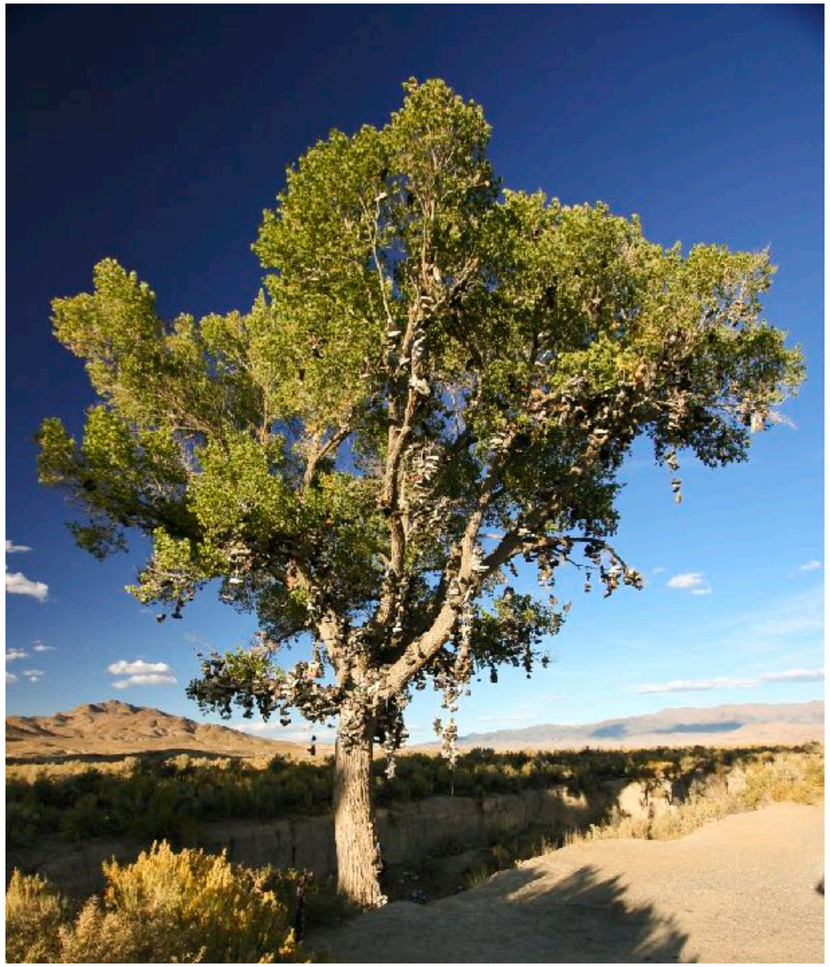
# Traits on trees



Cornell Arbor workshop 2018

Luke Harmon

Part 1:
Continuous
traits

Part 2:
Discrete
traits

Part 1:
Continuous
traits

Part 2:
Discrete
traits

# What is Brownian motion?

# Brownian Motion: The Model

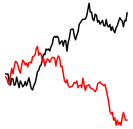
- Describes a “random walk” of evolution for continuously-valued characters

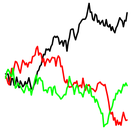
# Three Facts Describe Brownian Motion

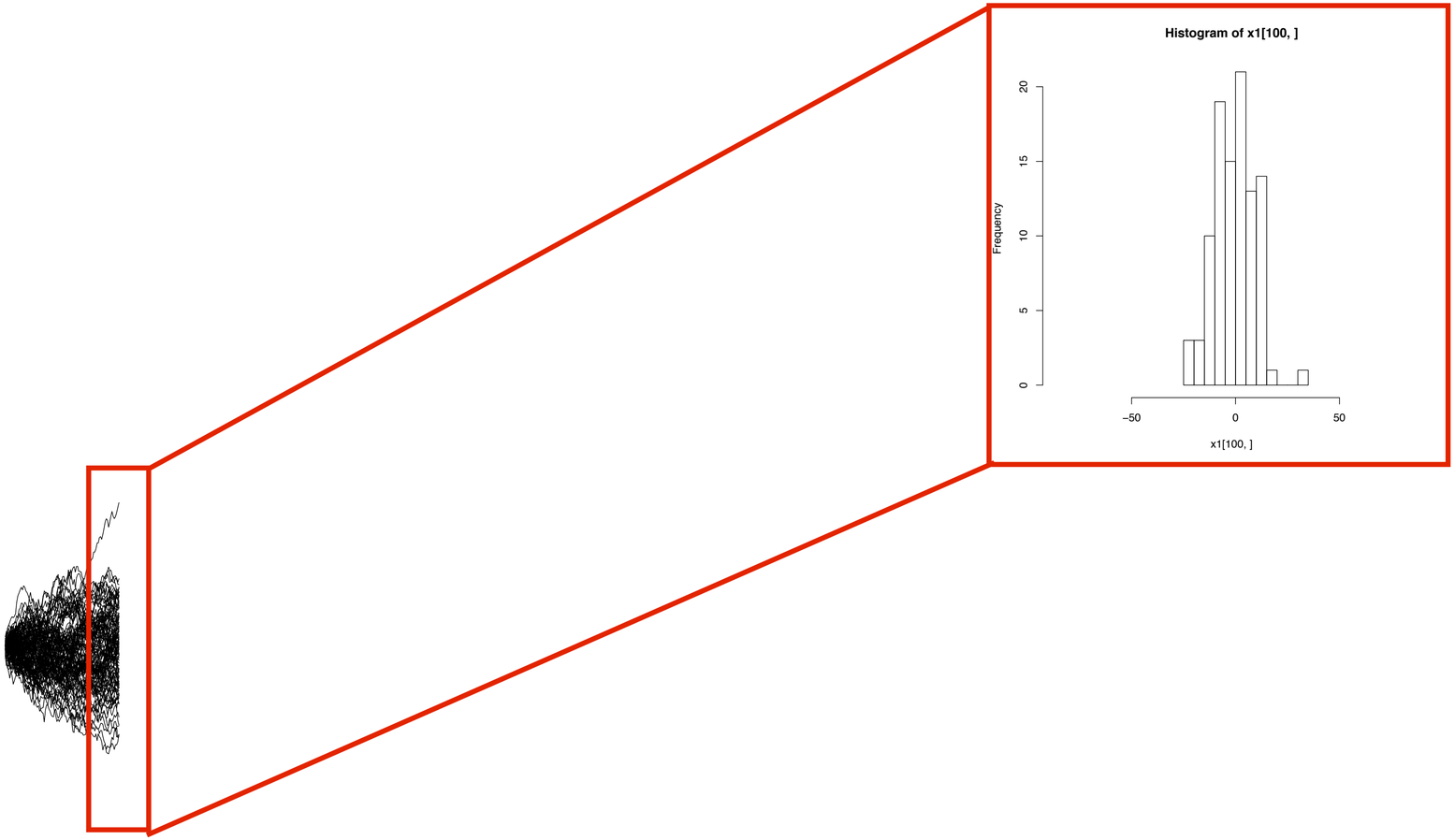
- No trend
- Successive steps independent
- Expected values follow normal distribution where variance depends on rate and time

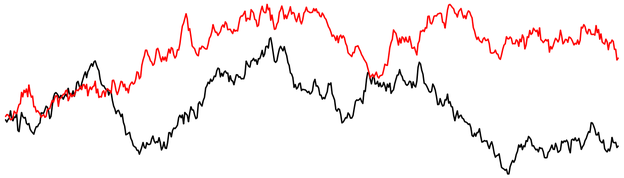
# Parameters of BM

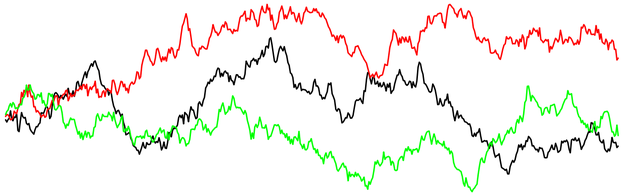
- Brownian motion models have two parameters:
  - $\Theta$, the starting value; $W(0) = \Theta$
  - $\sigma^2$, the rate parameter

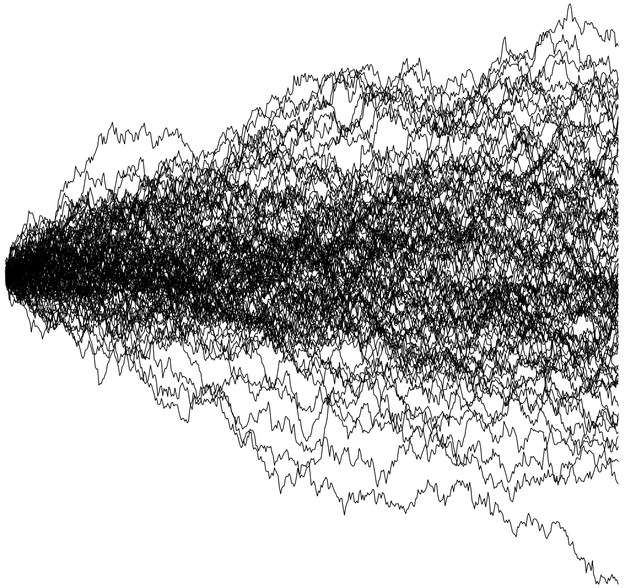


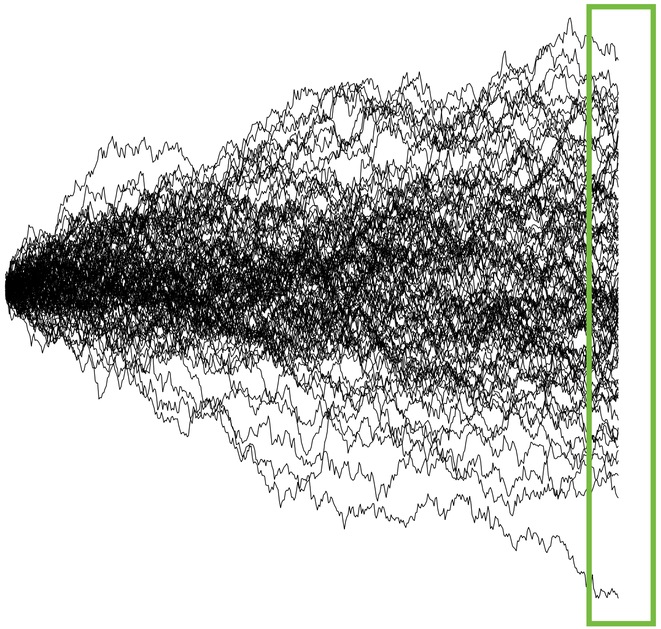


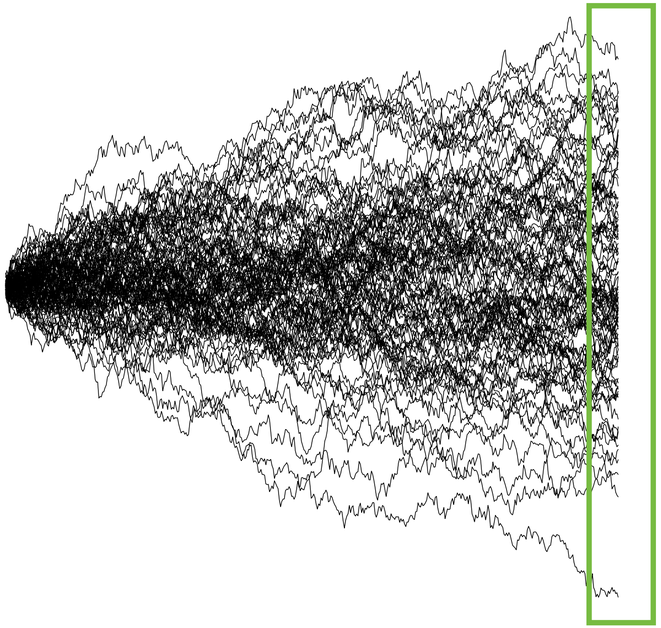

Histogram of x1[100, ]
Frequency
0
5
10
15
20
−50
0
50
x1[100, ]

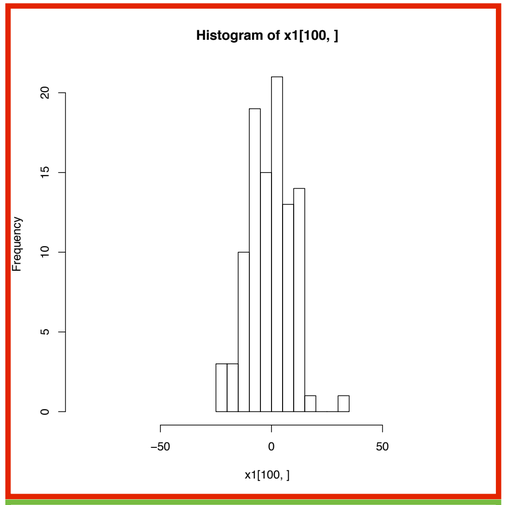


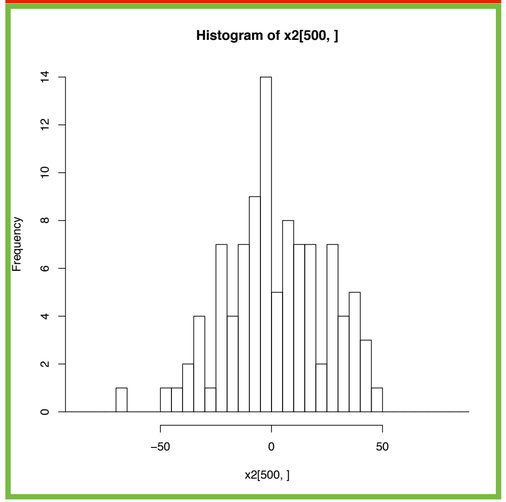


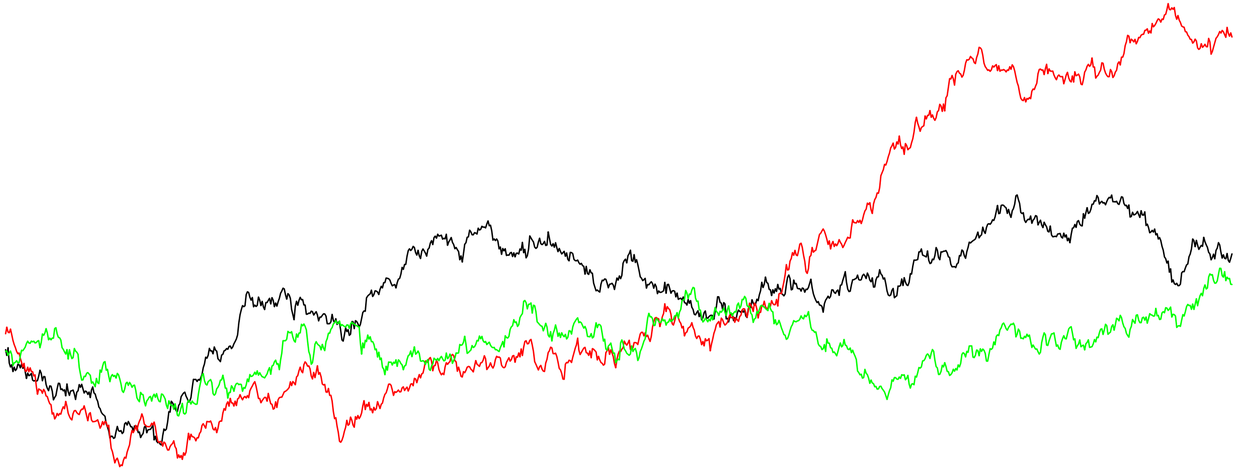


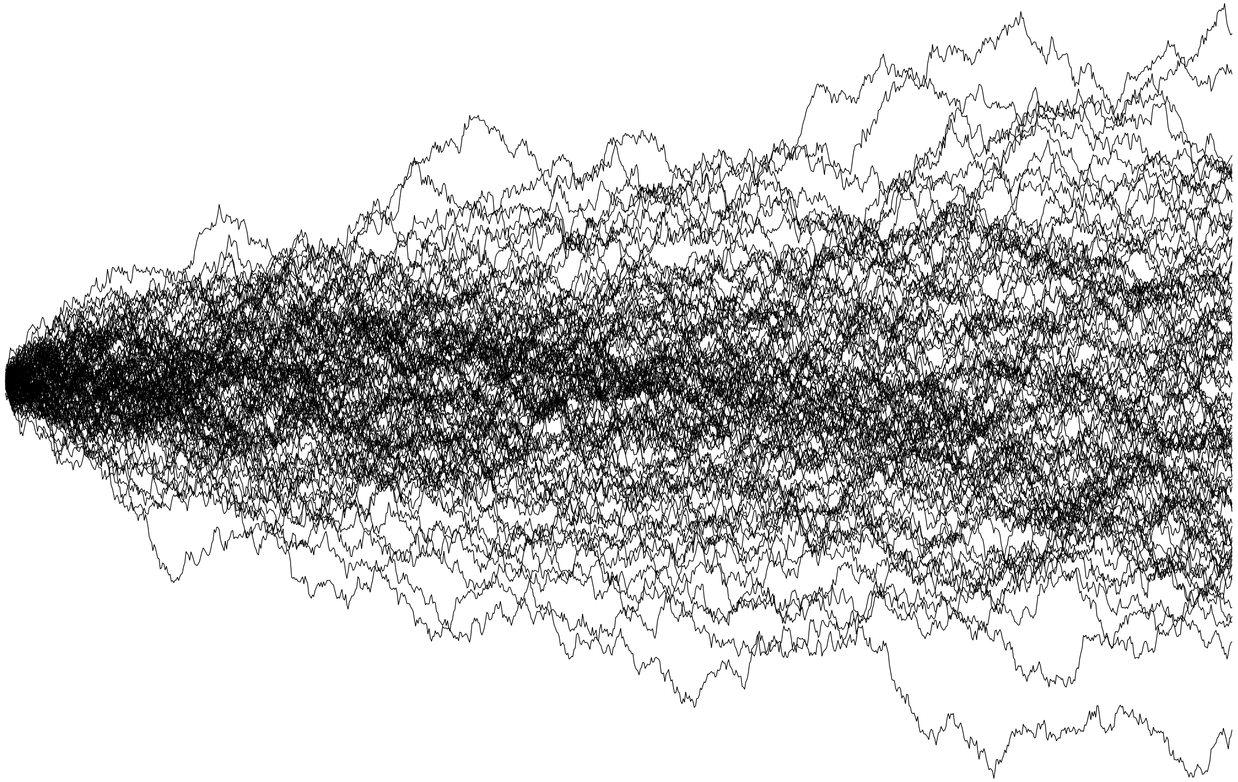


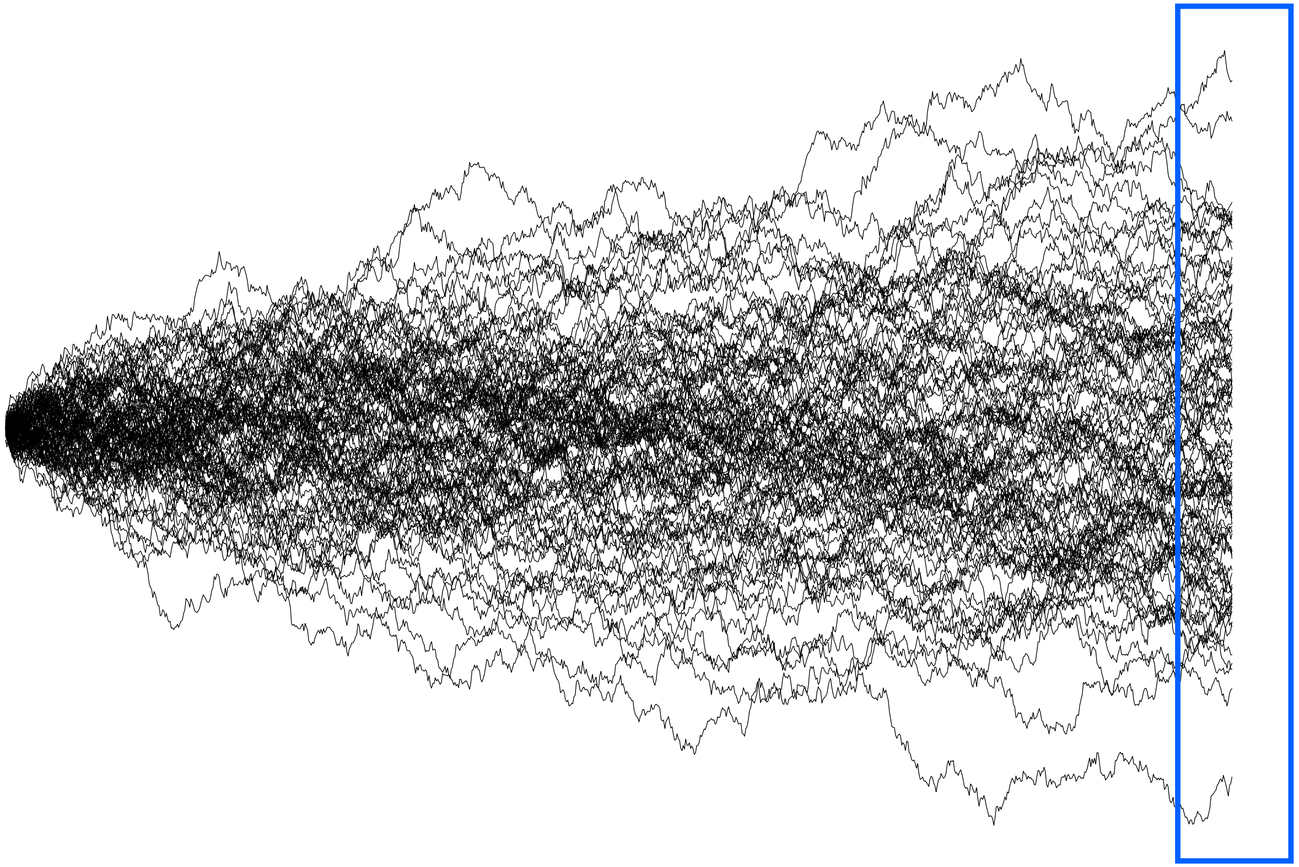


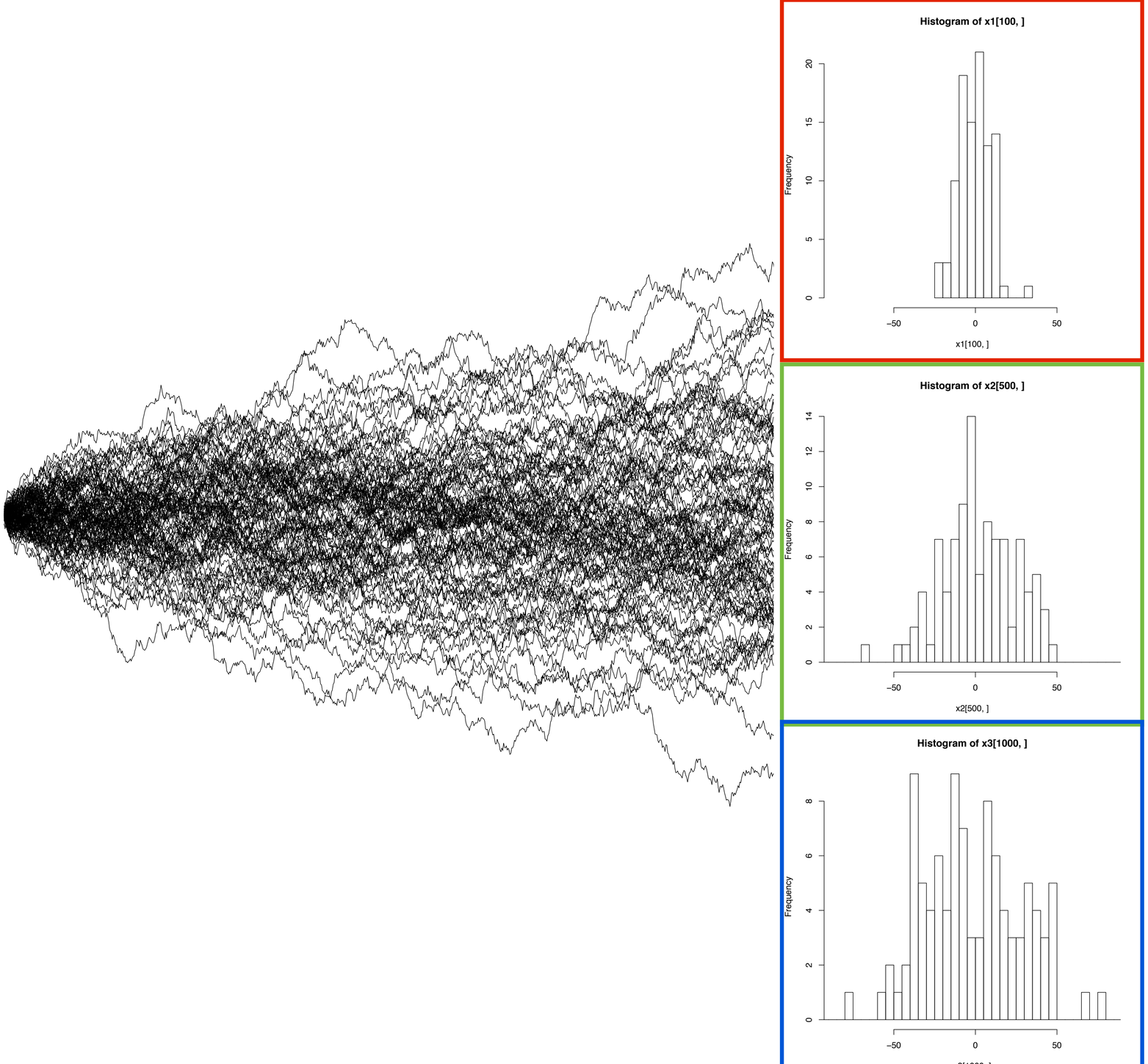

Histogram of x1[100, ]
Frequency
0
5
10
15
20
−50
0
50
x1[100, ]

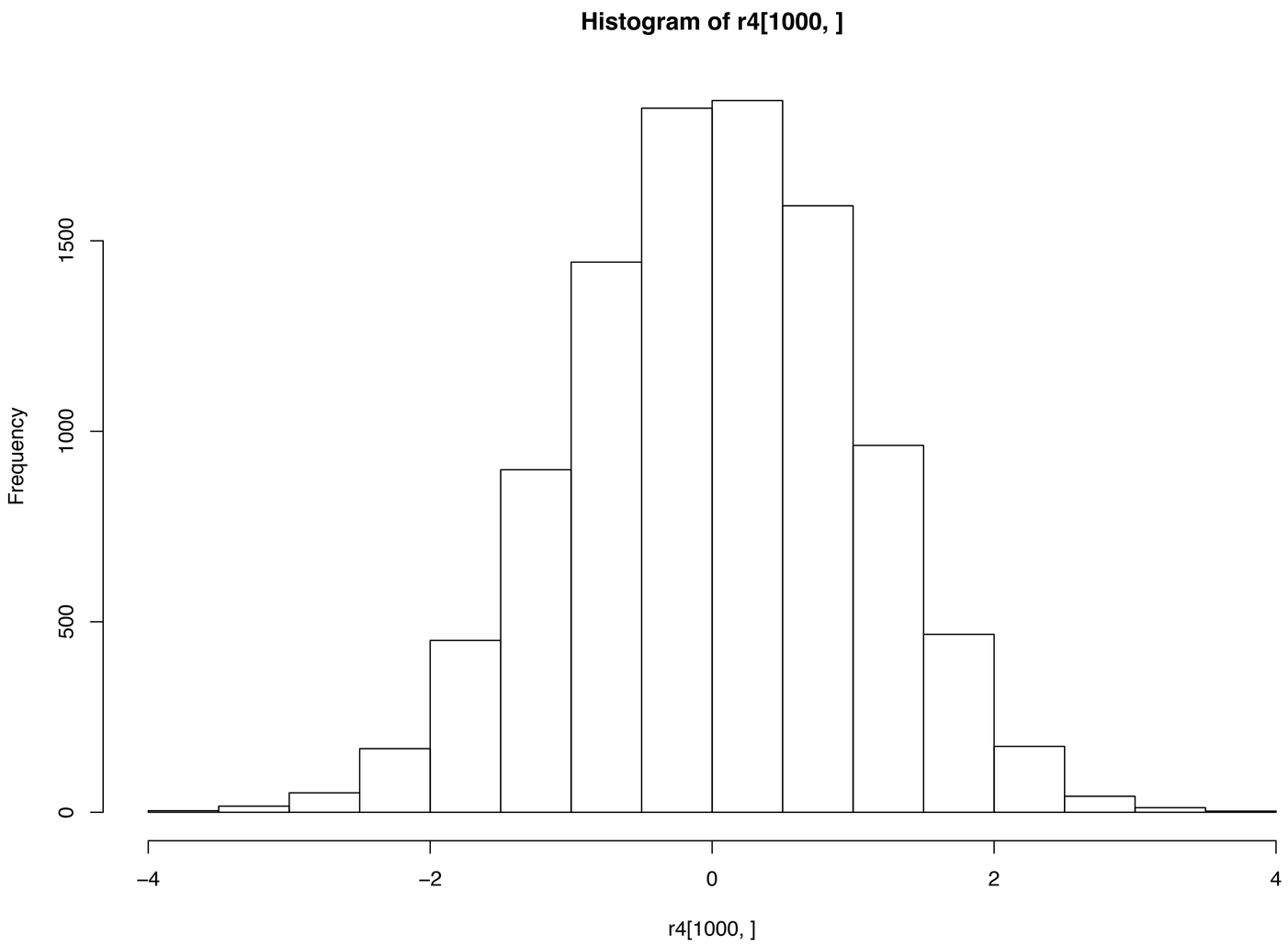

Histogram of x2[500, ]
Frequency
0
2
4
6
8
10
12
14
−50
0
50
x2[500, ]

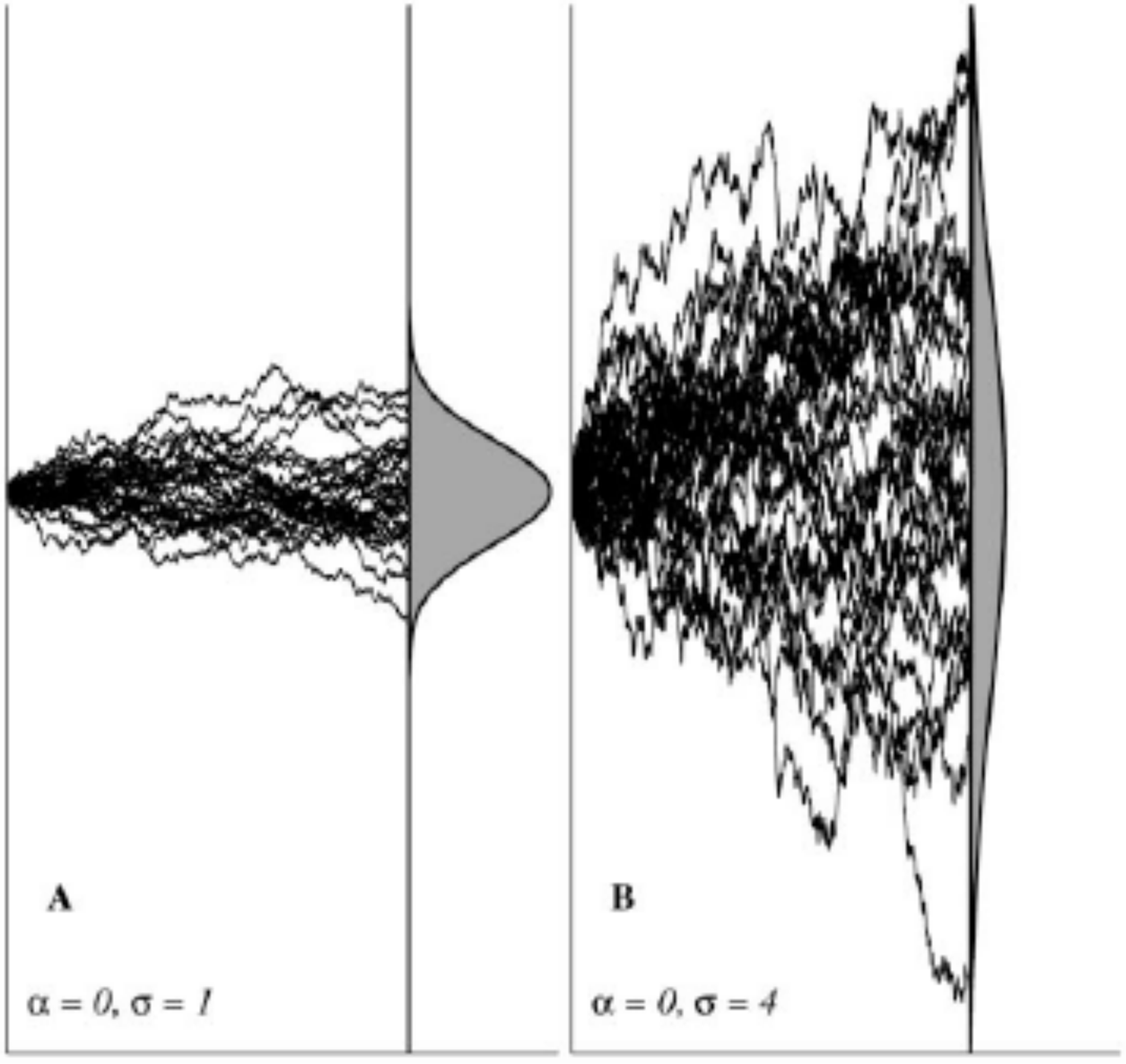







Histogram of x1[100, ]
Frequency
0
5
10
15
20
−50
0
50
x1[100, ]
Histogram of x2[500, ]
Frequency
0
2
4
6
8
10
12
14
−50
0
50
x2[500, ]
Histogram of x3[1000, ]
Frequency
0
2
4
6
8
−50
0
50


Histogram of r4[1000, ]
Frequency
0
500
1000
1500
−4
−2
0
2
4
r4[1000, ]


A
α = 0, σ = 1
B
α = 0, σ = 4

# Why use this model for trait evolution?

# A physical model for BM

# Why Normal?

- BM can be used to describe motion that results from the combination of a large number of independent weak forces
- Adding many small independent variables result in normal distributions, no matter the original distribution (Central limit theorem)

# Evolution might approximate BM...

- Genetic drift
- Random punctuated change
- Selection that is weak relative to the time interval considered
- Selection that changes randomly through time

# Phylogenetic signal

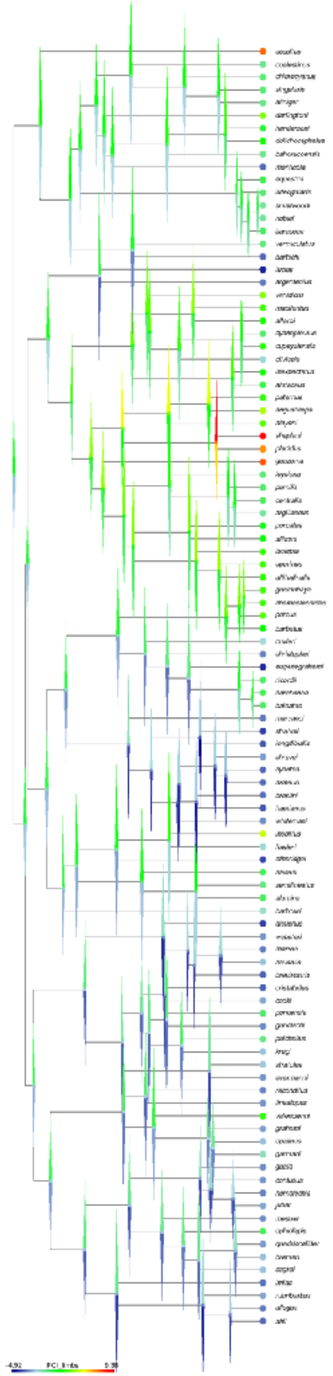
- Phylogenetic signal = close relatives resemble one another
- Can be model based or test-statistic-based
- Often based on comparisons to Brownian motion (which represents “high signal”)

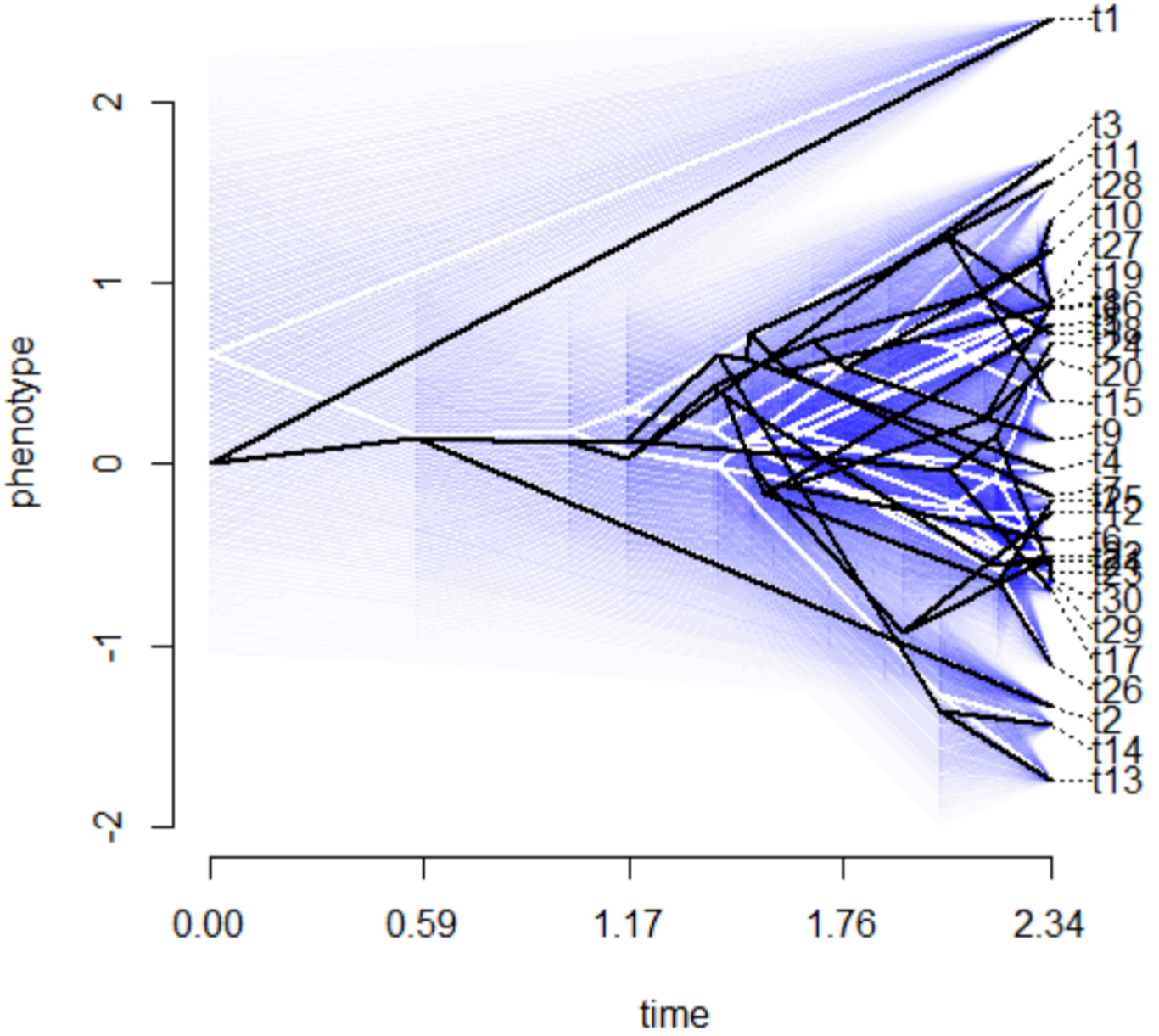
Phylogenetic signal is a pattern not a process

Phylogenetic constraint?

# Ancestral states

- Reconstruct ancestral states at internal nodes on the tree
- Assume a particular model of evolution
- Important to capture uncertainty

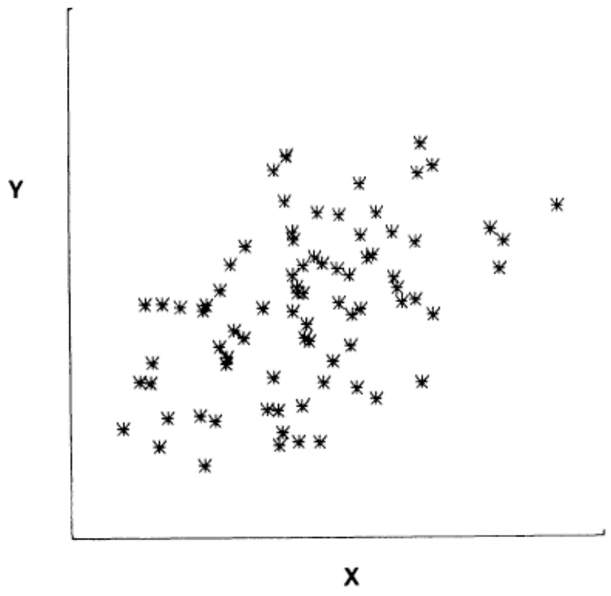

-4.92
PC1 Entry
5.39


phenotype
2
1
0
-1
-2
0.00
0.59
1.17
1.76
2.34
time
t1
t3
t11
t28
t10
t27
t19
t20
t15
t9
t4
t2
t14
t13

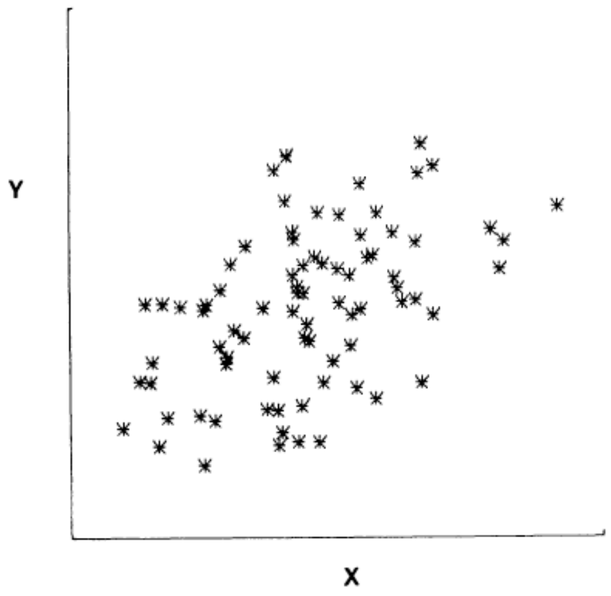
# Phylogenetic Independent Contrasts

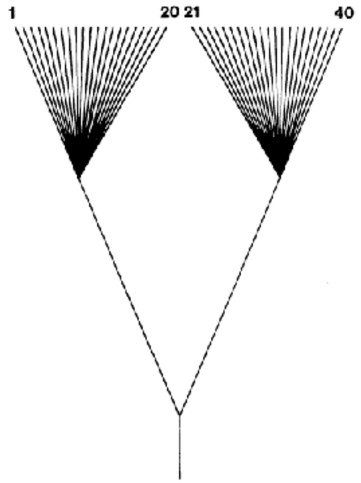
*A simple way to understand the evolution of traits evolving under Brownian motion*

# Felsenstein's "Worst Case"

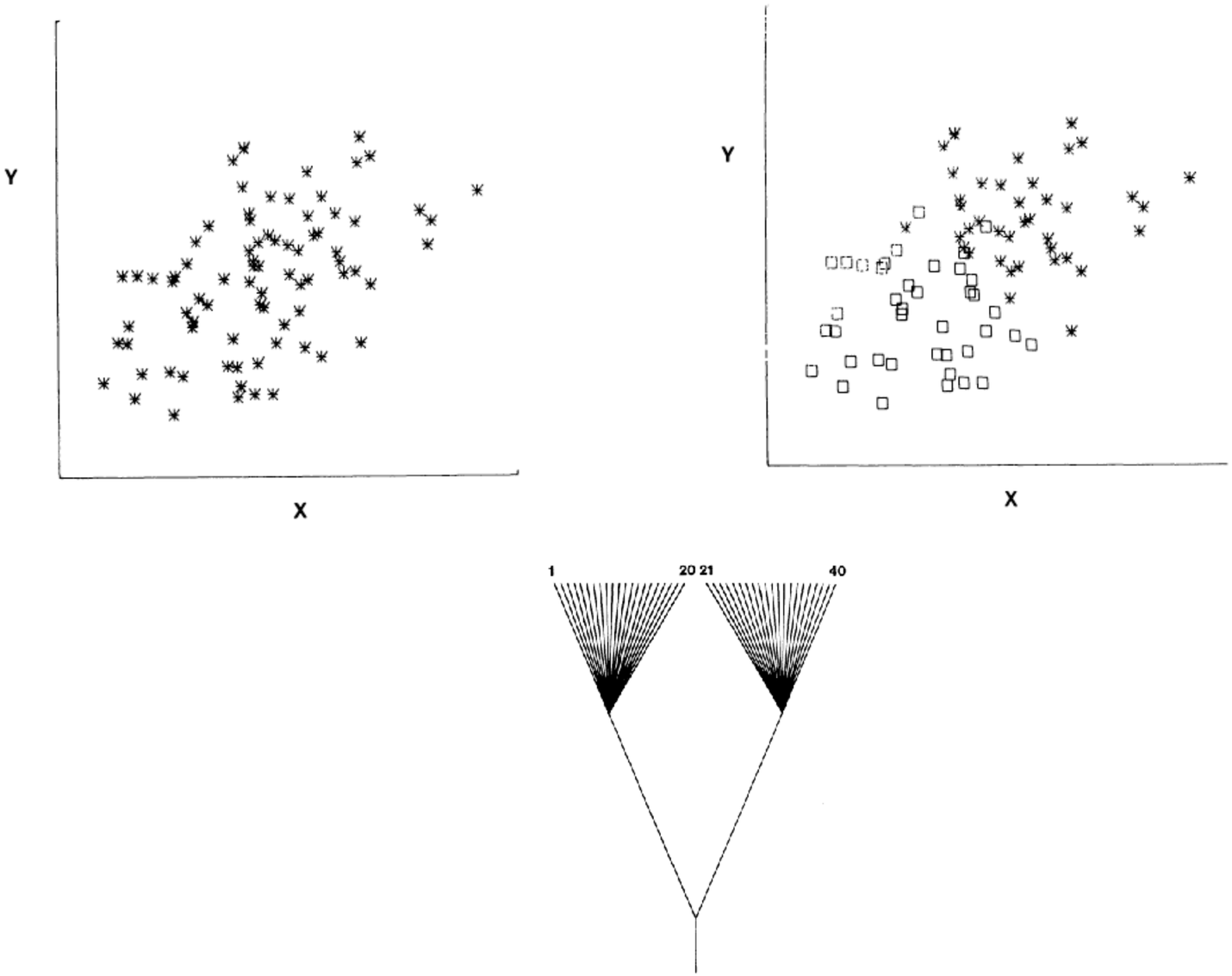


# Felsenstein's "Worst Case"



# Felsenstein's "Worst Case"



# Phylogenetic Independent Contrasts

- PICs are a way to analyze data that comes from phylogenetic trees
- Test for *evolutionary correlations* among characters
- We can think of PICs as a statistical transformation that creates independent data points

# Standard correlation

- Can we predict Y from X?
- We might be able to do this for two reasons: species are related, or X and Y tend to evolve together

# Evolutionary correlation

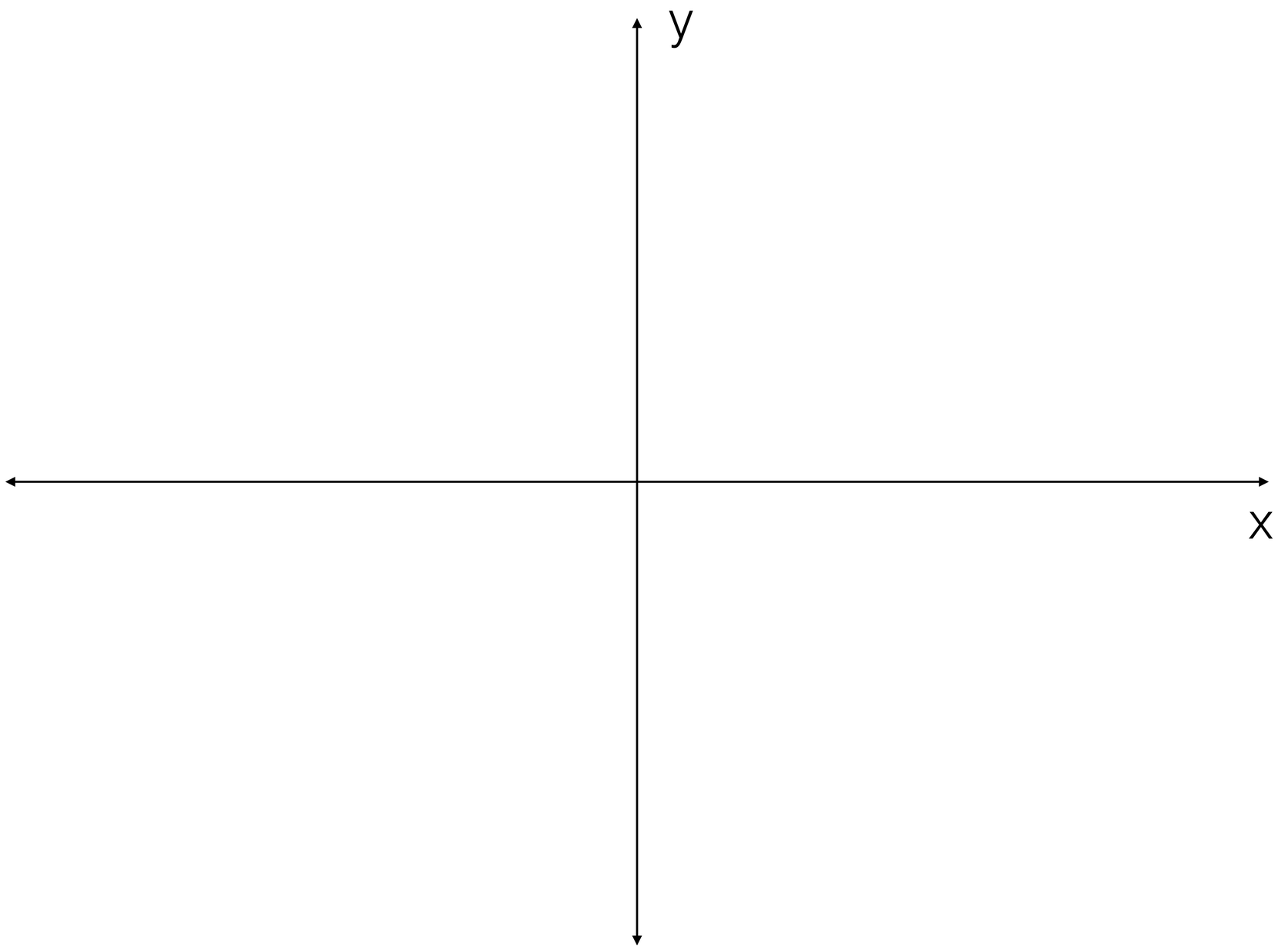
- X and Y evolve in a correlated fashion
- When X changes, Y tends to change in a predictable way

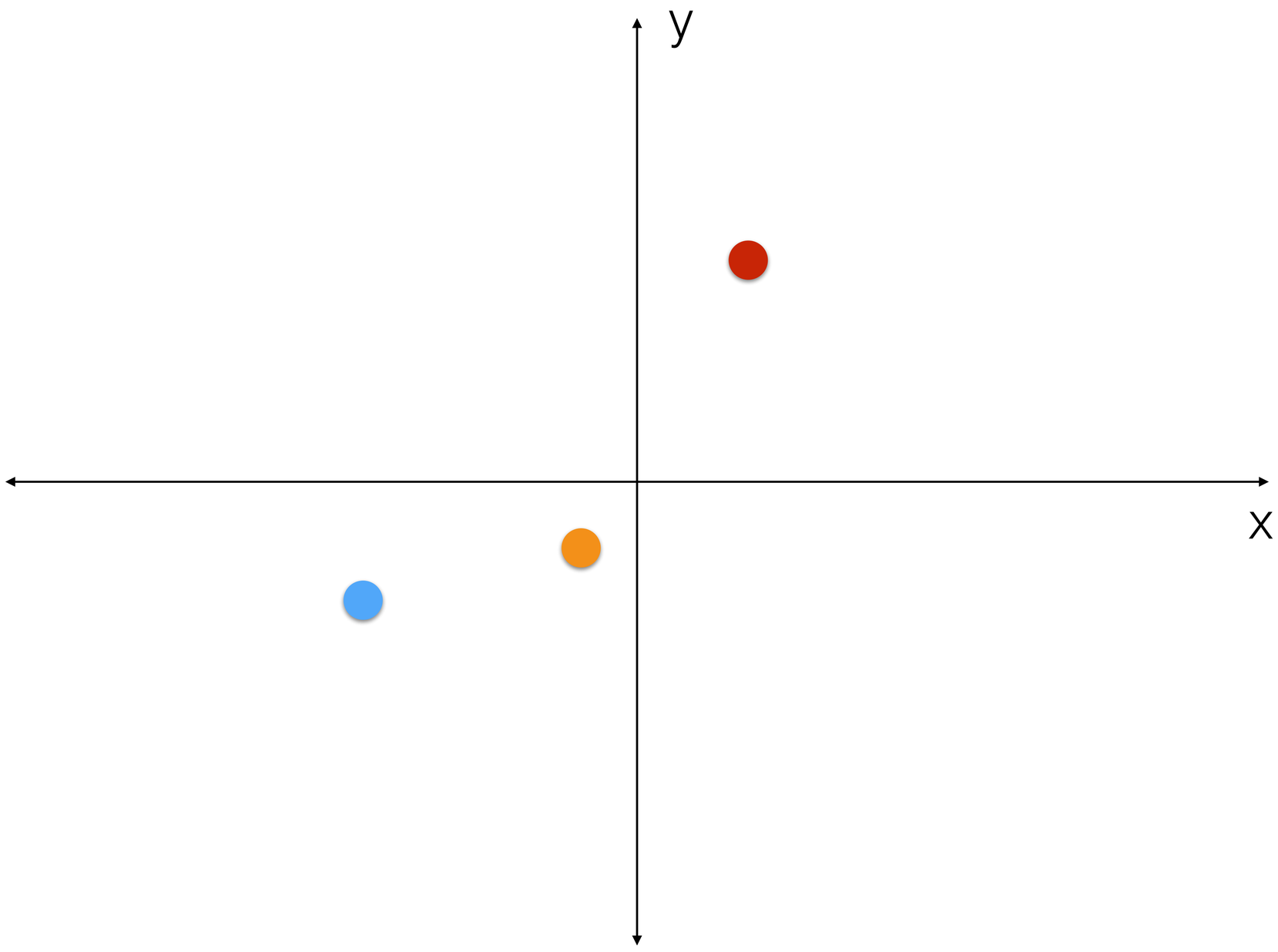
# Character correlations

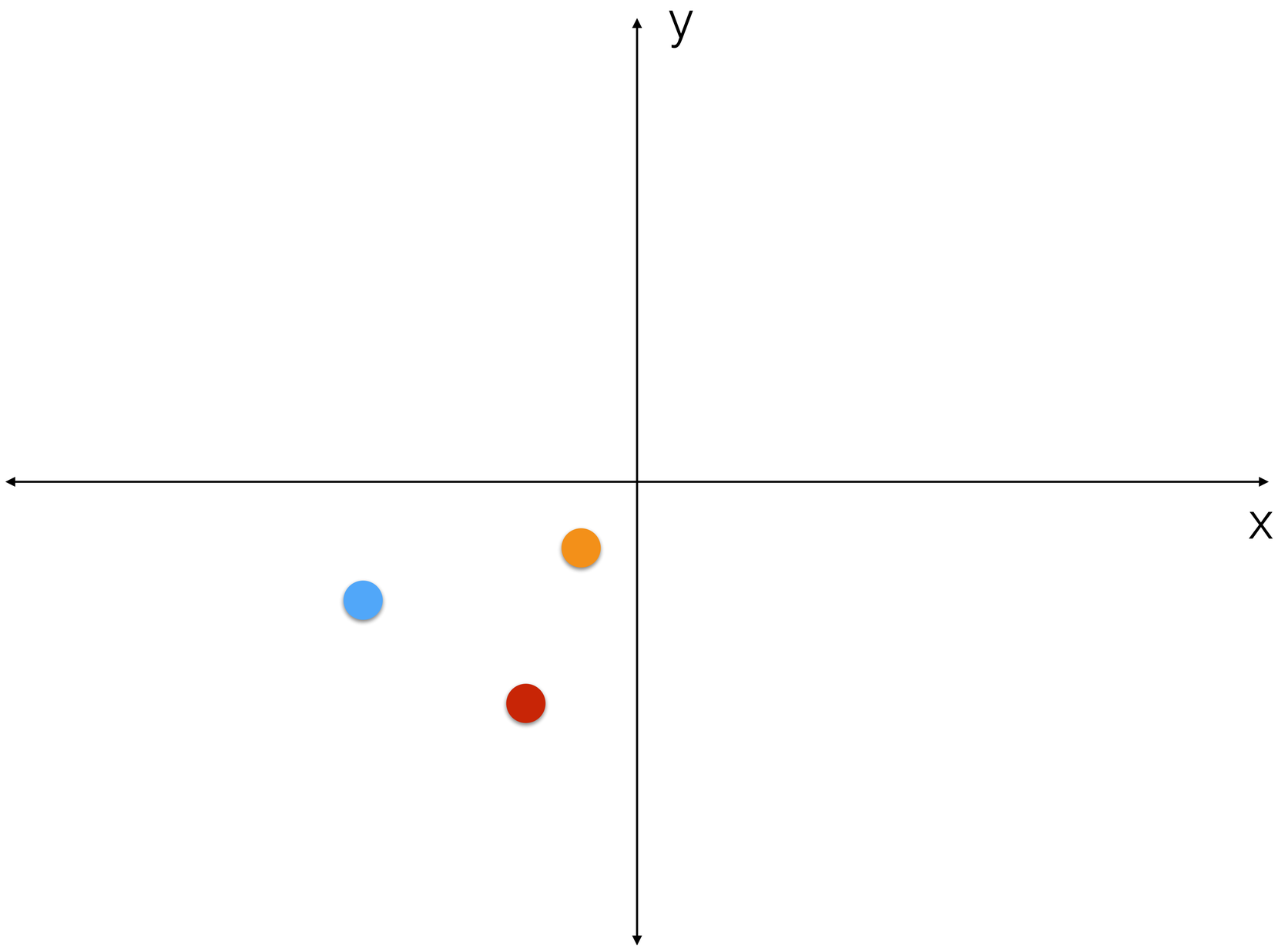
- Most common use for ICs: testing for character correlations
- Are two characters evolving in a correlated fashion?

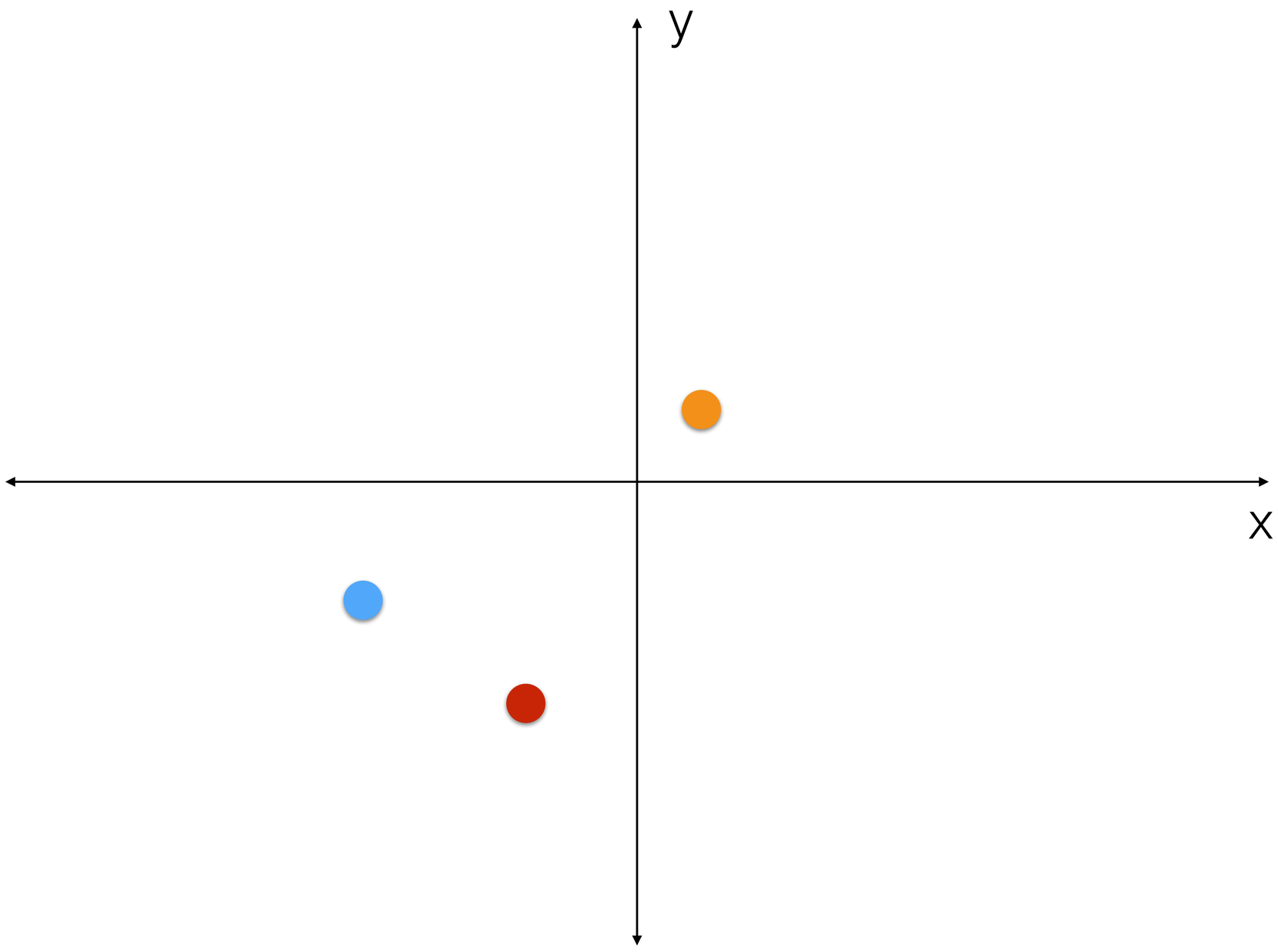
# Character correlations

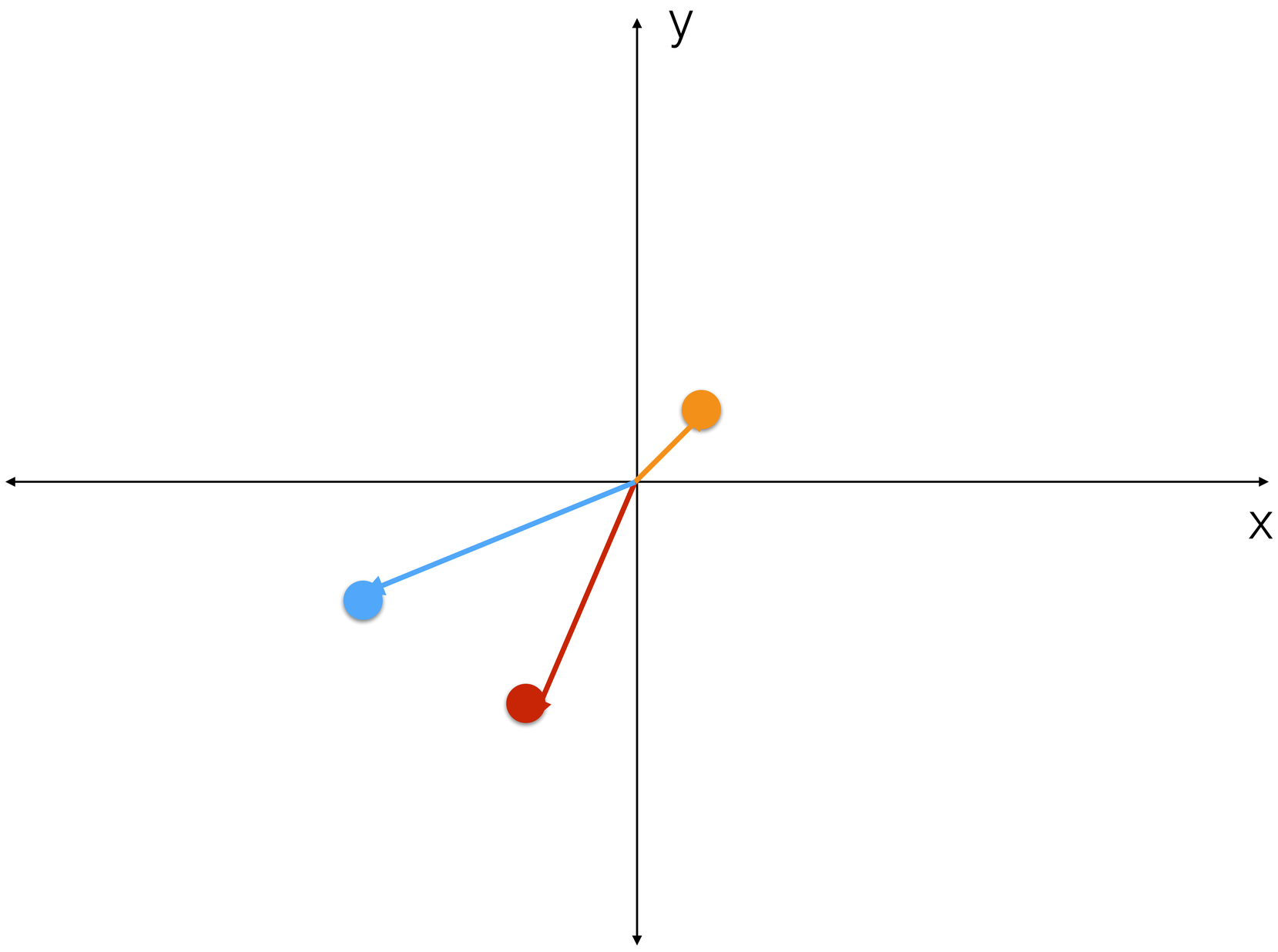
- Calculate independent contrasts for two characters, *x* and *y*
- Carry out a regression analysis of *y* on *x* with **no intercept** (force regression line through the origin)
- $P < 0.05$, then reject the null hypothesis of no evolutionary correlation

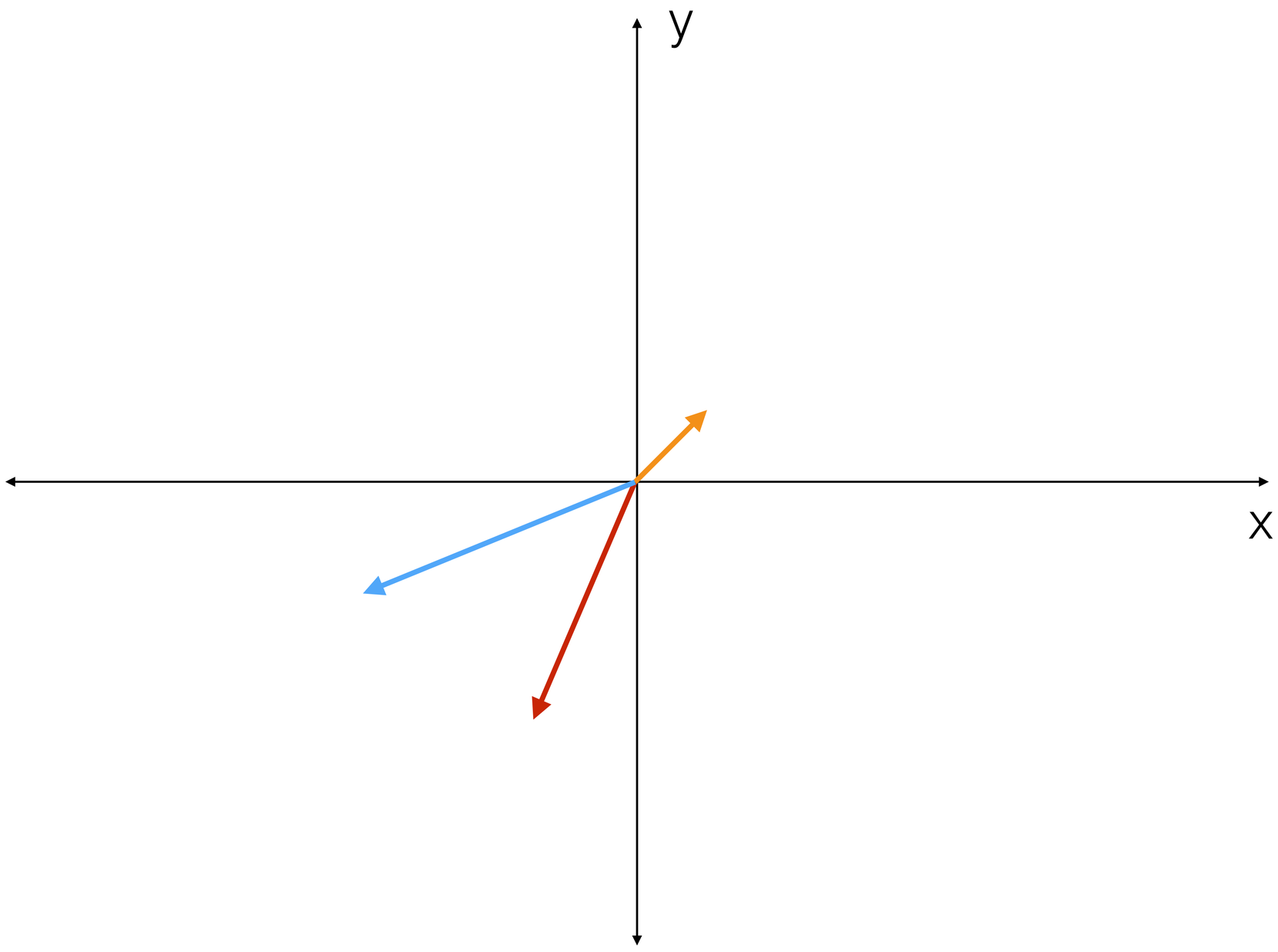

y
x

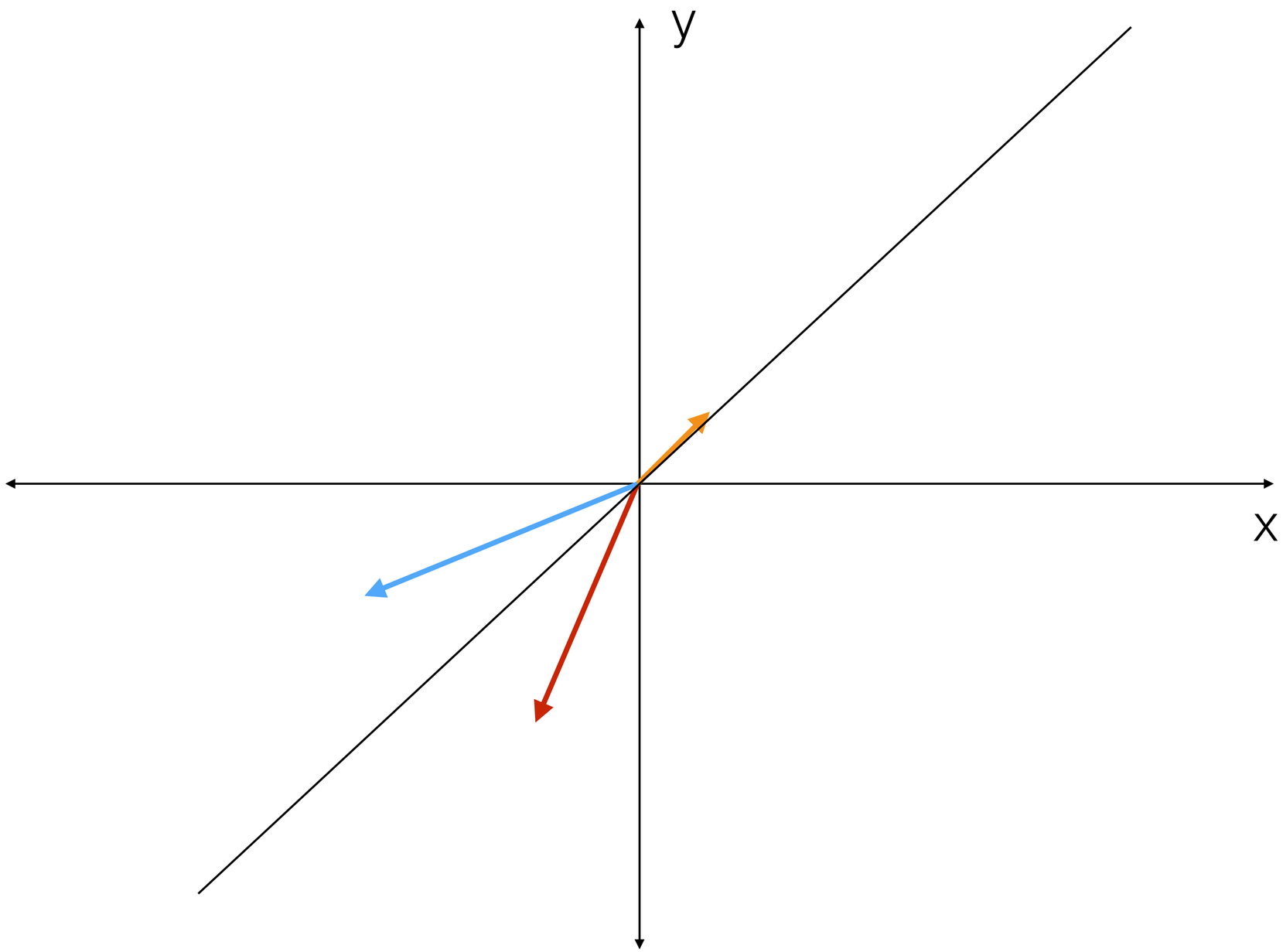

y
x


y
x


y
x


y
x


y
x


y
x

# Character correlations

- Why force contrasts through the origin?
- Because, for each contrast, the direction of subtraction is arbitrary; the signs of all the contrasts could be reversed
- Regression through the origin treats the contrasts as vectors

Part 1:
Continuous
traits

Part 2:
Discrete
traits

# Evolution of discrete characters

*Varanus salvator*

*Bitis gabonica*

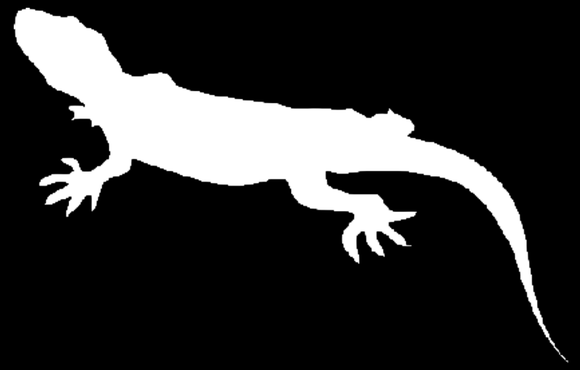
*Gehyra mutilata*

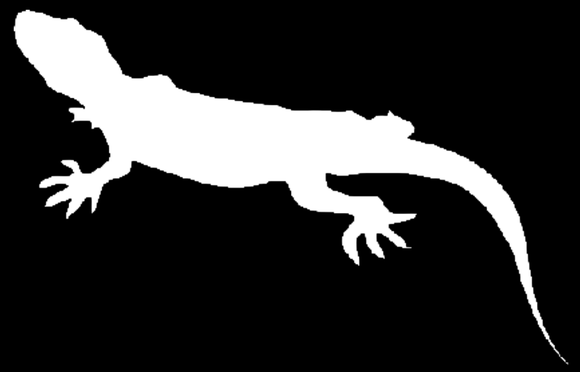
*Lialis burtonis*

*Plestiodon fasciatus*

*Acontias meleagris*

How do we model the process
by which squamates
(snakes and lizards)
lose their limbs?



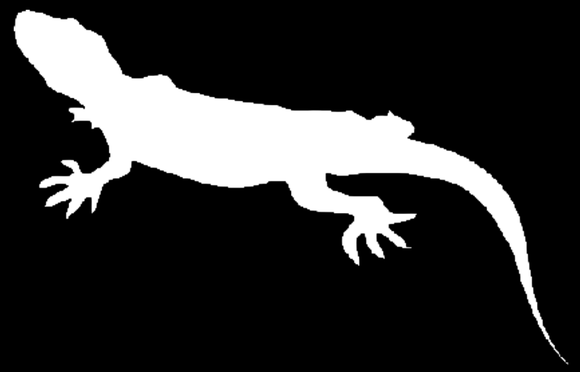


No limbs

State 0

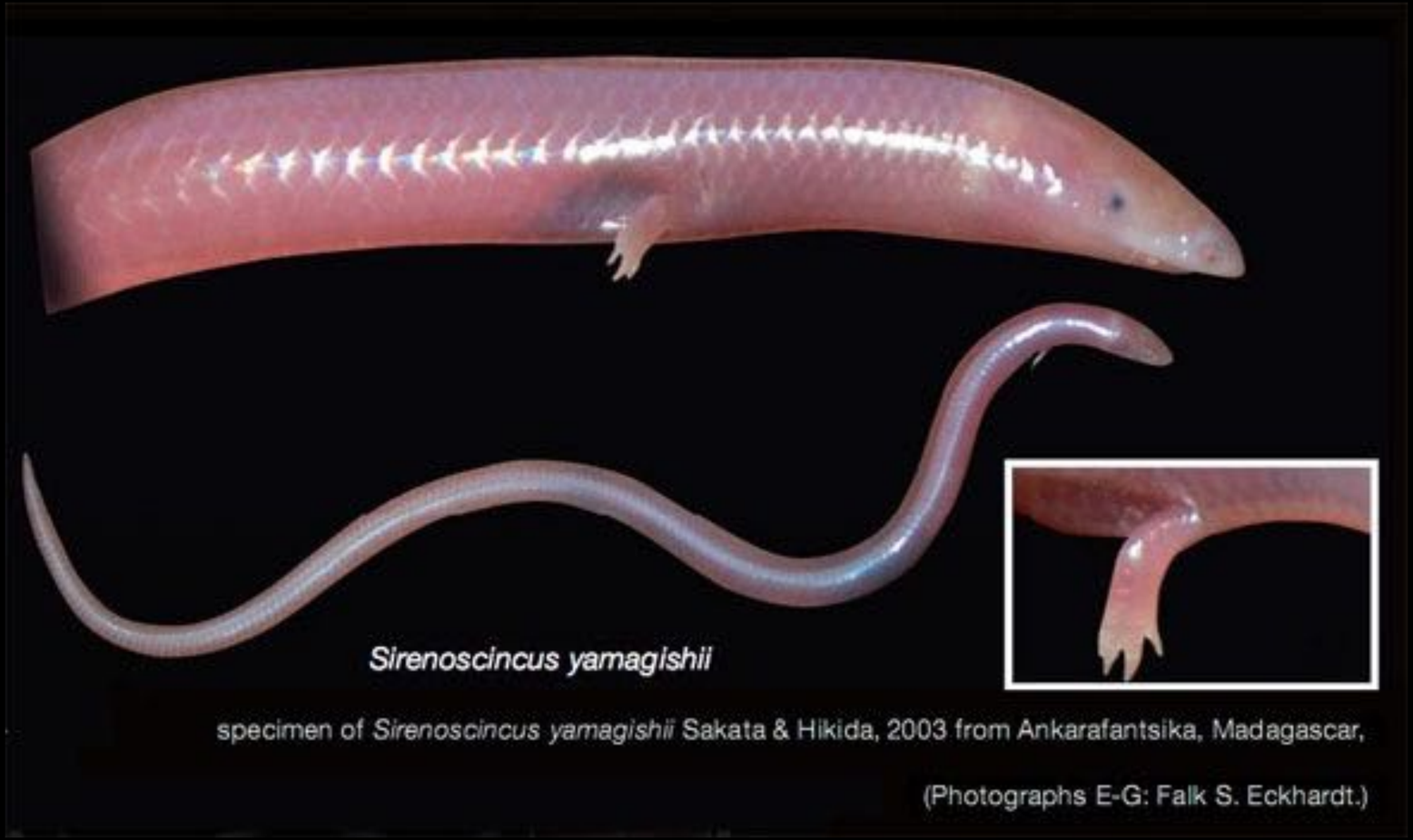
No limbs

State 0



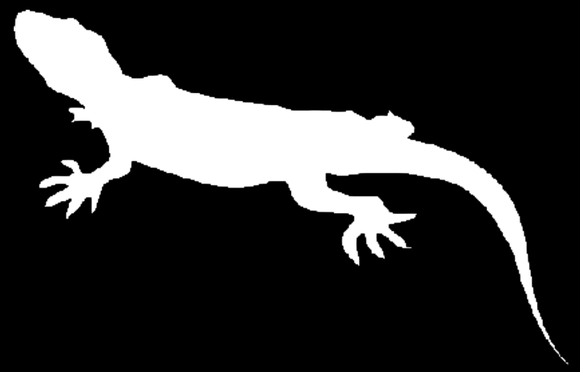
Limbs

State 1


*Sirenoscincus yamagishii*
specimen of *Sirenoscincus yamagishii* Sakata & Hikida, 2003 from Ankarafantsika, Madagascar,
(Photographs E-G: Falk S. Eckhardt.)

No limbs

State 0



Limbs

State 1

# The Mk model

# The Mk model

*“Markov k-state”*

Lewis 2001 Sys Bio

0

1

0

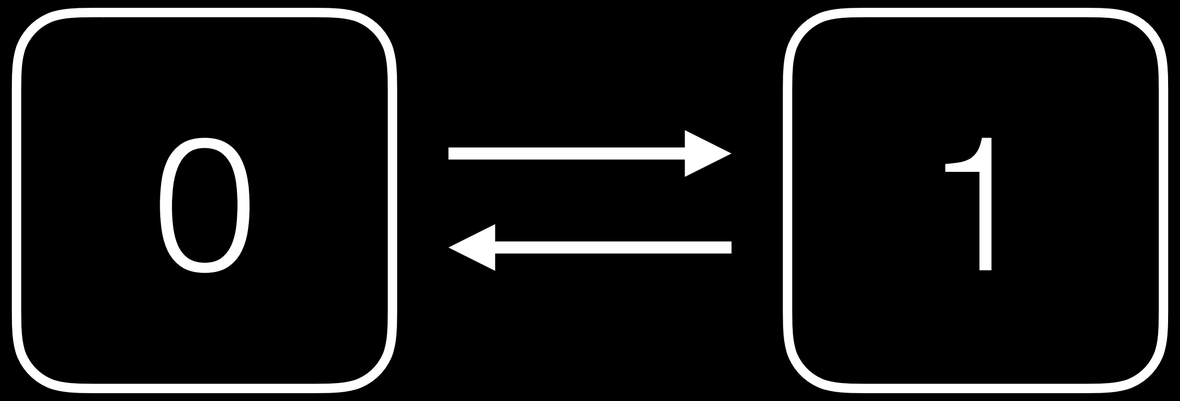
no limbs

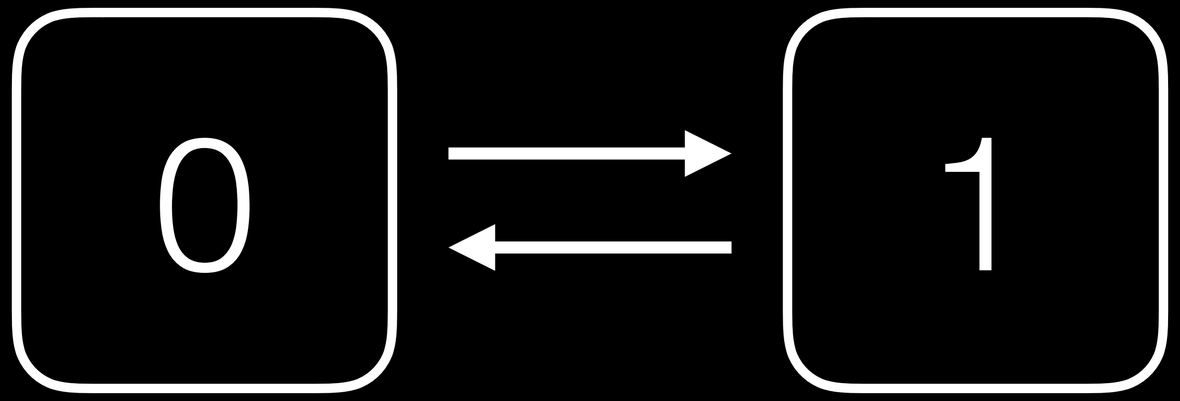
1

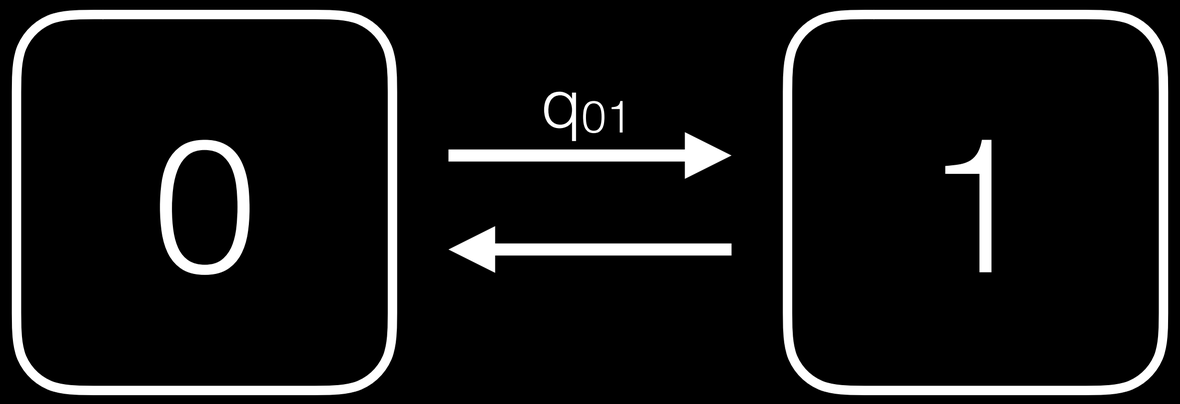
limbs

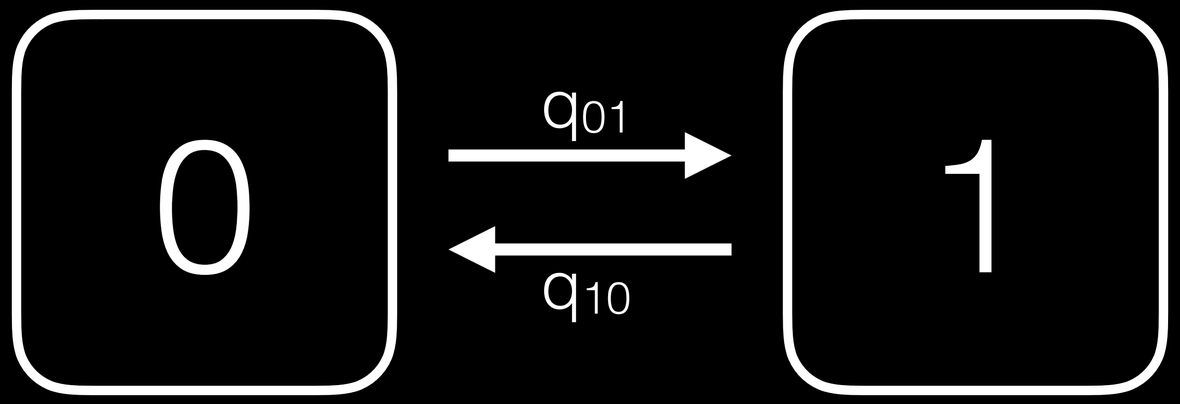
0

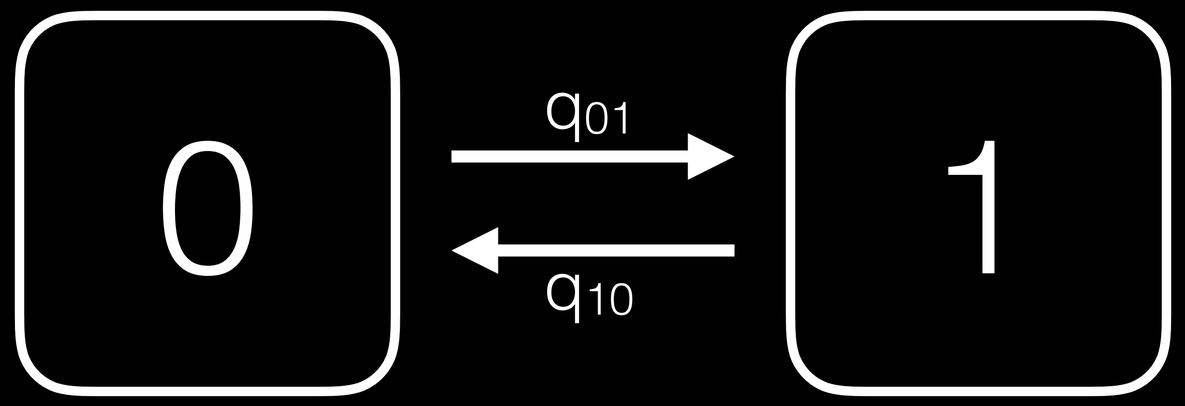
1


0
1


0
1


0
$q_{01}$
1


0
$q_{01}$
$q_{10}$
1




$q_{01}$ = instantaneous rate of change from 0 to 1

$$Q = \begin{bmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{bmatrix}$$

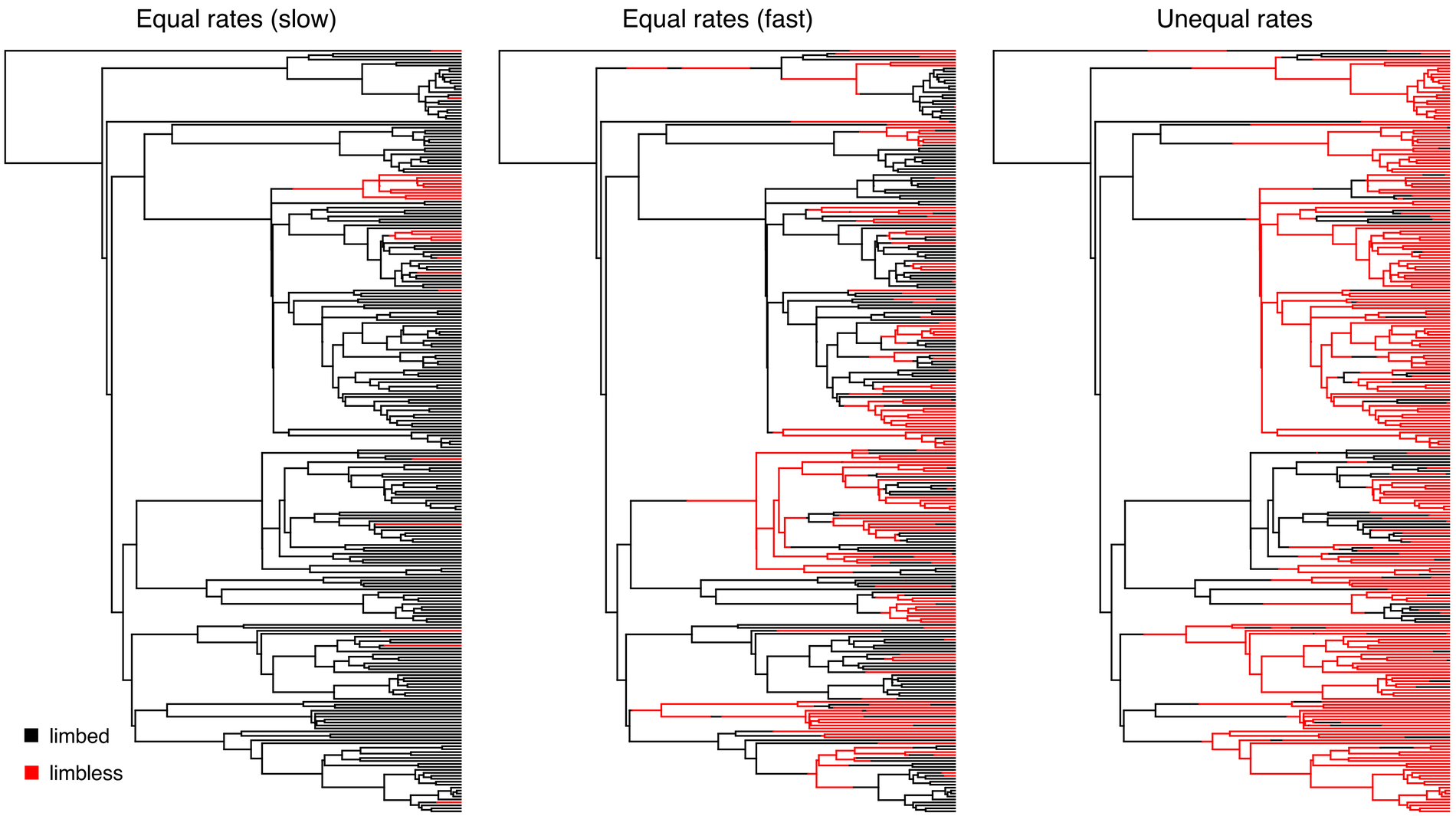
$$Q = \begin{bmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{bmatrix}$$

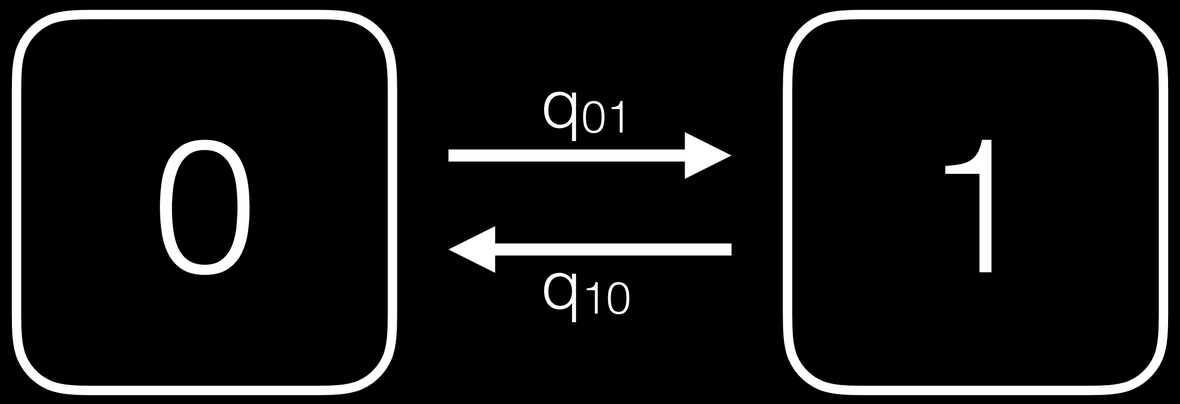
*The instantaneous rate matrix*

$$P(t)=e^{Qt}$$

$$P(t)=\underbrace{e^{Qt}}$$

matrix exponentiation

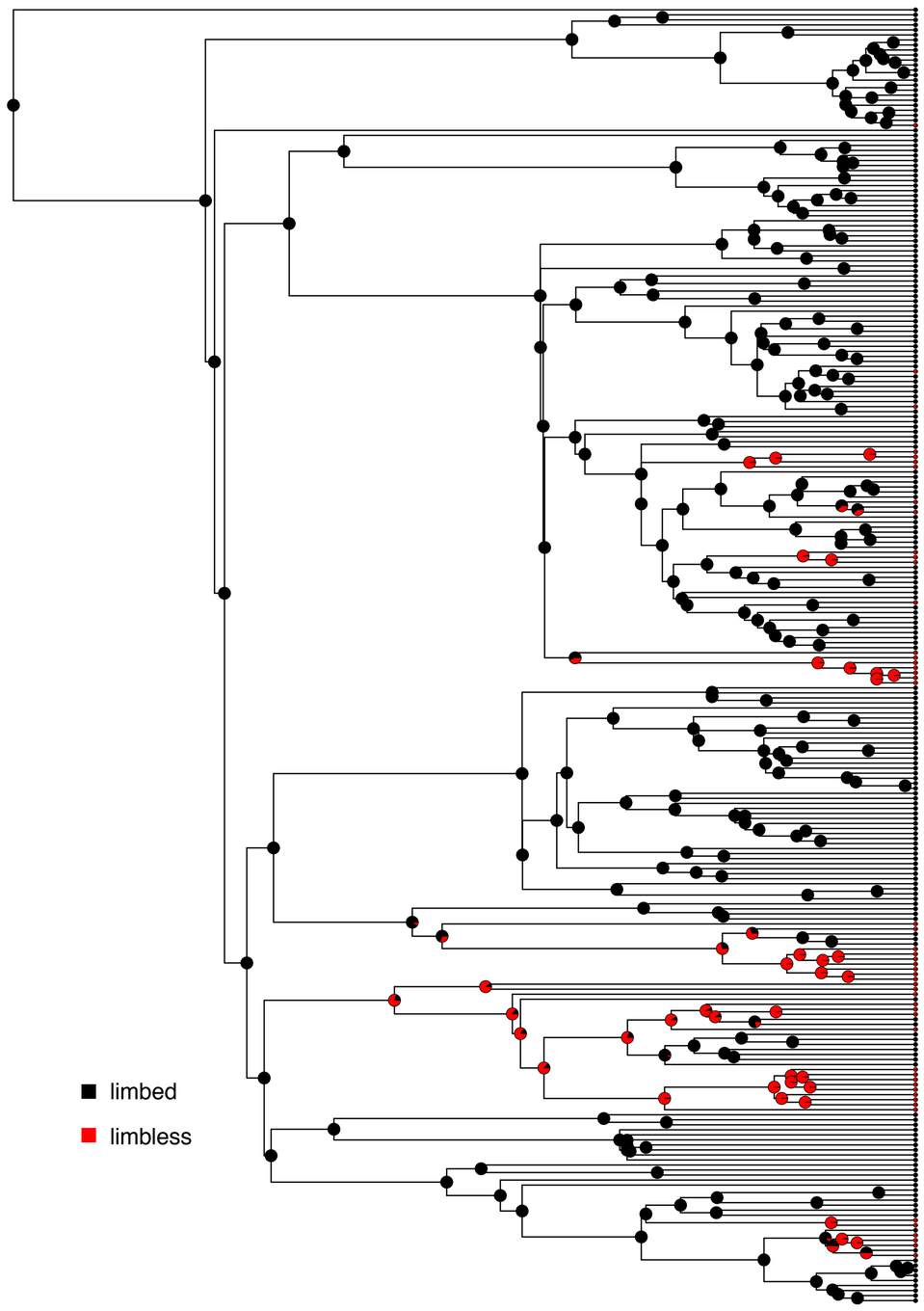

Equal rates (slow)
Equal rates (fast)
Unequal rates
limbed
limbless


0
$q_{01}$
$q_{10}$
1

You must be **explicit** about the probability of each state at the root of the tree!

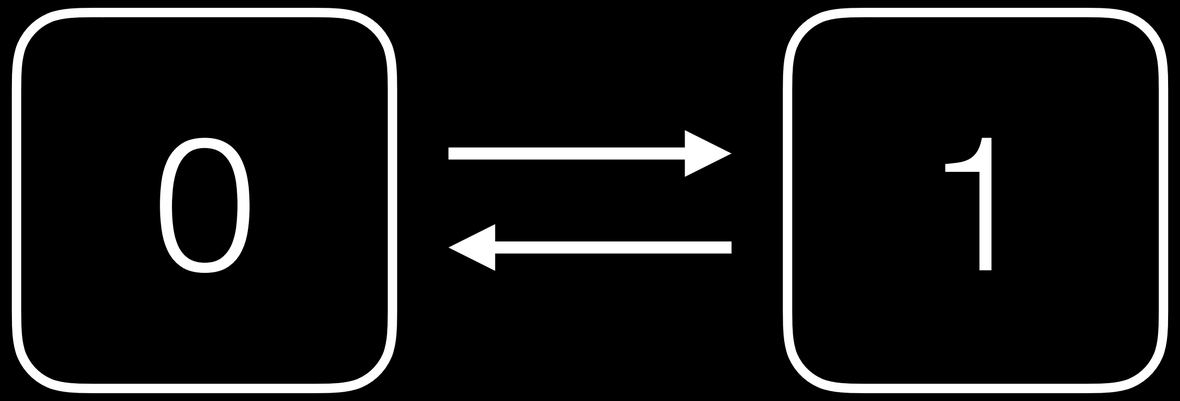
Very common to assume either **equal probability** for all states or that the current state frequencies are in **equilibrium**

# Ancestral state reconstructions

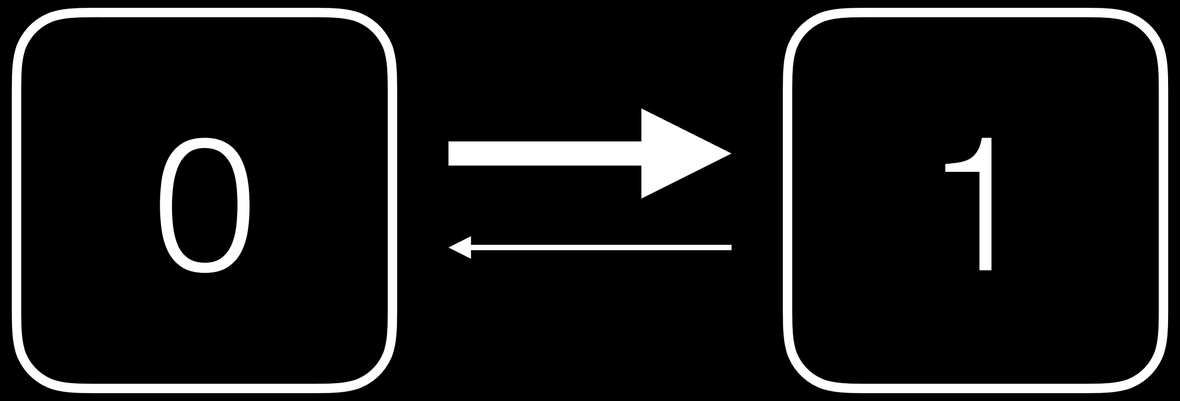

limbed
limbless

Let’s try to fit the model to some example data!

# Equal Rates Model



# Unequal Rates Model



Equal Rates Model

Unequal Rates Model

Equal Rates Model

lnL = -80.5

$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

Unequal Rates Model

Equal Rates Model

lnL = -80.5

$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

Unequal Rates Model

lnL = -79.4

$$Q_{ARD} = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0038 & -0.0038 \end{bmatrix}$$

Equal Rates Model

lnL = -80.5

$$Q_{ER} = \begin{bmatrix} -0.0019 & 0.0019 \\ 0.0019 & -0.0019 \end{bmatrix}$$

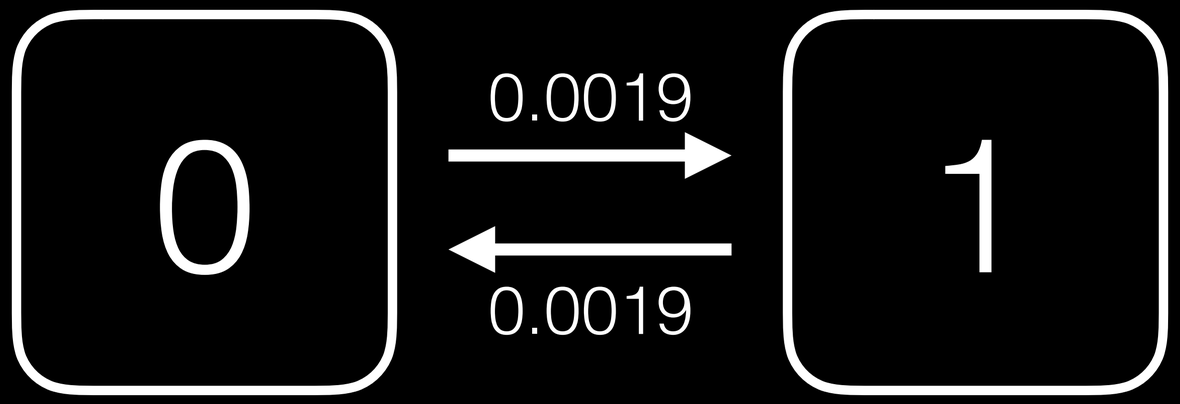
Unequal Rates Model

lnL = -79.4

$$Q_{ARD} = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0038 & -0.0038 \end{bmatrix}$$

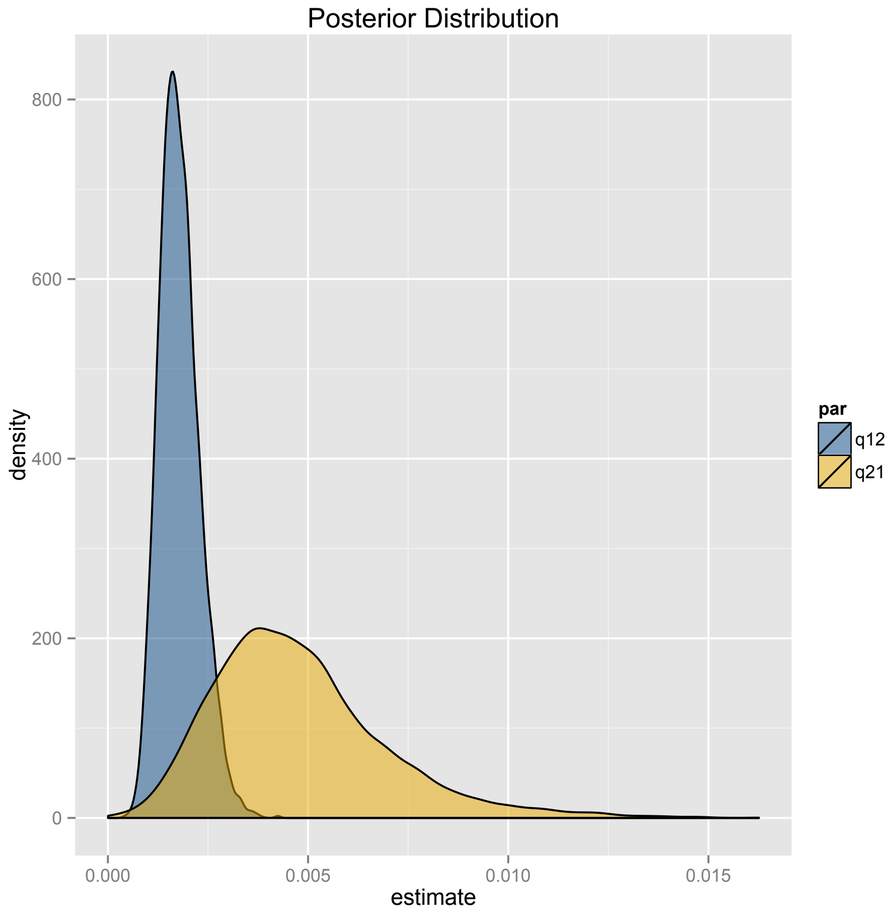
Conclusion: cannot reject ER model

# Equal Rates Model




We have to wait about 500 my on average to see a transition

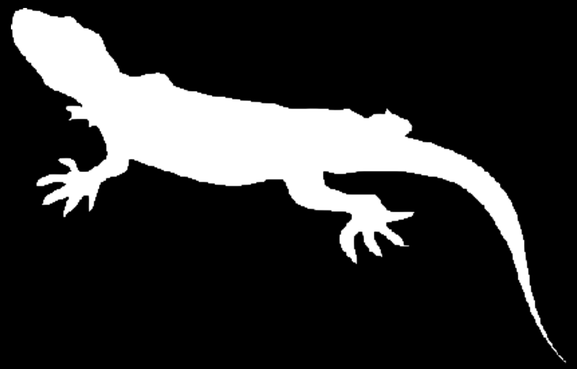
# We can also do this with Bayesian analyses

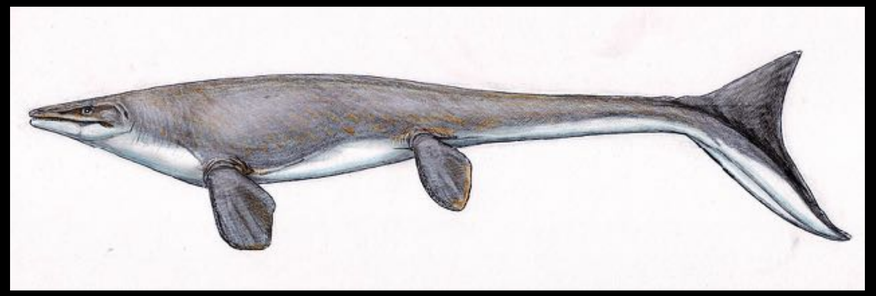


We can also fit a model with more character states

Front limbs only

State 1



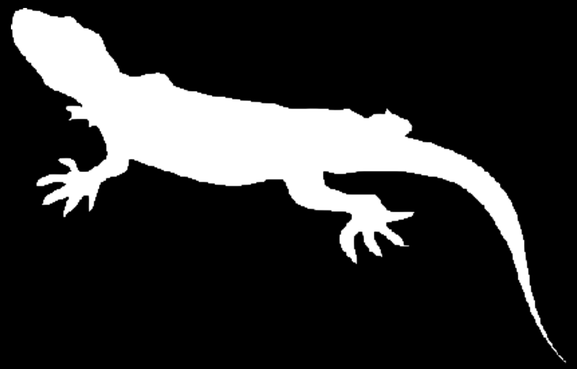


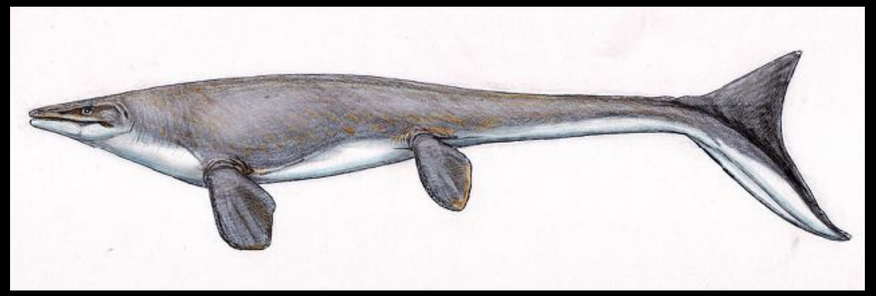
No limbs

State 0

Front limbs only

State 1



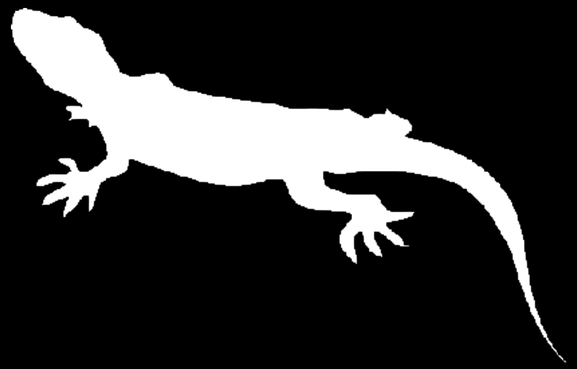


No limbs

State 0

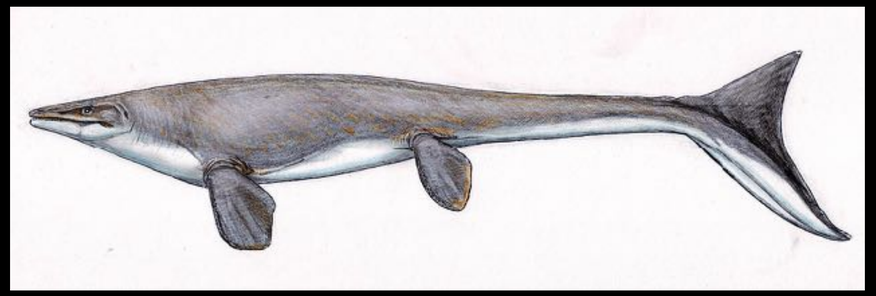
Front limbs only

State 1



Limbs

State 2

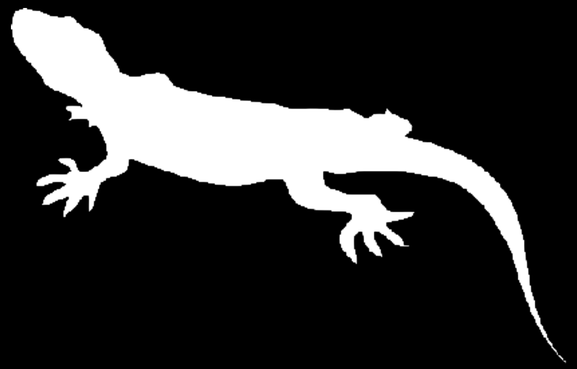


No limbs

State 0

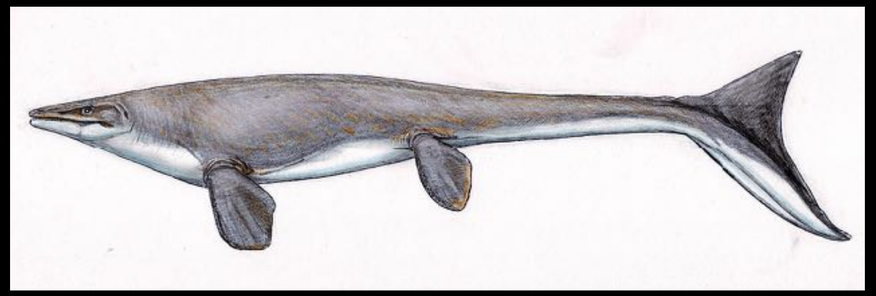
Front limbs only

State 1



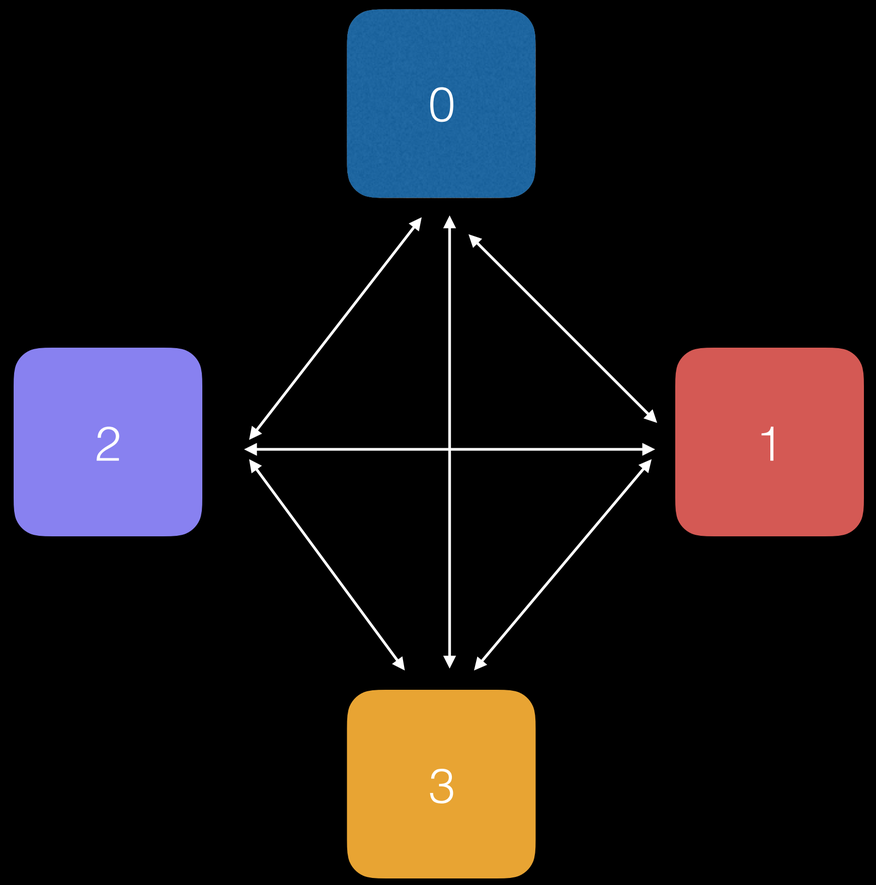
Limbs

State 2



“Fins”

State 3


0
2
1
3

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

— is the number that makes the row sums 0

e.g. $-(q_{01} + q_{02} + q_{03})$

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

ER = equal rates

1 rate parameter

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

SYM = symmetric

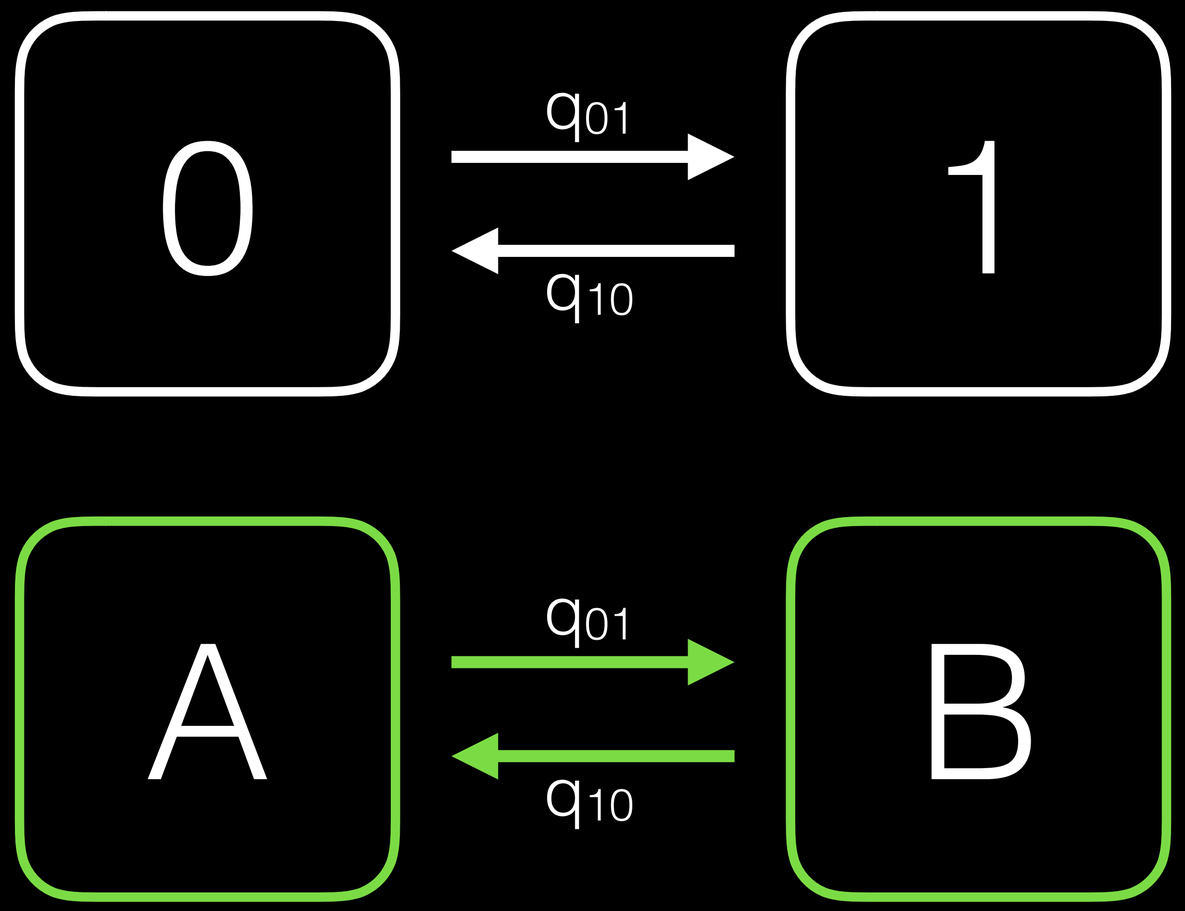
6 rate parameters

$$Q = \begin{bmatrix} — & q_{01} & q_{02} & q_{03} \\ q_{10} & — & q_{12} & q_{13} \\ q_{20} & q_{21} & — & q_{23} \\ q_{30} & q_{31} & q_{32} & — \end{bmatrix}$$

ARD = all rates different

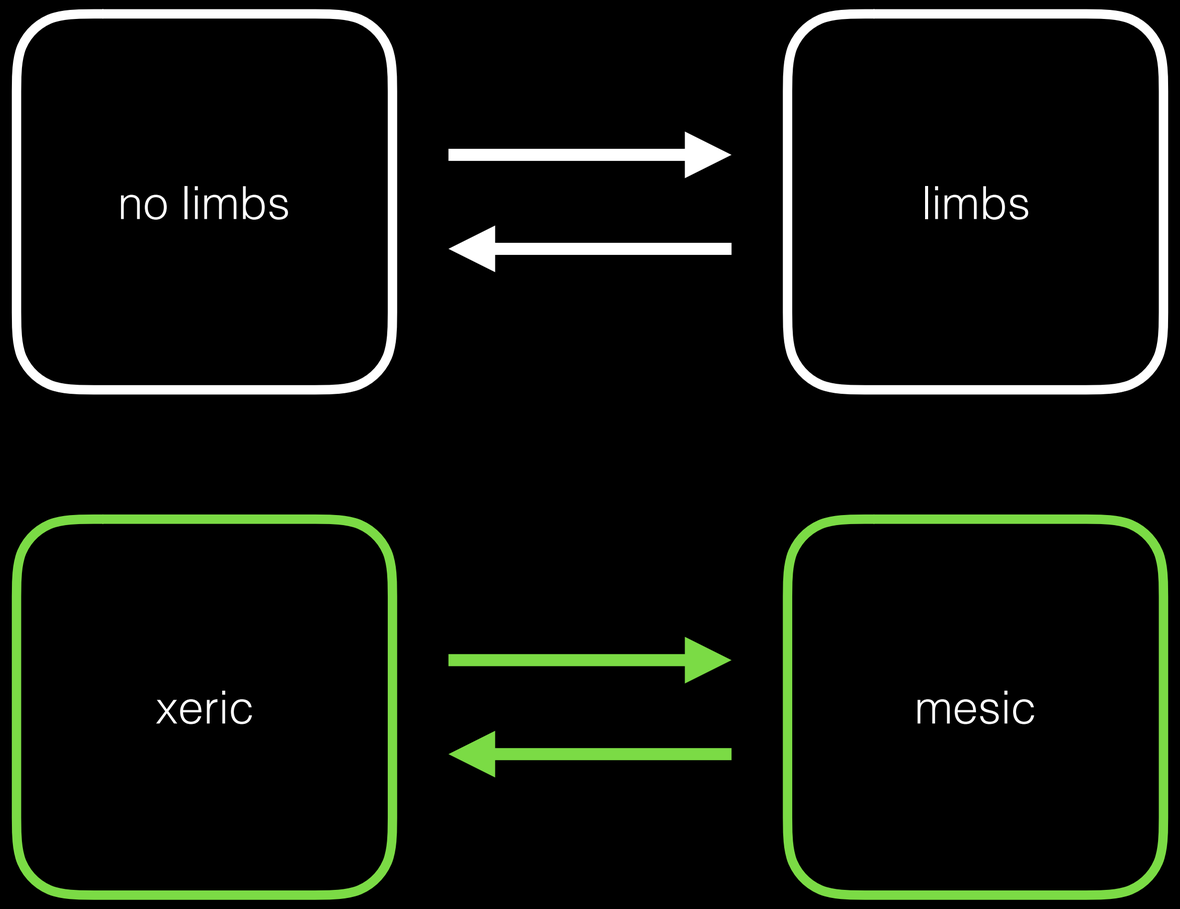
12 rate parameters

We can use the Mk model to test a wide variety of evolutionary hypotheses about discrete characters




Correlated or not?

*Lialis burtonis*



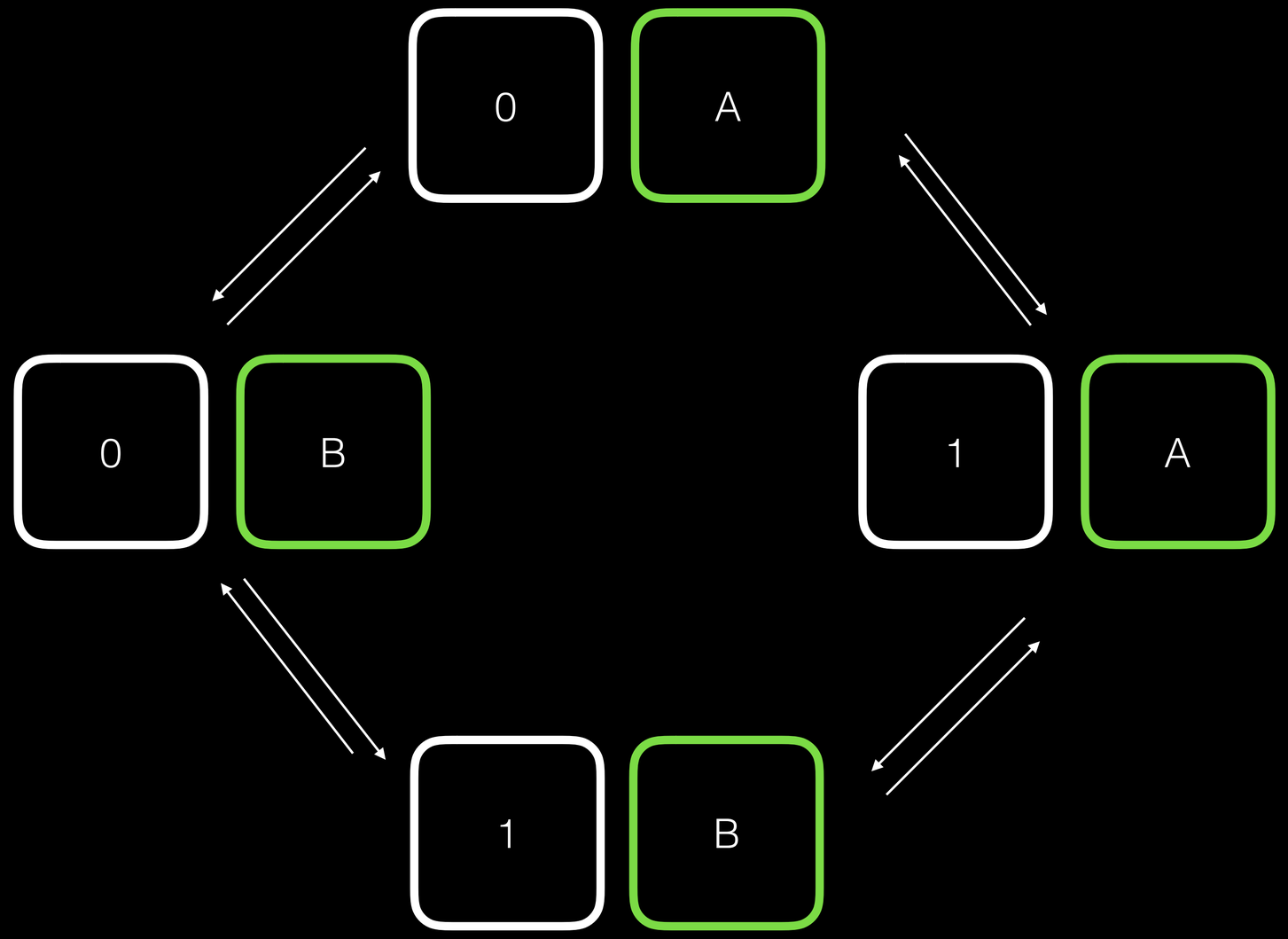

Correlated or not?

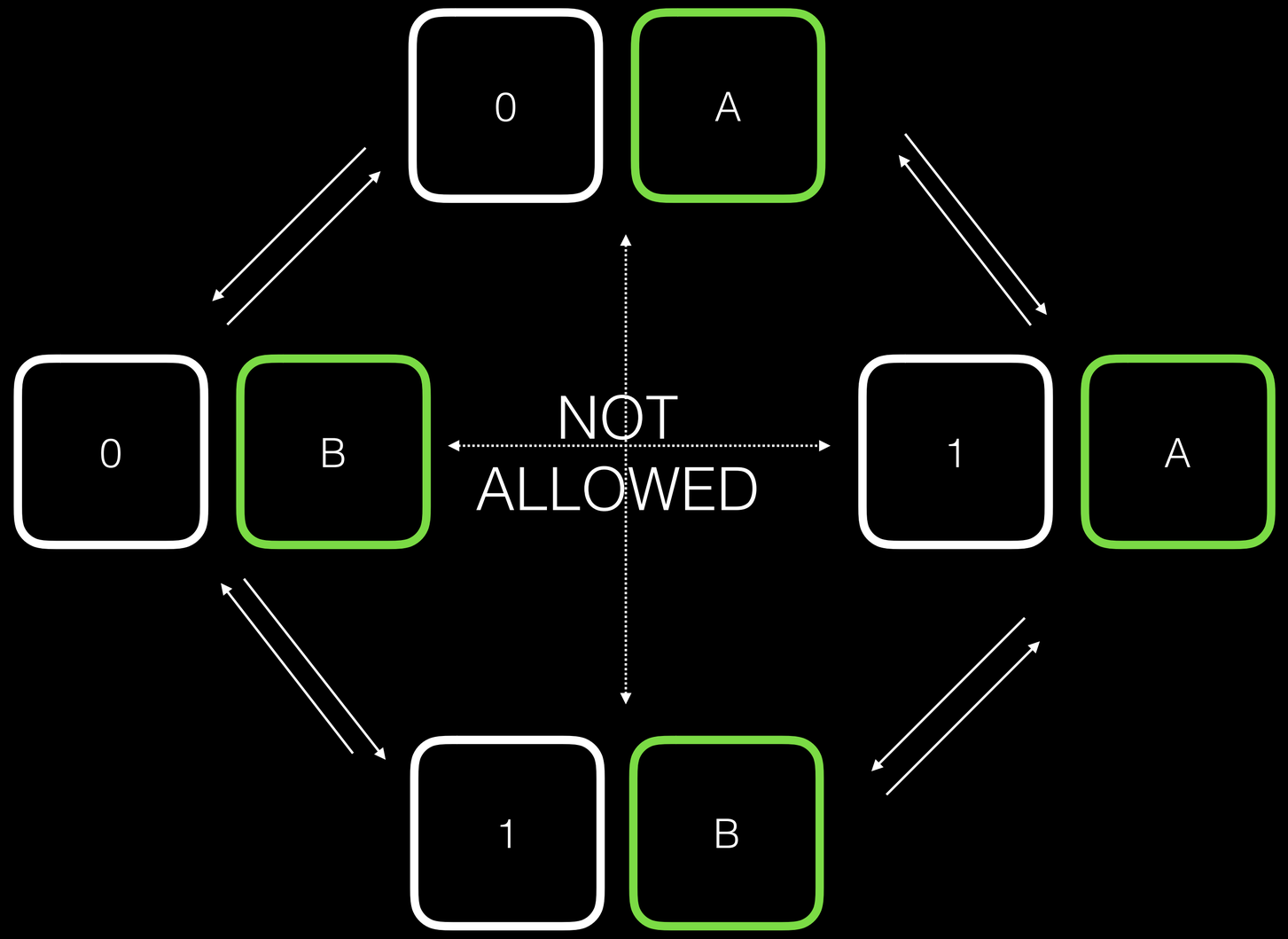
0

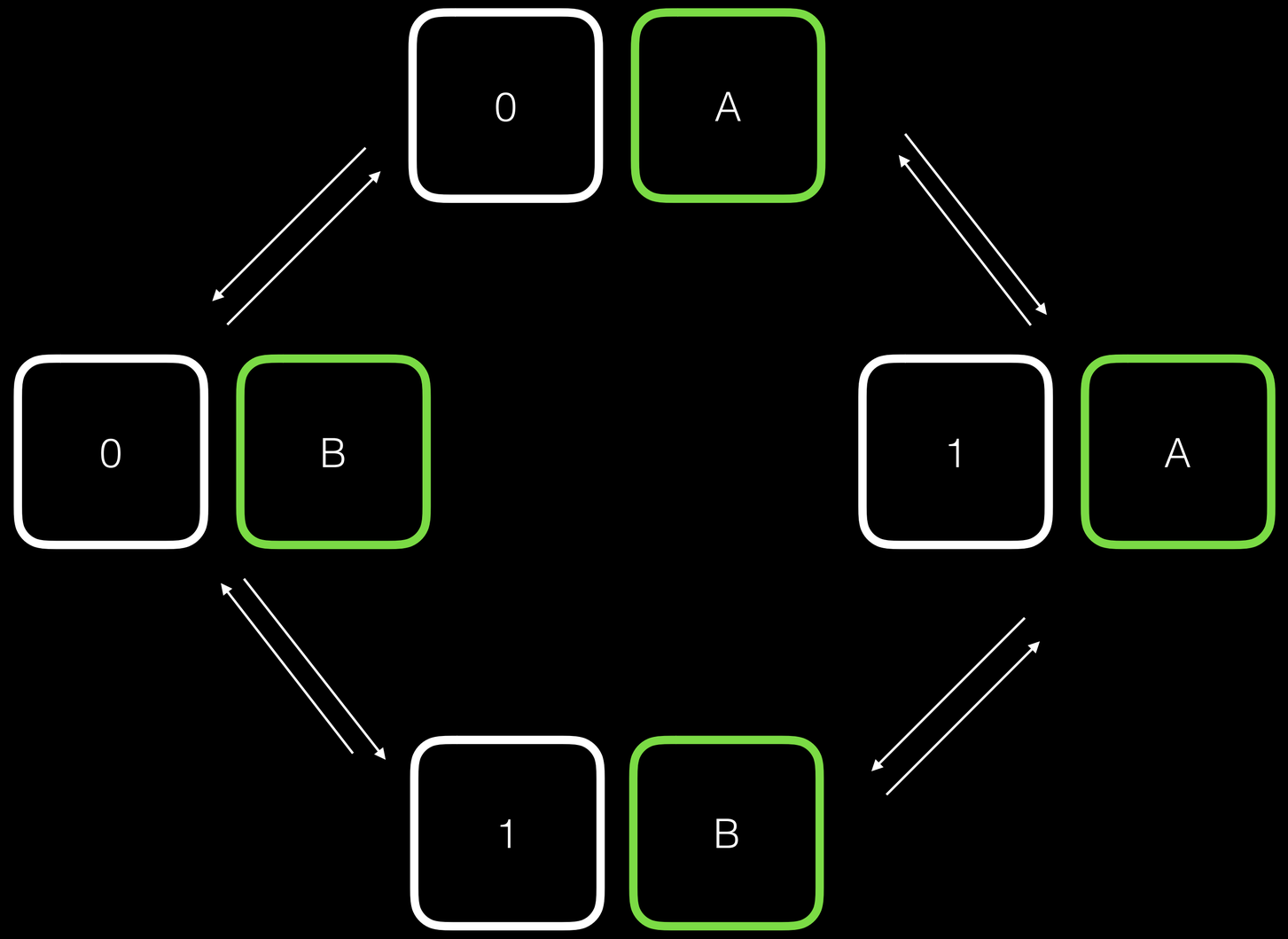
1

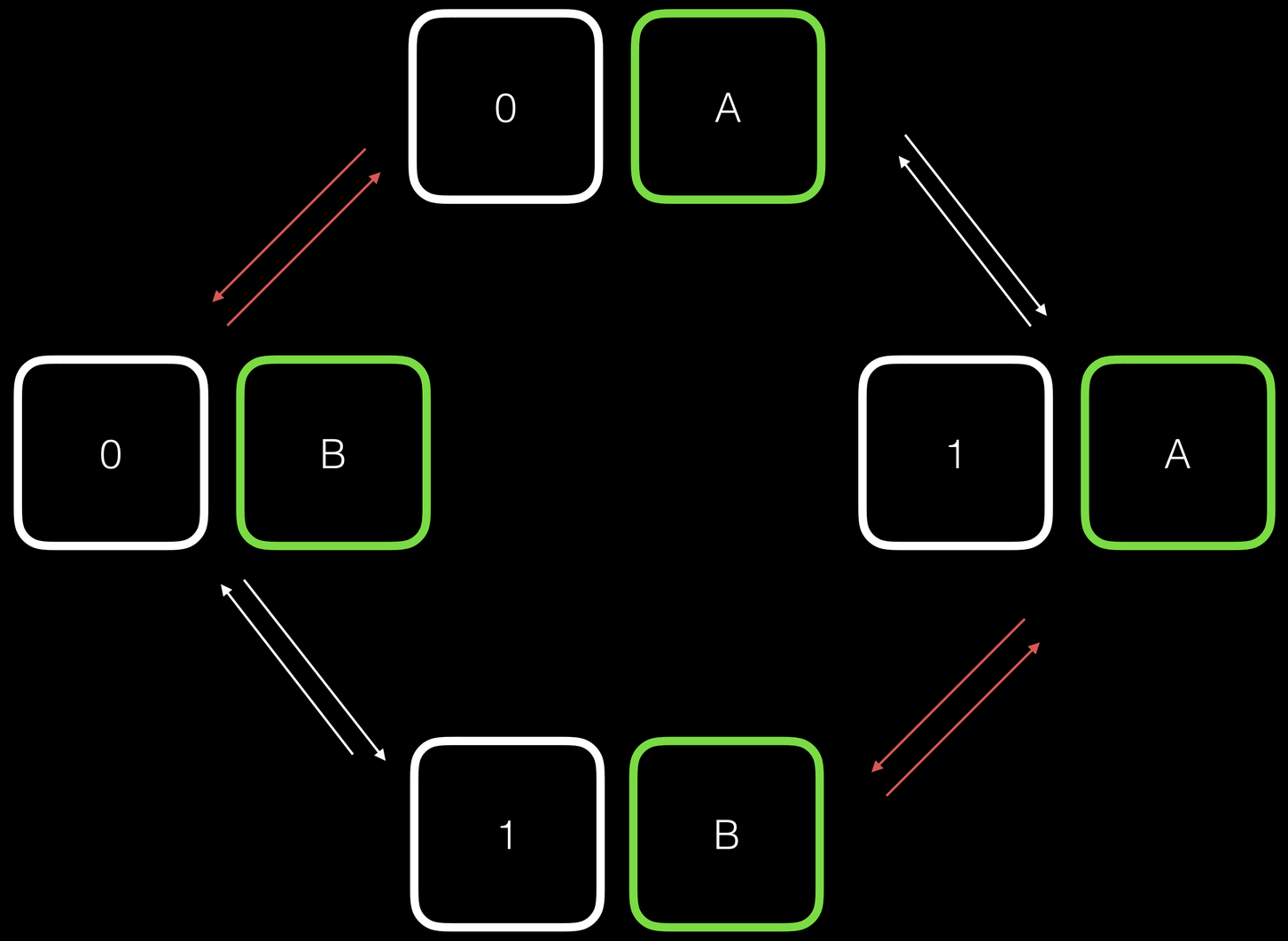
A

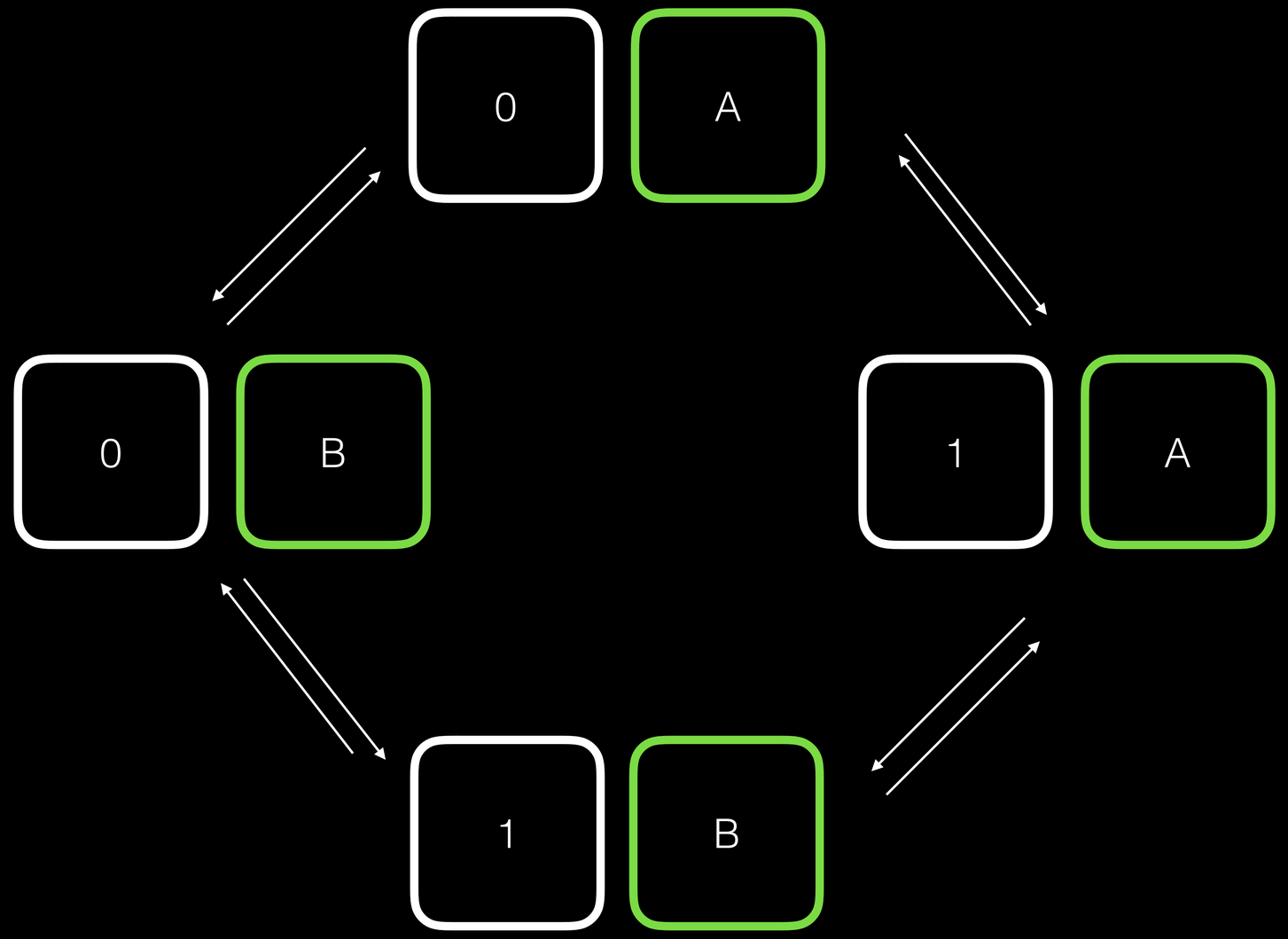
B

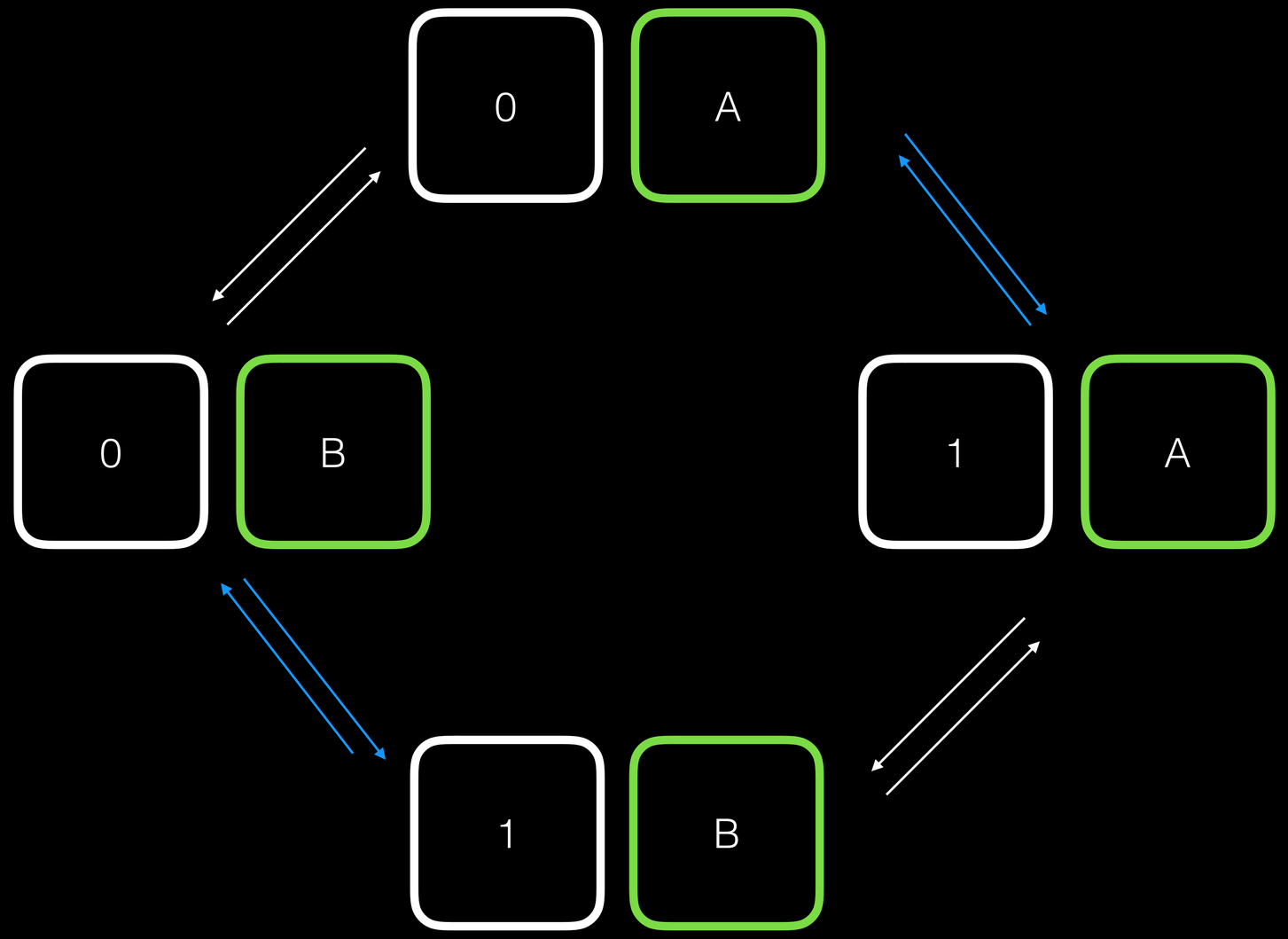

0
A
0
B
1
A
1
B

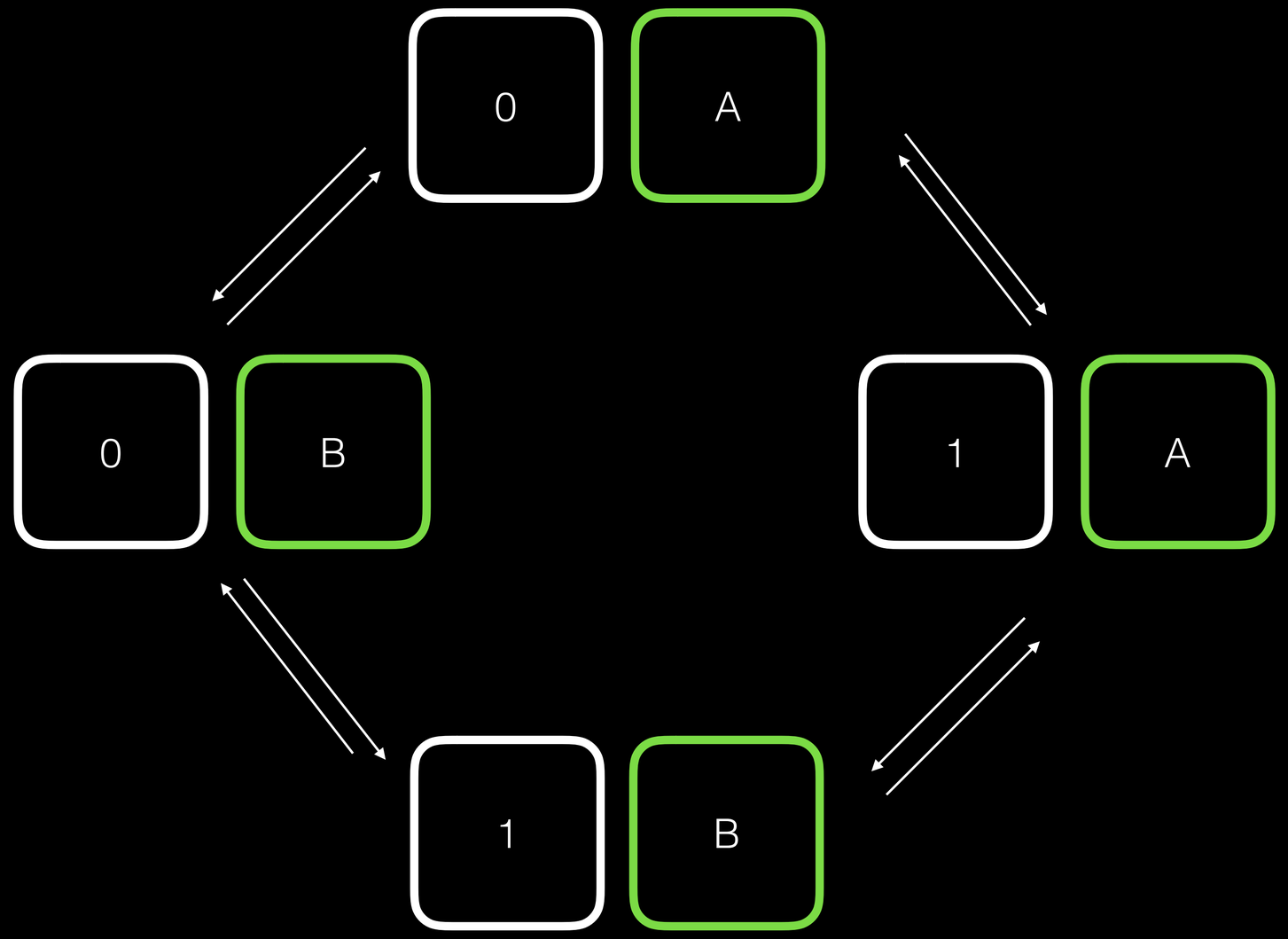

0
A
0
B
NOT
ALLOWED
1
A
1
B

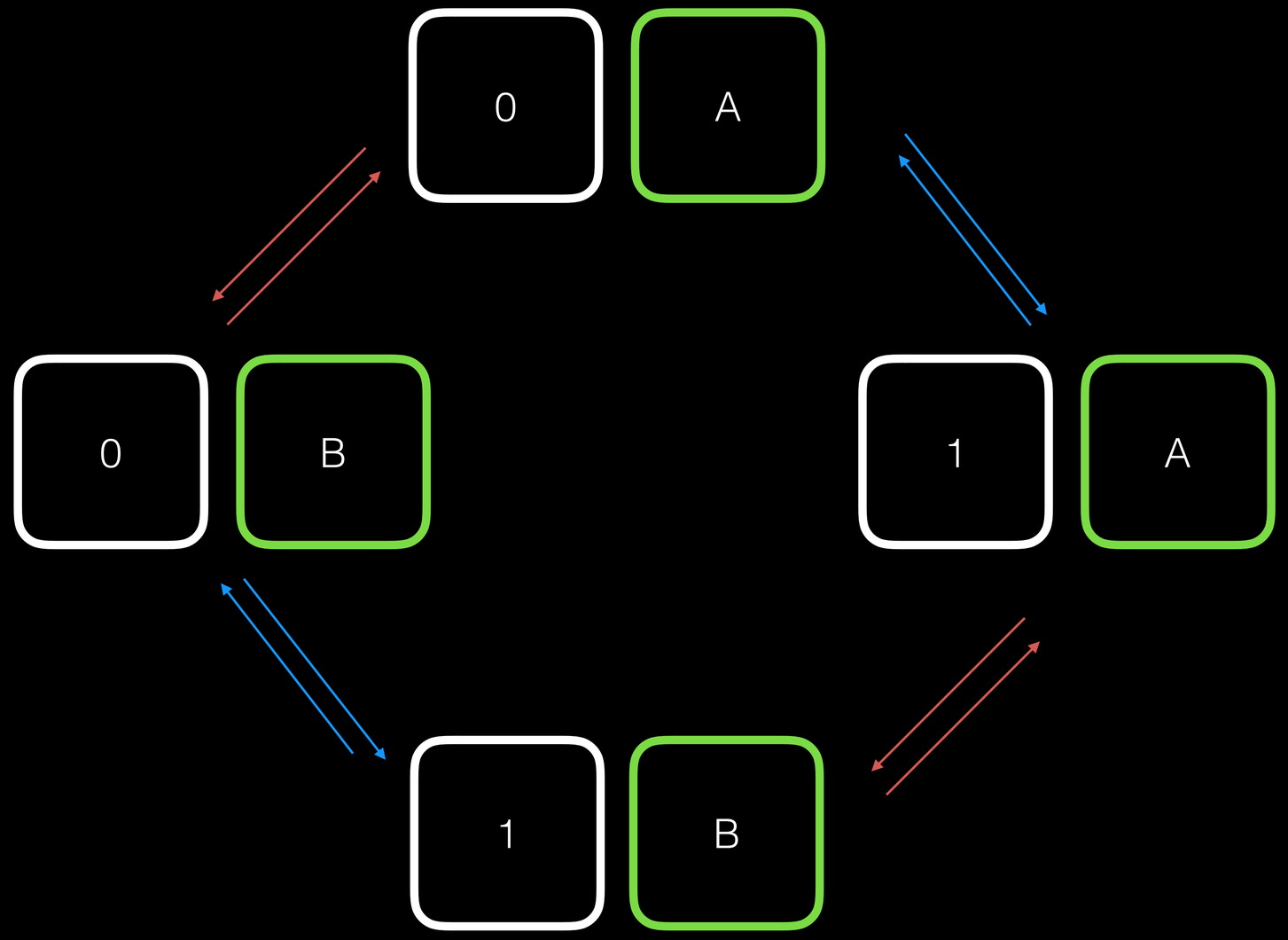

0
A
0
B
1
A
1
B

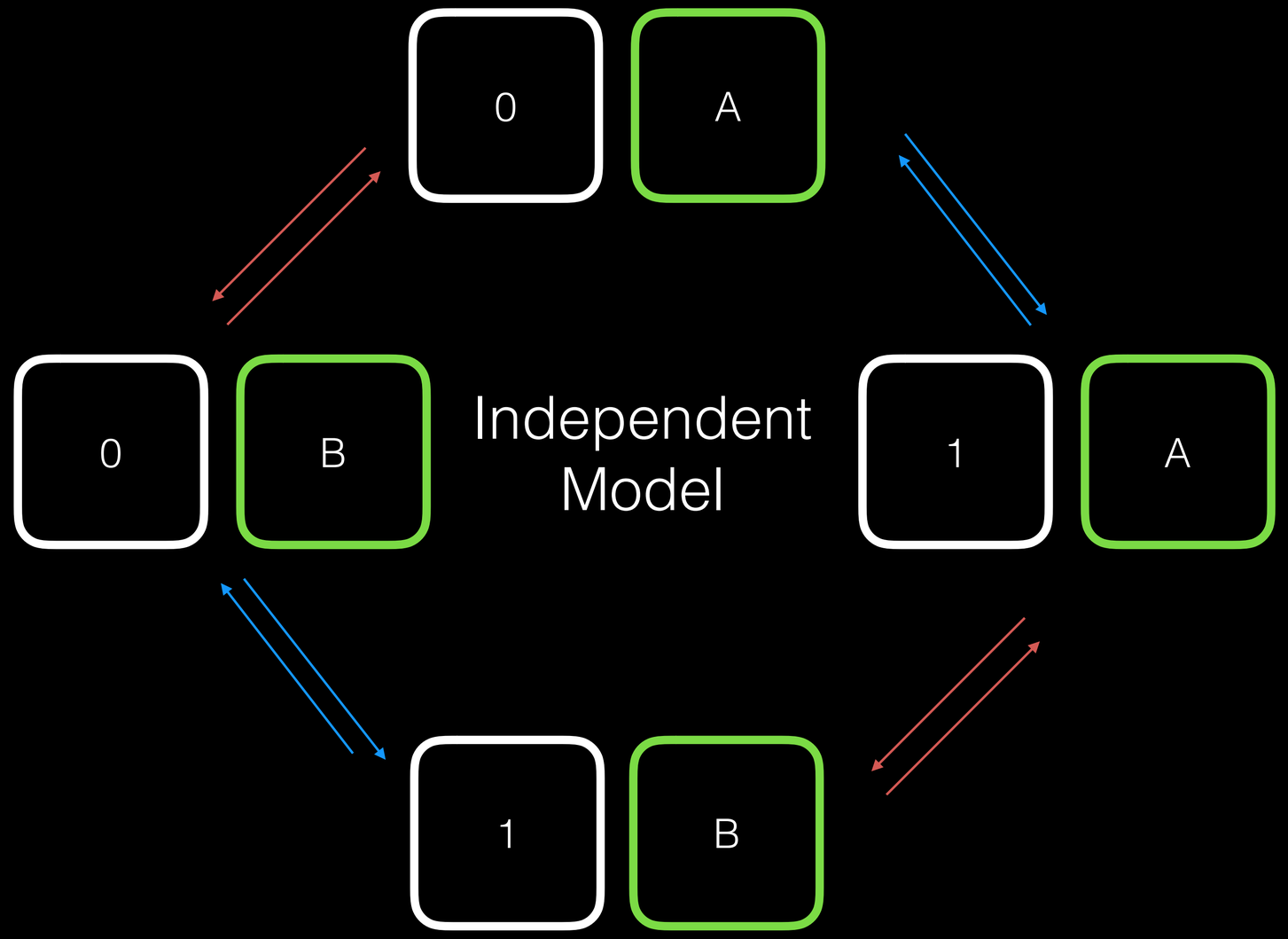

0
A
0
B
1
A
1
B

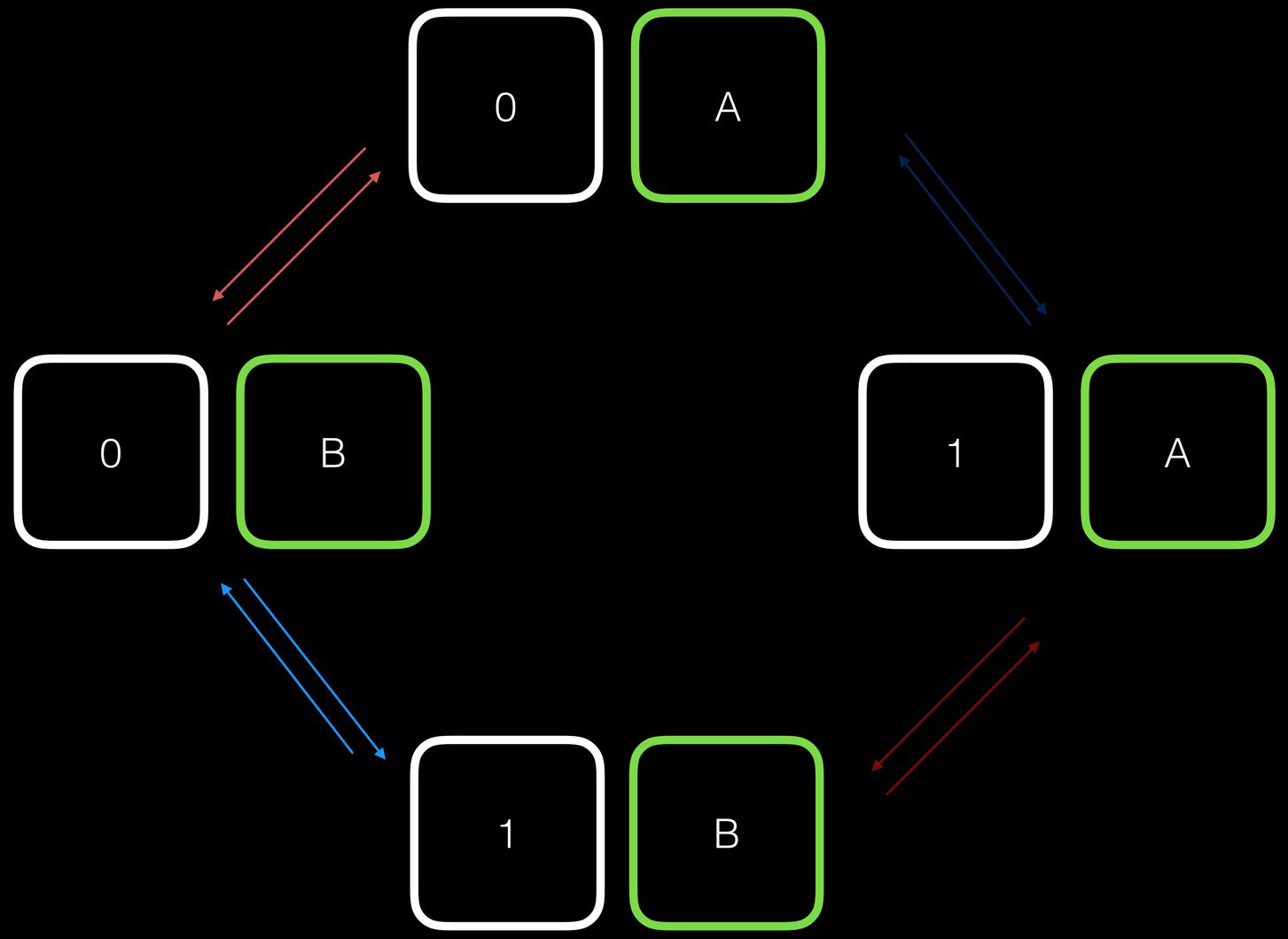

0
A
0
B
1
A
1
B

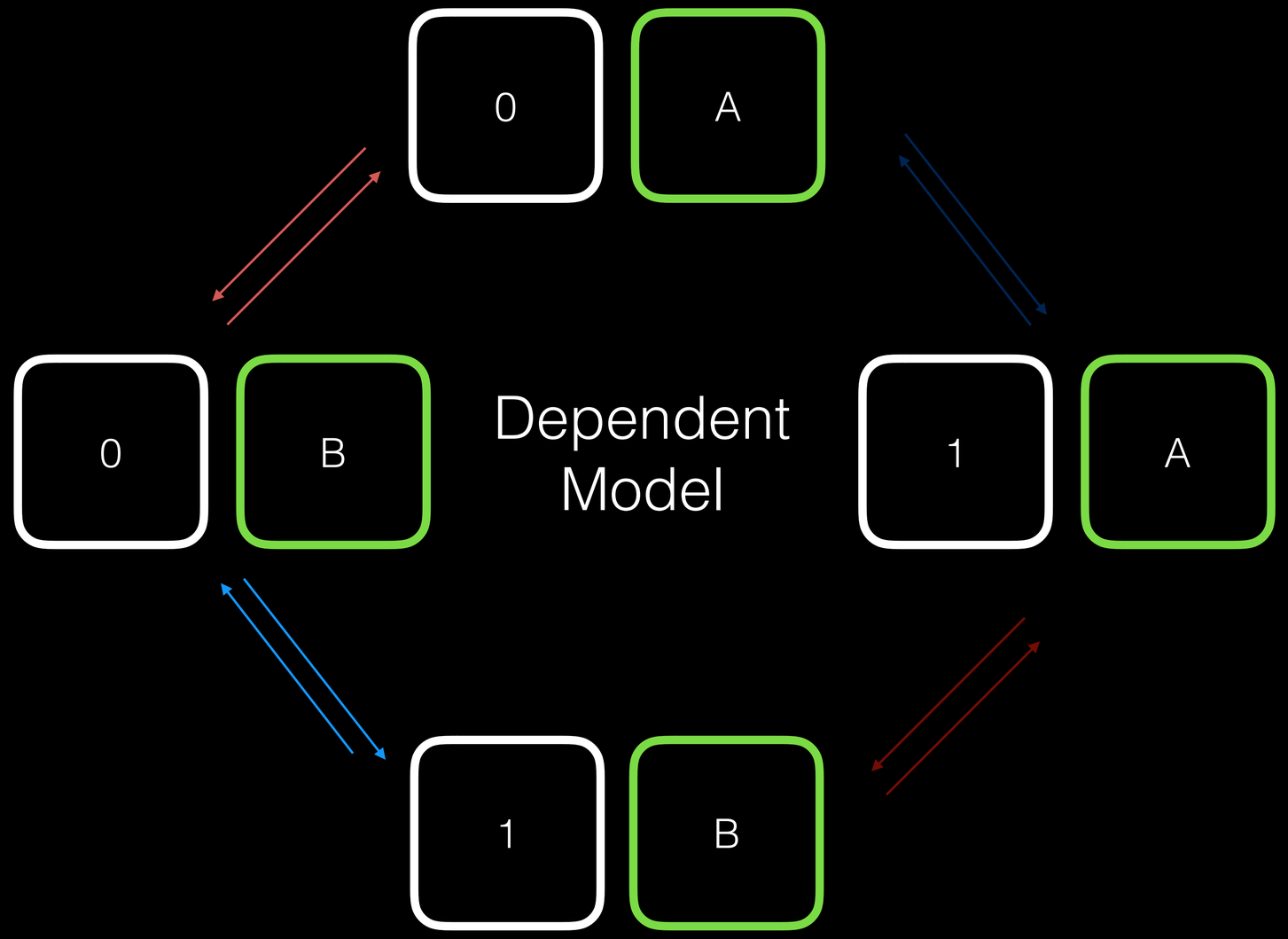

0
A
0
B
1
A
1
B


0
A
0
B
1
A
1
B


0
A
0
B
1
A
1
B


0
A
0
B
Independent
Model
1
A
1
B


0
A
0
B
1
A
1
B


0
A
0
B
Dependent
Model
1
A
1
B

Independent Model

vs.

Dependent Model

Independent Model

vs.

Dependent Model

Likelihood ratio test,
AICc, etc.

Part 1:
Continuous
traits

Part 2:
Discrete
traits