Bayes' rule: continuous case

Likelihood

Prior probability density

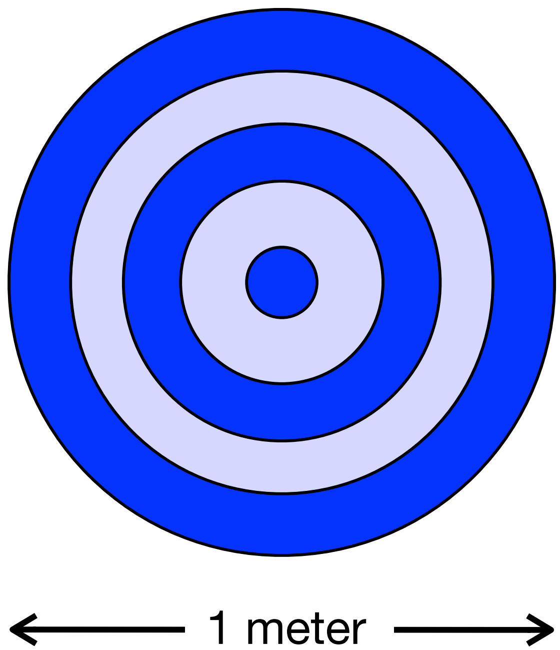
$$p(\theta|D) = \frac{p(D|\theta)\, p(\theta)}{\int p(D|\theta')\, p(\theta') d\theta'}$$

Posterior probability density

Marginal probability of the data

(a.k.a. marginal likelihood)

If you had to guess...

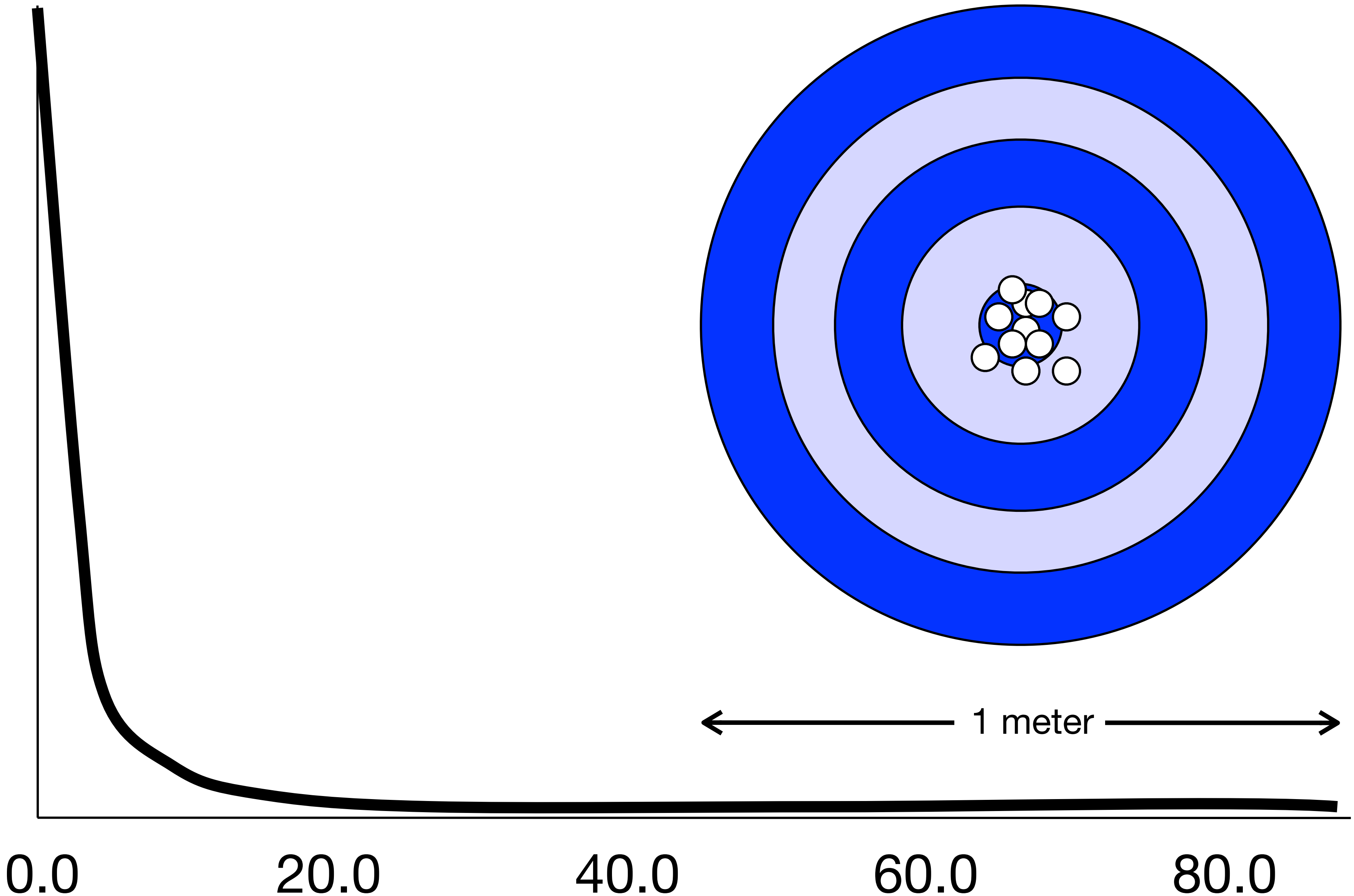


Not knowing anything about my archery abilities, draw a curve representing your view of the chances of my arrow landing a distance *d* from the center of the target (if it helps, I'm standing 50 meters away from the target)

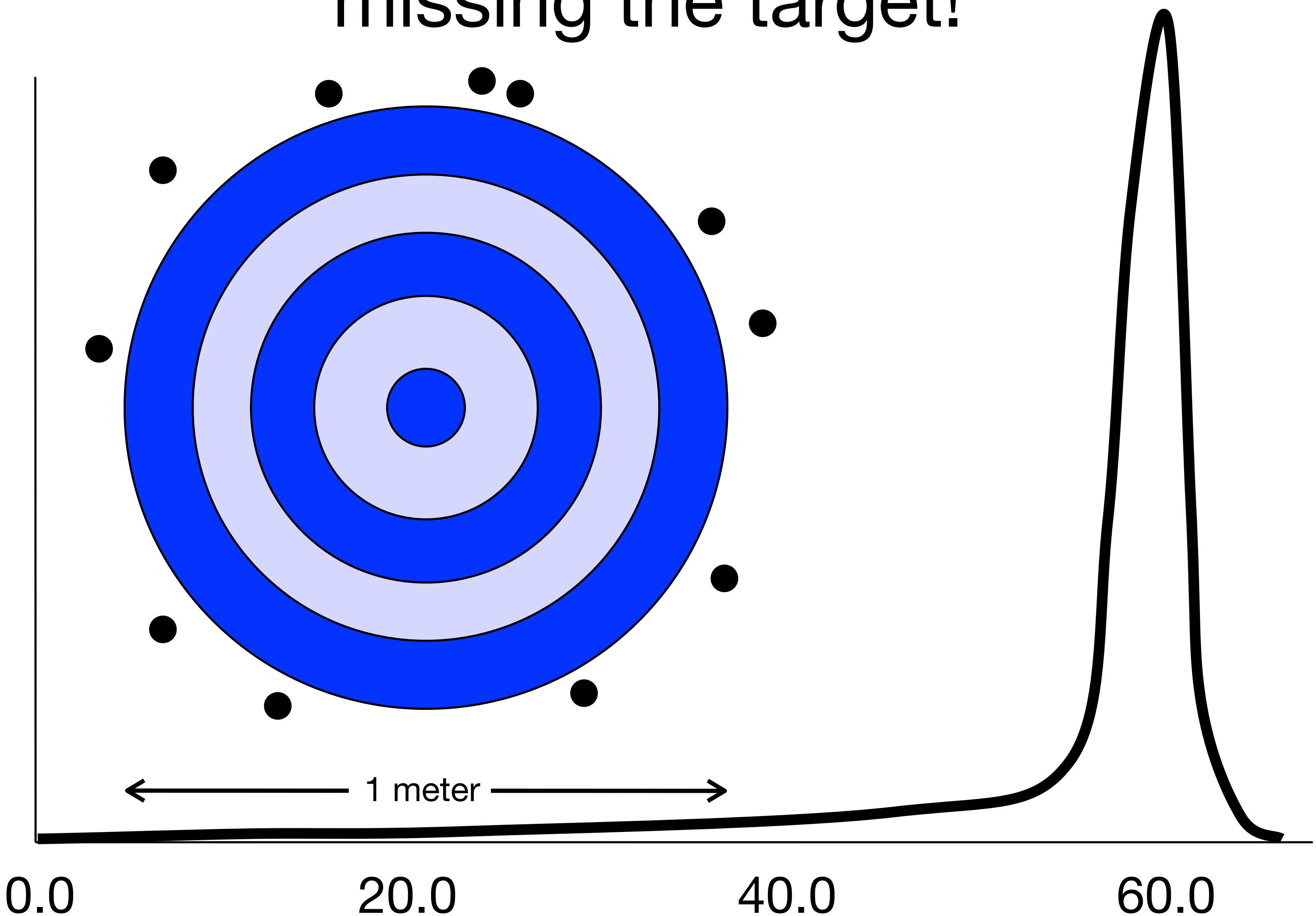
0.0   *d*   ∞

distance in centimeters from target center

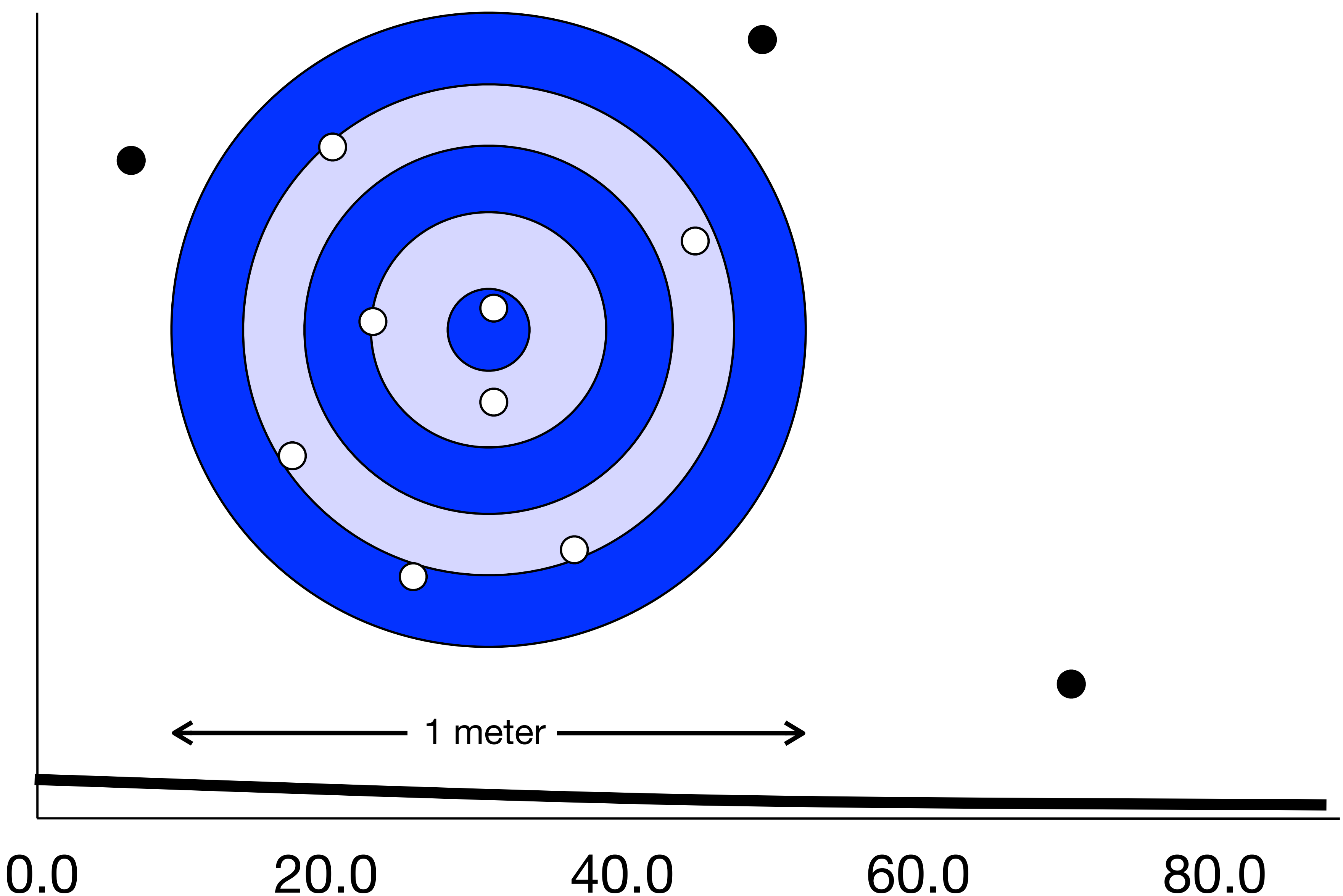
Case 1: assume I have talent



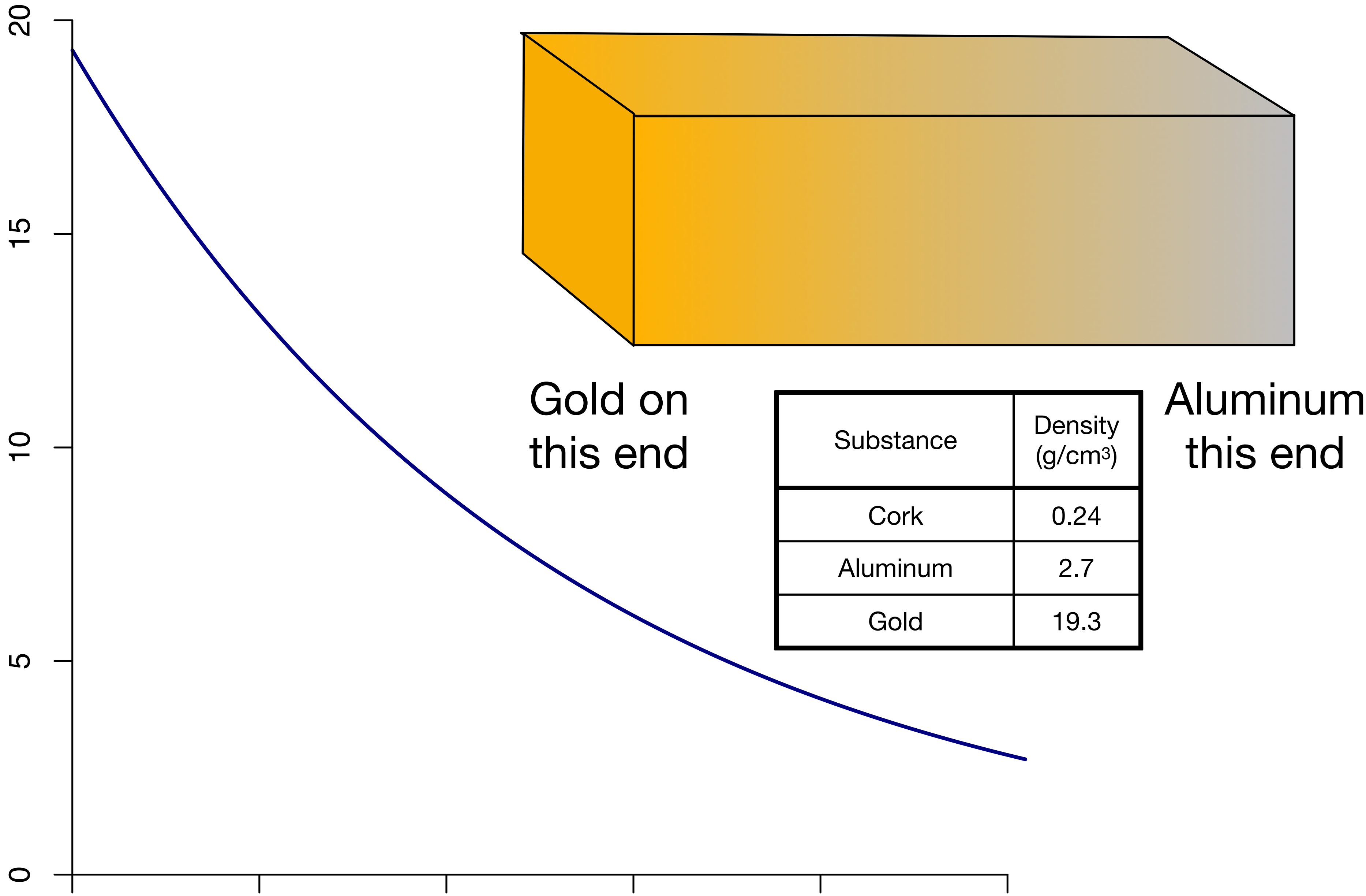
Case 2: assume I have a talent for missing the target!



Case 3: assume I have no talent

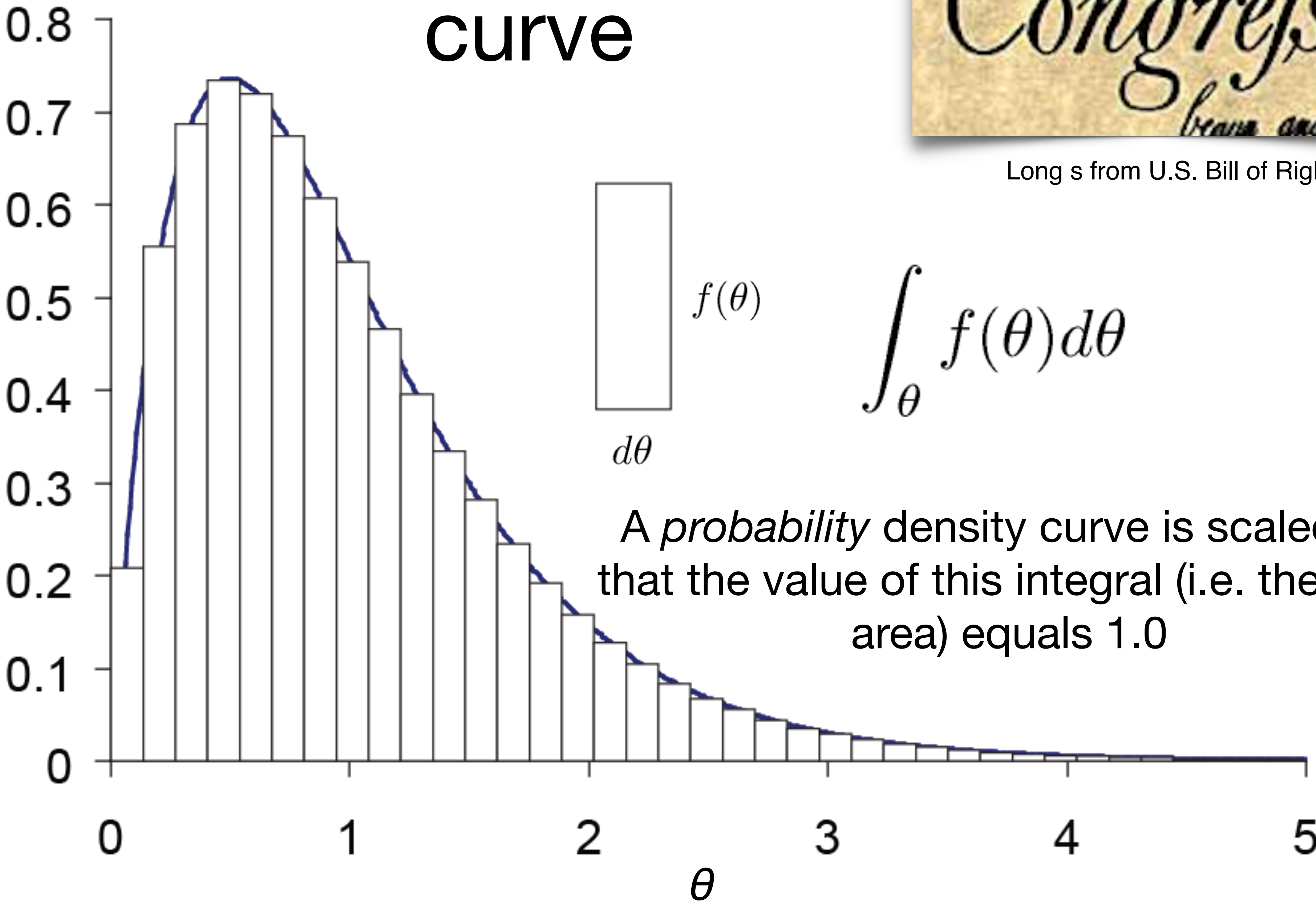


Probabilities vs. probability *densities*



Substance	Density (g/cm^3)
Cork	0.24
Aluminum	2.7
Gold	19.3

Probability density curve



Long s from U.S. Bill of Rights

$f(\theta)$

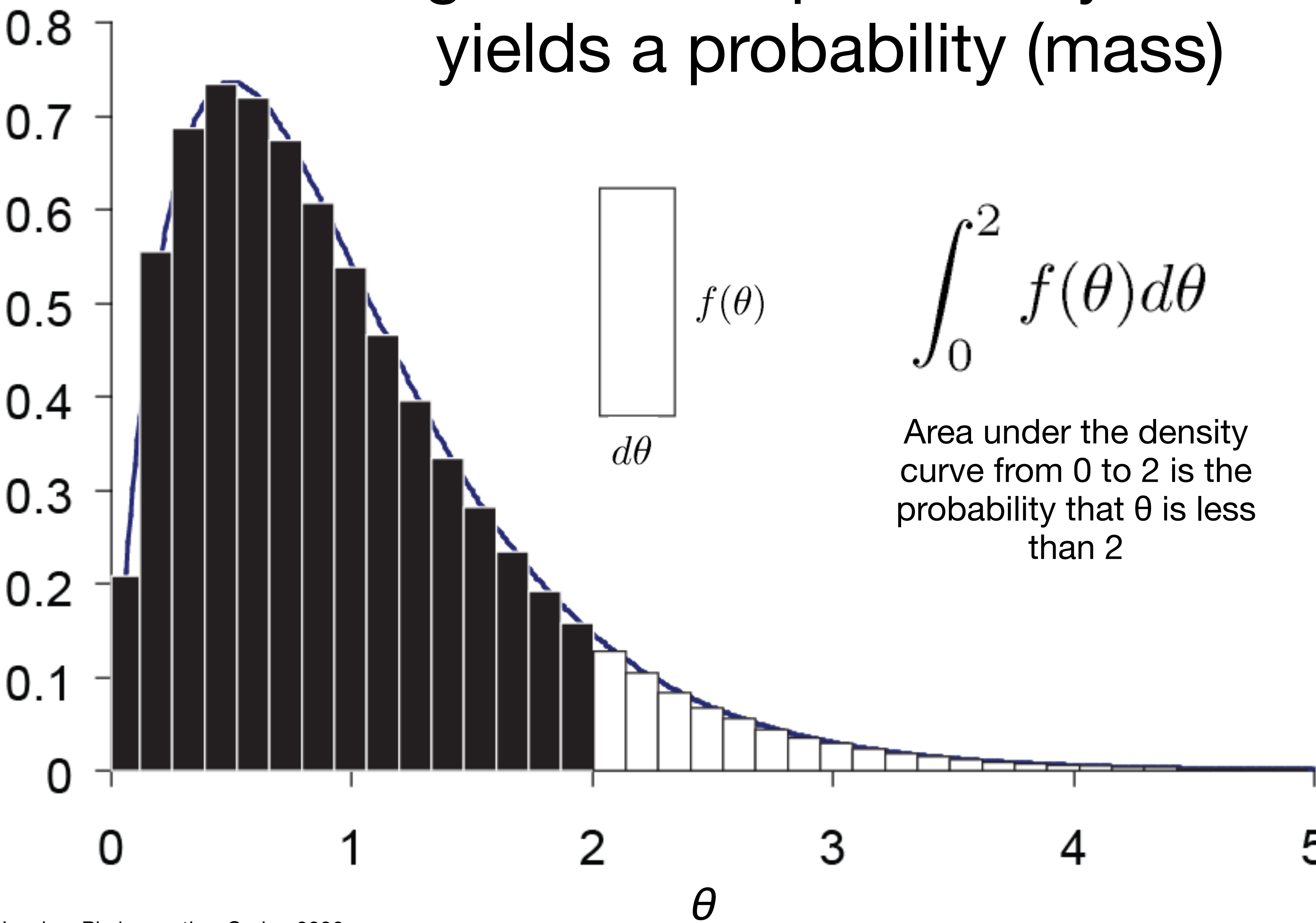
$d\theta$

$$\int_{\theta} f(\theta)d\theta$$

A *probability* density curve is scaled so that the value of this integral (i.e. the total area) equals 1.0

probability of a single point

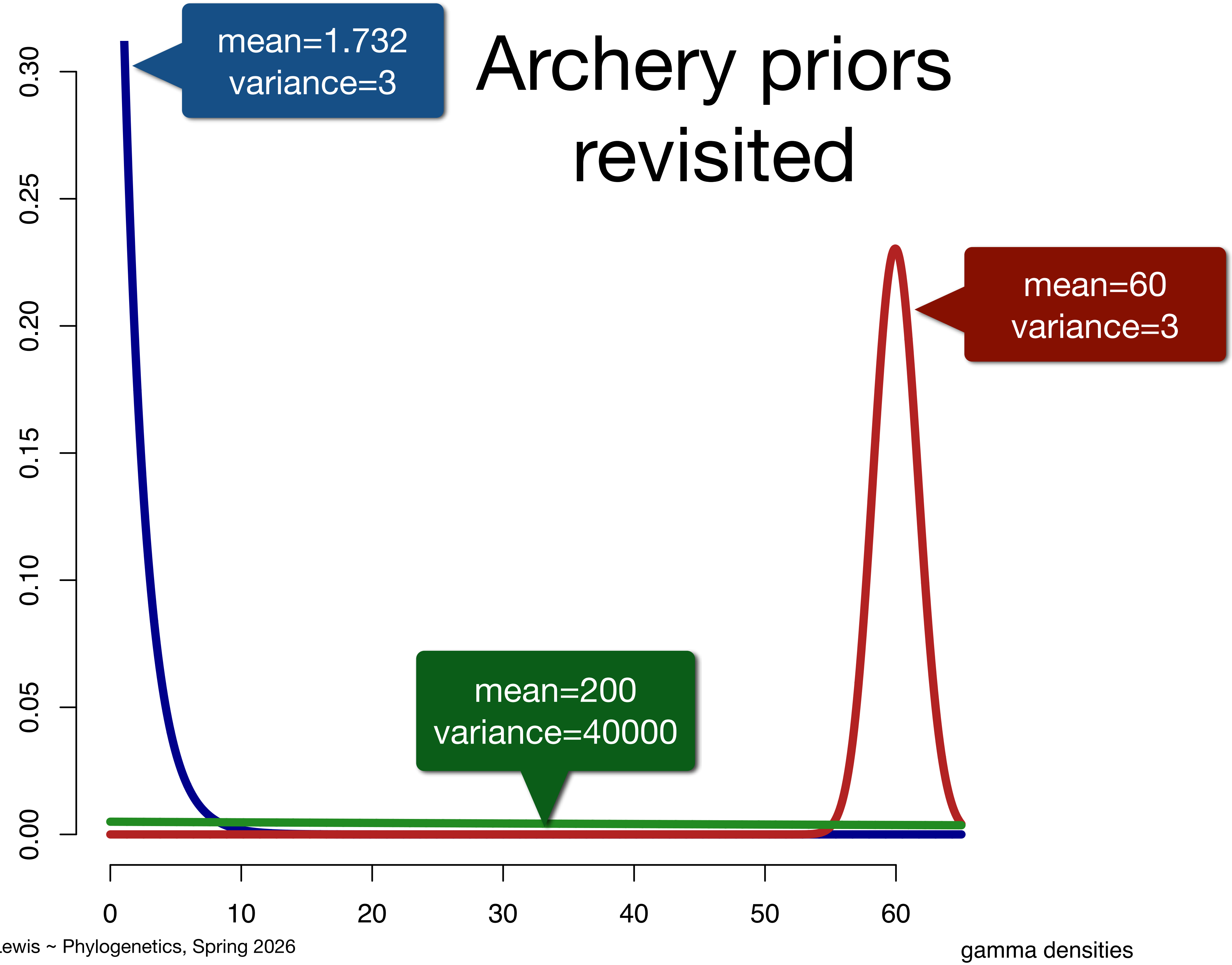
Integration of a probability density yields a probability (mass)



$$\int_0^2 f(\theta)d\theta$$

Area under the density curve from 0 to 2 is the probability that θ is less than 2

Archery priors revisited



Usually there are many parameters...

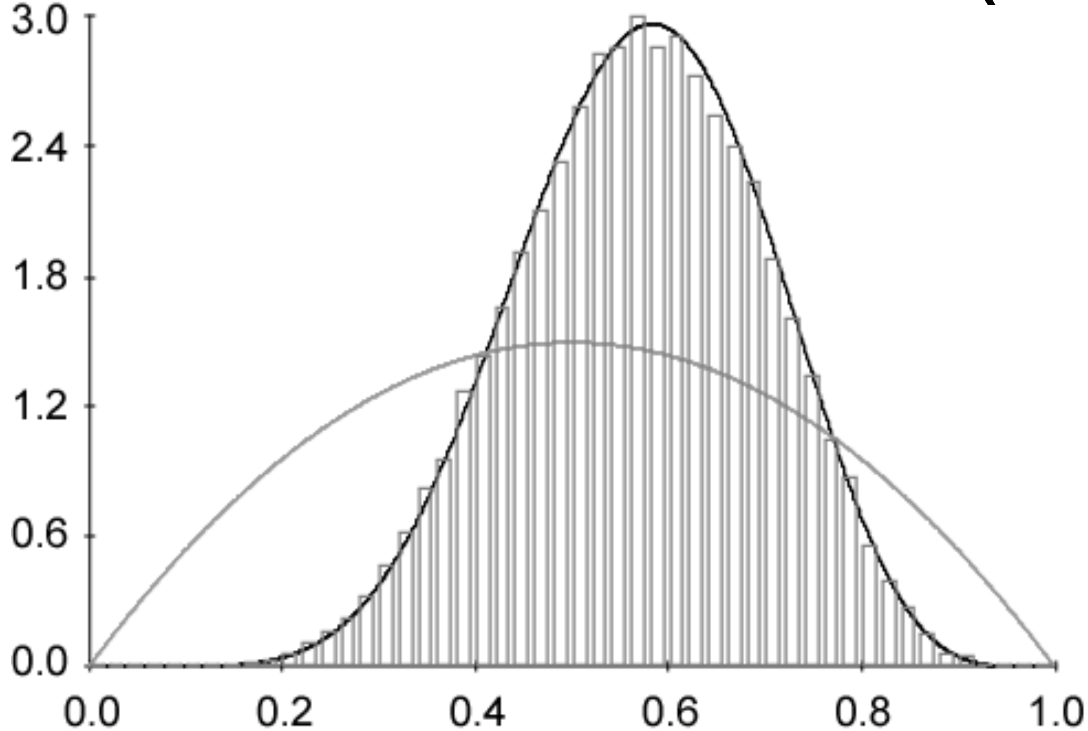
A 2-parameter example

$$f(\theta,\phi|D) = \frac{\overbrace{f(D|\theta,\phi)}^{\text{Likelihood}}\ \overbrace{f(\theta)f(\phi)}^{\text{Prior probability density}}}{\underbrace{\int_\theta \int_\phi f(D|\theta)\ f(\theta)f(\phi)\ d\theta d\phi}_{\text{Marginal probability of data}}}$$

Posterior probability density

An analysis of **100 sequences** under the simplest model (JC69) requires 197 branch length parameters. The denominator is a **197-fold integral** in this case! Now consider summing over **all possible tree topologies**! It would thus be nice to avoid having to calculate the marginal probability of the data...

Markov chain Monte Carlo (MCMC)

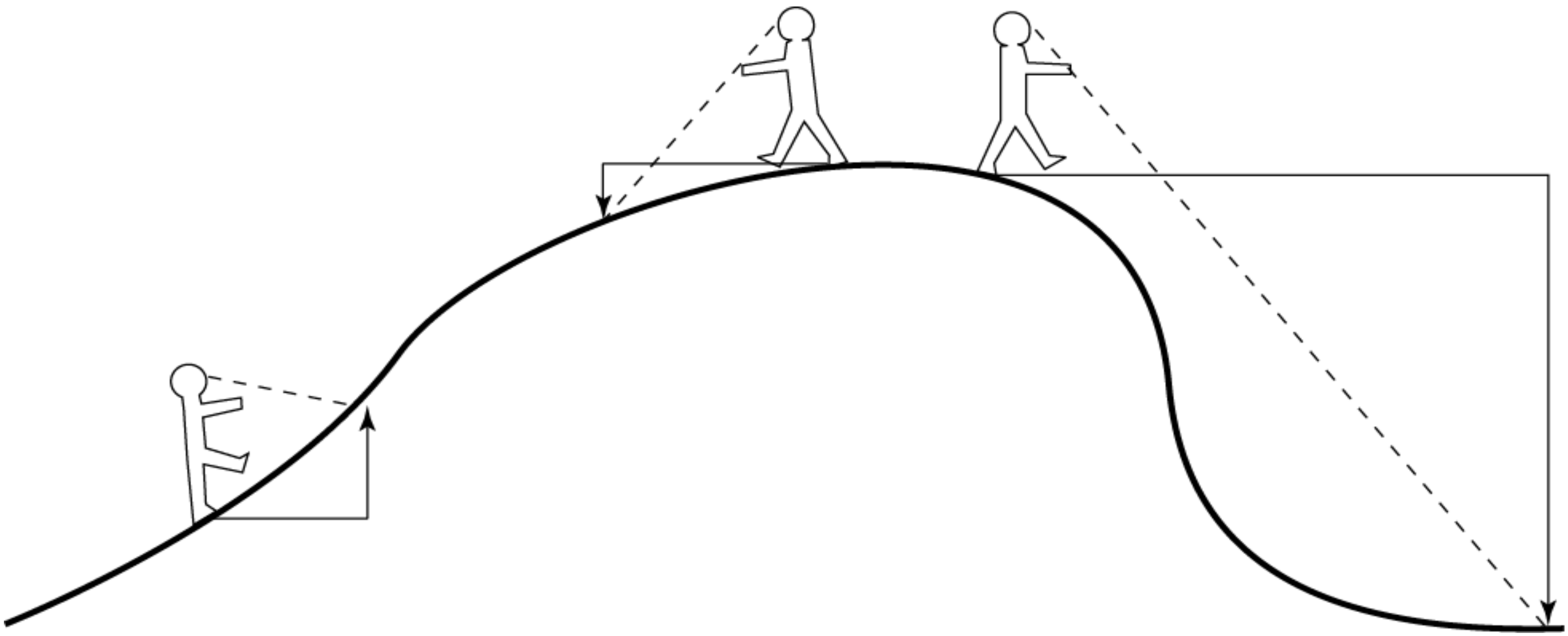


For more complex problems, we might settle for a

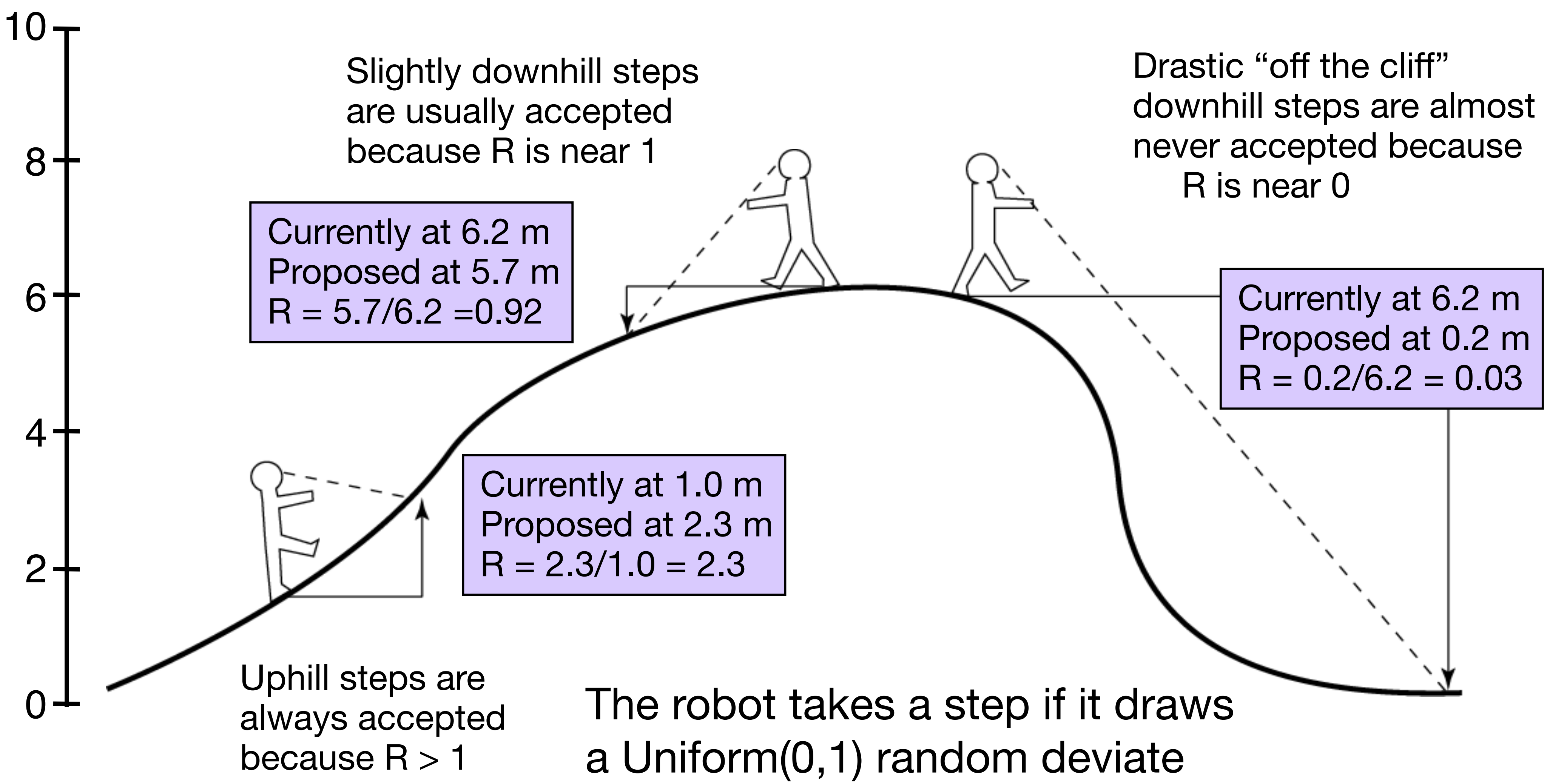
good approximation

to the posterior distribution

MCMC robot's rules



(Actual) MCMC robot rules

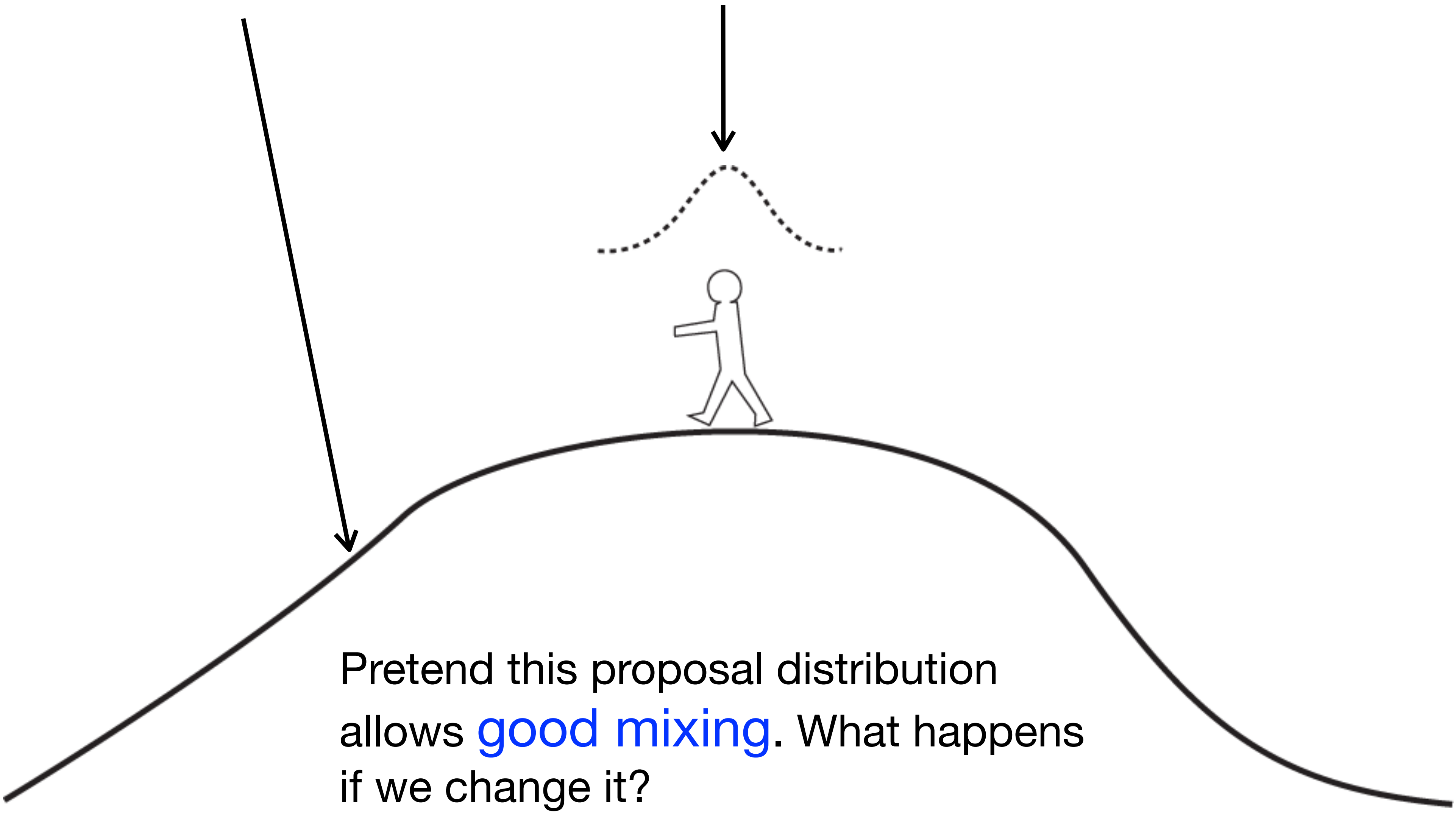


The robot takes a step if it draws a Uniform(0,1) random deviate that is less than or equal to R

Look! Ugly denominator gone!

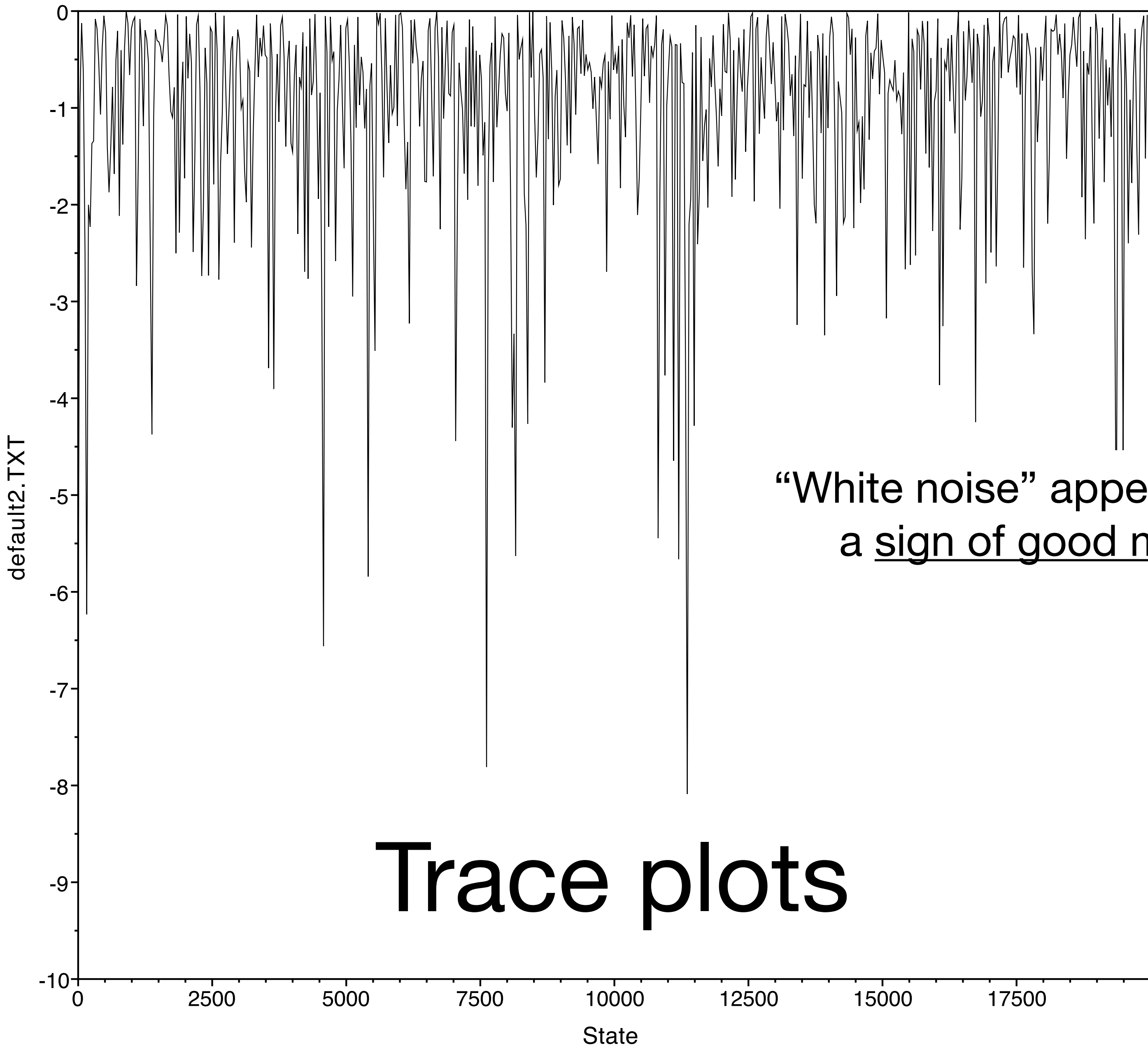
$$R = \frac{\Pr(B|D)}{\Pr(W|D)} = \frac{\frac{\Pr(B,D)}{\cancel{\Pr(B,D)+\Pr(W,D)}}}{\frac{\Pr(W,D)}{\cancel{\Pr(B,D)+\Pr(W,D)}}}$$

Target vs. Proposal Distributions



Pretend this proposal distribution allows good mixing. What happens if we change it?

Trace plots

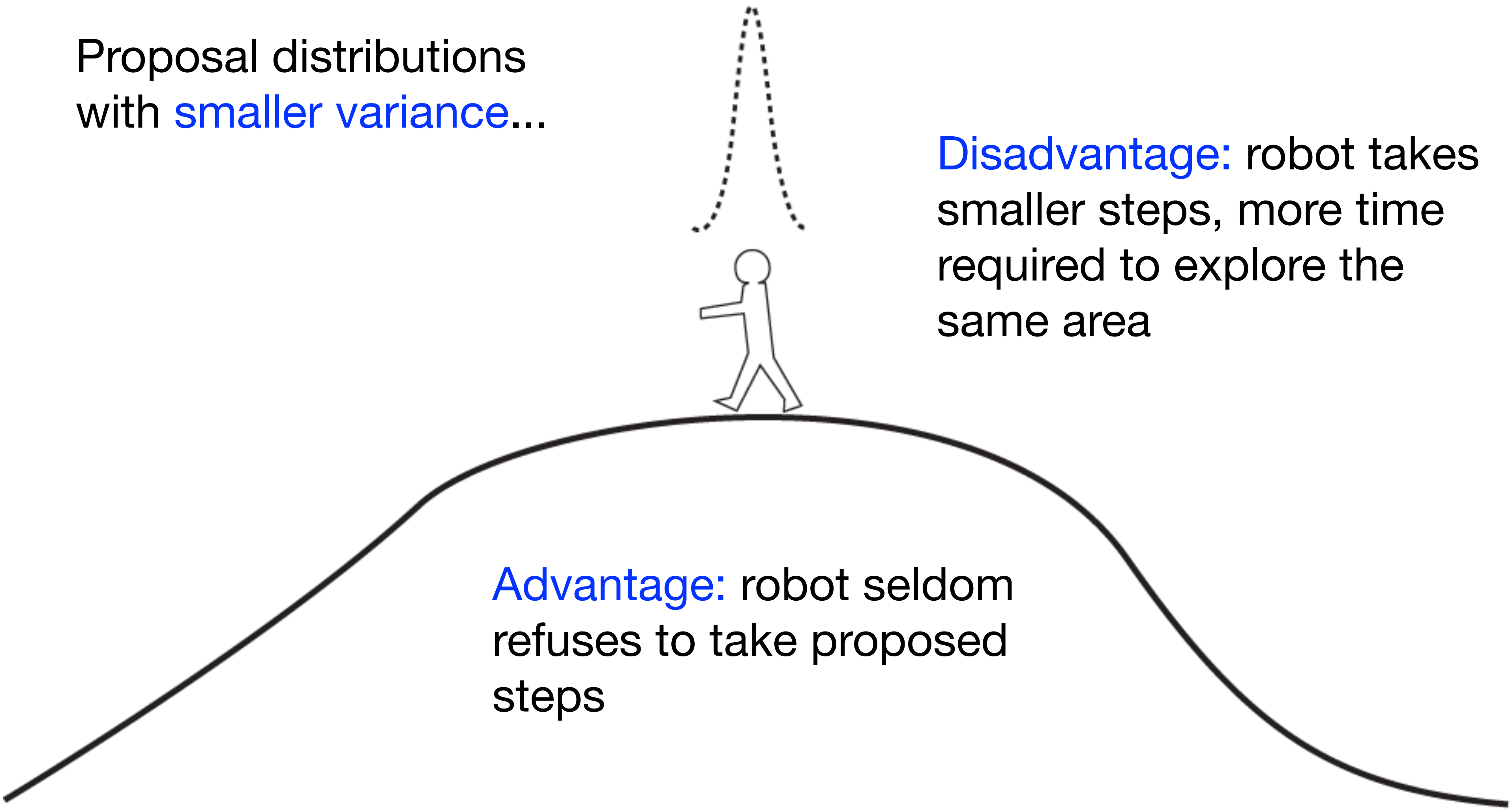


"White noise" appearance is a sign of good mixing

I used the program Tracer to create this plot:
http://tree.bio.ed.ac.uk/software/tracer/

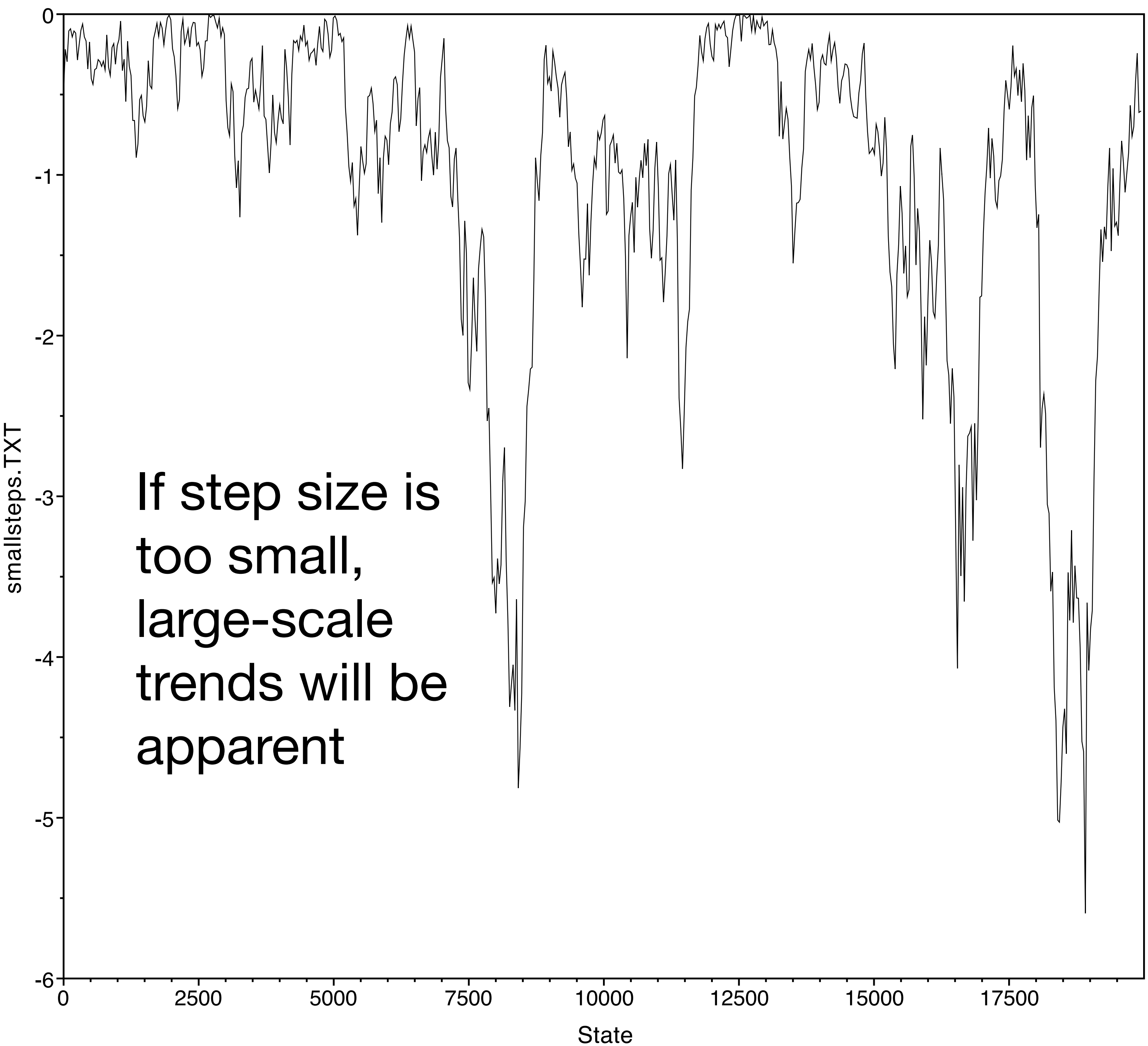
Target vs. Proposal Distributions

Proposal distributions with smaller variance...



Disadvantage: robot takes smaller steps, more time required to explore the same area

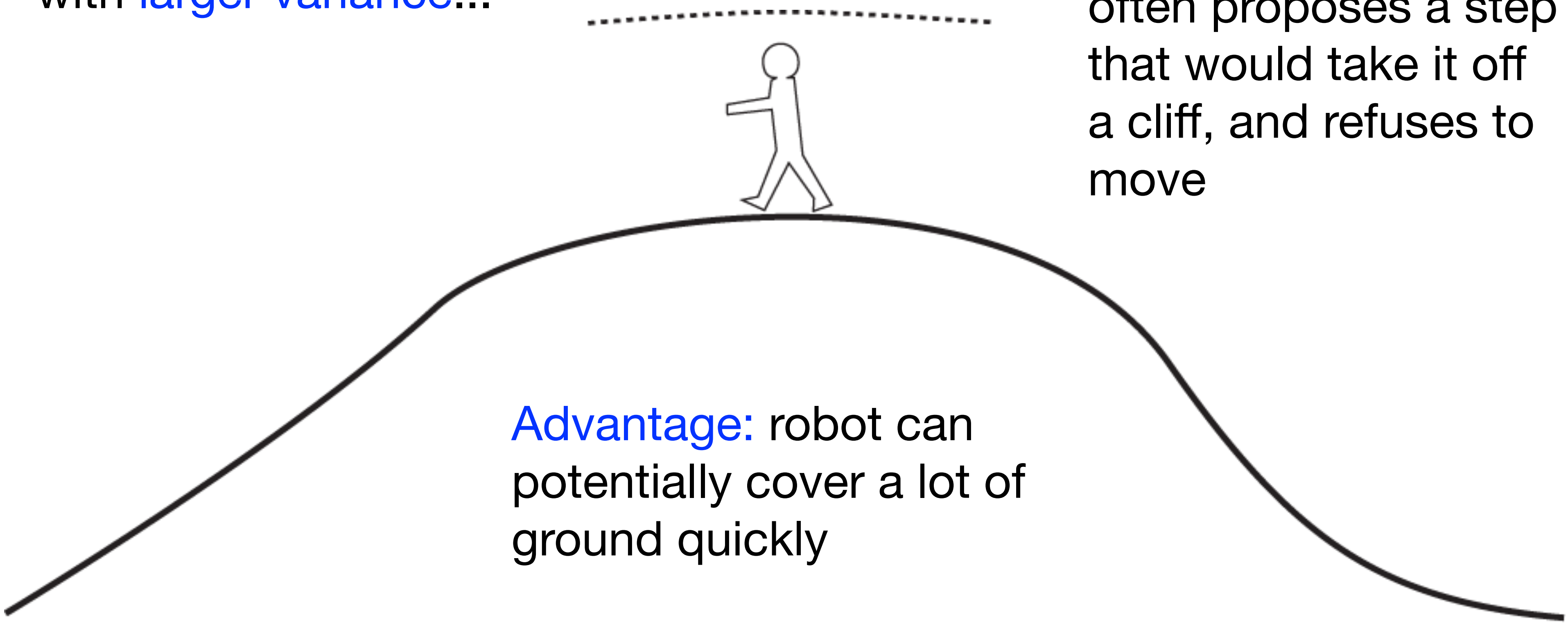
Advantage: robot seldom refuses to take proposed steps


0
-1
-2
-3
-4
-5
-6
smallsteps.TXT
0
2500
5000
7500
10000
12500
15000
17500
State
If step size is too small, large-scale trends will be apparent

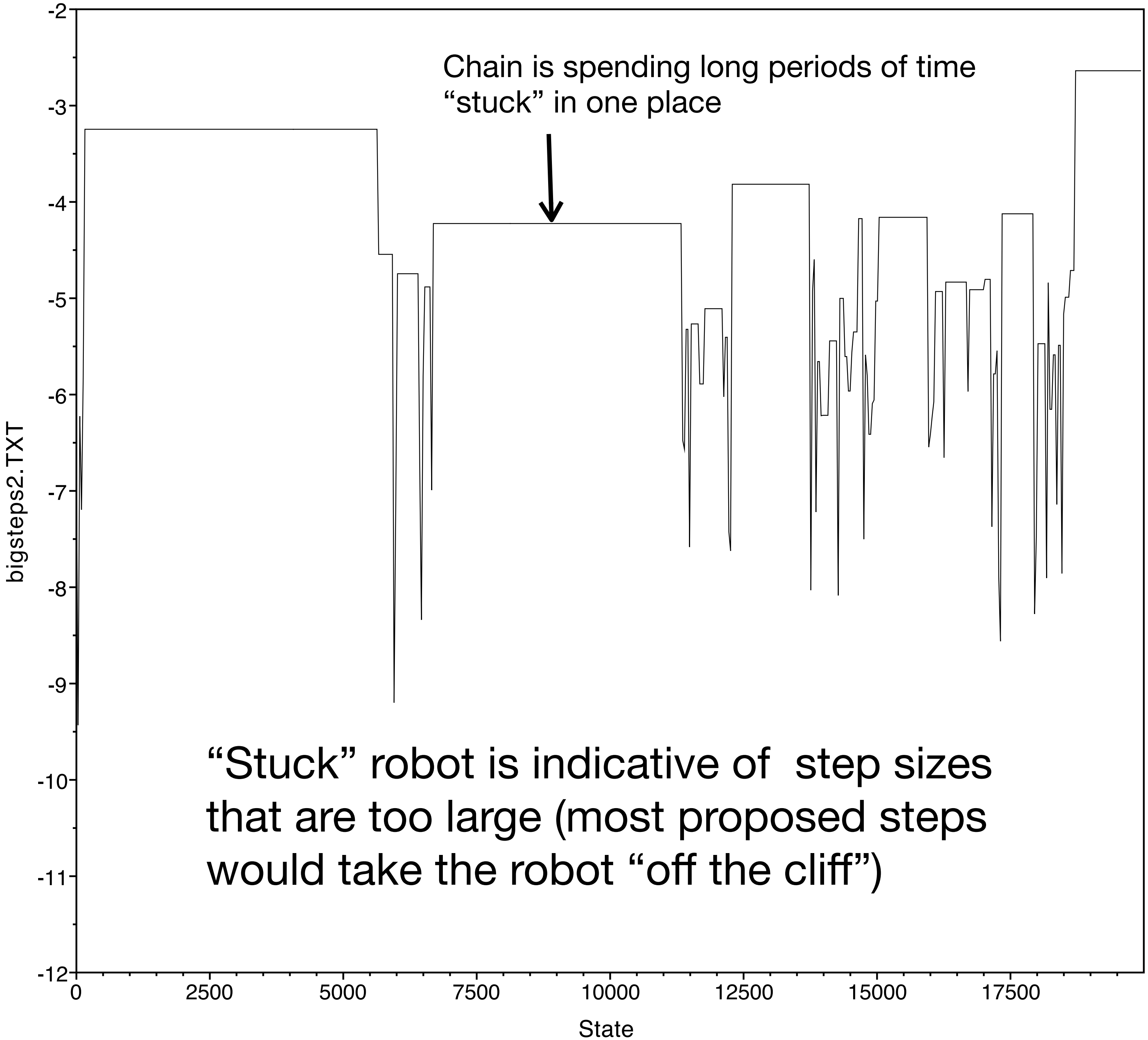
Target vs. Proposal Distributions

Proposal distributions with larger variance...

Disadvantage: robot often proposes a step that would take it off a cliff, and refuses to move



Advantage: robot can potentially cover a lot of ground quickly


Chain is spending long periods of time
“stuck” in one place
“Stuck” robot is indicative of step sizes
that are too large (most proposed steps
would take the robot “off the cliff”)
bigsteps2.TXT
State
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
-12
0
2500
5000
7500
10000
12500
15000
17500