Formal Models of the Evolution of Language

Jelle Zuidema

Institute for Logic, Language & Computation

Agenda: To talk about...

- how human language compares to animal communication;
- what the main evolutionary questions are;
- how formal models can contribute to answering those questions;

Agenda: To talk about...

- how human language compares to animal communication;
- what the main evolutionary questions are;
- how formal models can contribute to answering those questions;

Hidden Agenda: To advertise...

- the wonderful world outside, full of empirical facts;
- "Evolutionary Game Theory & Language Evolution"

Beware of human uniqueness claims...

- *Man is a rational animal*

- Human language is unique in Nature

Beware of human uniqueness claims...

- *Man is a rational animal who always loses his temper when he is called upon to act in accordance with the dictates of reason." - Oscar Wilde*

- Human language is unique in Nature

Unique design features of human language

- Compositional Semantics
- Phonemic Coding
- Cultural Transmission
- Arbitrariness
- Displacement
- Stimulus freedom
- Discreteness
- Recursion

Bee dance

(von Frisch'65, '74)

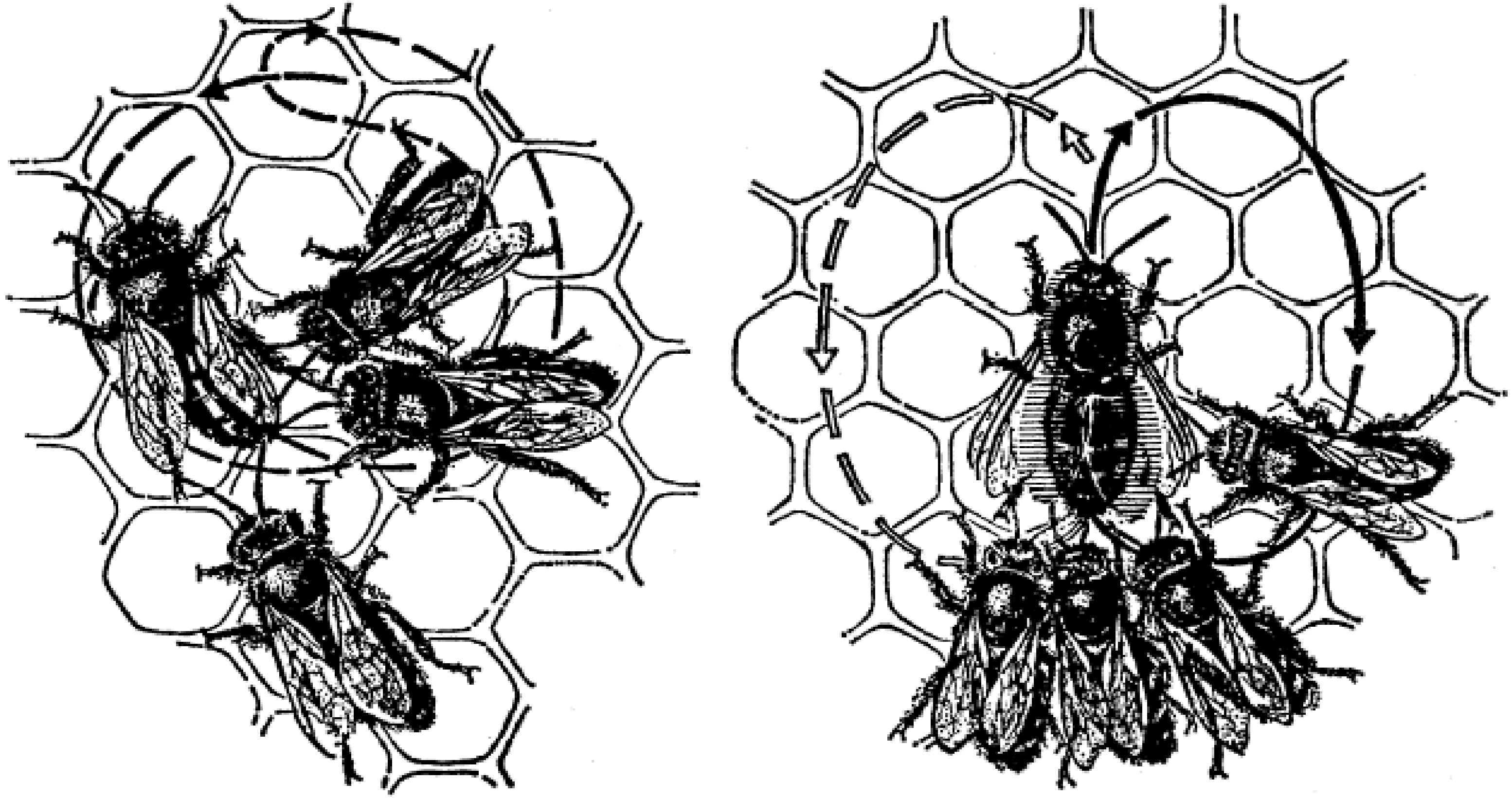


Fig. 1. Running curve of the bee (a) during round dance and (b) during tail-wagging dance. Bees that follow the dancer take in information.

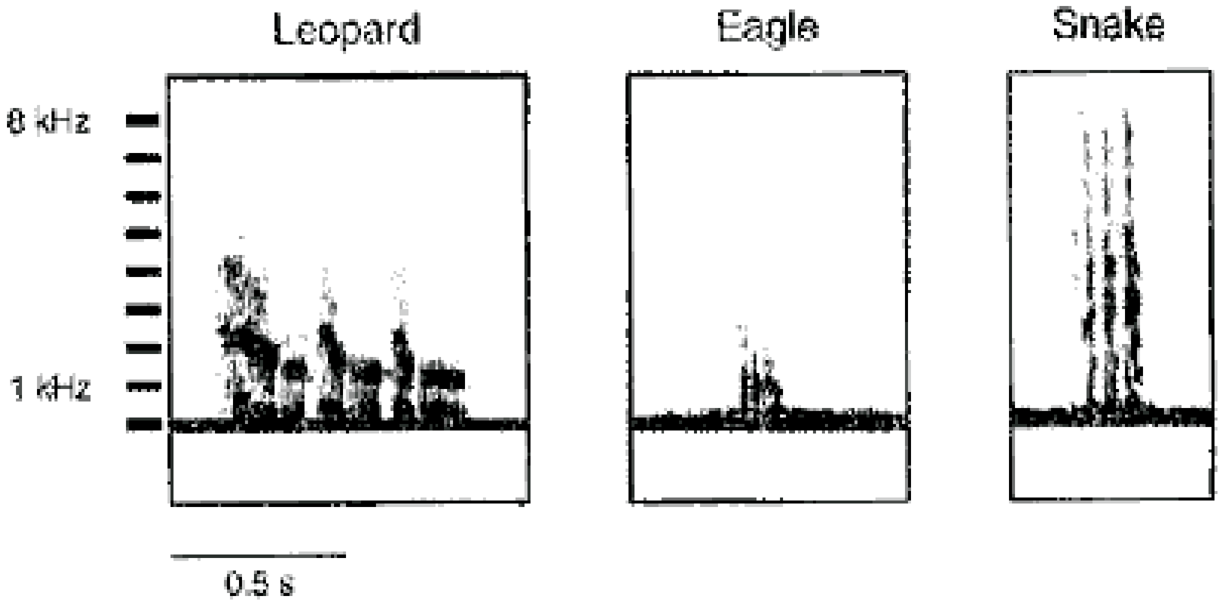
von Frisch'74, *Science* 4152:66

Unique design features of human language

- ~~Compositional Semantics~~
- Phonemic Coding
- Cultural Transmission
- Arbitrariness
- ~~Displacement~~
- Stimulus freedom
- Discreteness
- Recursion

Vervet monkey alarm calls

(Struhsaker'67; Seyfarth, Cheney & Marler'80; squirrels: Sherman'77)



"chirp" "grunt" "chutter"

(Zuberbühler'02 An. Beh. 63)

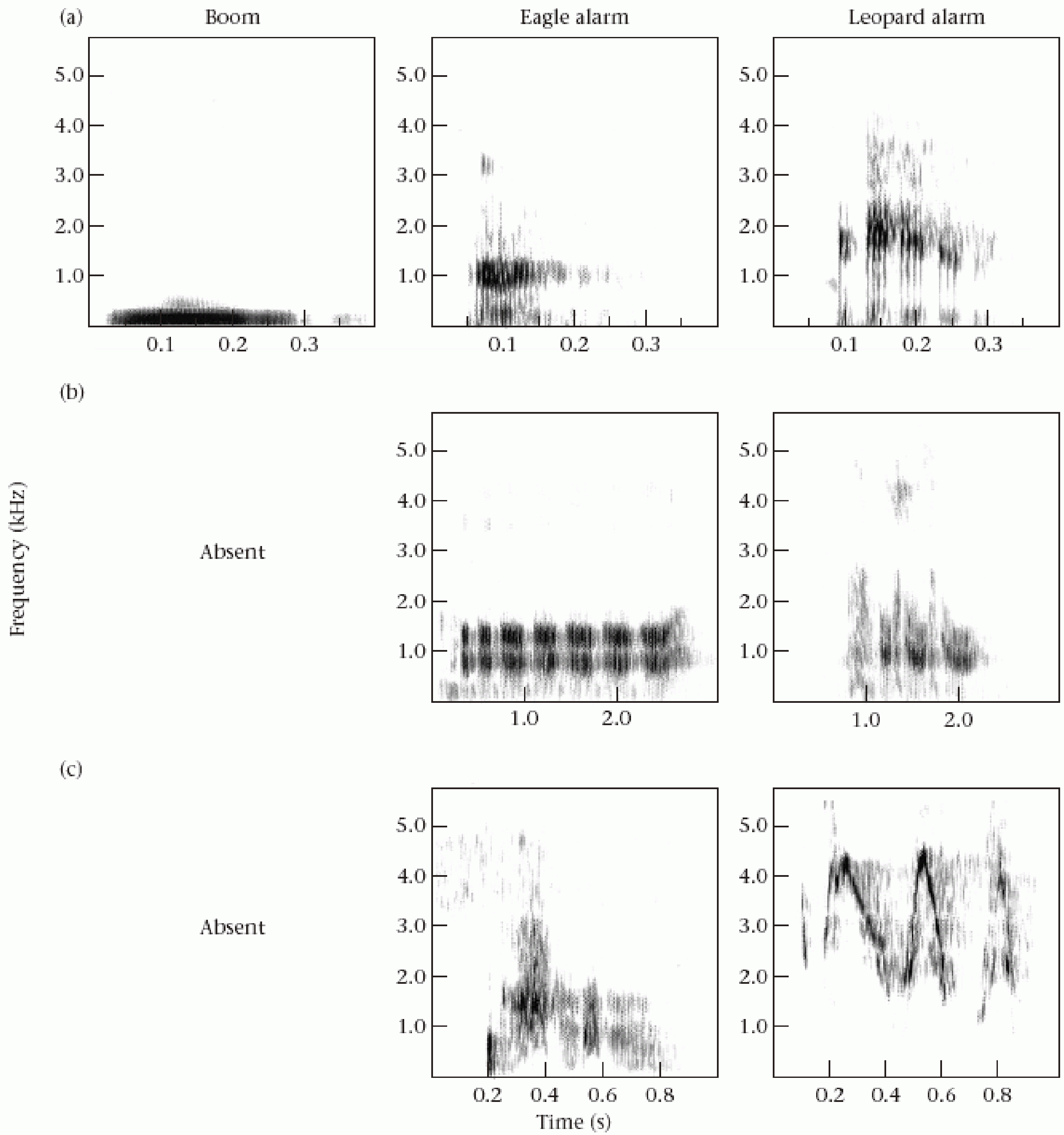
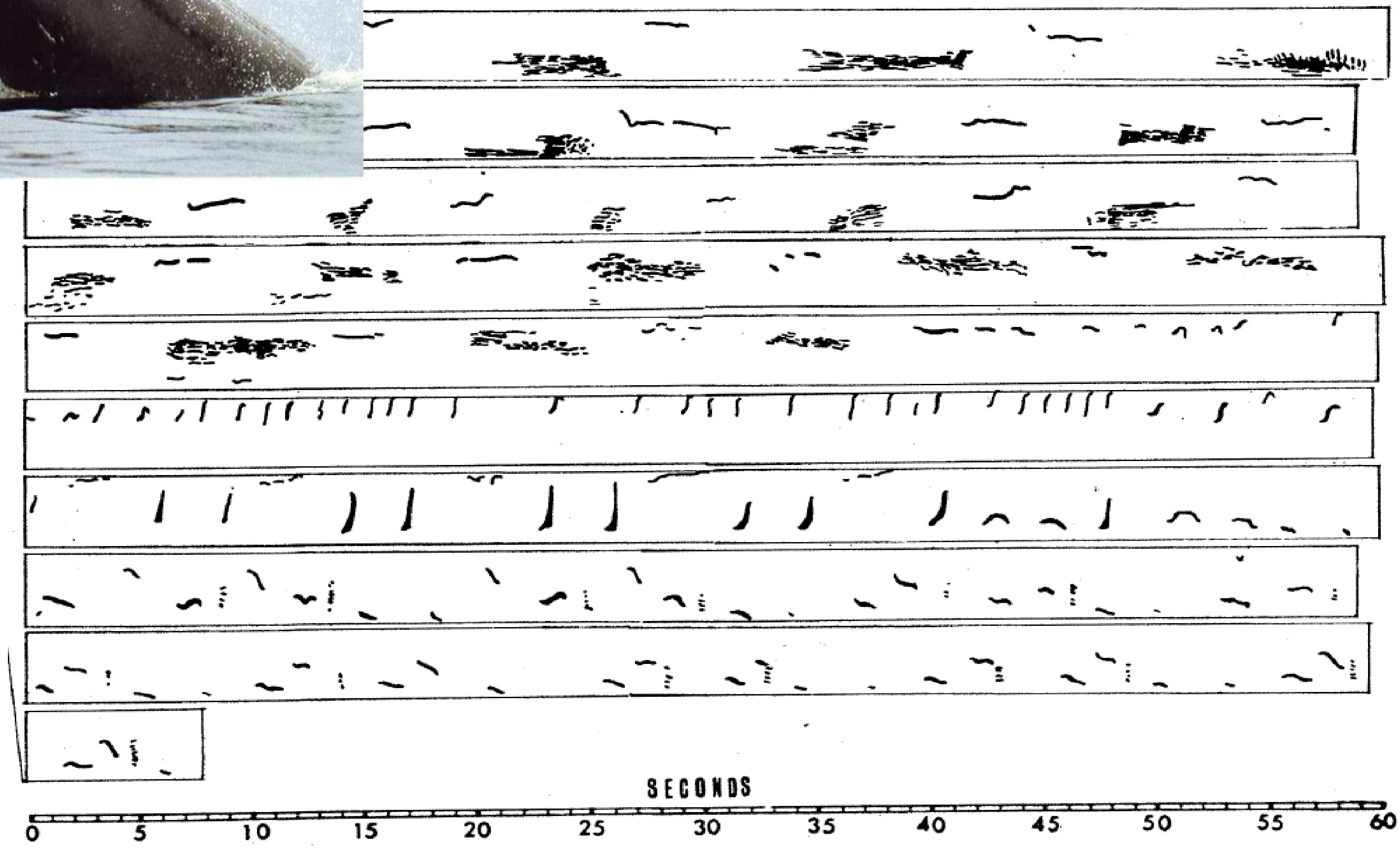


Figure 1. Spectrographic illustrations of the vocalizations used in this study. (a) Male Campbell's monkey, (b) male Diana monkey, (c) female Diana monkey. Male Campbell's and male Diana monkey alarm calls were used as playback stimuli (see text). Recordings were digitized at a sampling rate of 44 kHz (16 bits accuracy) and displayed using a 256-point Fourier transformation (Hamming window function) that resulted in wide-band spectrograms (analysis resolution: 700 Hz with 21.5 Hz/0.72 ms accuracy).

Unique design features of human language

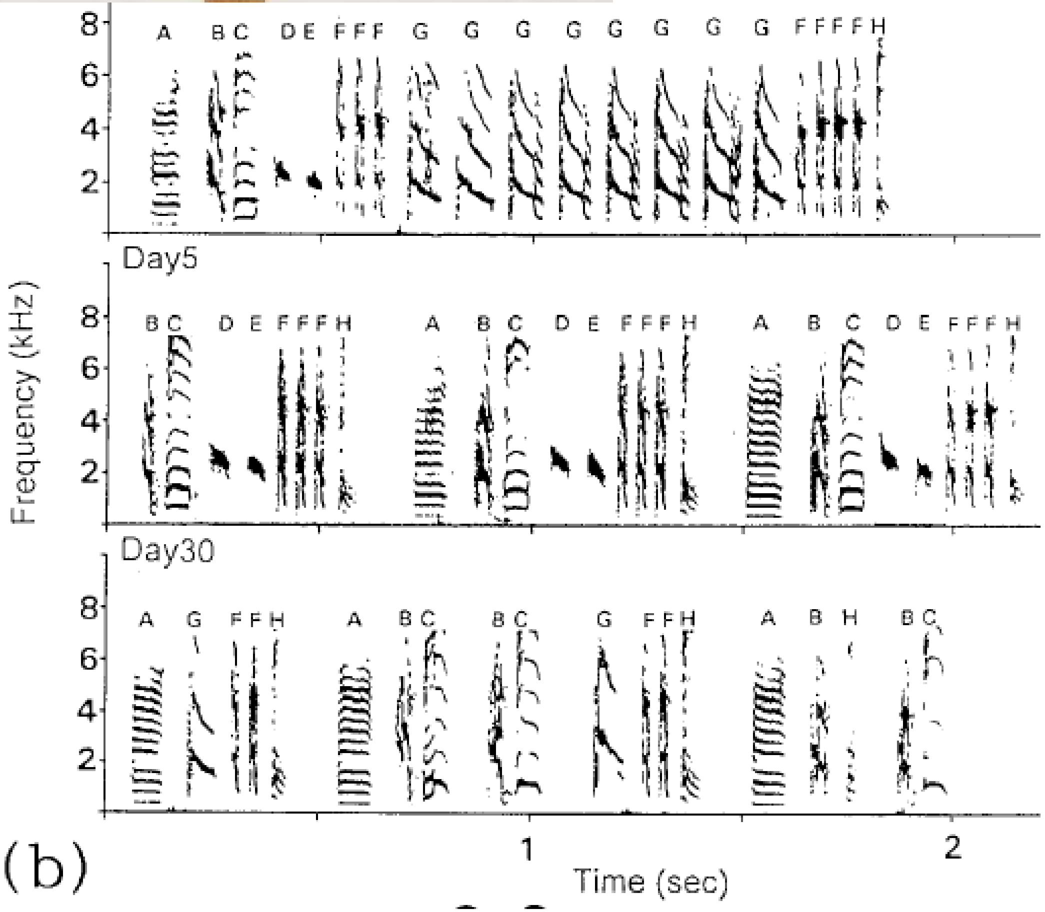
- ~~Compositional Semantics~~
- Phonemic Coding
- Cultural Transmission
- ~~Arbitrariness~~
- ~~Displacement~~
- Stimulus freedom
- ~~Discreteness~~
- Recursion

Humpback whale song

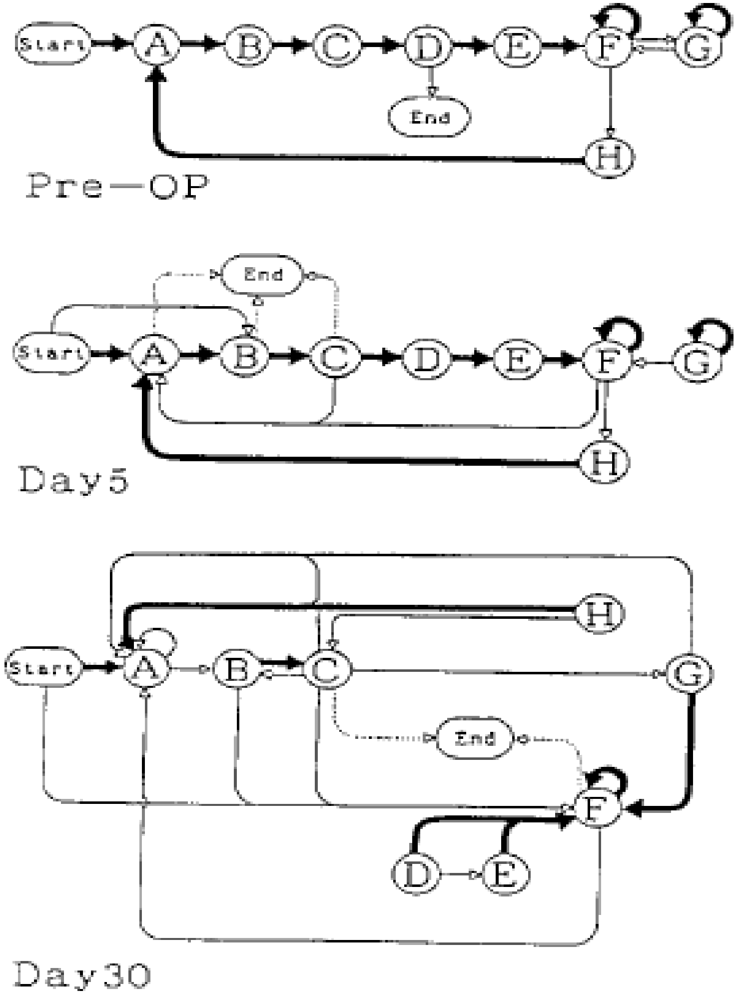


(Watlington'63; Payne & McVay'71)

Bengalese finch song



(b)



(Okanoya & Yamaguchi'9
J. Neurobio. 33,4)

Unique design features of human language

- ~~Compositional Semantics~~
- ~~Phonemic Coding~~
- ~~Cultural Transmission~~
- ~~Arbitrariness~~
- ~~Displacement~~
- ~~Stimulus freedom~~
- ~~Discreteness~~
- Recursion

(Examples from Comrie, 1981)

Isolating languages: No morphological variation for tense, case or plurality; Each word typically consists of a single morpheme. E.g. Vietnamese "*Khi tôi dên nhà ban tôi, chúng tôi*"

Khi	*tôi*	*dên*	*nhà*	*ban*	*tôi,*	*chúng*	*tôi*	*băt*	*dâu*	*làm*	*bài.*
when	I	come	house	friend	I	PLURAL	I	begin		do	lesson

"When I came to my friend's house, we began to do lessons."

Polysynthetic languages: Many lexical morphemes combined in a single word. E.g. Chukchi (Siberia) "*tϵmey$\eta\epsilon$levtϵp$\epsilon\gamma$tϵrkϵn*"

tϵ-	*mey$\eta\epsilon$-*	*levtϵ-*	*p$\epsilon\gamma$t-*	*ϵrkϵn*
great	head	ache	1stSINGULAR	IMPERFECT

"I have a fierce head-ache"

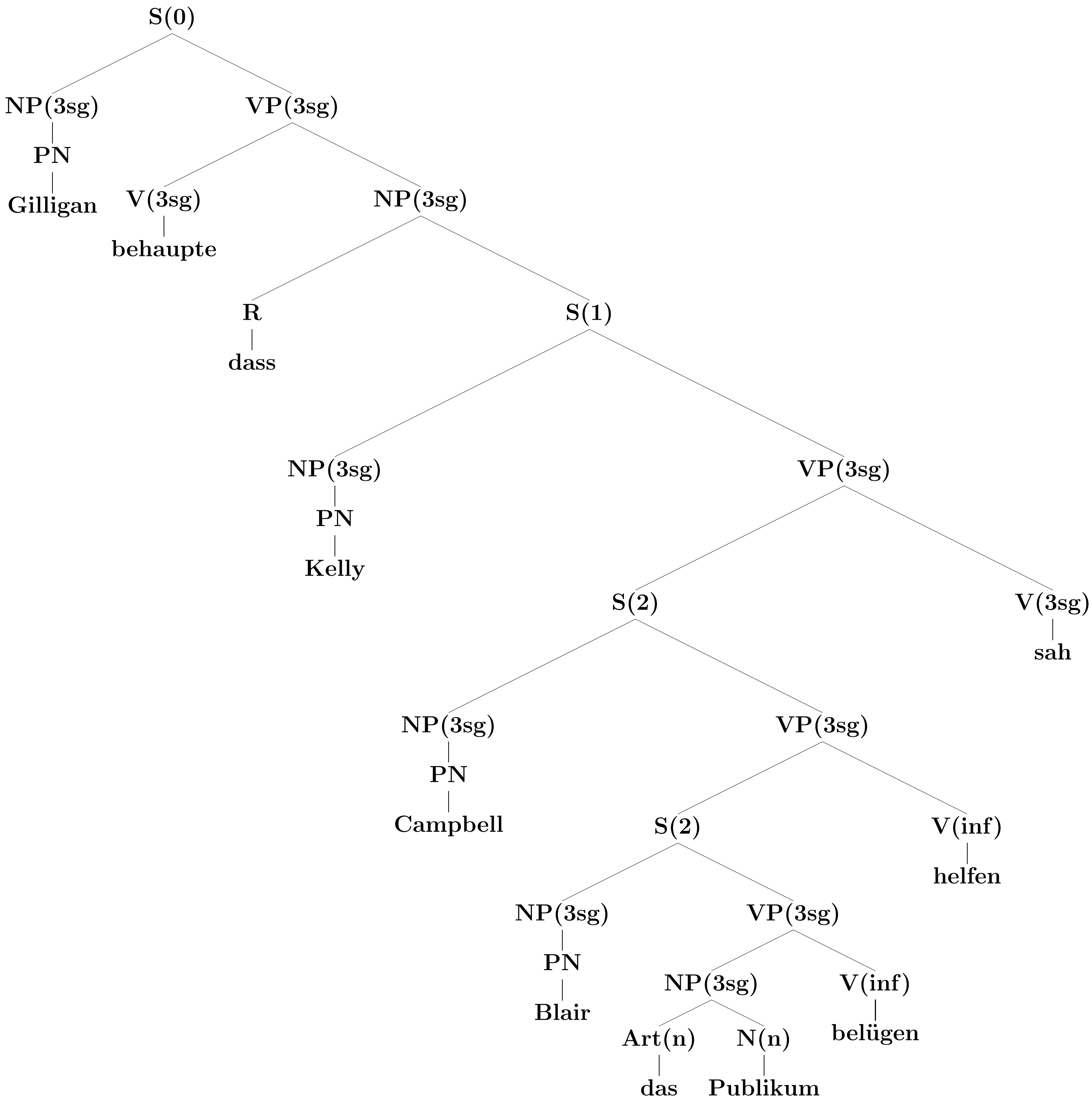
(1) a. Gilligan claims that Blair deceived the public.

b. Gilligan claims that Campbell helped Blair deceive the public.

c. Gilligan claims that Kelly saw Campbell help Blair deceive the public. (tail recursion)

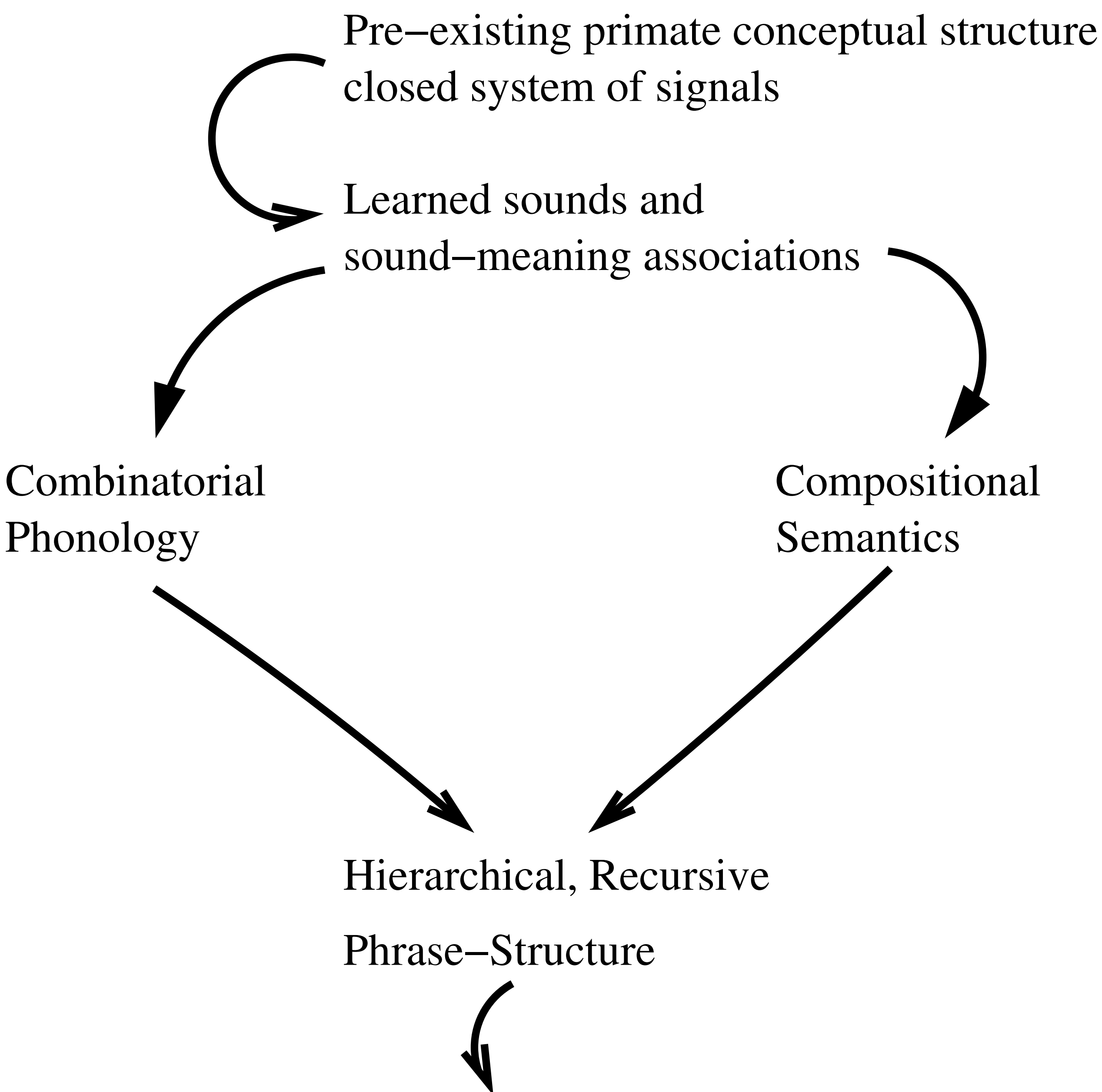
(2) a. Gilligan behaupte dass Kelly Campbell Blair das Publikum belügen helfen sah. (center embedding)

b. Gilligan beweert dat Kelly Campbell Blair het publiek zag helpen bedriegen. (crossing dependencies)


S(0)
NP(3sg)
VP(3sg)
PN
Gilligan
V(3sg)
behaupte
NP(3sg)
R
dass
S(1)
NP(3sg)
PN
Kelly
VP(3sg)
S(2)
V(3sg)
sah
NP(3sg)
PN
Campbell
VP(3sg)
S(2)
V(inf)
helfen
NP(3sg)
PN
Blair
VP(3sg)
NP(3sg)
Art(n)
das
N(n)
Publikum
V(inf)
belügen

Major Transitions in the Evolution of Language

(Jackendoff'02)



Evolutionary Questions

- Why and how did each of these transitions occur?

Evolutionary Questions

- Why and how did each of these transitions occur?

  (in humans)

- Why and how did each of these transitions not occur?

  (in non-human primates)

That’s all fascinating...

... but what does it have to do with logic?

A simple mathematical model

(based on Nowak, Komarova and Niyogi, *Science*, 2001)

a finite set $\mathcal{L}$ of possible languages.

$$l_1, l_2, l_3, \ldots, l_N$$

Learning (or Genetic Transmission)

a probabilistic mapping from parent's language to the child's.

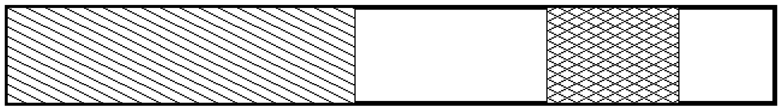
$$Q = \begin{pmatrix} & l_1 & l_2 & l_3 & l_4 \\ \hline l_1 & q & r & r & r \\ l_2 & r & q & r & r \\ l_3 & r & r & q & r \\ l_4 & r & r & r & q \end{pmatrix},$$

where $Q_{ii} = q$ is the learning accuracy, and $r = \frac{1-q}{N-1}$.

Population

a vector of relative frequencies $\vec{x}$. E.g.:

$$\begin{array}{cccccccc} & & & l_1 & l_2 & l_3 & l_4 & \\ \vec{x}(t) & = & (& .45 & .25 & .2 & .1 &) \end{array}$$



Payoff

General

$$F_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$$
$$f_i = \sum_j x_j F_{ij}$$

Simplified

$$F = \begin{array}{c|cccc} & l_1 & l_2 & l_3 & l_4 \\ \hline l_1 & 1 & a & a & a \\ l_2 & a & 1 & a & a \\ l_3 & a & a & 1 & a \\ l_4 & a & a & a & 1 \end{array}$$
$$f_i = x_i + (1 - x_i)a$$

If $a < 1$ every diagonal state is a Nash equilibrium: where no player can improve by unilaterally changing strategy.

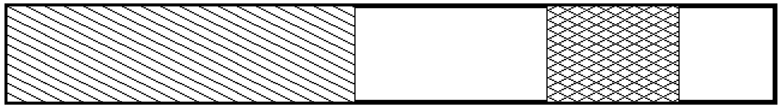
Evolution

change in frequencies as a function of learning and payoff

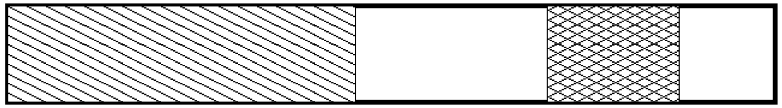
$$\Delta x_i = \sum_{j=0}^{N} x_j Q_{ji} f_j - \phi x_i$$

where ϕ is the average payoff, $\phi = \sum_{i=1}^{N} f_i x_i$

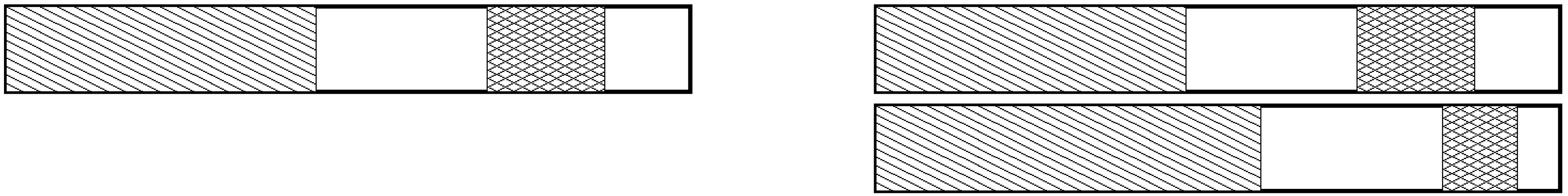
Random guessing ($q = 0$)



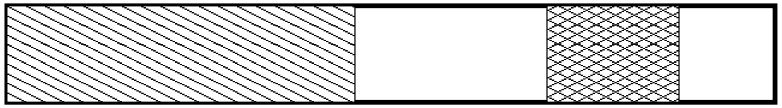
Perfect learning ($q = 1$)



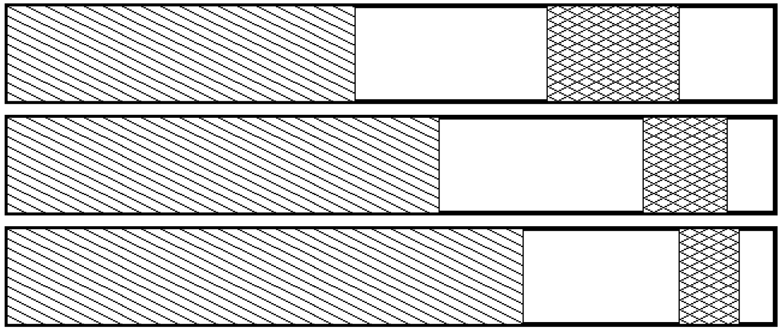
Random guessing ($q = 0$) **Perfect learning ($q = 1$)**



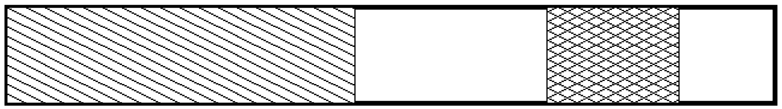
Random guessing ($q = 0$)



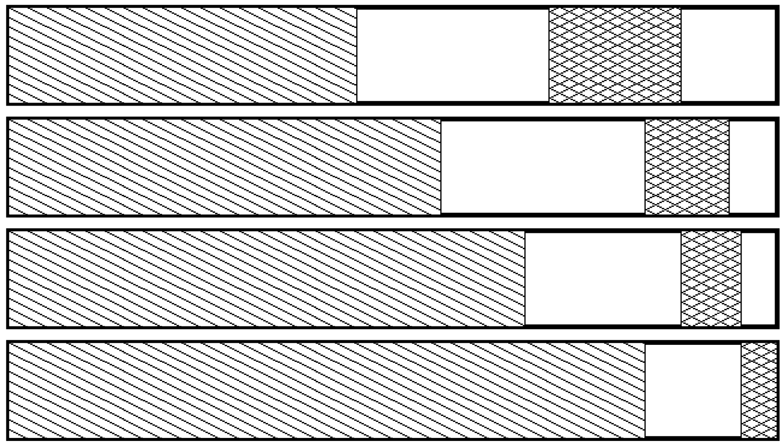
Perfect learning ($q = 1$)



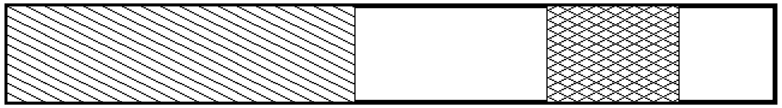
Random guessing ($q = 0$)



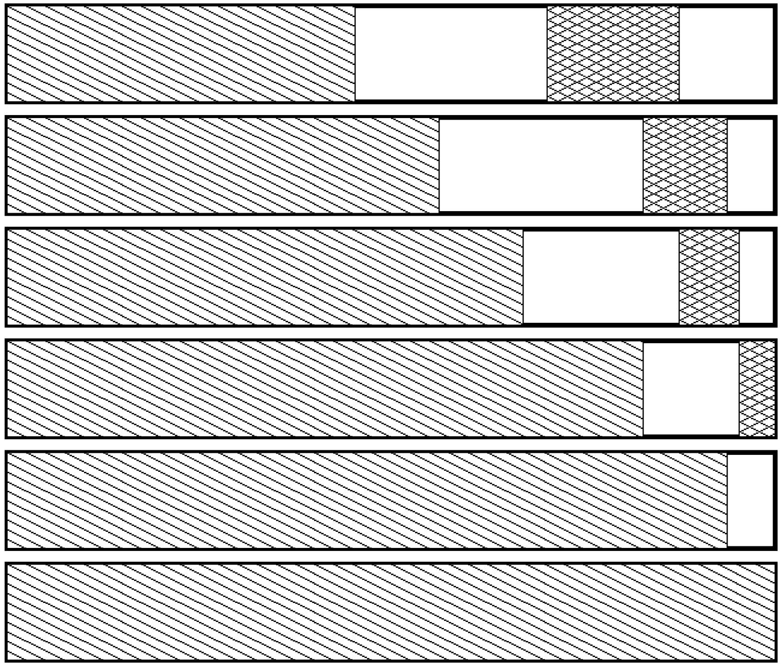
Perfect learning ($q = 1$)



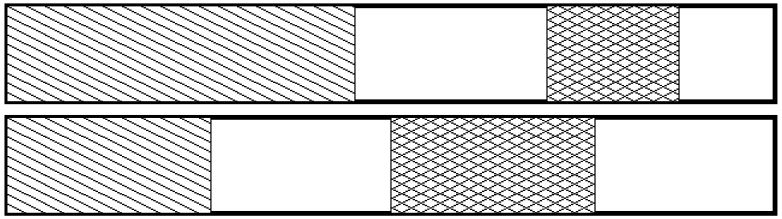
Random guessing ($q = 0$)



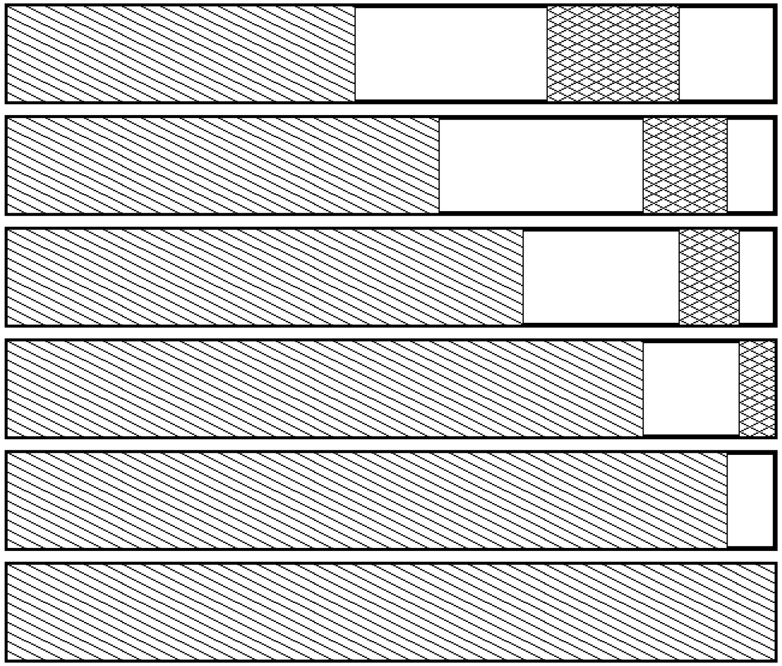
Perfect learning ($q = 1$)



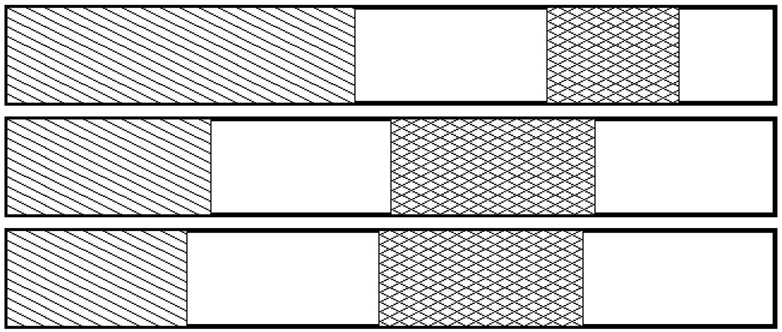
Random guessing ($q = 0$)



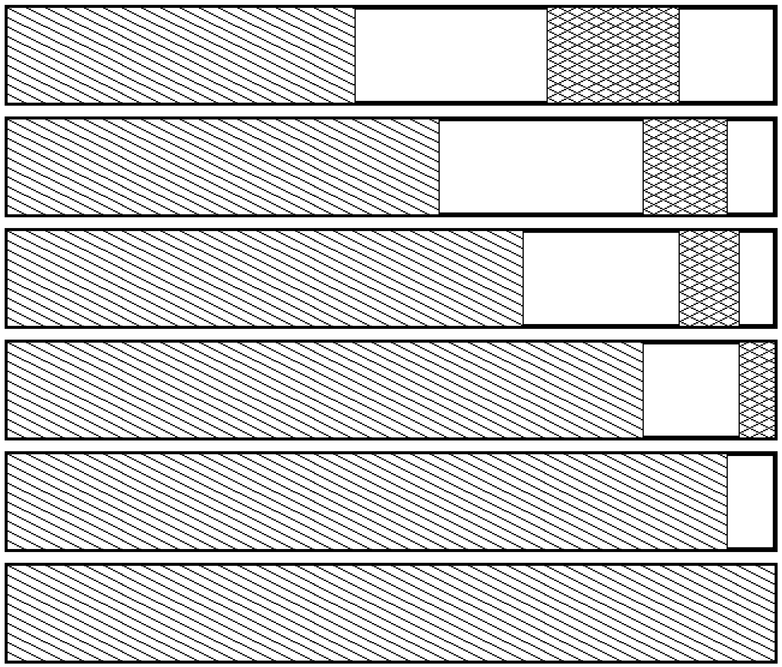
Perfect learning ($q = 1$)



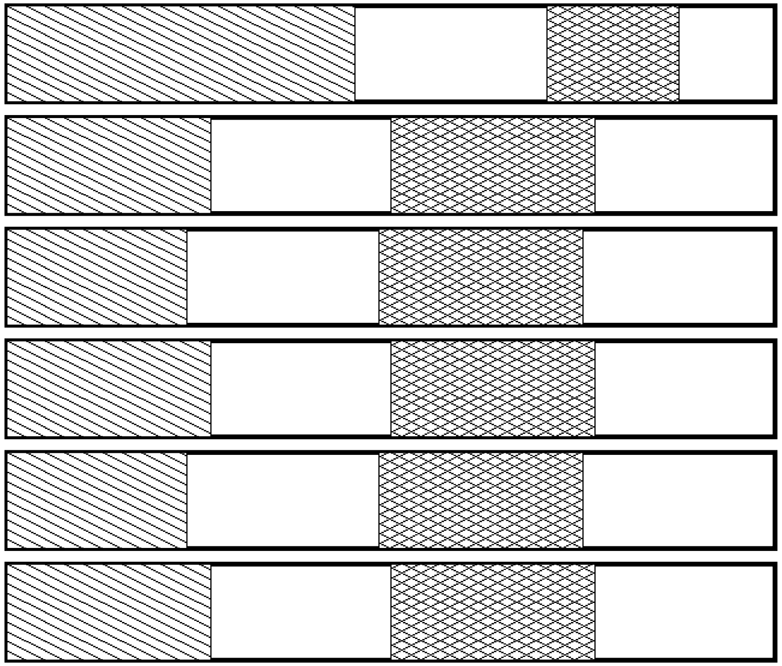
Random guessing ($q = 0$)



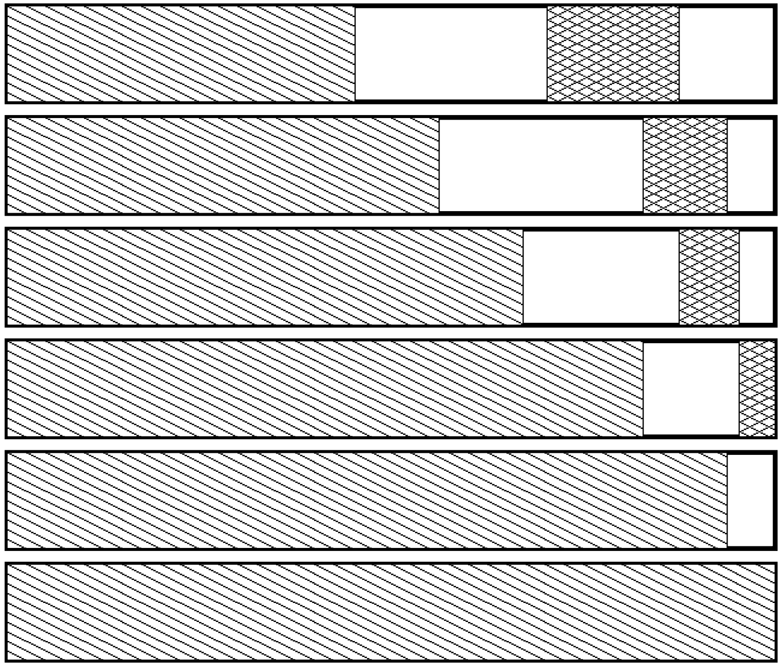
Perfect learning ($q = 1$)



Random guessing ($q = 0$)



Perfect learning ($q = 1$)



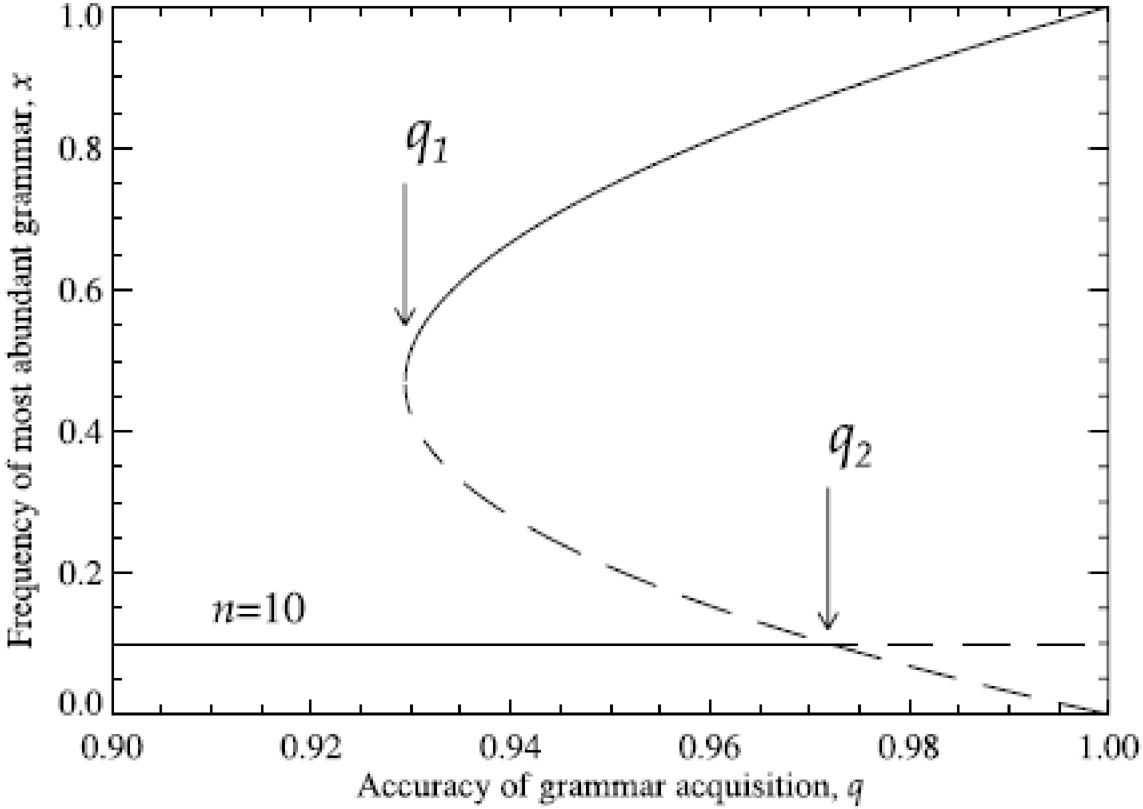
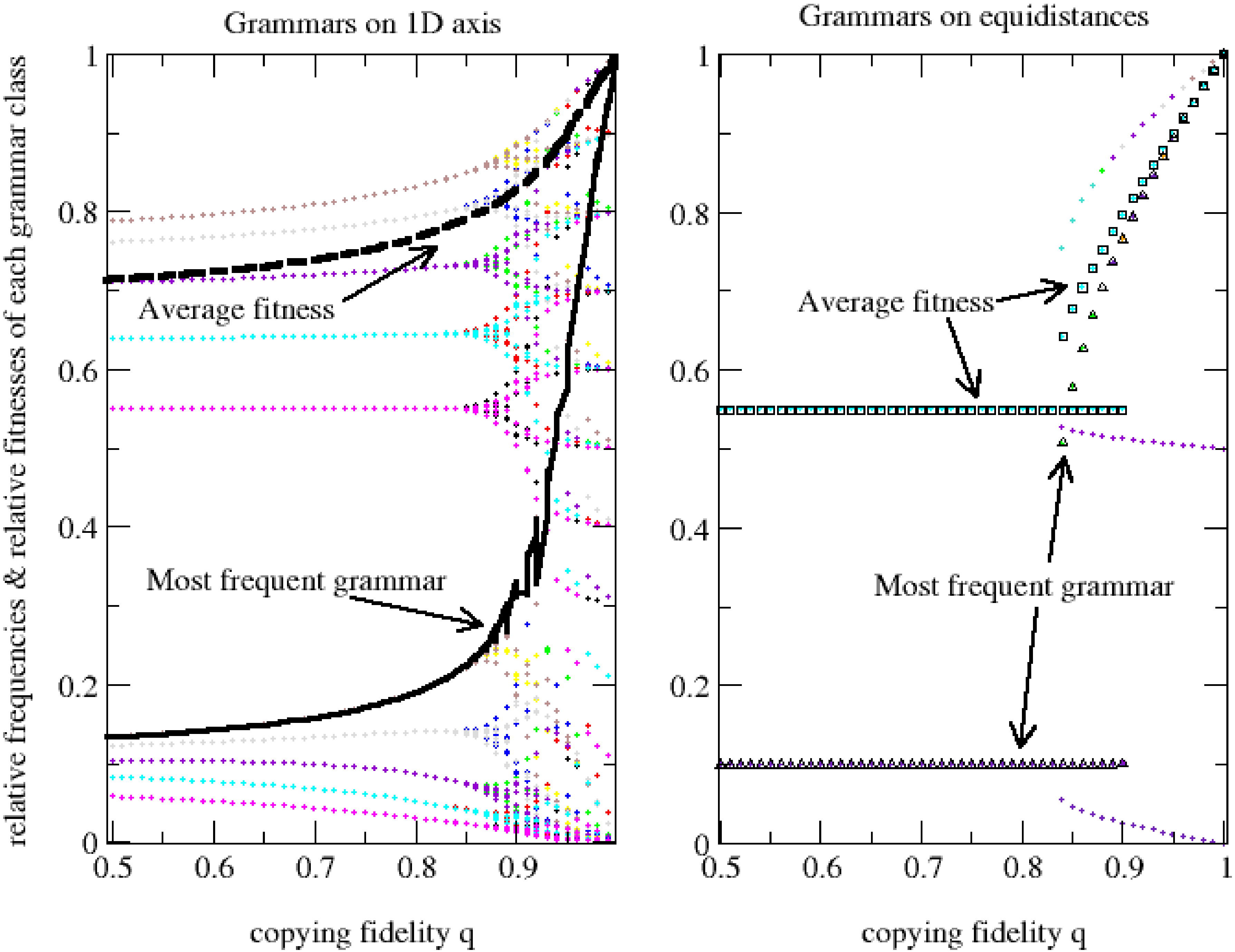


Fig. 1. Bifurcation diagram showing the frequency of the most abundant grammar in a population versus the probability, q, that children correctly acquire the grammar of their parents. All n grammars have the same distance from each other. We have $a_{ij} = a$ for $i \neq j$ and $a_{ii} = 1$. The n asymmetric solutions are given by $x_i = X$ and $x_j = (1 - X)/(n - 1)$ for a specific i and all $j \neq i$. For $n \gg 1$, we have $X = (q/2)(1 \pm \sqrt{D})$, where $D = 1 - 4[(1 - q)/q^2][a/(1 - a)]$. The equilibrium frequency of the predominating grammar does not depend on n. In the limit where q approaches 1, the frequency of the most common grammar is simply given by q and grammatical coherence is given by q^2. The asymmetric solutions exist for $q > q_1$. Each solution has a stable and an unstable branch. The symmetric solution, $x_i = 1/n$, loses stability for $q > q_2$. Parameter values are $a = 1/2$ and $n = 10$. Dashed lines correspond to unstable solutions.


Grammars on 1D axis
Grammars on equidistances
relative frequencies & relative fitnesses of each grammar class
Average fitness
Most frequent grammar
Average fitness
Most frequent grammar
copying fidelity q
copying fidelity q
0
0.2
0.4
0.6
0.8
1
0.5
0.6
0.7
0.8
0.9
1

Coherence Threshold

There exists a precise threshold value q_1 for the “copying fidelity” q of language

Below this “coherence threshold”, $q < q_1$, no evolutionary optimization takes place.

E.g. the MEMORY-LESS LEARNER receives examples sentences, and checks if they are consistent with the presently hypothesized grammar. If not, it changes the hypothesis to a random other grammar.

Can we calculate the probability that the algorithm has found the right grammar (out of N possible ones) after b example sentences?

$$\begin{aligned} q_{memoryless} &= 1 - \frac{(N-1)}{N}\left(a + \frac{(N-2)(1-a)}{N-1}\right)^b \\ &= 1 - \frac{(N-1)}{N}\left(1 - \frac{(1-a)}{N-1}\right)^b \end{aligned}$$

The BATCH LEARNER, in contrast, memorizes all received sentences and finds all grammars from the set of possible ones that are consistent with these sentences.

The probability that the batch learner has found the correct grammar after b input sentences is found by Nowak et al. (2001) to be

$$q_{batch} = \frac{\left(1 - \left(1 - a^{b}\right)^{N}\right)}{\left(Na^{b}\right)}$$

Major Transitions in the Evolution of Language

(Jackendoff'02)

Pre–existing primate conceptual structure
closed system of signals

Learned sounds and
sound–meaning associations

Combinatorial
Phonology

Compositional
Semantics

Hierarchical, Recursive

Phrase–Structure

Take-home Messages

1. Human language as a whole is unique, but for all its design features we see echos in animal communication;
2. There is a lot of comparative empirical evidence, that allows us to formulate hypotheses on stages and transitions in the evolution of human language;
3. We need additional theoretical constraints to evaluate theories – the “coherence threshold” is just the first such constraint;

- `http://staff.science.uva.nl/~jzuidema`
  → research (phd-thesis & bibliography)
- `jzuidema@science.uva.nl`
- `http://staff.science.uva.nl/~vanrooy/`
- *Evolutionary Game Theory & Language Evolution* (MOLEGT6), 6 credits, semester II(1&2), Jelle Zuidema & Robert v. Rooij

- NOWAK, M. A., KOMAROVA, N. & NIYOGI, P. (2001). Evolution of universal grammar. *Science* **291**, 114–118.
- JACKENDOFF, R. (2002). *Foundations of Language*. Oxford, UK: Oxford University Press.