Human simulation

Kenny Smith
University of Edinburgh
kenny@ling.ed.ac.uk
@kennysmithed

The need for human simulation

- How do we know our models are relevant to human language?
  - "*Powerful and potentially interesting although this approach is, its failure to incorporate more realistic conditions (perhaps because these would be more difficult to simulate) sharply reduces any contribution it might make toward unraveling language evolution. So far, it is a classic case of looking for your car-keys where the street-lamps are*" Bickerton (2007) p. 522
- Can we show the same processes at play in real language learners?

You have already seen some human simulation on this course

An artificial language learning study

Hudson-Kam & Newport (2005)

- Adults trained and tested on an artificial language
  - 36 nouns, 12 verbs, negation, **2 determin**
- Multiple training sessions
- Variable (unpredictable) use of 'determiners'

flern
rams
"the e

Greenbergian Universals and Bayesian inference

Jenny Culbertson

Simulating Language, 11 March, 2015

Why do experimentalists need models?

- In our models, we know exactly what we put in
  - We wrote the code
- For experiments with humans, we don't
  - We don't know how humans work
  - We know that any human we can get into the lab will have extensive linguistic experience

The ideal

- Simulation models to build nice, clean accounts of the mechanisms involved
- Human simulation to test whether the assumptions of the models, or the results, apply to ‘real people’

This talk: an example

Language structure as a consequence of cultural evolution in response to two pressures

- Learning
  - Favours simplicity/compression
- Use
  - Favours complexity/expressivity

Collaborators

Simon Kirby, Monica Tamariz
University of Edinburgh

Hannah Cornish
University of Stirling

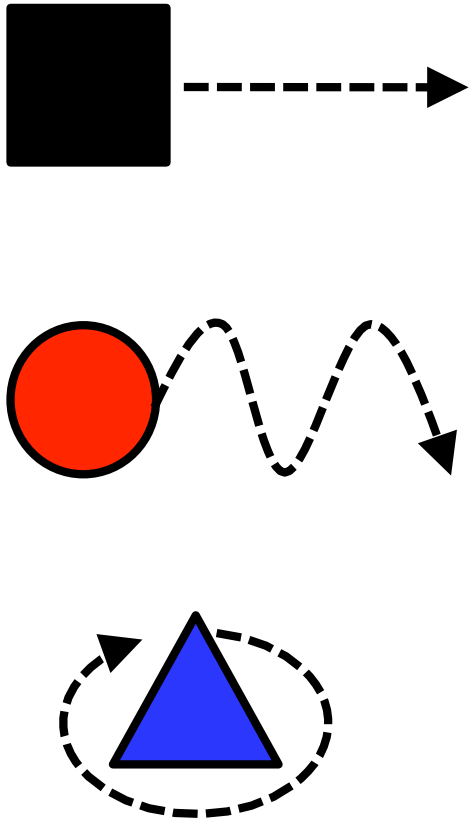
ESRC for funding Experiment 2

Experiment 1

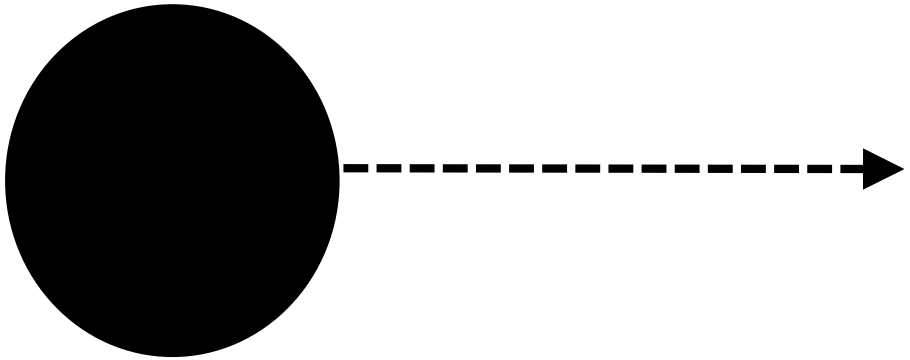
Kirby, Cornish & Smith (2008)

An artificial language learning experiment

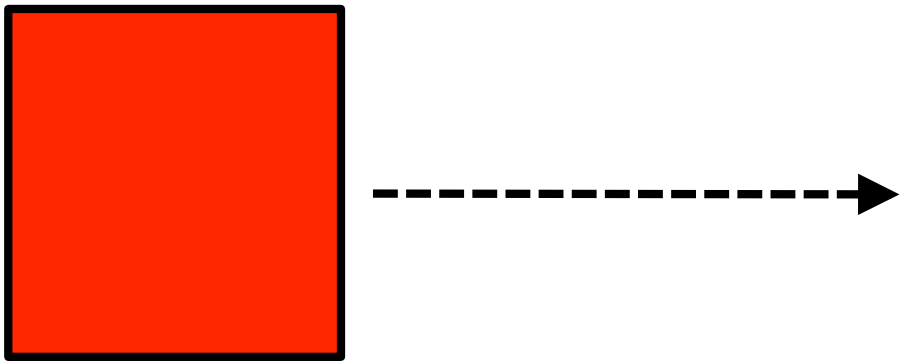
- 27 pictures
  - 3 colours, 3 shapes, 3 movements
- Each picture associated with a label
  - Typed text
- Adult participant repeatedly trained on set of picture-label pairs
  - Presented with picture paired with label
- Then tested
  - Presented with picture only, enter label
  - No feedback, no communicative task
- Based on model from Kirby (2002)



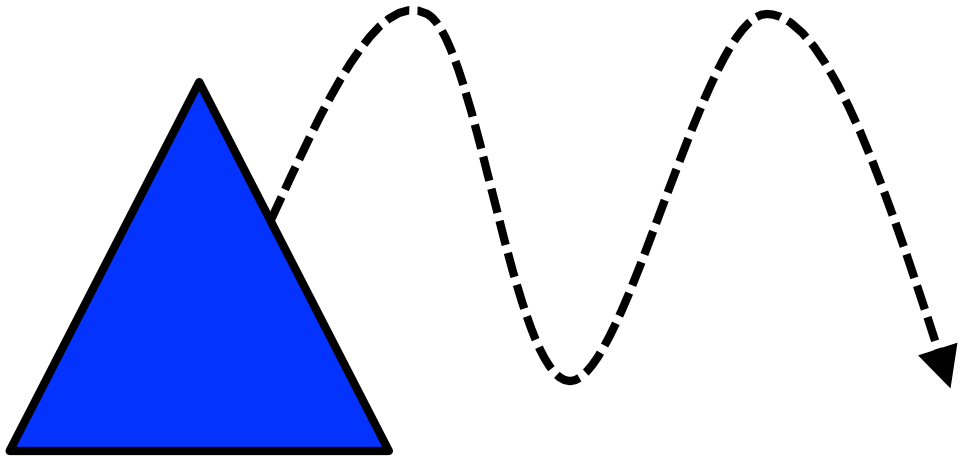
nihepi

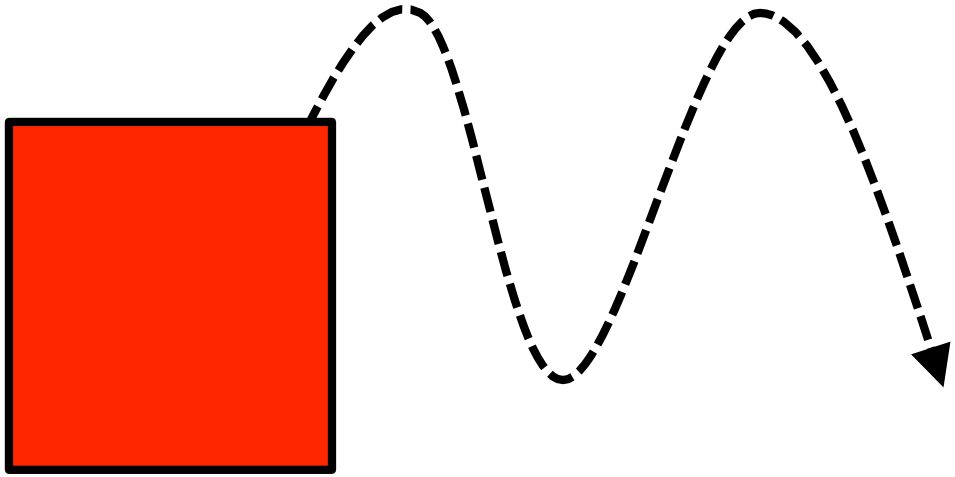


gepini



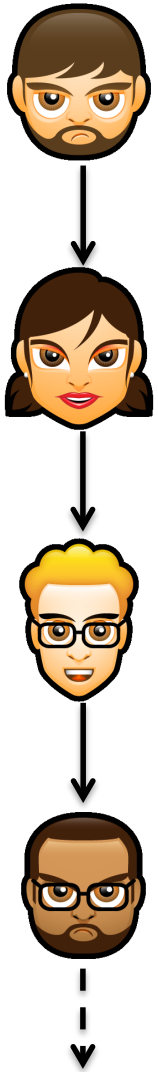
wige



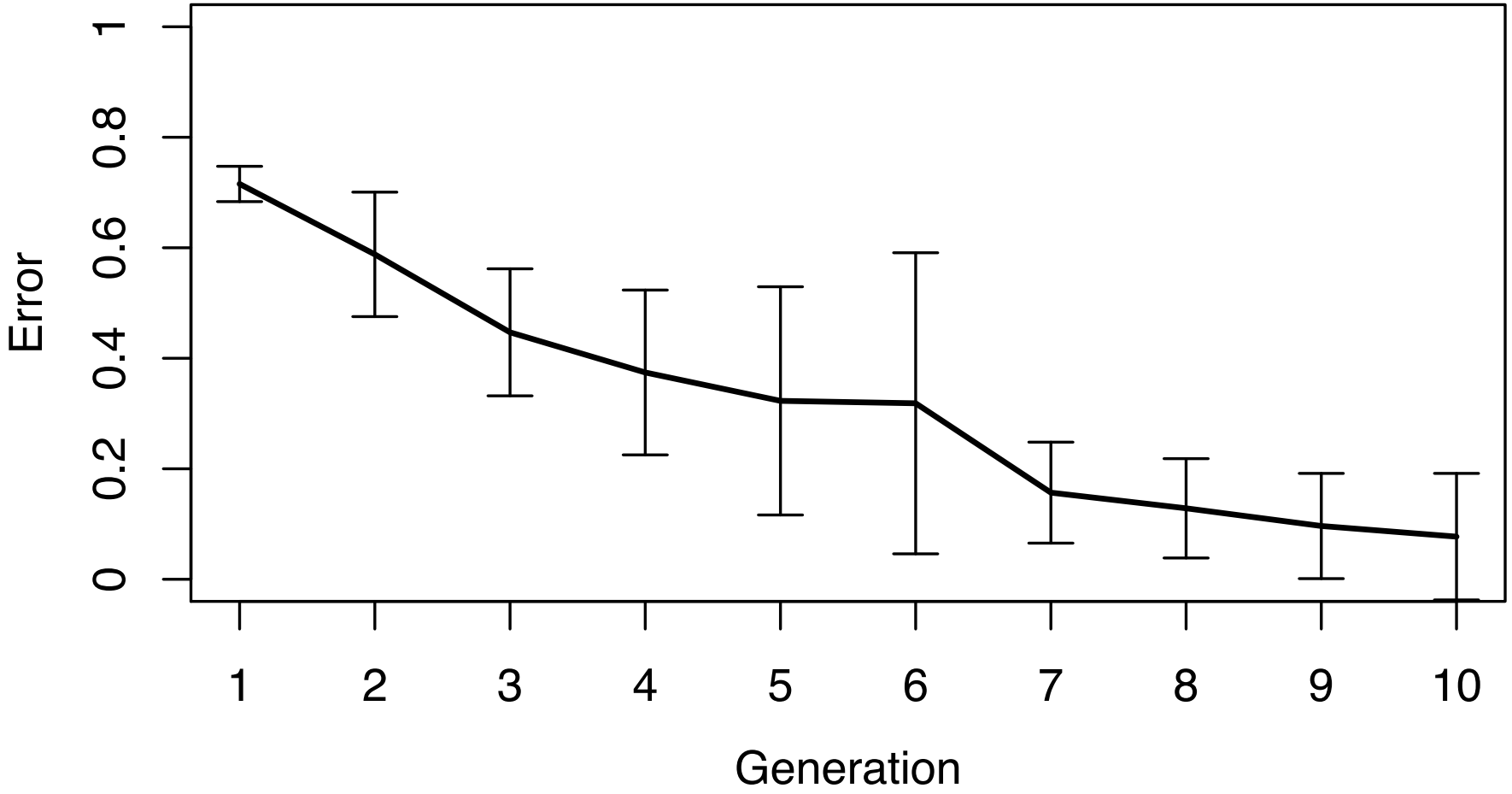


pon

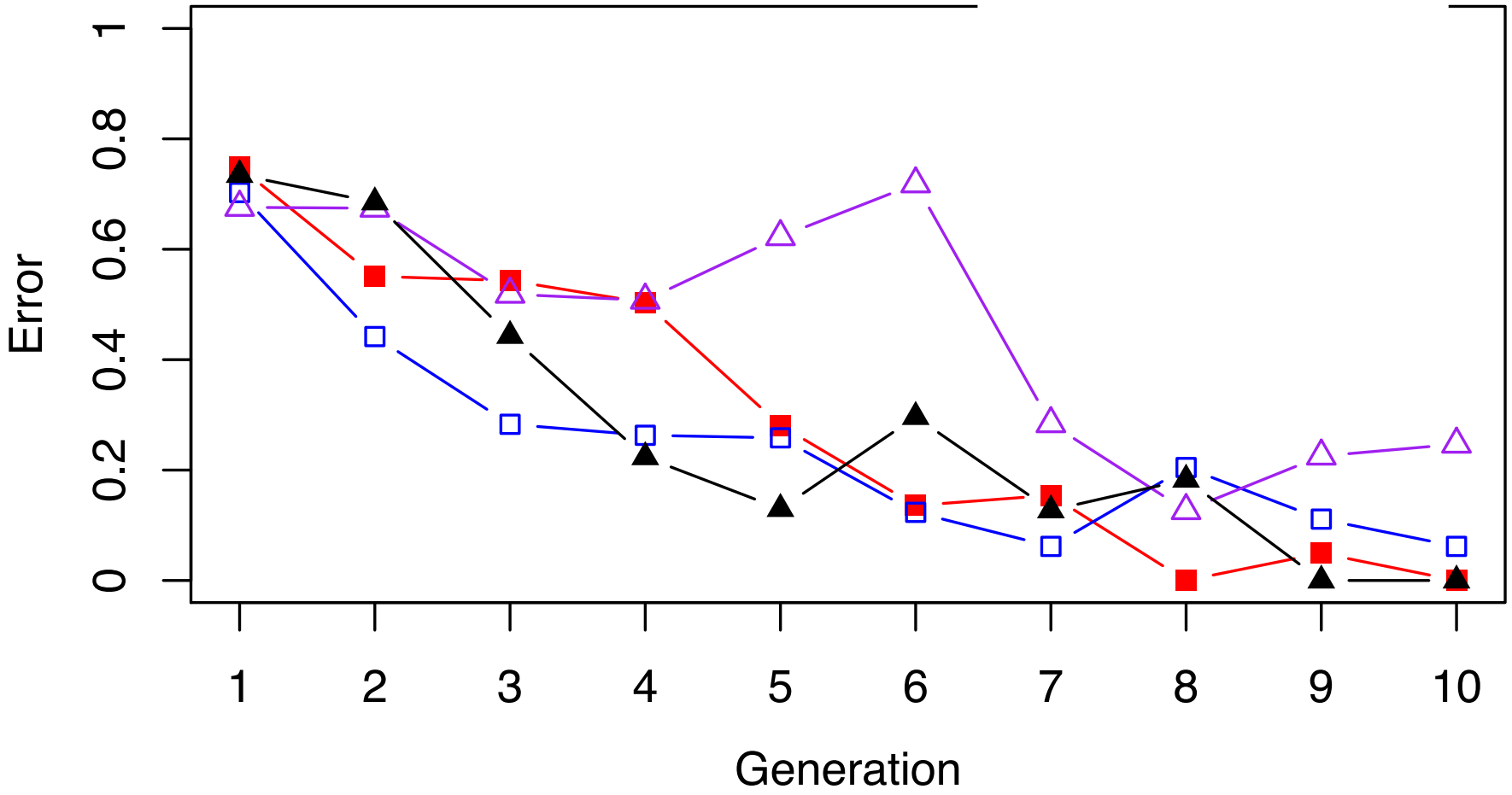
Iterated Learning



Results: decrease in error

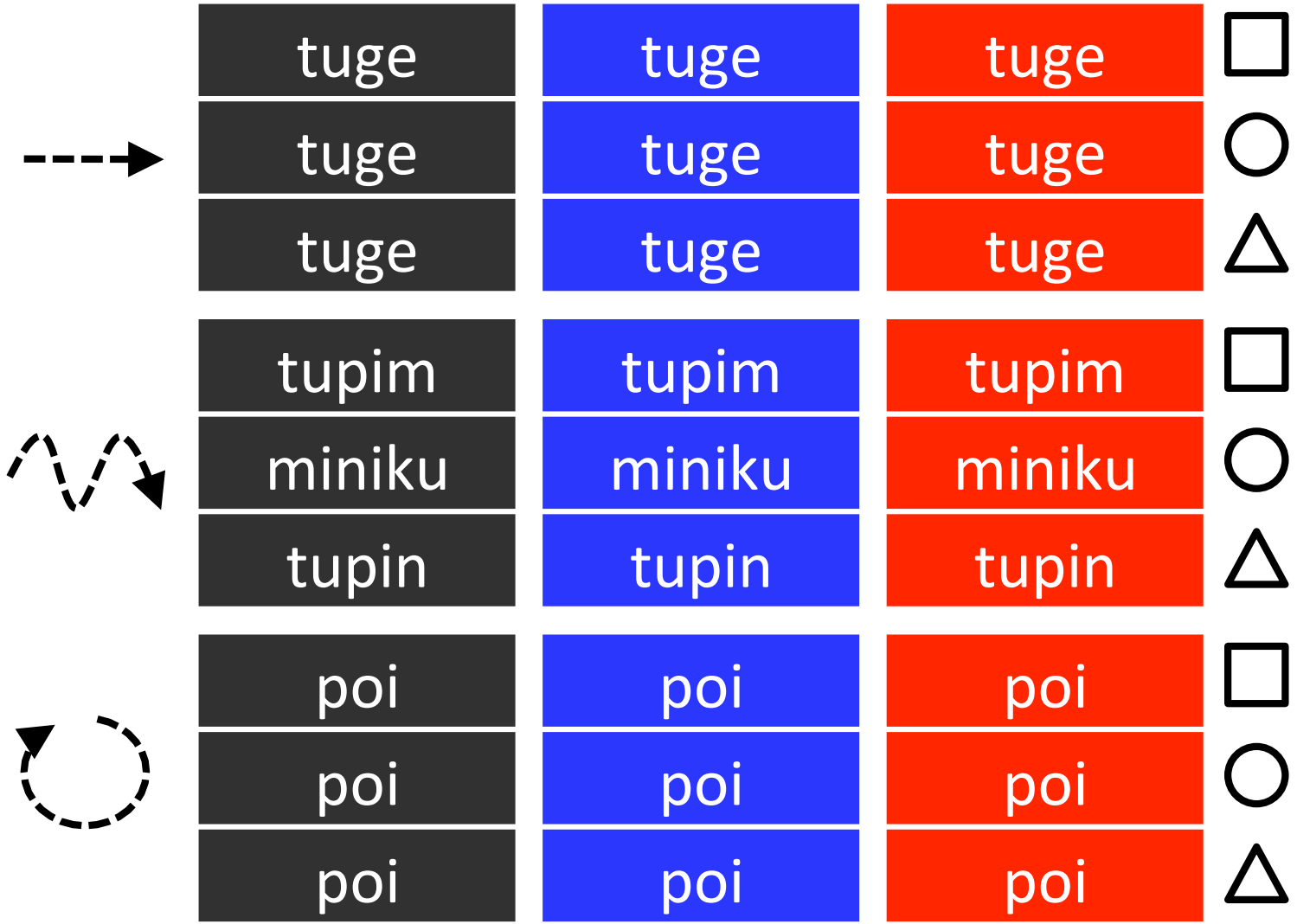


Results: decrease in error

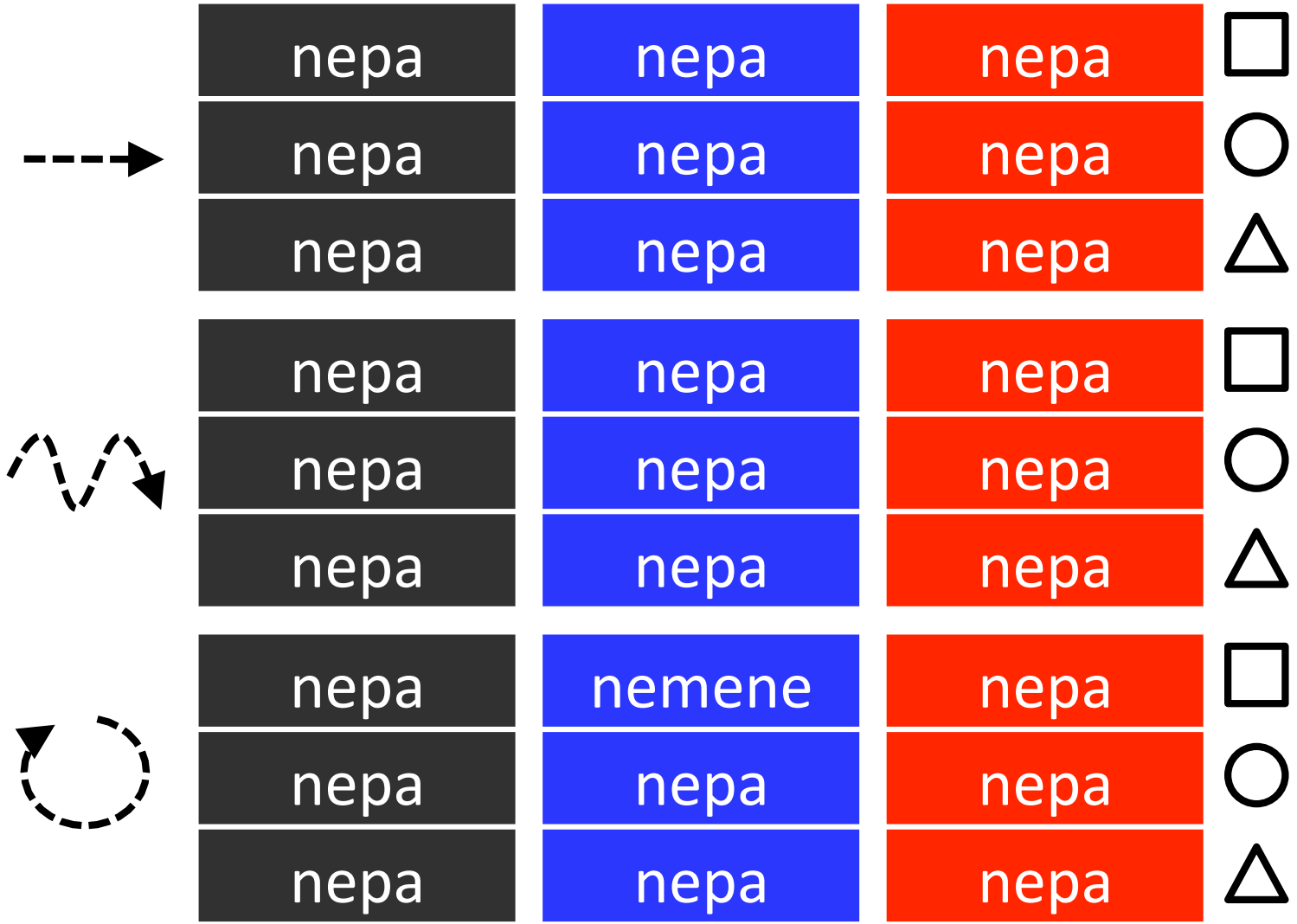


Initial language from chain 4

Final language from chain 4



Final language from chain 1 (!)



The languages become **degenerate**

Learnability and degeneracy

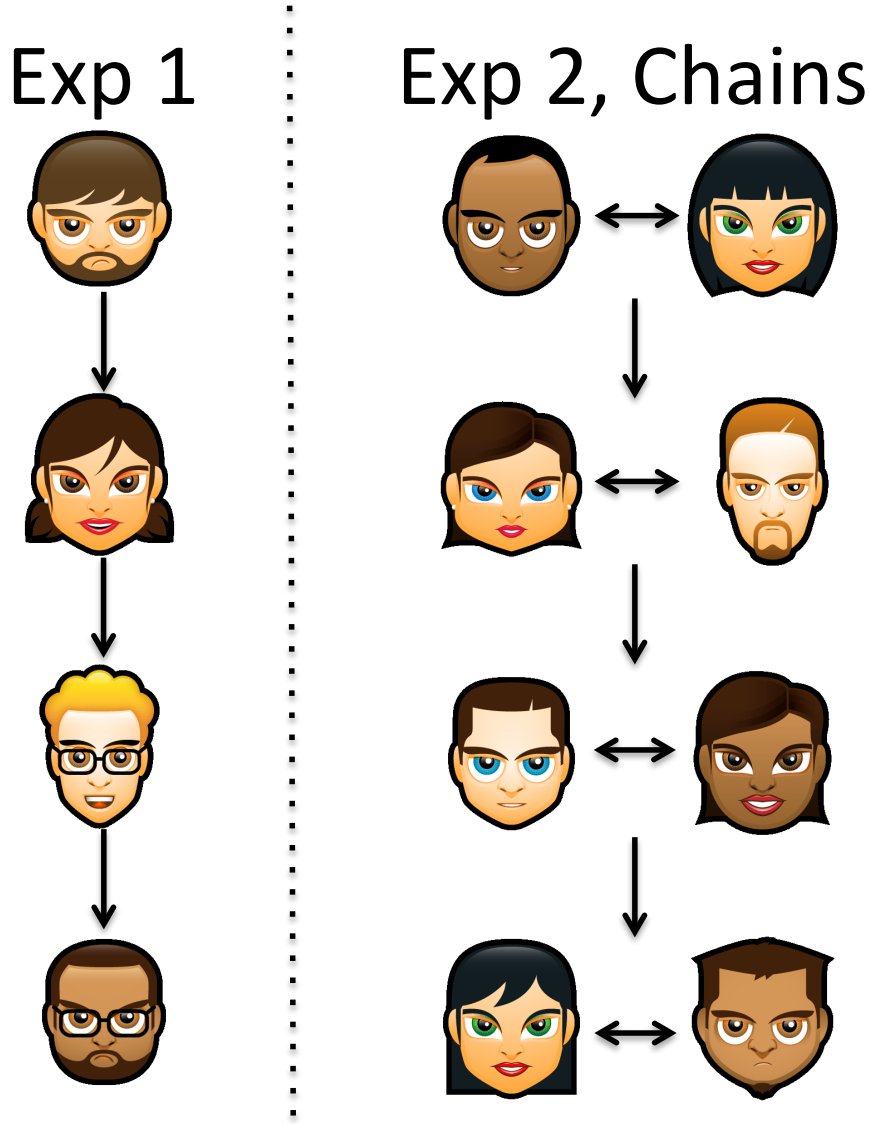
Learners prefer simpler languages

The only pressure in Experiment 1 is **learnability**

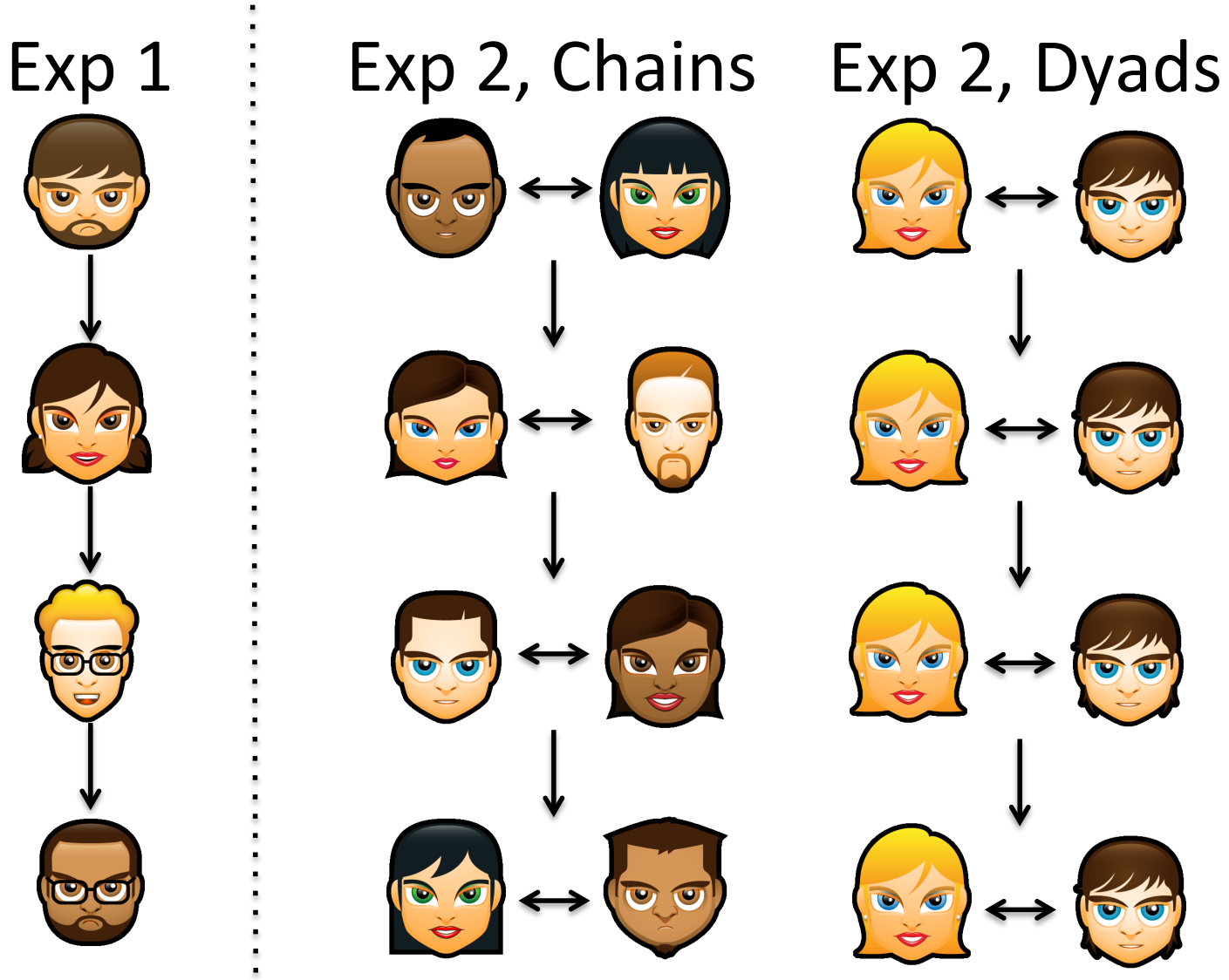
- The languages don't need to be **expressive**
- They get very simple

Can we add in a pressure for expressivity?

Experiment 2: Adding communication



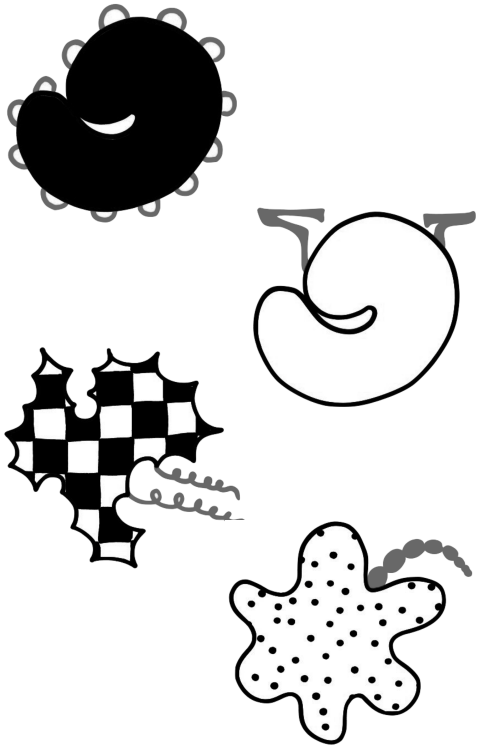
Experiment 2: Adding communication, removing learning



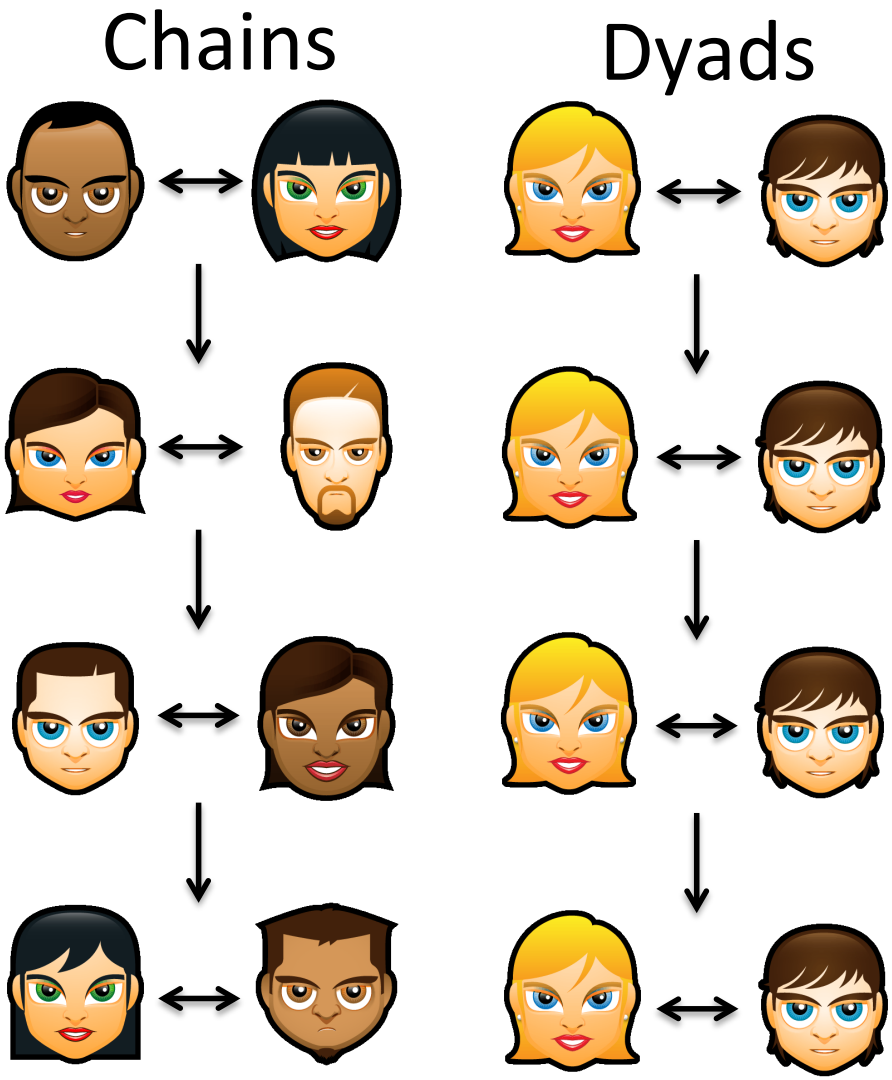
Experiment 2

Kirby, Tamariz, Cornish & Smith (forthcoming)

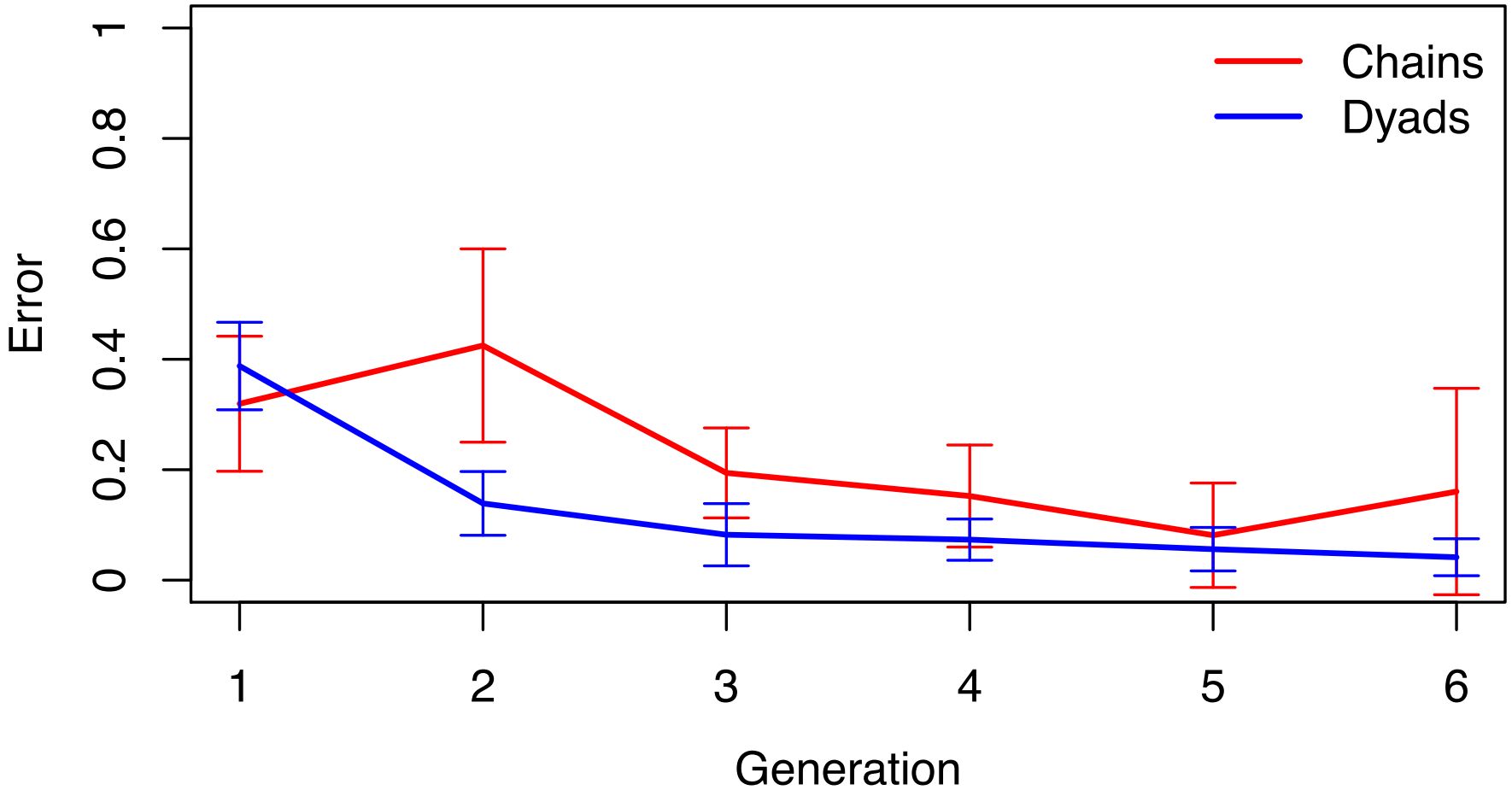
- 12 pictures
  - 3 shapes, 4 patterns
- Each picture associated with a label
  - Typed text
- Repeated training on 12 picture-label pairs
  - As before
- **Communicative testing**
  - Director: given picture, provide label
  - Matcher: given label, identify picture from array
  - Feedback provided, prizes available!



Iterated learning

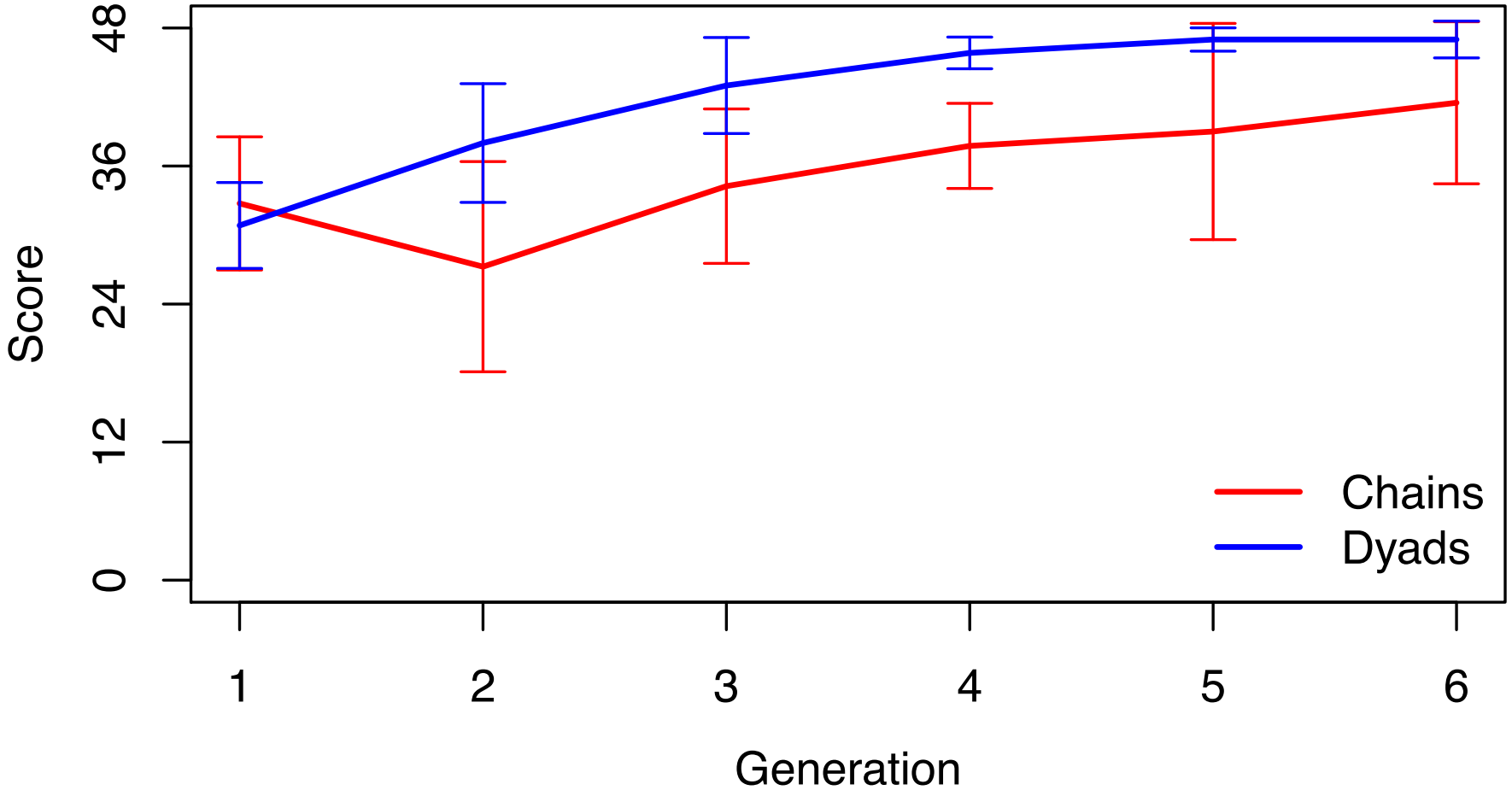


Results: decrease in error



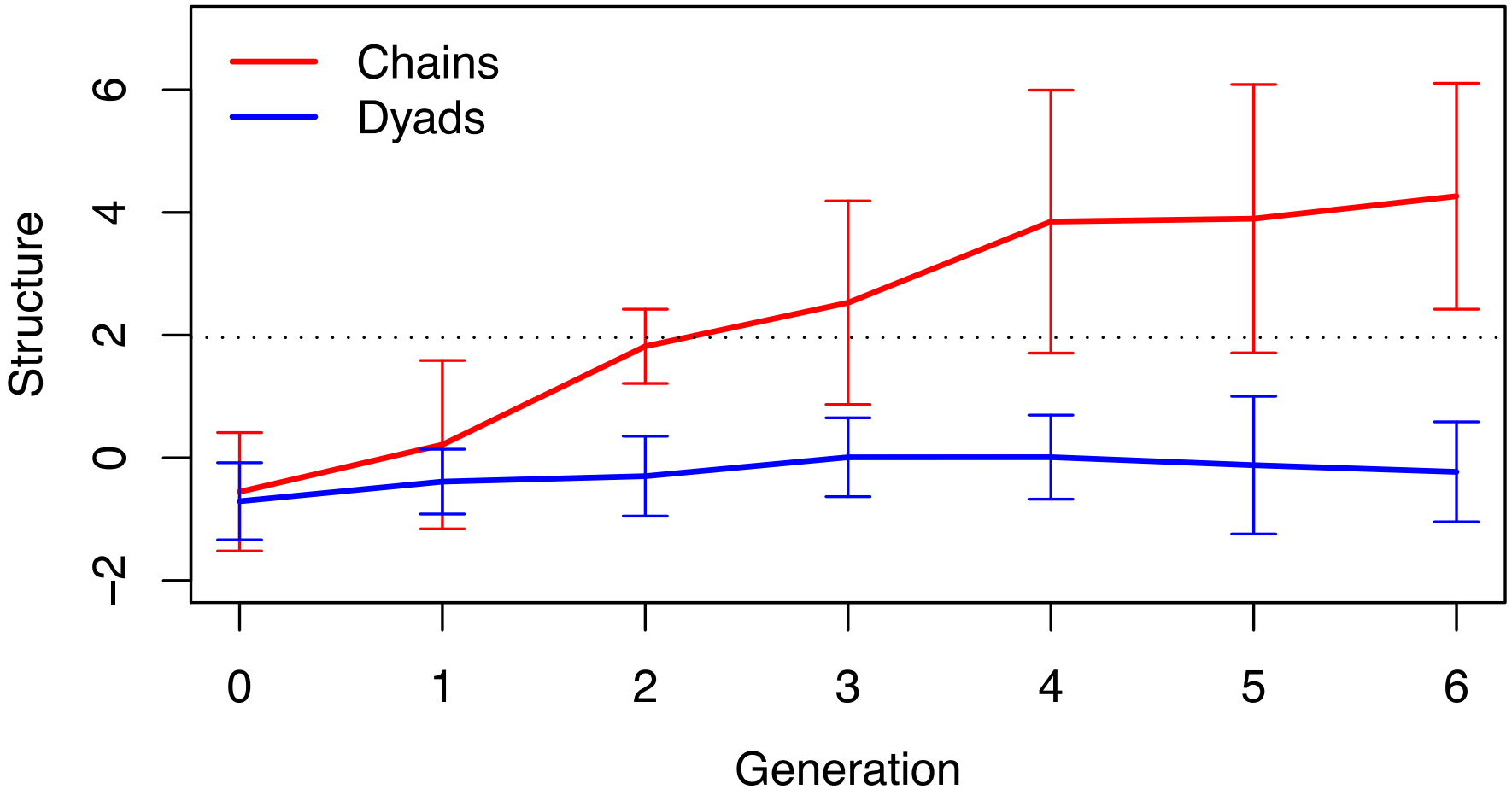
Cumulative decrease in error (Chains: L=333, m=4, n=6, p=0.006; Dyads: L=516, m=6, n=6, p<0.001)

Results: increase in score



Cumulative increase in score (Chains: L=332, m=4, n=6, p=0.008; Dyads: L=533, m=6, n=6, p<0.001)

Results: structure in chains, not dyads



An initial language

 megemume	 megi	 lameme
 mugimemu	 giwulami	 nomenoge
 wugi	 wumume	 gemulawu
 lamege	 wulamugi	 megiwuwa

A final language from a chain

 egewawu	 mega	 gamenewawu
 egewawa	 megawawa	 gamenewawa
 egewuwu	 megawuwu	 gamenewuwu
 ege	 wulagi	 gamane

Learnability + expressivity = **structure**

Learnability, expressivity and structure

Learners prefer simpler languages

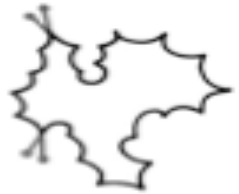
Users prefer expressive languages

The **simplest expressive** language is compositional

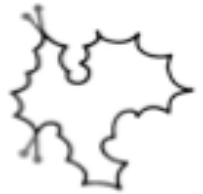
- 7 words plus 1 rule is the most compressed language which generates 12 distinct labels

What if we take away the pressure for learnability?

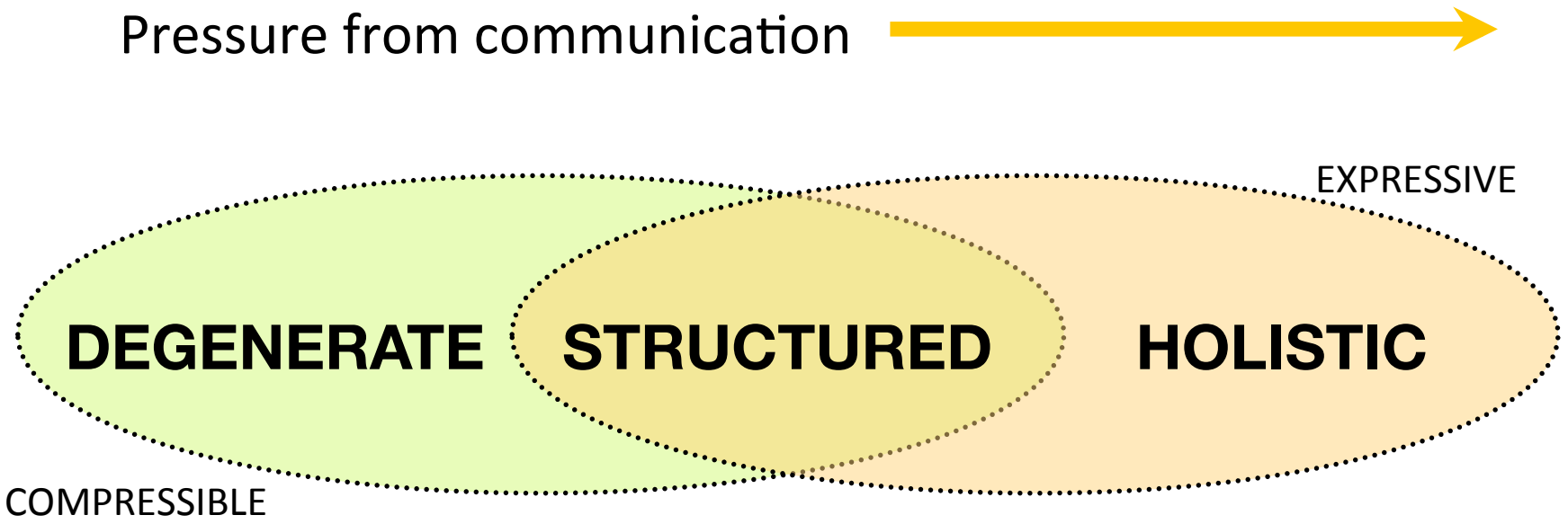
An initial language

 mokimu	 moko	 konu
 kimuwahu	 wahuhu	 lawa
 kinuki	 wekihu	 mohumu
 mukimuwa	 numu	 wakimu

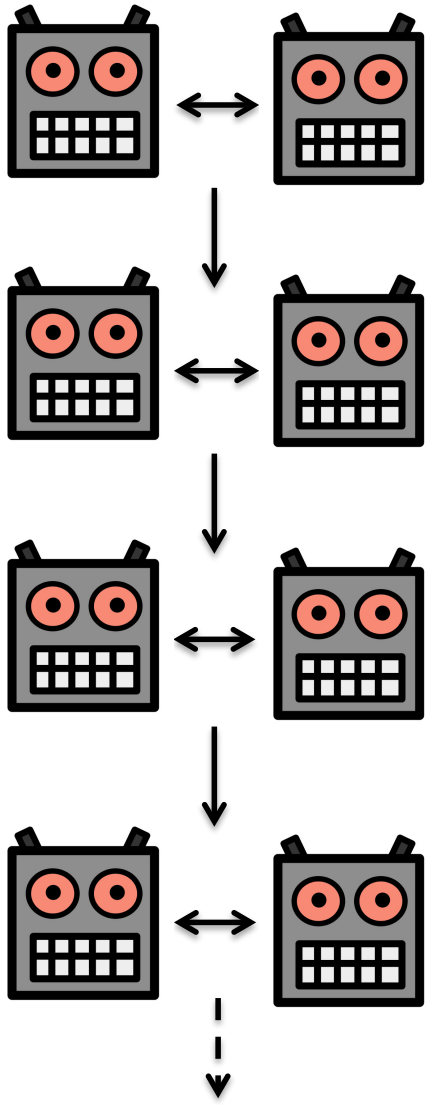
A final **holistic** language from a dyad

 manunumoko	 moko	 konu
 wekihumanunu	 mokowekihu	 lawa
 makihu	 mahiku	 wekihulawa
 manunumonu	 nomu	 wekihu

The theory


Pressure from communication
EXPRESSIVE
DEGENERATE
STRUCTURED
HOLISTIC
COMPRESSIBLE
Pressure from learning/transmission

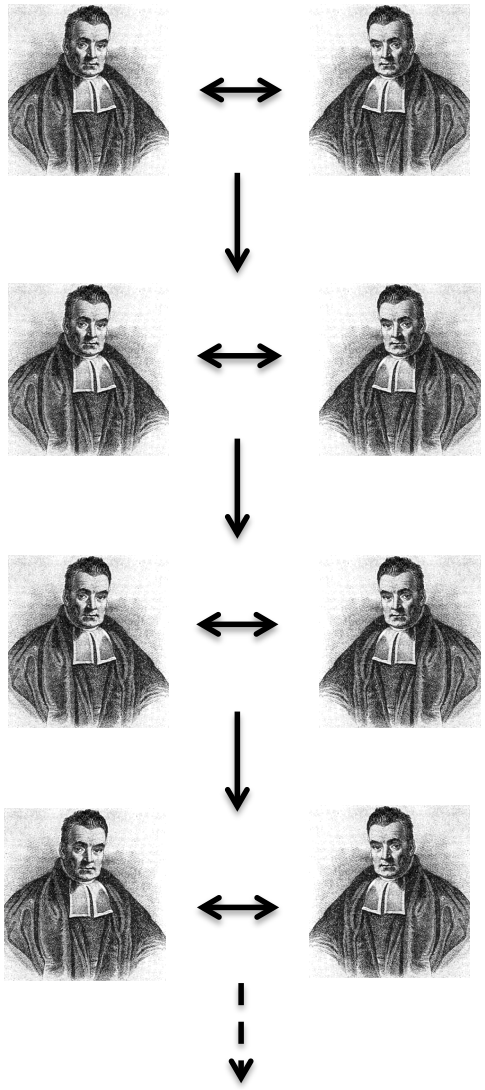
Testing the theory: Simulating Iterated Learning



Allows us to manipulate:

- **Pressure for expressivity**
  - Communication?
  - **Unambiguous** communication?
- **Pressure from new learners**
  - Transmission to naïve individuals?
- Biases of learners
  - **Experiments involve English speakers**
  - Strength of preference for simplicity
- Population structure
  - Beyond dyads and chains
- Memory
- …

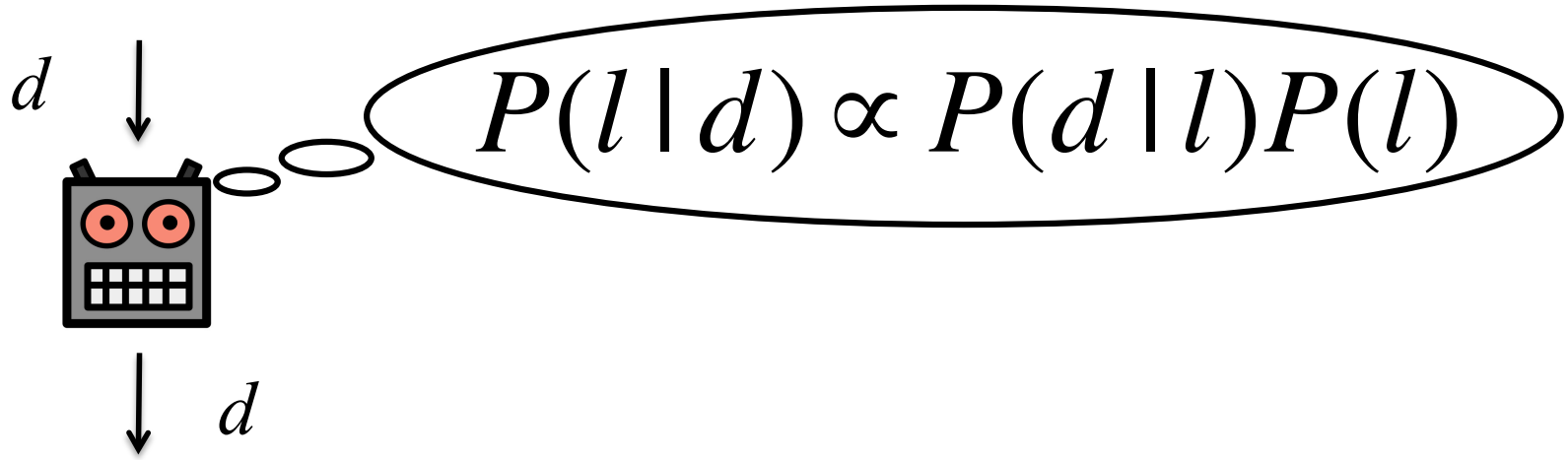
Testing the theory: **Bayesian** Iterated Learning



Allows us to manipulate:

- **Pressure for expressivity**
  - Communication?
  - **Unambiguous** communication?
- **Pressure from new learners**
  - Transmission to naïve individuals?
- Biases of learners
  - **Experiments involve English speakers**
  - Strength of preference for simplicity
- Population structure
  - Beyond dyads and chains
- Memory
- …

Learning as Bayesian inference



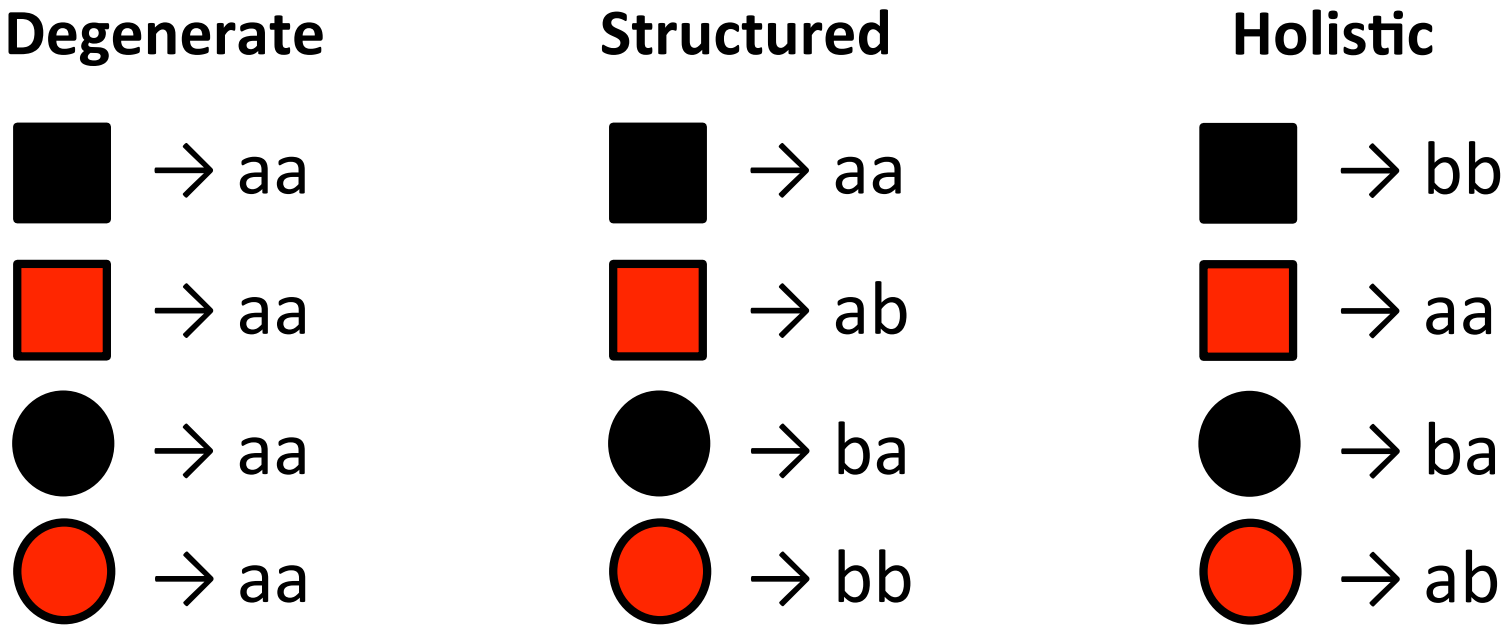
Data: meaning-signal pairs

Languages: specify signals for meanings

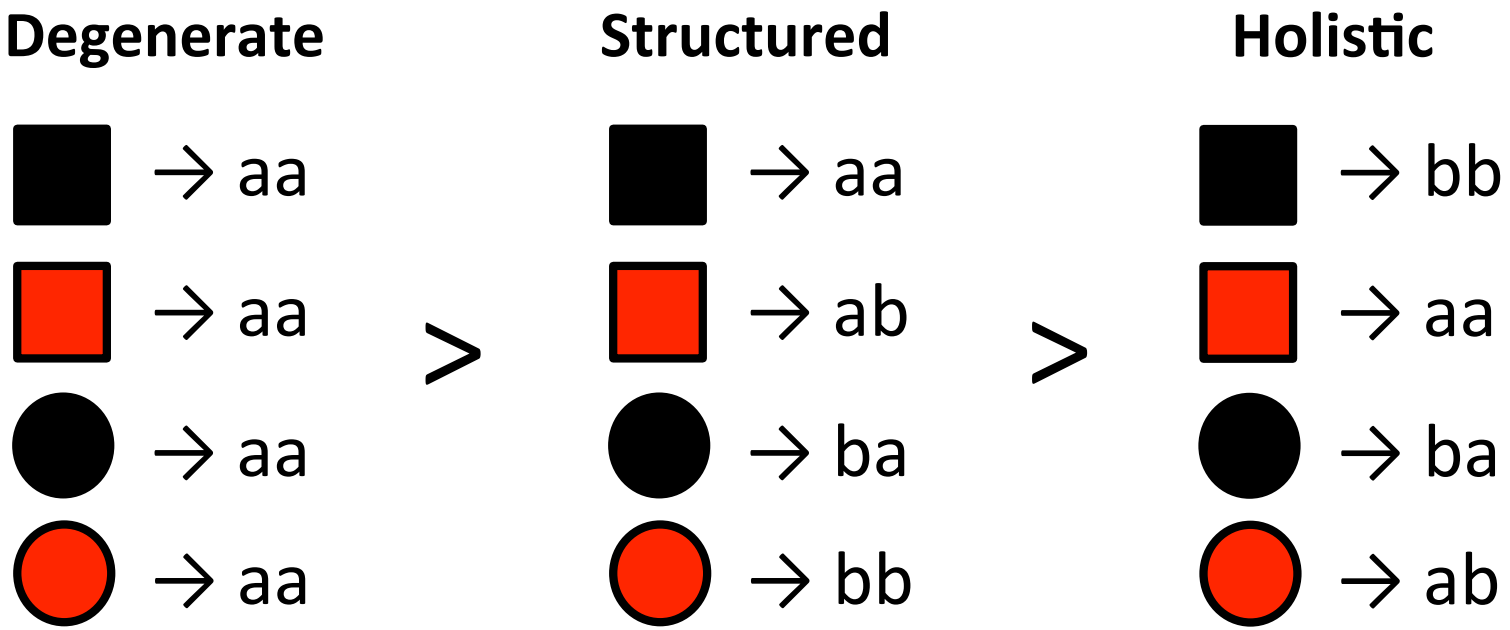
Likelihood p(d|l): model of language use

Prior p(l): learner preference for compressible languages (shorter coding length)

Example languages (from set of 256)



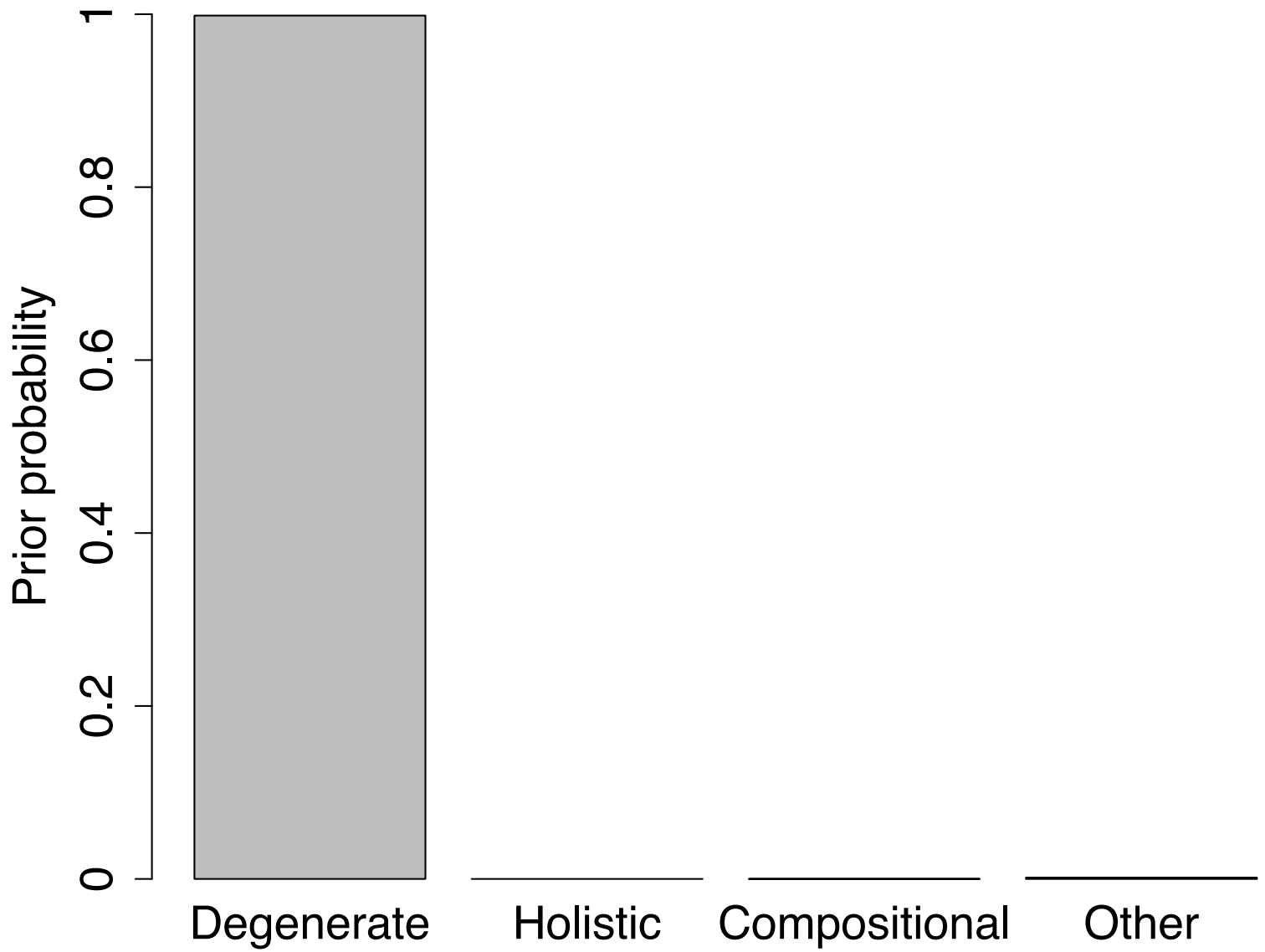
The prior



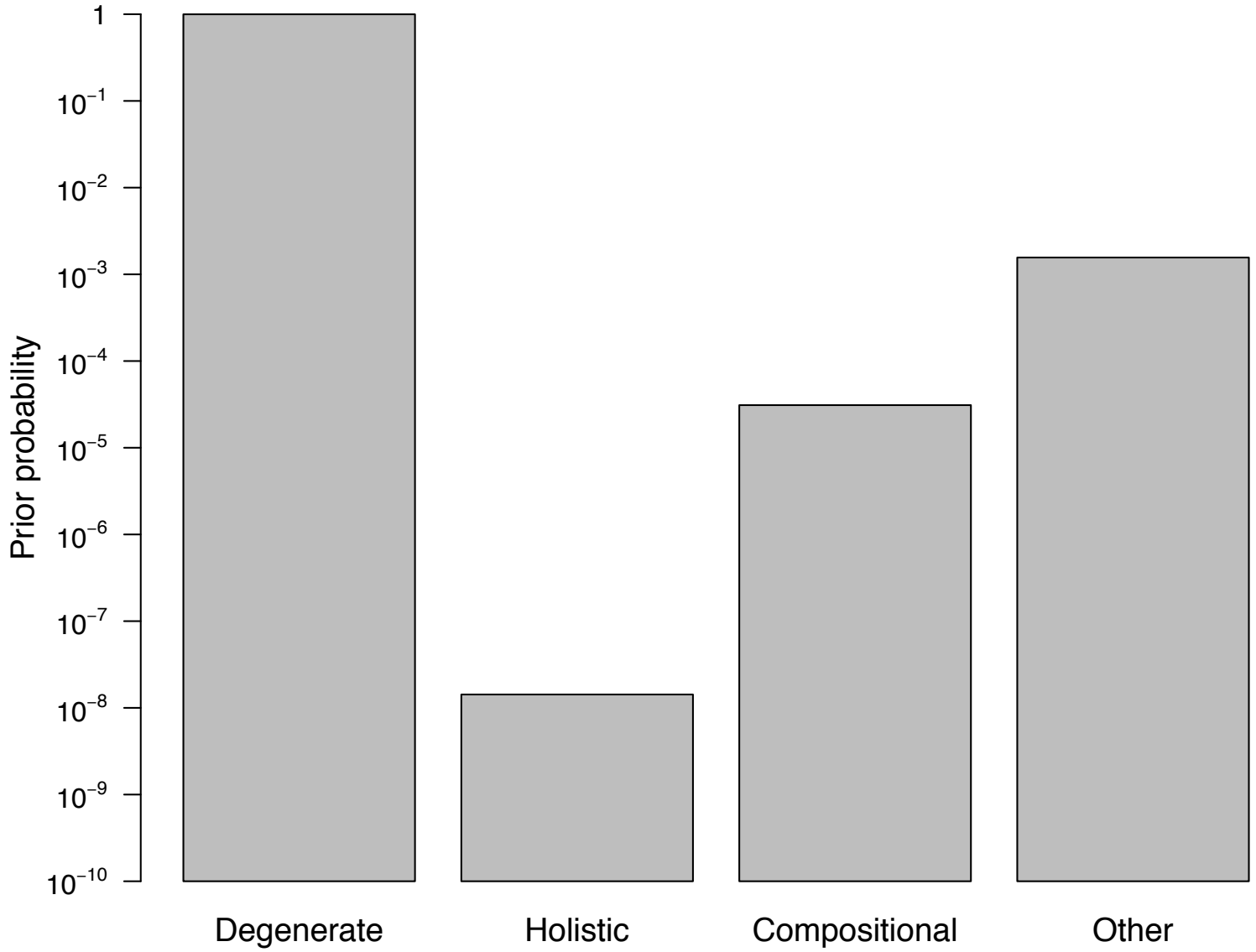
Preference for compressibility

- Coding length of grammar for each language

The prior


Prior probability
0
0.2
0.4
0.6
0.8
1
Degenerate
Holistic
Compositional
Other

The prior (log scale)



Bayesian inference

$$P(h|d) \propto P(d|h)P(h)$$

Data: meaning-signal pairs (in context)

Hypotheses: (distributions over) languages

Likelihood: model of language use

Prior: preference for simple languages

Meanings and languages

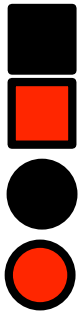
- Meanings and signals: vary along 2 binary dimensions
  - **Meanings:** 00,01,10,11
  - **Signals:** aa, ab, ba, bb
- Language: specifies a signal for each meaning
  - 256 possible languages

Likelihood

$P(\text{signal} \mid \text{language, context}) \propto (1/\text{ambiguity in context})^{\alpha}$

Frank & Goodman (2012)

My language



→ aa

→ ab

→ bb

→ bb

The context

Likelihoods

$P(\text{aa} \mid \text{My language}) \propto 1-\varepsilon$

$P(\text{ab} \mid \text{My language}) \propto \varepsilon/3$

$P(\text{ba} \mid \text{My language}) \propto \varepsilon/3$

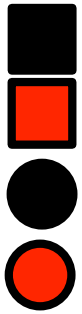
$P(\text{bb} \mid \text{My language}) \propto \varepsilon/3$

Likelihood

$P(\text{signal} \mid \text{language, context}) \propto (1/\text{ambiguity in context})^{\alpha}$

Frank & Goodman, 2012, *Science*

My language



→ aa

→ ab

→ **bb**

→ **bb**

The context

Likelihoods

P(aa | My language) ∝ ε/3

P(ab | My language) ∝ ε/3

P(ba | My language) ∝ ε/3

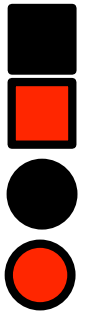
P(bb | My language) ∝ 1-ε

Likelihood

$P(\text{signal} \mid \text{language, context}) \propto (1/\text{ambiguity in context})^{\alpha}$

Frank & Goodman, 2012, *Science*

My language



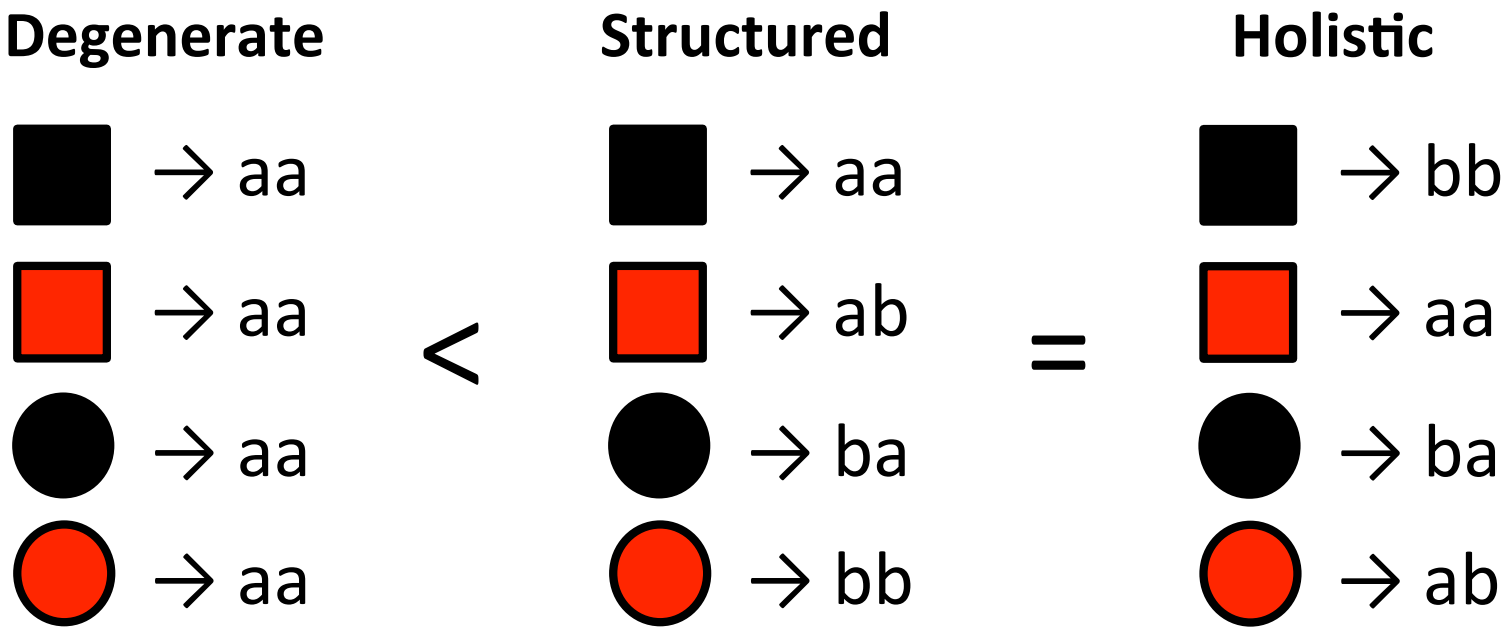
→ aa
→ ab
→ **bb**
→ **bb**

The context

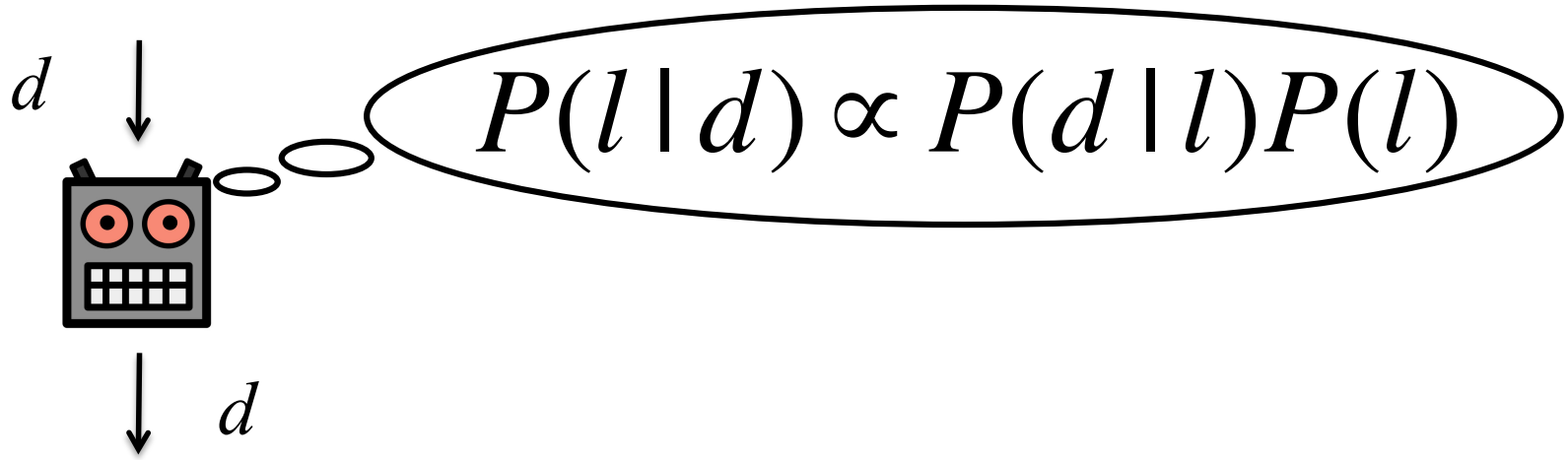
Likelihoods

P(aa | My language) ∝ ε/3
P(ab | My language) ∝ ε/3
P(ba | My language) ∝ ε/3
P(bb | My language) ∝ ½$^{\alpha}$(1-ε)

Likelihood


Degenerate
→ aa
→ aa
→ aa
→ aa
<
Structured
→ aa
→ ab
→ ba
→ bb
=
Holistic
→ bb
→ aa
→ ba
→ ab

Learning as Bayesian inference



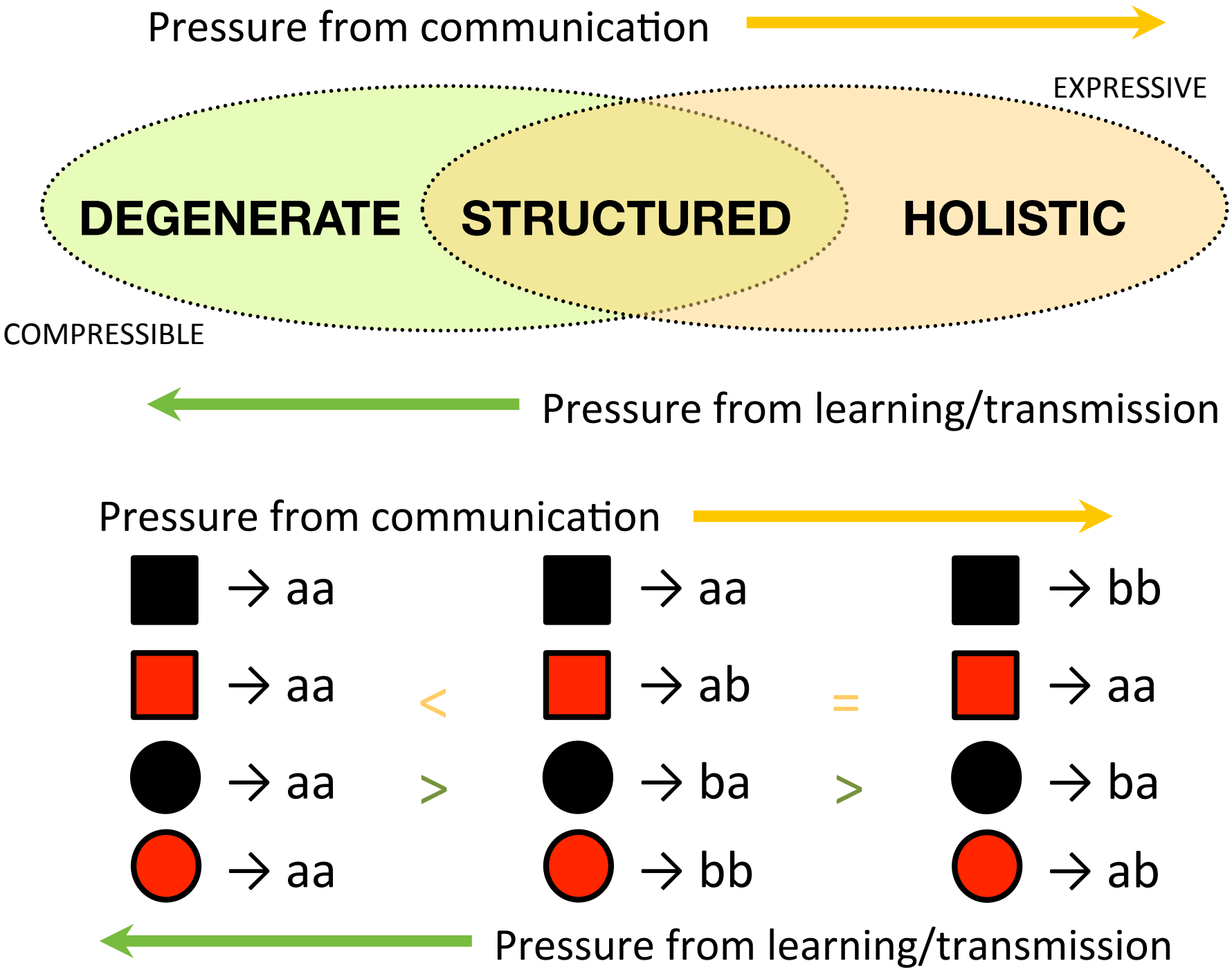
Data: meaning-signal pairs

Languages: specify signals for meanings

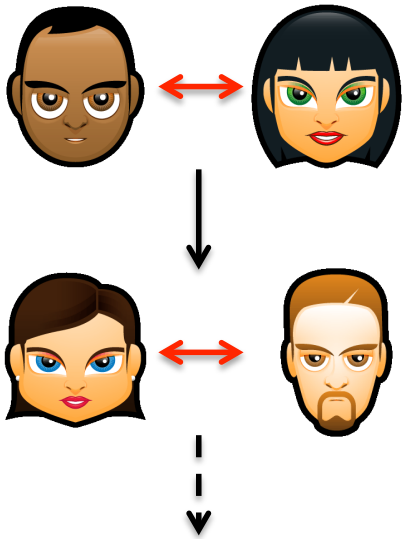
Likelihood p(d|l): model of language use

- Avoid utterances which are ambiguous in context

Prior p(l): learner preference for compressible languages (shorter coding length)

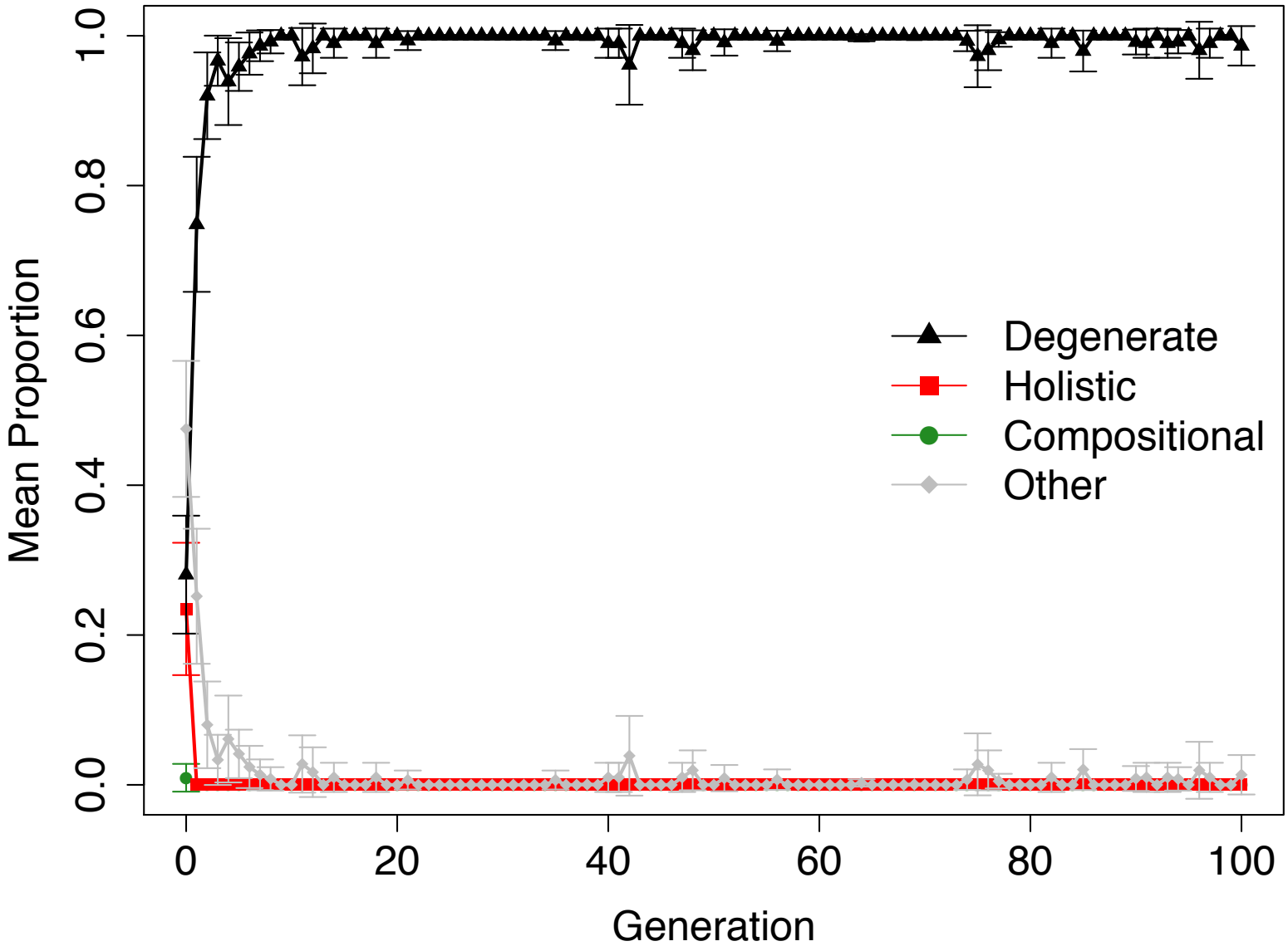

Pressure from communication
EXPRESSIVE
DEGENERATE
STRUCTURED
HOLISTIC
COMPRESSIBLE
Pressure from learning/transmission
Pressure from communication
→ aa
→ aa
→ aa
→ aa
<
>
→ aa
→ ab
→ ba
→ bb
=
>
→ bb
→ aa
→ ba
→ ab
Pressure from learning/transmission

Simulation of Exp 1

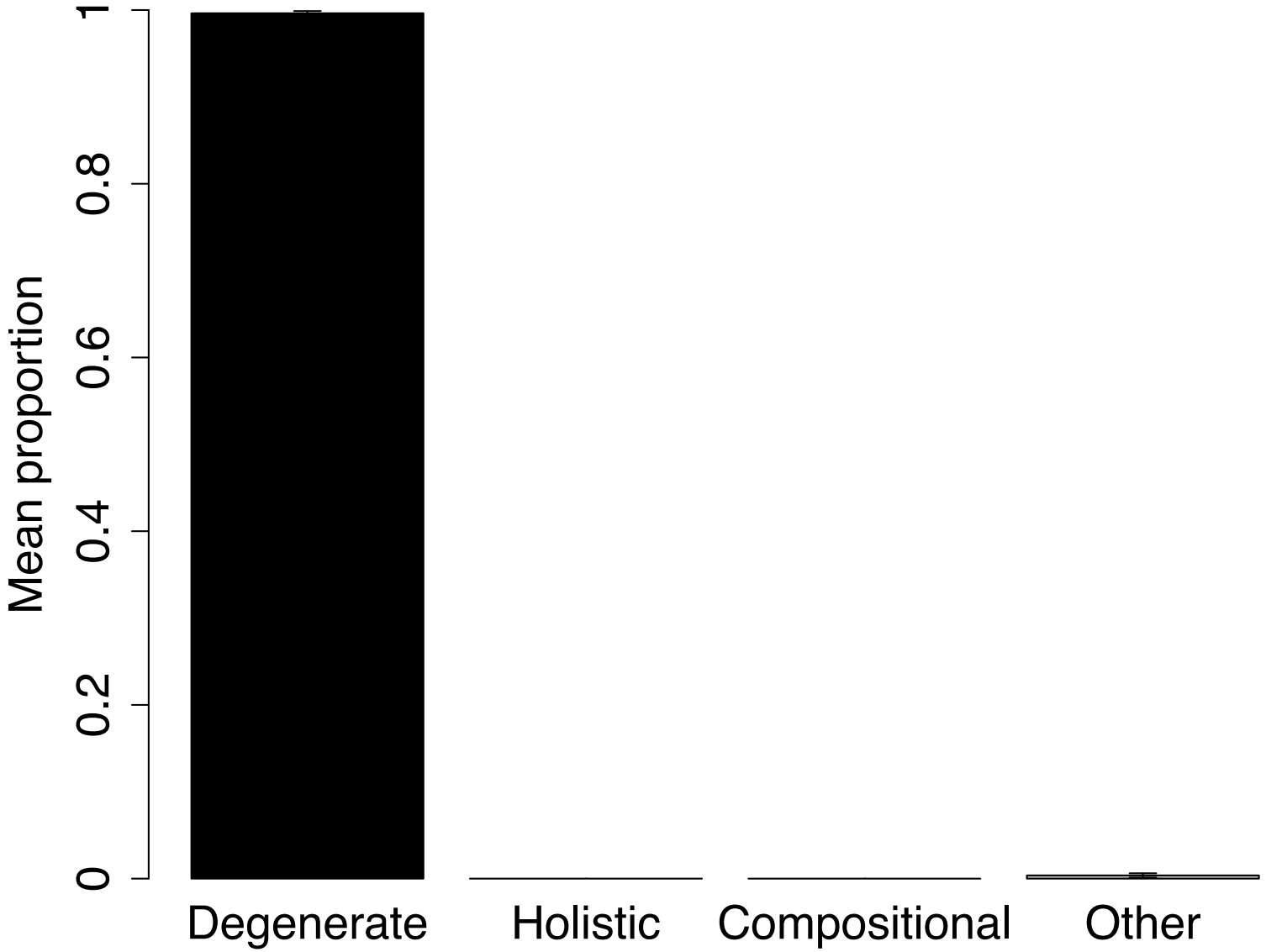


- Members of pairs interact
- But they don't care about communication: $\alpha=0$

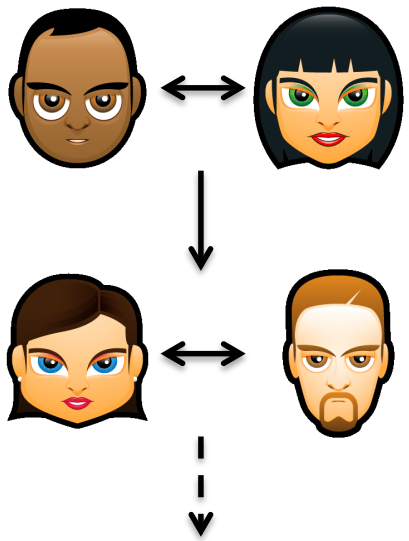
Simulation of Exp 1



Simulation of Exp 1

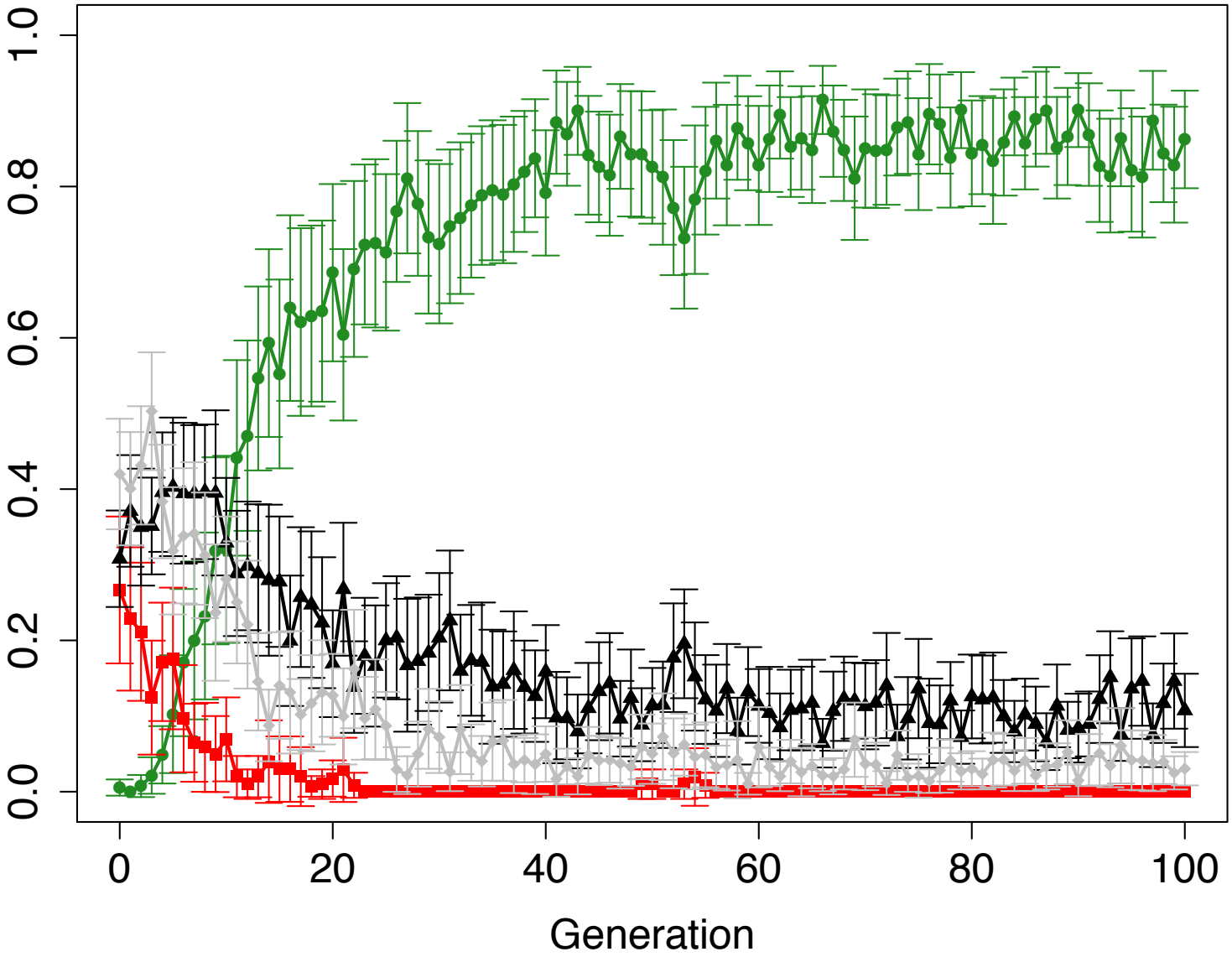


Simulation of Exp 2, Chains

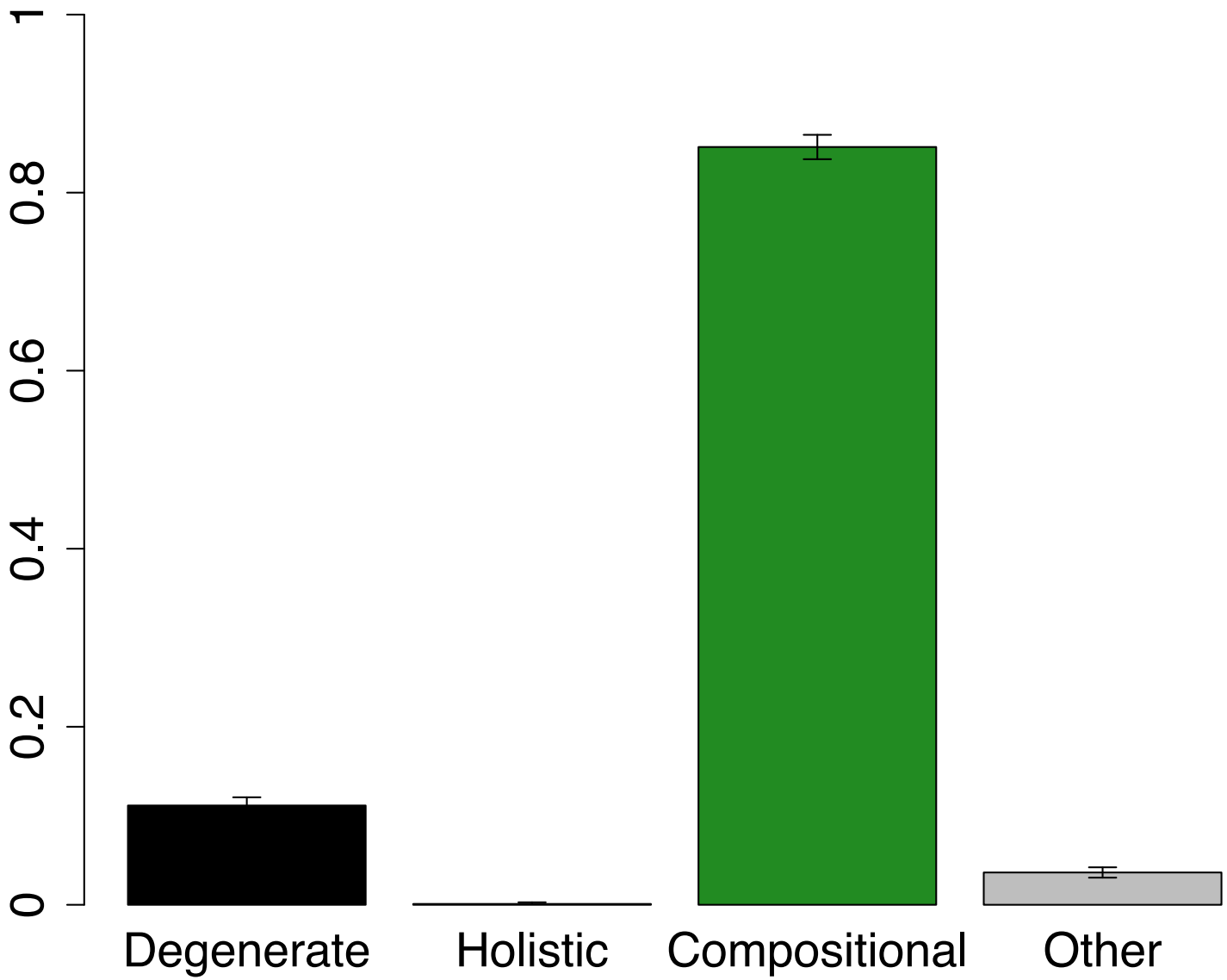


- Members of pairs interact
- They care about communication: $\alpha=2$

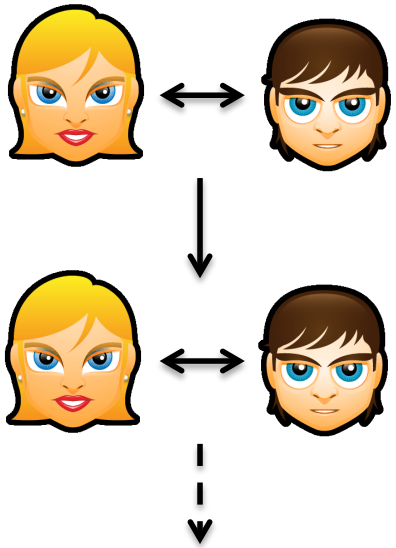
Simulation of Exp 2, Chains



Simulation of Exp 2, Chains

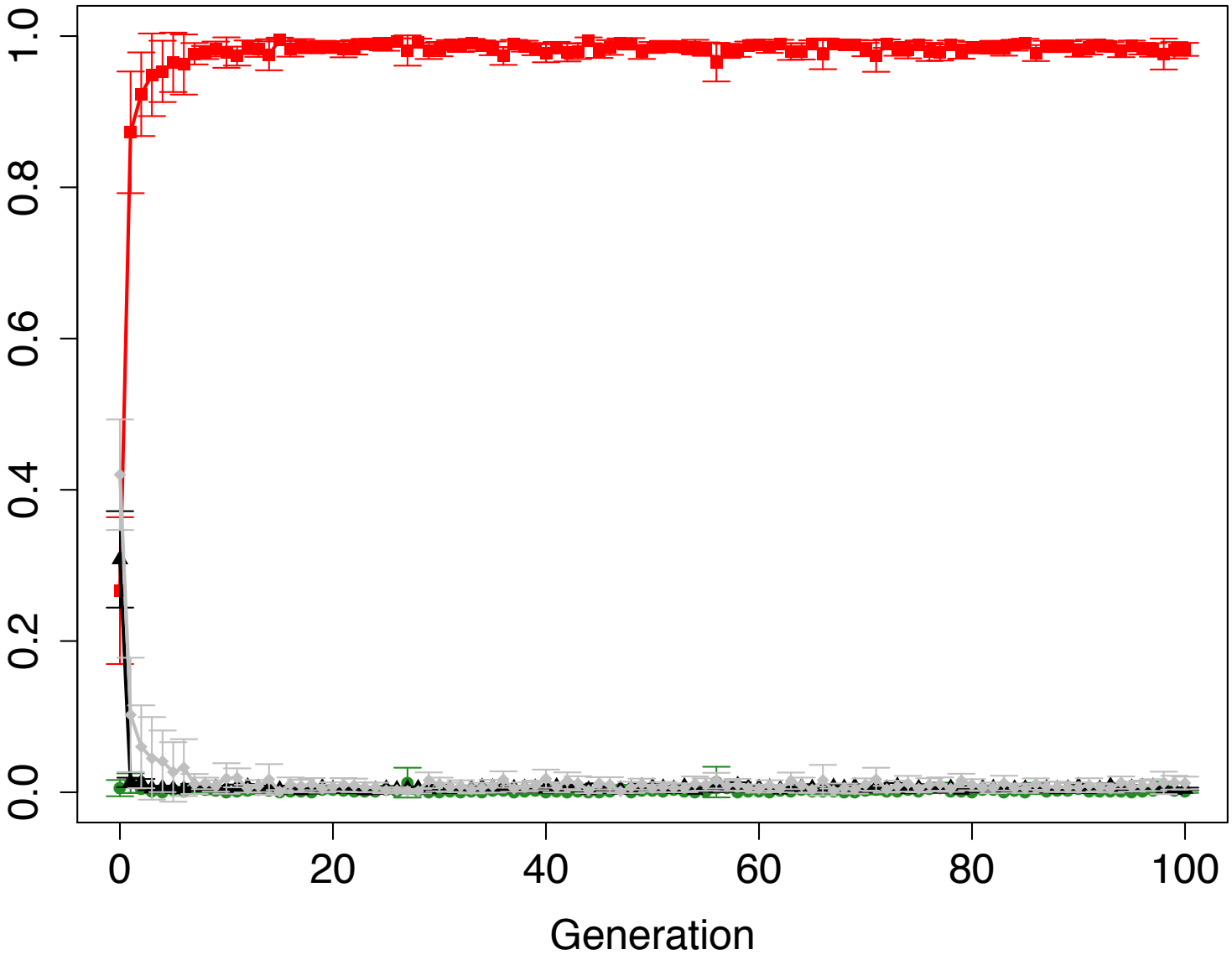


Simulation of Exp 2, Dyads

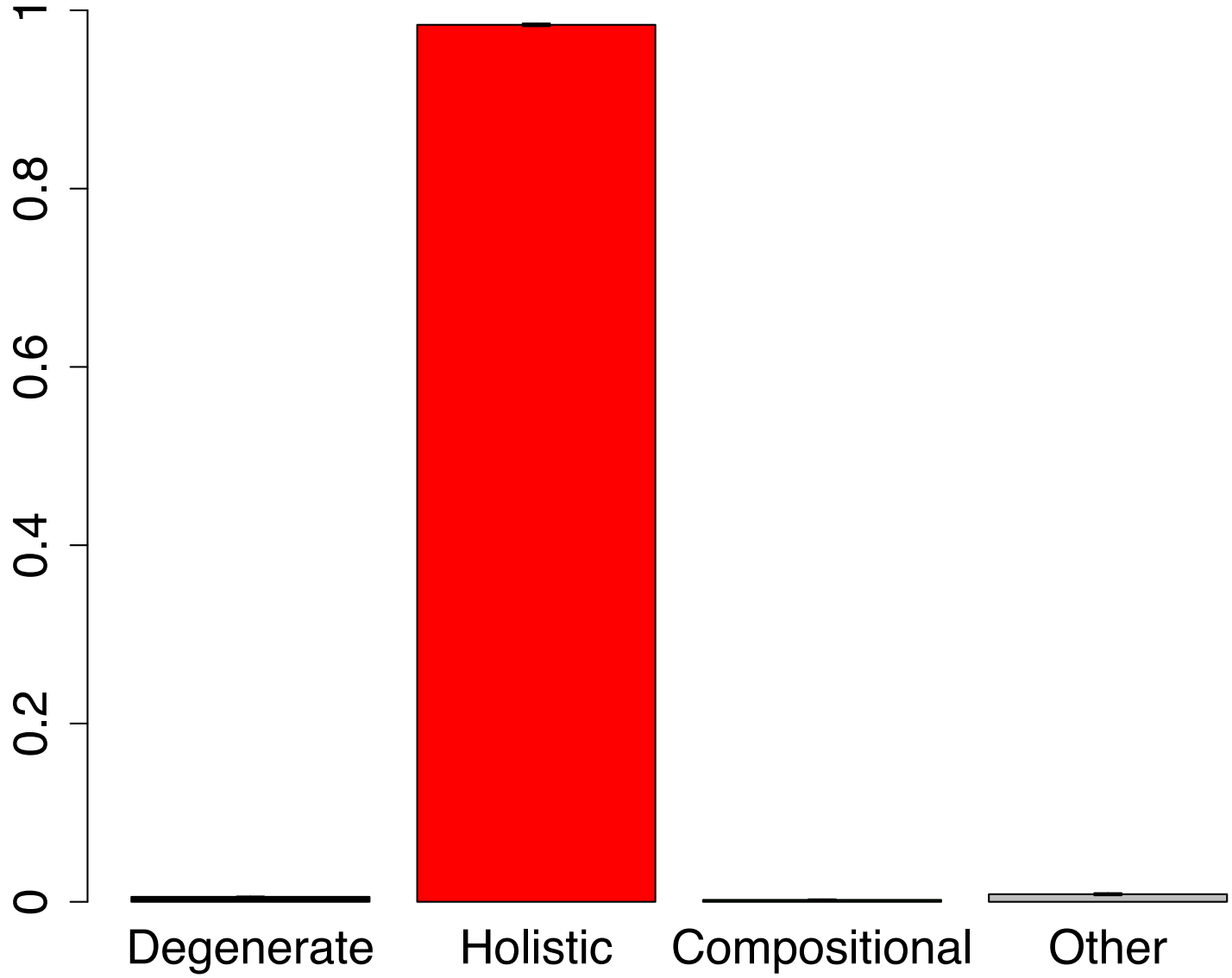


- Members of pairs interact
- They care about communication: $\alpha=2$
- They learn from their own previous output: no naïve learners

Simulation of Exp 2, Dyads



Simulation of Exp 2, Dyads



What's required for structure in language?

- A general preference for simplicity/ compressibility
- A communicative task
- Communication with naïve individuals
  - This might differ across populations

Once these are in place, cultural evolution delivers structure

The value of modelling

- Clean implementation of theory
- Allows you to generate predictions of theory, check against real-world data
- Particularly valuable when theories are hard to understand through purely verbal reasoning
  - e.g. systems involving interactions between learning, culture, biology

The value of human simulation

- Scientifically: it's important to be confident that our models are relevant to the behaviour/species we are interested in
- Pragmatically: people seem to find experiments more digestible than models

The ideal: a combination of the two

Liked this course? Do the MSc!

References

- Bickerton, D. (2007) Language evolution: a brief guide for linguists. *Lingua, 117,* 510 –526.
- Frank, M. C., & Goodman, N. D. (2012) Predicting pragmatic reasoning in language games. *Science, 336,* 998.
- Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. J. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models* (pp. 173-204). Cambridge: Cambridge University Press.
- Kirby, S., Cornish, H., & Smith, K. (2008) Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *PNAS, 105,* 10681–10686.
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (forthcoming). Compression and Communication in the Cultural Evolution of Linguistic Structure. To appear in *Cognition.*