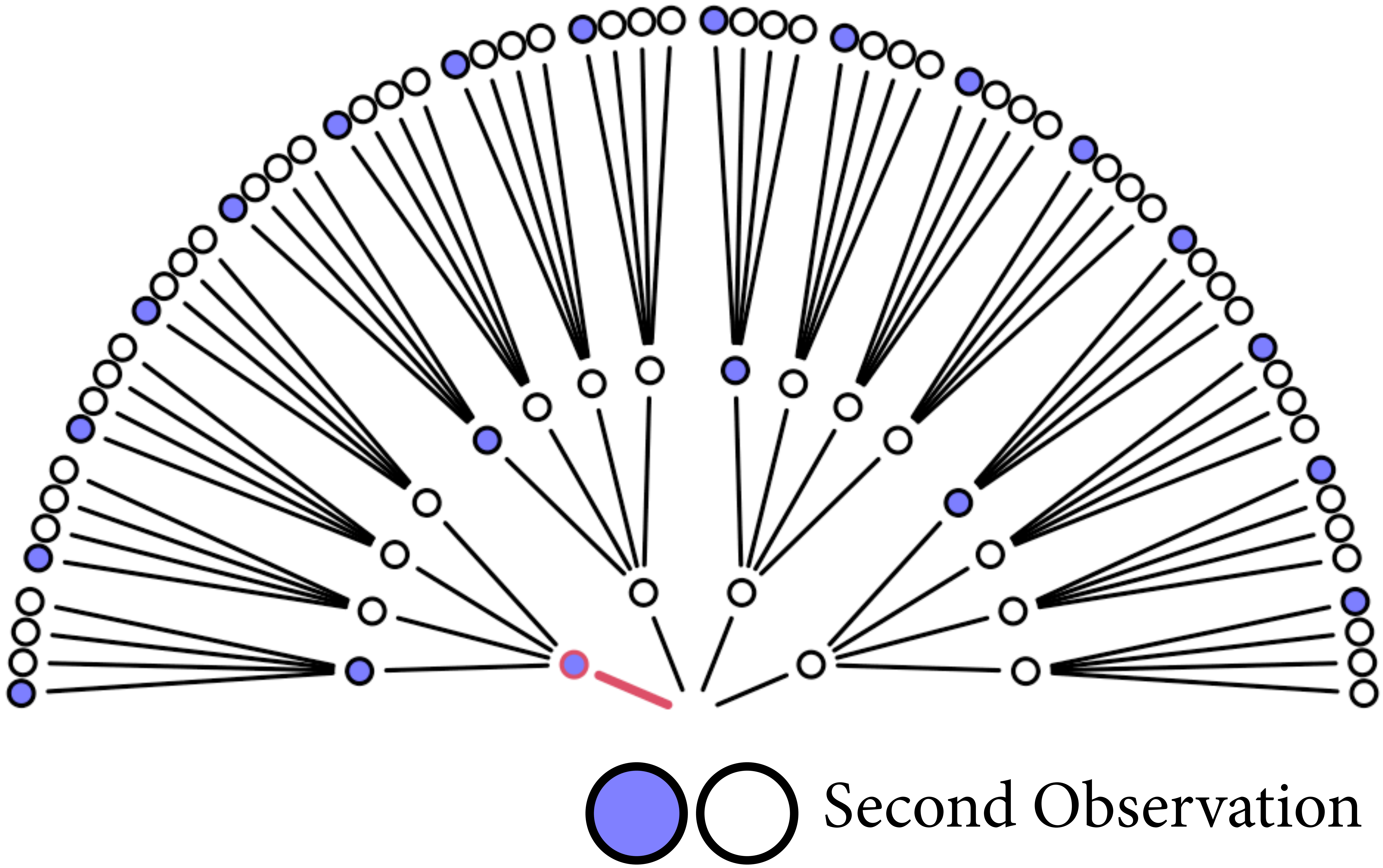


Figure 2.2

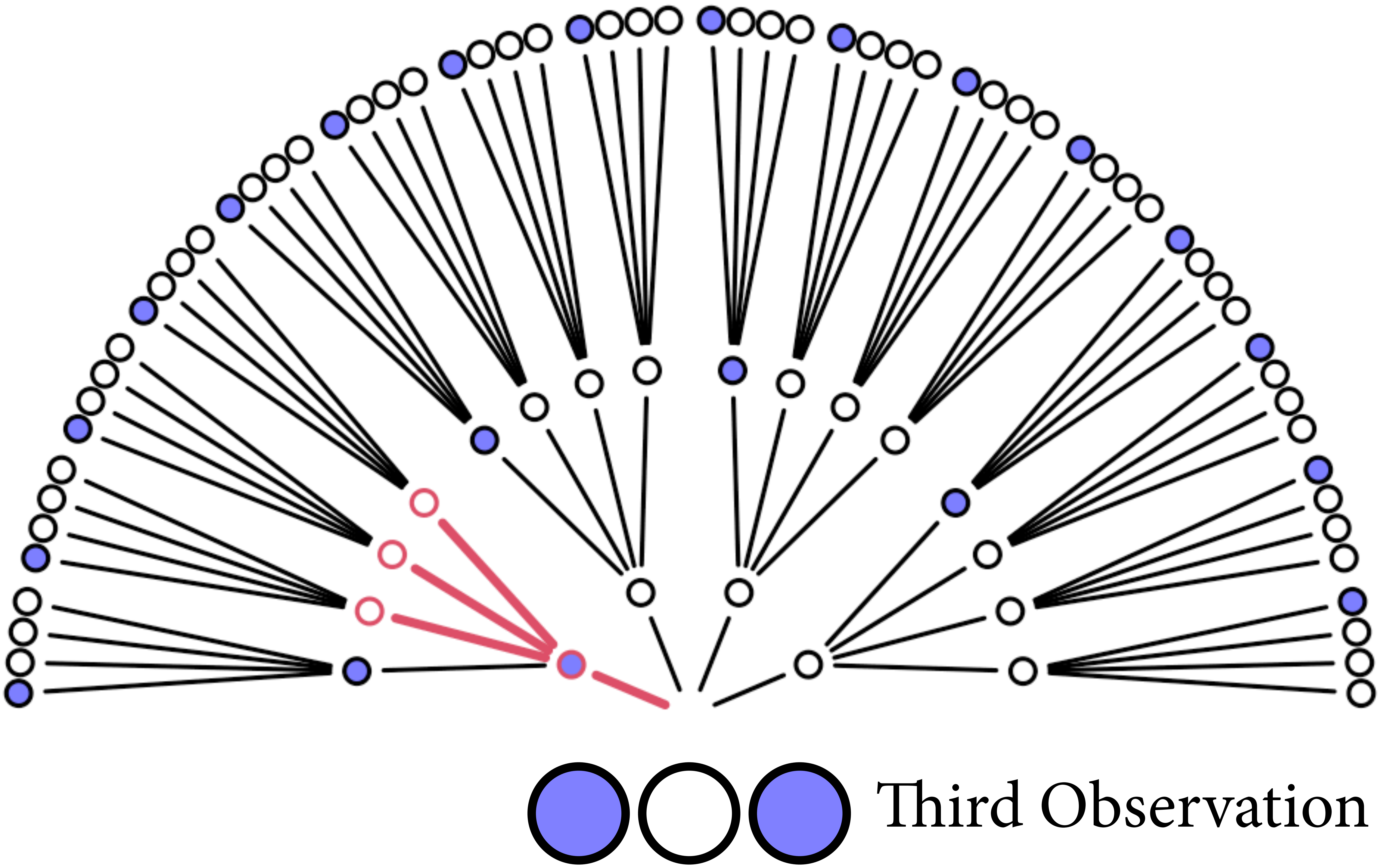
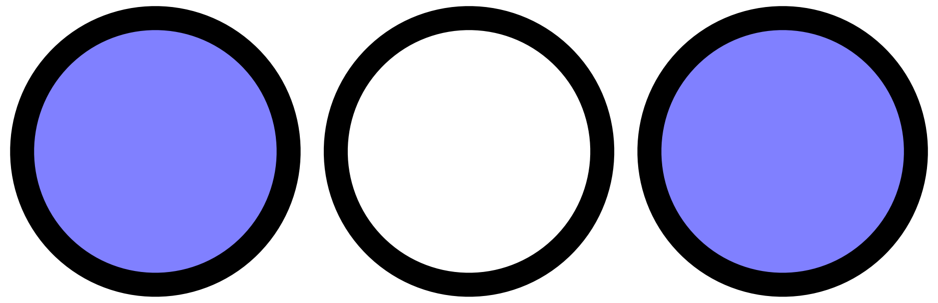
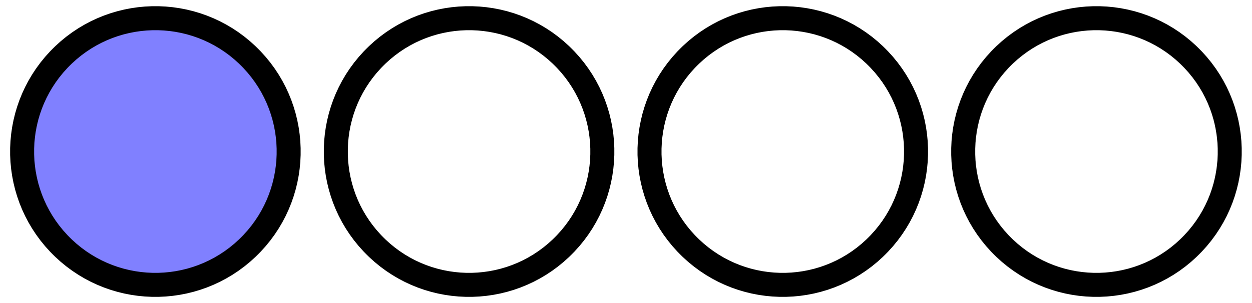


Figure 2.2

3 Ways to see



if the bag contains



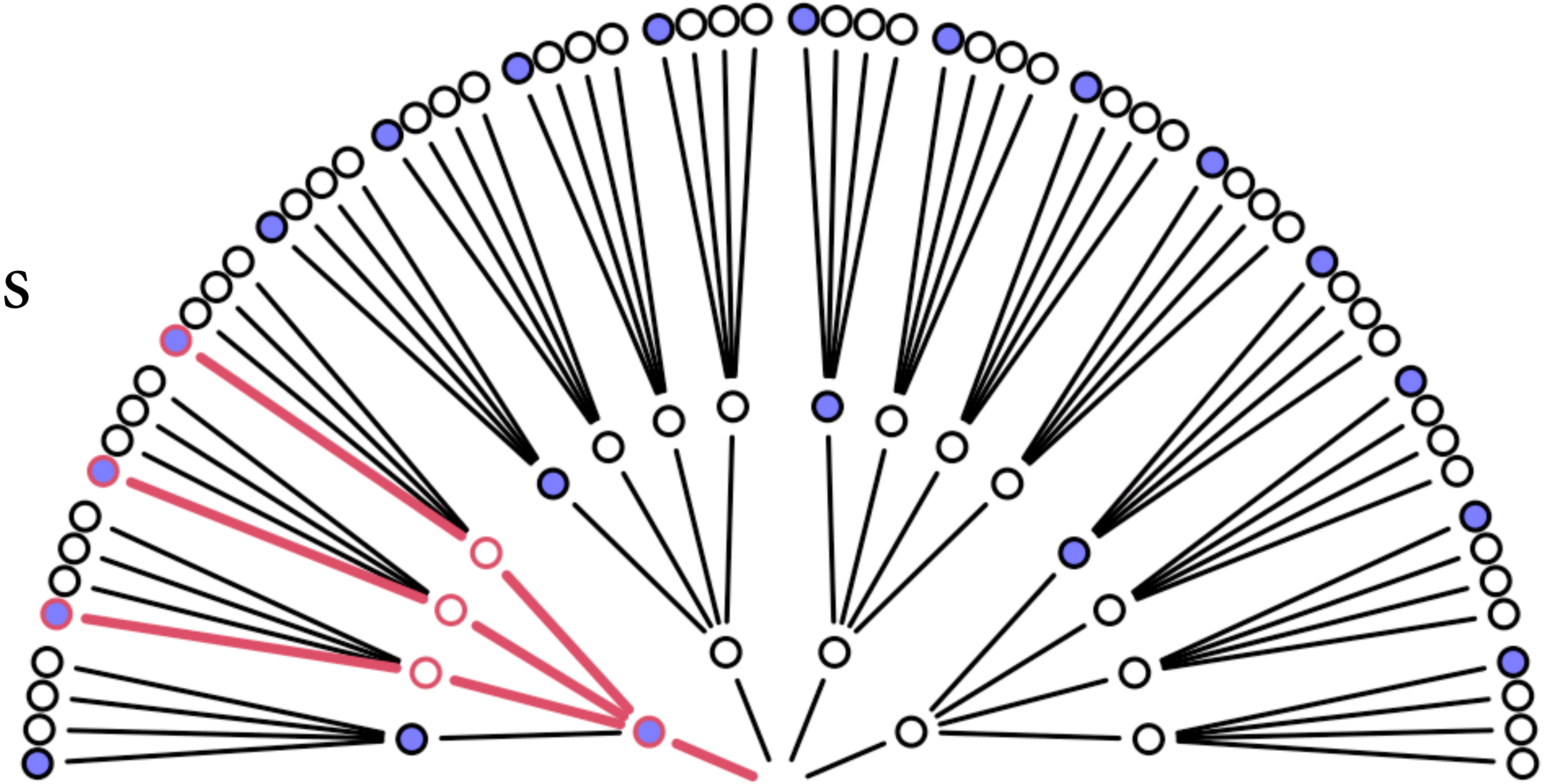
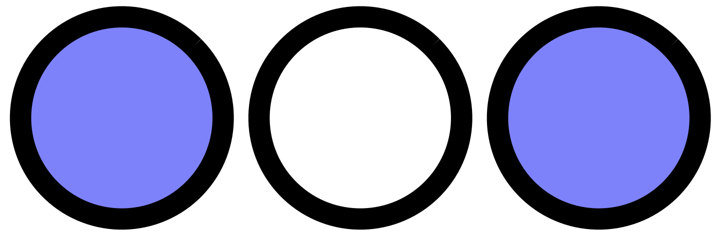
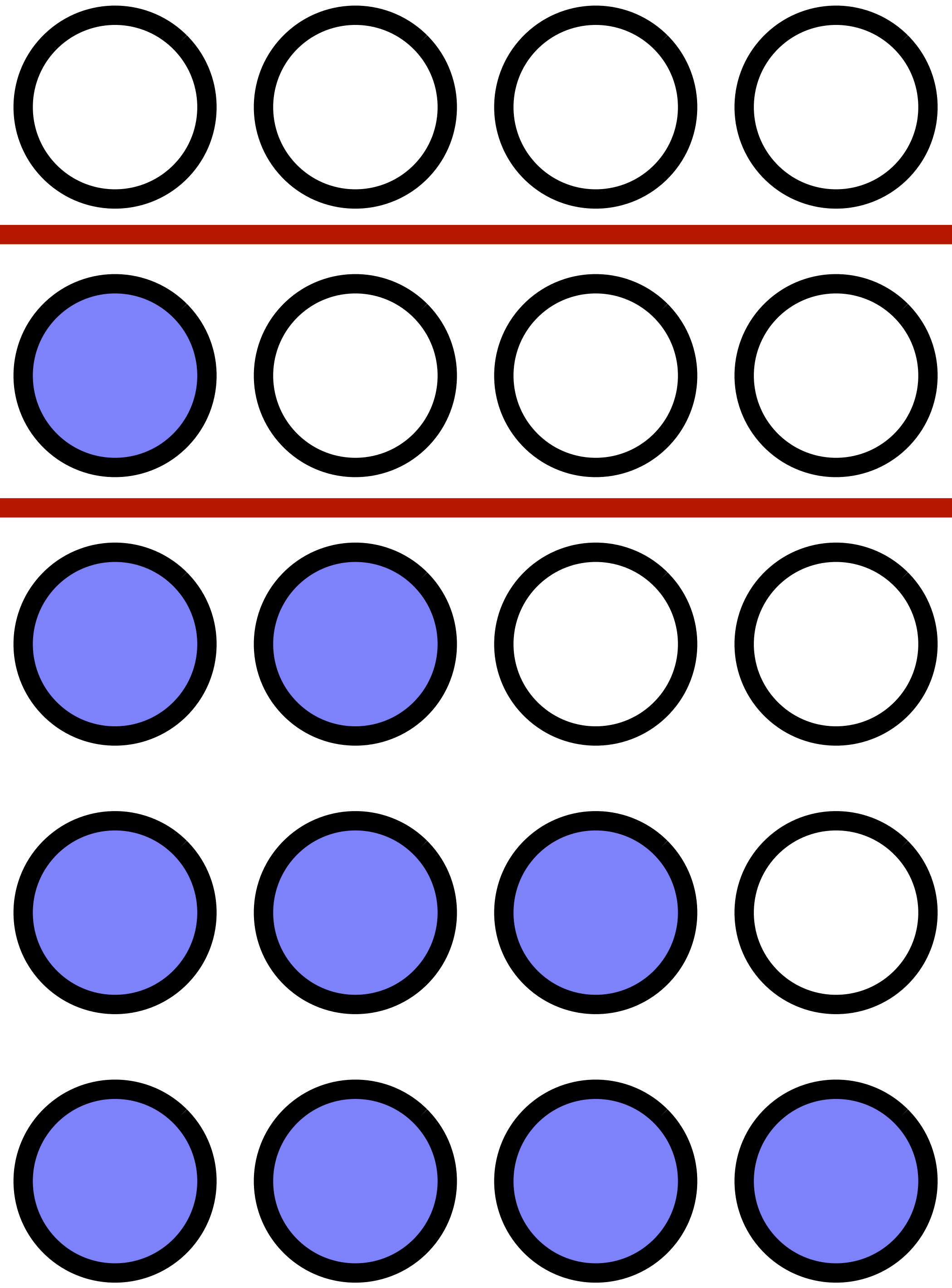


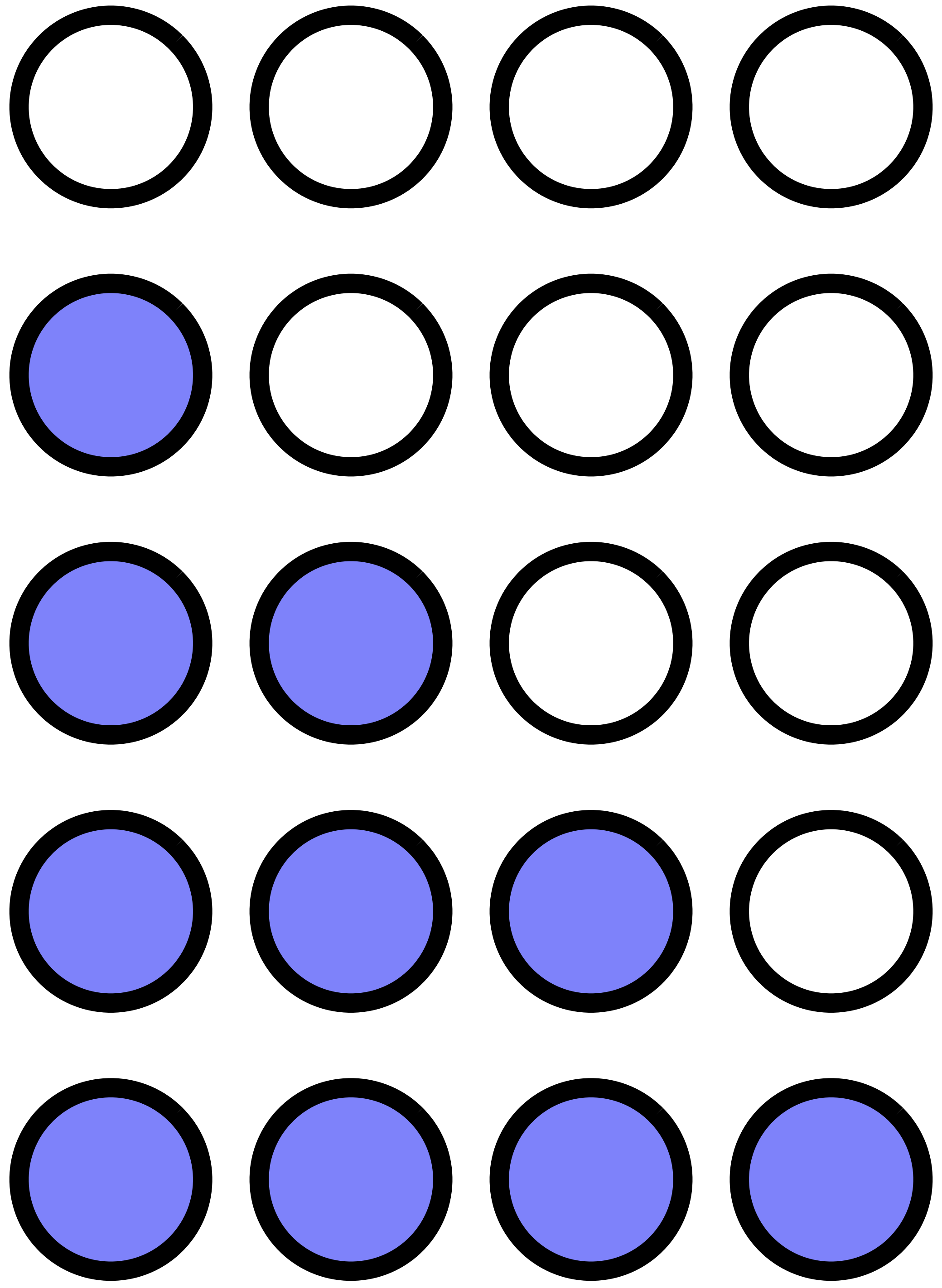
Figure 2.2

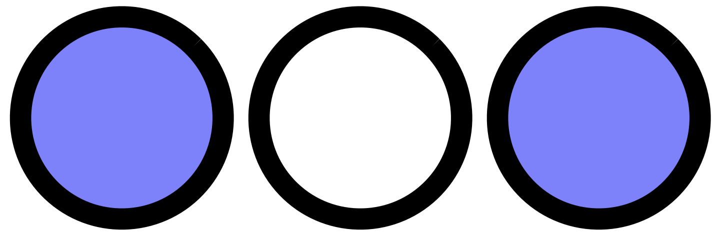
Garden of Forking Data

Possible contents:		Ways to produce 
(1)		?
(2)		3
(3)		?
(4)		?
(5)		?

Garden of Forking Data

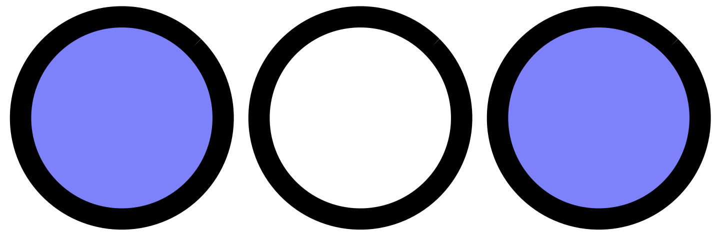
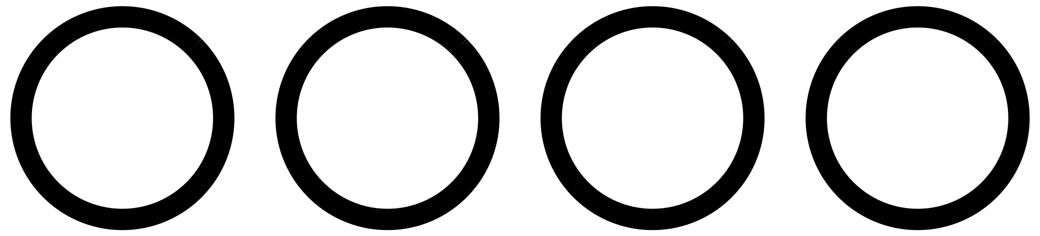
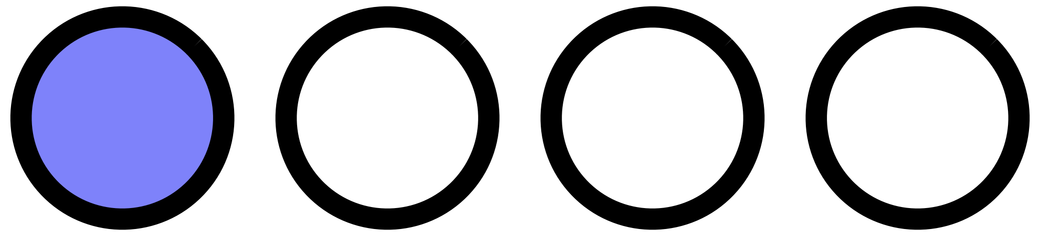
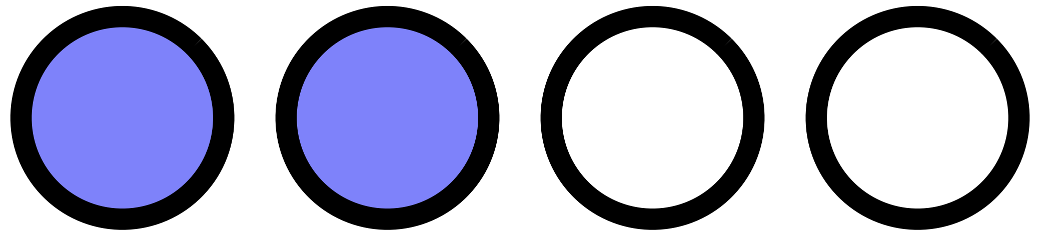
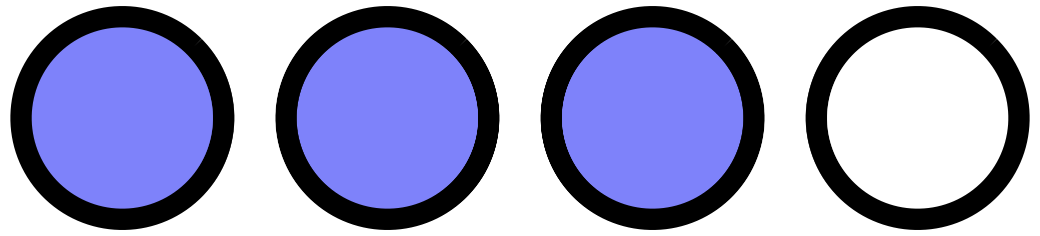
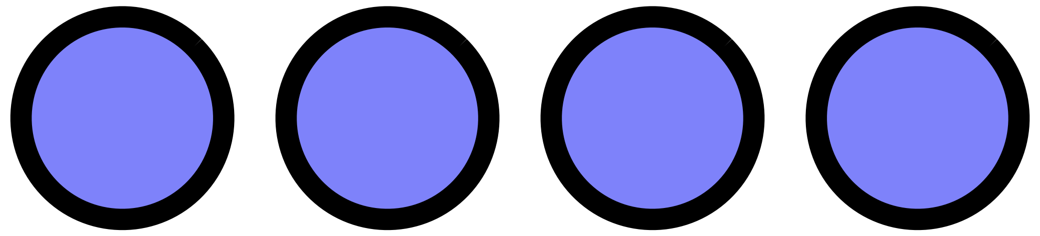
Possible contents:

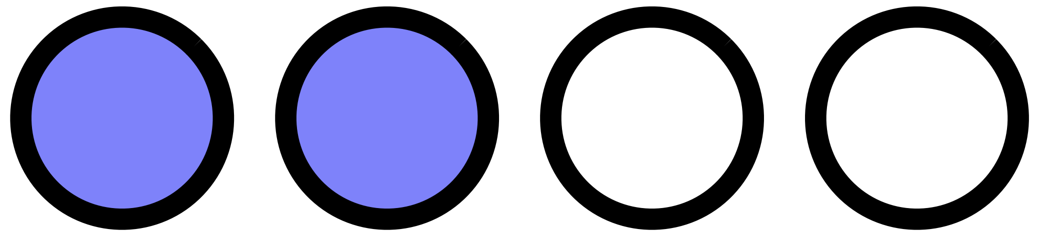


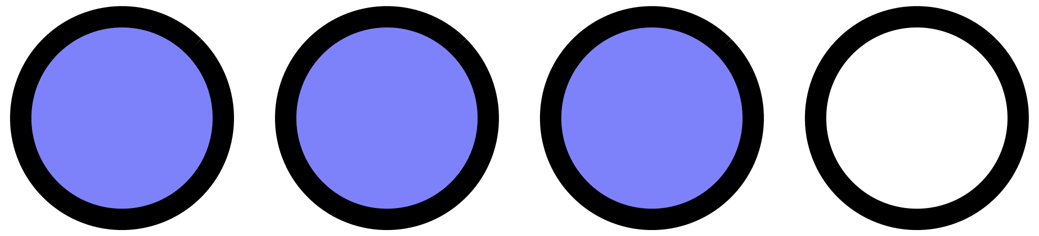
Ways to produce 

(1)	0
(2)	3
(3)	?
(4)	?
(5)	?

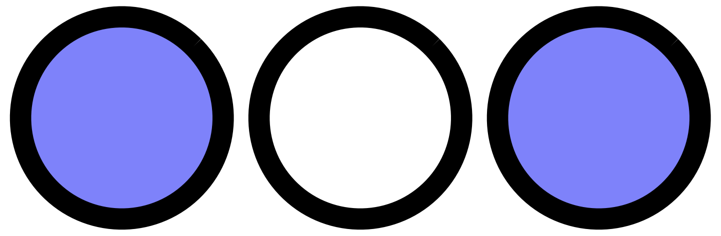
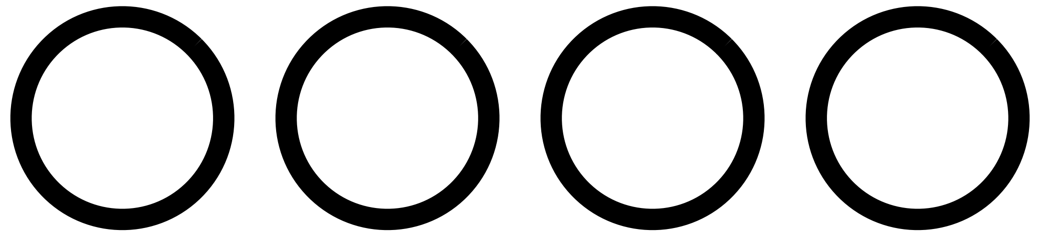
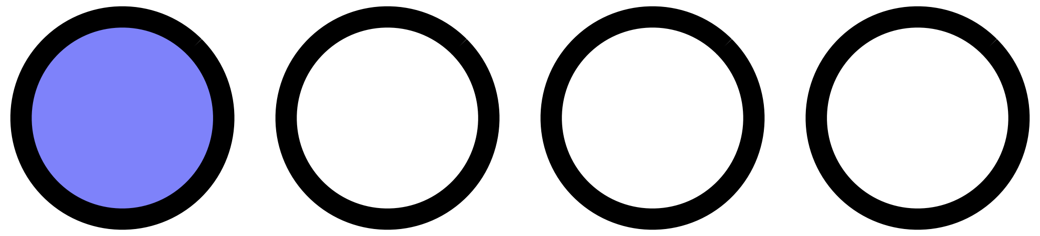
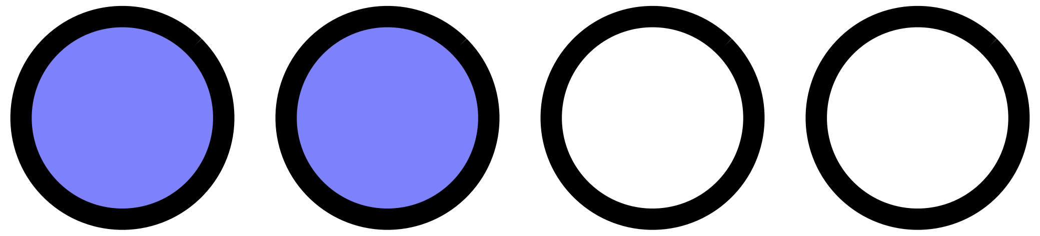
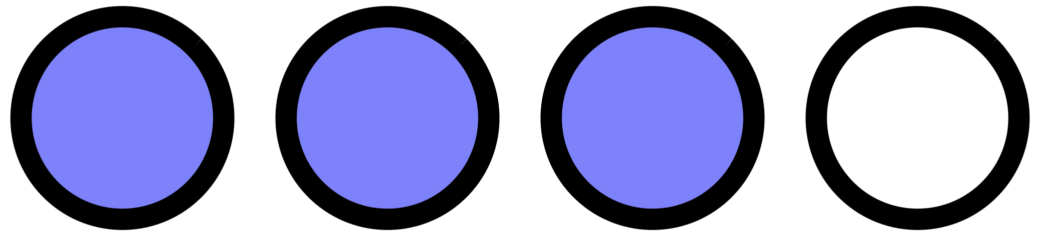
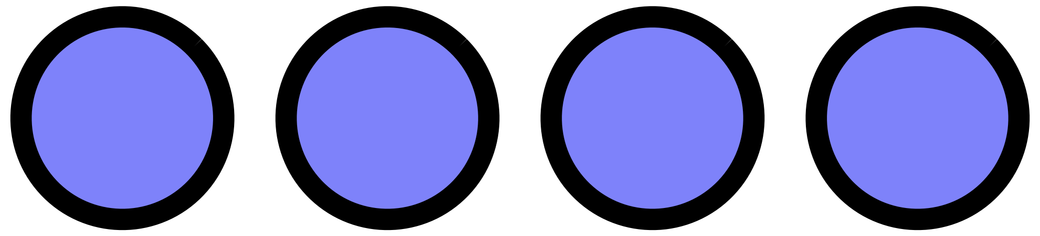
Garden of Forking Data

Possible contents:		Ways to produce 
(1)		0
(2)		3
(3)		?
(4)		?
(5)		0

(3) 

(4) 

Garden of Forking Data

	Possible contents:	Ways to produce 
(1)		0
(2)		3
(3)		8
(4)		9
(5)		0

Counts to plausibility

Unglamorous basis of applied probability:
Things that can happen more ways are more plausible.

Possible composition
[○○○○]
[●○○○]
[●●○○]
[●●●○]
[●●●●]

Counts to plausibility

Unglamorous basis of applied probability:
Things that can happen more ways are more plausible.

Possible composition	p	ways to produce data
[○○○○]	0	0
[●○○○]	0.25	3
[●●○○]	0.5	8
[●●●○]	0.75	9
[●●●●]	1	0

Counts to plausibility

Unglamorous basis of applied probability:
Things that can happen more ways are more plausible.

Possible composition	p	ways to produce data	plausibility
[○○○○]	0	0	0
[●○○○]	0.25	3	0.15
[●●○○]	0.5	8	0.40
[●●●○]	0.75	9	0.45
[●●●●]	1	0	0

Counts to plausibility

Possible composition	p	ways to produce data	plausibility
[○○○○]	0	0	0
[●○○○]	0.25	3	0.15
[●●○○]	0.5	8	0.40
[●●●○]	0.75	9	0.45
[●●●●]	1	0	0

```
ways <- c( 3 , 8 , 9 )
ways/sum(ways)
```

R code 2.1

```
[1] 0.15 0.40 0.45
```

Updating

Another draw from the bag: ●

Conjecture
[○○○○]
[●○○○]
[●●○○]
[●●●○]
[●●●●]

Updating

Another draw from the bag: ●

Conjecture	Ways to produce ●
[○○○○]	0
[●○○○]	1
[●●○○]	2
[●●●○]	3
[●●●●]	4

Updating

Another draw from the bag: ●

Conjecture	Ways to produce ●	Previous counts
[○○○○]	0	0
[●○○○]	1	3
[●●○○]	2	8
[●●●○]	3	9
[●●●●]	4	0

Updating

Another draw from the bag: ●

Conjecture	Ways to produce ●	Previous counts	New count
[○○○○]	0	0	$0 \times 0 = 0$
[●○○○]	1	3	$3 \times 1 = 3$
[●●○○]	2	8	$8 \times 2 = 16$
[●●●○]	3	9	$9 \times 3 = 27$
[●●●●]	4	0	$0 \times 4 = 0$

Bayesian updating

The rules:

1. State a causal model for how the observations arise, given each possible explanation

2. Count ways data could arise for each explanation

3. Relative plausibility is relative value from (2)

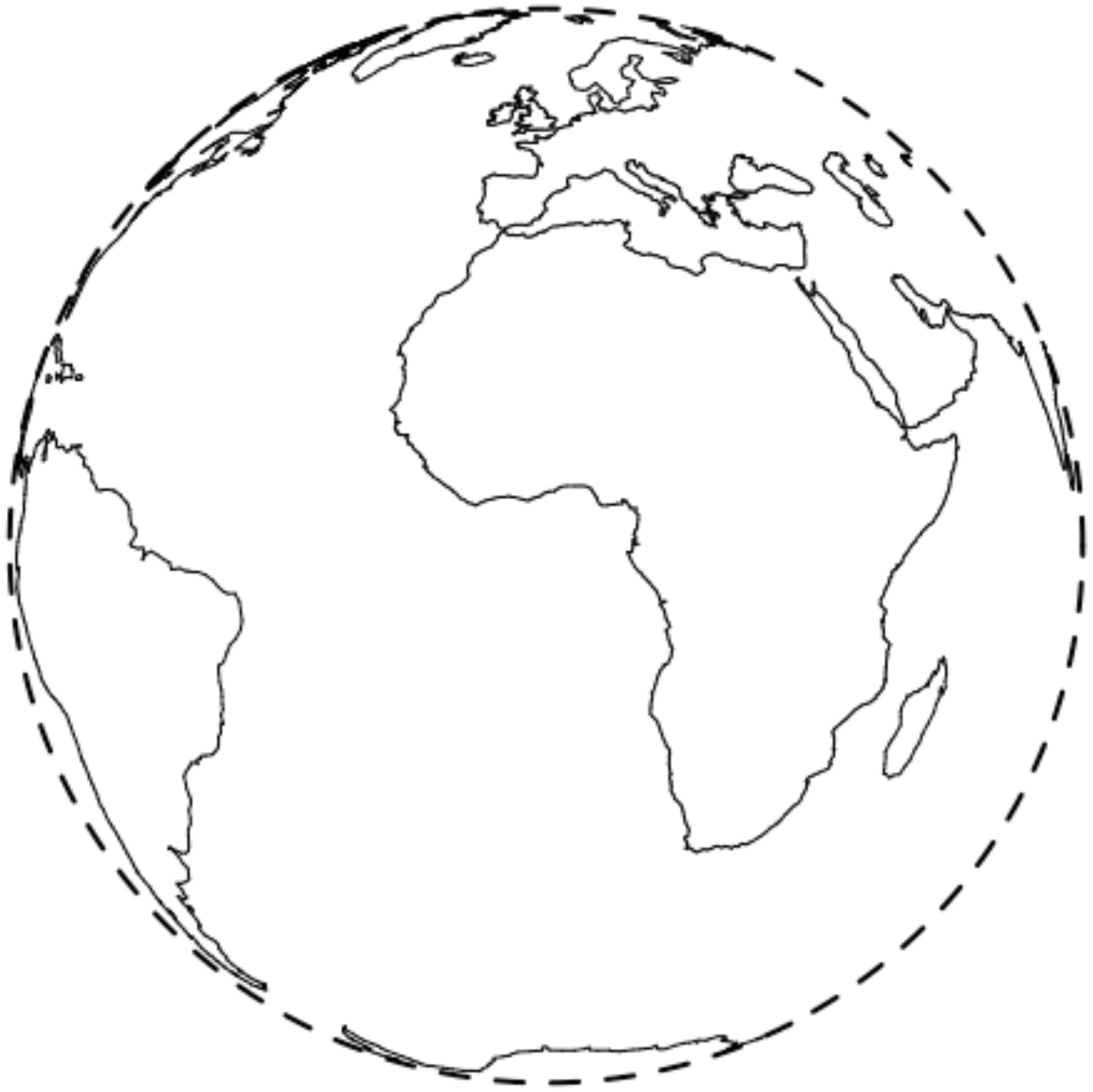
Globe of Forking Water

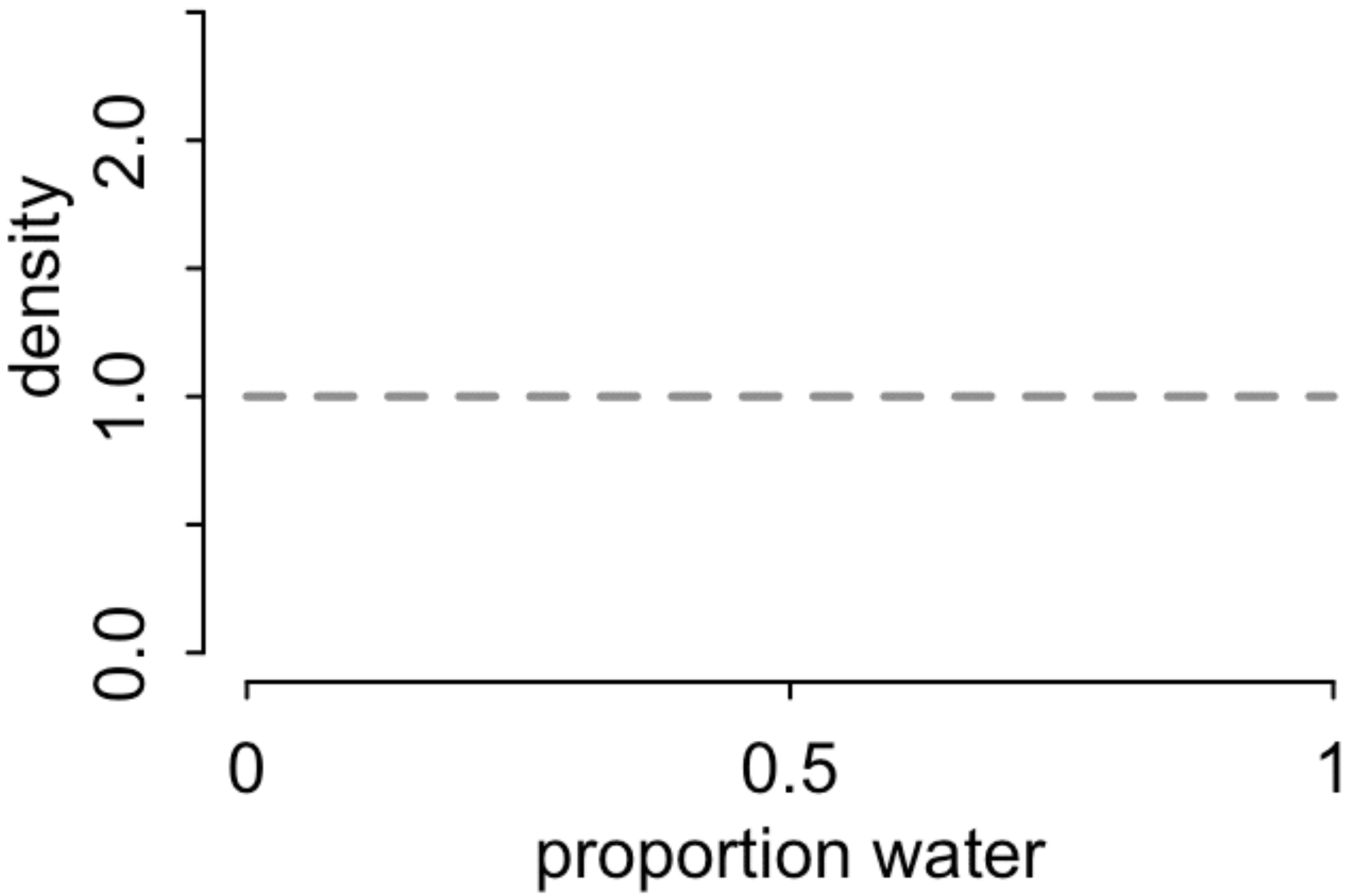
For each possible proportion of water,

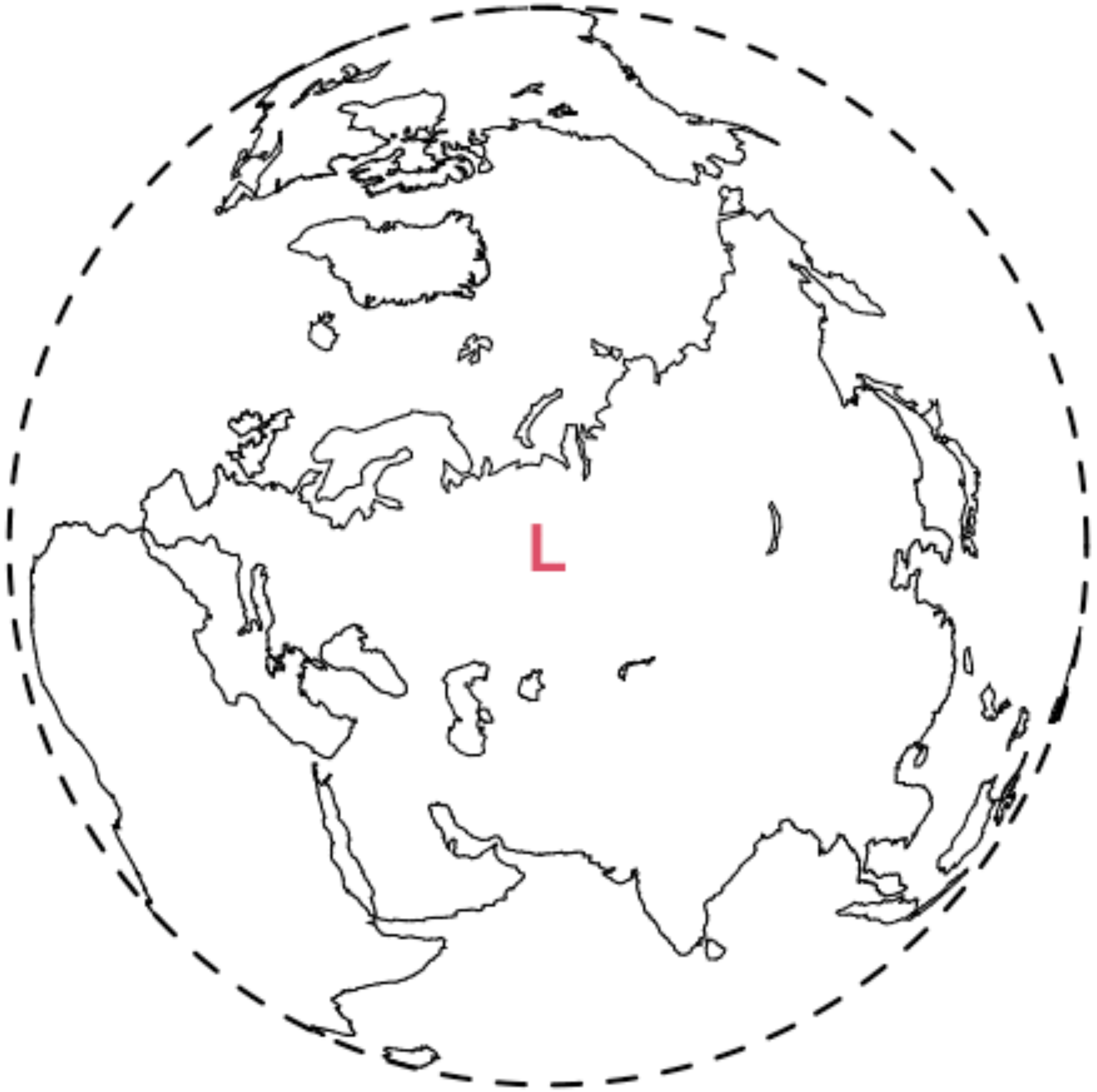
Count number of ways data could happen.

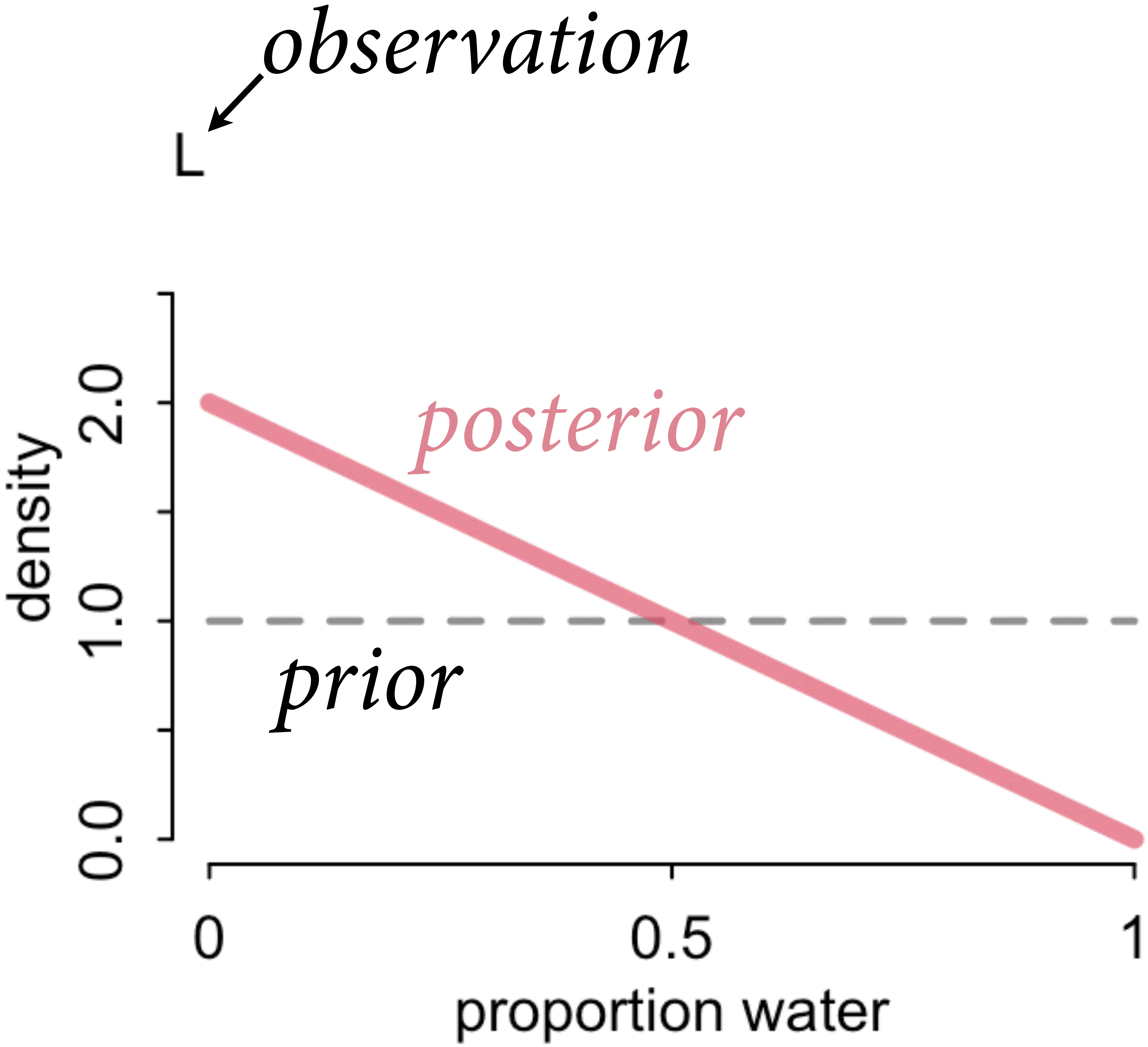
Must state how observations are generated

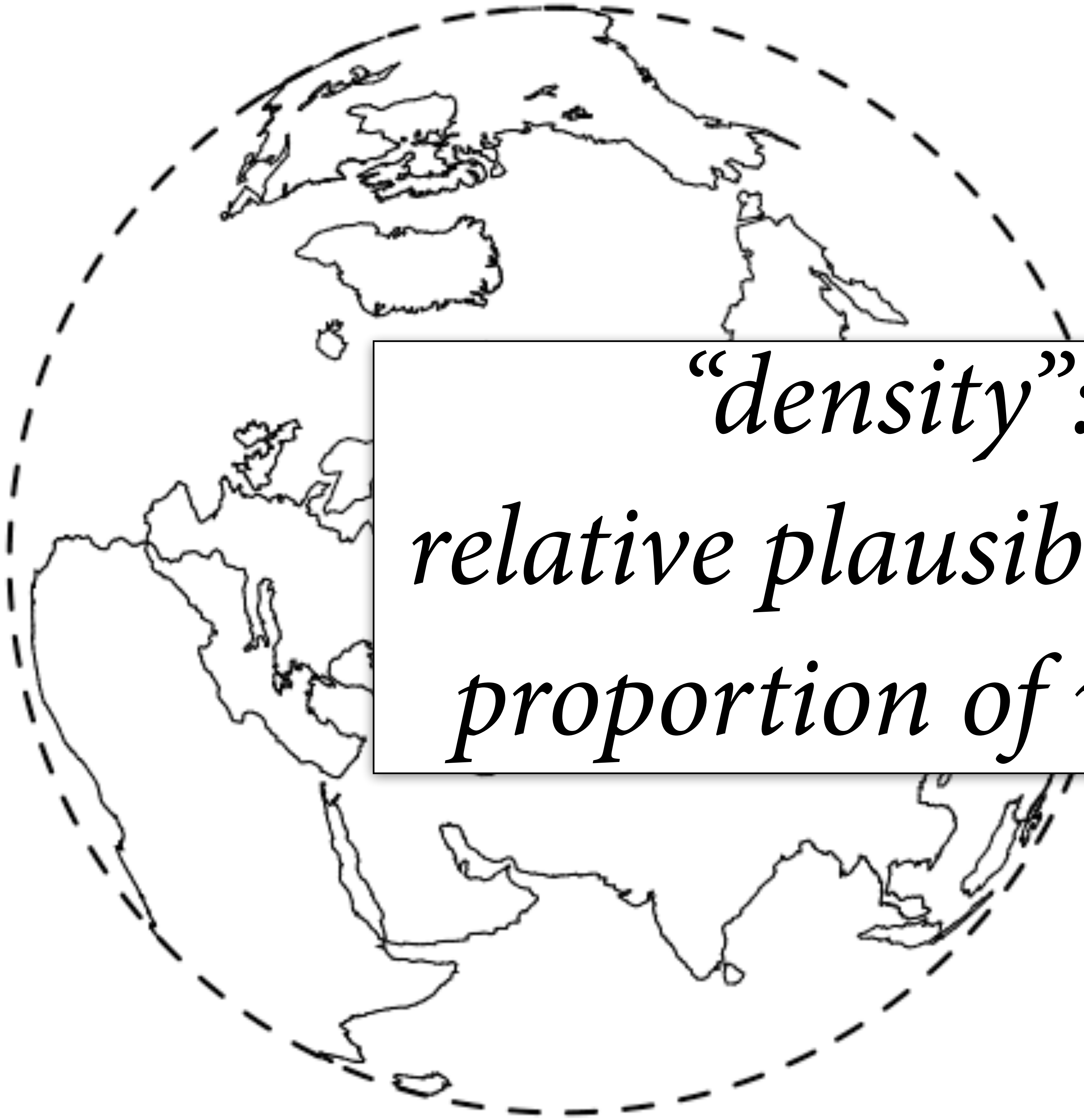
Toss The First



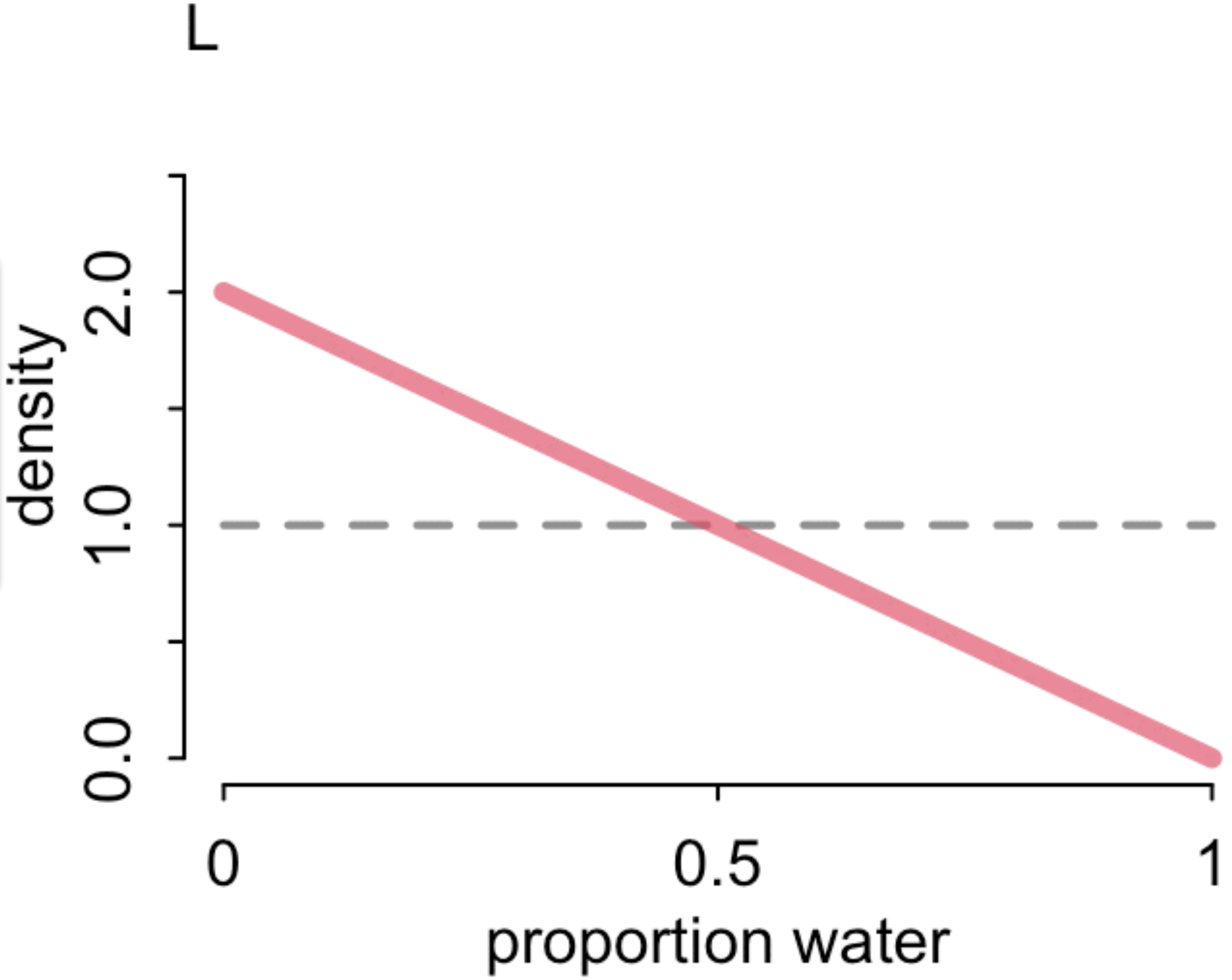



L

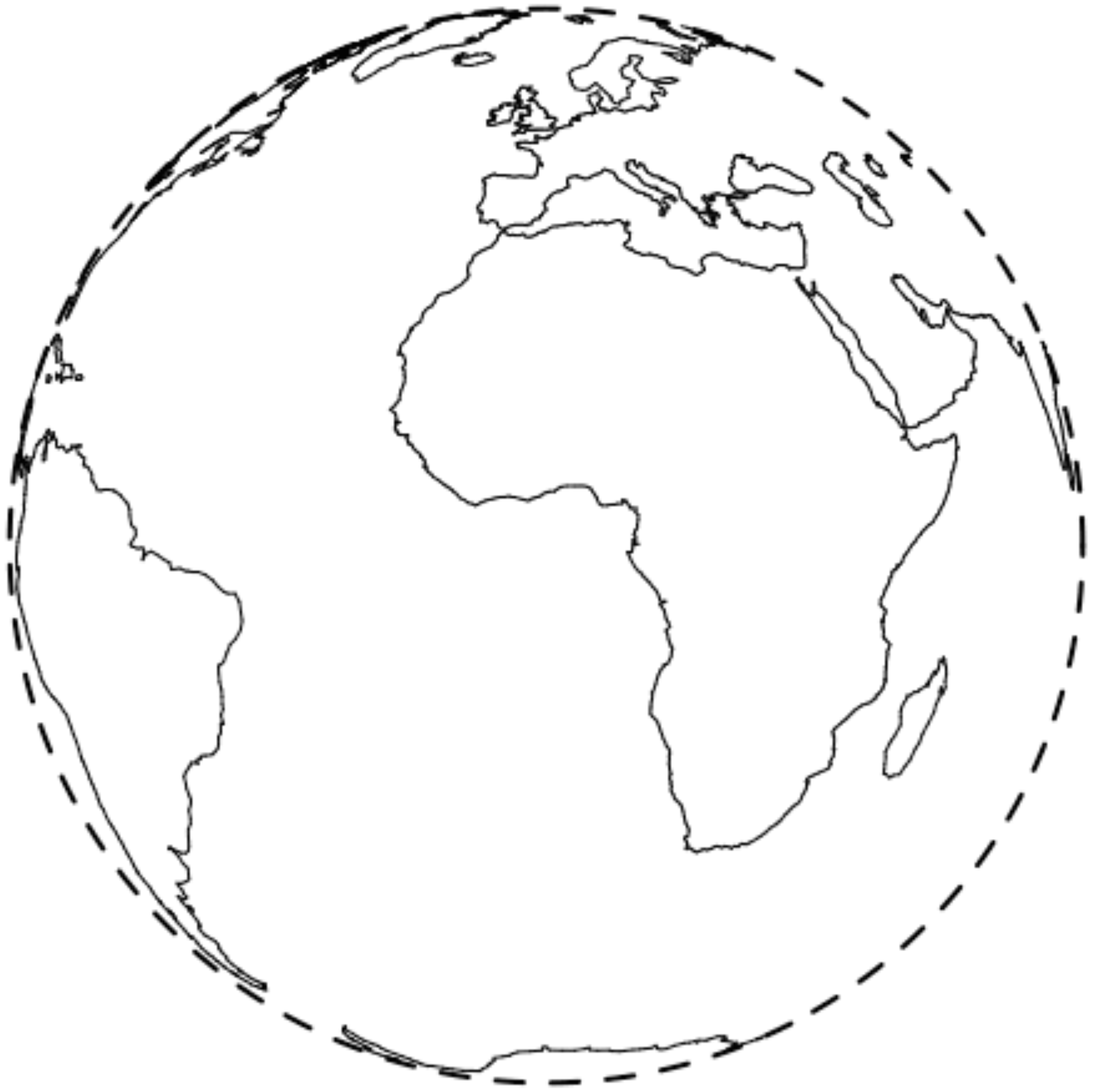

observation
L
posterior
prior
density
0.0
1.0
2.0
0
0.5
1
proportion water

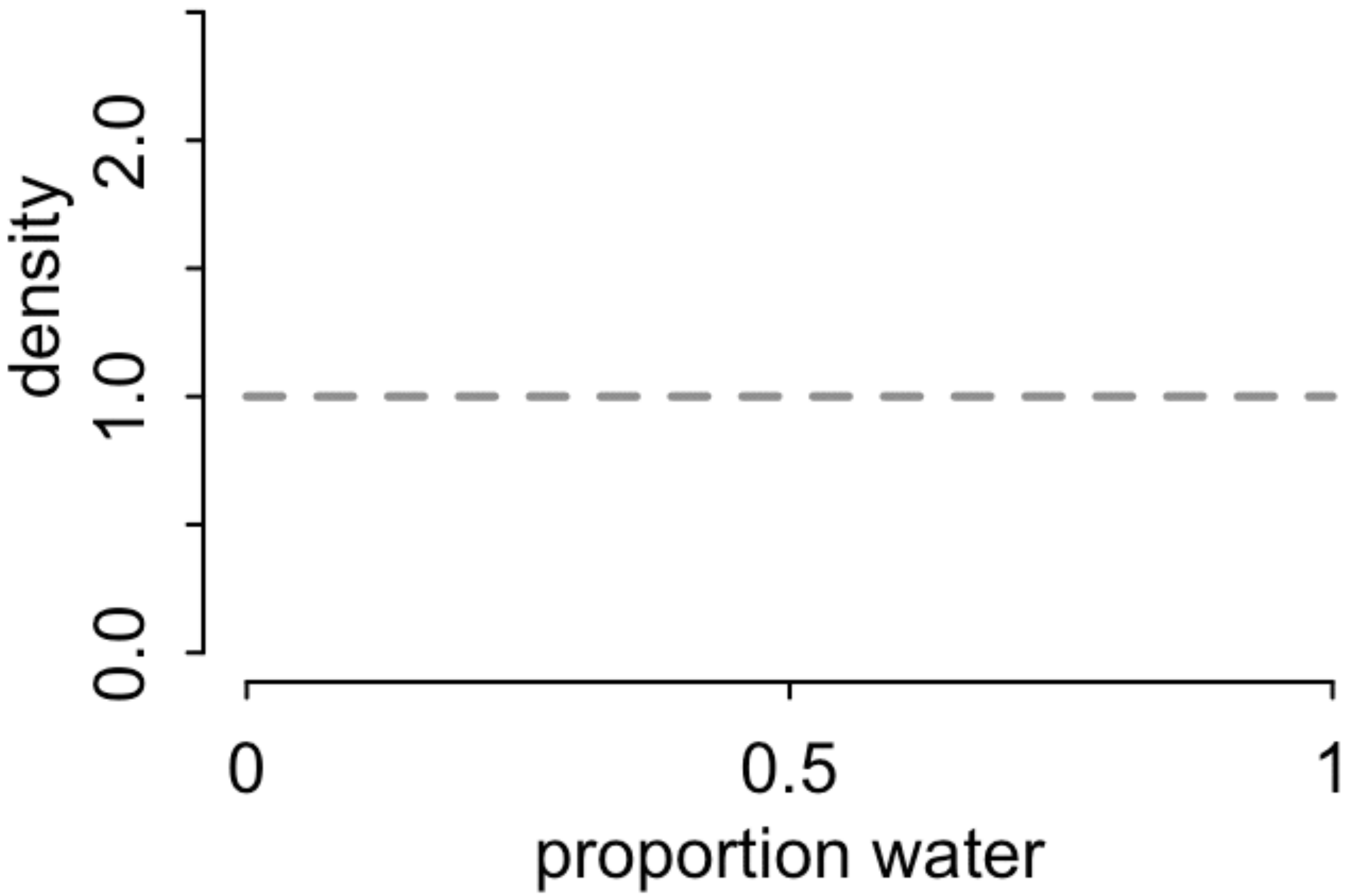


"density": relative plausibility of proportion of water

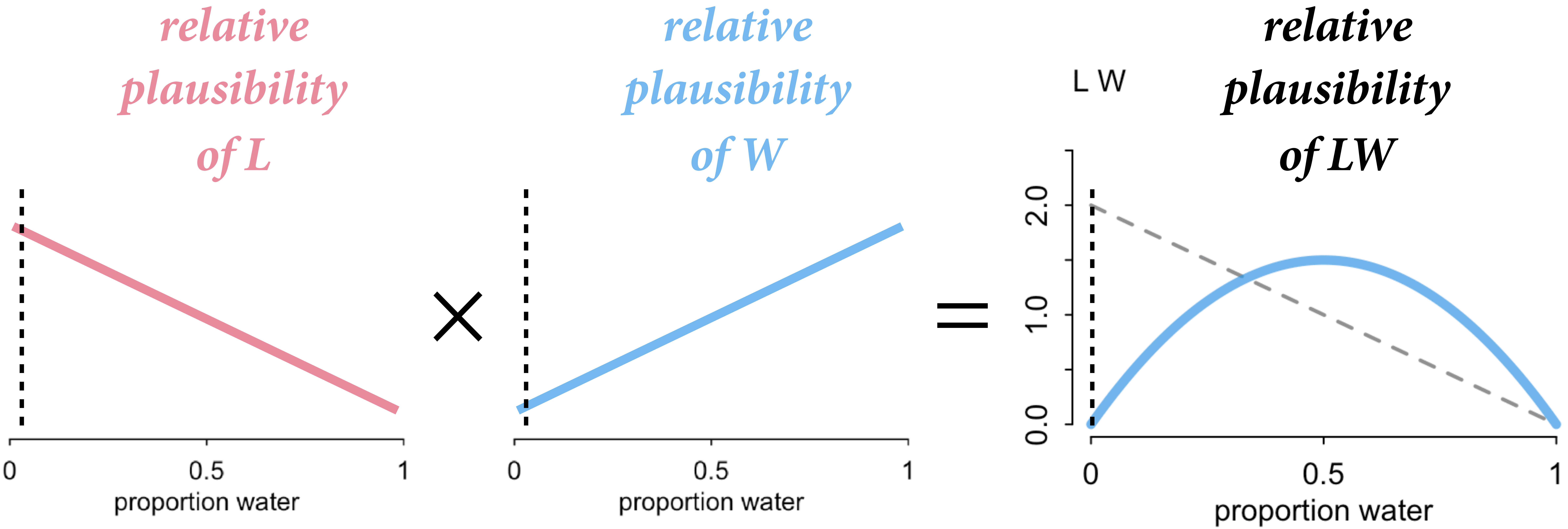


Toss The Second

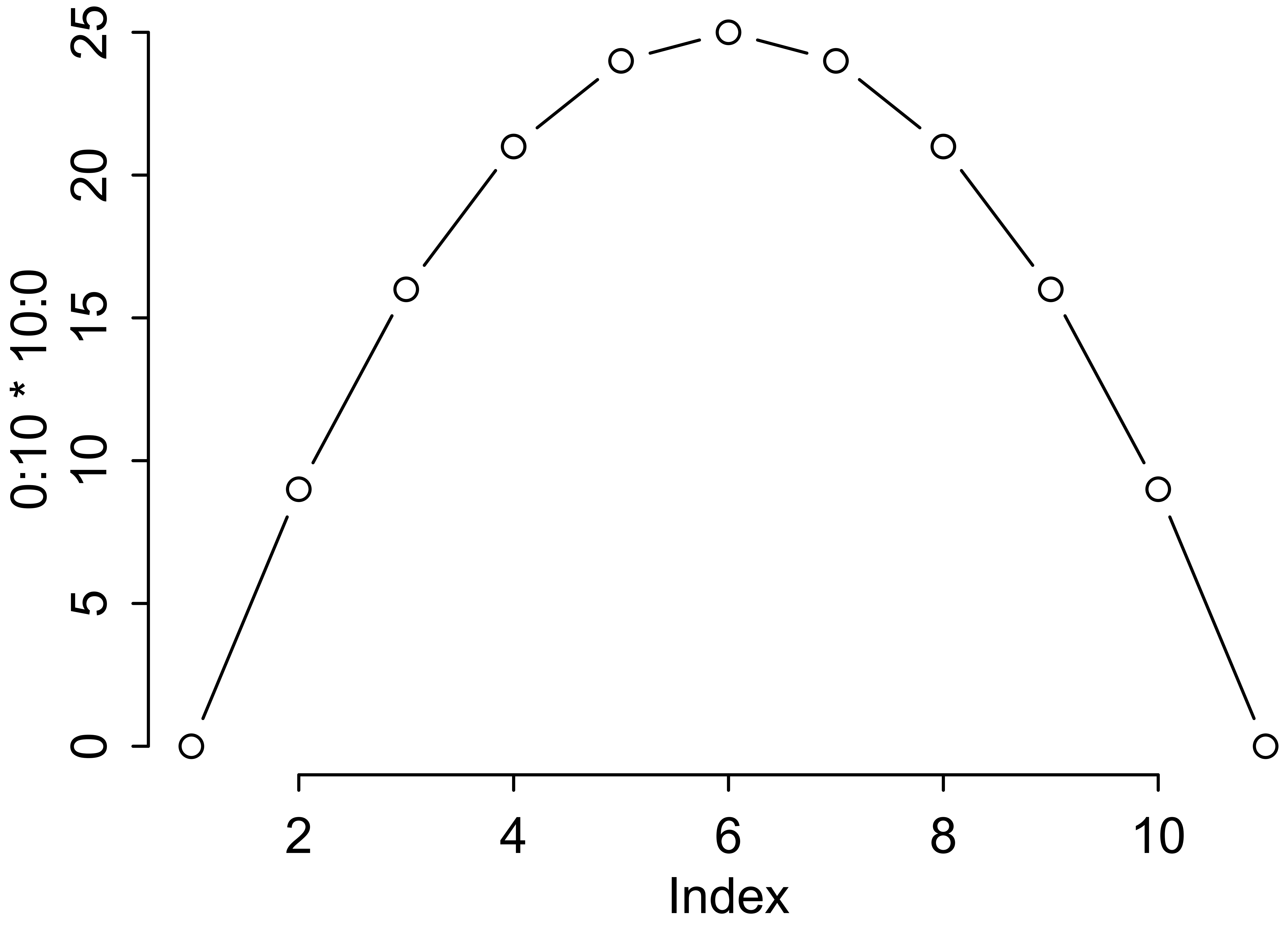




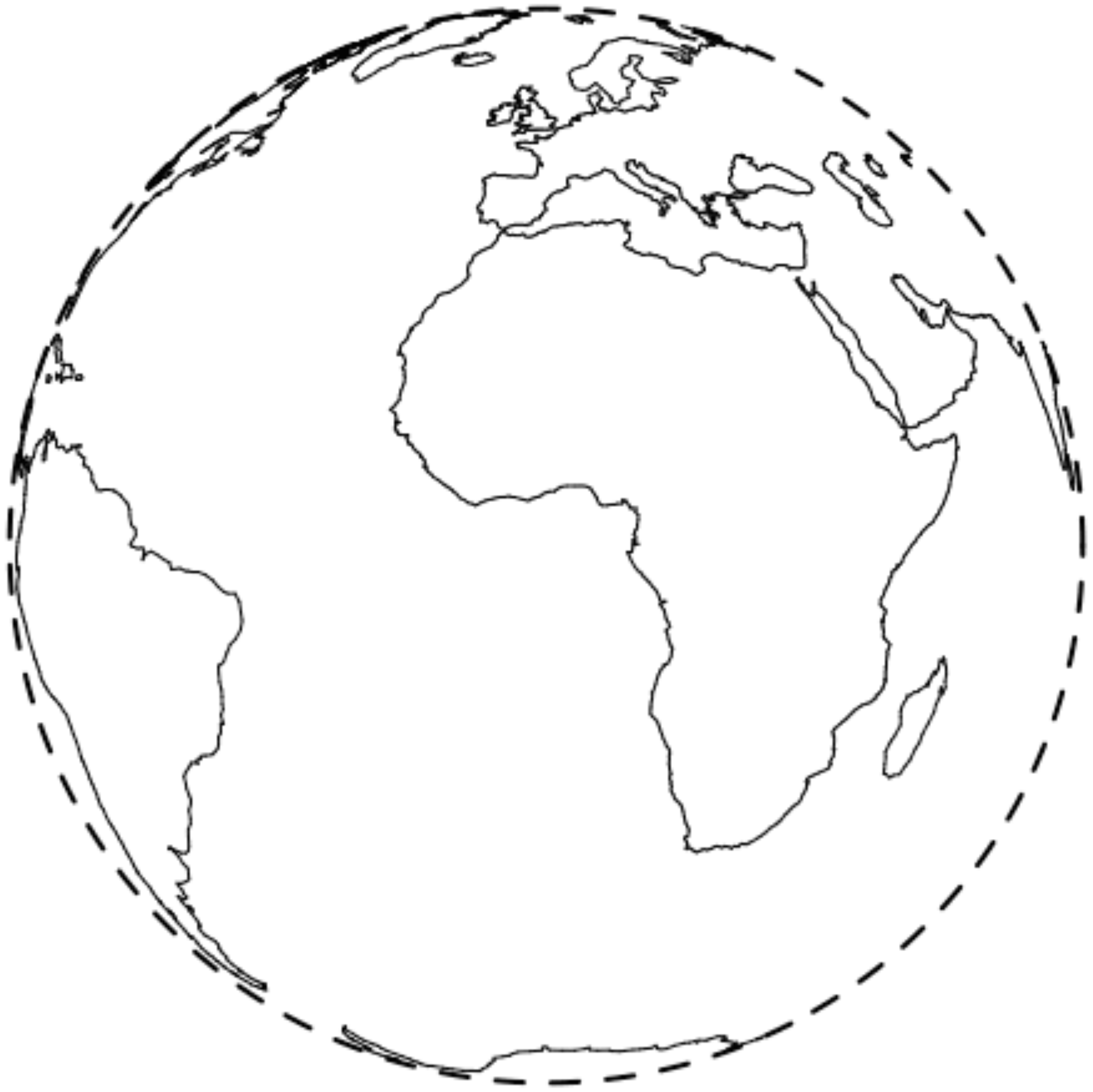
Toss The Second

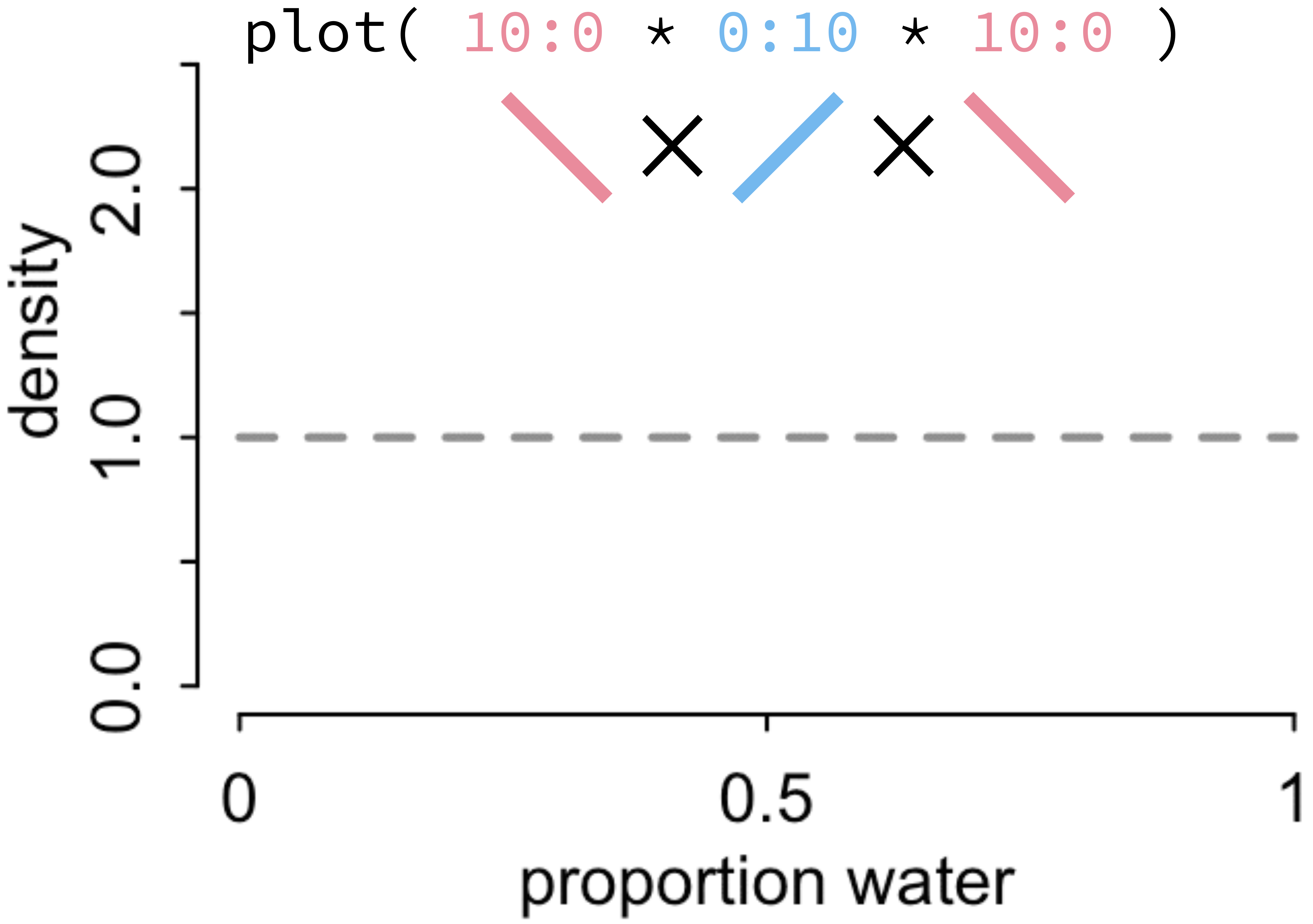


plot(0:10 * 10:0)

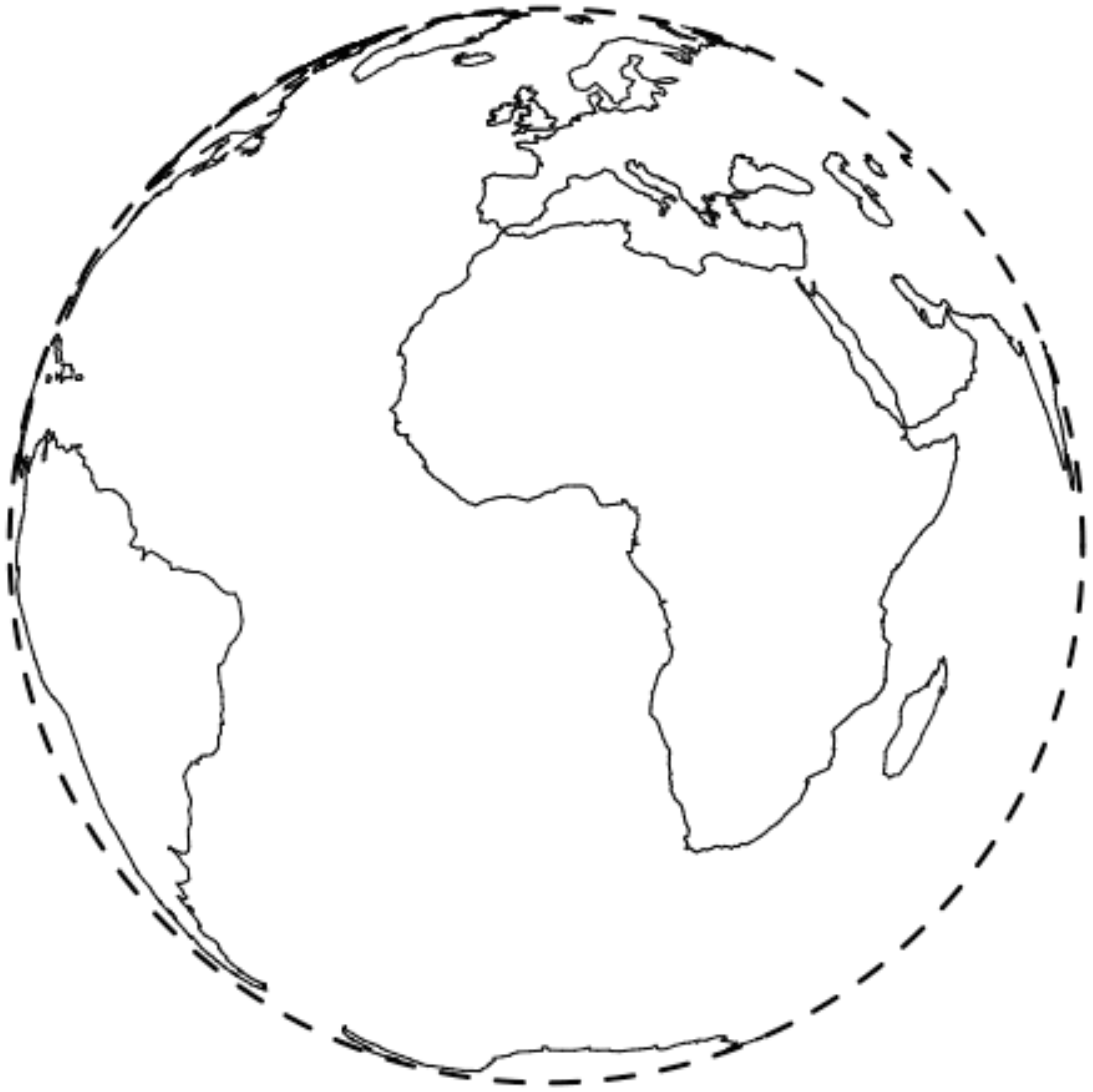


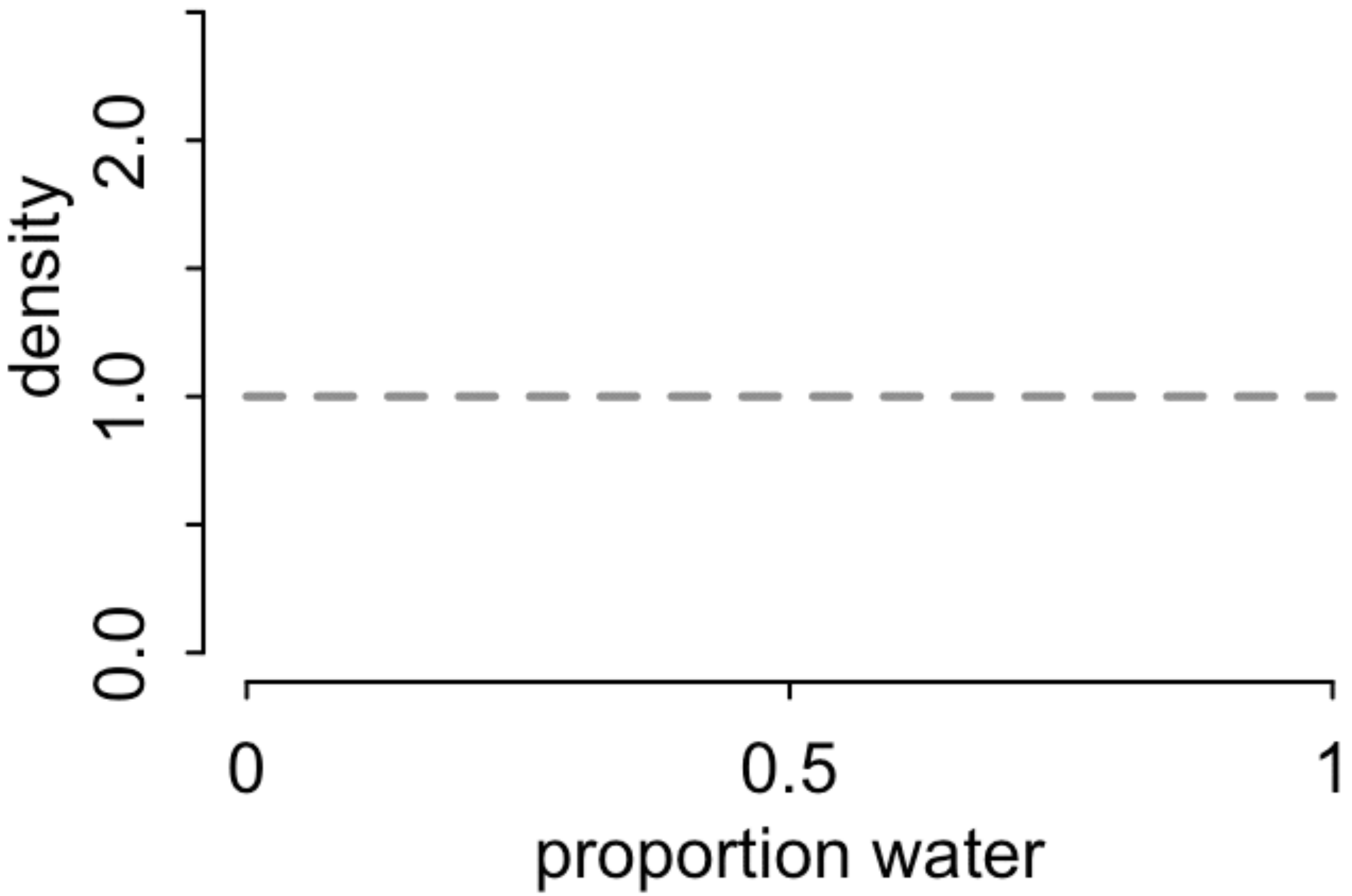
Toss The Third

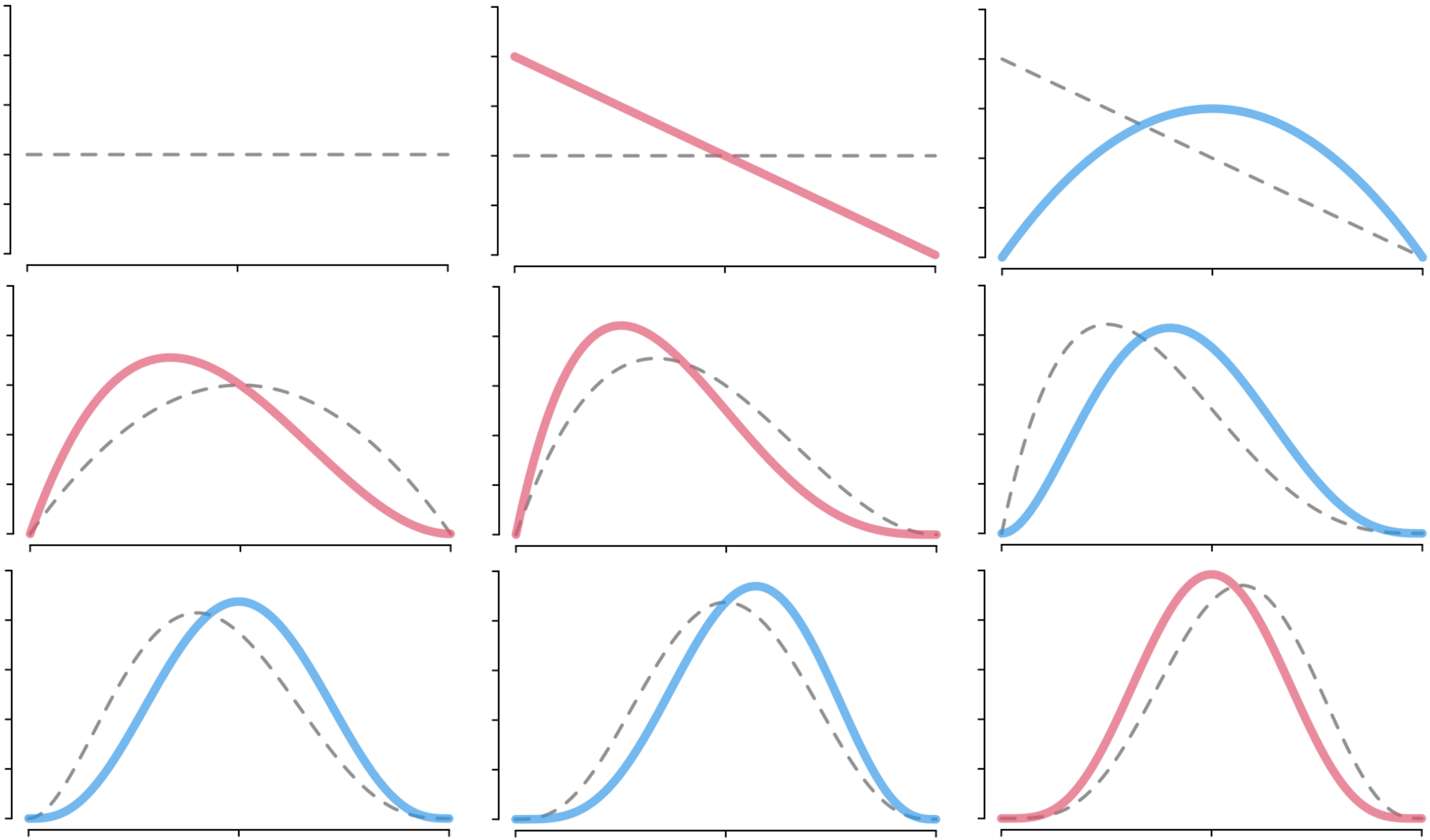




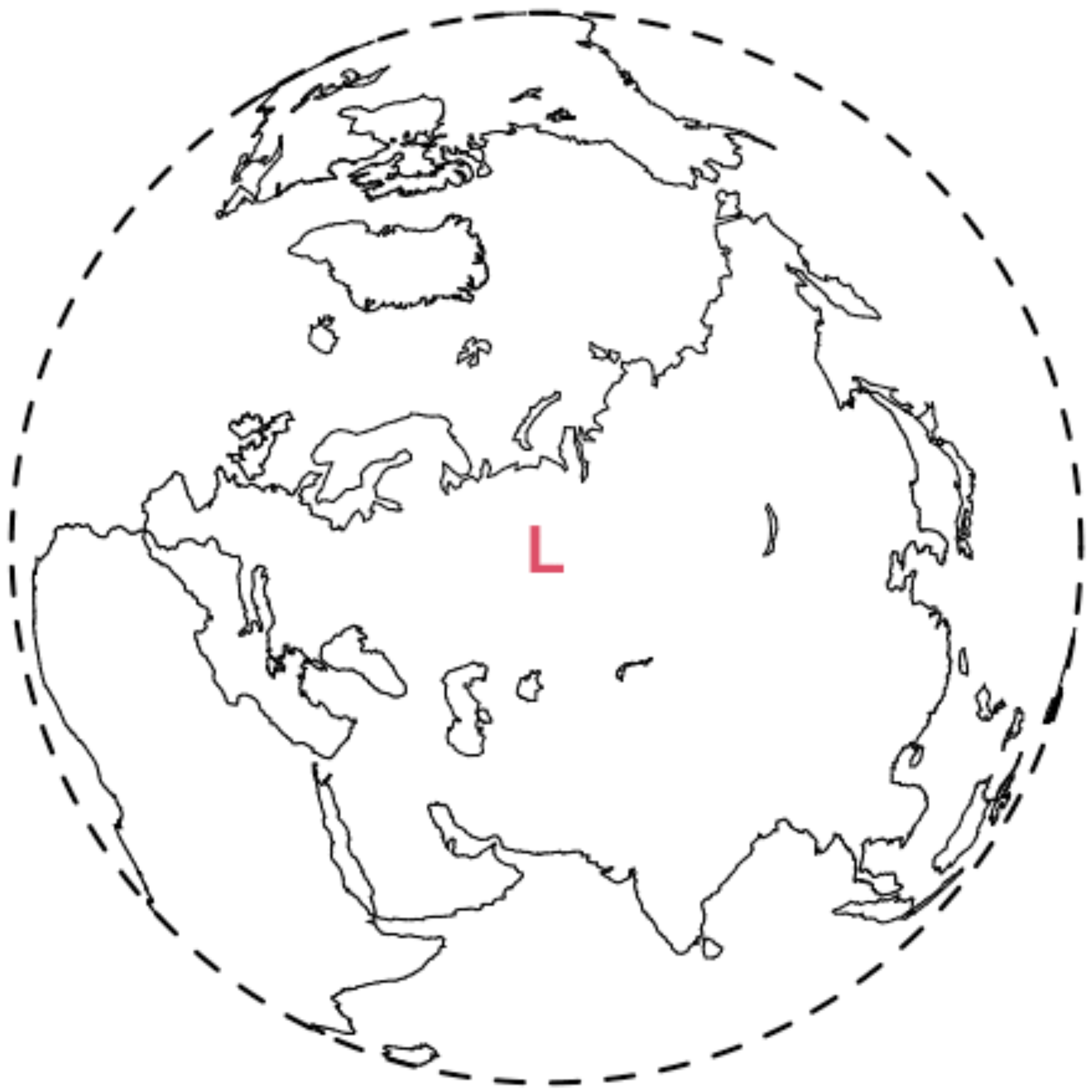
Toss The Ten

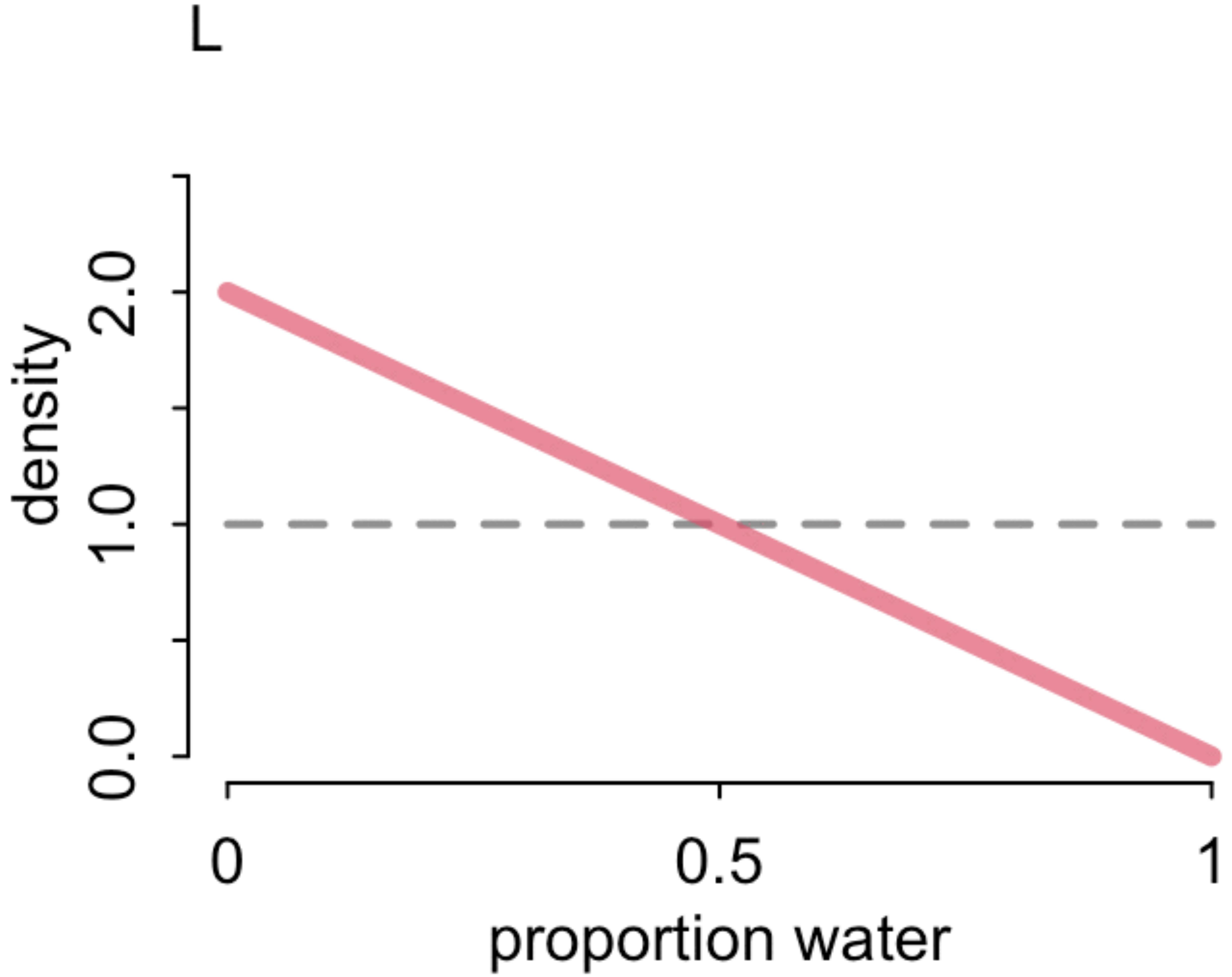




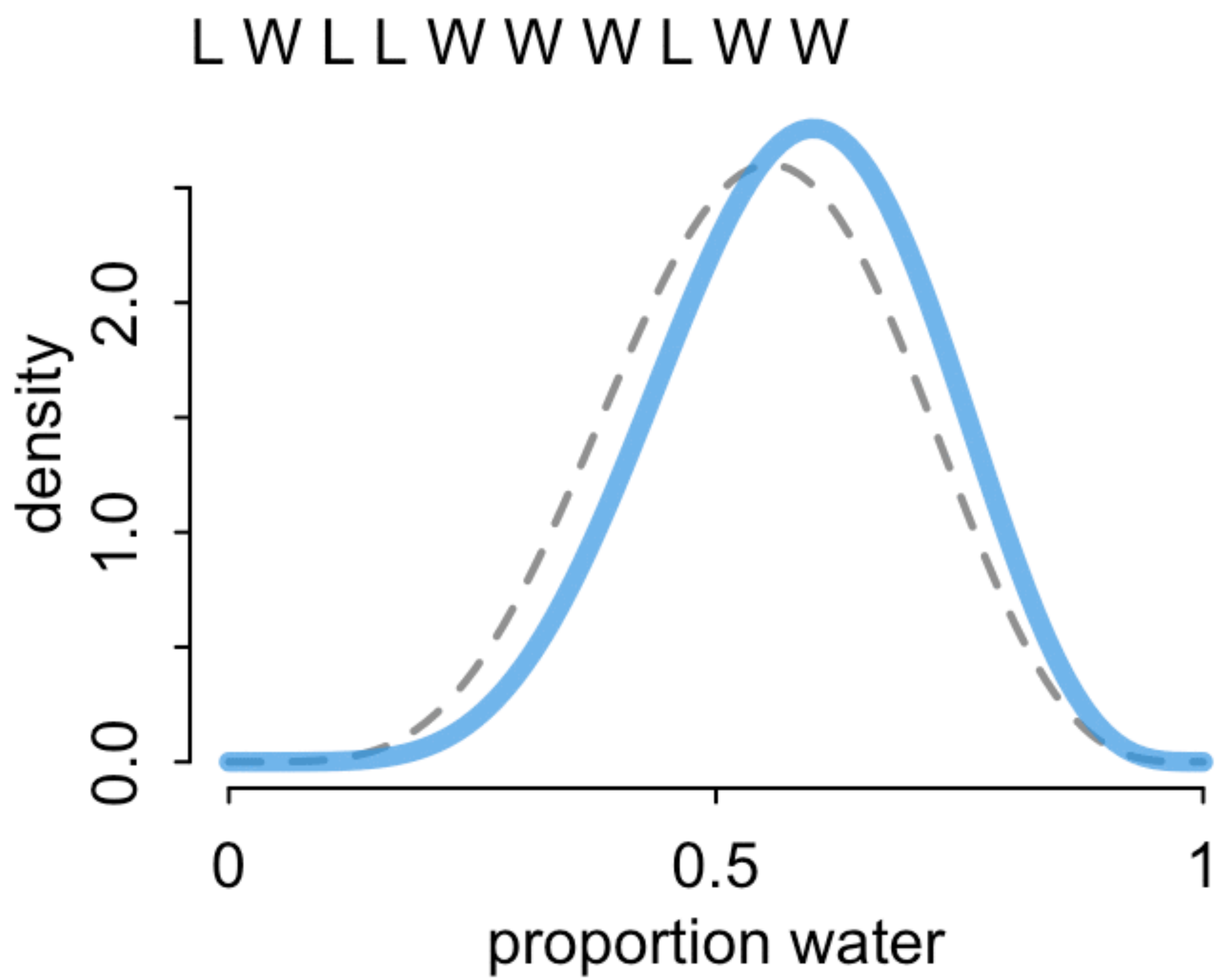


(1) No minimum sample size

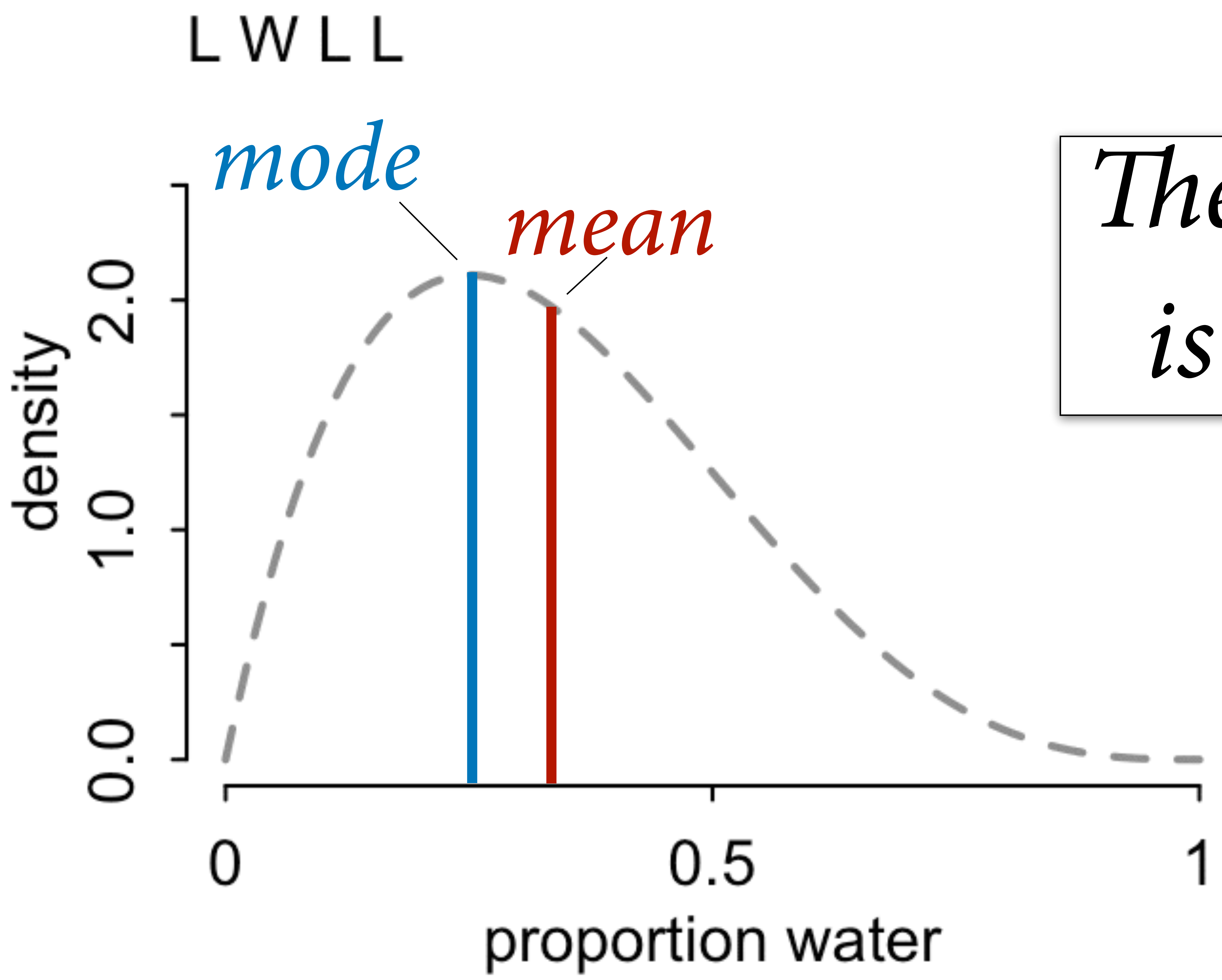




(2) Shape embodies sample size



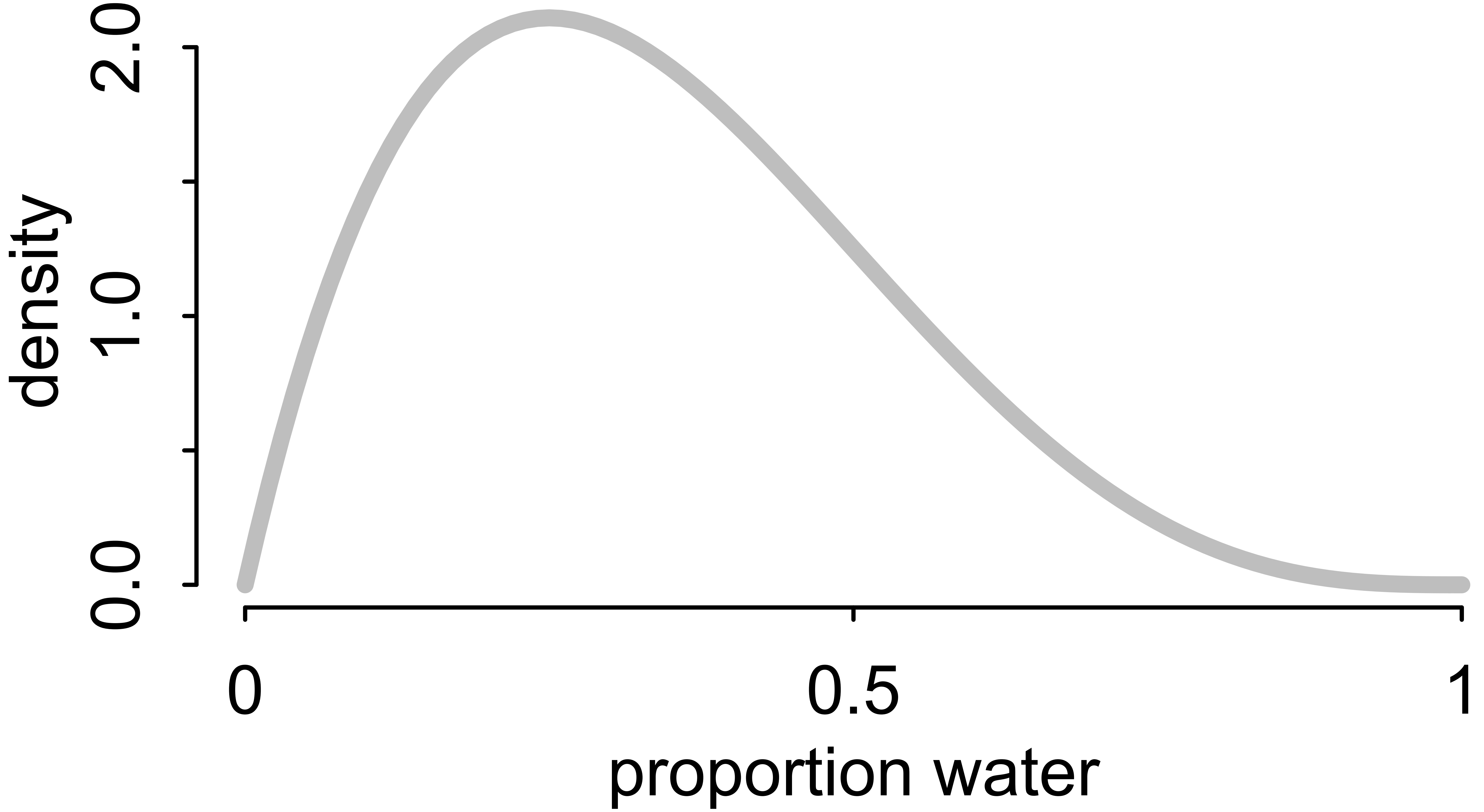
(3) No point estimate



*The **distribution** is the estimate*

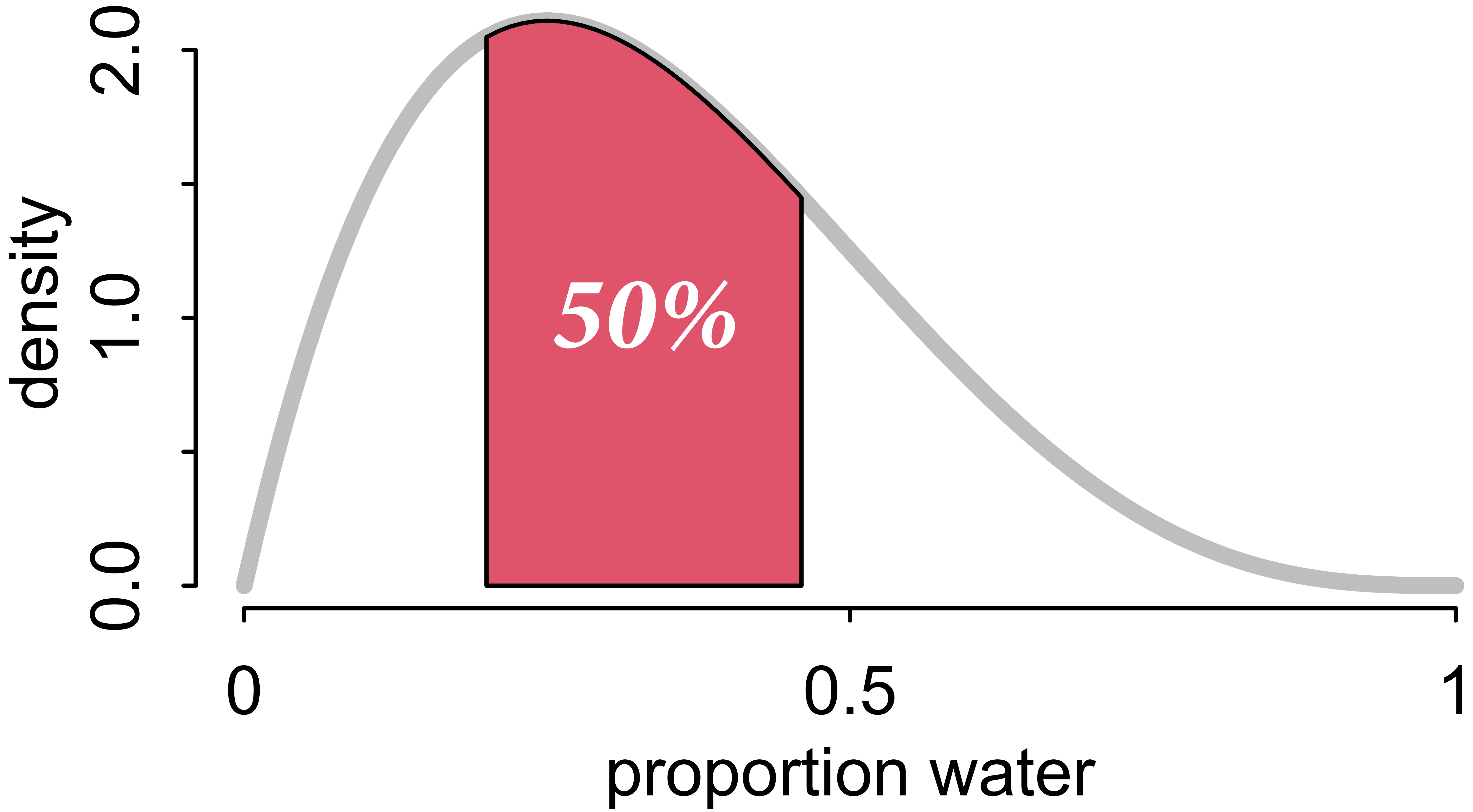
Always use the entire distribution

(4) No one true interval



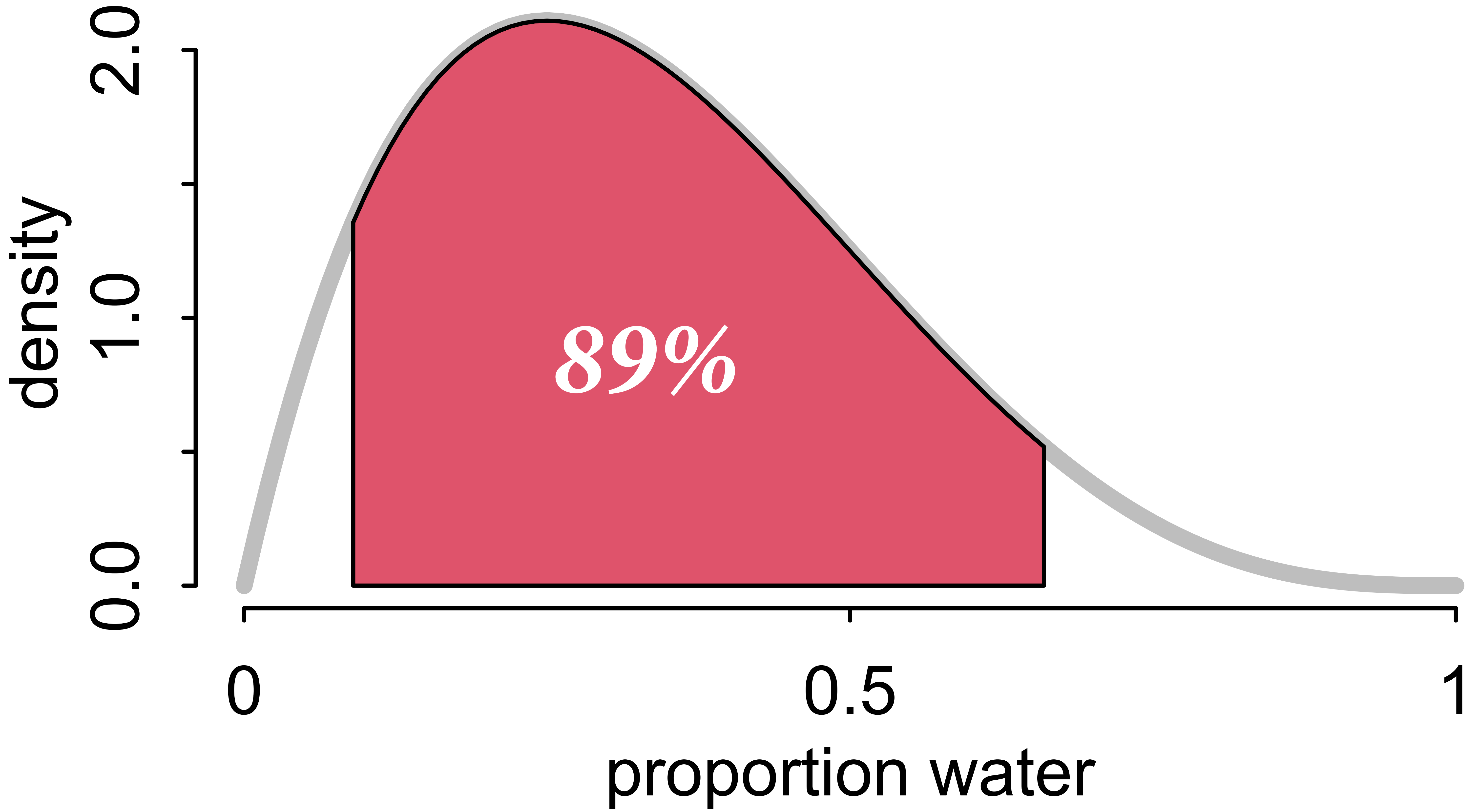
Intervals communicate shape of posterior

(4) No one true interval



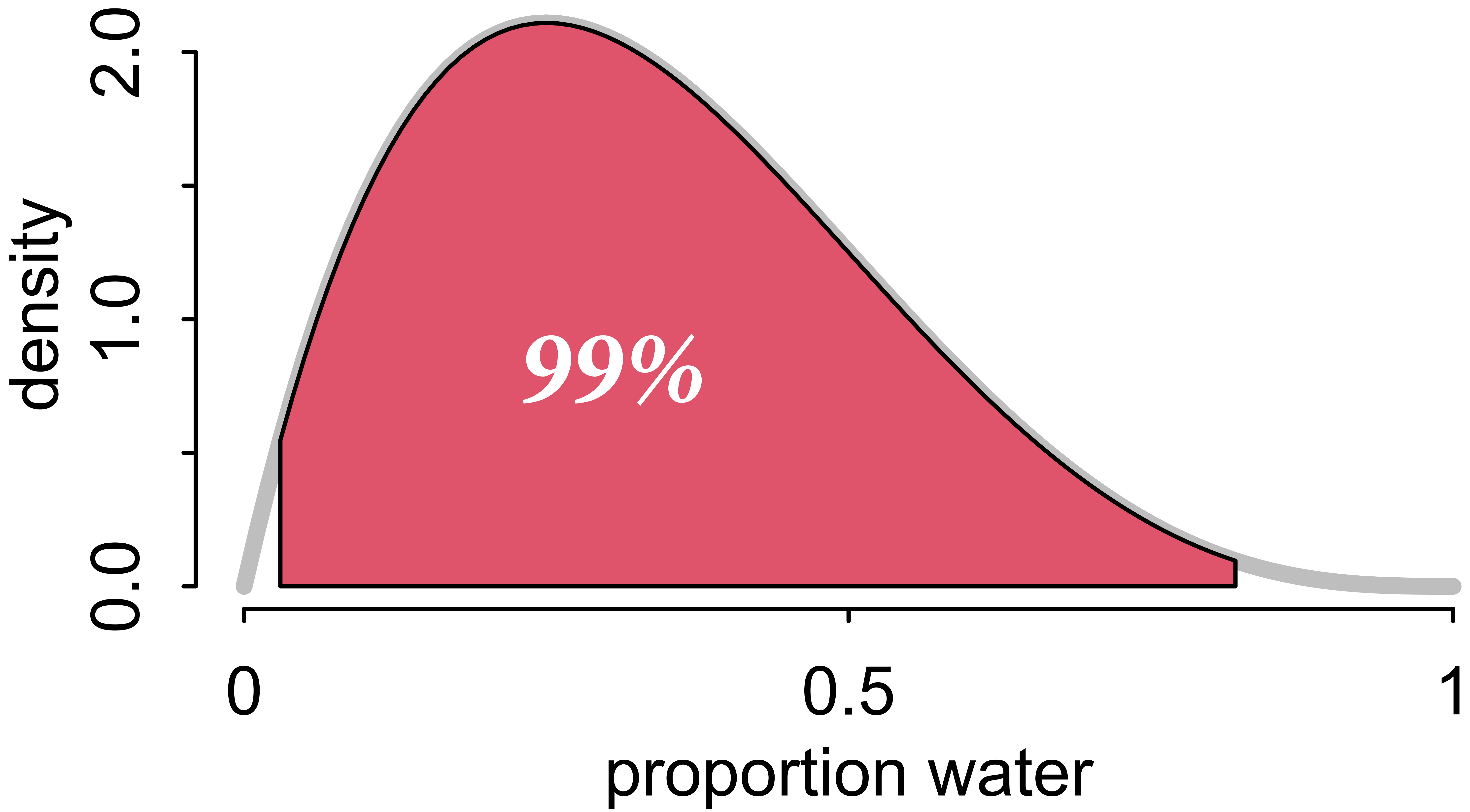
Intervals communicate shape of posterior

(4) No one true interval



Intervals communicate shape of posterior

(4) No one true interval



Intervals communicate shape of posterior

95% is obvious superstition. Nothing magical happens at the boundary.

Letters From My Reviewers

"The author uses these cute 89% intervals, but we need to see the 95% intervals so we can tell whether any of the effects are robust."

That an arbitrary interval contains an arbitrary value is not meaningful. Use the whole distribution.

PAUSE

Coding

This course involves a lot of scripting. Students can engage with the material using either the original R code examples or one of several conversions to other computing environments. The conversions are not always exact, but they are rather complete. Each option is listed below.

Original R Flavor

For those who want to use the original R code examples in the print book, you need to install the `rethinking` R package. The code is all on github https://github.com/rmcelreath/rethinking/ and there are additional details about the package there, including information about using the more-up-to-date `cmdstanr` instead of `rstan` as the underlying MCMC engine.

R + Tidyverse + ggplot2 + brms

The <Tidyverse/brms> conversion is very high quality and complete through Chapter 14.

Python and PyMC3

The <Python/PyMC3> conversion is quite complete.

Julia and Turing

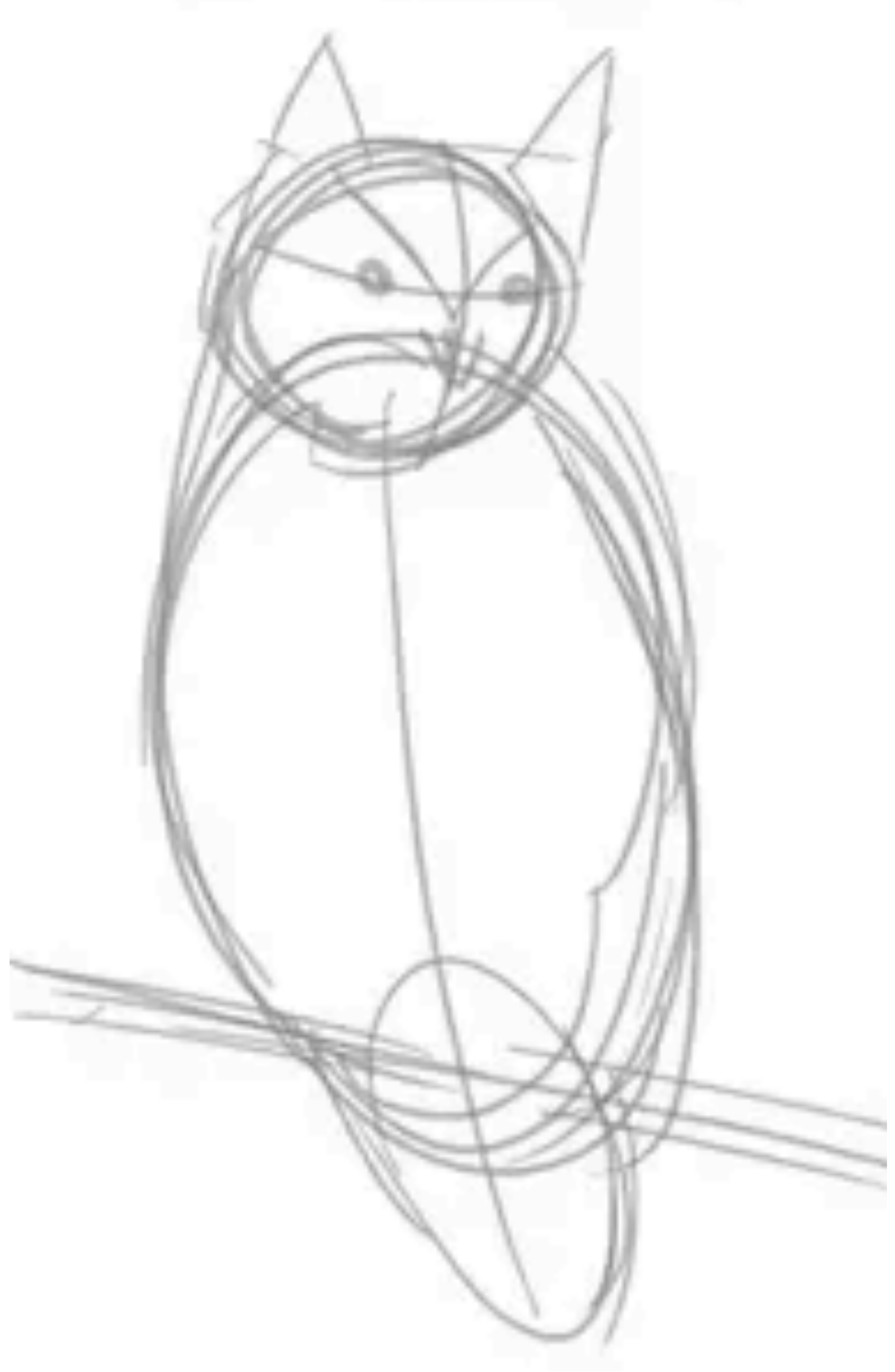
The <Julia/Turing> conversion is not as complete, but is growing fast and presents the Rethinking examples in multiple Julia engines, including the great <TuringLang>.

The Formalities

In practice, we write the model in a way that communicates all of the probability assumptions.

The observations (data) and explanations (parameters) are variables

For each variable, must say how it is generated



The Formalities

Data: W and L, the number of water and land observations

$$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

The number of ways to realize W,L given p

Binomial probability function

```
dbinom( W , W+L , p )
```

```
> dbinom( 6 , 9 , 0.7 )
[1] 0.2668279
>
```

The Formalities

Data: W and L, the number of water and land observations

$$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

The number of ways to realize W,L given p

Parameters: p, the proportion of water on the globe

$$\Pr(p) = \frac{1}{1-0} = 1.$$

Relative plausibility of each possible p

The Formalities

$$\Pr(W, L|p) = \frac{(W+L)!}{W!L!} p^W (1-p)^L$$

$$\Pr(p) = \frac{1}{1-0} = 1.$$

Posterior is (normalized) product:

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

Relative plausibility of each possible p, after learning W,L

We multiply because that's how the garden counts!

With Numbers

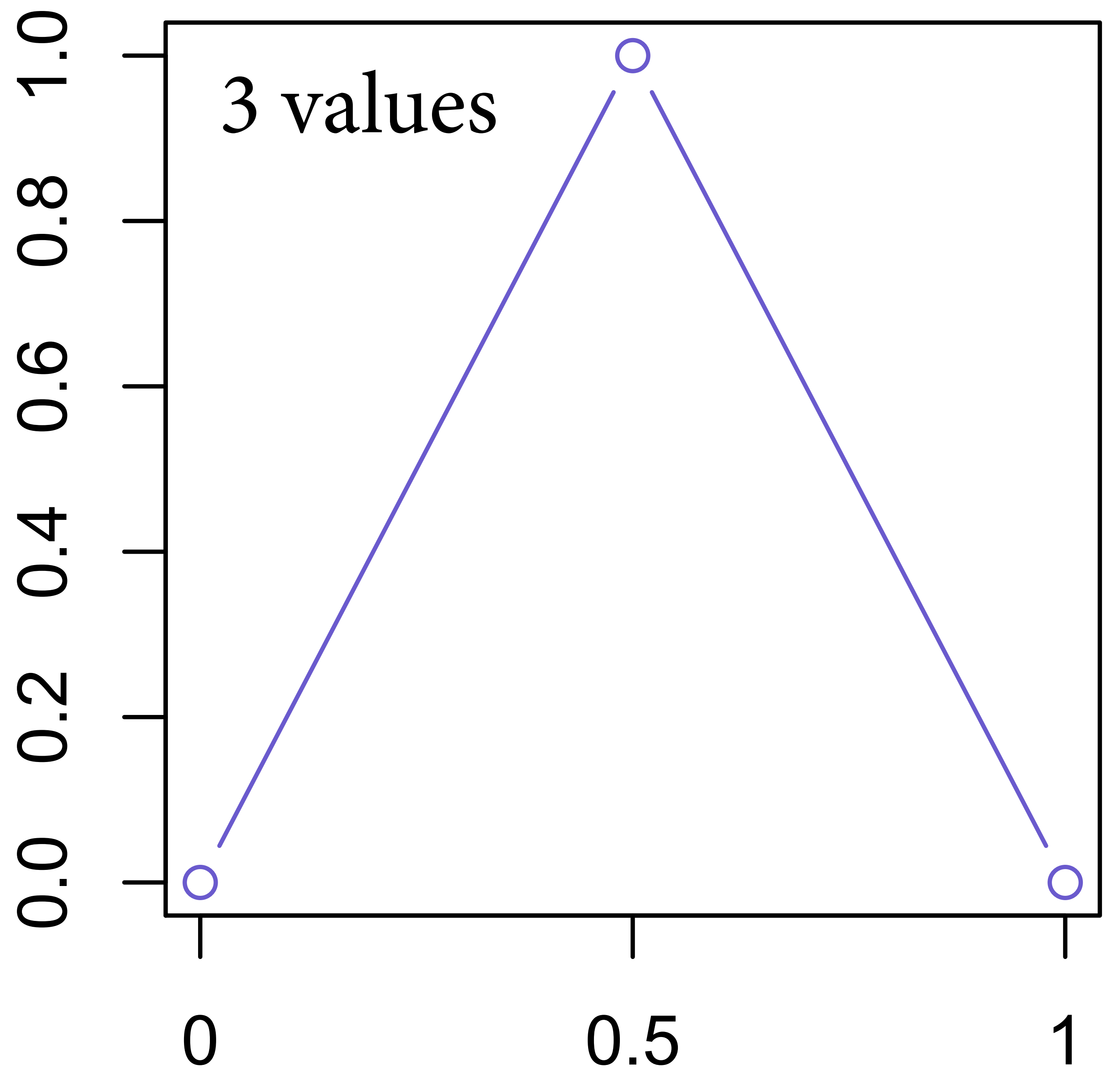
Ignore the mathematics for the moment and just draw the owl with numbers

1. For each possible value of p

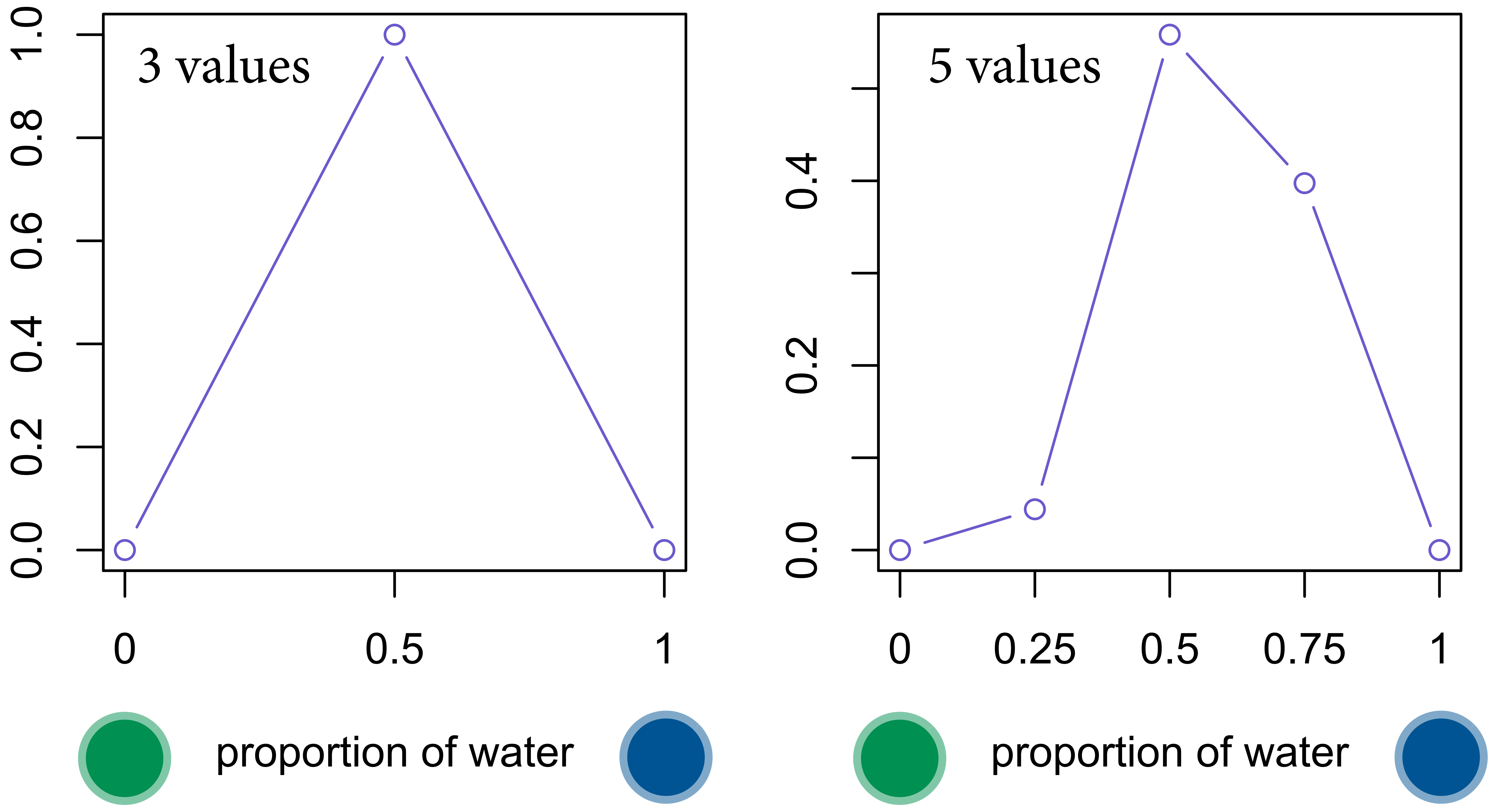
2. Compute product $\Pr(W,L|p)\Pr(p)$

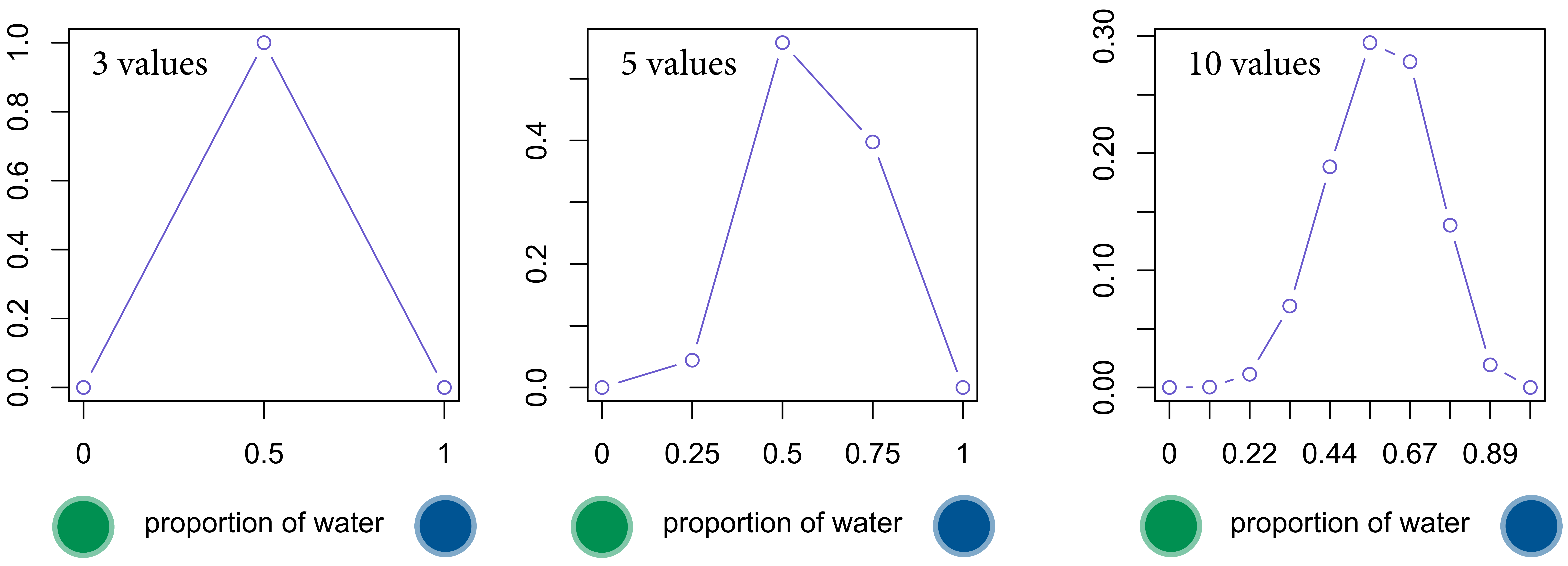
3. Relative sizes of products in (2) are posterior probabilities

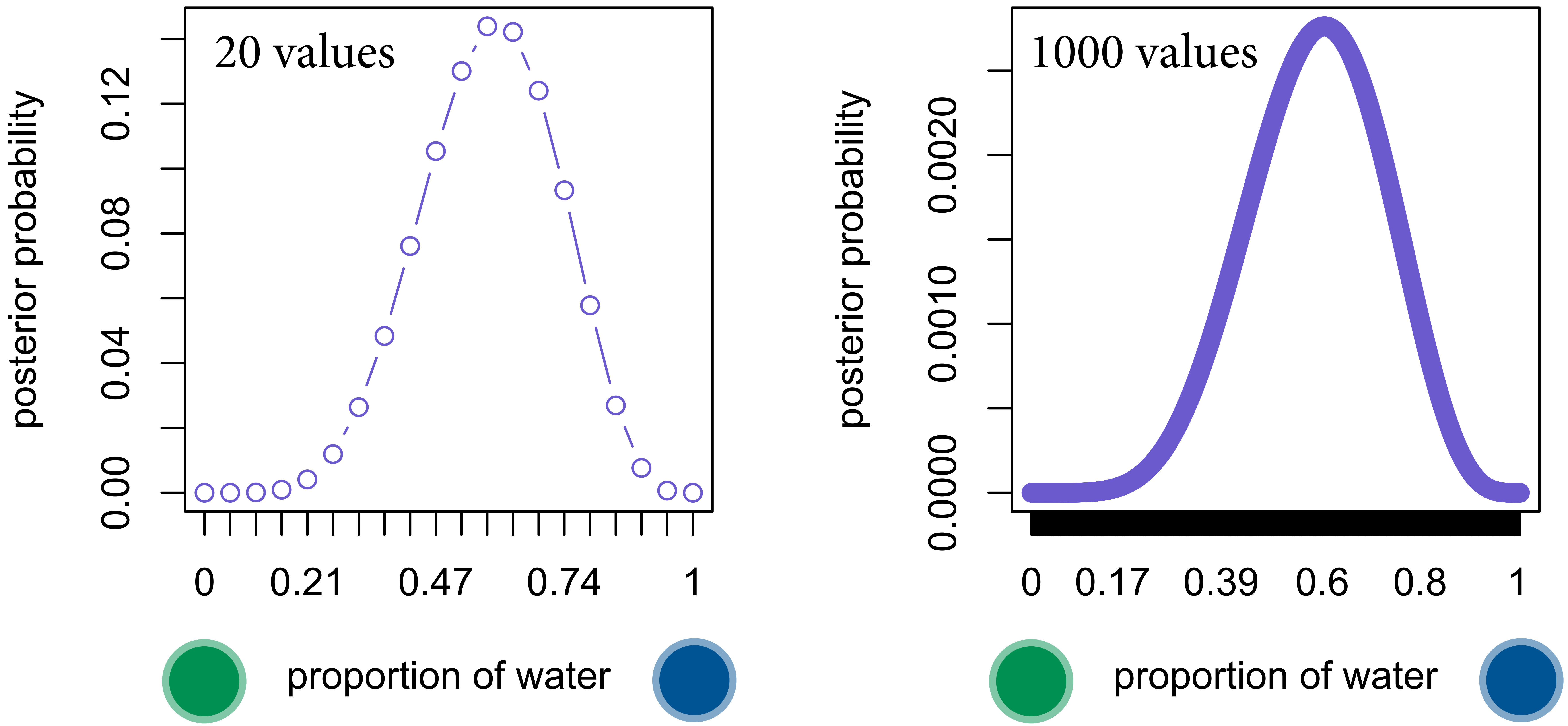
Bayesian owl



proportion of water


3 values
1.0
0.8
0.6
0.4
0.2
0.0
0
0.5
1
proportion of water
5 values
0.4
0.2
0.0
0
0.25
0.5
0.75
1
proportion of water


3 values
5 values
10 values
1.0
0.8
0.6
0.4
0.2
0.0
0.4
0.2
0.0
0.30
0.20
0.10
0.00
0
0.5
1
0
0.25
0.5
0.75
1
0
0.22
0.44
0.67
0.89
proportion of water
proportion of water
proportion of water


20 values
posterior probability
0.00
0.04
0.08
0.12
0
0.21
0.47
0.74
1
proportion of water
1000 values
posterior probability
0.0000
0.0010
0.0020
0
0.17
0.39
0.6
0.8
1
proportion of water

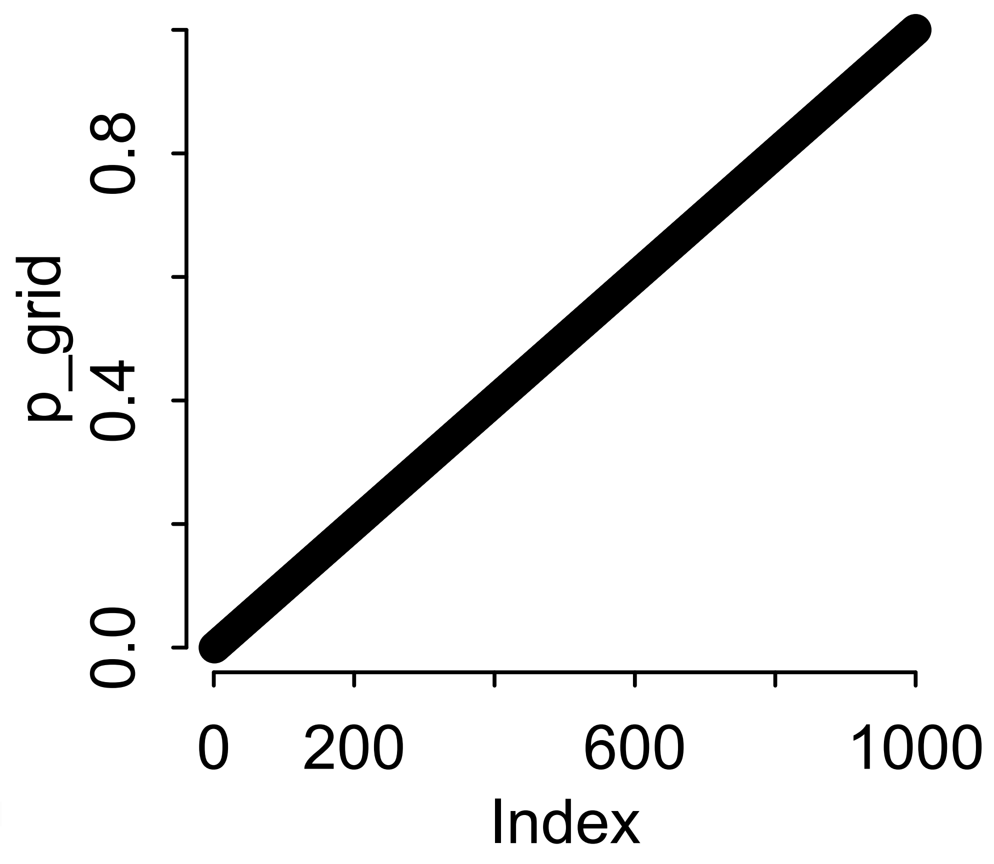
Grid Approximation

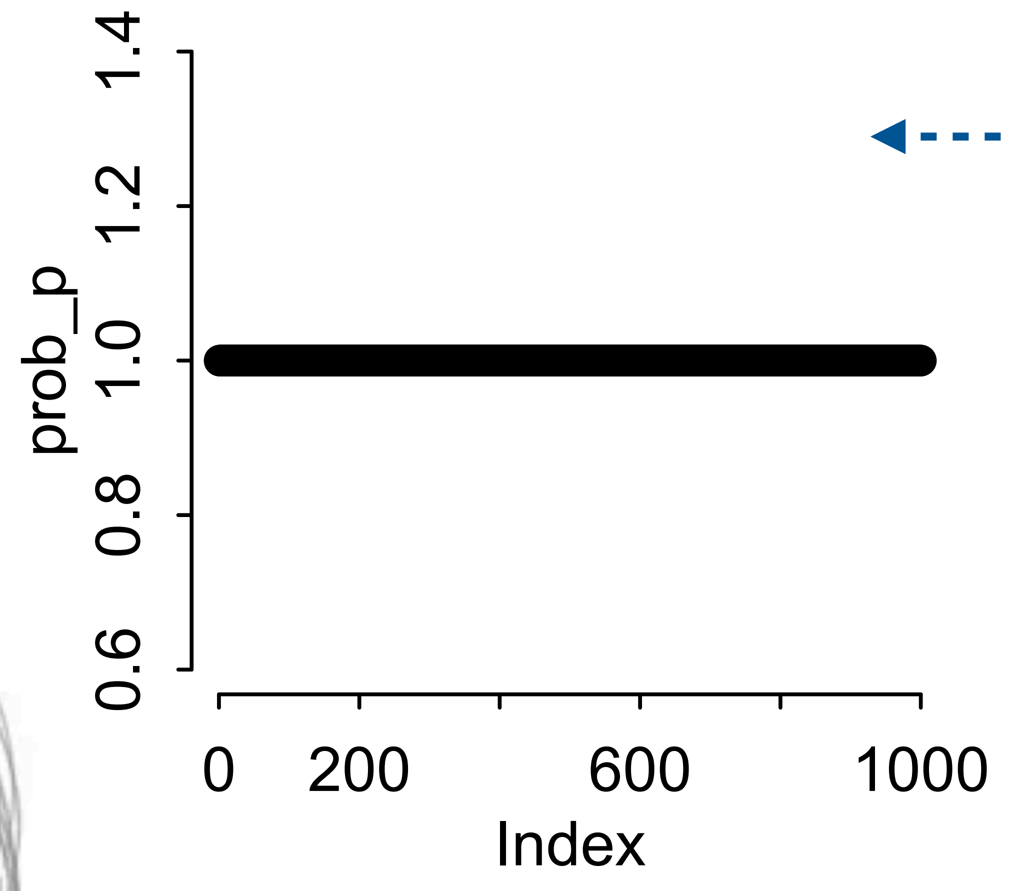
R code 3.2

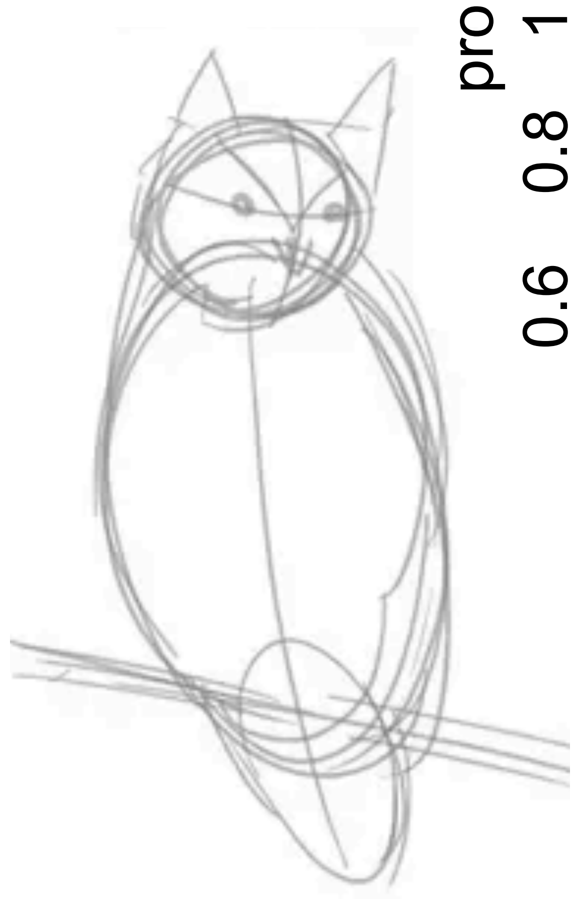
```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

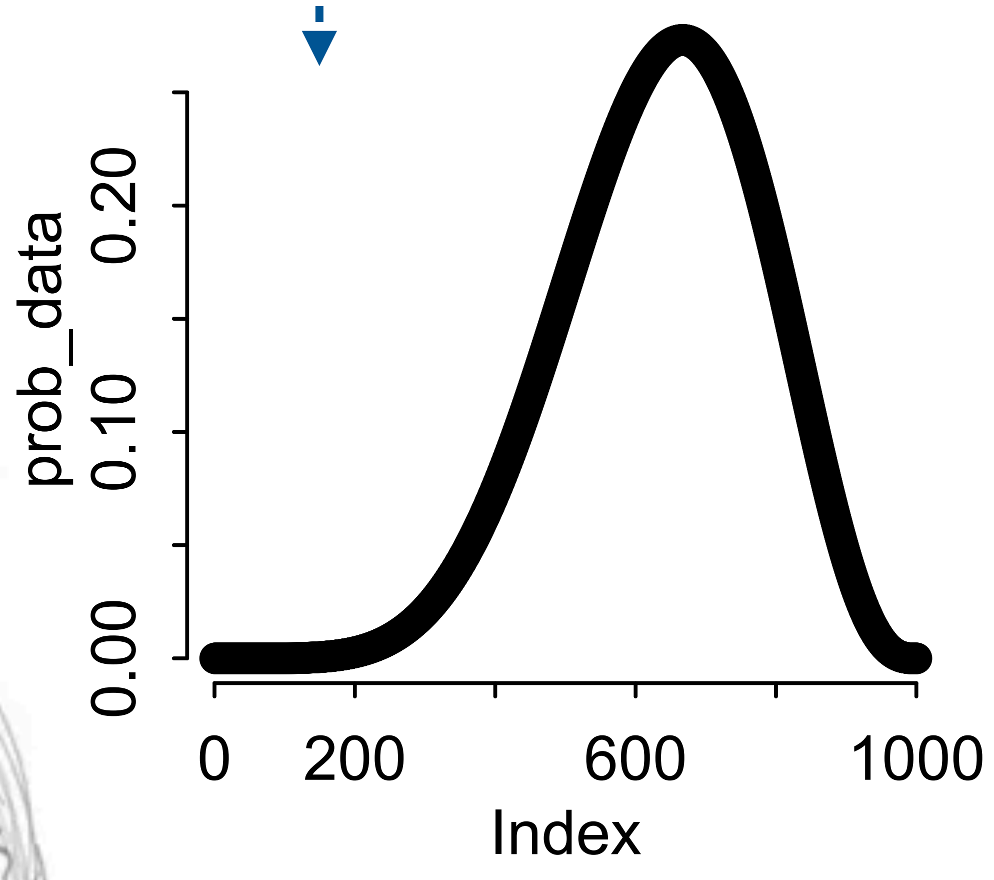
Let's draw the owl

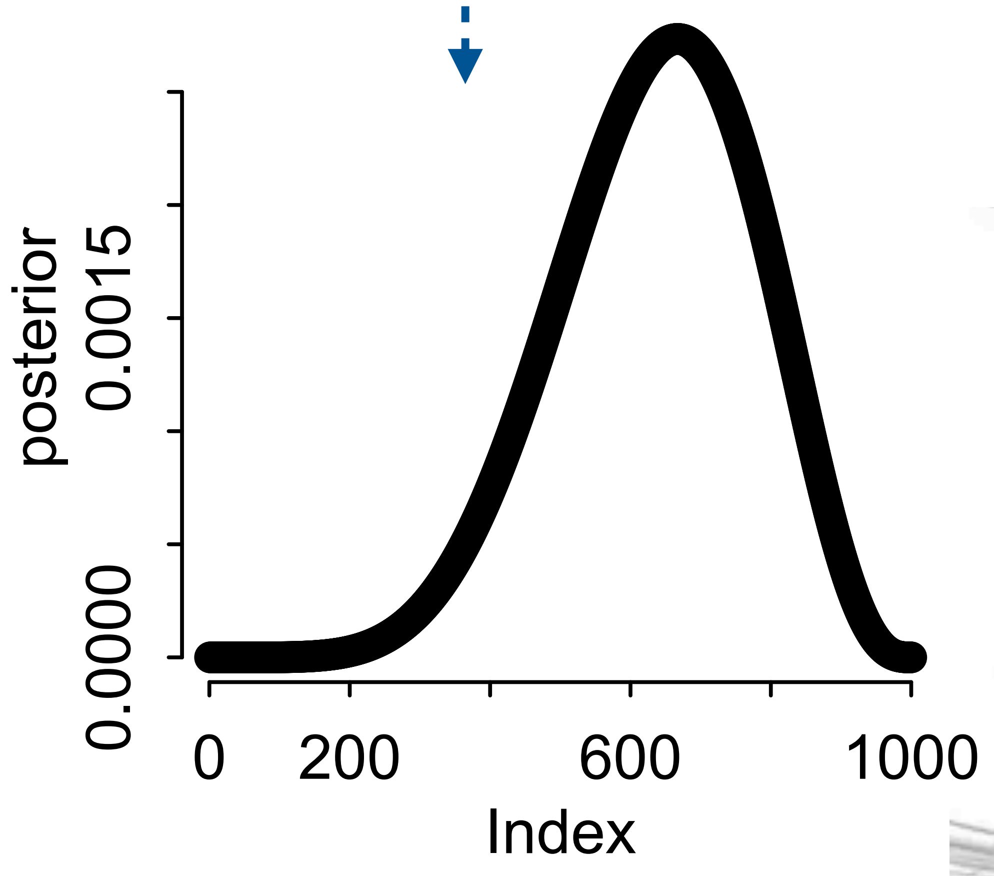
```r
p_grid ← seq( from=0 , to=1 , len=1000 )
prob_p ← rep( 1 , 1000 )
prob_data ← dbinom( 6 , 9 , prob=p_grid )
posterior ← prob_data * prob_p
posterior ← posterior / sum(posterior)
```



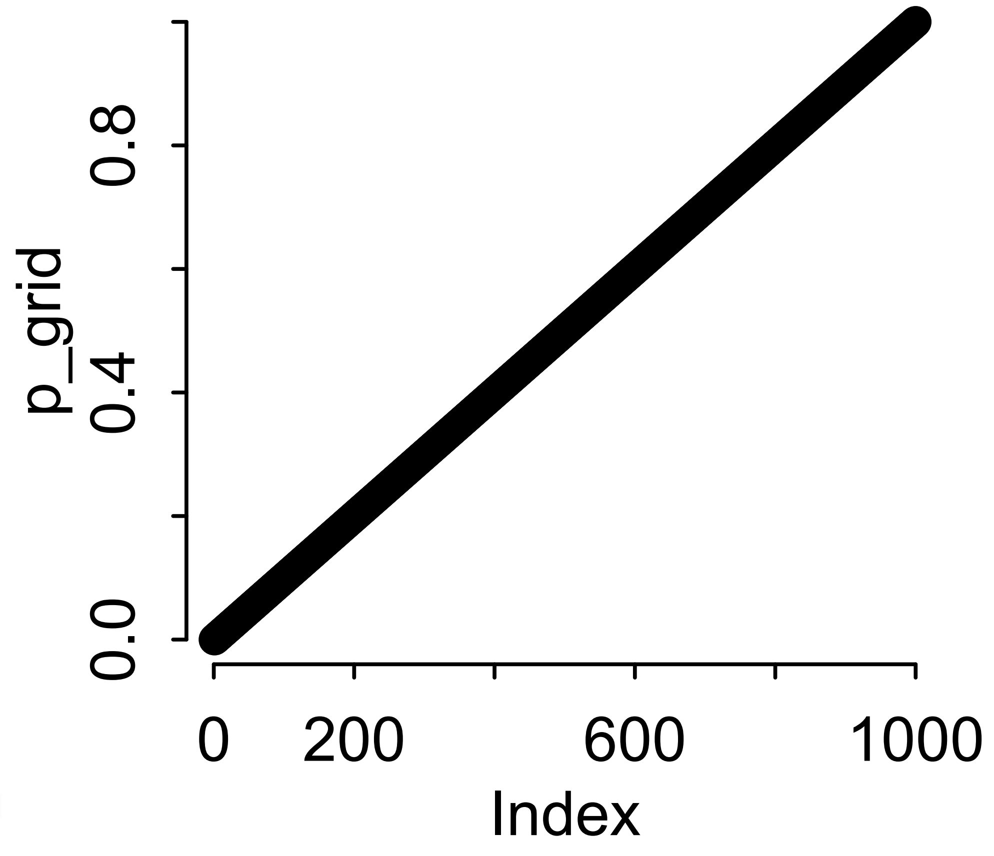


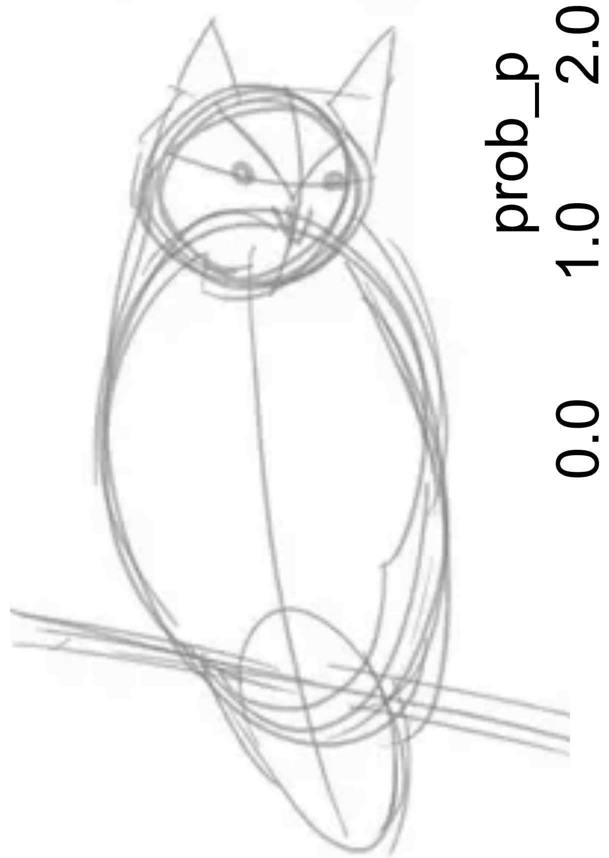


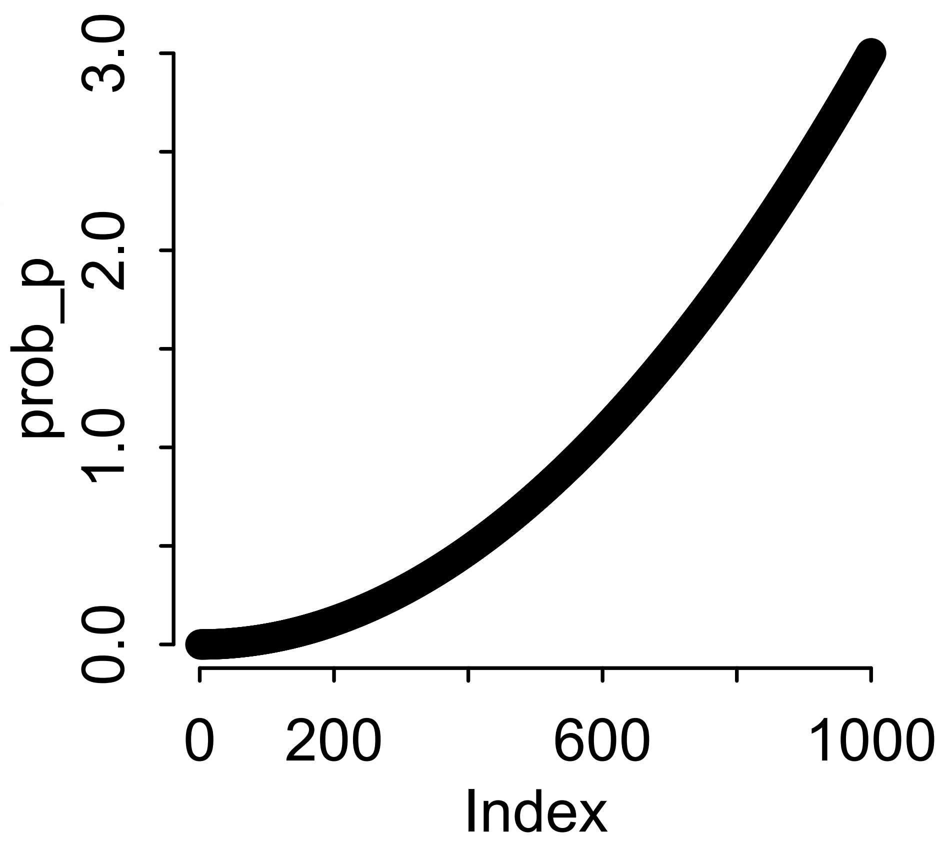


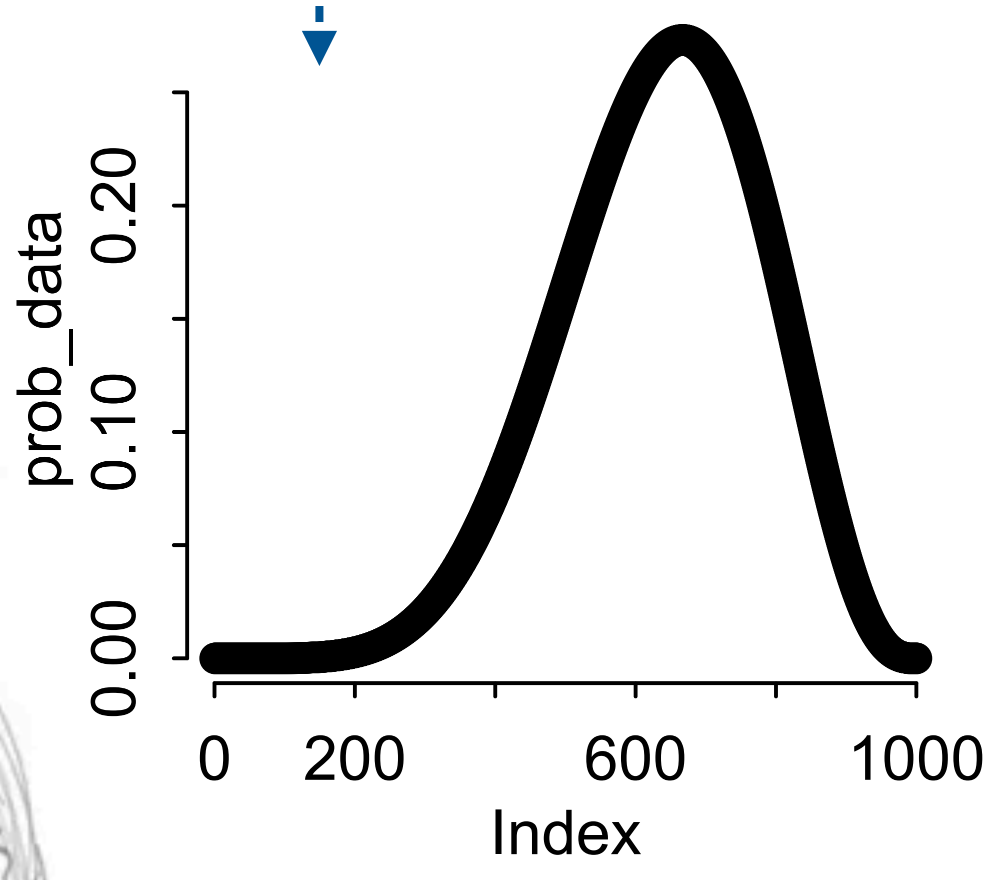


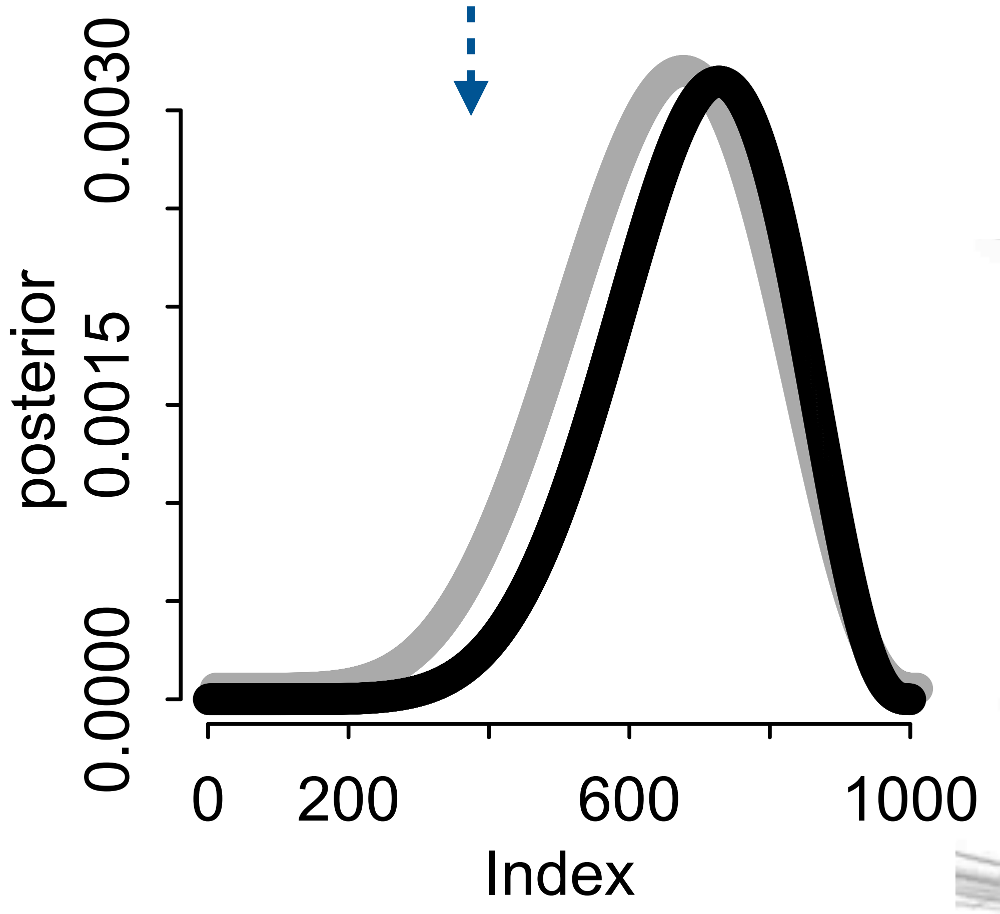
```r
p_grid ← seq( from=0 , to=1 , len=1000 )
prob_p ← dbeta( p_grid , 3 , 1 )
prob_data ← dbinom( 6 , 9 , prob=p_grid )
posterior ← prob_data * prob_p
posterior ← posterior / sum(posterior)
```











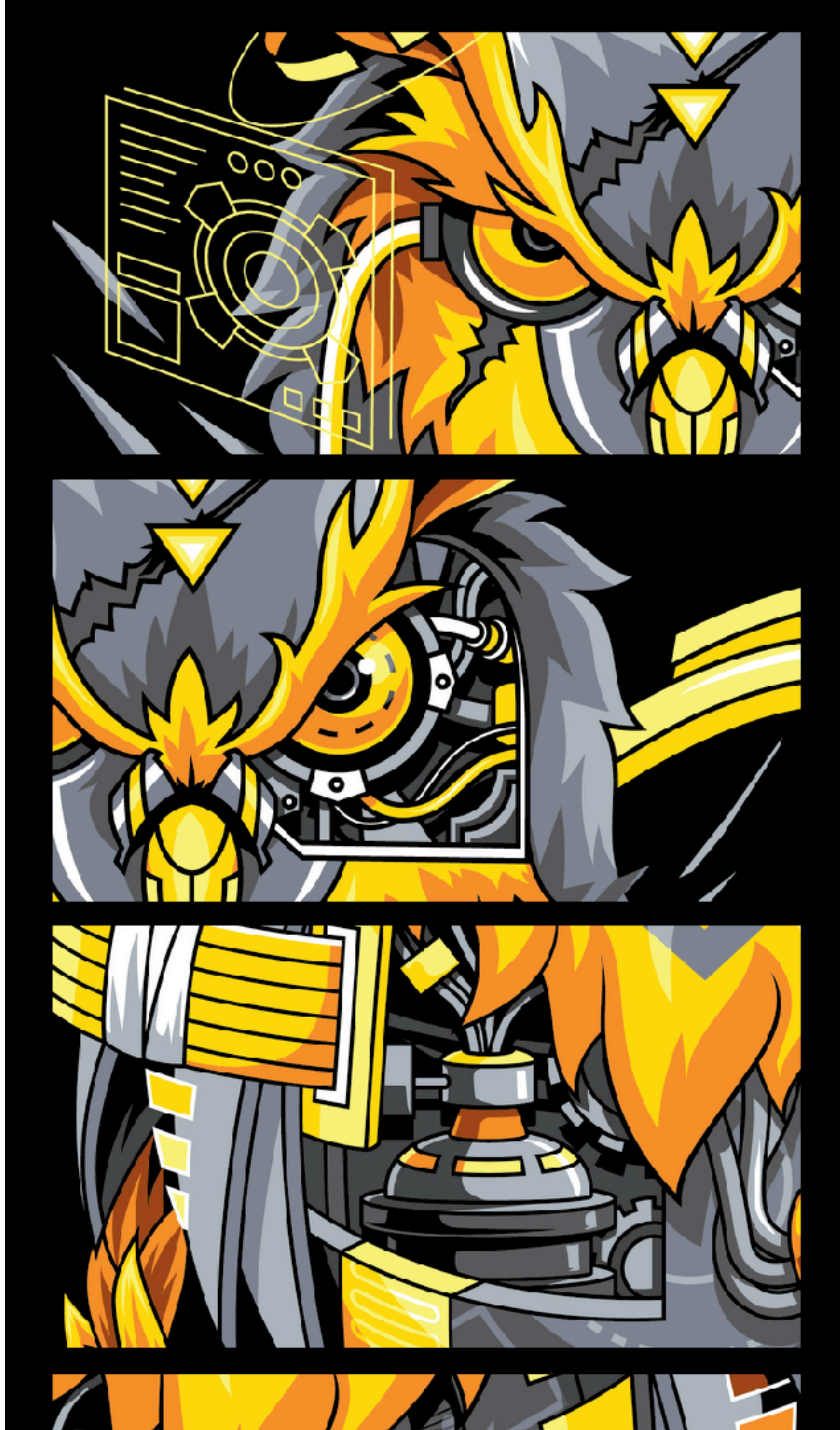
Many Ways to Count

Grid Approximation inefficient

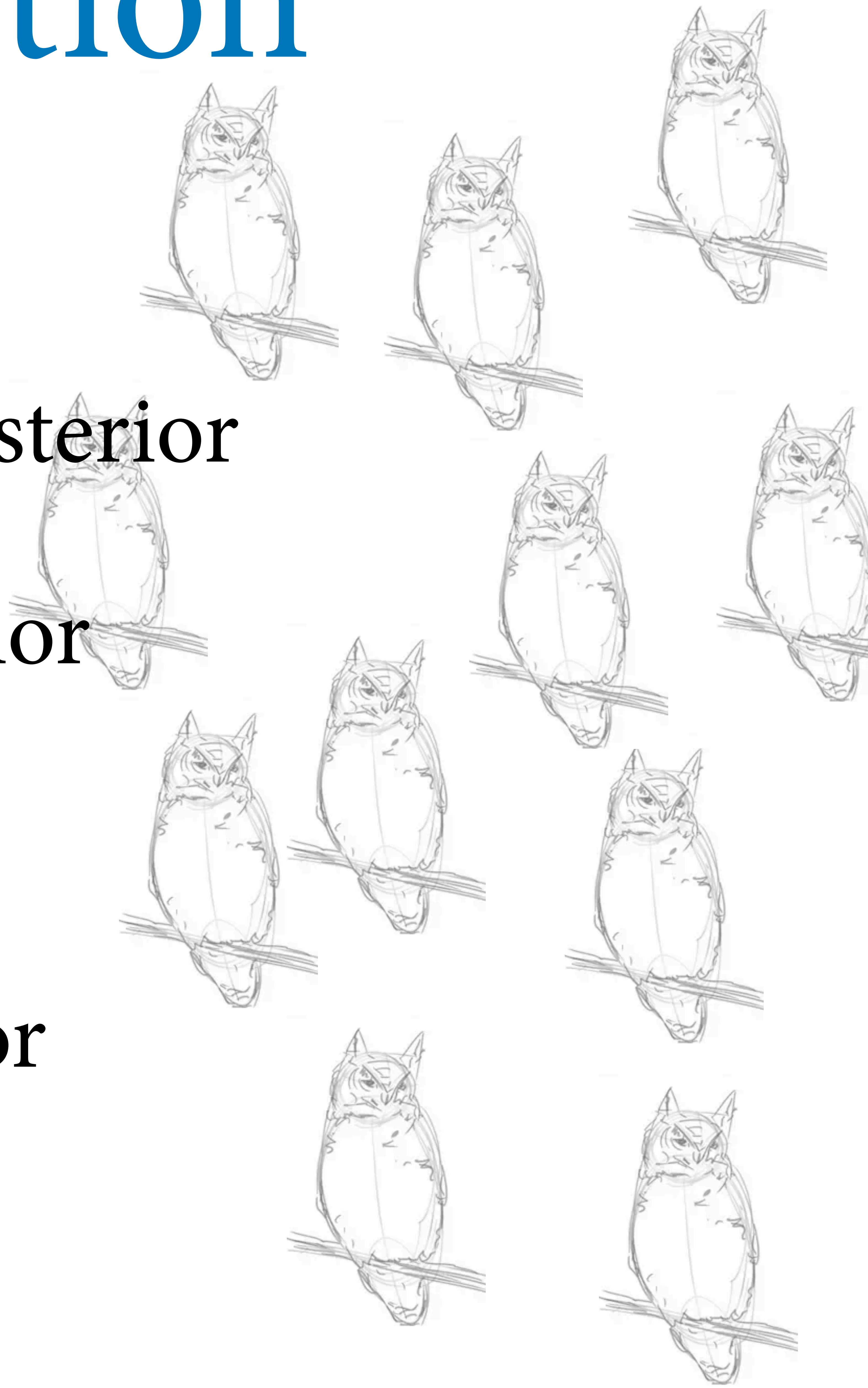
Other methods:

Quadratic approximation

Markov chain Monte Carlo (MCMC)



From Posterior to Prediction



Implications of model depend upon **entire** posterior

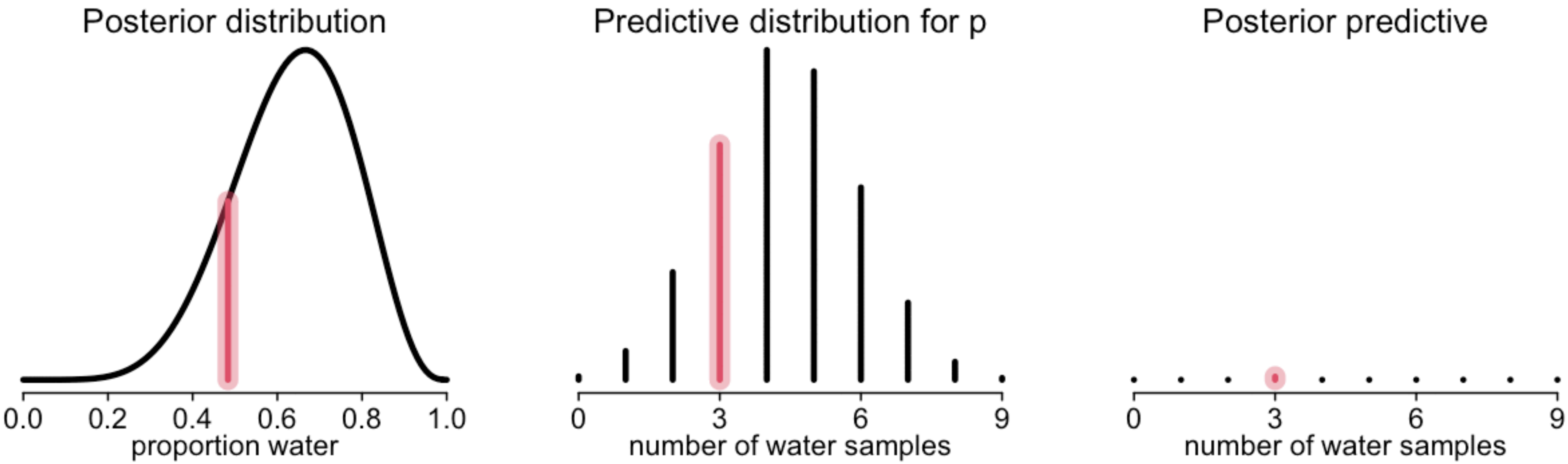
Must average any inference over entire posterior

This usually requires integral calculus

OR we can just take samples from the posterior

Uncertainty ⇨ Causal model ⇨ Implications

Uncertainty ⇨ Causal model ⇨ Implications



Sample from posterior

R code 3.2

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prob_p <- rep( 1 , 1000 )
prob_data <- dbinom( 6 , size=9 , prob=p_grid )
posterior <- prob_data * prob_p
posterior <- posterior / sum(posterior)
```

R code 3.3

```
samples <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

Sample from posterior

R code 3.3

```
samples <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

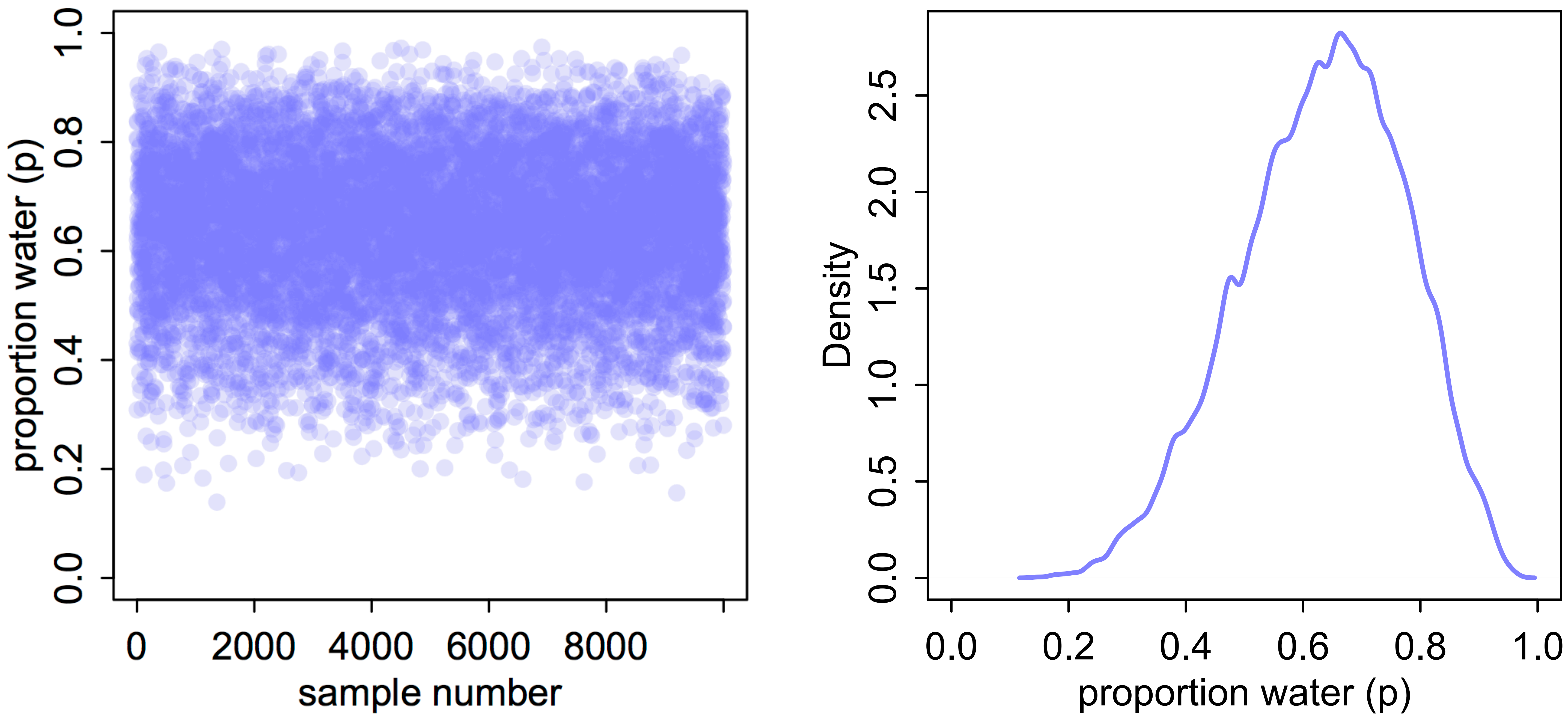


Figure 3.1

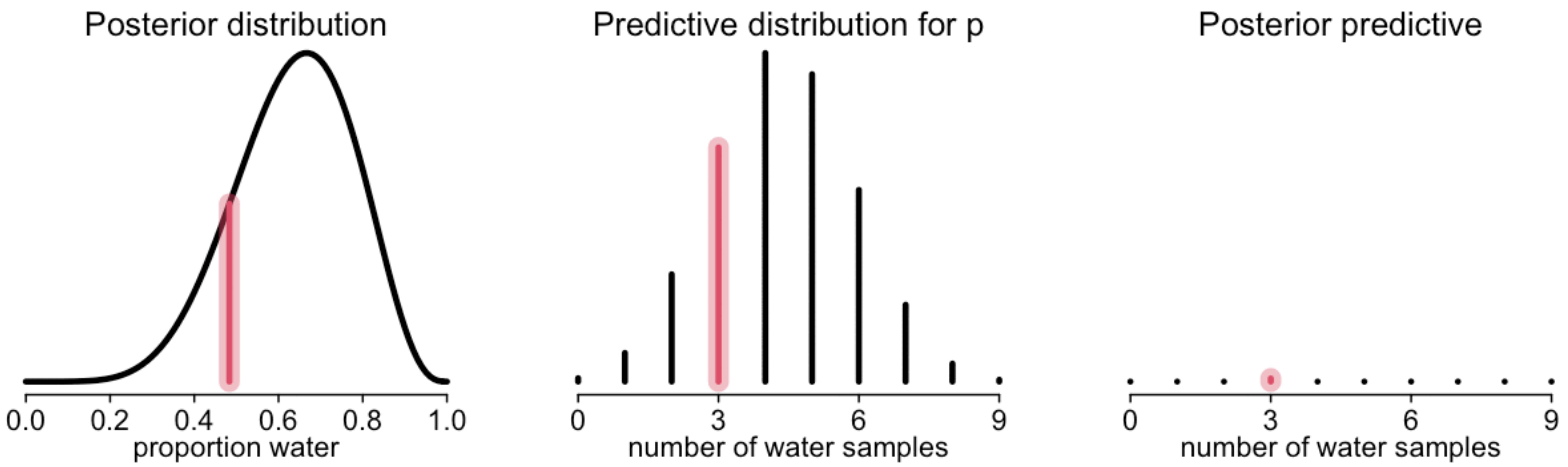
Sample predictions

R code 3.3

```
samples <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

R code 3.26

```
w <- rbinom( 1e4 , size=9 , prob=samples )
```



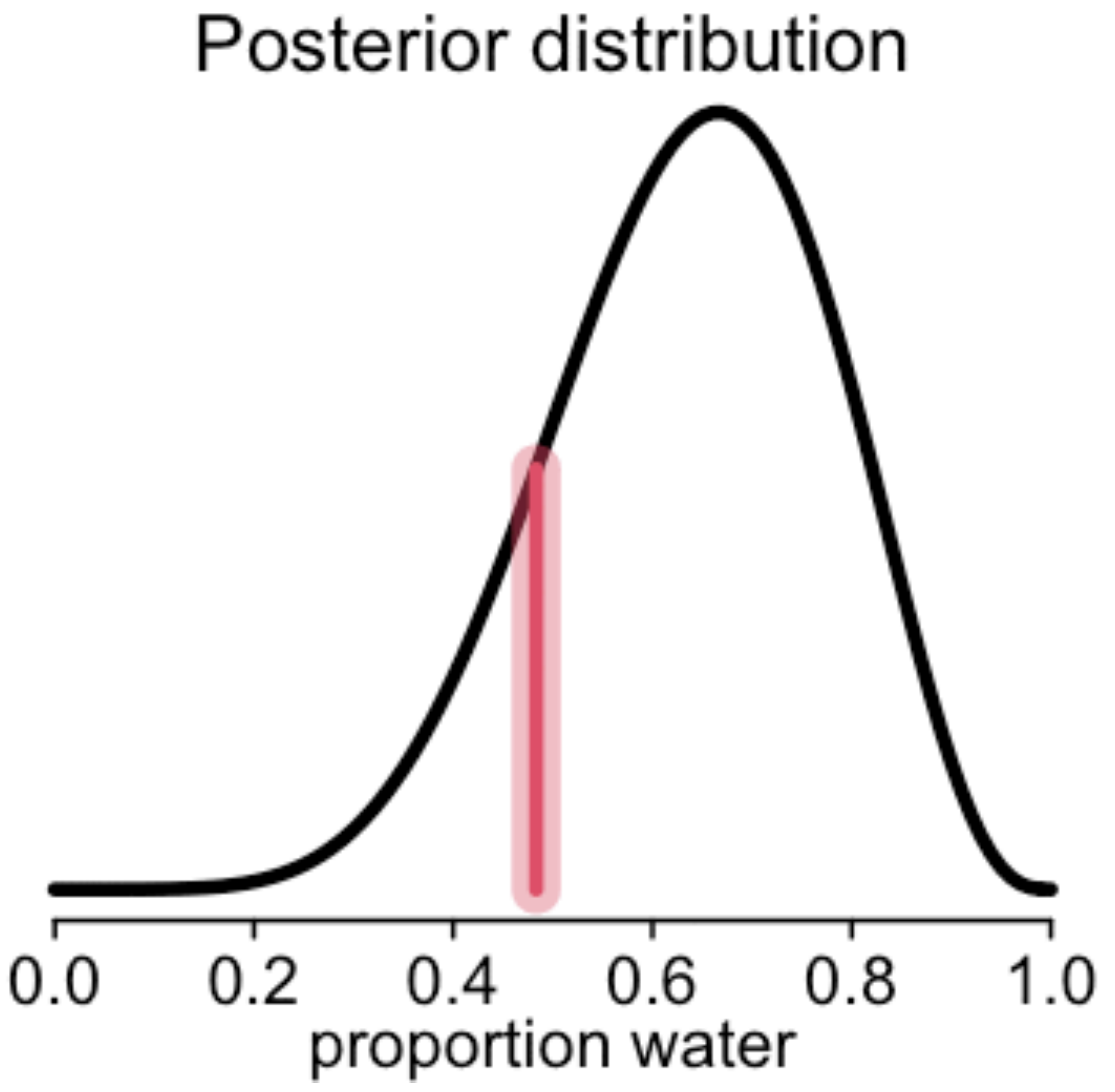
Sampling is Fun & Easy

Sample from posterior, compute desired quantity for each sample, profit

Much easier than doing integrals

Turn a **calculus problem** into a **data summary problem**

MCMC produces only samples anyway



Sampling is Handsome & Handy

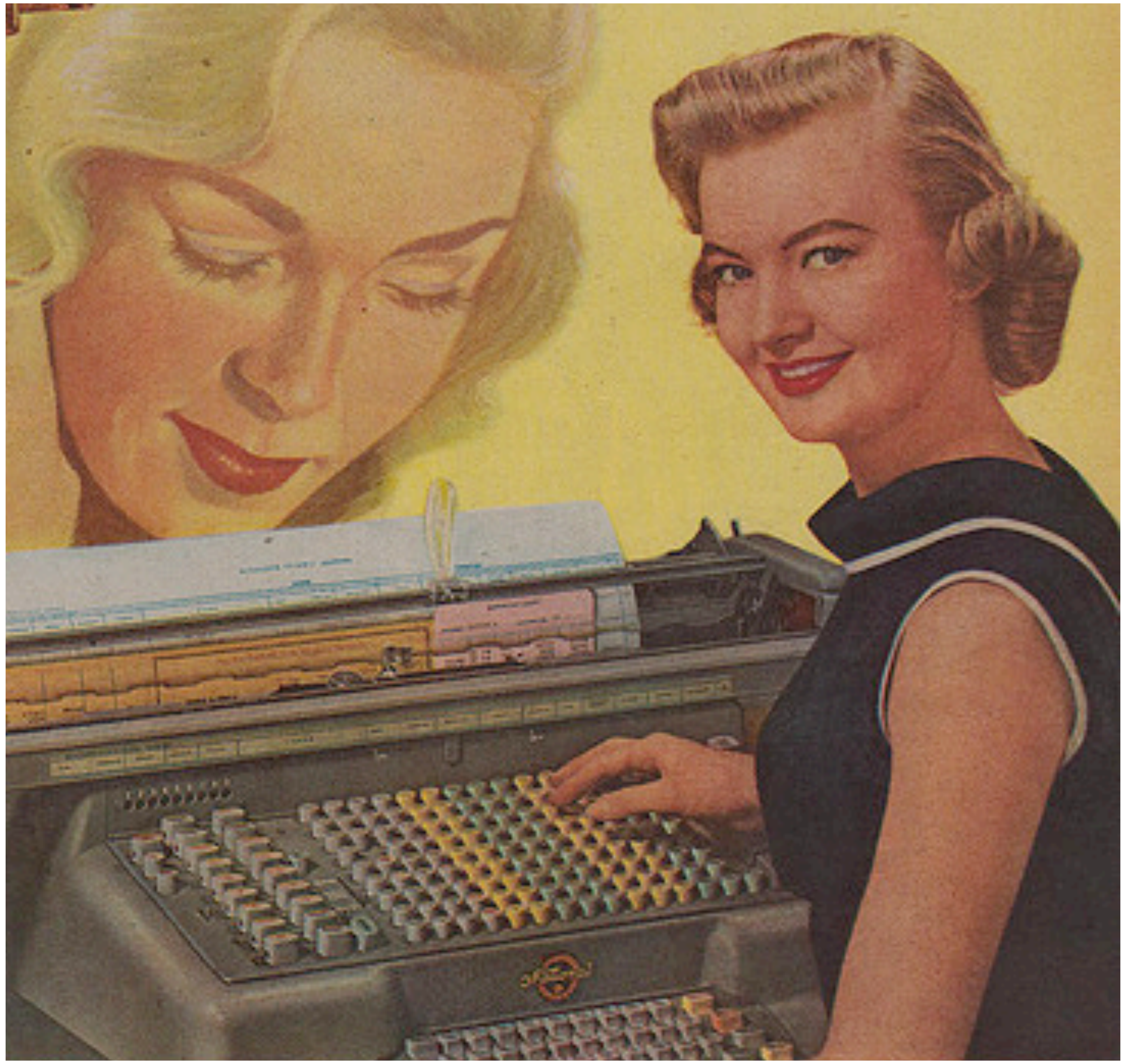
Things we'll compute with sampling:

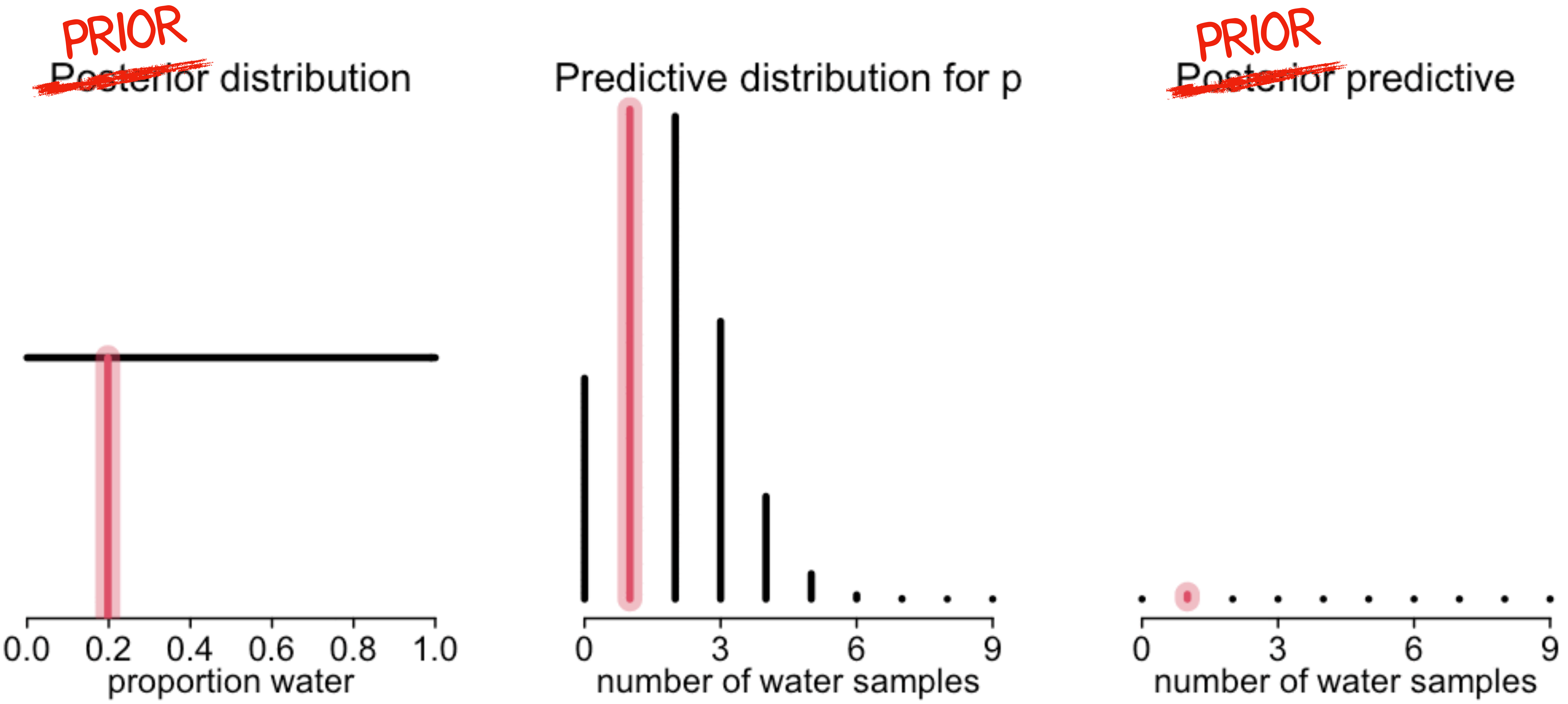
Model-based forecasts

Causal effects

Counterfactuals

Prior predictions ???




PRIOR
~~Posterior~~ distribution
0.0 0.2 0.4 0.6 0.8 1.0
proportion water
Predictive distribution for p
0 3 6 9
number of water samples
PRIOR
~~Posterior~~ predictive
0 3 6 9
number of water samples

Bayesian data analysis

For each possible explanation of the data,

Count all the ways data can happen.

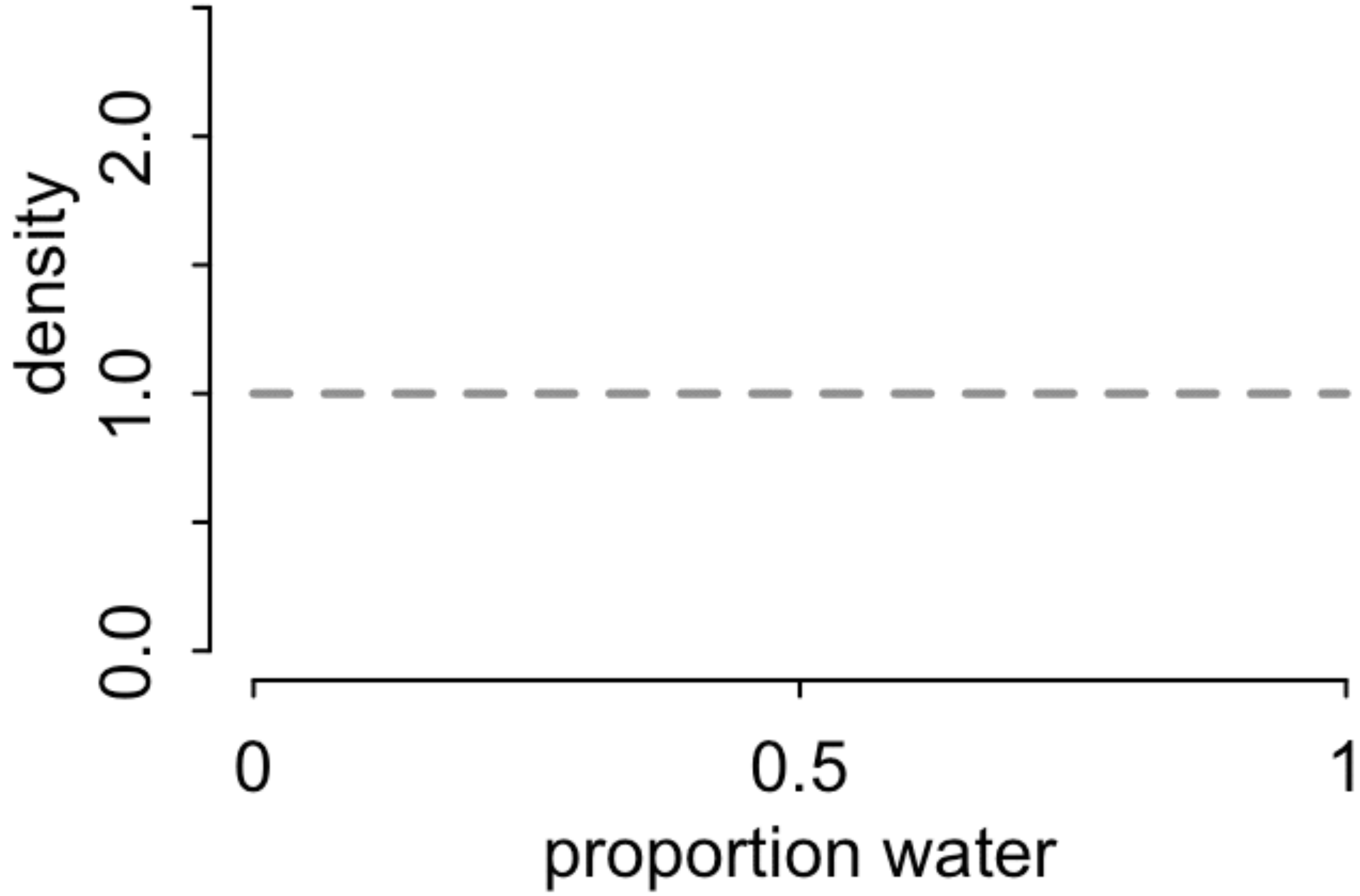
Explanations with more ways to produce the data are more plausible.

Bayesian modesty

No guarantees except ***logical***

Probability theory is a method of logically deducing ***implications of data*** *under assumptions that you must choose*

Any framework selling you more is hiding assumptions



Course Schedule

Week 1	Bayesian inference	Chapters 1, 2, 3
Week 2	Linear models & Causal Inference	Chapter 4
Week 3	Causes, Confounds & Colliders	Chapters 5 & 6
Week 4	Overfitting / Interactions	Chapters 7 & 8
Week 5	MCMC & Generalized Linear Models	Chapters 9, 10, 11
Week 6	Integers & Other Monsters	Chapters 11 & 12
Week 7	Multilevel models I	Chapter 13
Week 8	Multilevel models II	Chapter 14
Week 9	Measurement & Missingness	Chapter 15
Week 10	Generalized Linear Madness	Chapter 16

`https://github.com/rmcelreath/statrethinking_2022`