

# Simulating Language

## Lecture 14: Iterated Bayesian Learning in populations

---

Simon Kirby

[simon@ling.ed.ac.uk](mailto:simon@ling.ed.ac.uk)



# A reminder of the Griffiths & Kalish result

---

- Given enough time, the end result of cultural evolution always reflects the prior bias and nothing else

Bottleneck does nothing

Noise does nothing

Details of language model do nothing

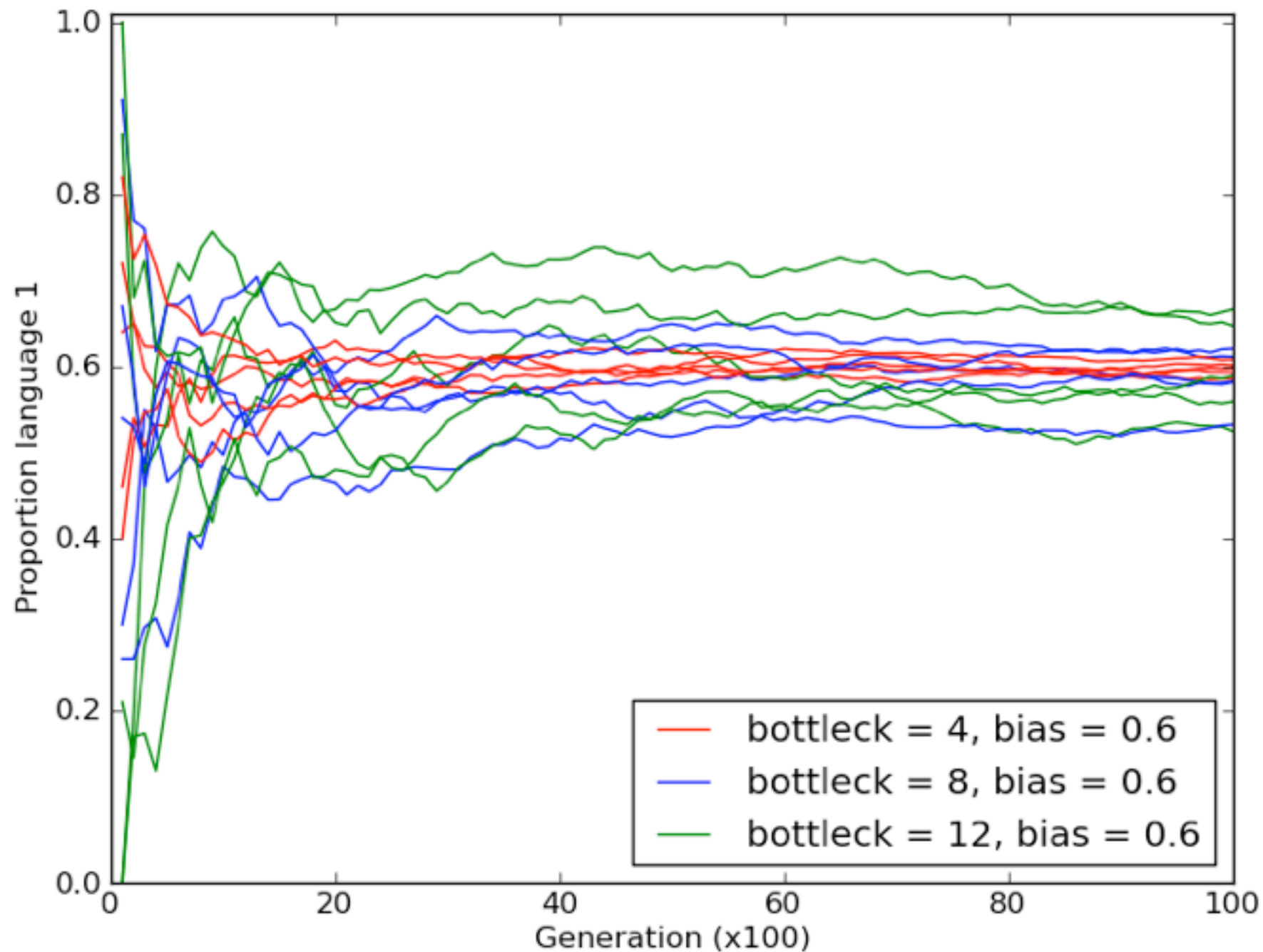
- If prior bias is innate, then this means that the universal properties of language are just a straightforward reflection of innateness
  - Contra all that stuff about culture doing interesting things

# An important detail: hypothesis selection

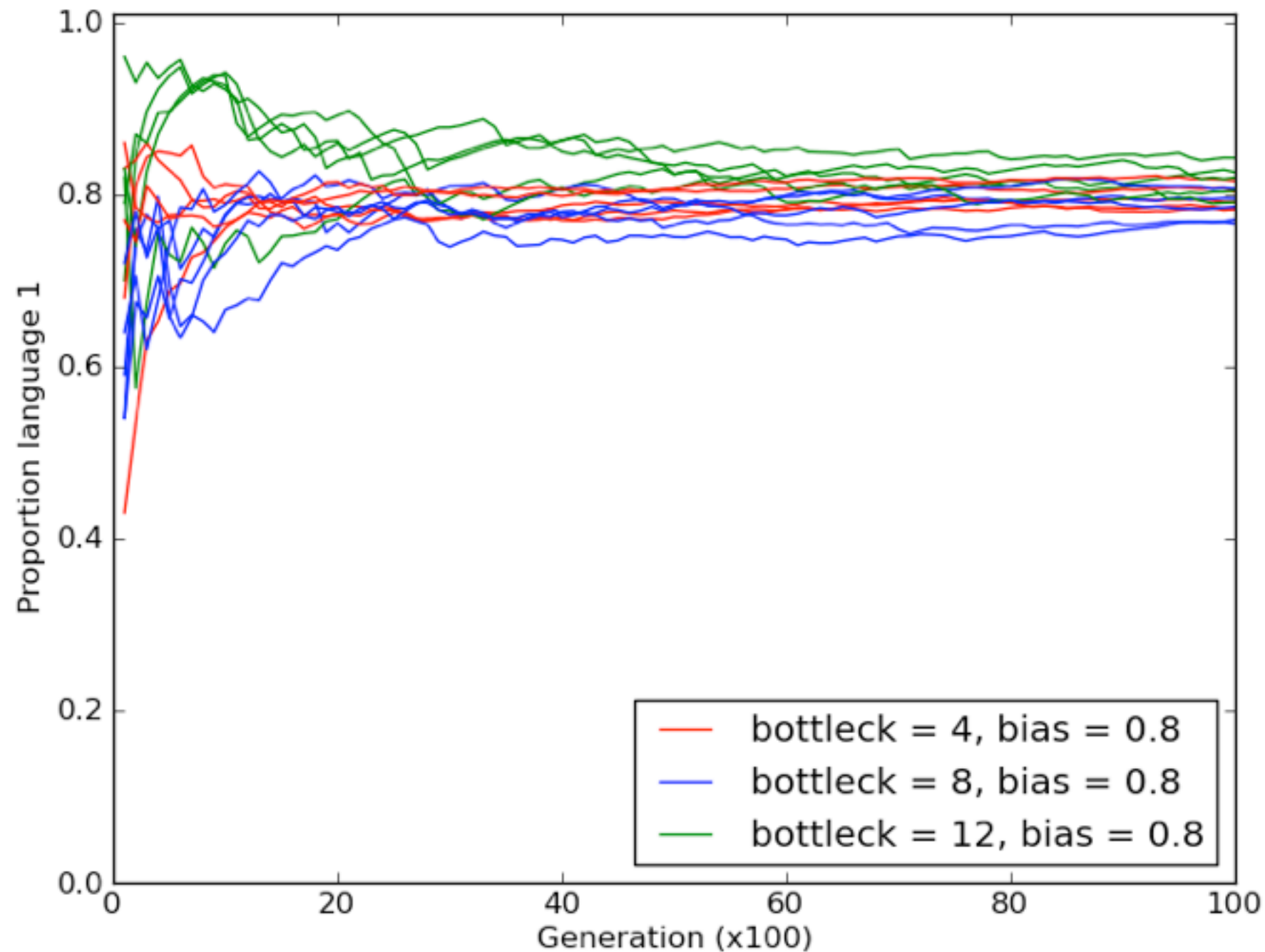
---

- How do you decide, given the posterior probabilities of various languages, which to select?
  - Sampling: given a particular distribution of probabilities, pick your hypothesis from the distribution proportionately.
  - MAP: given a particular distribution of probabilities, pick the best.
- Griffiths & Kalish's result as stated is for **samplers**.

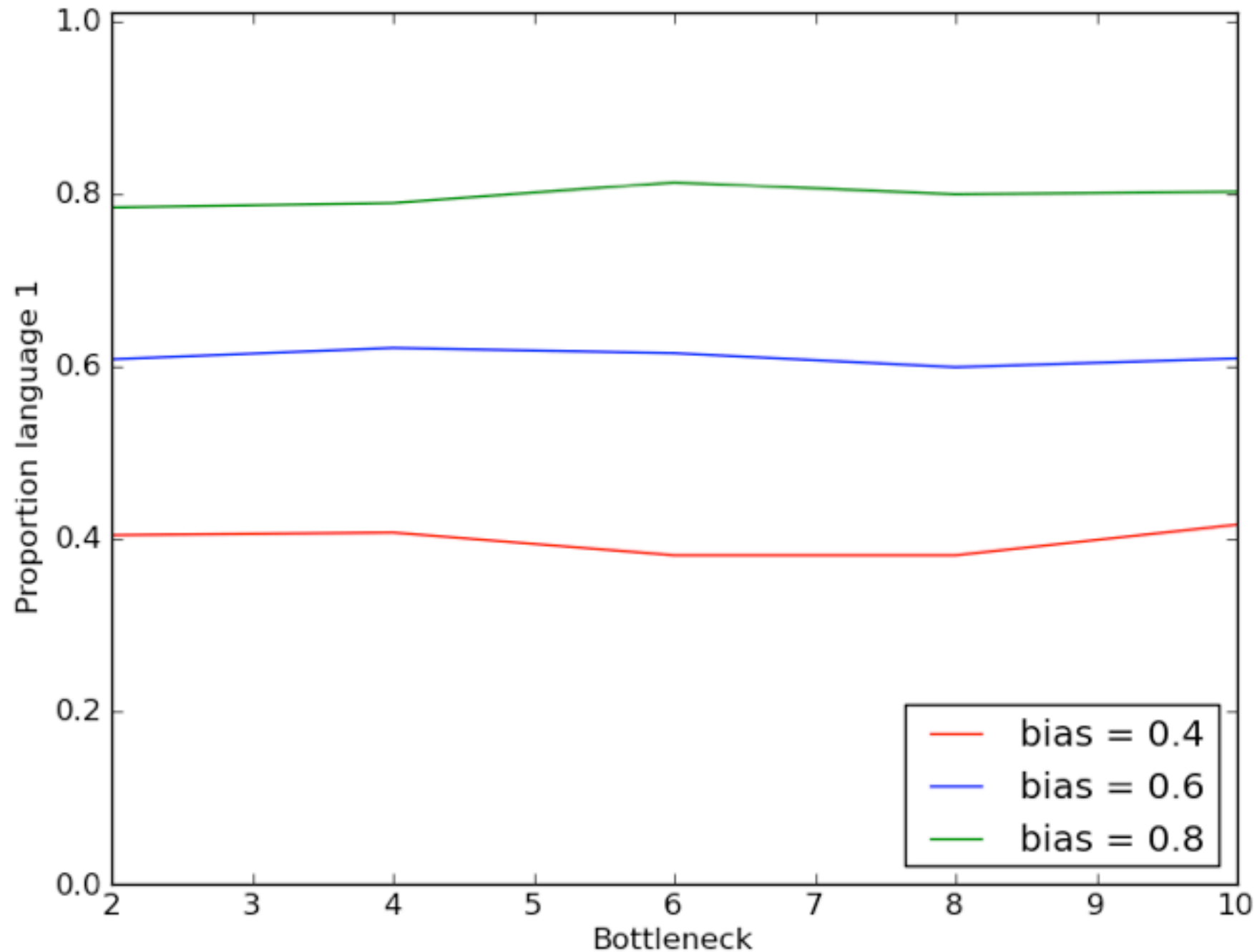
# Friday's lab: replicating the Griffiths & Kalish result for samplers



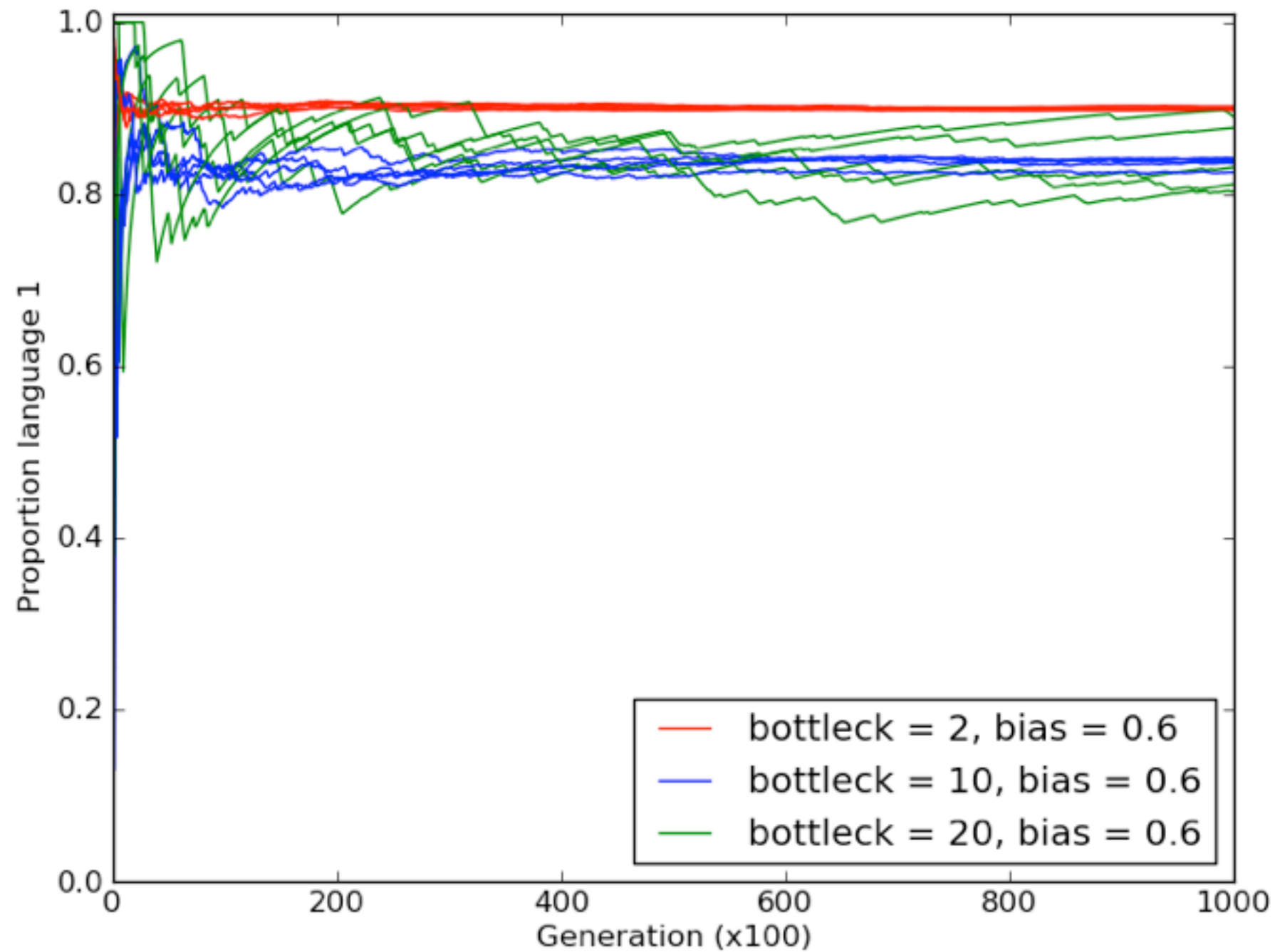
# Friday's lab: replicating the Griffiths & Kalish result for samplers



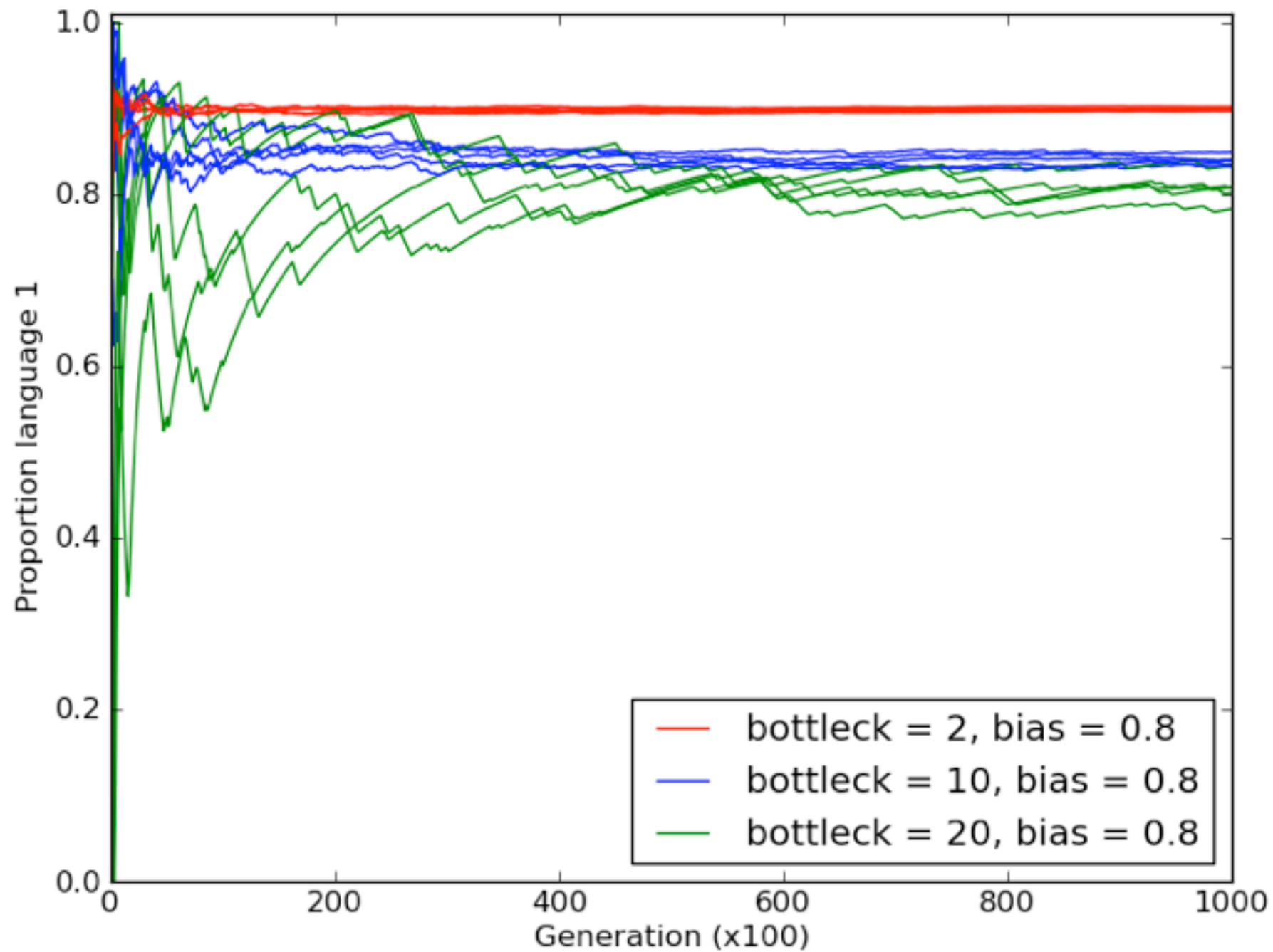
# Friday's lab: replicating the Griffiths & Kalish result for samplers



# Friday's lab: MAP learning

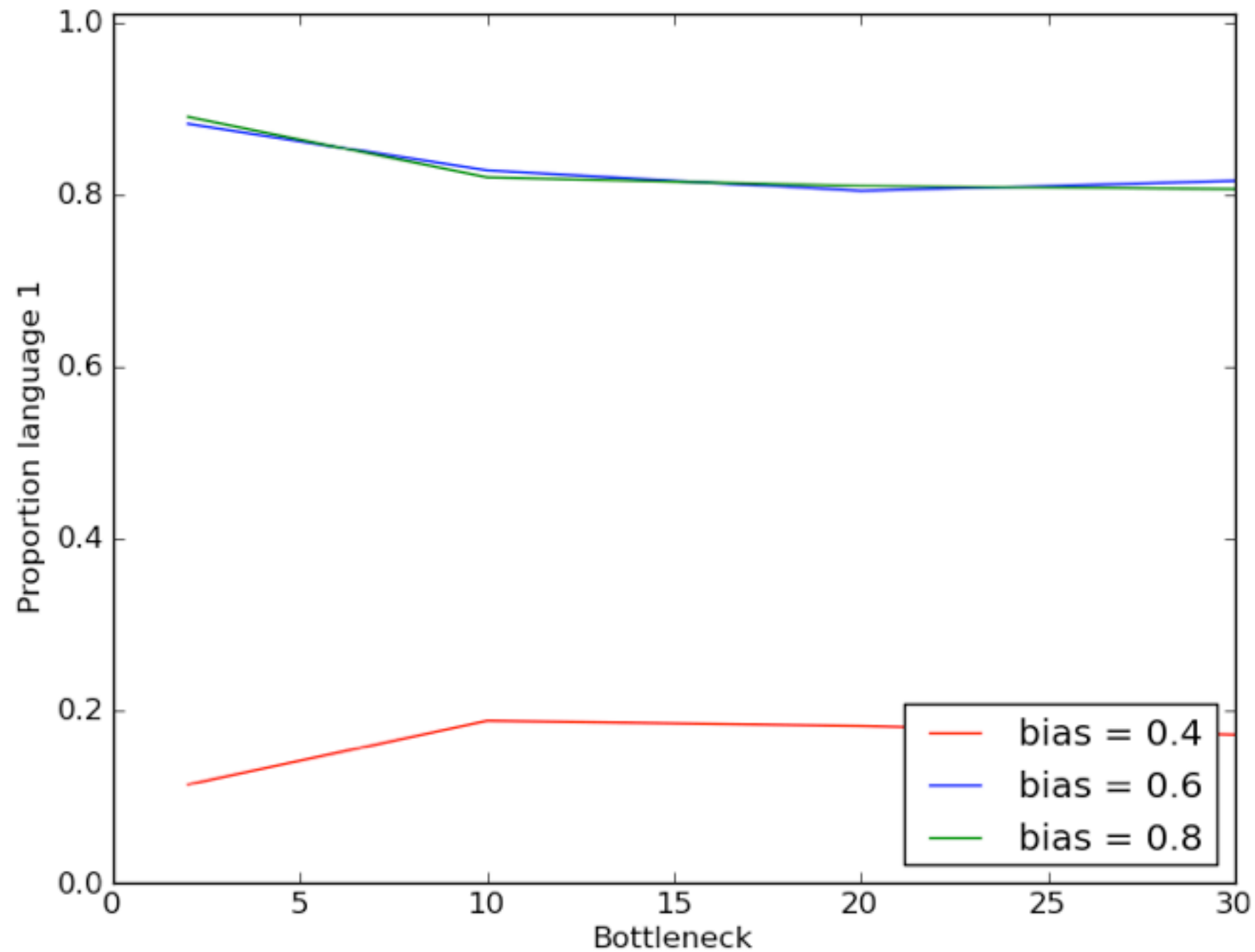


# Friday's lab: MAP learning





# Friday's lab: MAP learning



# A slightly weird feature of the two grammar model + MAP learning

---

- In general, for MAP learners the **strength** of their bias isn't important
  - Although the difference between 0.499999, 0.5 and 0.500001 does
- But **in the two grammar model**, bias can sometimes be entirely irrelevant

[0, 0]	learn 0
[0, 1]	prior chooses
[1, 1]	learn 1

[0, 0, 0]	learn 0
[0, 0, 1]	learn 0
[0, 1, 1]	learn 1
[1, 1, 1]	learn 1

- This is a bit untidy, but solely a (slightly odd) feature of this language model

# Sampling vs MAP: summary so far

---

- Iterated Bayesian Learning allows us to more precisely understand the relationship between learning bias and eventual language structure
- If you assume social learning is about maximising the chance of converging on what other people are doing (i.e. selecting the MAP hypothesis), then cultural evolution does a lot of work for you
- Very weak innate biases are all that's needed to explain strong linguistic universals
  - If people are MAP learners
- If we see universals in language, then we should not assume that these are hard-coded as strong constraints in the genes
  - If people are MAP learners

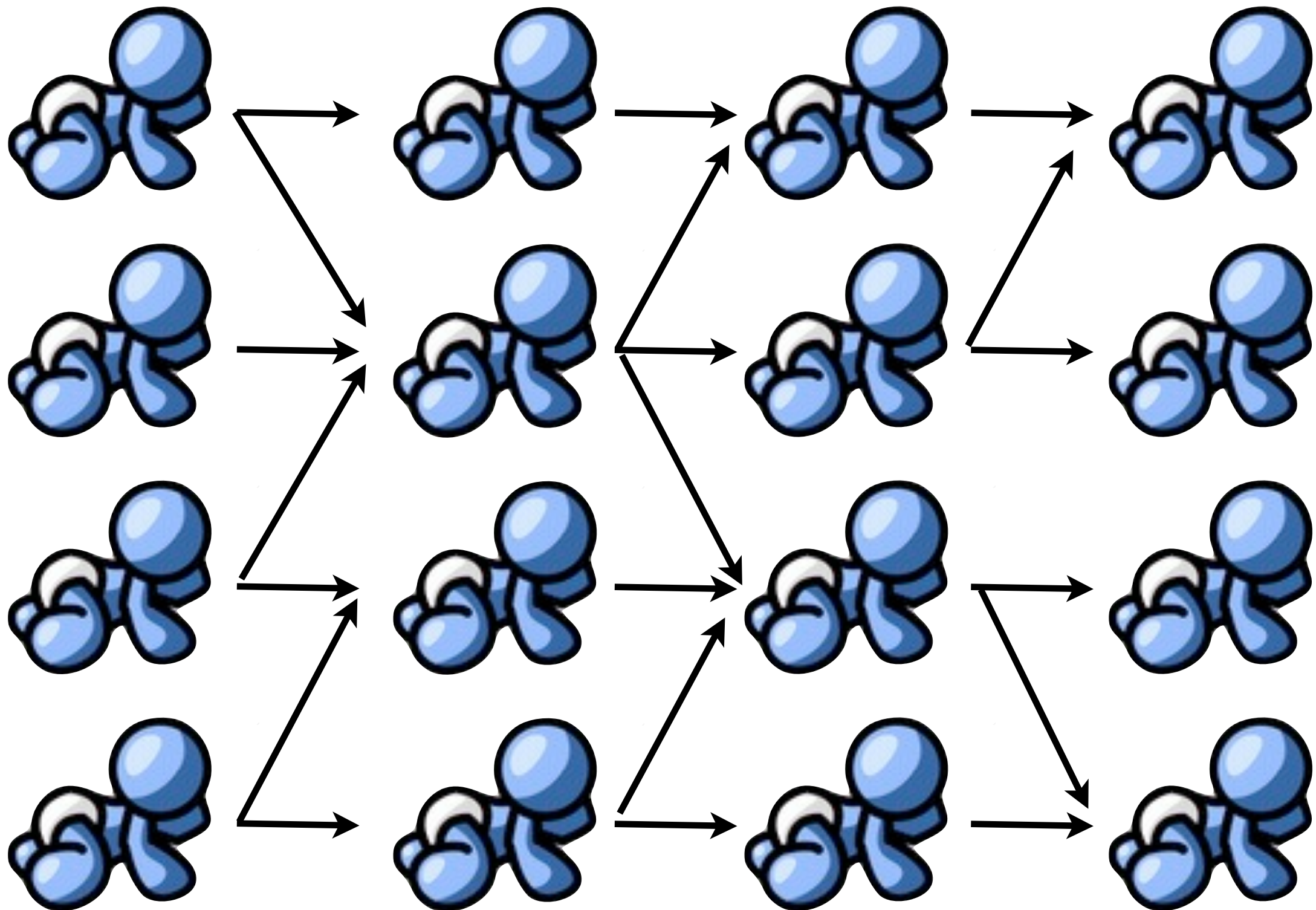
# It's really important we get this right!

---

- If language learning is like sampling, language universals probably closely reflect learner biases. If it's like MAP, they don't.
- How can we tell which is right?
  - Run experiments on real people to see if they behave like they are sampling or selecting the MAP language
  - Maybe evolution will favour one alternative over the other?
    - See final lecture
  - Maybe one of these results is an unrepresentative special case
    - For instance: what happens if we go beyond long skinny diffusion chains and look at transmission in populations?
    - Smith (2009), Burkett & Griffiths (2010)

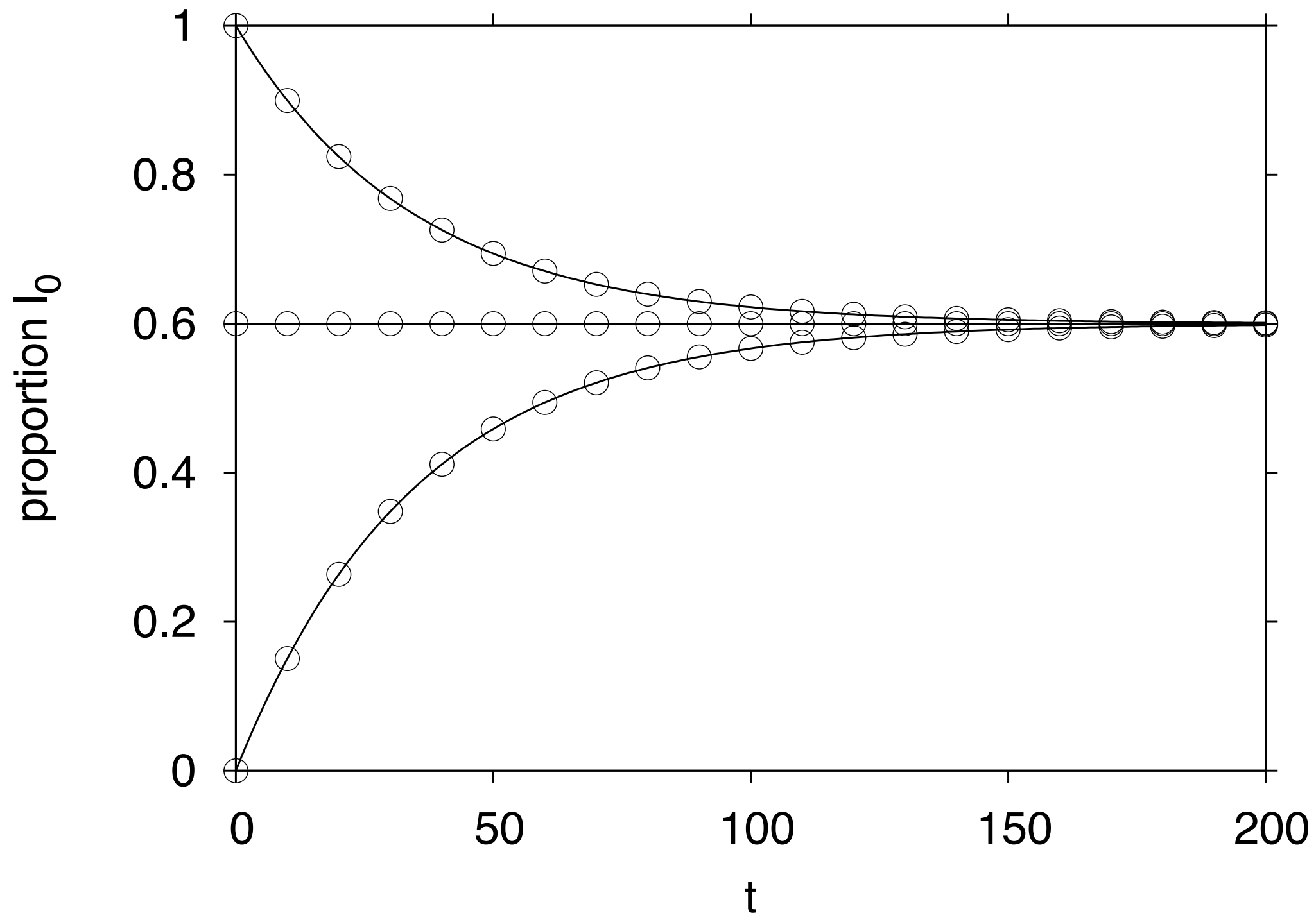
# Moving to populations

---



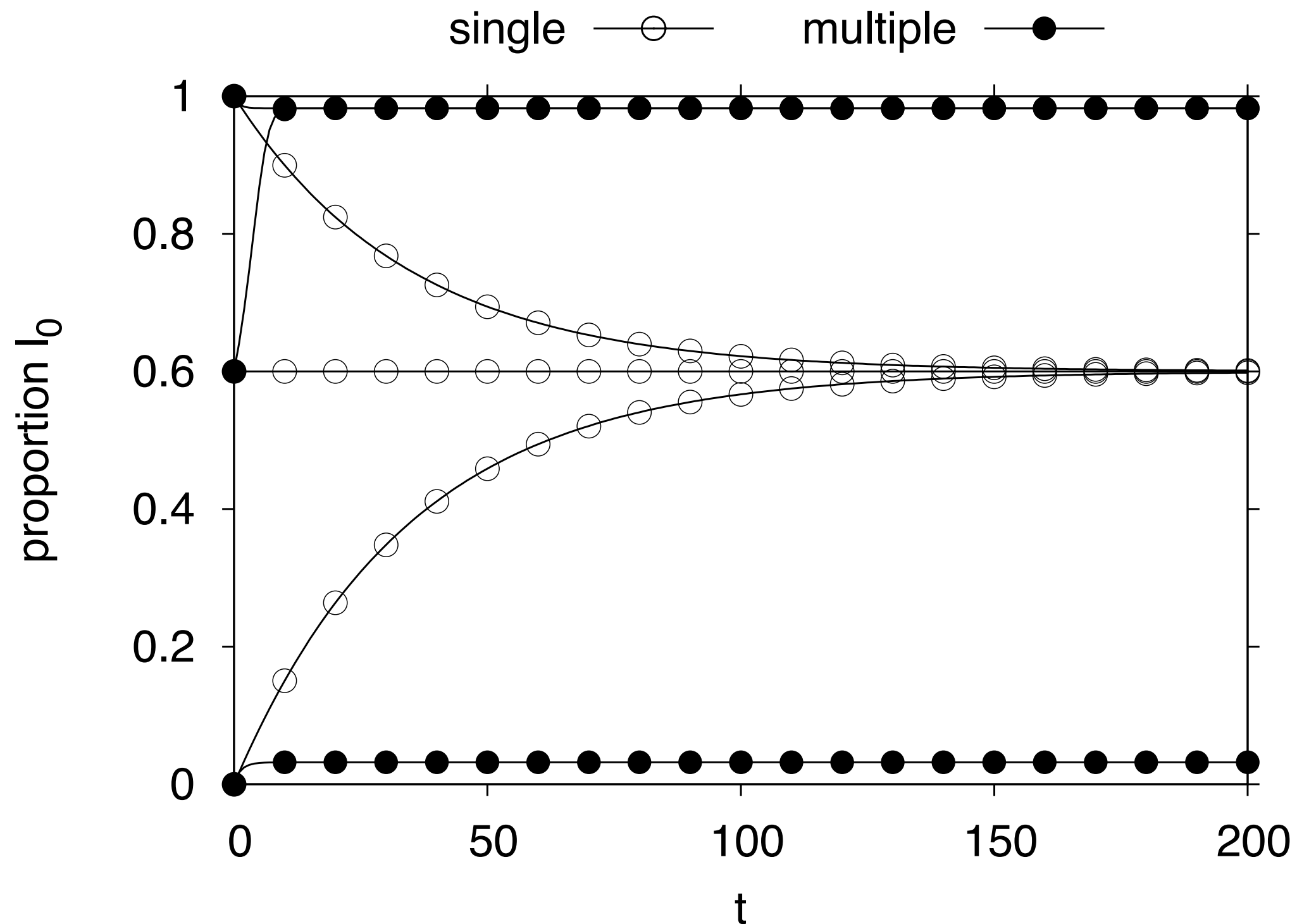
Samplers, everyone learns from one teacher (bias for  $L_0 = 0.6$ )

---

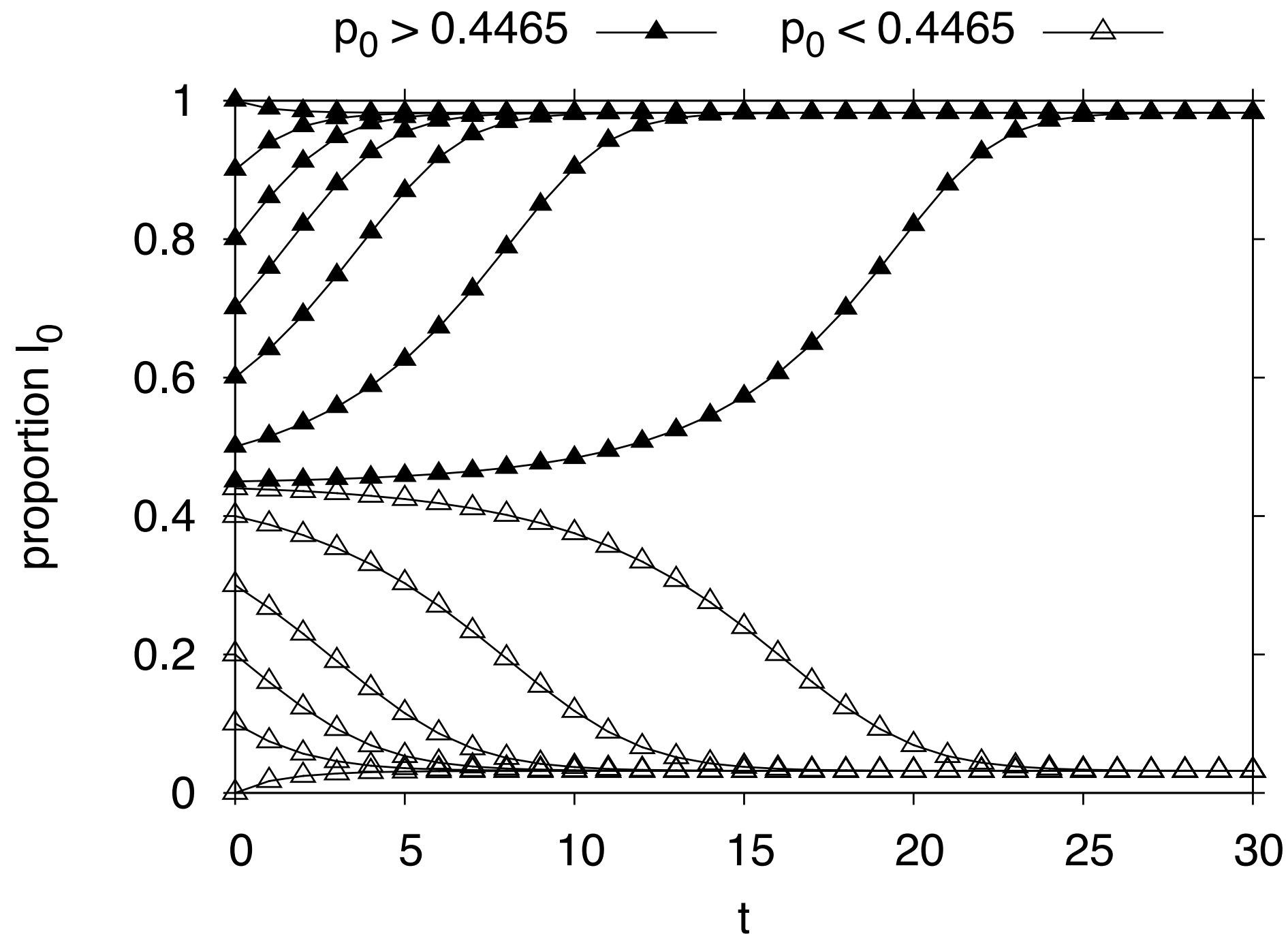


# Samplers, everyone learns from multiple teachers (bias = 0.6)

---



# Samplers, everyone learns from multiple teachers (bias = 0.6)





# Sampler populations look like MAP populations!

---

- In populations, when samplers learn from multiple teachers:

No convergence to the prior

Amplification of weak biases

Bottleneck effects

...

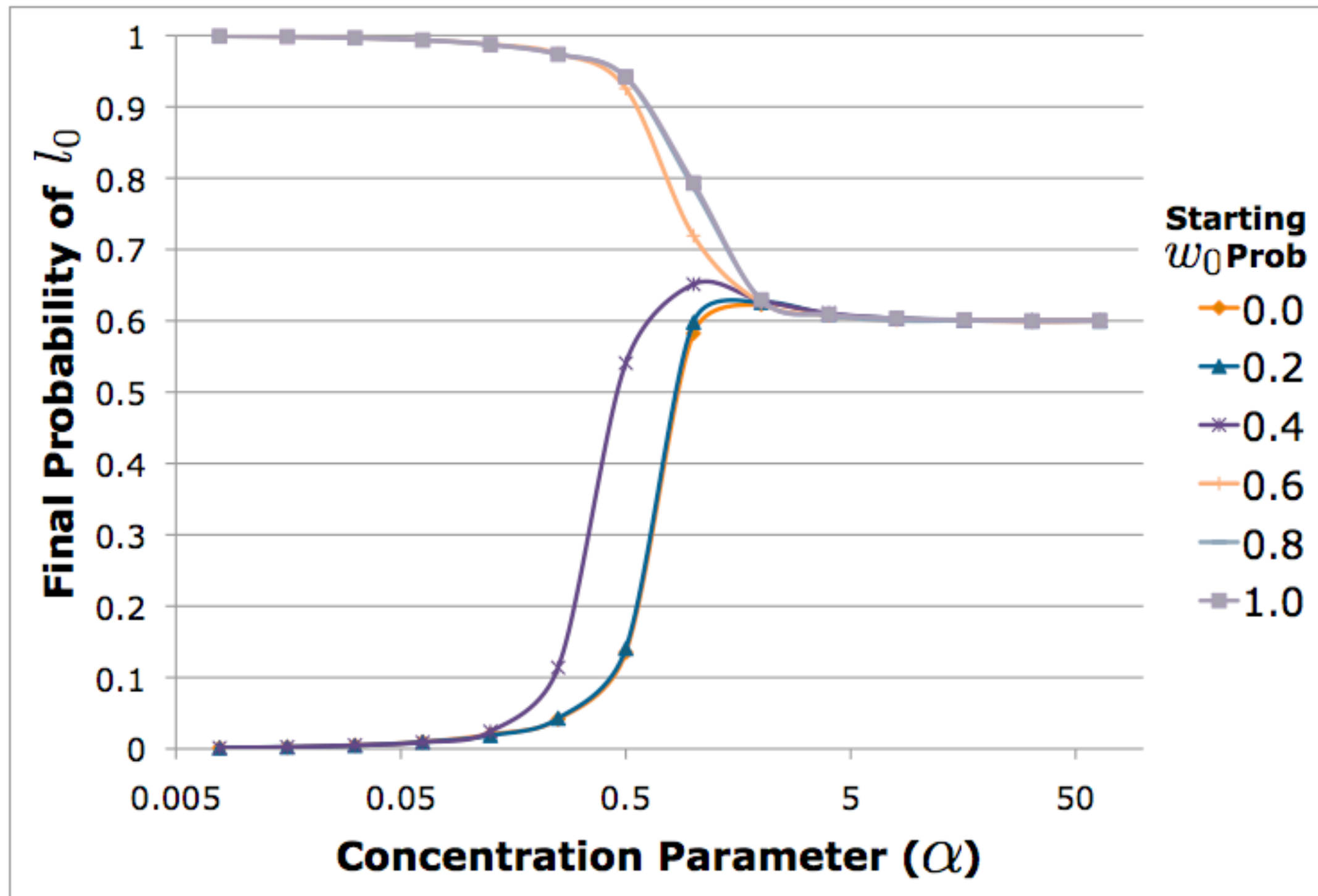
- In this context, Bayesian learning is **conformist**
  - Disproportionately likely to learn the more common language
  - Known result of conformist learning: convergence on single language

# Learning one language versus learning multiple languages?

---

- That's based on the assumption that learners try to find a single grammar to account for their data
  - Even if it was generated by multiple people
- Burkett & Griffiths (2010): we can just add this as a parameter of the model
  - Low  $\alpha$ : learners tend to learn a single language
  - High  $\alpha$ : learners learn multiple languages

# Burkett & Griffiths' result



# Summary

---

- An active area of ongoing research
- My hunch is that the Griffiths & Kalish sampling result will turn out to be a special case
  - We should not expect to see a straightforward relationship between language universals and learner bias
- But in either case, Iterated Bayesian Learning has been key to clarifying our understanding of what cultural evolution might be like

# Up next

---

- Monday: last lab
  - More Bayesian stuff
- Thursday: putting it all together
  - Learning, culture, biological evolution
  - Evolution of the language faculty?