Linking Language Development and Language Transmission

Vera Kempe
Abertay University
v.kempe@abertay.ac.uk


Aberdeen
Dundee
Glasgow
Edinburgh
Newcastle
Belfast
Manchester
Dublin
Liverpool
Birmingham
Cardiff
London

Overview

1. Studying Language Evolution in the Lab:
   Overview and Demonstration

   Iterated learning: What's different in children?

2. Negotiating Meaning:
   Communicative Constraints in Children and Adults

   Can children invent a novel communication system?

3. Transmitting Symbolic Signals:
   Learnability Constraints in Children and Adults

   Who are the agents of language change?

4. Accommodating the Learner:
   The Role of Teaching in Language Transmission

   How do experts transmit linguistic knowledge?

Overview

1. Studying Language Evolution in the Lab:
   Overview and Demonstration

   Iterated learning: What's different in children?

2. Negotiating Meaning:
   Communicative Constraints in Children and Adults

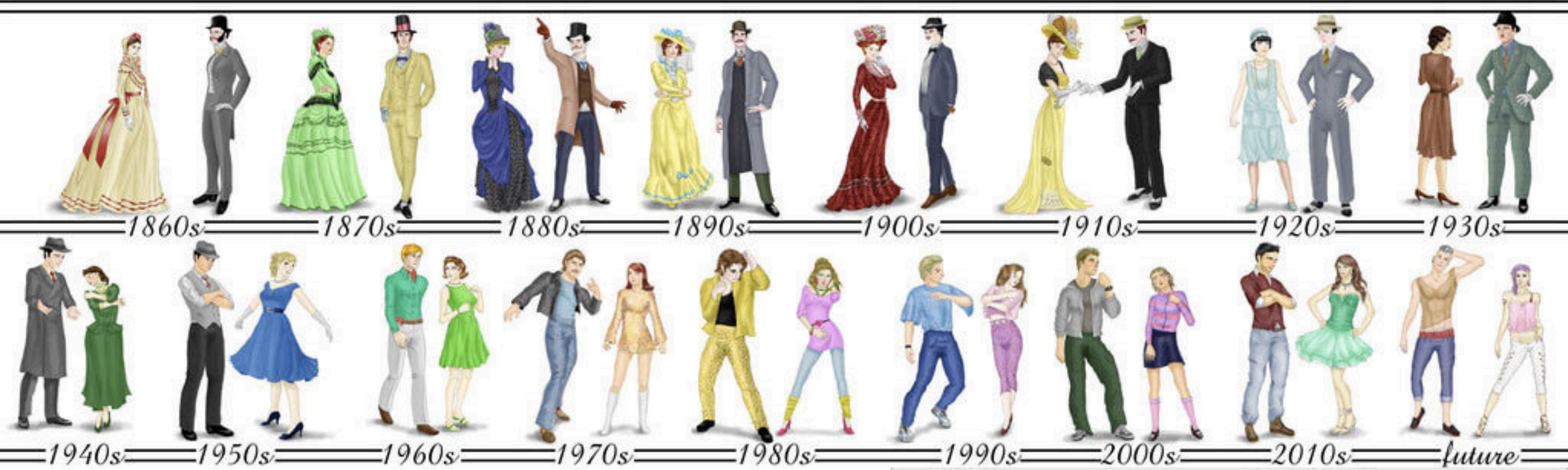
   Can children invent a novel communication system?

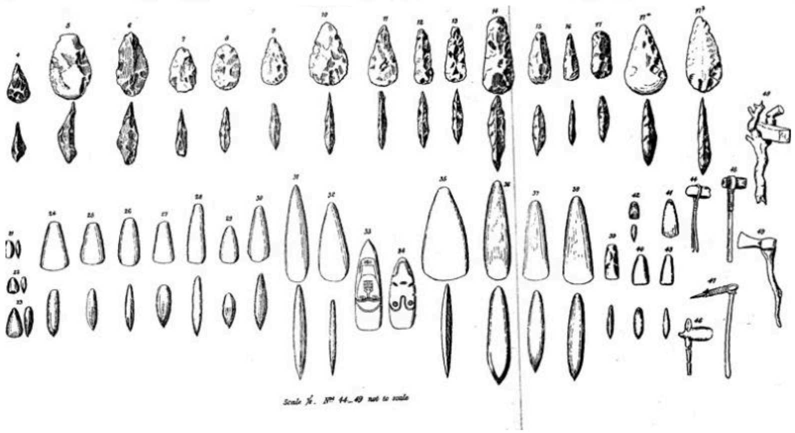
3. Transmitting Symbolic Signals:
   Learnability Constraints in Children and Adults

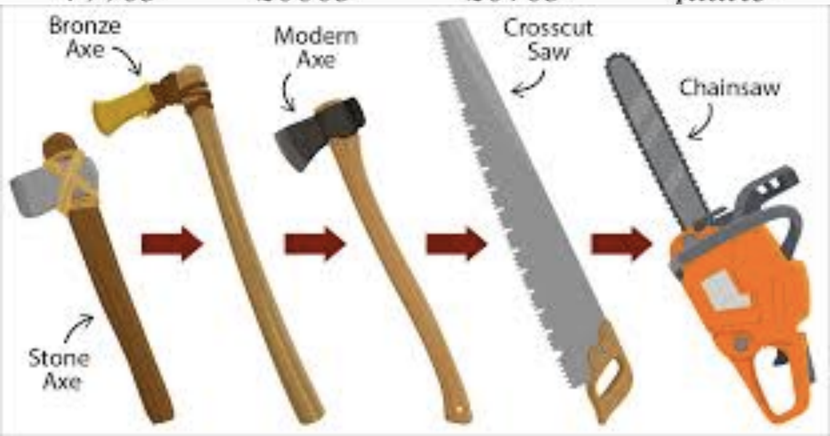
   Who are the agents of language change?

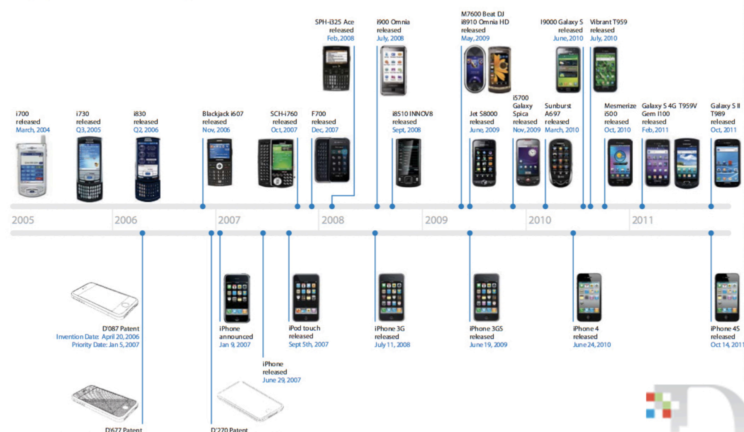
4. Accommodating the Learner:
   The Role of Teaching in Language Transmission

   How do experts transmit linguistic knowledge?


1860s
1870s
1880s
1890s
1900s
1910s
1920s
1930s
1940s
1950s
1960s
1970s
1980s
1990s
2000s
2010s
future

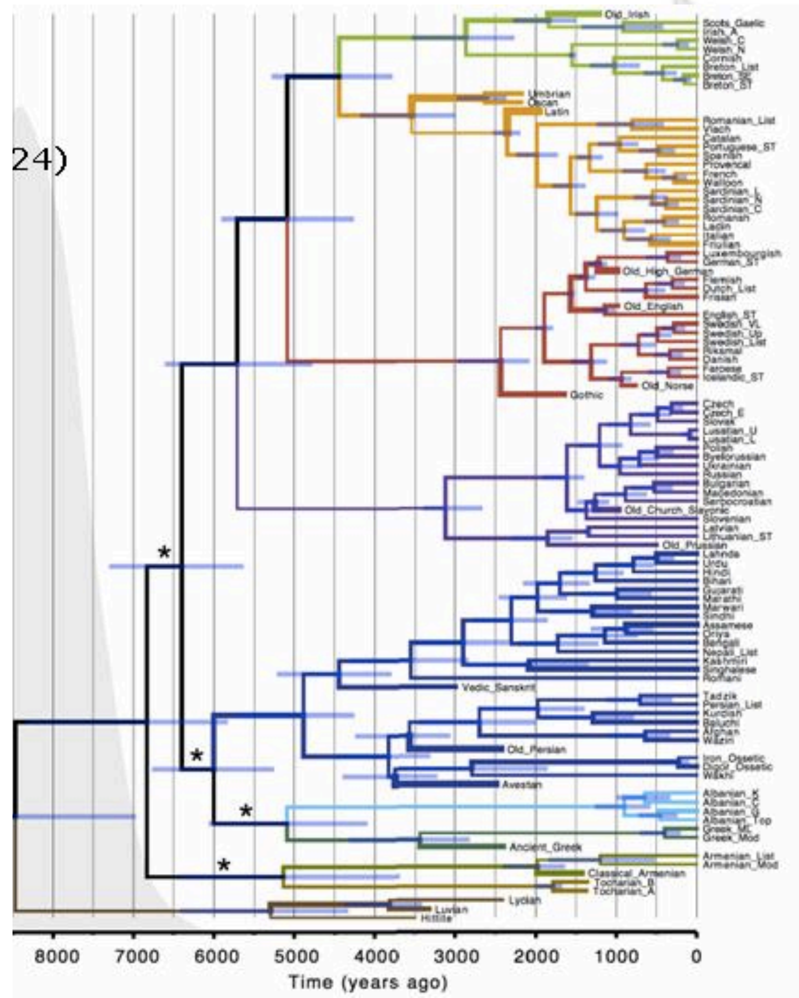



Bronze Axe
Modern Axe
Crosscut Saw
Chainsaw
Stone Axe


2005
2006
2007
2008
2009
2010
2011


HIGHLIGHTS OF THE
jazz
STORY IN USA

24)


Old_Irish
Scots_Gaelic
Irish_A
Welsh_C
Welsh_N
Cornish
Breton_List
Breton_SE
Breton_ST
Umbrian
Oscan
Latin
Romanian_List
Vlach
Catalan
Portuguese_ST
Spanish
Provencal
French
Walloon
Sardinian_L
Sardinian_N
Sardinian_C
Romansh
Ladin
Italian
Friulian
Luxembourgish
German_ST
Old_High_German
Flemish
Dutch_List
Frisian
Old_English
English_ST
Swedish_VL
Swedish_Up
Swedish_List
Riksmal
Danish
Faroese
Icelandic_ST
Old_Norse
Gothic
Czech
Czech_E
Slovak
Lusatian_U
Lusatian_L
Polish
Byelorussian
Ukrainian
Russian
Bulgarian
Macedonian
Serbocroatian
Old_Church_Slavonic
Slovenian
Latvian
Lithuanian_ST
Old_Prussian
Lahnda
Urdu
Hindi
Bihari
Gujarati
Marathi
Marwari
Sindhi
Assamese
Oriya
Bengali
Nepali_List
Kashmiri
Singhalese
Romani
Vedic_Sanskrit
Tadzik
Persian_List
Kurdish
Baluchi
Afghan
Waziri
Old_Persian
Iron_Ossetic
Digor_Ossetic
Wakhi
Avestan
Albanian_K
Albanian_C
Albanian_G
Albanian_Top
Greek_Mc
Greek_Mod
Ancient_Greek
Armenian_List
Armenian_Mod
Classical_Armenian
Tocharian_B
Tocharian_A
Lycian
Luvian
Hittite
*
*
*
*
8000
7000
6000
5000
4000
3000
2000
1000
0
Time (years ago)

Language as Shaped by the Brain

Morten H. Christiansen
Department of Psychology
Cornell University
Ithaca, NY 14853
email: mhc27@cornell.edu

and

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501

Nick Chater
Department of Psychology
University College London
London, WC1E 6BT
email: n.chater@ucl.ac.uk

Abstract

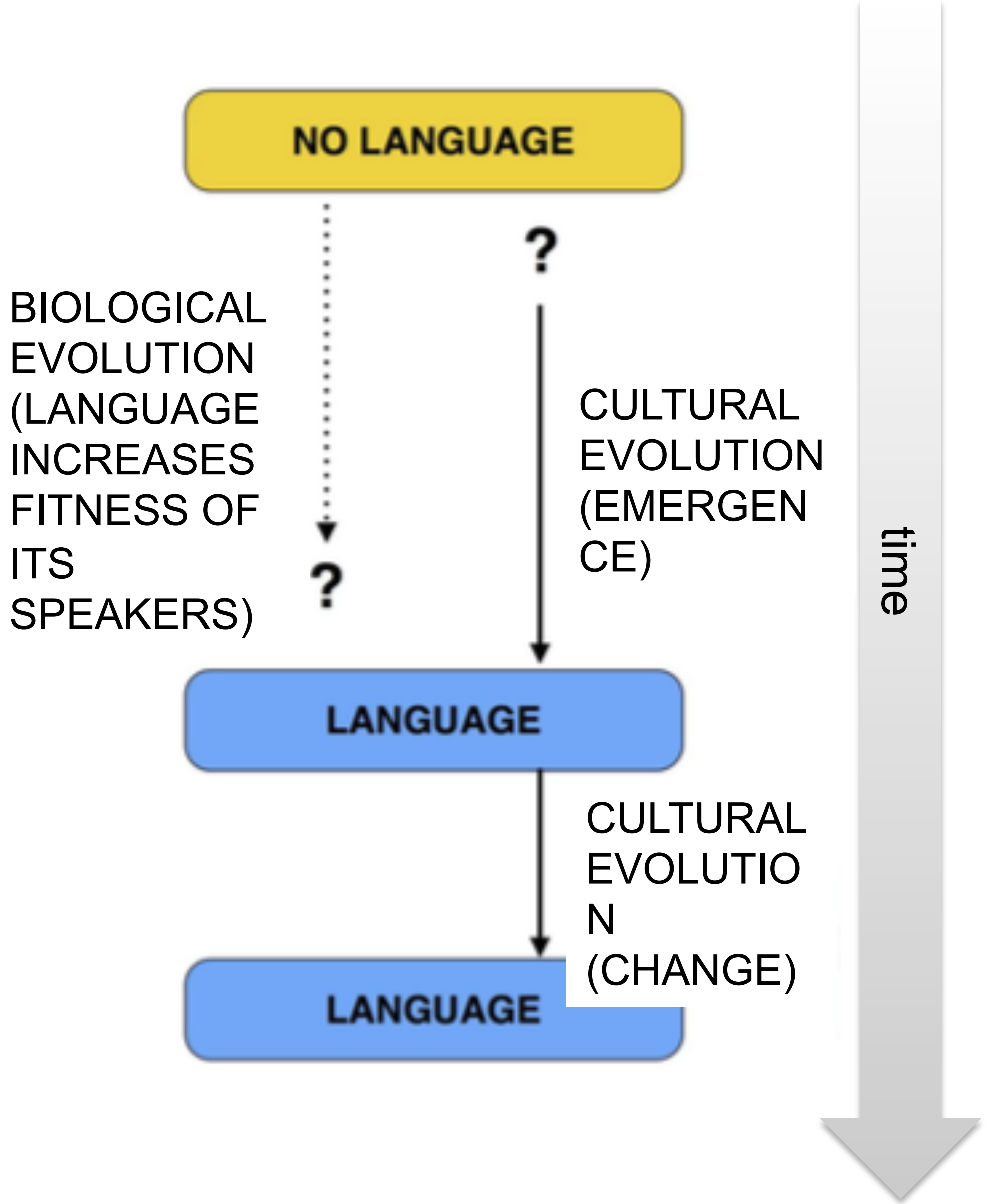
It is widely assumed that human learning and the structure of human languages are intimately related. This relationship is frequently suggested to be rooted in a language-specific biological endowment, which encodes universal, but arbitrary, principles of language structure (a universal grammar or UG). How might such a UG have evolved? We argue that UG could not have arisen either by biological adaptation or non-adaptationist genetic processes. The resulting puzzle concerning the origin of UG we call the *logical problem of language evolution*. Because the processes of language change are much more rapid than processes of genetic change, language constitutes a "moving target" both over time and across different human populations, and hence cannot provide a stable environment to which UG genes could have adapted. We conclude that a biologically determined UG is not evolutionarily viable. Instead, the original motivation for UG—the mesh between learners and languages—arises because language has been shaped to fit the human brain, rather than vice versa. Following Darwin, we view language itself as a complex and interdependent "organism," which evolves under selectional pressures from human learning and processing mechanisms. That is, languages are themselves undergoing severe selectional pressure from each generation of language users and learners. This suggests that apparently arbitrary aspects of linguistic structure may result from general learning and processing biases, independent of language. We illustrate how this framework can integrate evidence from different literatures and methodologies to explain core linguistic phenomena, including binding constraints, word order universals, and diachronic language change.

Christiansen & Chater (2008)

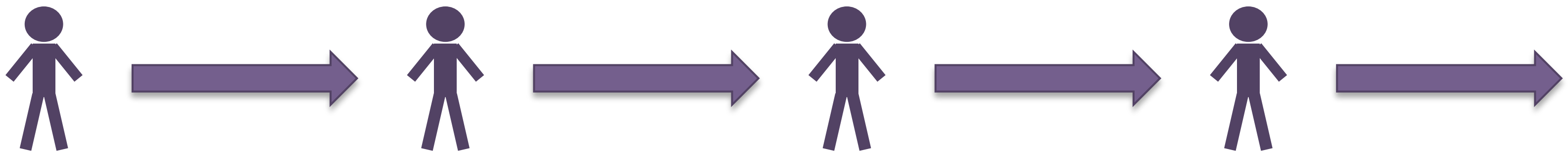
Language = product of cultural evolution.

Language evolution is shaped by constraints on

- learnability
- usage
- transmission



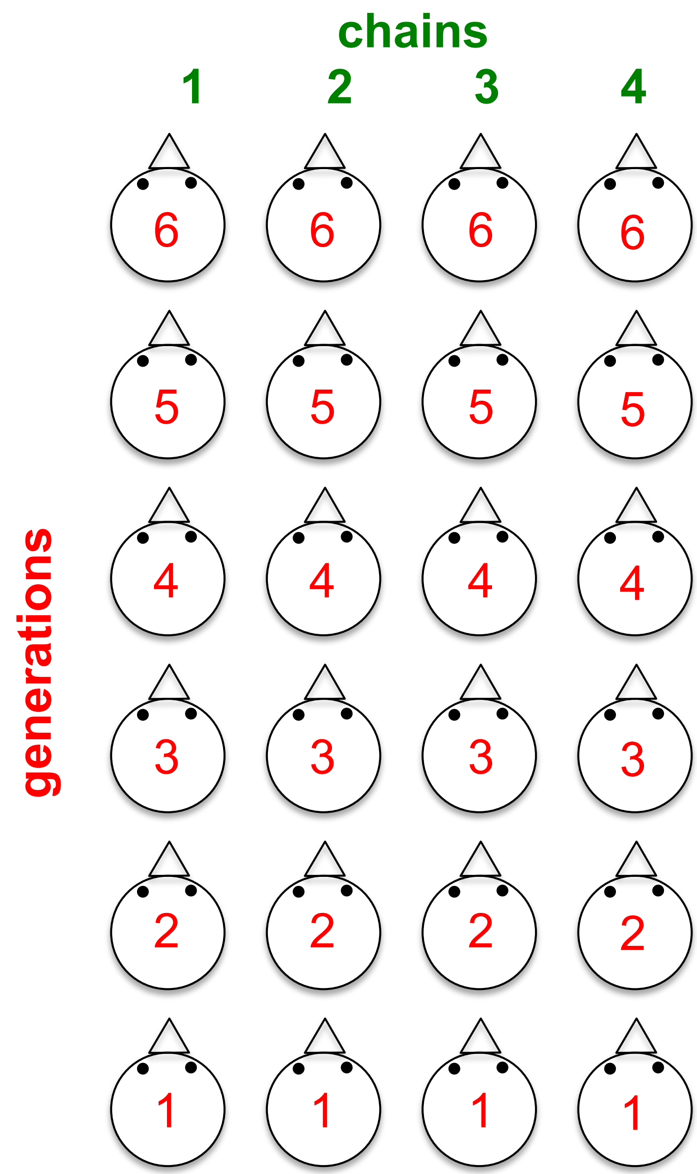
Simulating Cultural Evolution Through Iterated Learning



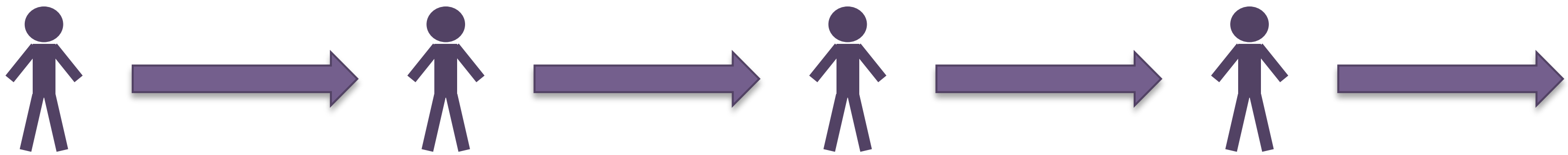
Iterated learning is an experimental paradigm that allows us to study aspects of language transmission in the lab.

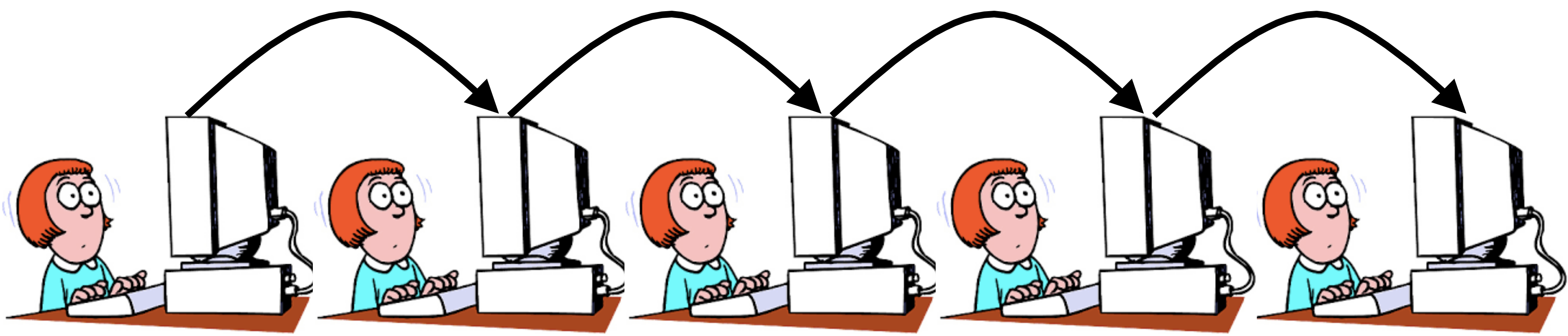
Demo Instructions

1. Assemble into X diffusion chains of Y 'generations'.
2. Get piece of paper and pen; mark paper with chain and generation number & turn it over.
3. Get stopwatch on your smartphone ready.
4. Receive paper with target drawing; look at it for 10 sec. **Then put it away!**
5. Draw what you remember seeing on the paper and hand your drawing to the next person ('generation') in your chain.
6. Hold on to your target drawing until I collect it.



Simulating Cultural Evolution Through Iterated Learning

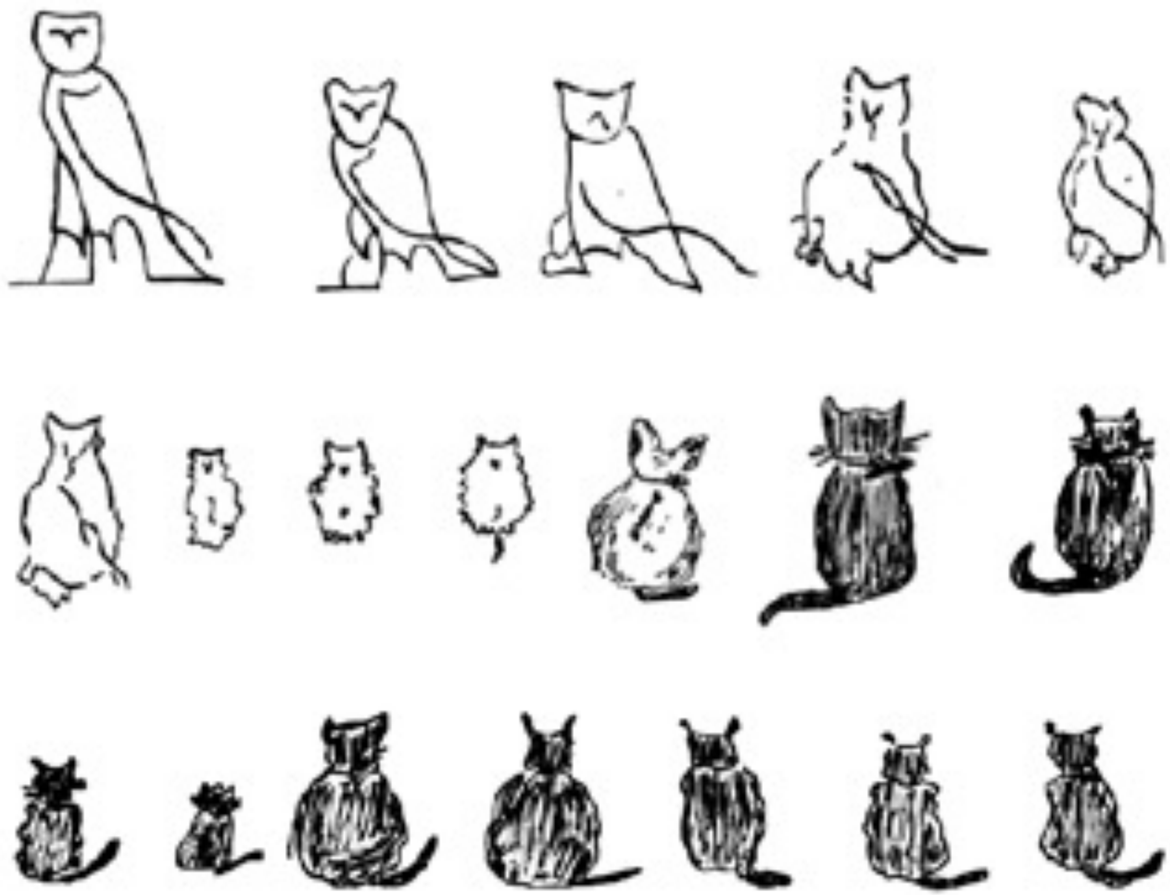




1. **How is the outcome of iterated learning experiments shaped by constraints imposed by different learners, e.g. children?**
2. **What can these experiments tell us about the potential role of different learners in language evolution and language change?**

Bartlett (1932)

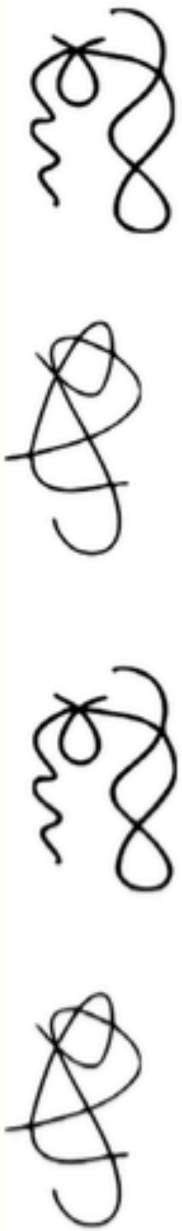
And there are more….



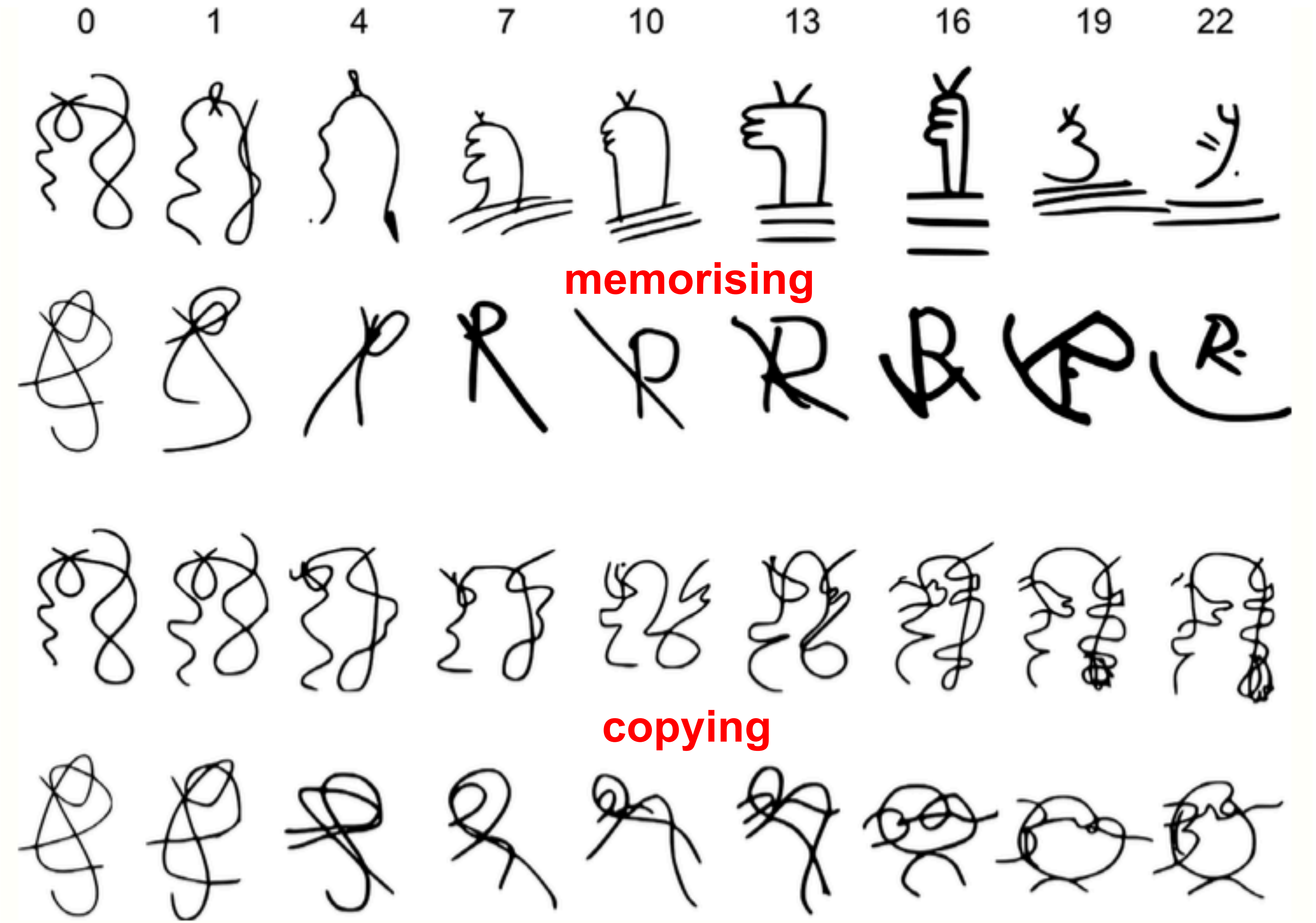
<u>Schema</u> = accepted conventional representation.

Bartlett (1932)

0

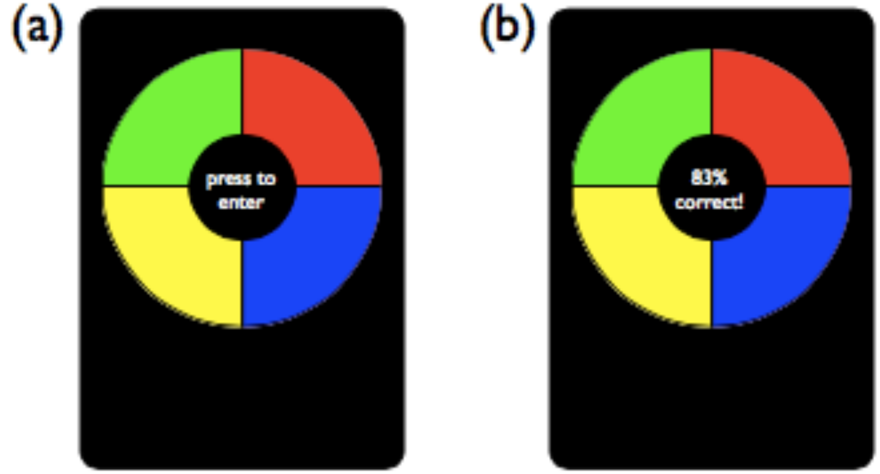


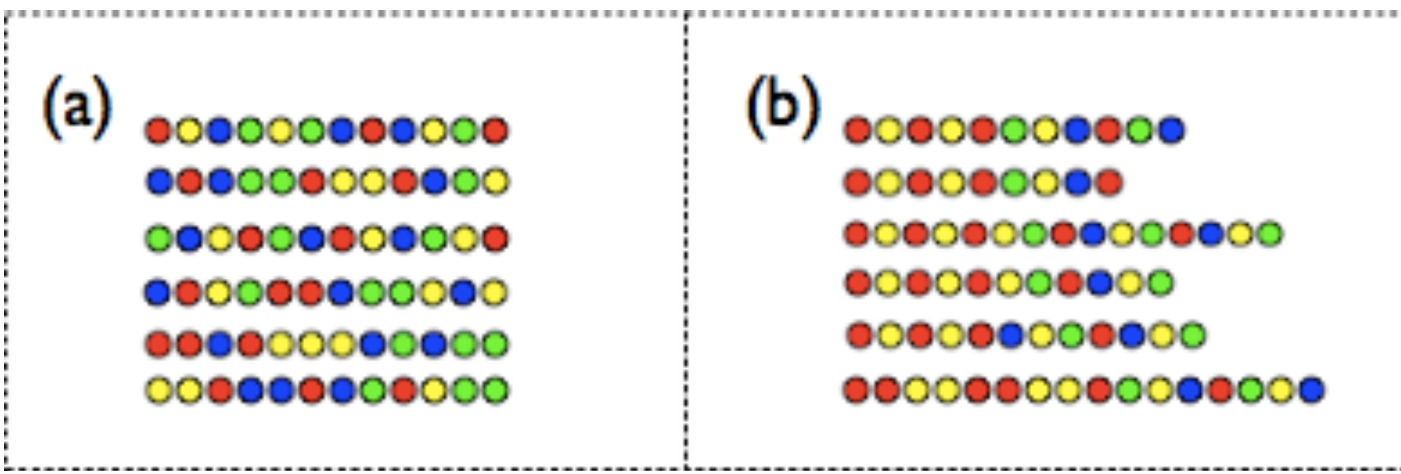
Tamariz & Kirby (2014)



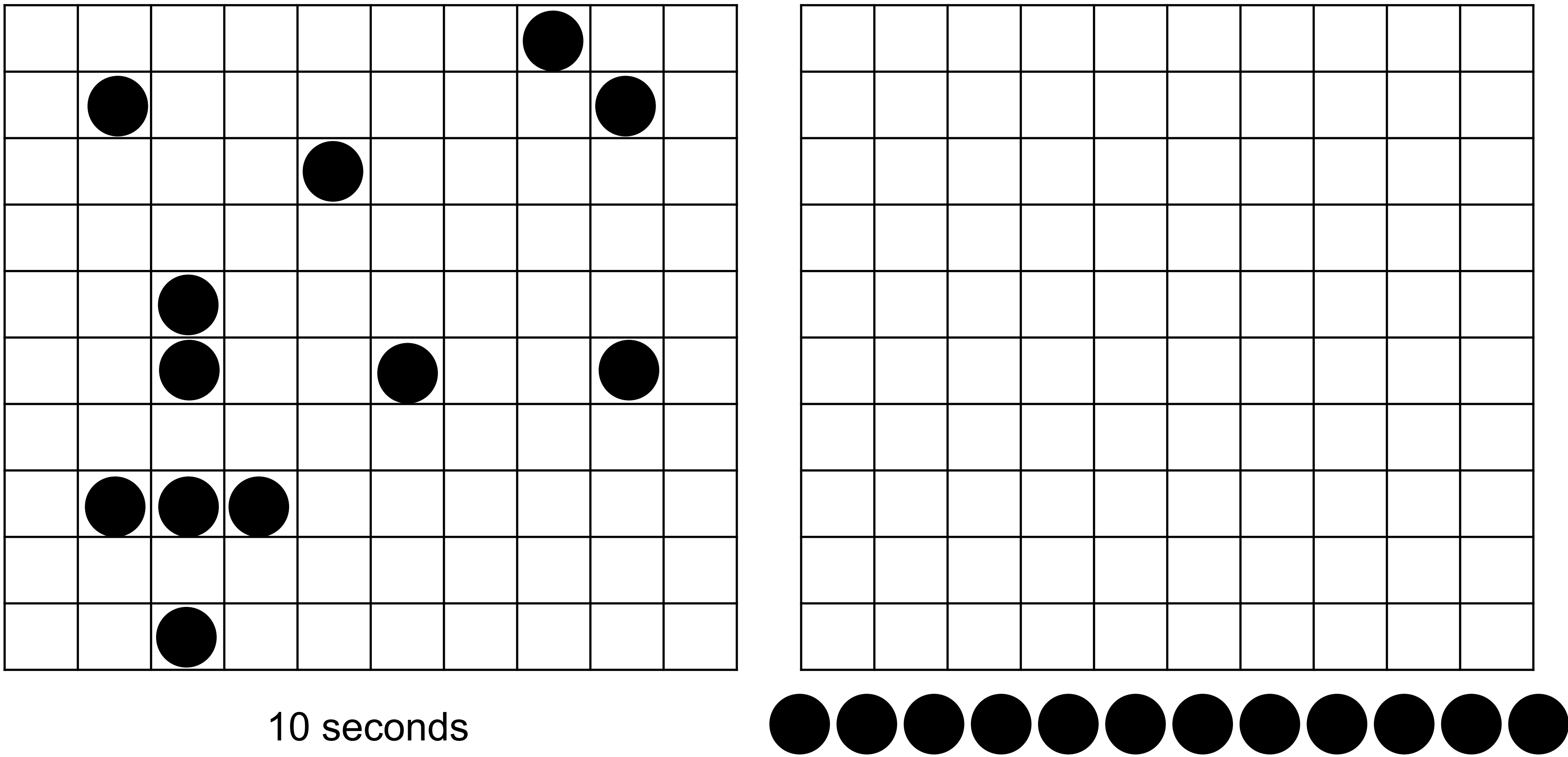
Tamariz & Kirby (2014)

More Examples:



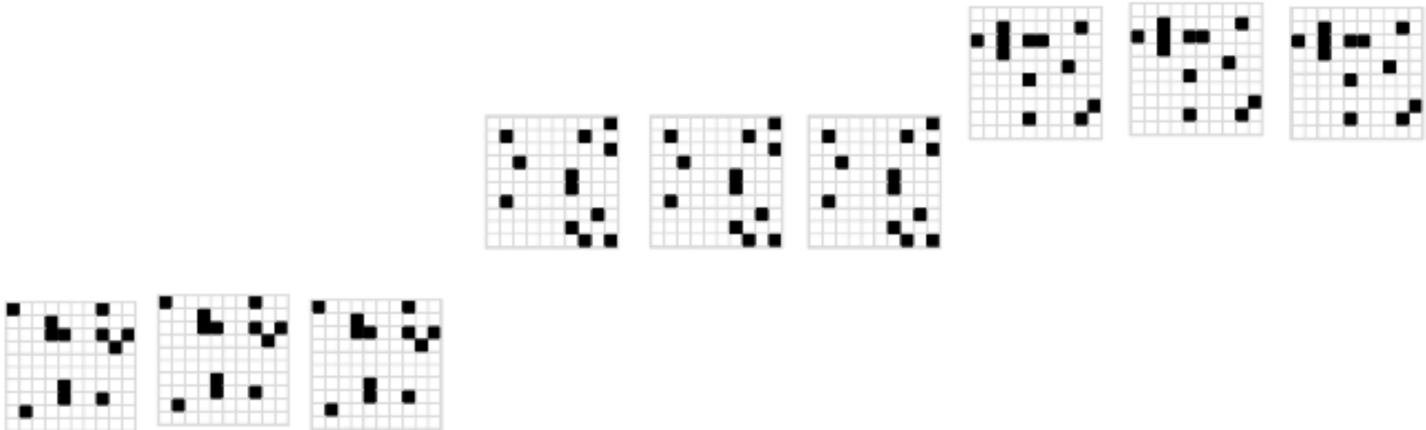


Cornish, Smith & Kirby (2013)

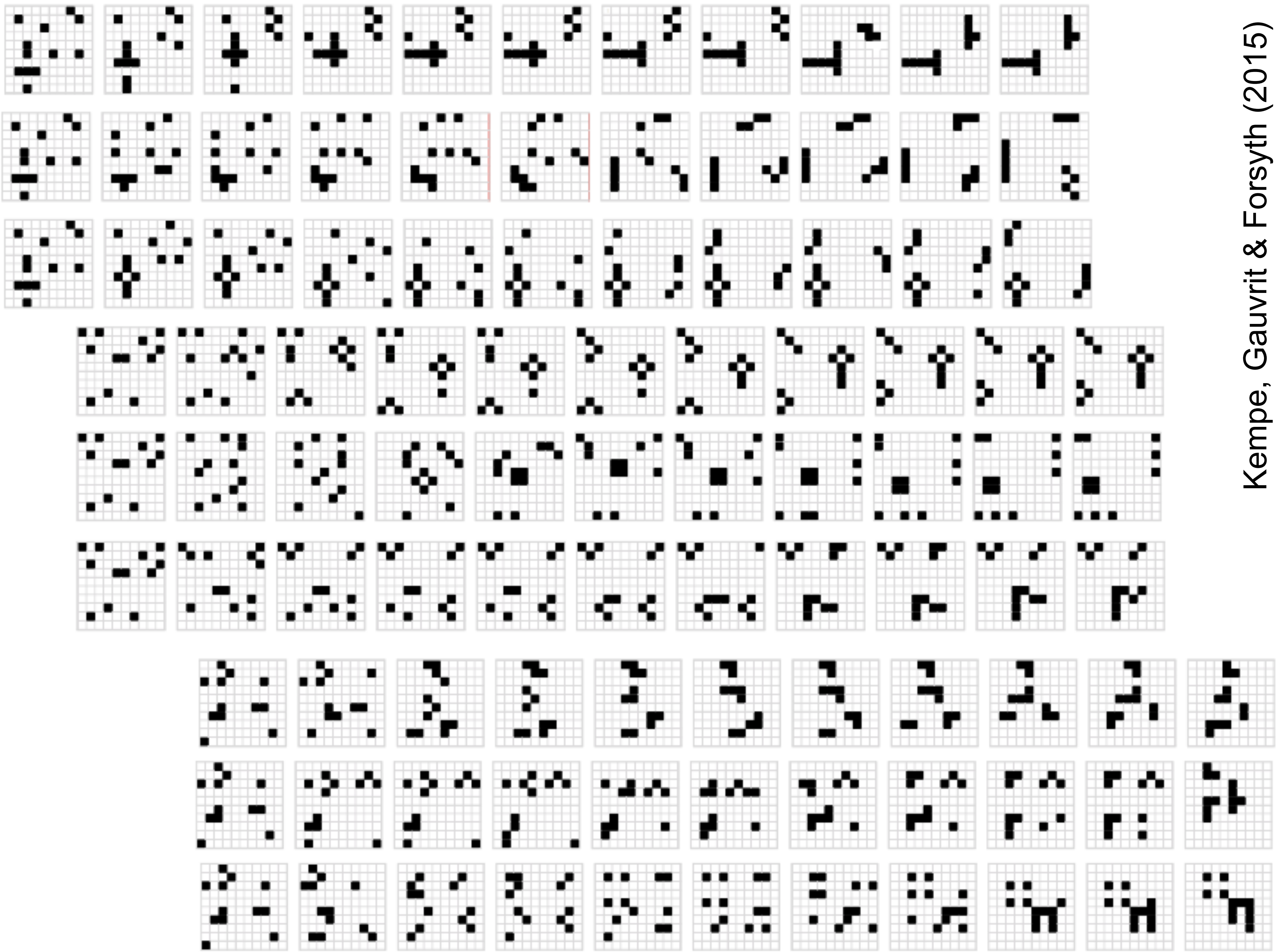


10 seconds

Kempe, Gauvrit & Forsyth (2015)

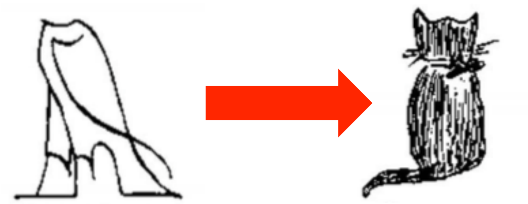


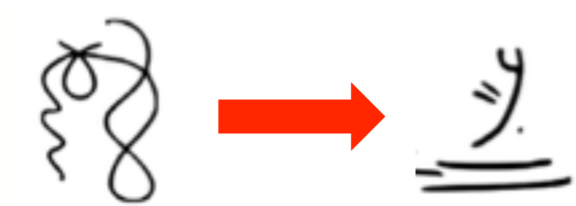
Kempe, Gauvrit & Forsyth (2015)

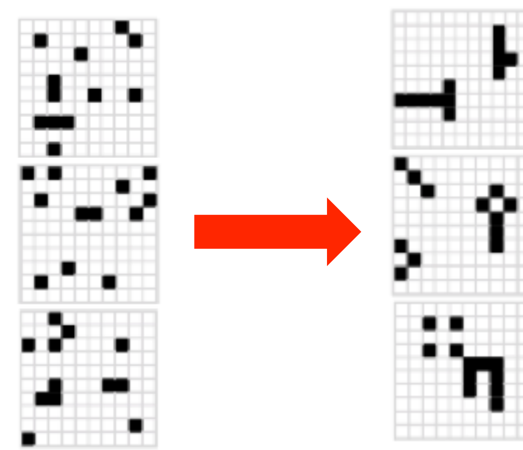


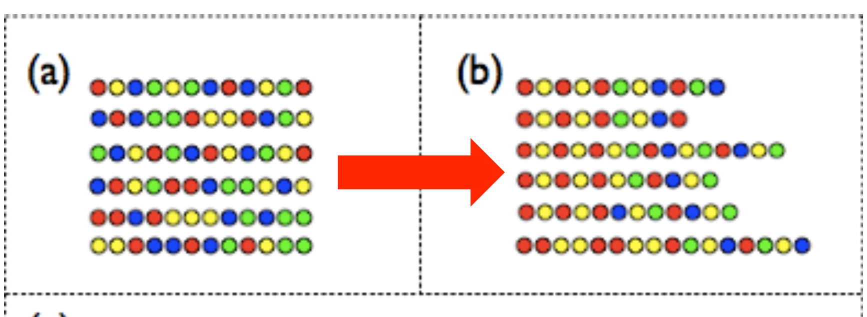
Kempe, Gauvrit & Forsyth (2015)

How to Measure Structure?



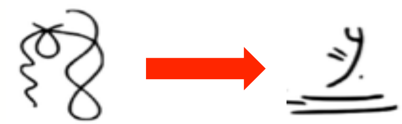


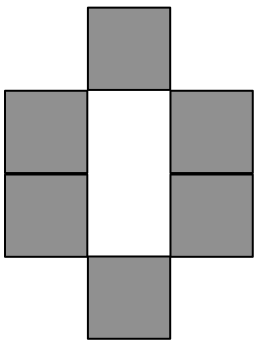




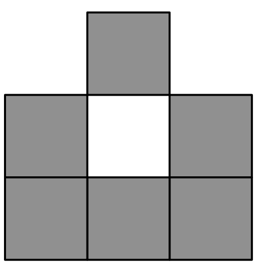
How to Measure Structure?

Perimetric Complexity:
(outer perimeter + inner perimeter)^2/ink area

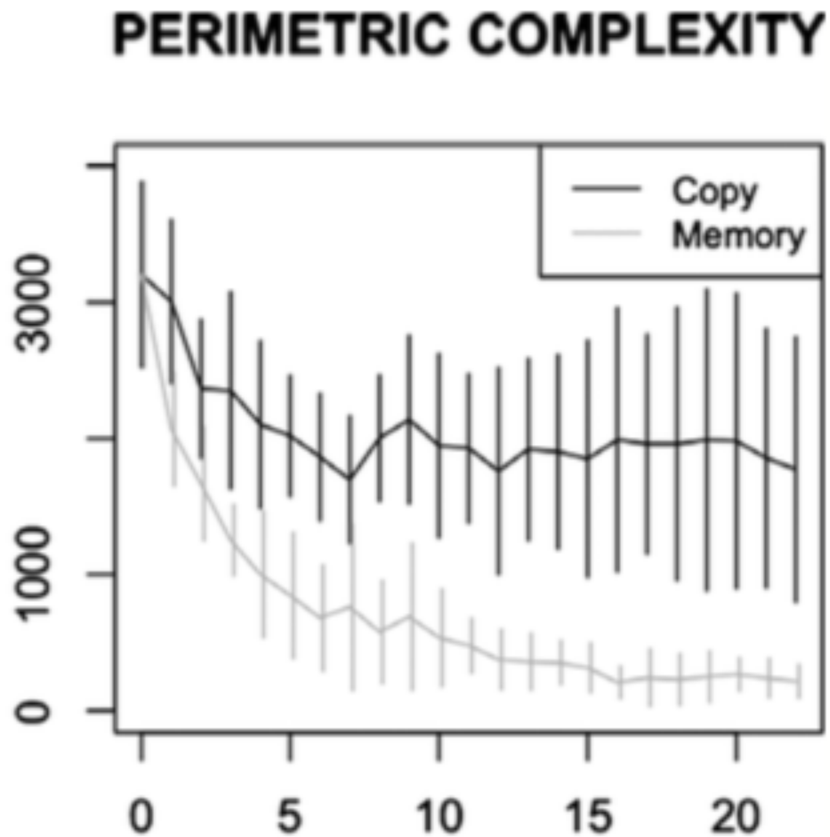




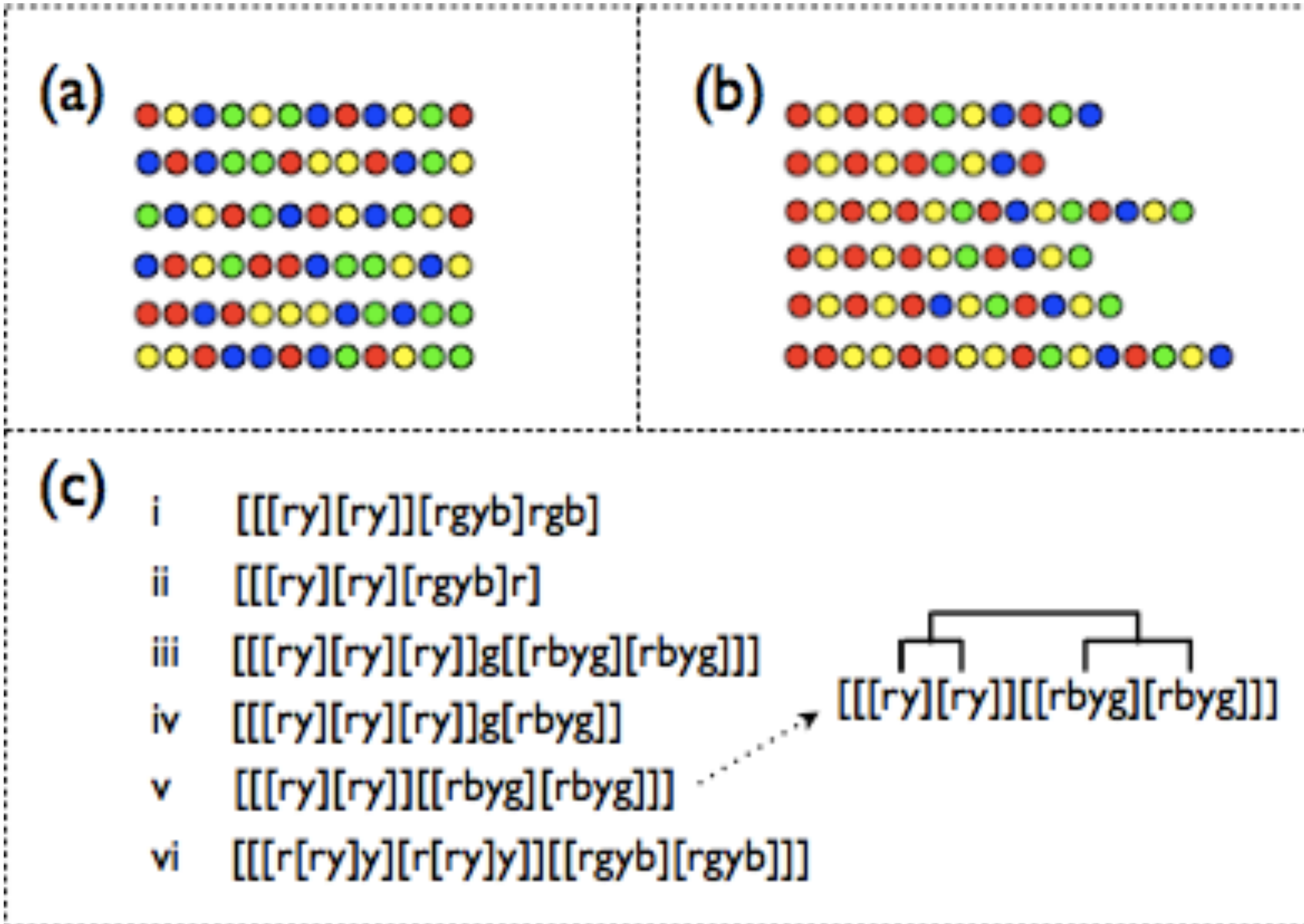
$(14+6)^2/6 = 66.7$



$(12+4)^2/6 = 42.7$

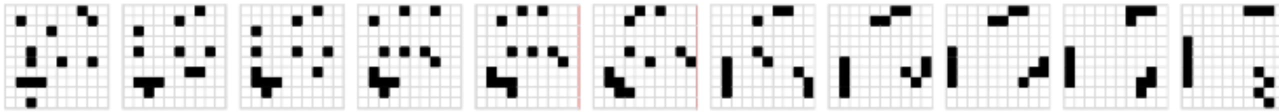


How to Measure Structure?

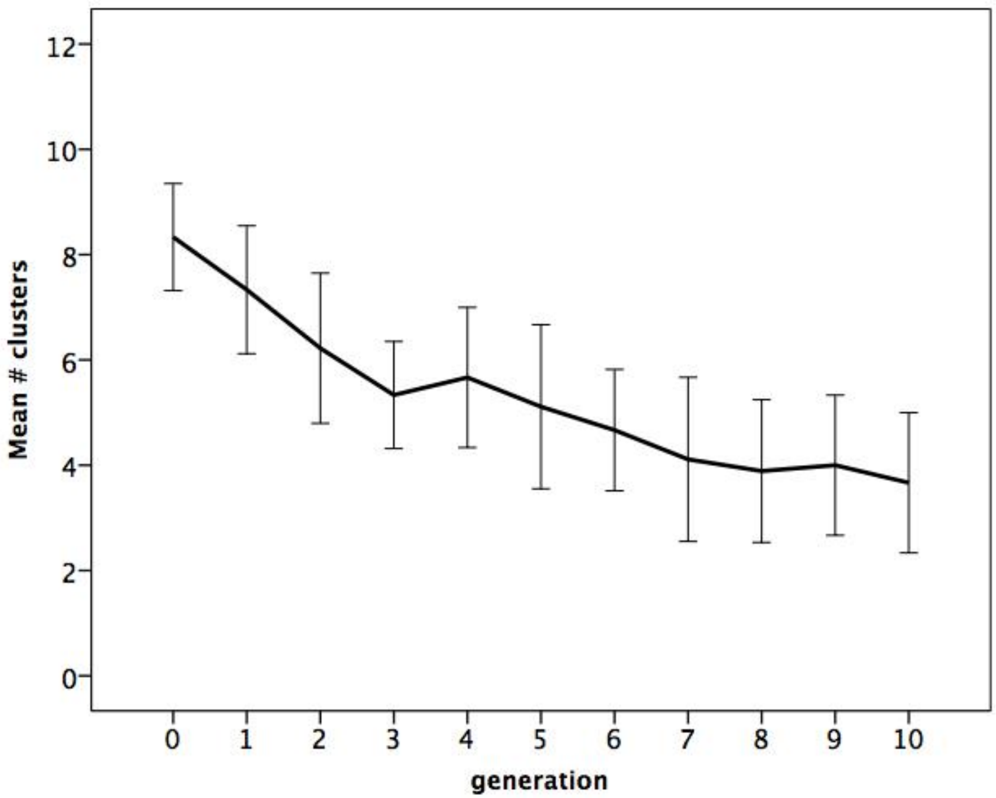


Cornish, Smith & Kirby (2013)

How to Measure Structure?



number of clusters



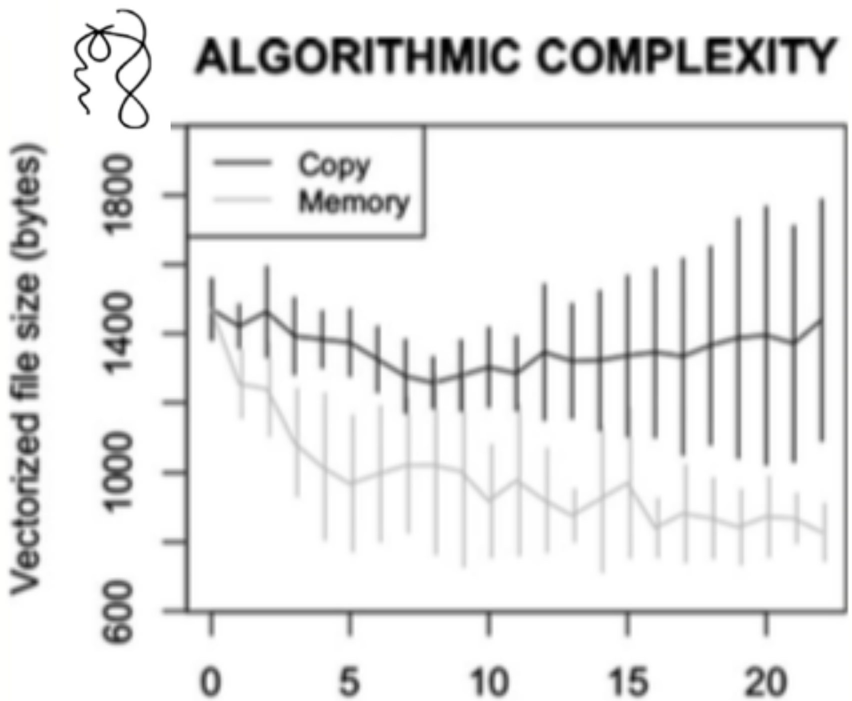
Kempe, Gauvrit & Forsyth (2015)

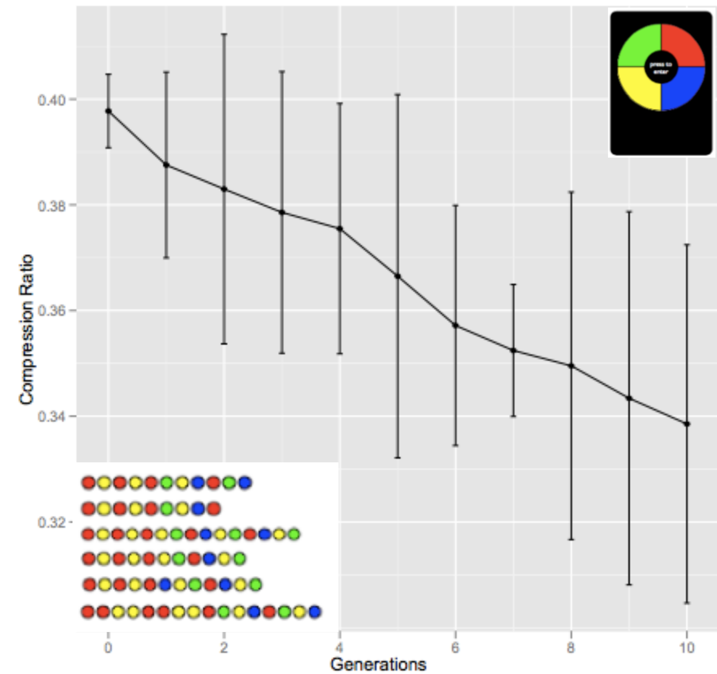
How to Measure Structure?

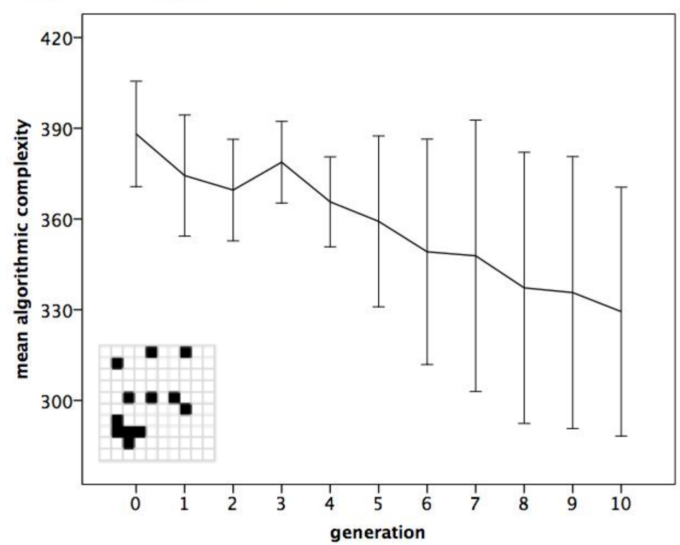
Algorithmic Complexity

- = length of the algorithm required to (re)produce a given stimulus/signal
- depends on underlying representation of the production/ generation mechanism
- proxy of structure (inverse algorithmic complexity): compression (e.g. zip) = looking for amount of redundancy in the stimulus/signal

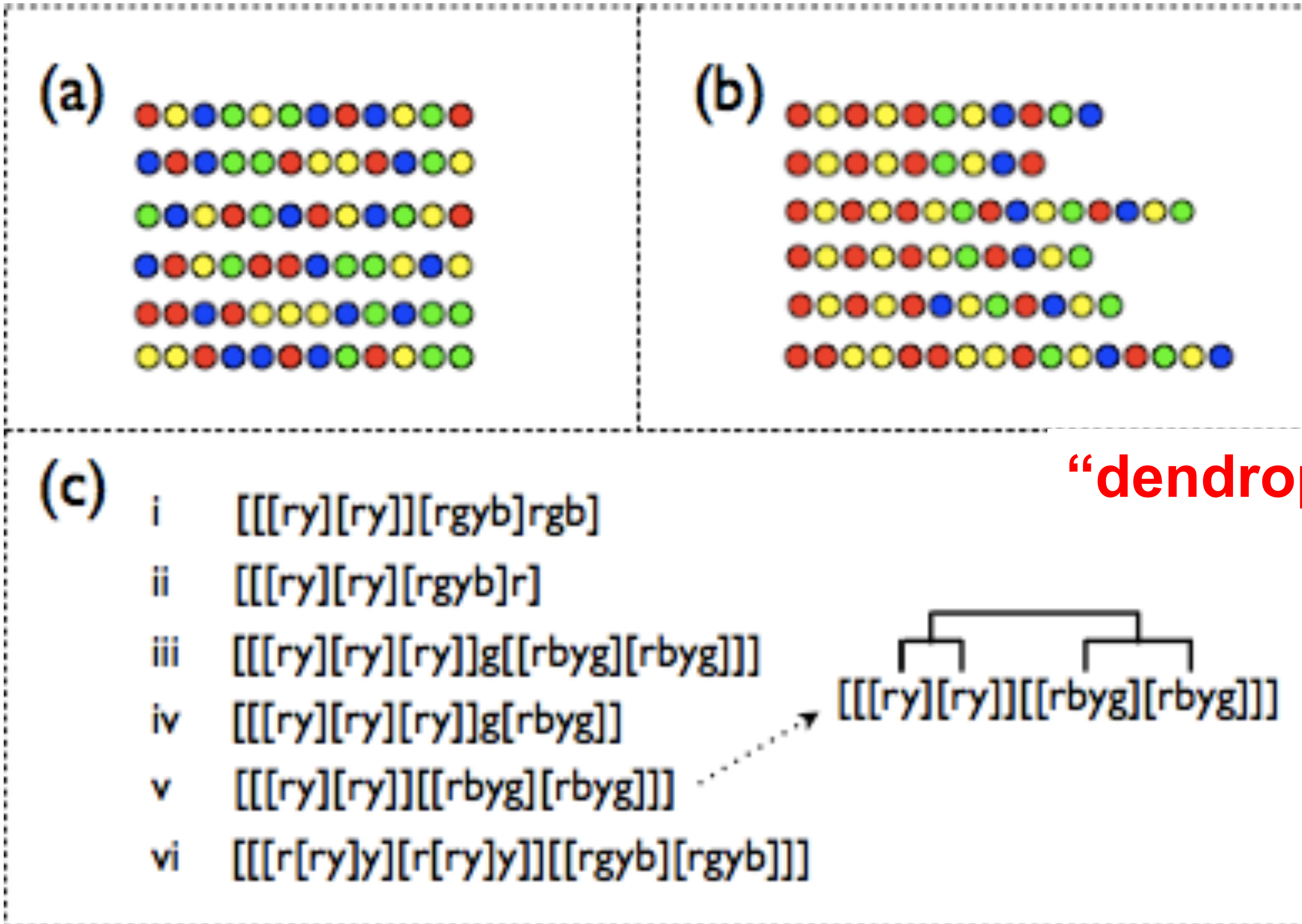
Measures of Structure







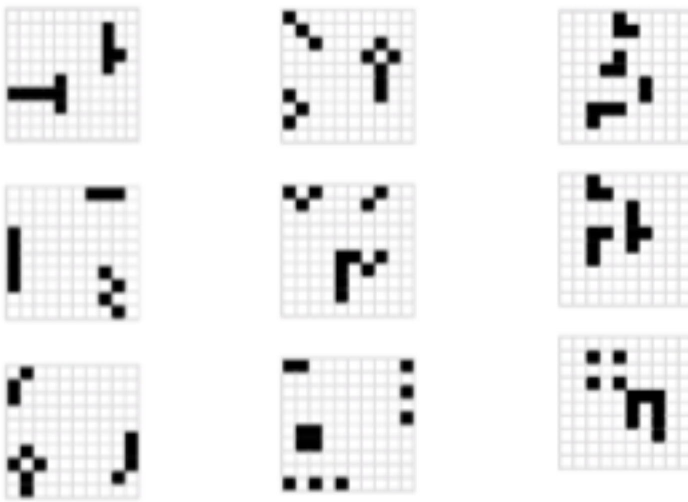
Where Does Structure Come From?



“dendrophilia”?
Fitch (2009)

Cornish, Smith & Kirby (2013)

Where Does Structure Come From?



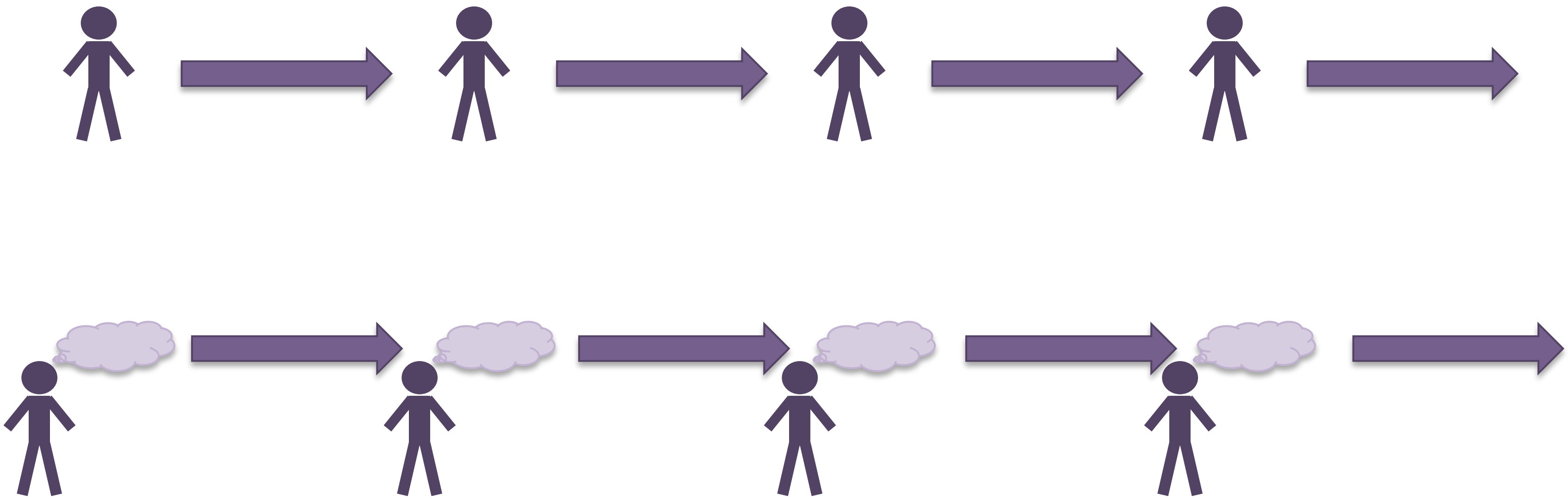
patters produced in final generation (G10)

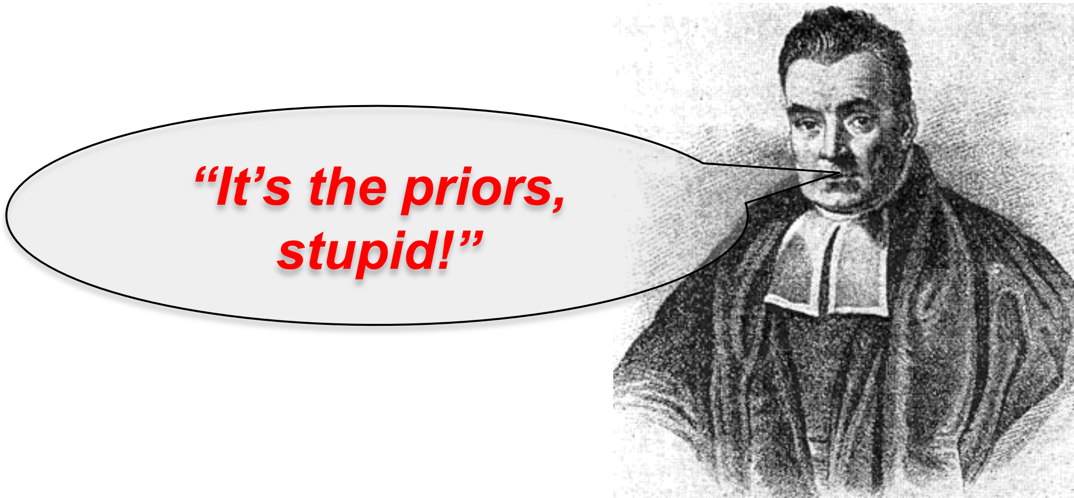
<u>observed schemas:</u>

- T-junctions
- lines
- zig-zags
- crosses
- triangles
- squares
- dotted lines
- corners
- dogs

Kempe, Gauvrit & Forsyth (2015)

Where Does Structure Come From?





You just returned from a summer holiday. Against your better knowledge, you spent a lot of time roasting in the sun. Upon your return you notice a small brown speck on your arm. Worried, you see you doctor who requests a test. While you are waiting, your doctor gives you the following information:

The probability that the test comes back positive if someone has cancer:

$p(\text{test+}|\text{cancer}) = .9$ **hit rate**

The probability that the test comes back positive if someone does not have cancer:

$p(\text{test+}|\text{no cancer}) = .2$ **false alarm rate**

The probability that someone has this type of cancer:

$p(\text{cancer}) = .01$ **base rate**

Your test came back positive.

What is the probability that you have cancer?

Belief Updating Based on Evidence: Bayes' Theorem

likelihood

prior (aka base rate)

$$p(H|D) = \frac{p(D|H) \times p(H)}{p(D)}$$

posterior

H = your hypothesis, theory, assumption, belief
D = the data

Belief Updating Based on Evidence: Bayes' Theorem

The extended form: partitioning up p(D)

base rate

$$p(H|D) = \frac{p(D|H) \times p(H)}{p(D|H) \times p(H) + p(D|\neg H) \times p(\neg H)}$$

hit rate

false alarm rate

H = your hypothesis, theory, assumption, belief

D = the data

Applying Bayes' Theorem:

$$p(\text{cancer}|\text{test+}) = \frac{p(\text{test+}|\text{cancer}) \times p(\text{cancer})}{p(\text{test+}|\text{cancer}) \times p(\text{cancer}) + p(\text{test+}|\text{no cancer}) \times p(\text{no cancer})}$$

$$\frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.2 \times 0.99} = \frac{0.009}{0.009 + 0.198} = 0.043$$

p = 0.043 = 4.3%

Reproduction / Learning as Bayesian Inference

Task: Extracting and storing information from a noisy signal.

How? Inference of what hypothesis of the state of the world (H) to extract and store based on the perceived data (D)

prior

$$p(H|D_x) = \frac{p(D_x|H) * p(H)}{p(D_{all})}$$

Reconstruction = compromise between noise in the data and uncertainty in the prior distribution.

Bayesian Inference

<u>**prior:**</u> **p of H in general = best understood as how much evidence learners need to adopt a particular H** (abstract computational-level approach agnostic to the nature and content of biases)

<u>**likelihood**</u>**: probability of observing a set of data if this particular hypothesis H holds true**

$$p(H|D) = \frac{p(D|H) \times p(H)}{p(D)}$$

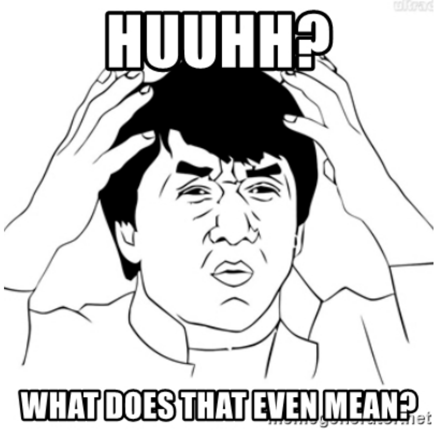
<u>**posterior:**</u> **probability of the H given the data**

probability of data averaged over all possible Hs

H = hypothesis about how to generate the data

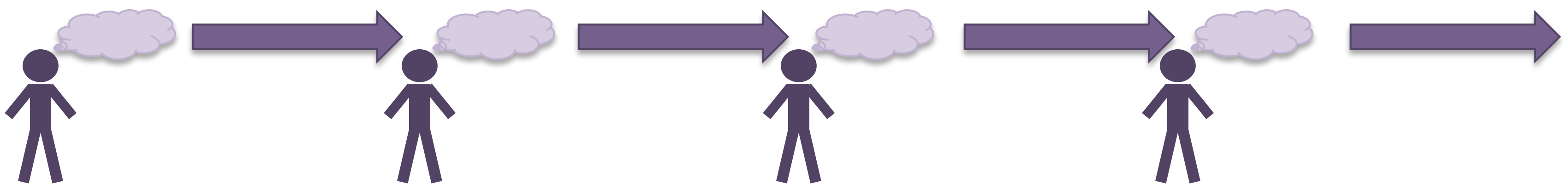
D = data

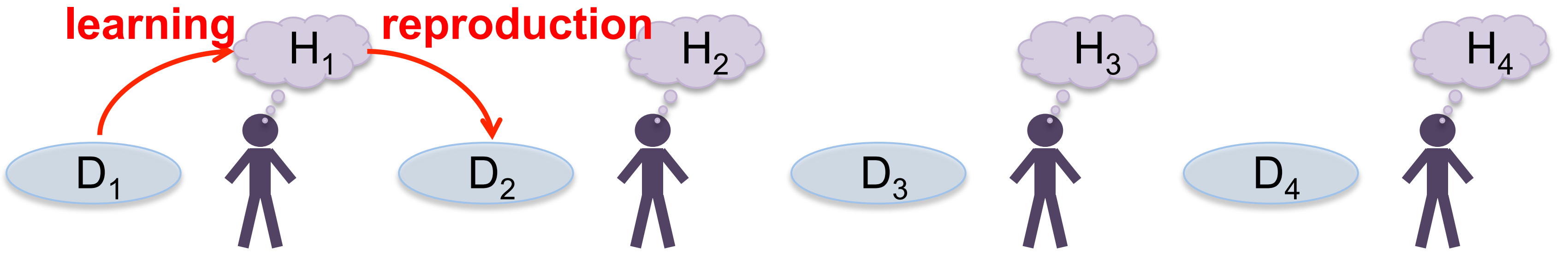
Griffiths & Kalish (2007)

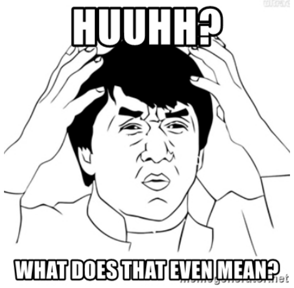


- Observers arrive at a **posterior probability** of a hypothesis given the data they have observed which -- according to Bayes’ Rule is -- dependent on the prior.

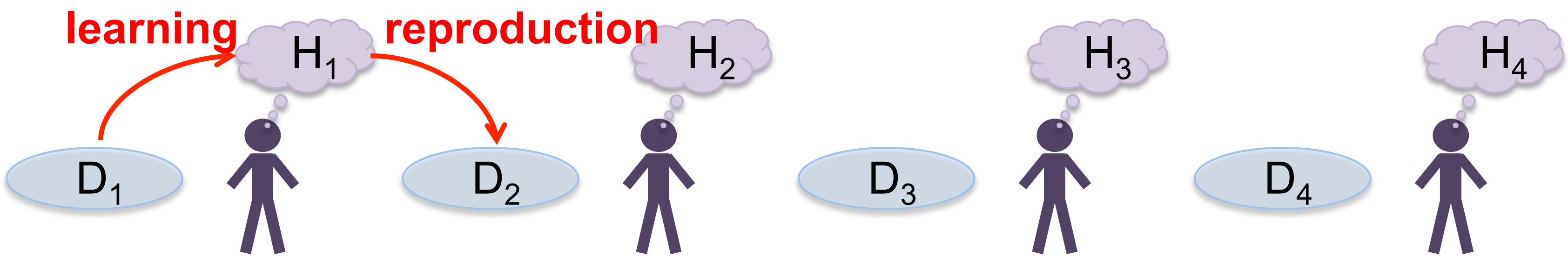
Iterated Reproduction



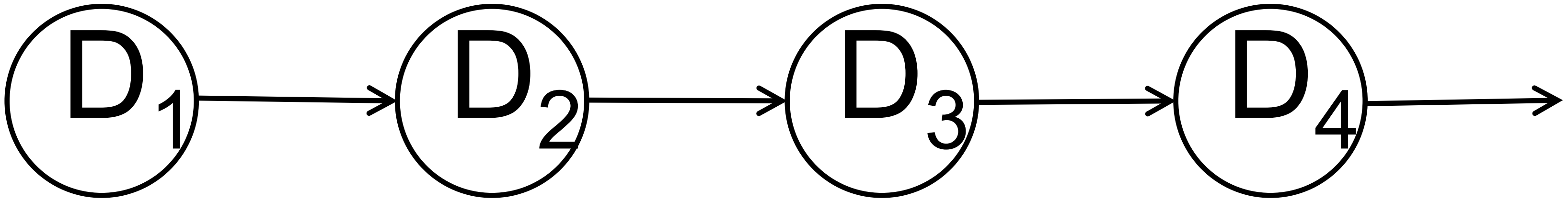




- Observers arrive at a **posterior probability** of a hypothesis given the data they have observed which -- according to Bayes' Rule is -- dependent on the prior.
- In iterated reproduction, observers then **sample** (i.e. probability-match) from the posterior probability distribution to generate the output for the next observer.



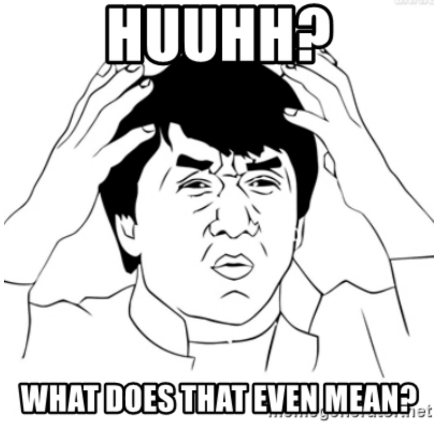
Iterated Reproduction as Markov Chain



$$p\,(D_t|D_{t-1}) = p(D_t|H) \times p(H|D_{t-1})$$

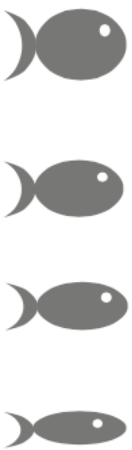
Reproduction of a stimulus D_t depending on the previous stimulus D_{t-1} is based on a combination of the prior and the previous reconstruction.

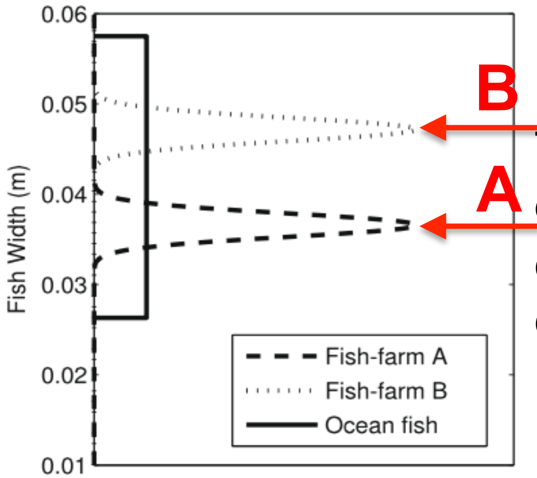
Griffiths & Kalish (2007)



- Observers arrive at a **posterior probability** of a hypothesis given the data they have observed which -- according to Bayes' Rule is -- dependent on the prior.
- Once observers have obtained the posterior probability, they **sample** (i.e. probability-match) from the posterior probability distribution to generate their output for the next observer.
- This sampling draws from the **combined distribution** of the probability of the current output state given the prior and the posterior determined from the previous input state.
- As the stationary distribution of the Markov chain is the prior (for proof see G&K, 2007), iteration of this process over time leads to **convergence of the generated state to the prior**.
- **Thus, the prior is exerting its influence on every single iteration while the data seen, or the hypotheses generated, by each observer are only one small piece of information.**

Iterated Reconstruction / Learning

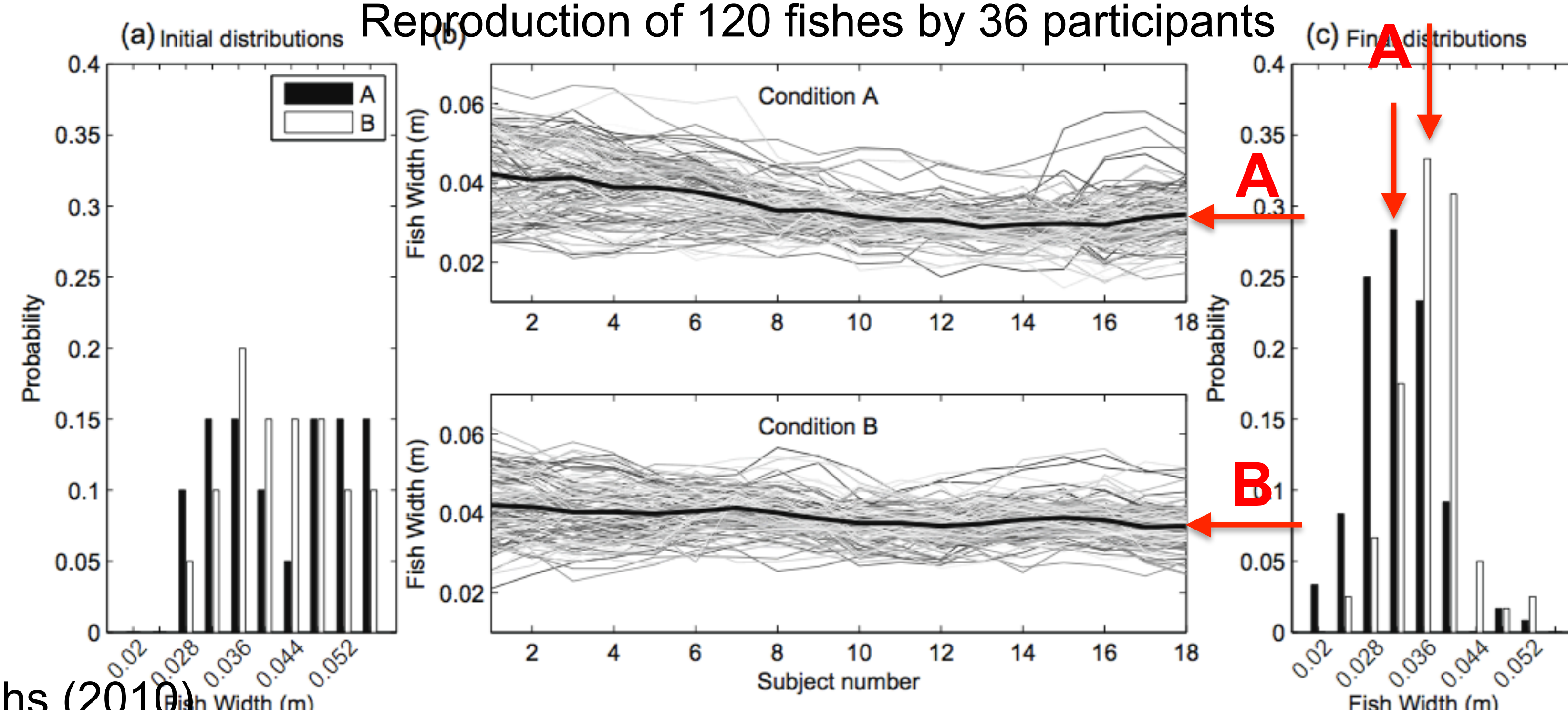




Training of two groups on distributions with different means.

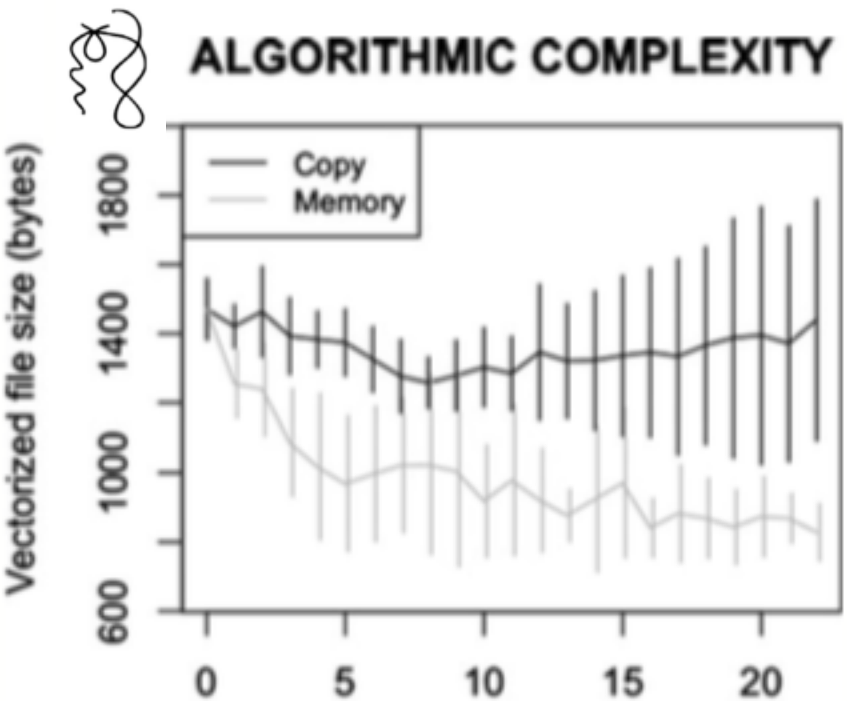
Results:

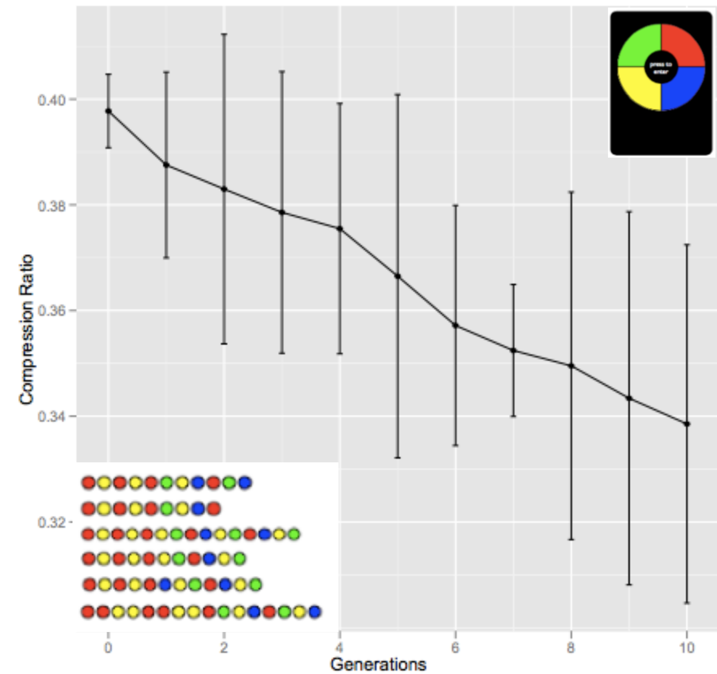
Reproduction of 120 fishes by 36 participants



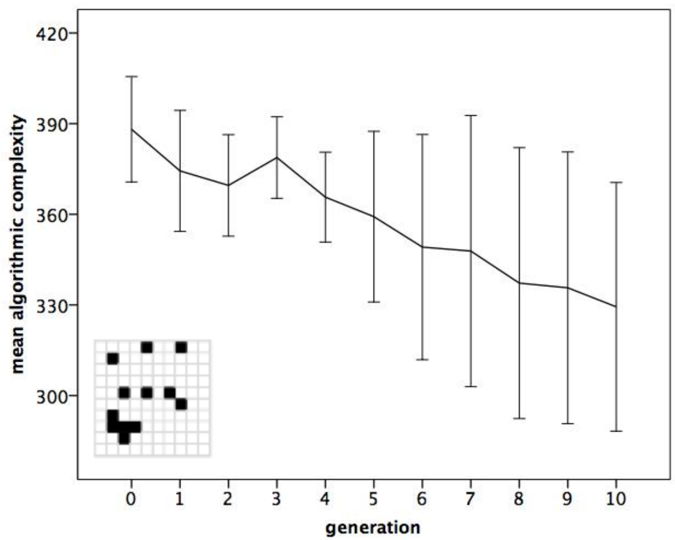
Xu & Griffiths (2010)

Amplification of Initial Biases

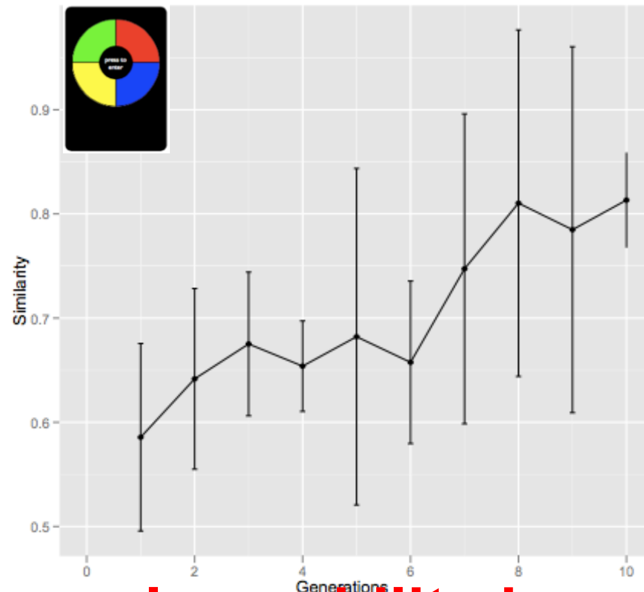




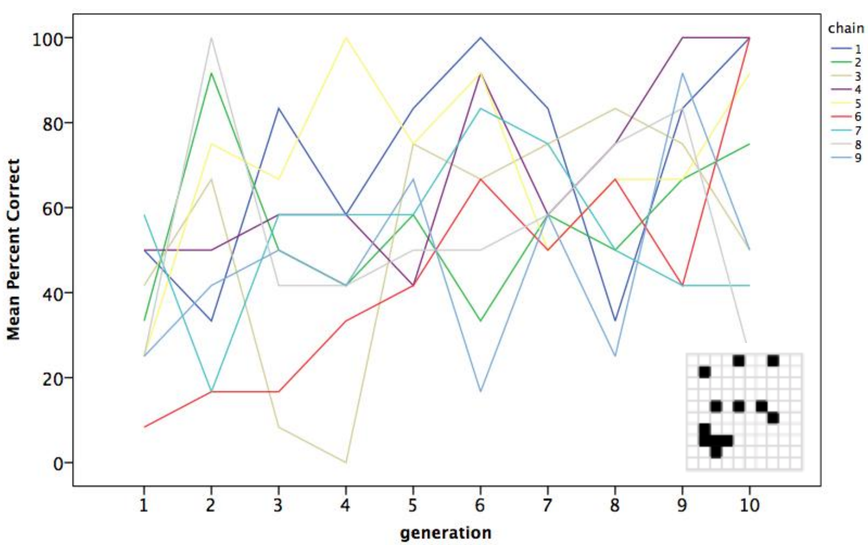
Initial biases towards structural simplicity (compressibility) are being amplified.

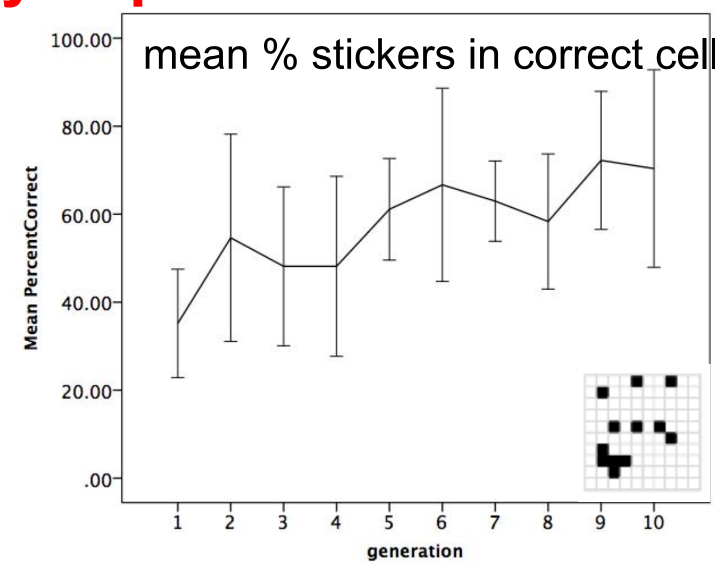


Transmission Fidelity / Learnability



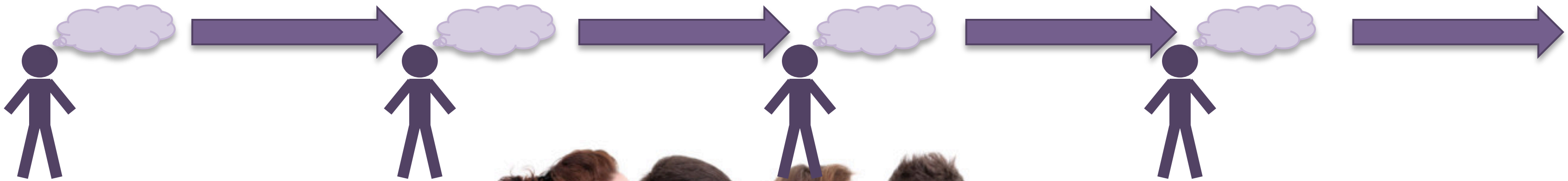
As structure increases, learnability improves.

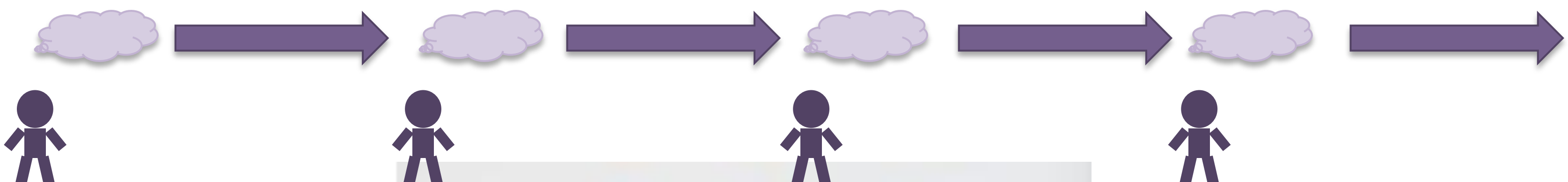




Summary: Adults

- Cultural transmission can be studied experimentally as a process of iterated reproduction/learning.
- Reproduction is biased: In experiments, participants show prior biases in favour of more compressibility/ structure.
- If reproduction and learning are viewed as Bayesian inference then iteration leads to convergence to these priors. As a result, weak biases get amplified.
- As structure increases learnability increases too.





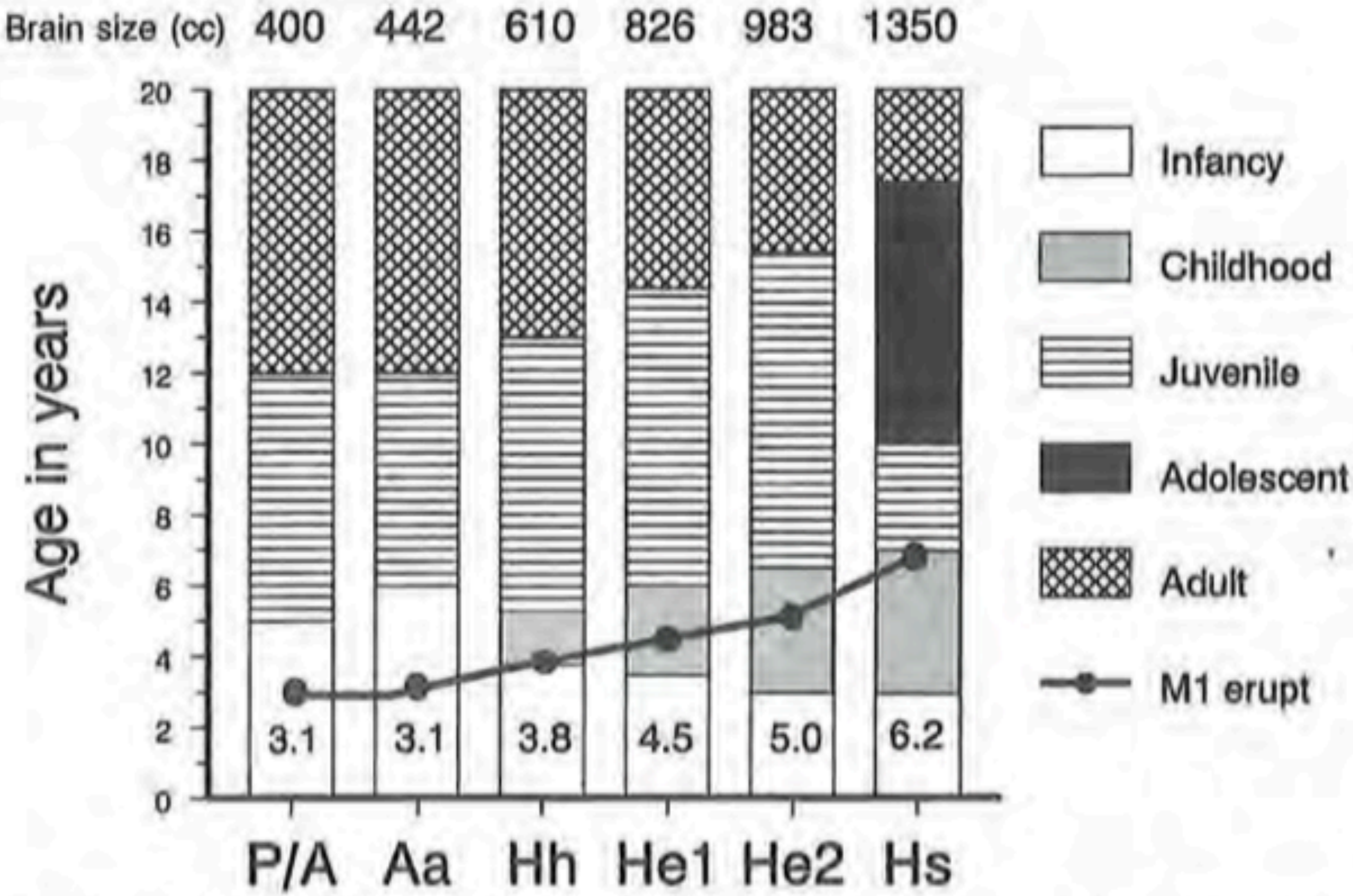
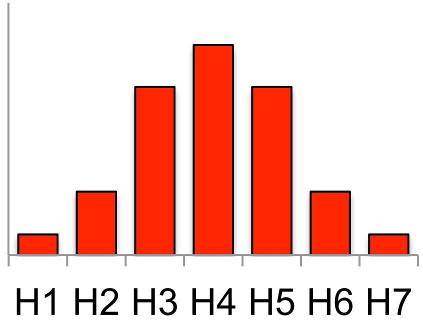



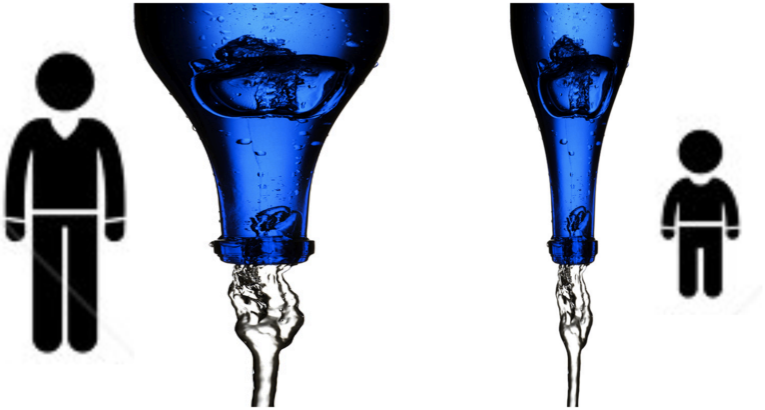
Figure 11.24 The evolution of hominin life history during the first 20 years of life. Abbreviation of the pongid and hominin taxa are P/A, *Pan, Australopithecus afarensis*; Aa, *Australopithecus africanus*; Hh, *Homo habilis*; He1, early *Homo erectus*; He2, late *Homo erectus*; Hs, *Homo sapiens*. *Source*: Bogin (1999).

Bogin & Smith (2000)

Children: What's Different?

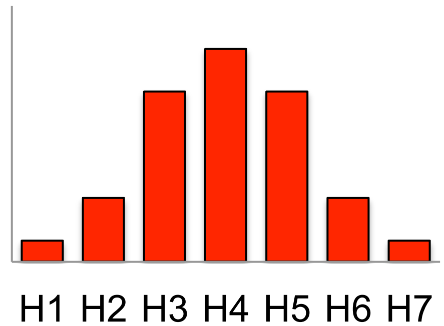


- Do children have biases / priors?
  - weaker biases?
  - stronger biases?
  - different biases?
- Effects of lower cognitive capacity
  - on learning?
  - on reproduction?

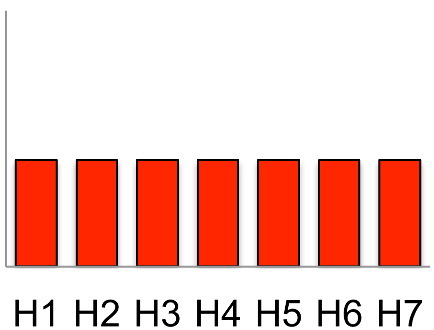


Possible Differences in Children's Priors

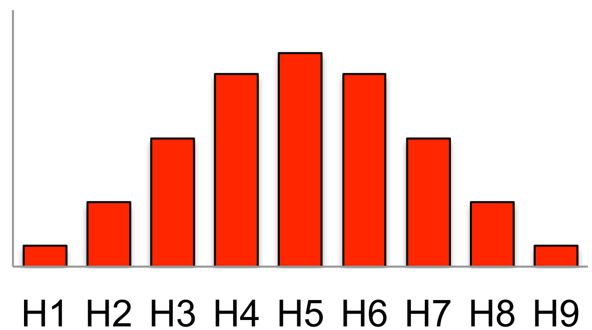
adults



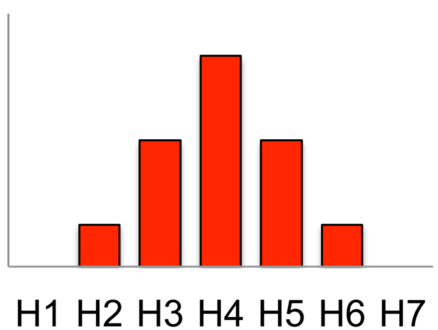
children



flat priors?



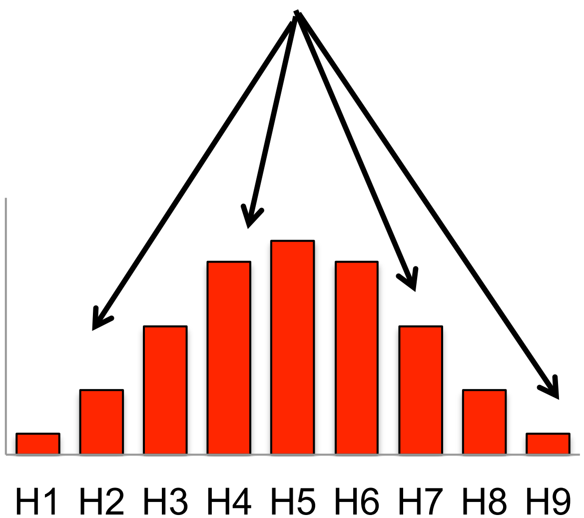
more hypotheses?



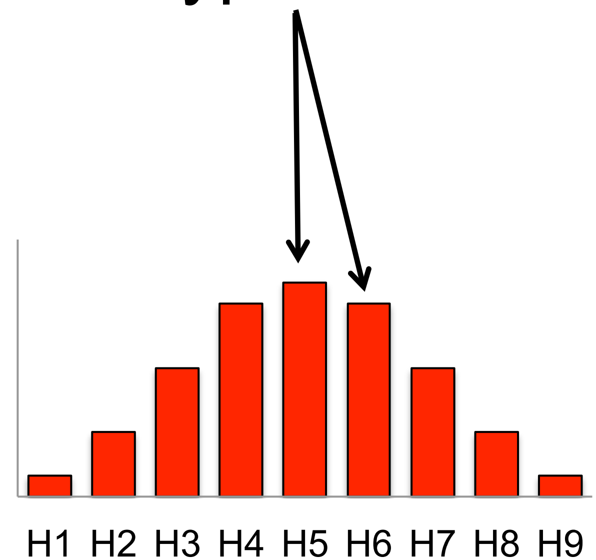
fewer hypotheses?

Differences in Children's Learning

Sampling from the distribution of hypotheses:



broad, "high temperature" search



narrow, "low temperature" search

Gopnik et al. (2017)

Hypothesis Search Across the Lifespan

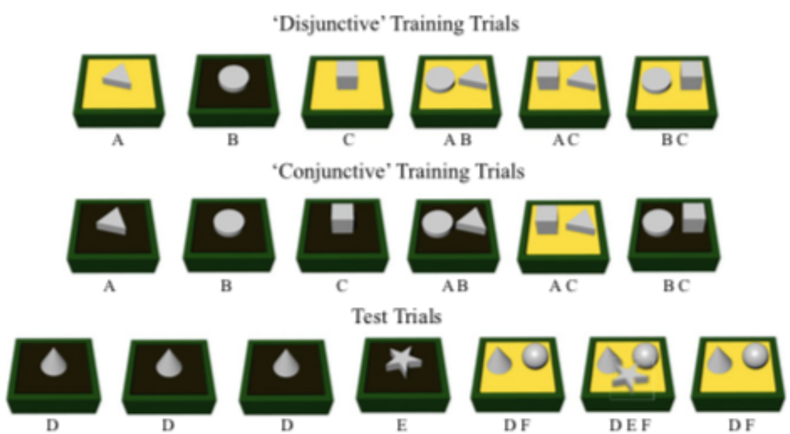


Fig. 1. Schematic of the procedure for Exp. 1. The yellow rectangle represents the machine's activation. "Disjunctive" training provides evidence of the more common, disjunctive hypothesis. "Conjunctive" training provides support for the less common conjunctive hypothesis. "Test" trials presented ambiguous evidence about the "D" object.

How likely are Ps to try object combinations?

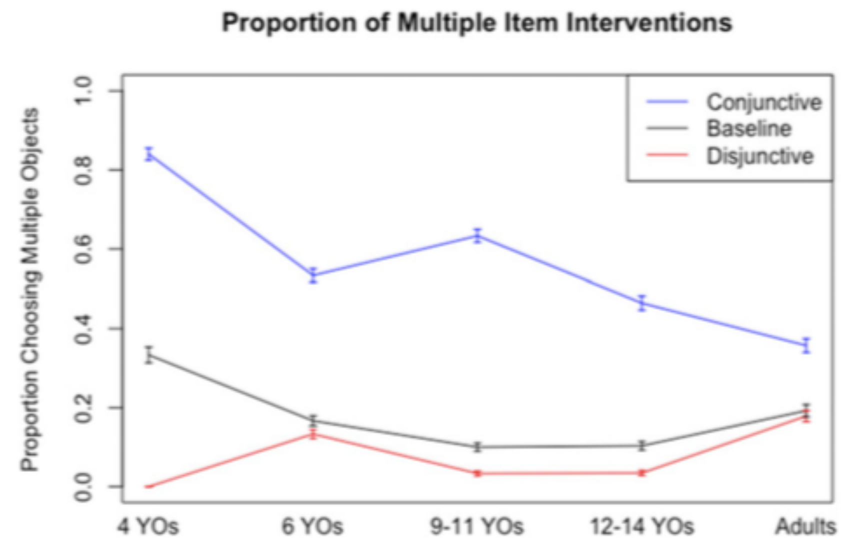


Fig. 3. Proportion of participants choosing either single or multiple items for intervention choice with SEs.

Gopnik et al. (2017)

Hypothesis Search Across the Lifespan

How likely are Ps to try blame the social context rather than a protagonist's traits (i.e. not committing the 'Fundamental Attribution Error?

Adolescents revise social attributions!

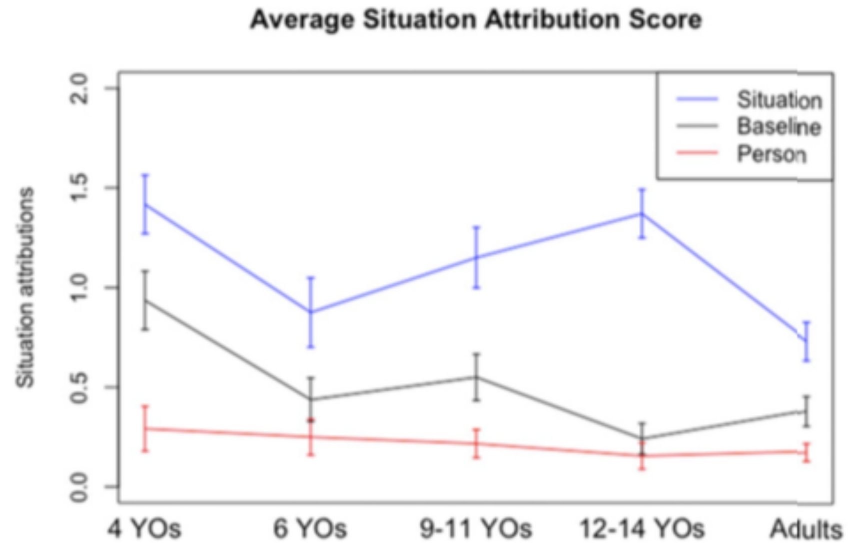
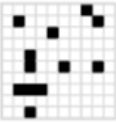


Fig. 4. Average attribution scores by age group and condition with SEs. YO, year old.

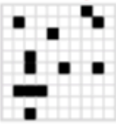
Gopnik et al. (2017)

Iterated Reproduction in Children

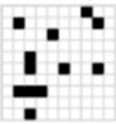
Primary 1



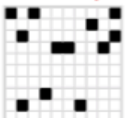
Primary 2



Primary 3

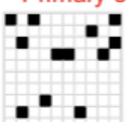


Primary 1

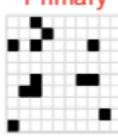


Primary 2

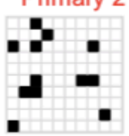
Primary 3



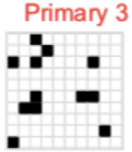
Primary 1



Primary 2

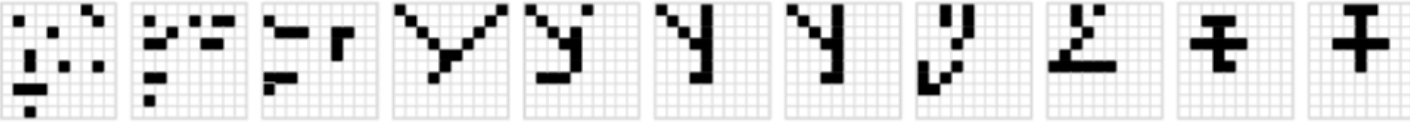


Primary 3

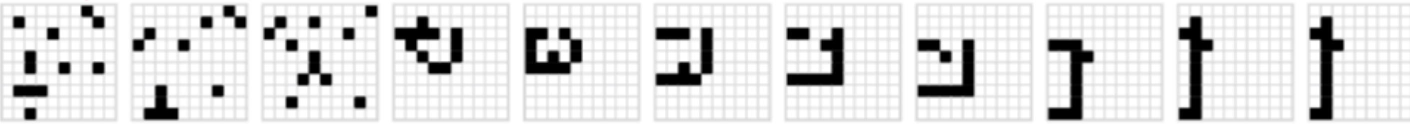


Children:

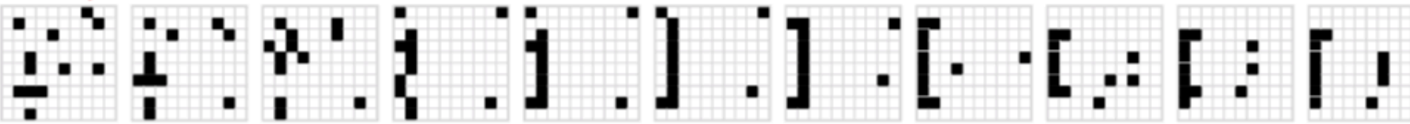
Primary 1



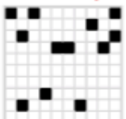
Primary 2



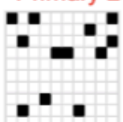
Primary 3



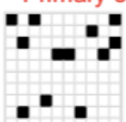
Primary 1



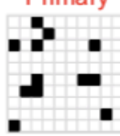
Primary 2



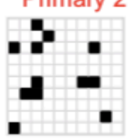
Primary 3



Primary 1



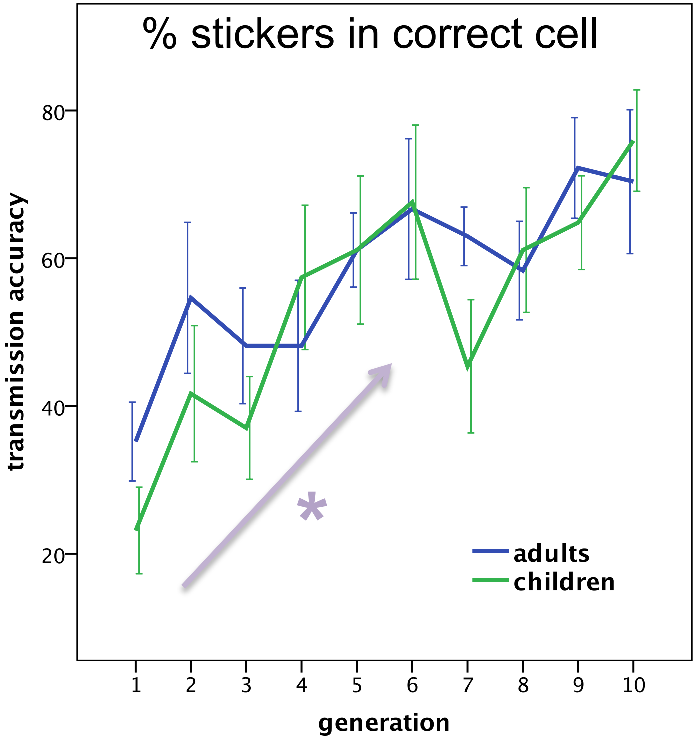
Primary 2



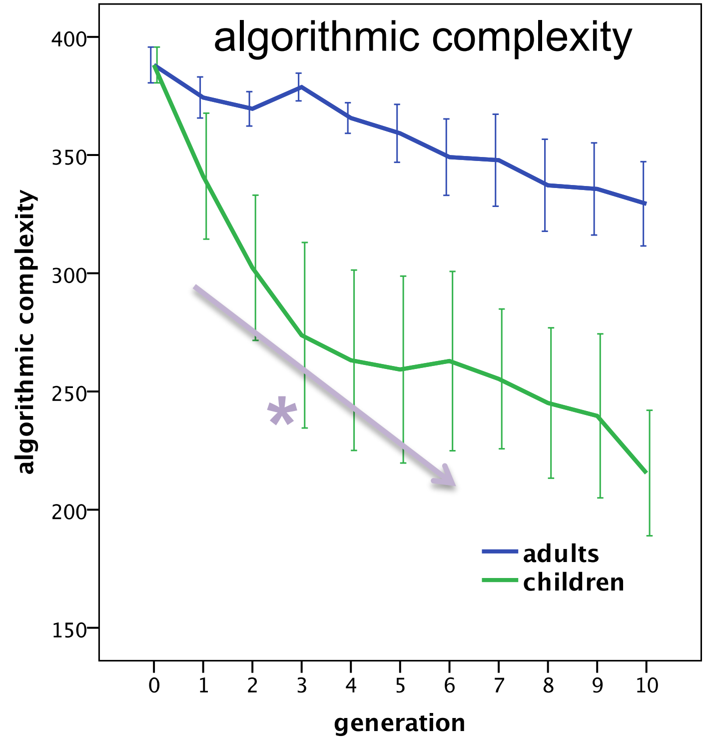
Primary 3

Kempe, Gauvrit & Forsyth (2015)

Transmission Accuracy

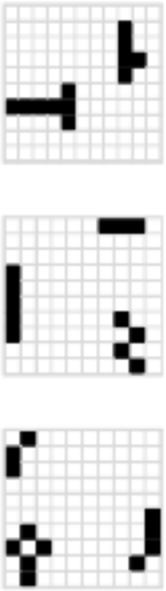


Combinatorial Structure

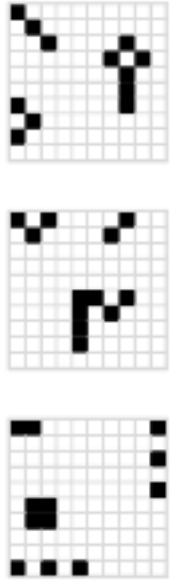


Structure emerges more readily in children.

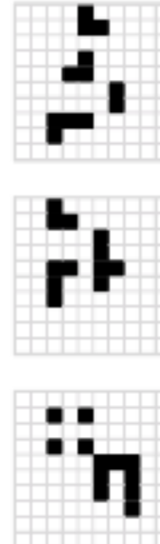
Kempe, Gauvrit & Forsyth (2015)



T-junction
line
zig-zag
cross
???



diagonal
triangle
cross
square
interrupted line
line
???

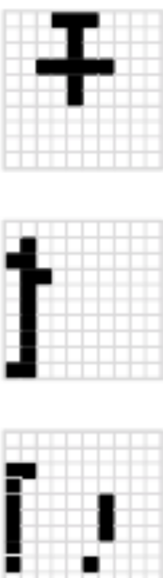


corner
line
T-junction
dog

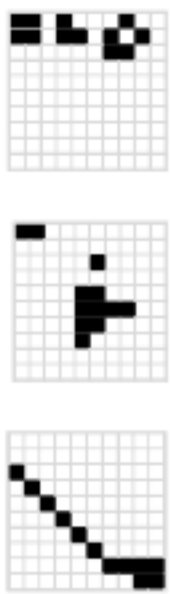
adults

---

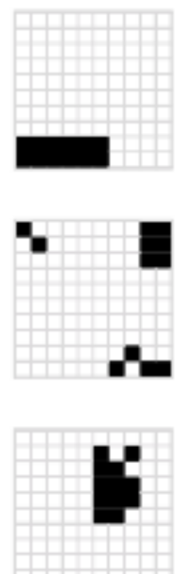
children



cross
line



square
corner
???
blob
diagonal



rectangle
diagonal
blob
???

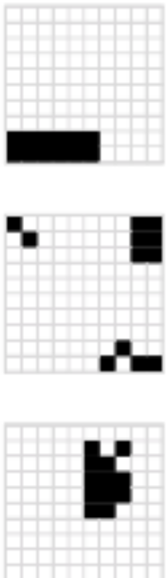
fewer patterns?
simpler patterns?
different patterns?

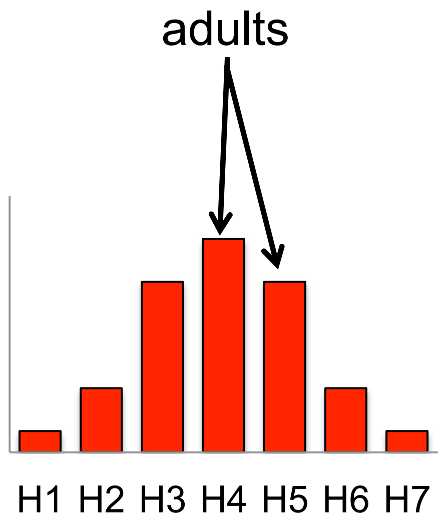
Final patterns (generation 10)

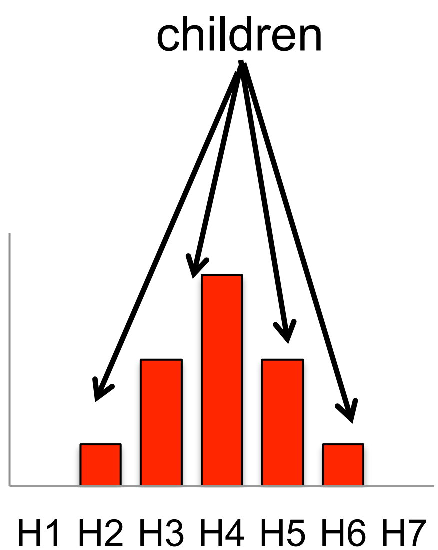
adults	**children**
T-junction	line
line	cross
zig-zag	square
cross	corner
triangle	diagonal lines
square	rectangles
dotted line	blobs
corner	???
dog	???
???	
???	

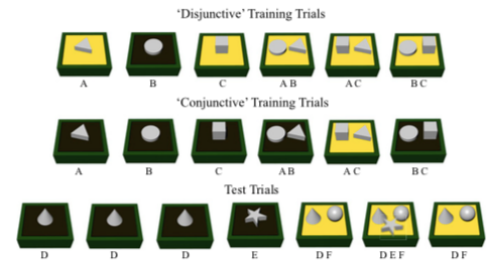
fewer patterns ✔
simpler patterns ✔
different patterns?

A Preliminary Hypothesis









Perhaps children sample more widely, but from a smaller set of initial hypotheses due to limited knowledge?

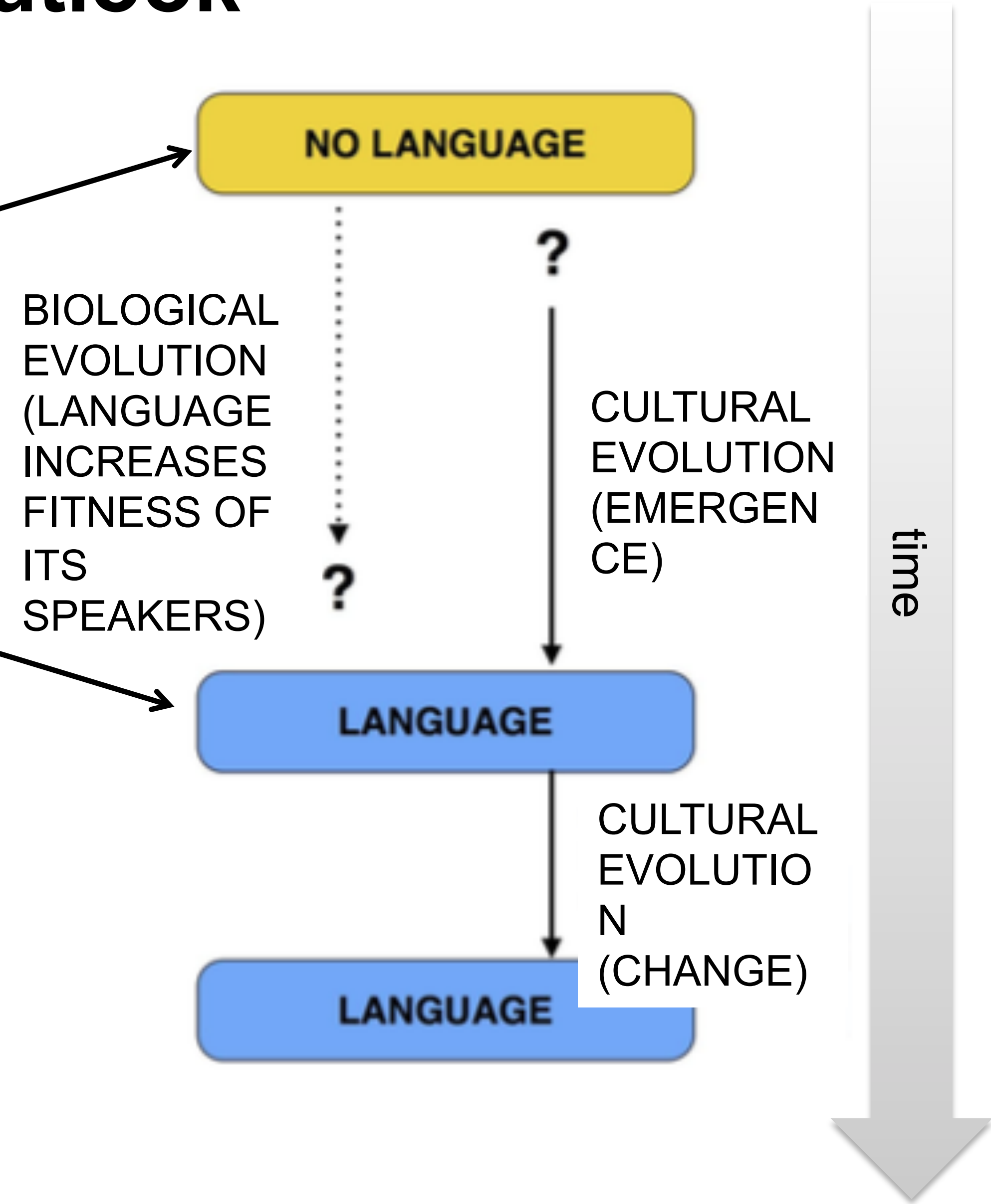
Summary: Children

- The extended childhood in humans provides opportunities for social learning.
- Children’s learning may differ in terms of their initial hypotheses space and/or in terms of how it is affected by limited cognitive capacity.
- Children may have a different hypothesis space.
- Children may differ in how broadly they sample from their hypothesis space.
- Preliminary findings from iterated reproduction suggest that children have fewer/simpler hypotheses.
- Hypothesis: Children sample more broadly from a smaller set of initial hypotheses.

Outlook

How did we get
from here

to here?



slides at: https://language.abertay.ac.uk/SSoL2018/

References

Bogin, B., & Smith, B. H. (1996). Evolution of the human life cycle*. American Journal of Human Biology:* The Official Journal of the Human Biology Association, 8(6), 703-716.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*(5), 489-509.Bartlett (1932)

Cornish, H., Smith, K., & Kirby, S. (2013). Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. In Proceedings of *the 35th Annual Conference of the Cognitive Science Society* (pp. 340-345).

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., ... & Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences, 114*(30), 7892-7899.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science, 31*(3), 441-480.

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition, 136*, 247-254.

Tamariz, M., & Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cognitive Science, 39(1)*, 171-183.

Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology, 60(2)*, 107-126.