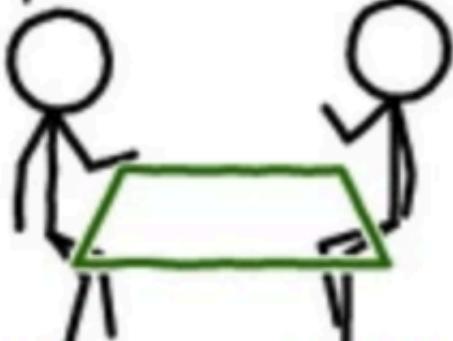


Bayesian humor

BAYESIAN VERSUS FREQUENTIST

I DON'T KNOW WHY WE
STILL HAVE THIS
"DEBATE" EACH YEAR.

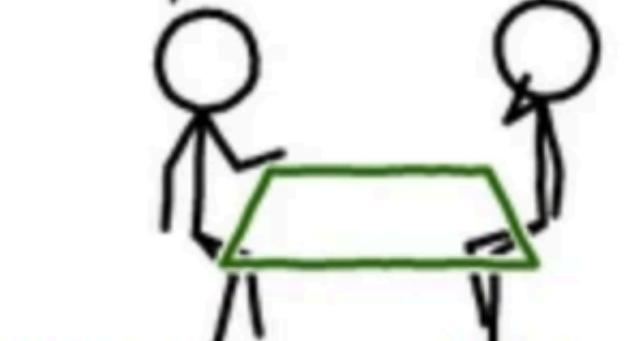
WANT TO TOSS
FOR IT?
I KNOW
ITS A BIT PASSÉ
SURE



MAYBE IT IS
MAYBE IT ISN'T

WAIT A MINUTE -
IS THAT YOUR
BIASED COIN?

HMM TRICKY



AHA! PRIOR
CONSTRUCTION!
BAYES WINS AGAIN!

DAMN

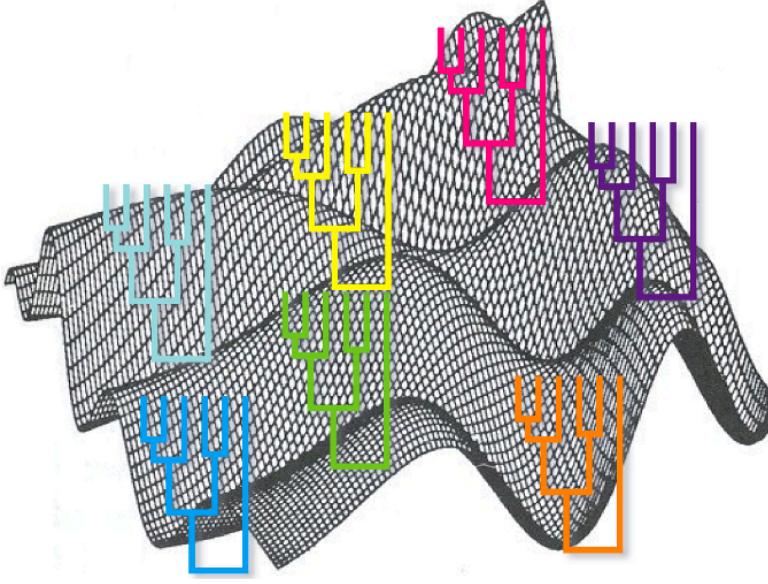
BAYESIAN

FREQUENTIST

BAYESIAN

FREQUENTIST

REVBAYES #005



Bayesian approaches

November 12th, 2019

Can we apply prior knowledge to inform our analyses?

Outline of lecture

- What is Bayesian?
- Basic introduction to Bayesian phylogenetics
- Using molecular data to date a phylogeny

ML vs Bayesian

- Maximum likelihood
 - Search for the tree that maximizes the chance of seeing the data
 $(P(\text{Data} \mid \text{Tree}))$
 - Data are random variables but parameters are fixed
- Bayesian
 - Search for the tree that maximizes the chance of seeing the tree given the data $(P(\text{Tree} \mid \text{Data}))$
 - Data are random variables and so are the model parameters
- If priors are well-behaved, then both ML and Bayesian inference converge on the same value (i.e. tree topology)

Expected outcomes

- For any ML analysis, we can expect a tree with the highest likelihood score
- For Bayesian any given run (sampler) may work well in most cases, all runs will fail in some cases, and not guaranteed to work for any particular case
- When do we know the run provides an accurate approximation for a given analysis?
 - We Never do

Bayes Theorem

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

Posterior probability Likelihood Prior probability of A

Prior of B, normalizing constant

The diagram illustrates the components of Bayes' Theorem. At the top, three labels are positioned above the formula: 'Posterior probability' points to $P(A|B)$, 'Likelihood' points to $P(B|A)$, and 'Prior probability of A' points to $P(A)$. Below the formula, an arrow points upwards to the term $P(B)$, which is labeled 'Prior of B, normalizing constant'.

Bayes theorem rewritten

$$P(\text{tree} | \text{data}) = \frac{P(\text{data} | \text{tree}) P(\text{tree})}{P(\text{data})}$$

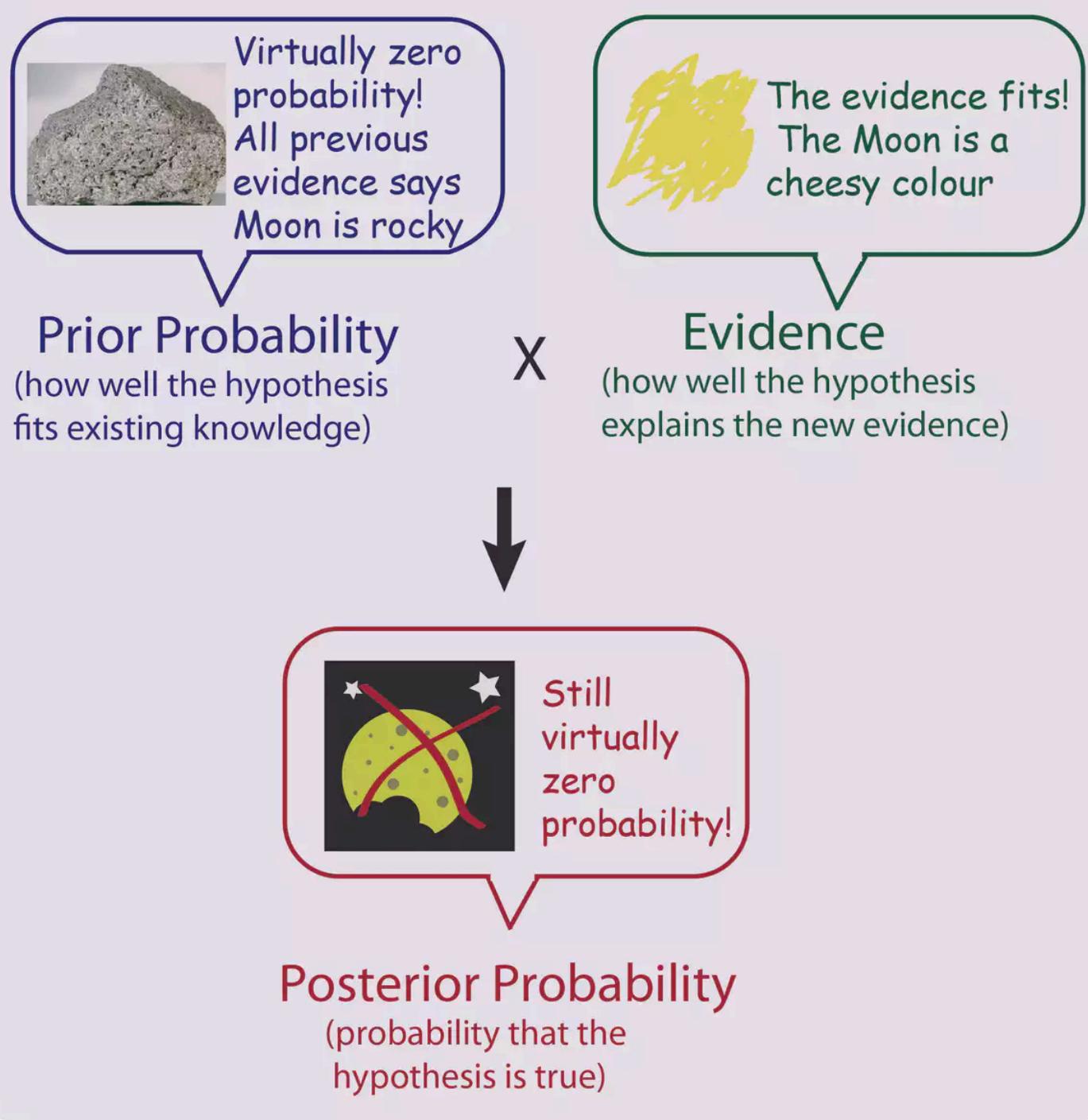
- $P(\text{Data})$ generally not computable
- Parameter space includes
 - Tree topology
 - Branch lengths
 - Substitution model parameters

Example



Hypothesis: The moon is made of cheese

New Evidence: I see the moon is yellow



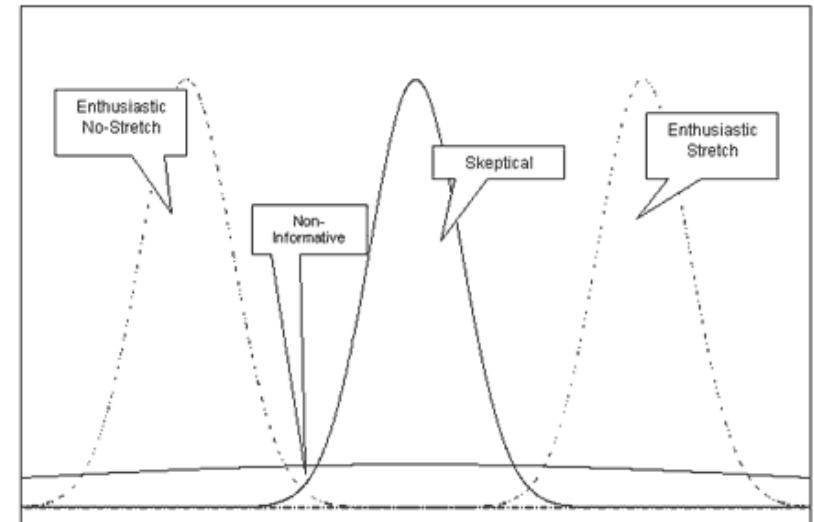
Bayesian inference

<http://theconversation.com/bayes-theorem-the-maths-tool-we-probably-use-every-day-but-what-is-it-76140>

Benefits of stretching before exercise?

Priors

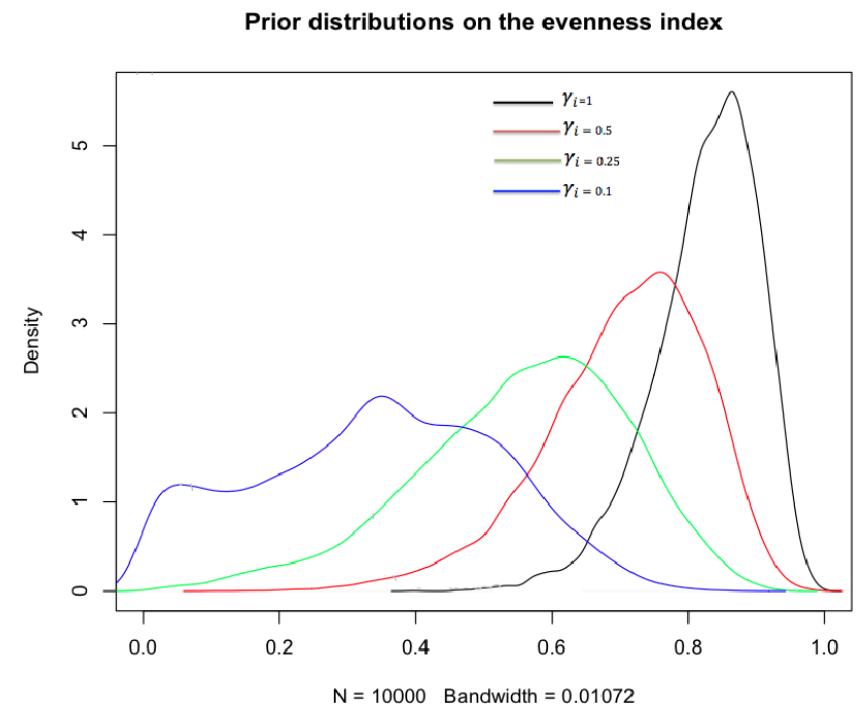
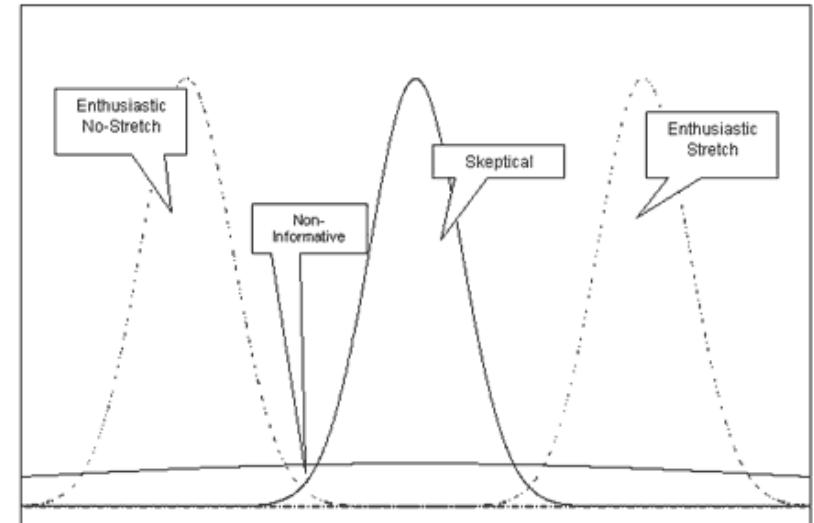
- Priors can benefit the analysis
 - Incorporate previous information
 - Make the analysis more conservative



Benefits of stretching before exercise?

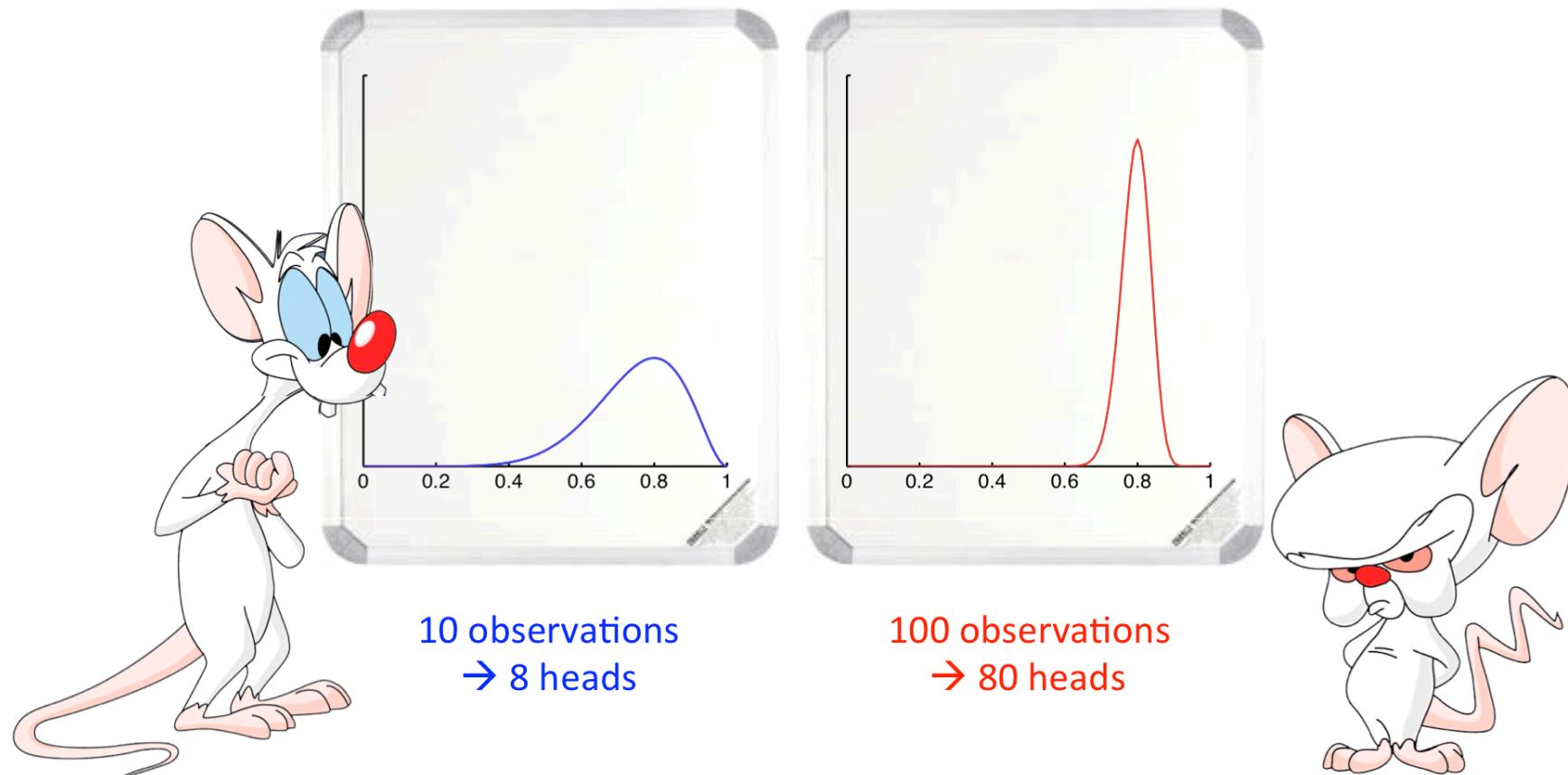
Priors

- Priors can benefit the analysis
 - Incorporate previous information
 - Make the analysis more conservative
- Priors can be detrimental if the data is not very informative
- Probability distributions of parameters
 - Reflect the action of random forces or reflect your uncertainty
 - What are the likely values?

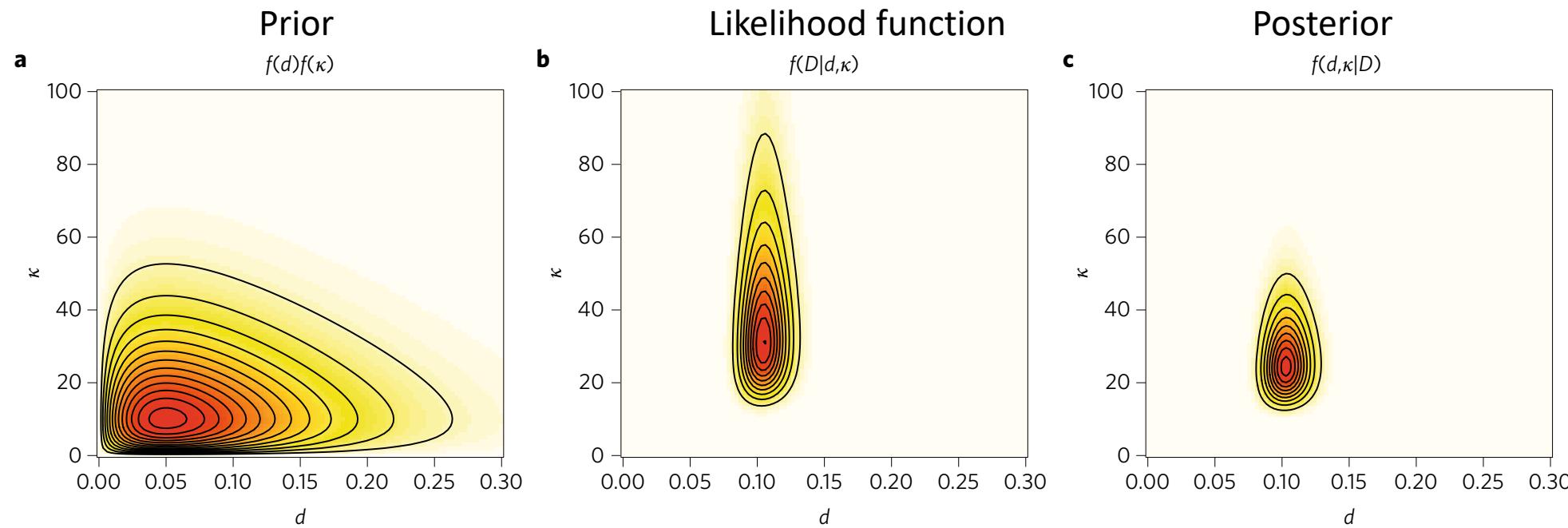


How much “faith” do we put in our priors?

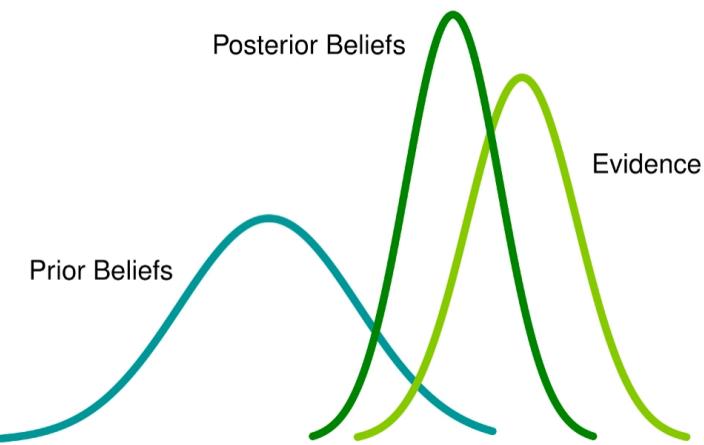
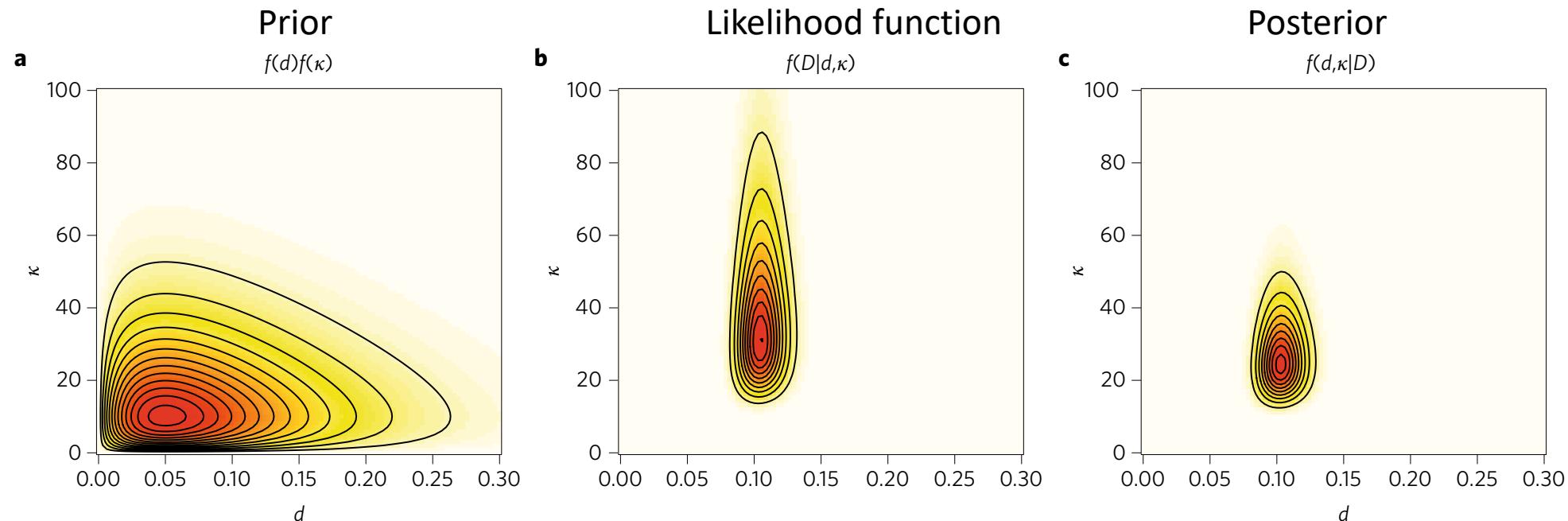
Precision-weighting: who are you going to believe more?



Impact of prior on posterior

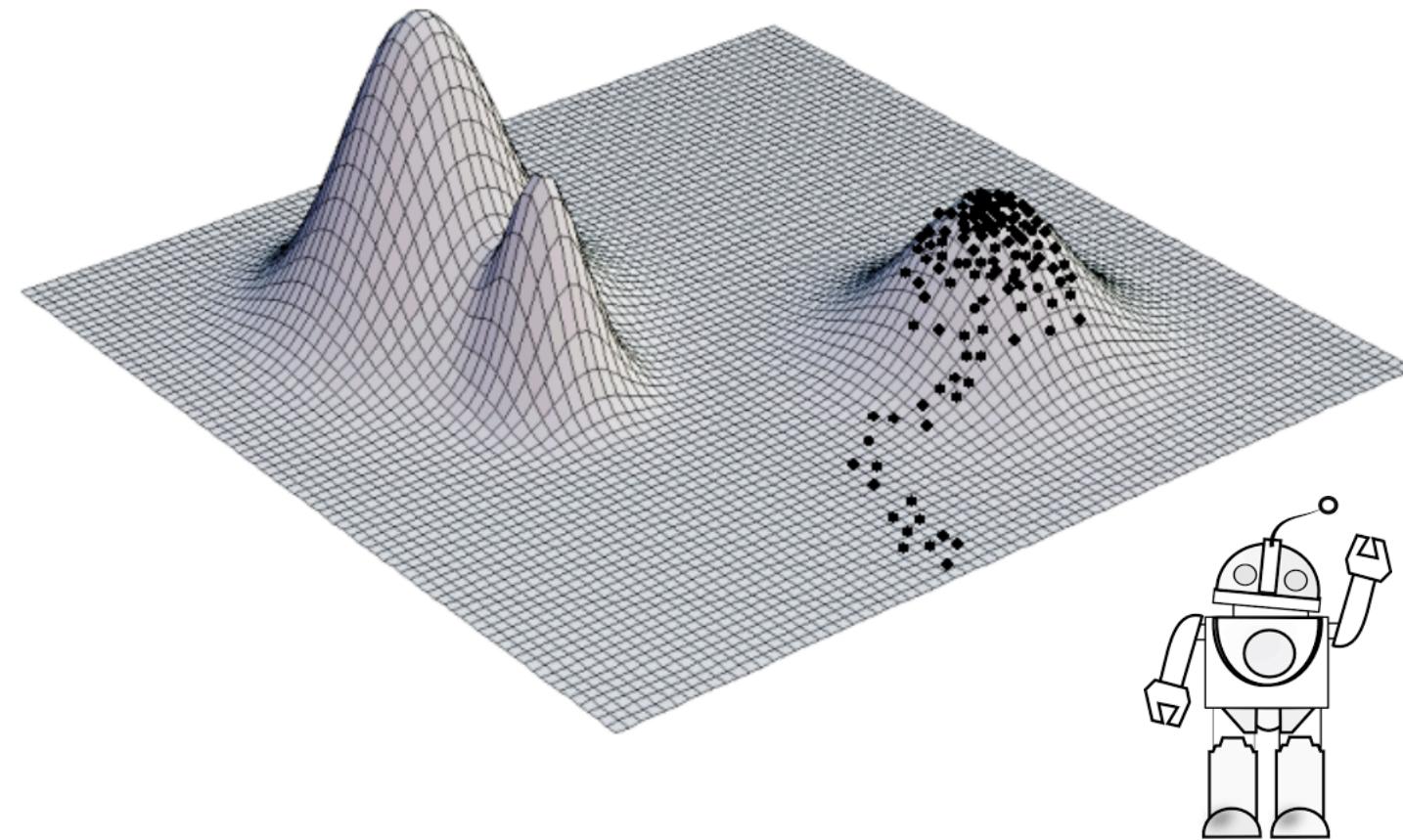


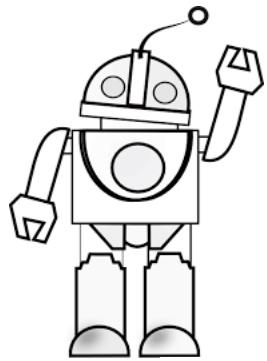
Impact of prior on posterior



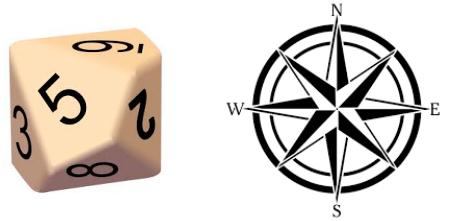
Tree space may be complex

Stuck at a local optimum





Markov Chain Monte Carlo (MCMC)



Proposal 1: 2 steps uphill → Accepted

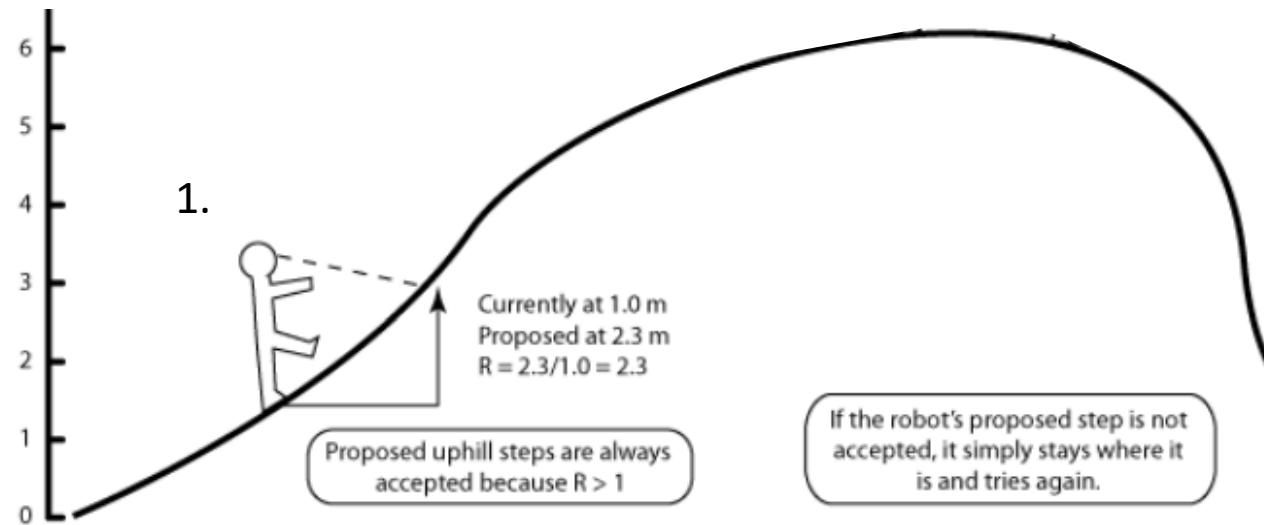
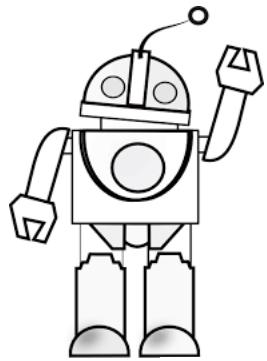


Illustration of MCMC method process (Lewis, 2011)



Markov Chain Monte Carlo (MCMC)



Proposal 1: 2 steps uphill → Accepted

Proposal 2: 1 step downhill → Accepted

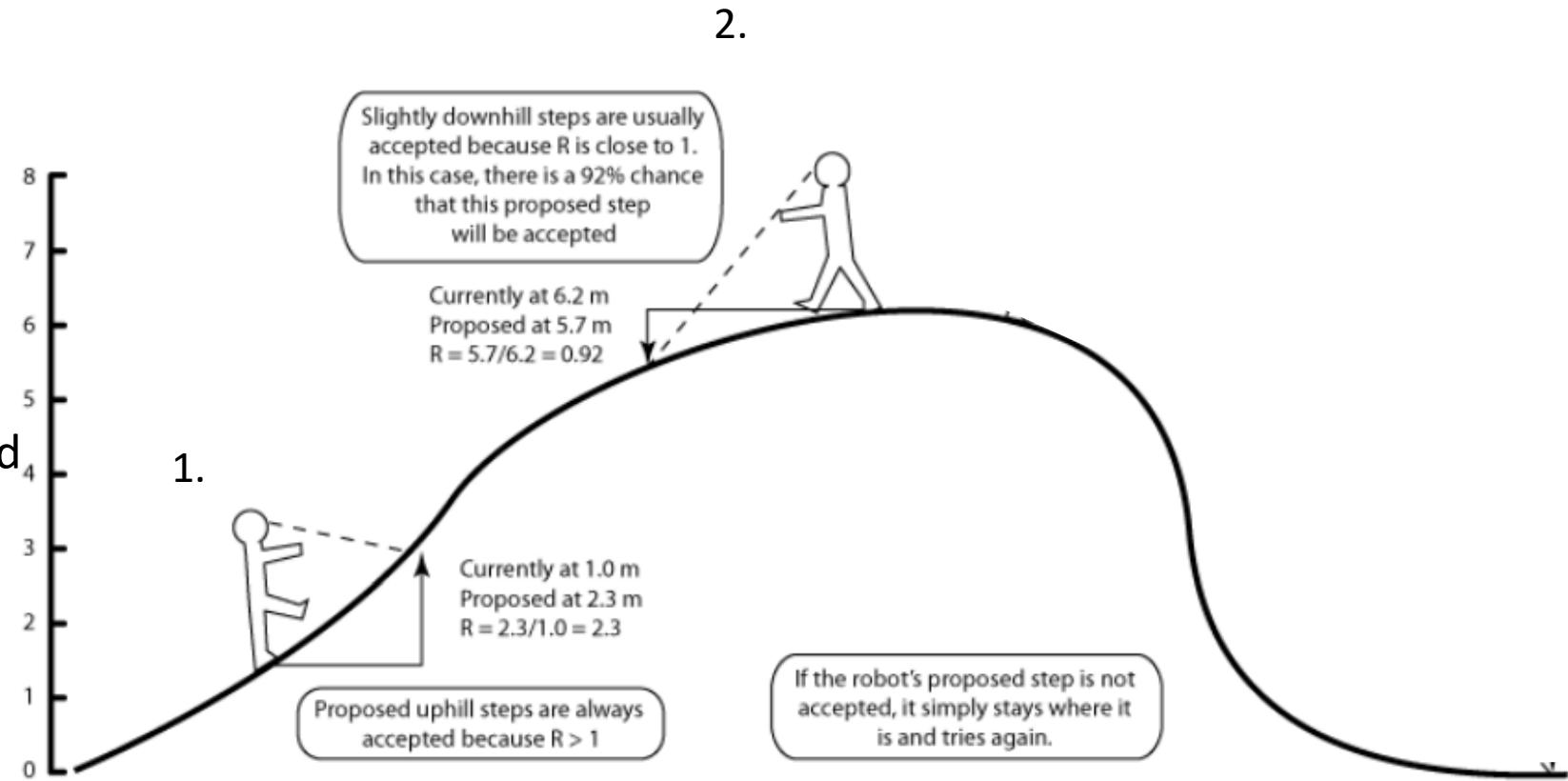
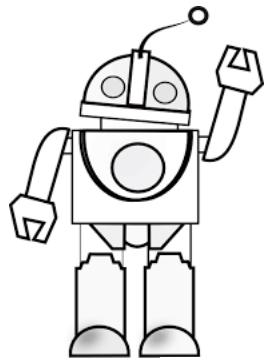


Illustration of MCMC method process (Lewis, 2011)



Markov Chain Monte Carlo (MCMC)



Proposal 1: 2 steps uphill → Accepted

Proposal 2: 1 step downhill → Accepted

Proposal 3: 10 steps downhill → Rejected

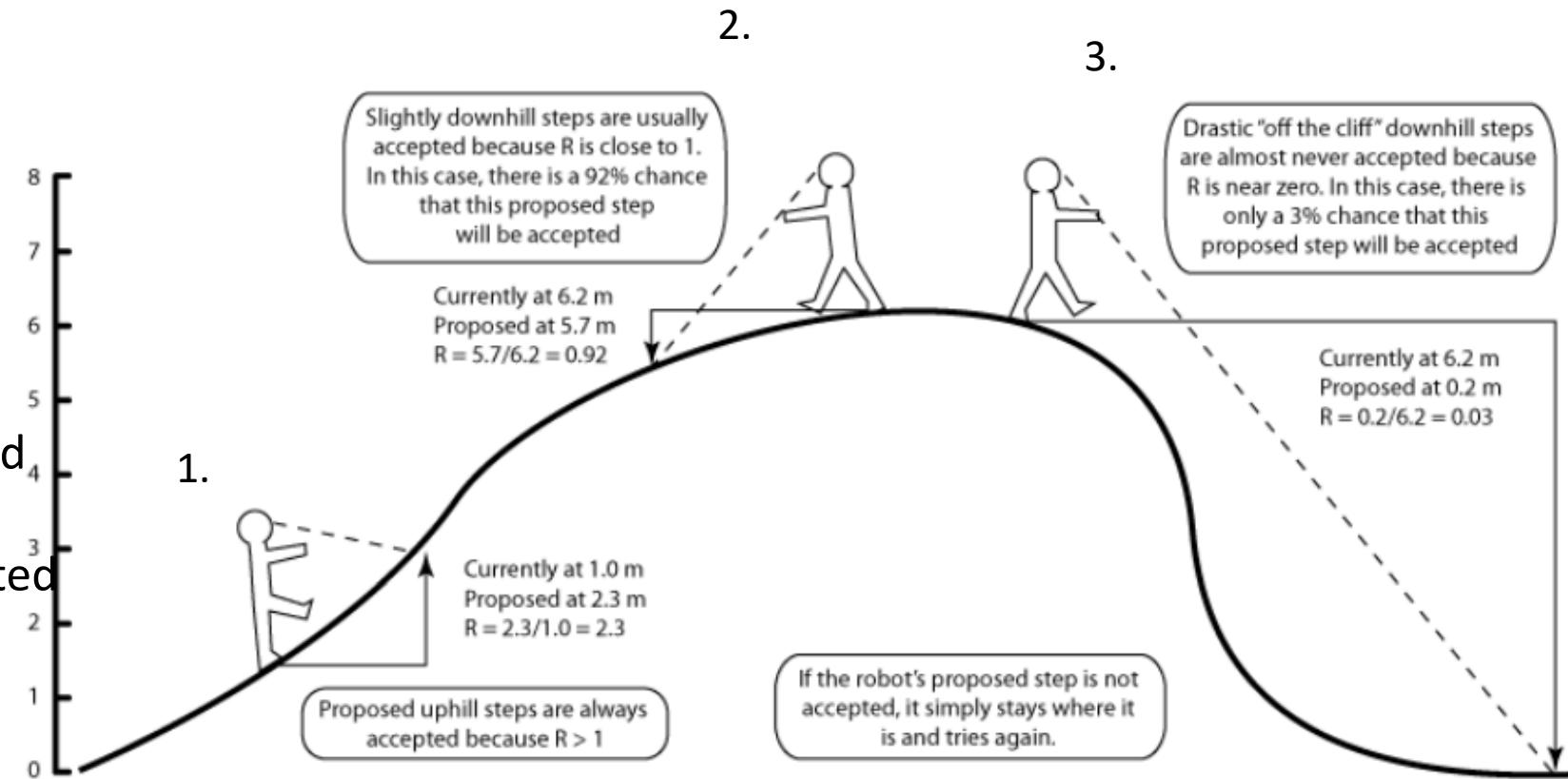
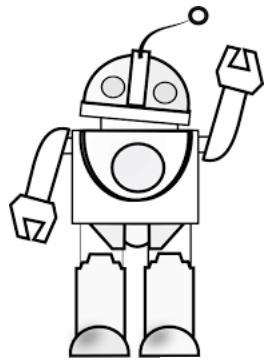


Illustration of MCMC method process (Lewis, 2011)



Markov Chain Monte Carlo (MCMC)



Proposal 1: 2 steps uphill → Accepted

Proposal 2: 1 step downhill → Accepted

Proposal 3: 10 steps downhill → Rejected

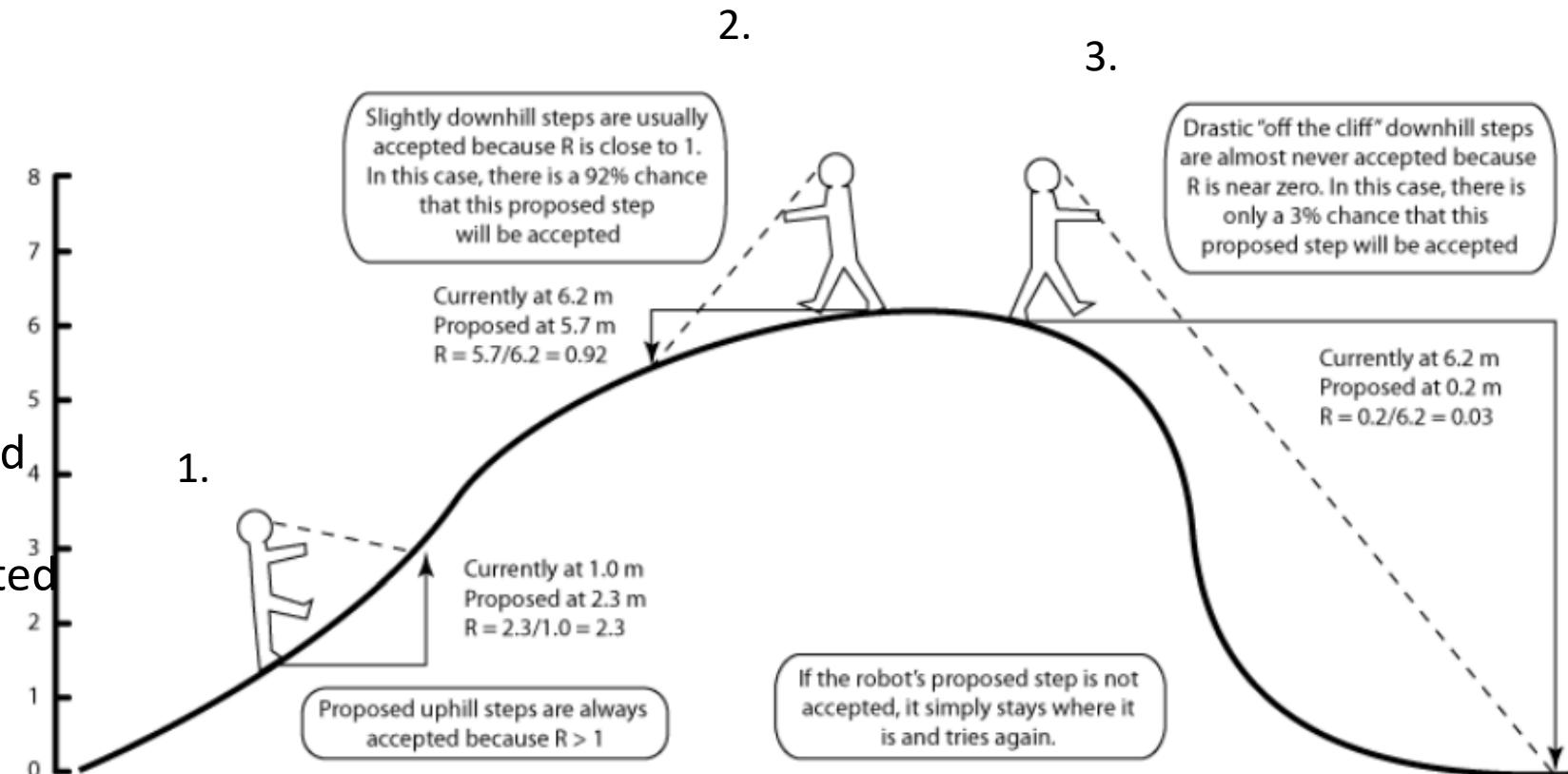
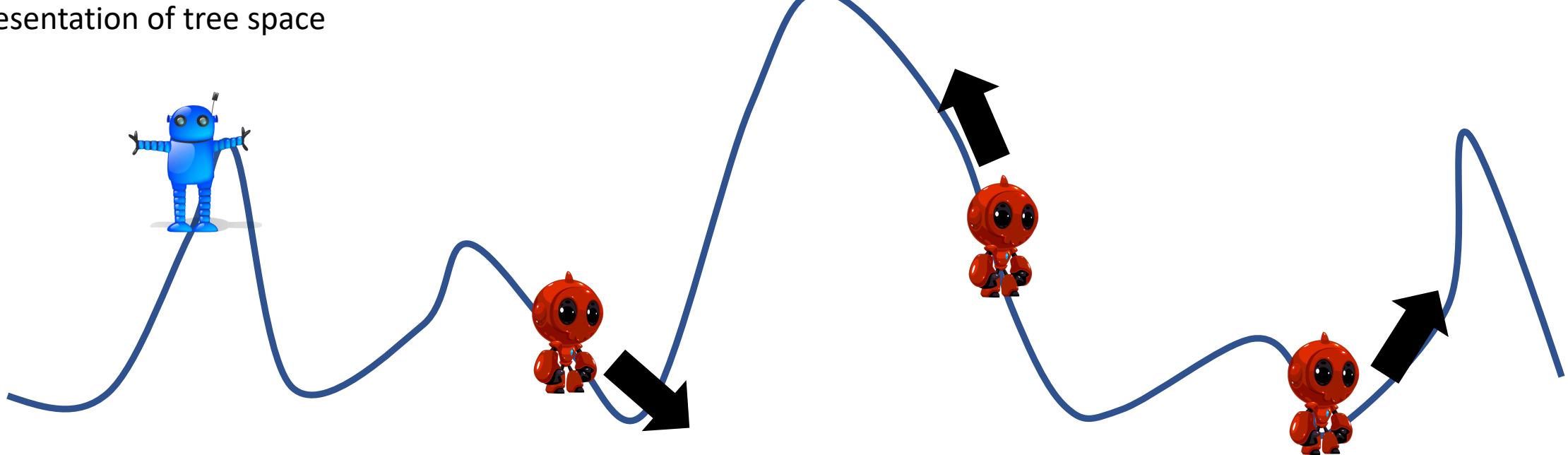


Illustration of MCMC method process (Lewis, 2011)

Markov Chain Monte Carlo (MCMC)

- Four chains (robots) exploring tree space
- 1 cold chain never takes a step with a worse score, 3 heated chains vary in chance of taking a worse step

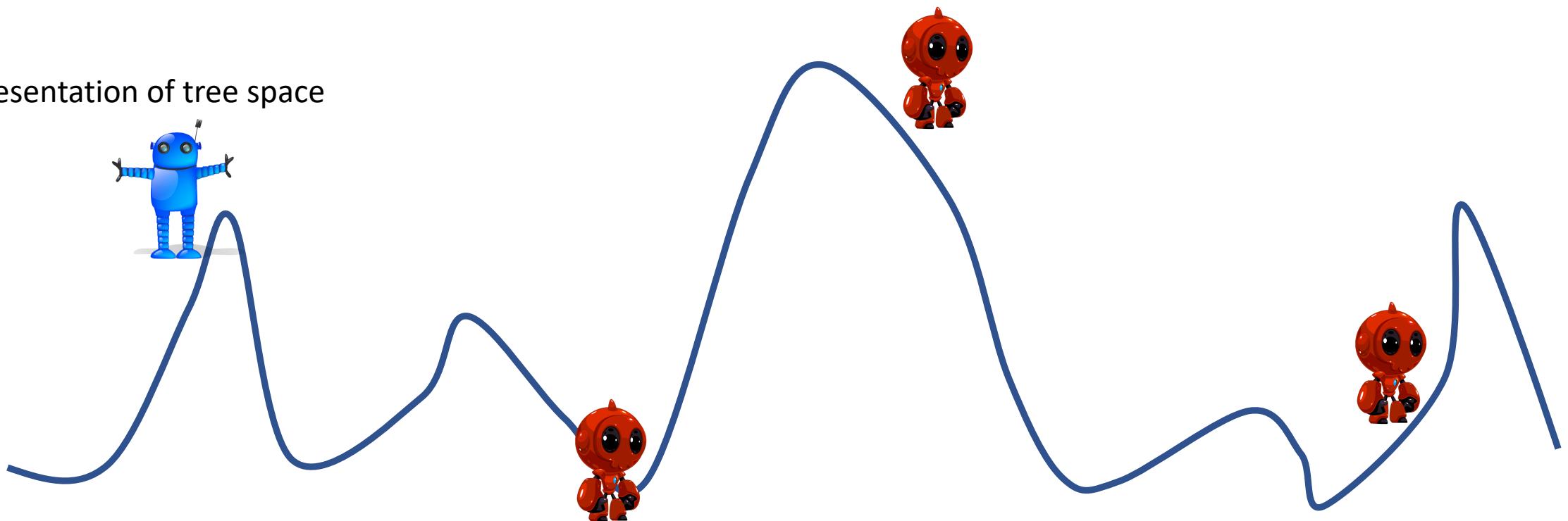
Representation of tree space



Markov Chain Monte Carlo (MCMC)

- Four chains (robots) exploring tree space
- The cold chain is the recorder, and if one of the hot chains finds a higher scoring place, they will switch positions

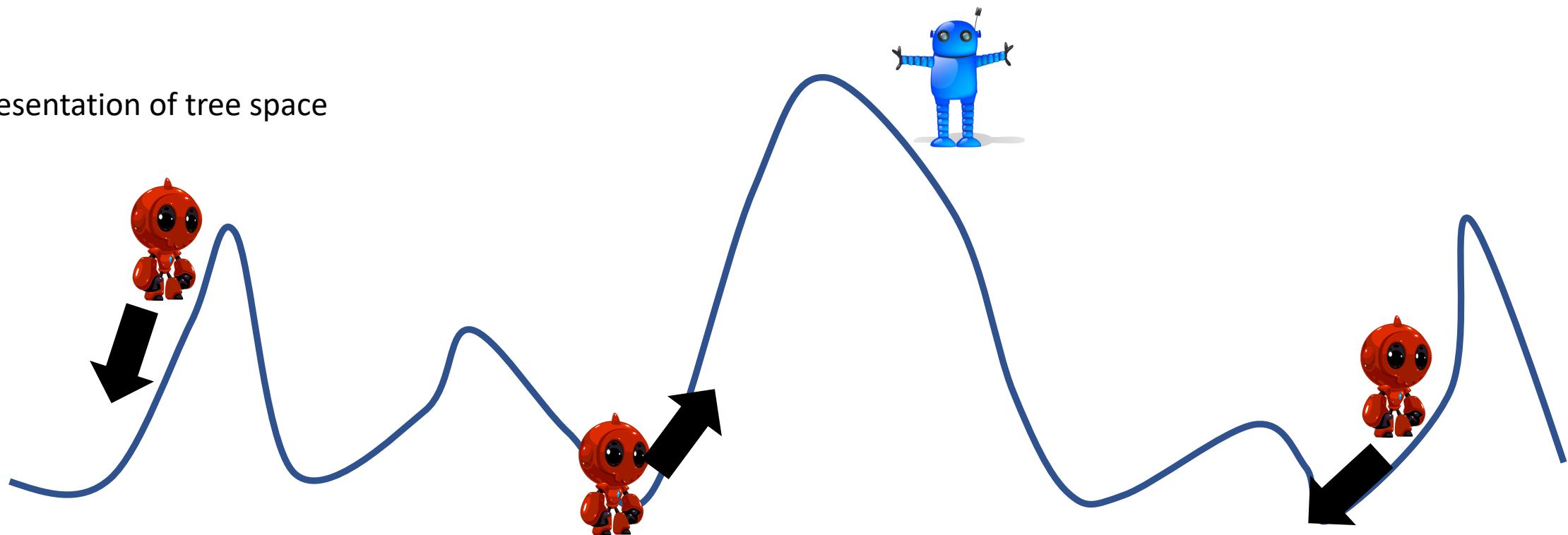
Representation of tree space



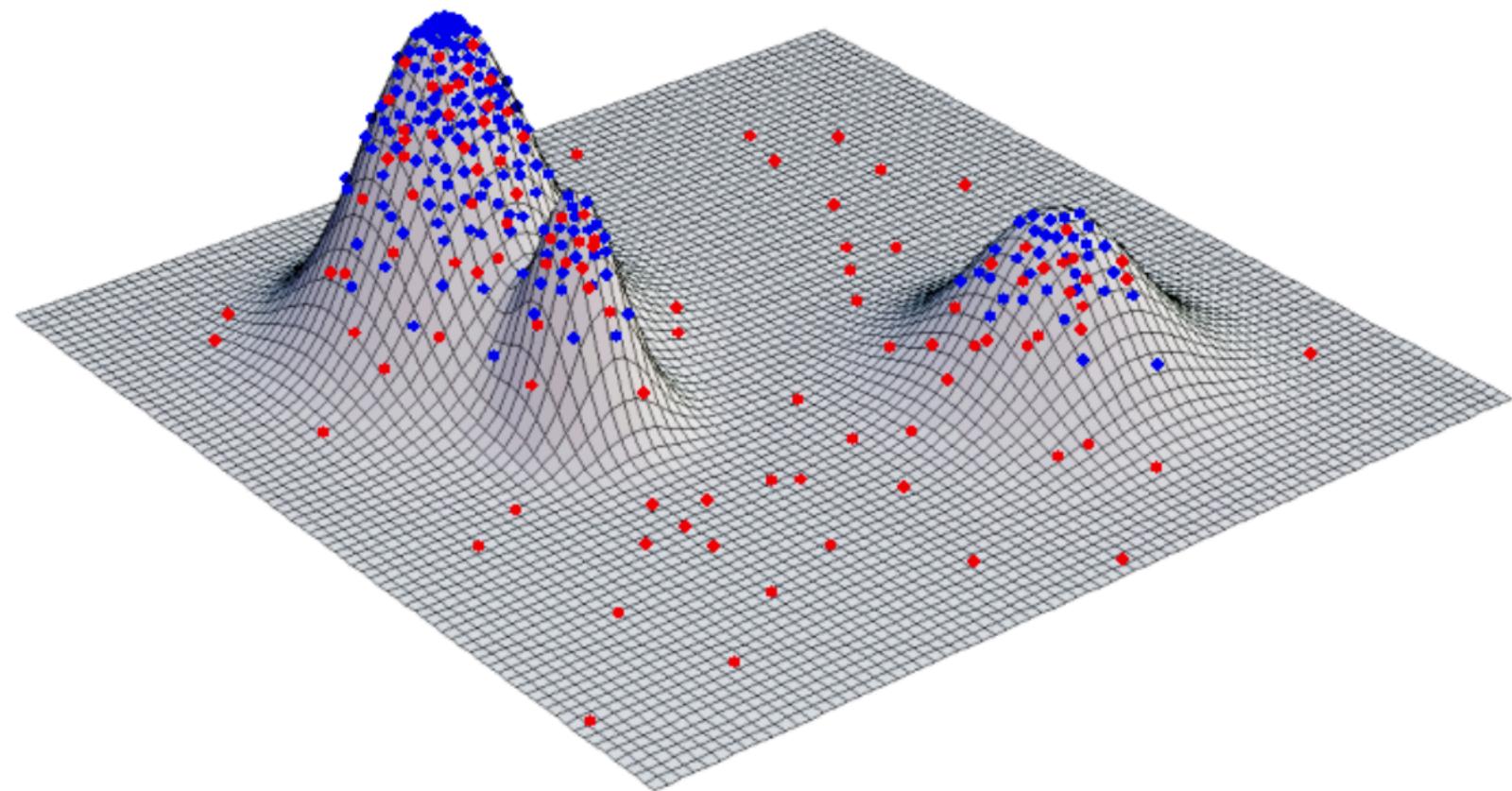
Markov Chain Monte Carlo (MCMC)

- Four chains (robots) exploring tree space
- The cold chain is the recorder, and if one of the hot chains finds a higher scoring place, they will switch positions

Representation of tree space

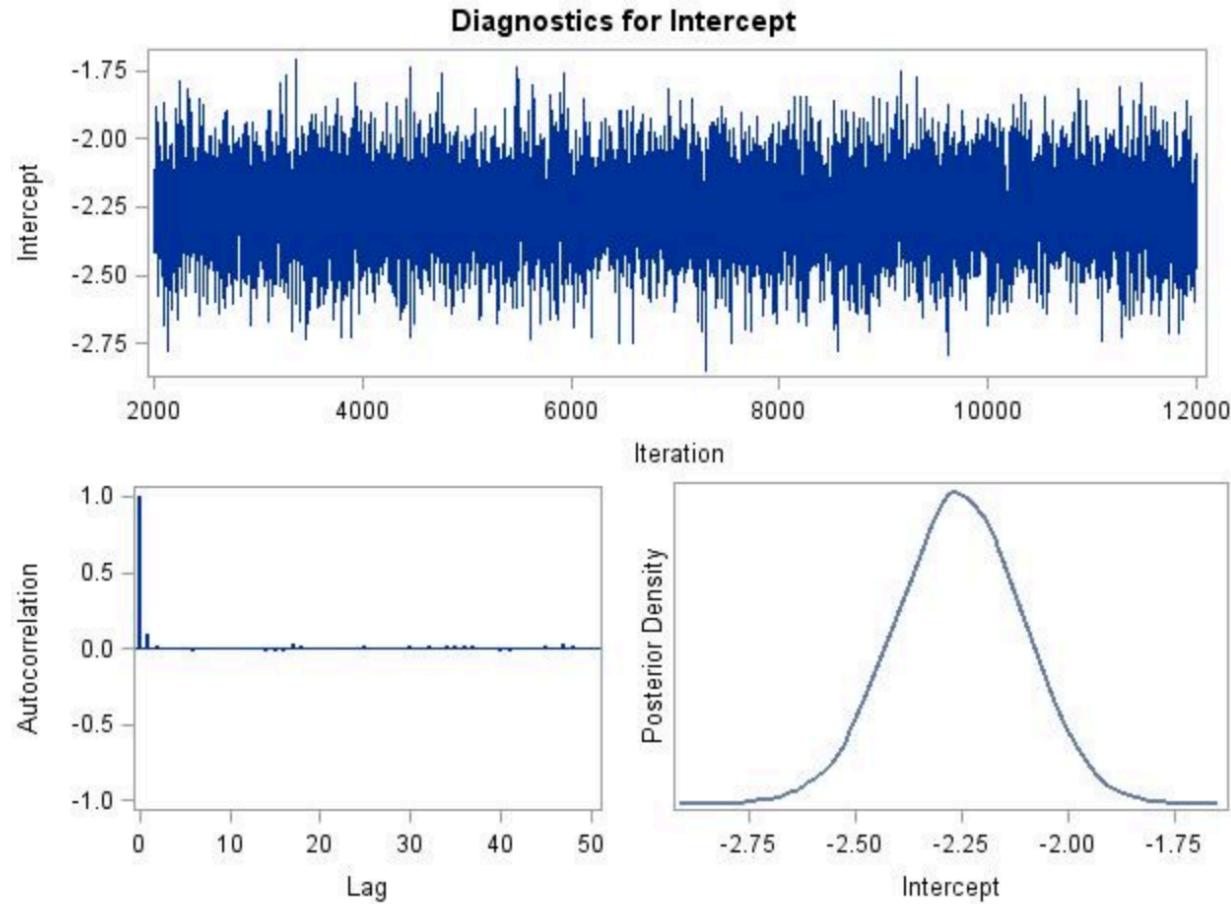


Exploration of tree space



MCMC

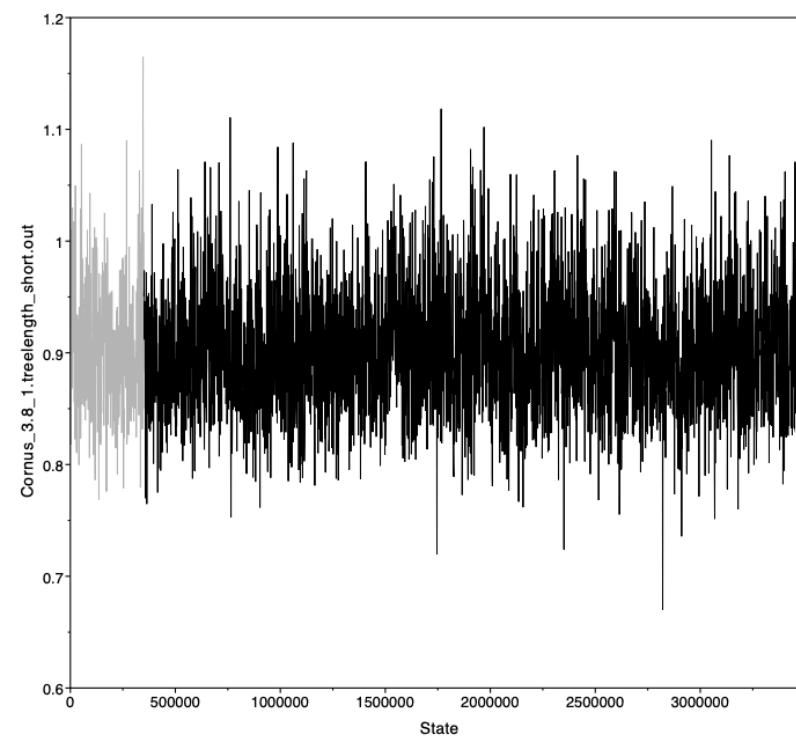
- Yields much larger sample of trees than ML because it produces one tree for every generation versus one tree per tree search
- Trees produced are highly auto-correlated
 - Sample relatively infrequently and discard trees early in the process
 - Millions of generations required compared to 1,000 bootstrap replicates to explore tree space



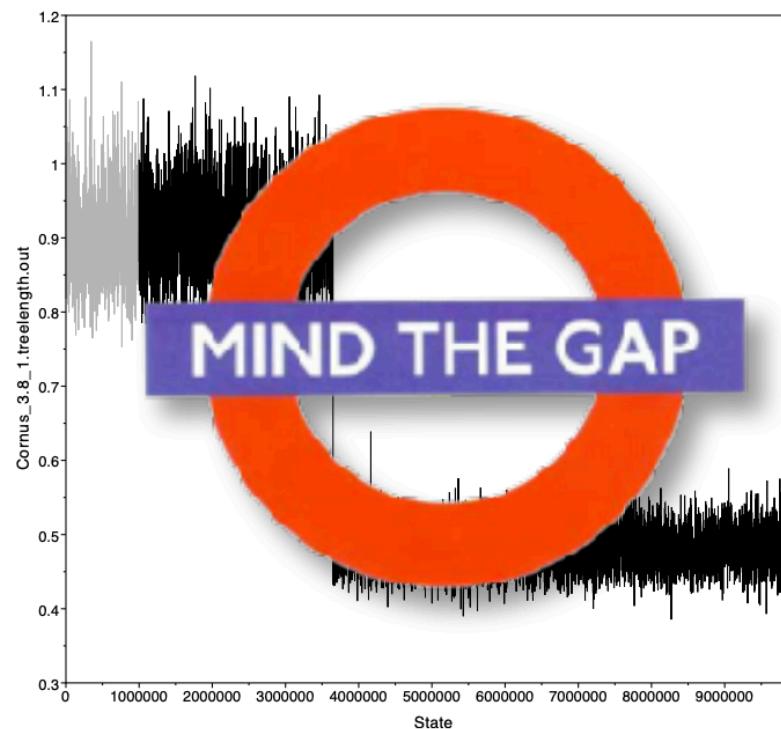
How the run may look

Tracer run after a Bayesian run

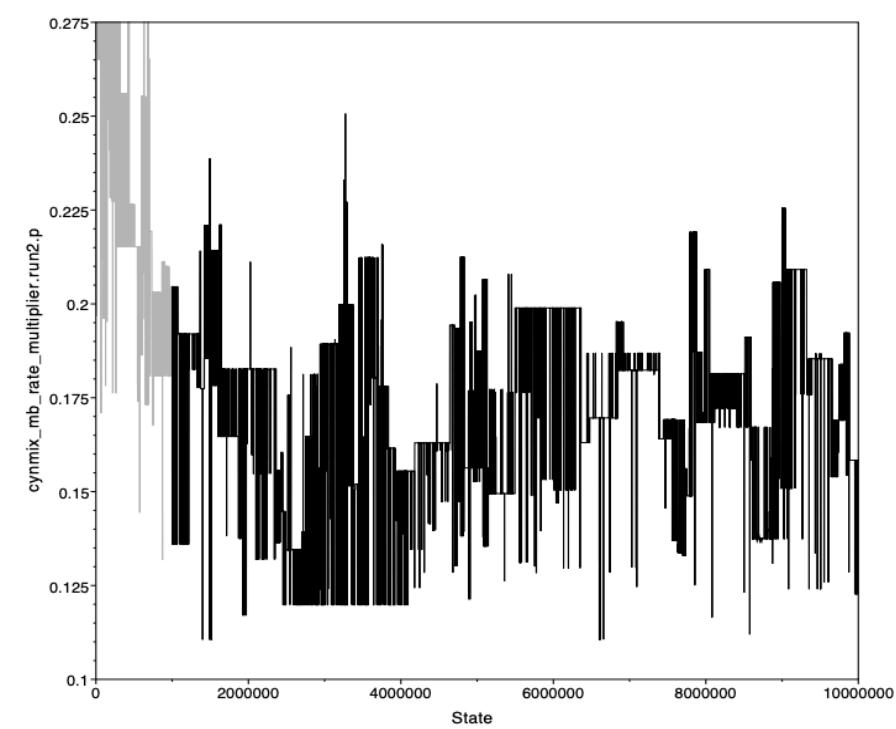
all looks good...



until it doesn't



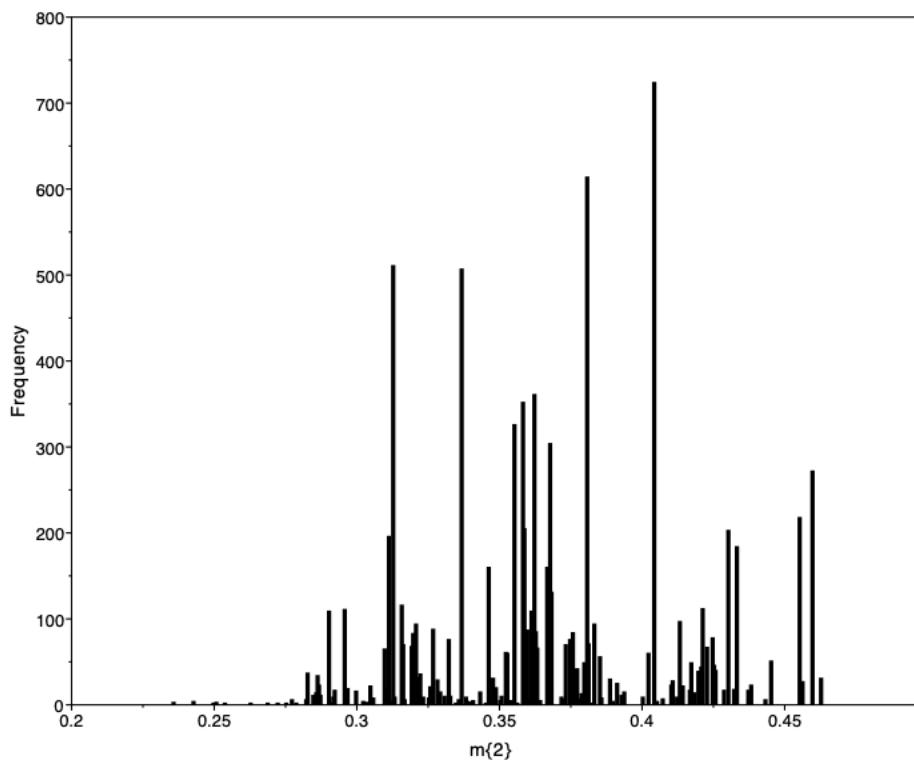
bad mixing



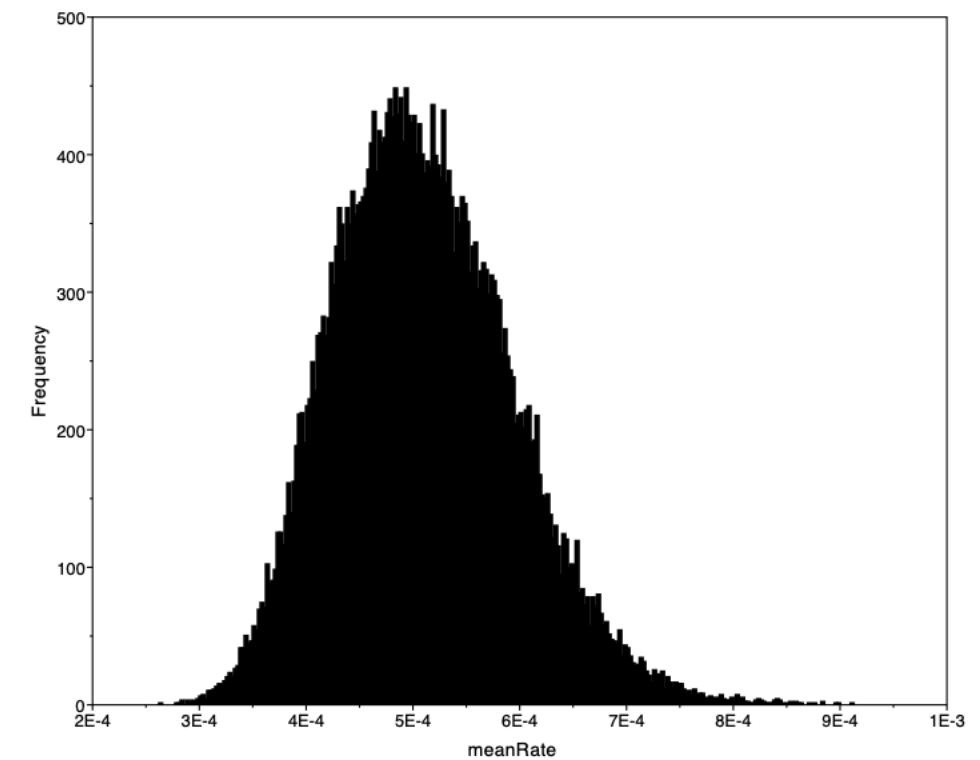
How the run may look

Posterior estimates of parameters

bad mixing

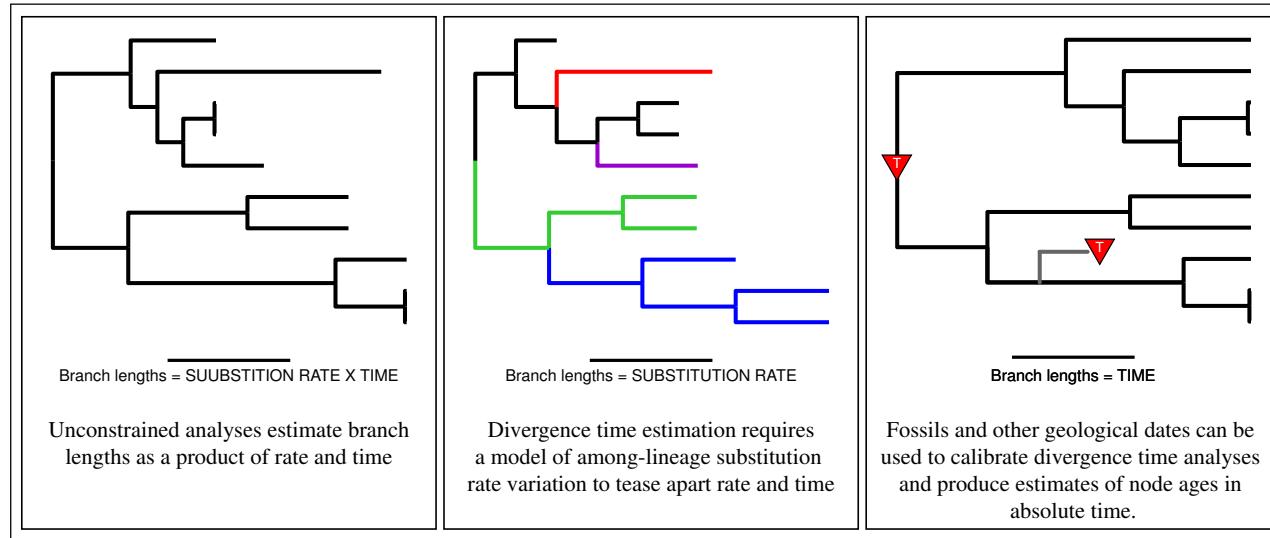


better mixing



Divergence time estimation

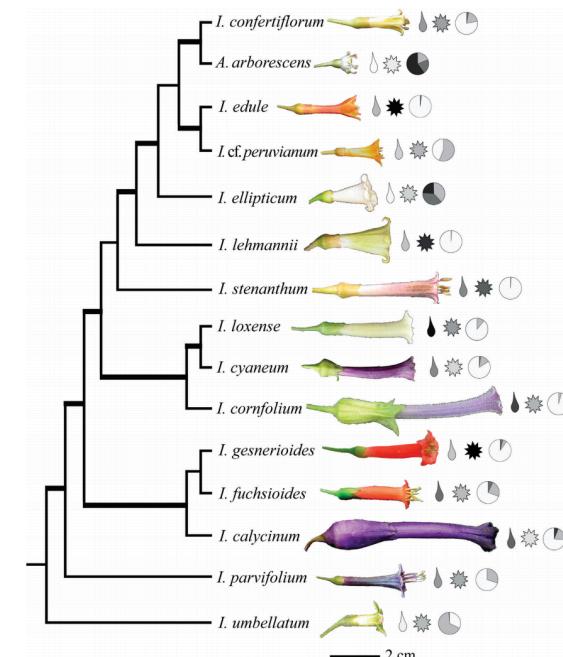
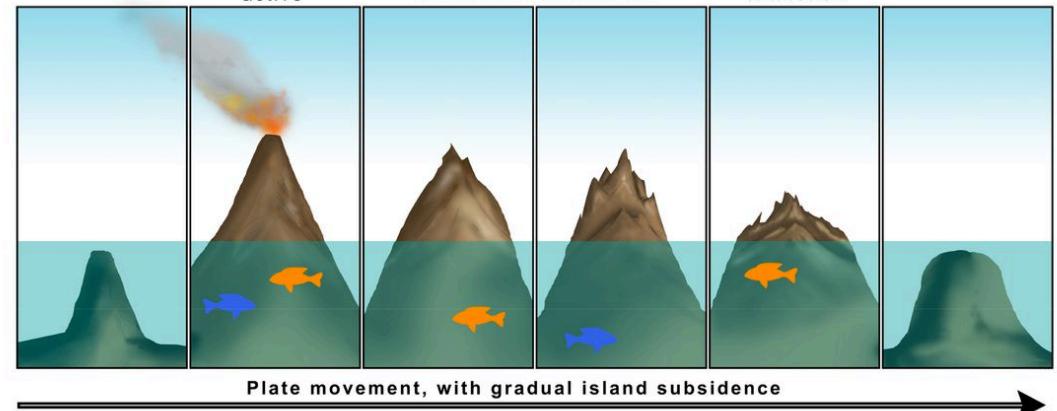
- The rate of evolution and time are linked when inferring the genetic difference between species



- Need a model of lineage-specific substitution rate variation to tease these apart
- To infer real-time (or absolute time) need fossil data, geographic timing (e.g. island formation) or other reliable sources

Absolute dating versus relative dating

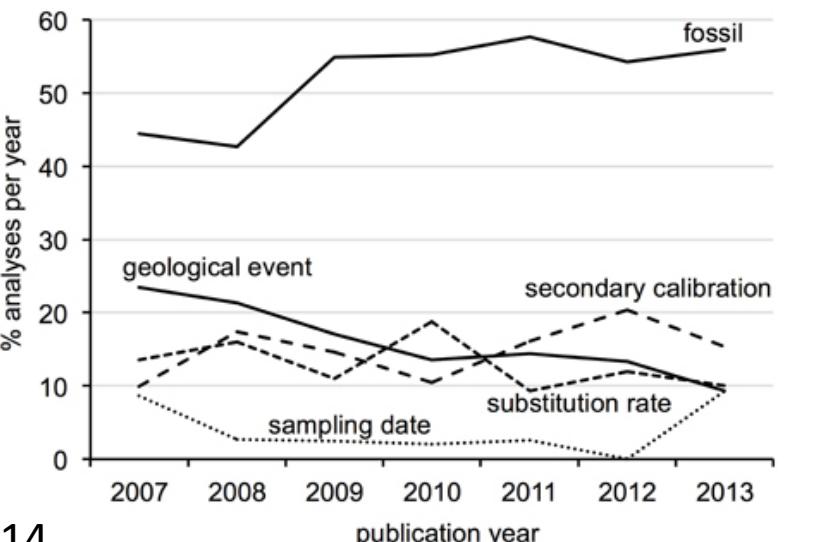
- Sometimes your group of interest does not have any fossils or other biogeographic information you can use for dates
- Relative dating may help for analyses looking at rates of change for continuous traits or relative rates of diversification
- However, if you want to know about biogeographical event and diversification, relative dating will not work for you



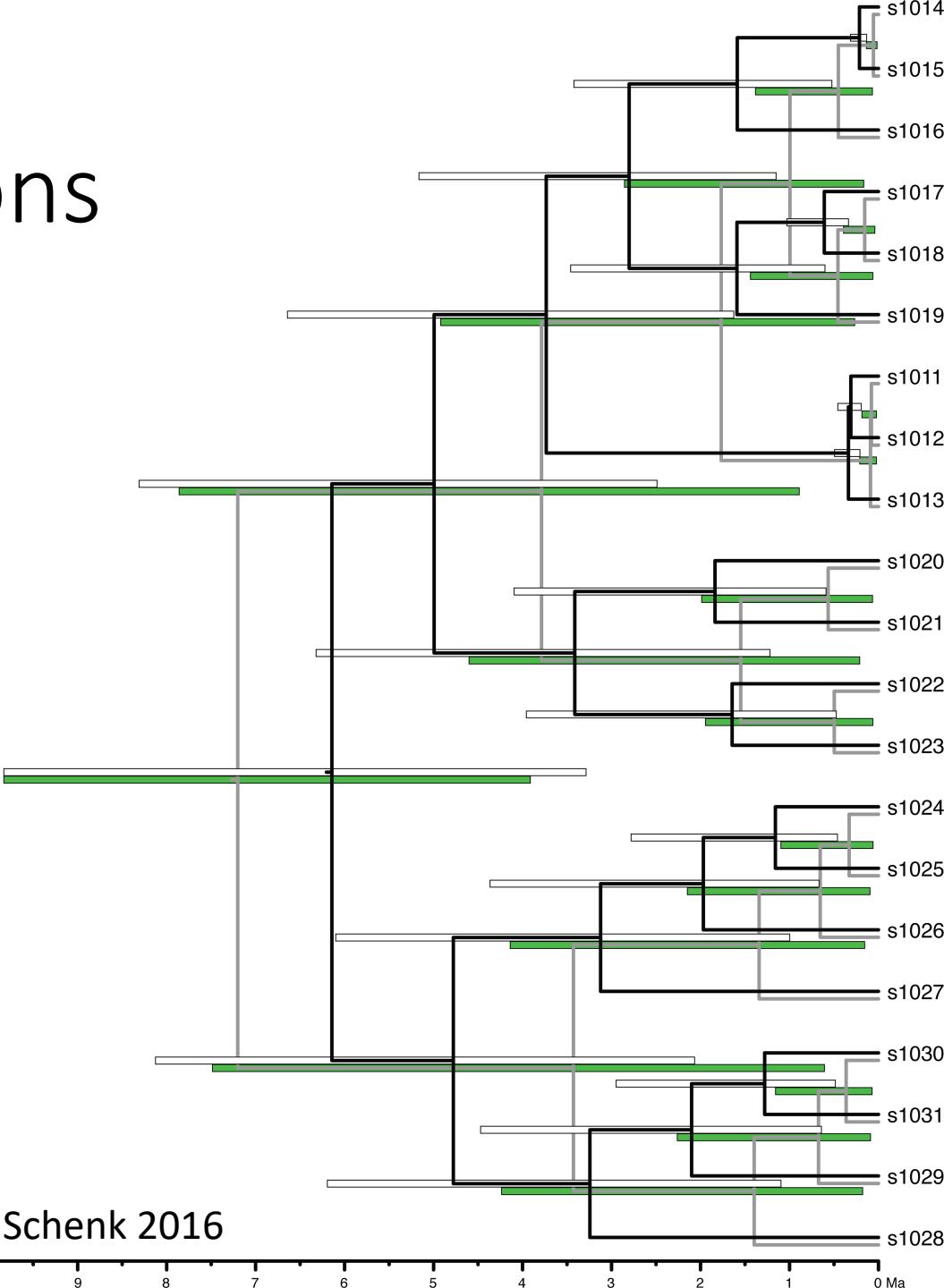
Smith et al. 2007

Dating options

- Can also use “known” substitution rates or dates from previous studies (secondary calibration; but there are several papers suggesting this may not be good), serially sampled data (virus data)



Hipsley and Muller 2014



Modeling substitution rates

- Substitution rate is influenced by many factors including mutation rate, population size, generation time, and selection
- Simplest model is a molecular clock or strict molecular clock
 - Constant rate through time
- Is this model realistic? Do you think these plants have the same substitution rates over long periods of time?



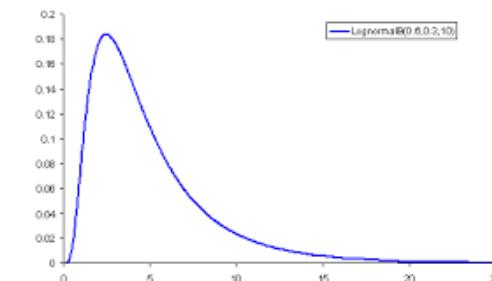
Arabidopsis thaliana



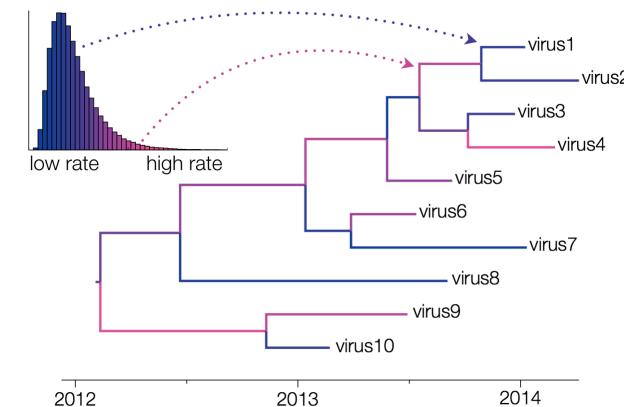
Redwoods

More advanced models for substitution rates

- Local molecular clocks - closely related lineages evolve at similar rates
- Autocorrelated rates – substitution rates evolve gradually over time and may be drawn from a log-normal distribution

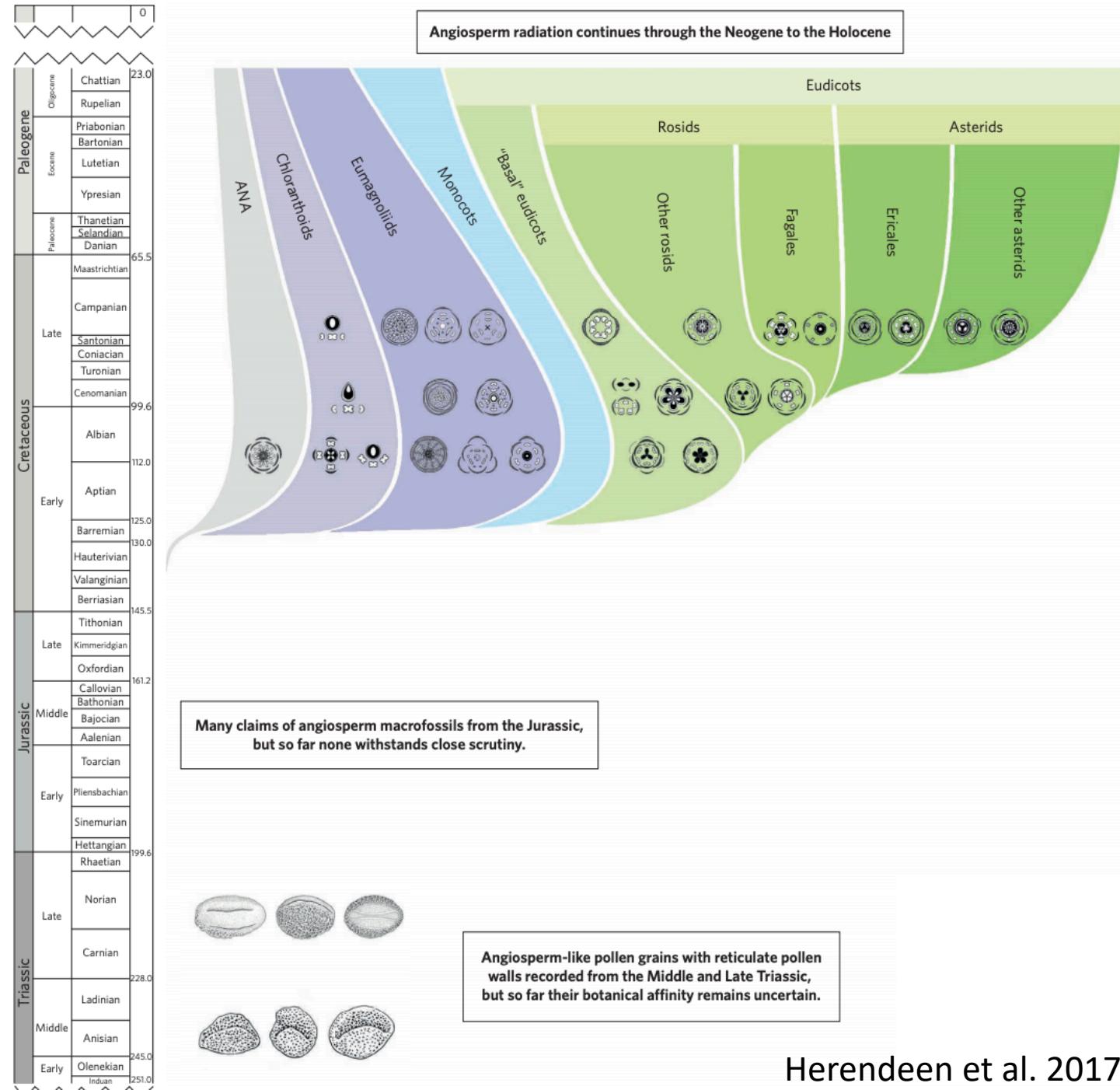


- Uncorrelated rates – each branch is drawn from an underlying distribution such as an exponential or log-normal
 - Uncorrelated log-normal relaxed clock is an option in BEAST



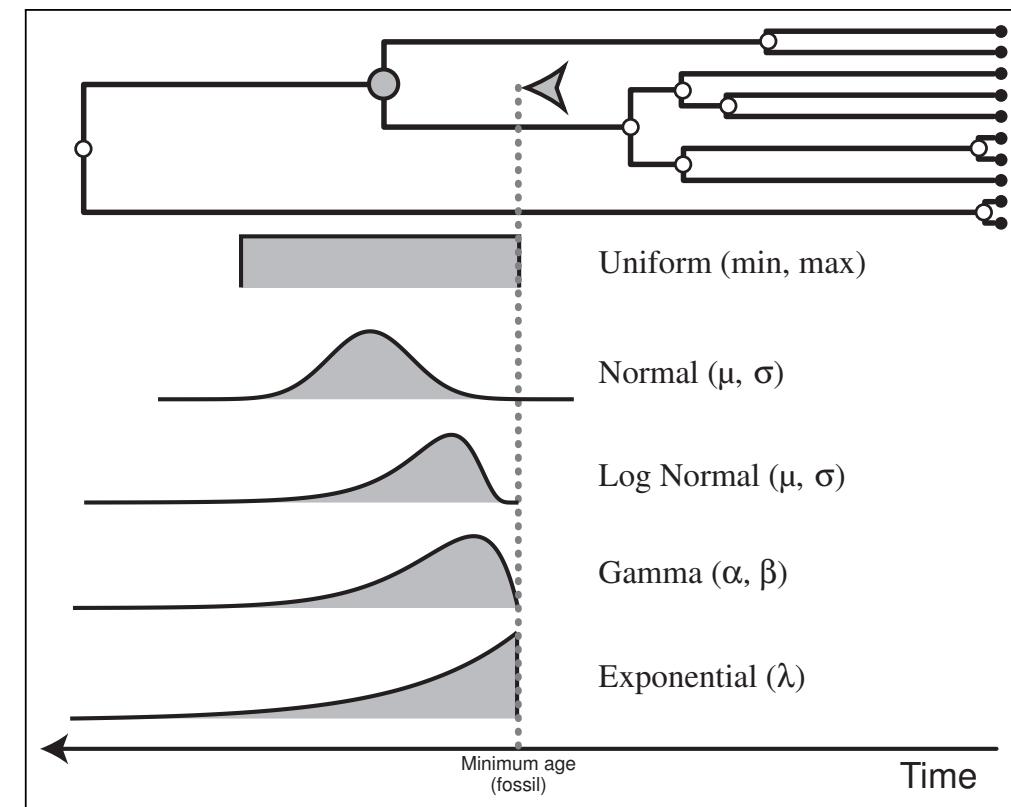
Angiosperm age

- Fossils from the Early Cretaceous, though their exact phylogenetic placement is not always certain
- Early molecular clock papers put the rise of the angiosperm around 180–140 MYA, but recent papers suggest closer to 130 MYA



Distributions for calibrating nodes

- Uniform – must have both maximum and minimum age bounds for this
- Normal – symmetrical and highest weight at mean; likely not appropriate for fossil calibrating but could be useful for biogeographical dates and secondary calibration
- Gamma – When scale parameter is large, approaches normal
- Exponential – weight is placed on nodes with ages very close to the age of the fossil
- Lognormal – highest probability on ages somewhat older than the fossil



Homework

- Assignment covering ML and Bayesian
- For most of the questions, you will need to use the alignment file from one gene and compare your trees from parsimony, ML, and Bayesian
- You will also compare one of your trees to that published with all the data in Ruhfel et al.
- There will be some questions regarding Beast, but you only need to use the 32-tip tree for that
 - Beast will probably take too long with a single gene