Language acquisition and Kolmogorov complexity:

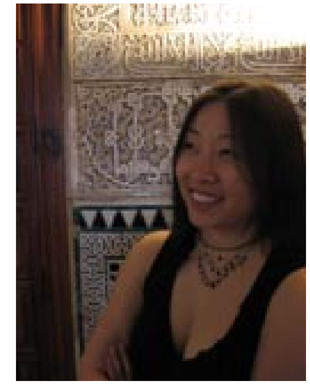
Why is language acquisition possible?

Nick Chater

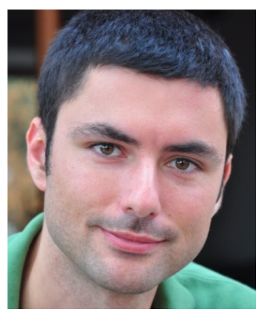
Morten Christiansen

Florencia Reali

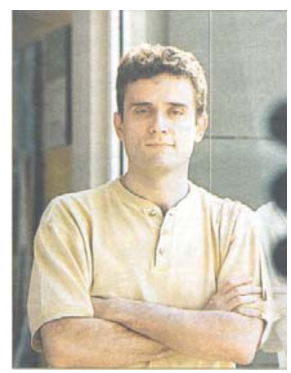
Paul Vitányi



Anne Hsu



Andrea
Baronchelli



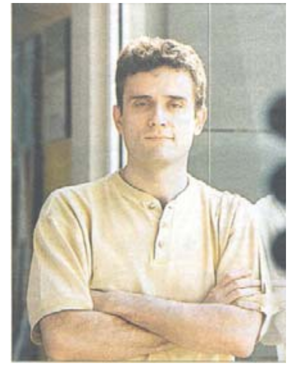
Romualdo
Pastor-Satorras

OVERVIEW

- A brain adapted for language…
- …or language shaped by the brain?
- What can be learned from positive data I
  - asymptotic results
- What can be learned from positive data II:
  - a recipe

Morten Christiansen

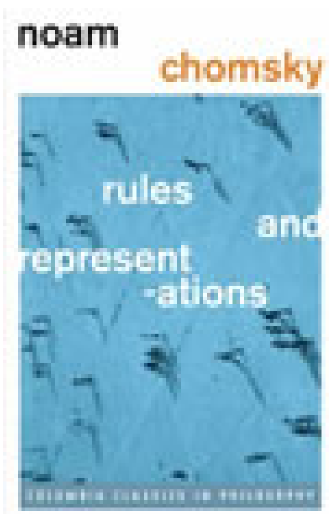
Andrea Baronchelli

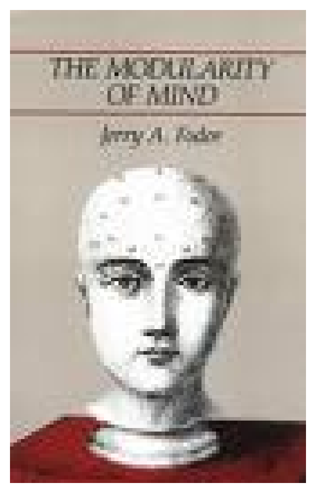


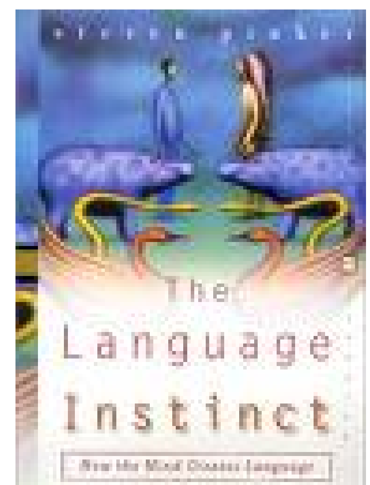
Romualdo Pastor-Satorras

Florencia Reali

1. A brain adapted for language?







COULD BRAINS BE *ADAPTED* FOR LANGUAGE?

- Language seems extremely *complex*
- And to have many highly specific and incredibly subtle properties
- How can children figure it out, while linguists can't?
- That is, how is language acquisition possible?
- Perhaps the triggering of a *genetically coded* language-specific faculty?
  - language instinct
  - language organ
  - language acquisition device
  - language module

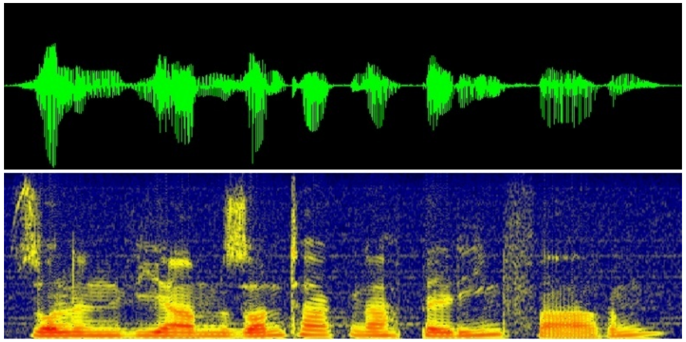
A *LANGUAGE-SPECIFIC* FACULTY IMPLIES THE BRAIN IS *ADAPTED* FOR LANGUAGE, JUST AS FOR VISION

- The visual environment today

- Visual environment of evolutionary adaptation

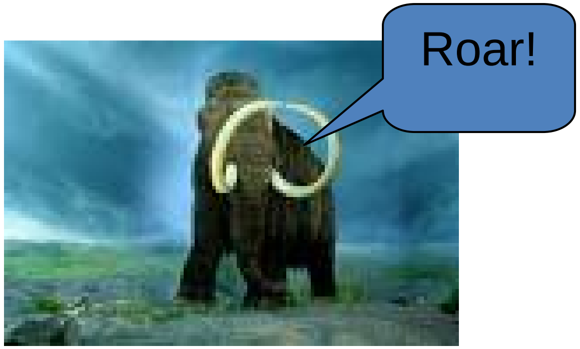
Strangely similar…

- The linguistic environment today



"…the cat sat on the mat…"

- Linguistic environment of evolutionary adaptation



Strangely dissimilar…

BUT PERHAPS LANGUAGE AND THE LANGUAGE FACULTY CO-EVOLVED VIA THE BALDWIN EFFECT (PINKER & BLOOM, 1990)

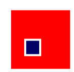


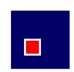
- Driving acquired traits into the genes--
  - It may work for ostrich calluses
  - perhaps it works for language

THE BALDWIN EFFECT: A VERY SIMPLE SIMULATION

- "Language" is a string of features

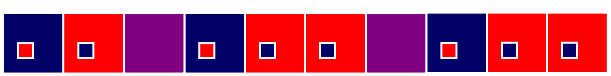
- Genes can express bias or neutrality on each feature:

  "fixed" .95 bias to red: 

  "fixed" .95 bias to blue: 

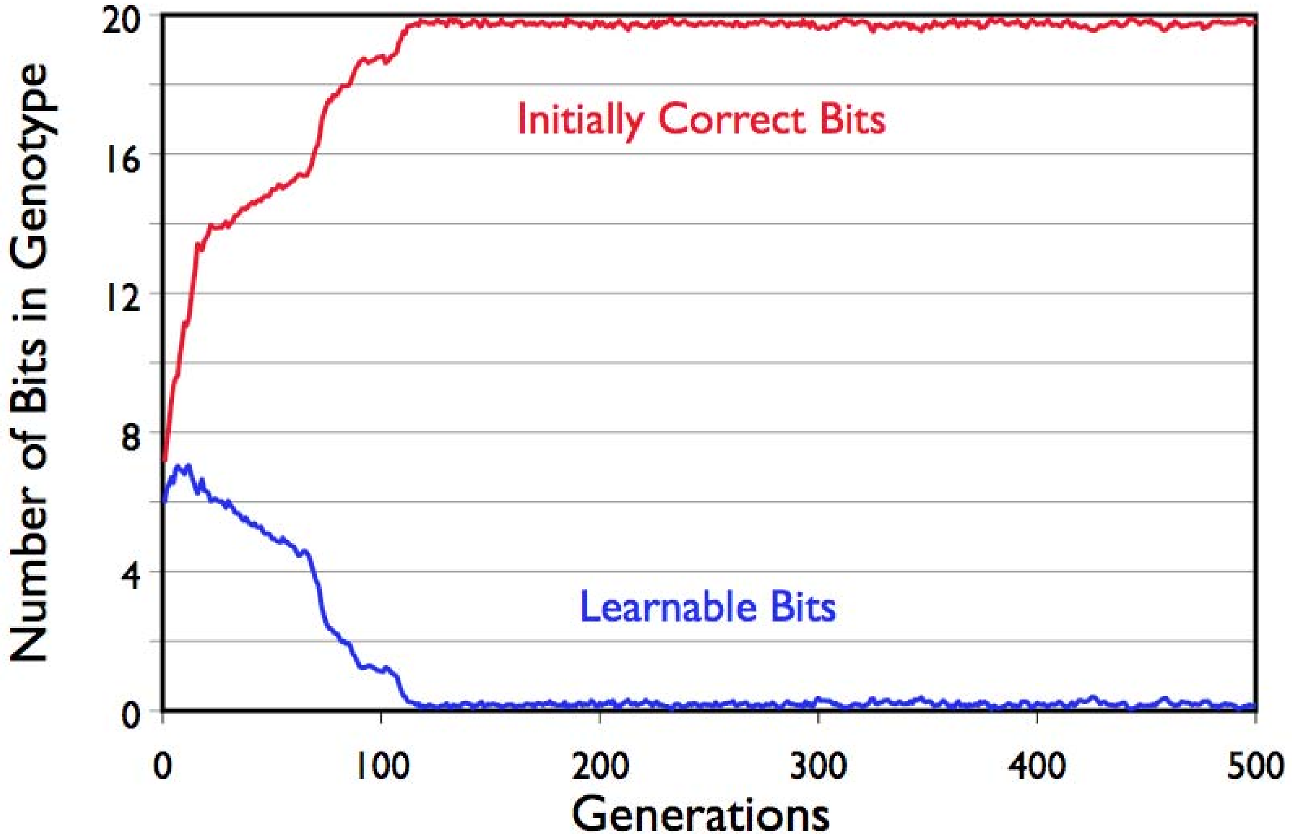
  "learnable": unbiased 

- "Genome":



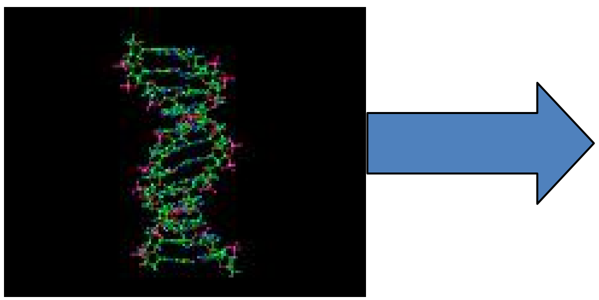
- Trial and error learning
- Only the fastest learners "reproduce"
- And create the next generation by sexual recombination and mutation
- Do the genes begin to adapt to the language???

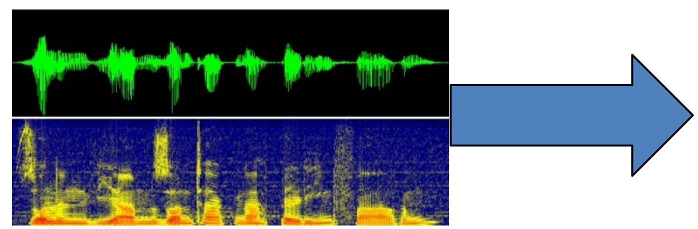
THE BALDWIN EFFECT IN ACTION



BUT CO-EVOLUTION REQUIRES GENETIC ADAPTATION TO A VARYING LANGUAGE

- Can language change *lead* language genes?

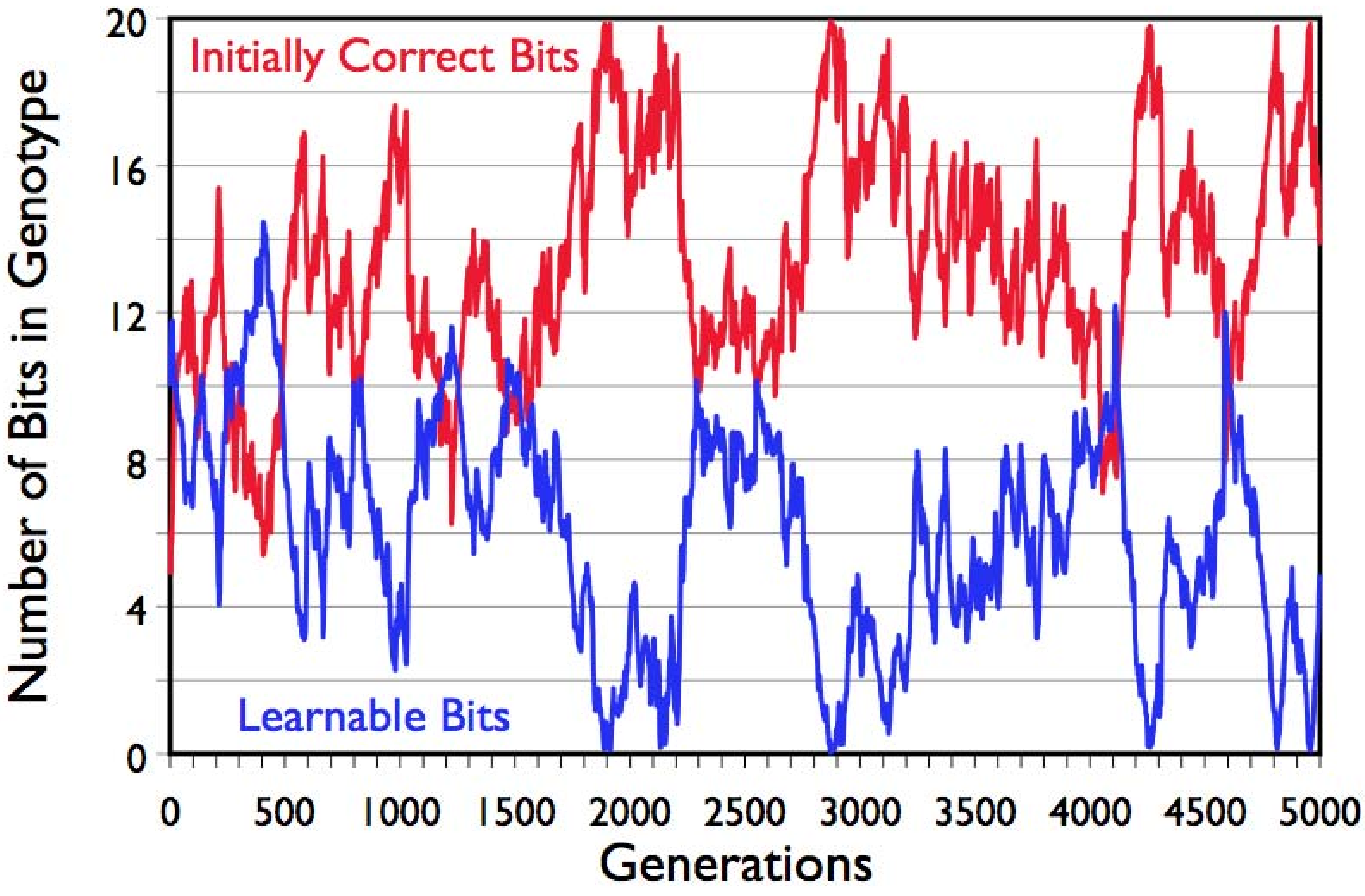




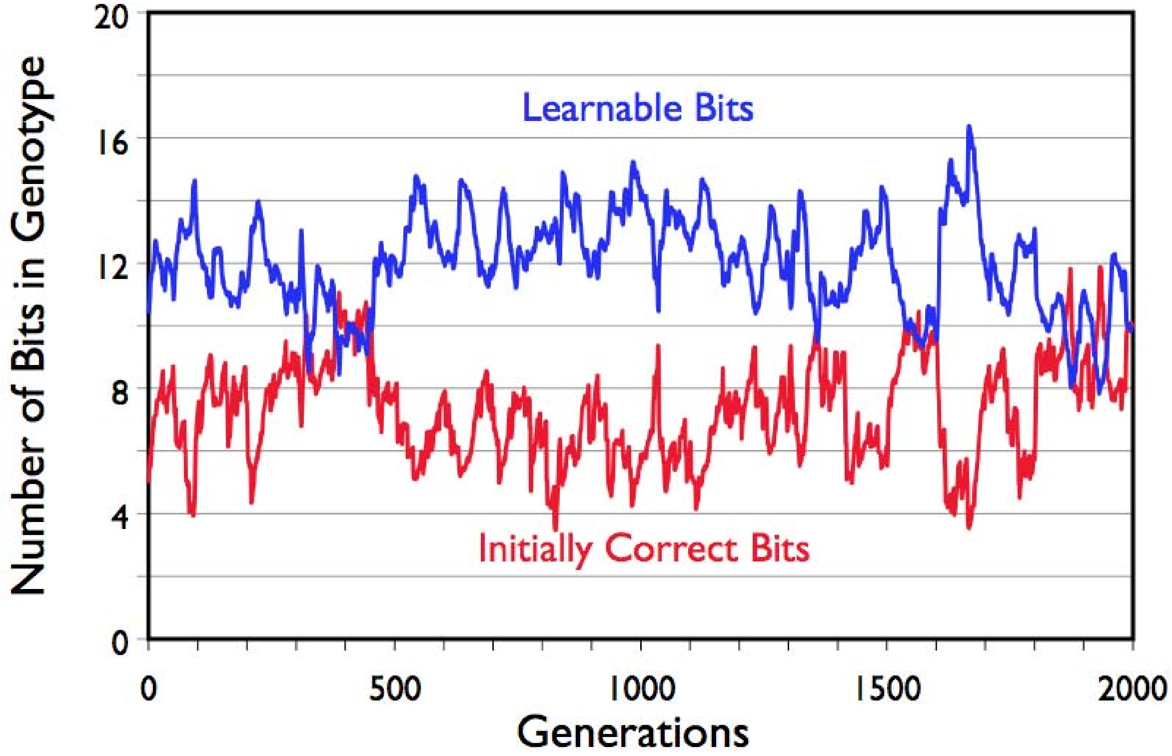
- Potential problem:
  - Language changes very fast, in relation to genetic change

- So what happens when language and genes can *both* change?

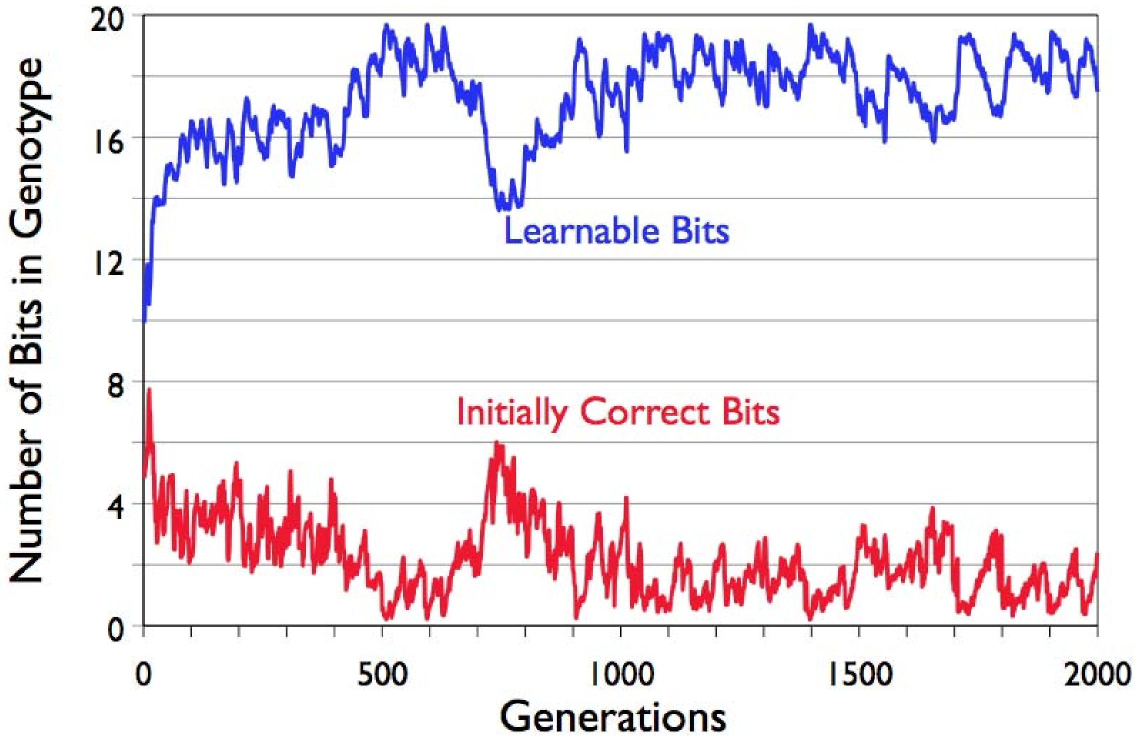
SAME SPEED FOR LANGUAGE AND GENETIC MUTATION RATE



LANGUAGES CHANGES TWICE AS FAST



LANGUAGE CHANGES 10 TIMES AS FAST

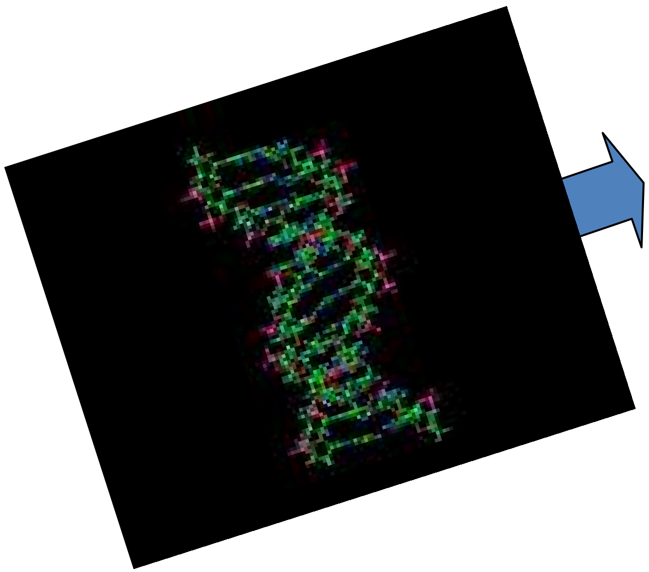


No Baldwin effect

No coevolution

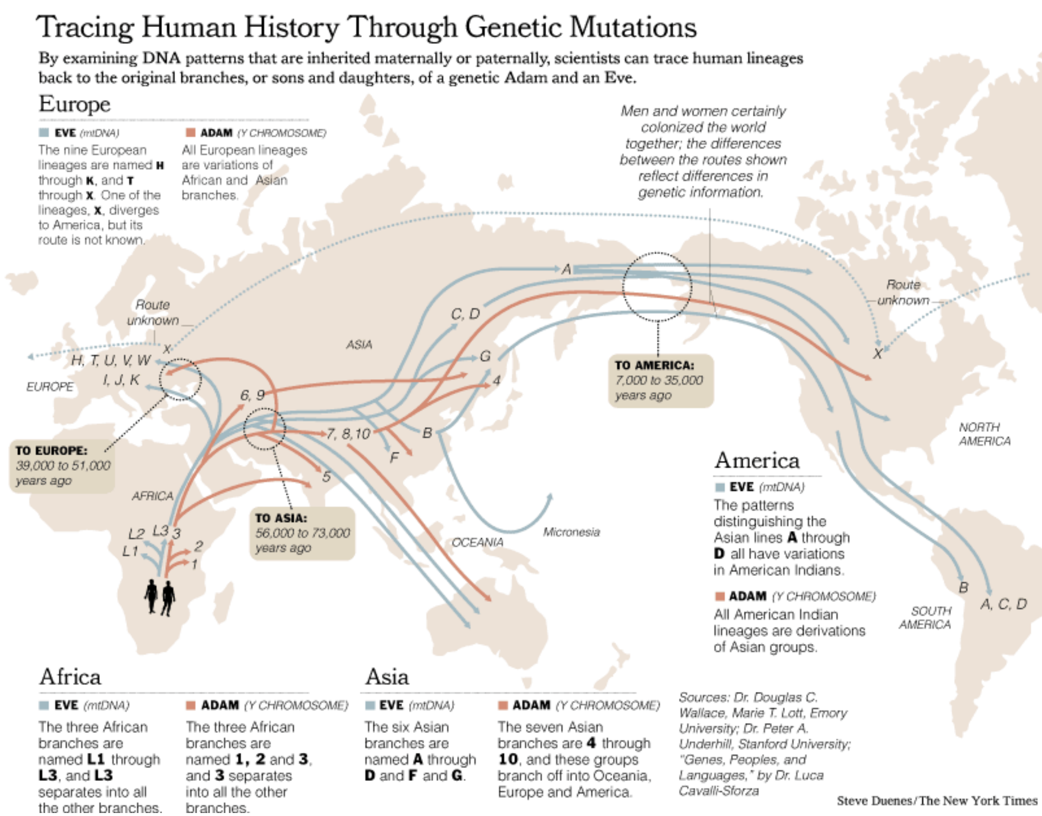
“Learnable” genes win out

GENES CANNOT CATCH A LINGUISTIC "MOVING TARGET"



Chater, N., Reali, F. & Christiansen, M.H. (2009). Restrictions on biological adaptation in language evolution. *PNAS, 106,* 1015-1020.
See also Kirby, Griffiths, Dowman, PNAS.

DIVERGING HUMAN POPULATIONS



Joint work with Andrea Baronchelli, Romualdo Pastor-Satorras, Morten Christiansen, in prepartion

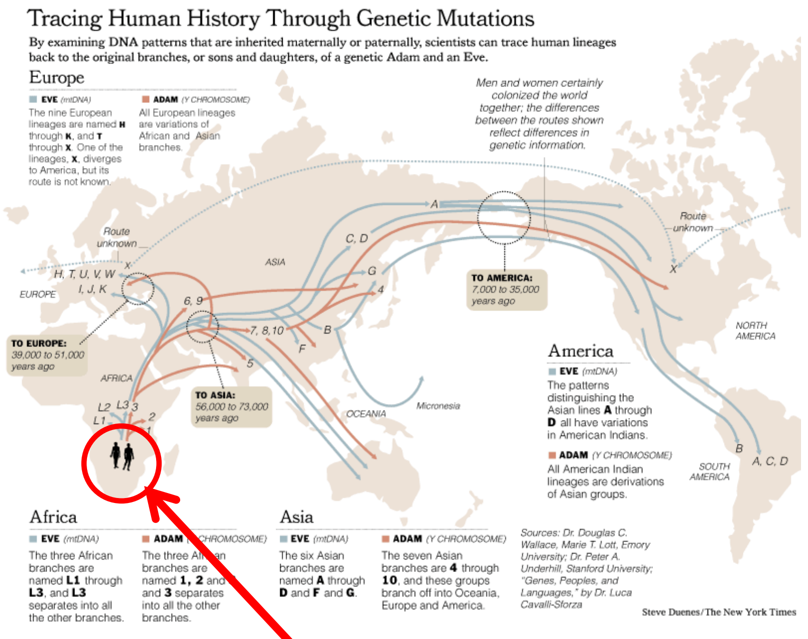
ONCE POPULATIONS ARE SPLIT, CO-EVOLUTION WILL BE SPECIFIC TO THE LOCAL LINGUISTIC ENVIRONMENT

If language-gene coevolution occurred, it had better stop, once populations diverge

But wide geographical separation occurred early, w.r.t., to presumed time-scale for language

(And even geographically nearby groups show very fast linguistic change)

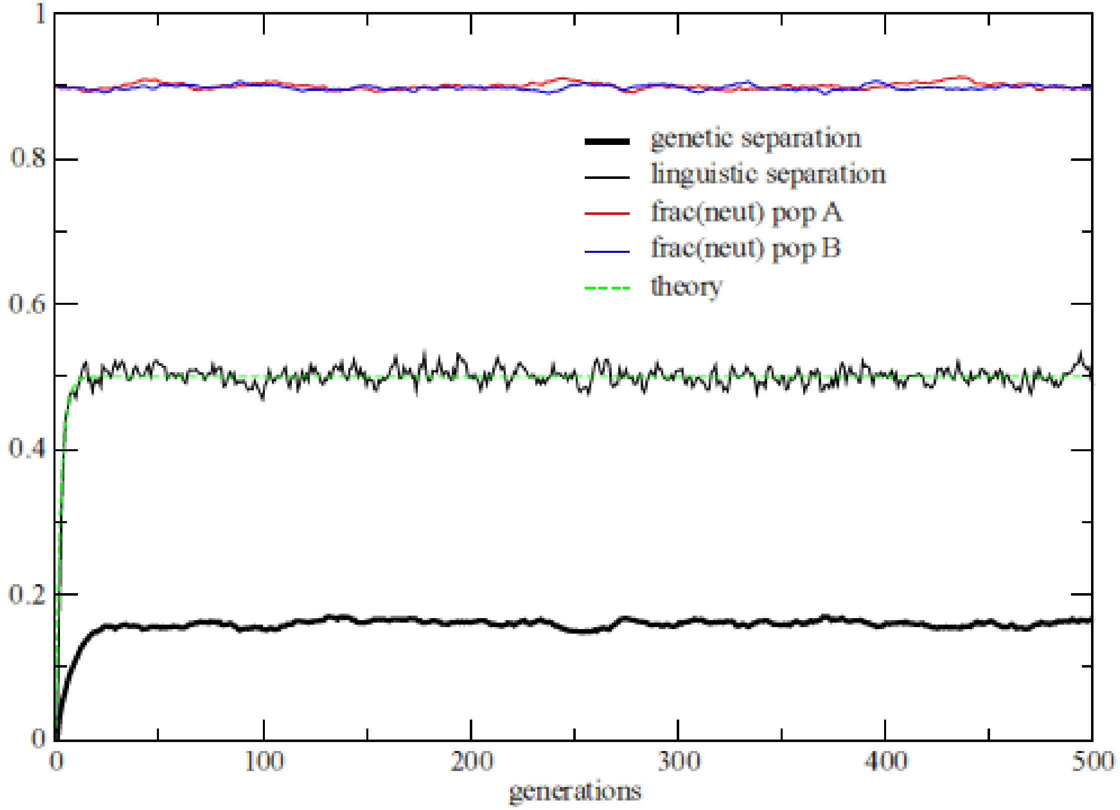
Test with population splitting simulations...



No co-evolution beyond this point!

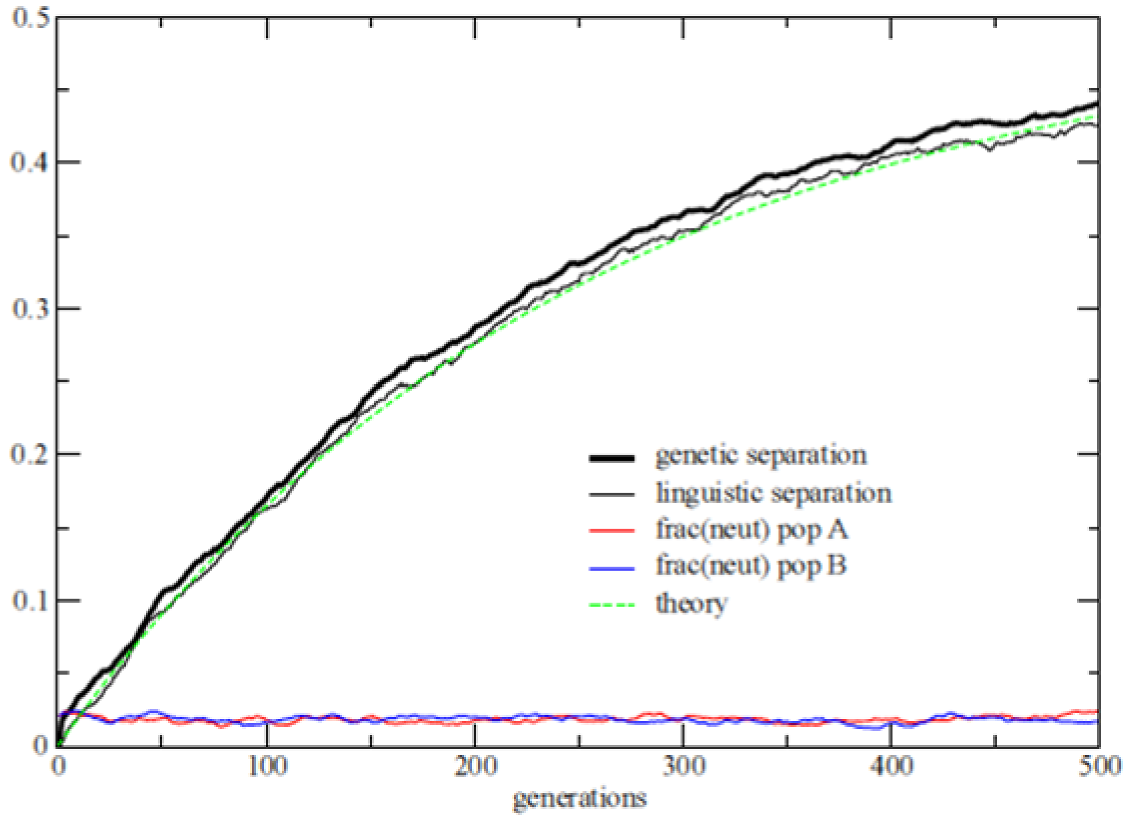
Joint work with Andrea Baronchelli, Romualdo Pastor-Satorras, Morten Christiansen (in prep)

Case 1: Language change is fast



- No coevolution
- Neutral “genes” dominate
- No UG

Case 2: Language change is slow



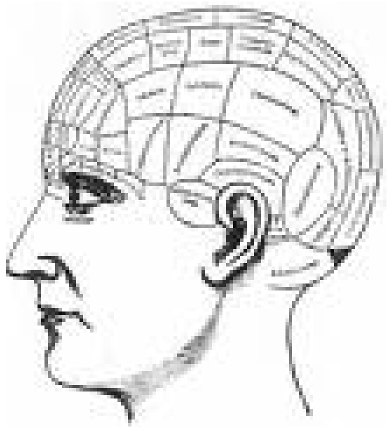
- Lots of local coevolution; few neutral genes
- Genetic divergence precisely *mirrors* linguistic divergence;
- No UG

1. A brain adapted for language?

2. …or is language shaped by the brain?

Morten Christiansen

TWO PROBLEMS OF INDUCTION



N-induction
1, 2, 4…

Hard: We may have
the wrong biases

C-induction
“1, 2, 4…”

Easy: we definitely have
the right biases

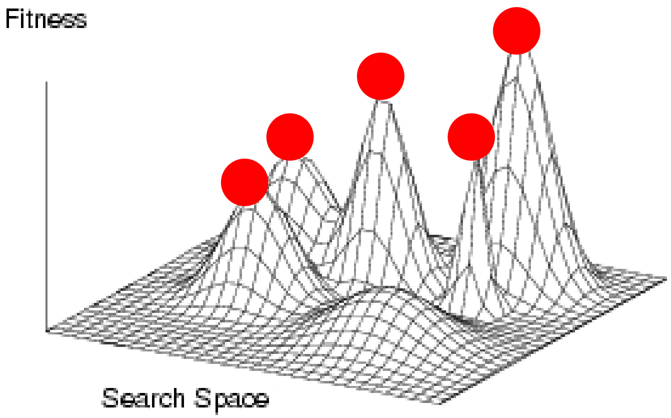
SO LANGUAGE EMERGES FROM INTERACTING CONSTRAINTS...

- Perceptual-motor
  - (speech/auditory apparatus)
- General learning mechanisms
- Cognitive/processing constraints

  (e.g., heuristics and “good enough” parsing, Ferreira)
- Semantics

  (perhaps including embodiment, e.g., Casasanto & Lozano)
- Pragmatics (e.g., Dowty, 1980; Levinson, 2000; Reinhart, 1983)
  - All of which then can become “fossilized” in the language
  - (cf grammaticalization, Hopper, Bybee, etc.)

LANGUAGE IS OPTIMISED TO THESE FACTORS OVER GENERATIONS



Hence language will end up in local minimum in a "fitness landscape"

LEARNING LANGUAGE IS SEARCHING FOR LOST KEYS IN A VAST CITY, ON A DARK NIGHT…

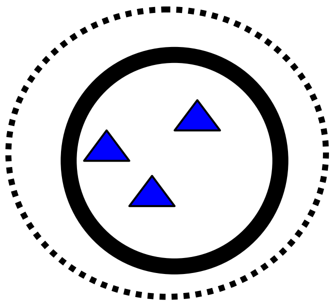
…at least the keys are always right under the lamp-posts

3. What can be learned from positive data: I. Asymptotic results

Paul Vitányi

THE LOGICAL PROBLEM OF LANGUAGE ACQUISITION
(e.g., Hornstein & Lightfoot, 1981; Pinker, 1979)

- Children appear able to learn from *positive* evidence alone
- *Over*general grammars predict that bad sentences are actually ok
- "Mere" non-occurrence of sentences is not enough
- Backed-up by formal results
  - Gold, 1967
  - though Feldman, Horning et al



cf. Alex Clark – learning complex grammars and time-complexity

Perfors, Tenenbaum & Regier on learning that language is based on phrase structure not strings

John Goldsmith, learning phonological structure

SPECIFYING AN "IDEAL" LEARNING SET-UP

- Linguistic environment
- Measures of learning performance
- Learning method

- *Positive evidence only; language generated by a computable-random process, μ, with shortest code length $K(\mu)$*
- *Statistical, not exact (PAC-style)*
- *Simplicity*

PREDICTION BY SIMPLICITY

- Find shortest ‘program/explanation’ for current ‘corpus’
- Predict using that program
    - Strictly, use ‘weighted sum’ of explanations, weighted by brevity

PREDICTION IS POSSIBLE! (SOLOMONOFF, 1978)
SUMMED ERROR HAS FINITE BOUND

$$\sum_{j=1}^{\infty} s_j^2 \leq \frac{\log_e 2}{2} K(\mu)$$

So prediction converges

[faster than 1/*n*log(*n*)], for corpus size *n*

An amazing, and fundamental, result; we assume only computability of the data

Admittedly, the method is uncomputable :--(

LOGICAL PROBLEM OF OVERGENERALIZATION IS SOLVABLE

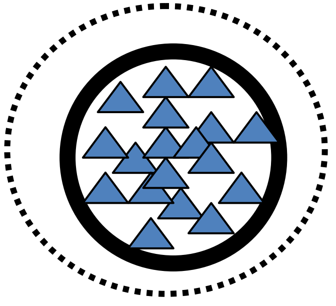
- Suppose learner has probability Δ_j of erroneously guessing an ungrammatical *j*th word

$$\sum_{j=1}^{\infty} \langle \Delta_j \rangle \leq \frac{K(\mu)}{\log_e 2}$$

- Intuitive explanation:
  - overgeneralization underloads probabilities of grammatical sentences;
  - Small probabilities implies longer code lengths

ABSENCE AS IMPLICIT NEGATIVE EVIDENCE

- Overgeneral grammars predict missing sentences
- And their absence is a clue that the grammar is wrong



This overgeneralization theorem makes this intuition rigorous

EXTENSIONS

- Convergence to learning to *generate* language
- Convergence to relate form and meaning
  - both to a high level of accuracy

Chater,N., Vitányi,P. (2007). Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology* 51, 135-163

- If we assume i.i.d. (or stationary distribution?)
- Can we *identify in the limit* the true generative model, with overwhelming probability
- i.e., given a stream of sentences,
- Learner generates successive guesses, until, ultimately
  - The learner never changes its mind
  - And it has precisely identified the generative model (extensionally)
- The learner's strategy is computable

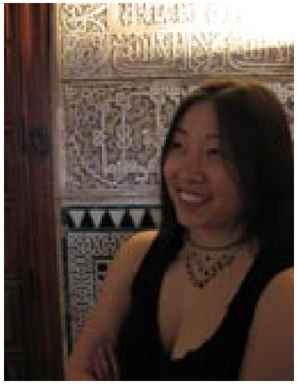
Hsu, Chater & Vitanyi, under review

BUT ABSTRACT RESULTS ARE ONLY A START…

- So far, our results show what an ‘ideal’ learner could do;
  - Uncomputable or intractable
  - Asymptotic
  - Not related to specific cognitive/linguistic phenomena

- How far can they be ‘scaled-down’ to deal with real language acquisition phenomena…?

4. What can be learned from positive data: II. A recipe

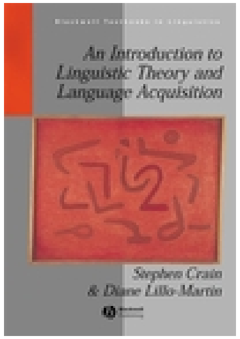


Anne Hsu

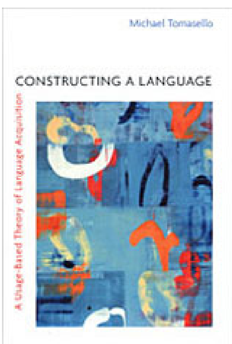
POVERTY OF THE STIMULUS: THE RECIPE

- Goal of the learner: find the shortest code for the linguistic data:
  - A “dual” of Bayesian inference
- Structure needs to pay its way
- Constraint C is learnable if code which
  - 1. “invests” I_c bits to encode C can…
  - 2. …recoup its investment by saving more than I_c bits in encoding the linguistic data

NATIVISM VS. EMPIRICISM, THE DEBATE



- Nativism
  - *c* is acquired
  - But key source of data X is not sufficent to recoup the cost of coding *c*
    - » Empiricist: But there may be hidden sources of data



- Empiricism?
  - Ample supply of data to recoup the cost
    - » Nativist: But the data may not be required or even registered

The recipe can be used for or against either side

EASY EXAMPLE: LEARNING SINGULAR-PLURAL

John loves tennis	x bits
They love_ tennis	y bits
John loves tennis	
*John love_ tennis	x+1 bits
They love_ tennis	
*They loves tennis	y+1 bits

If constraint applies to proportion p of n sentences, constraint saves $n\log_2 1/p$ bits

LEXICAL ALTERNATIONS

I snapped the pencil
I made the pencil snap
I made the pencil disappear
*I disappeared the pencil

I gave a book to the library
I gave the library a book
I donated a book to the library
*I donated the library a book

These *must* be learned, on any theory, of course…

CONTRACTIONS

- Gonna contraction

  I’m going to help her

  I’m gonna help her

  I’m going to the store

  *I’m gonna the store

- Wanna contraction

  Which team do you want to beat?

  Which team do you wanna beat?

  Which team do you want to win?

  *Which team do you wanna win?

These might arise from ‘deep’ syntactic principles, which might, or might not, be learned…

PRELIMINARY RESULTS

	Investment (bits)	Encoding savings per occurrence (bits)	Occurrences in 1 year of child experience	Years to learn
wanna	158.2	0.4	1.2	330
gonna	112.3	1.0	108.6	1
donate (the library)	44.9	4.9	7.2	1.4
disappear (a rabbit)	44.9	.2	61.8	3

∴ wanna contraction is not learnable in isolation, from distribution evidence alone

Hsu, A. & Chater, N. (in press, Cognitive Science). The logical problem of language acquisition goes probabilistic.

SUMMARY

- Why are learners and language aligned?
  - 1. Innate UG?; but UG is evolutionarily implausible?
  - 2. Language is shaped by the brain
    - And culture more broadly, including categories, rituals, social conventions, skills
    - Crucial to understand the inductive biases that have created these structures, in order to learn them (cf Tom Griffiths, on uncovering inductive biases experimentally)

- Is learning possible from positive evidence?
  - 3. Hopeful, but idealized, learnability results
  - 4. And a recipe for studying learnability of specific constraints