The Evolution of Language: From Bodily Mimesis to Grammar

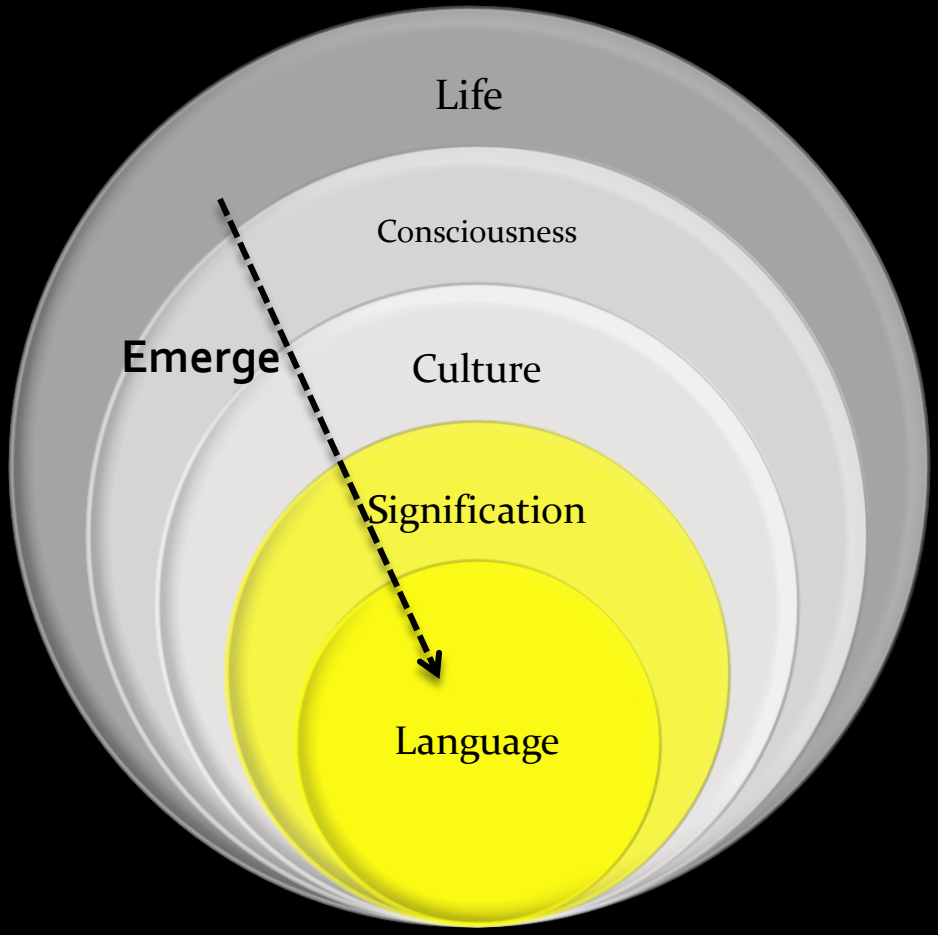
Lectures in Cognitive Semiotics, ENS, Paris 2012

Lecture 4

Jordan Zlatev

Department of Linguistics
Centre for Cognitive Semiotics (CCS)
Lund University

The Ladder of Meaning



Three transitions

1. From “mirroring” and (practical) imitation to sign use
2. From whole-body sign use to speech
3. From “protolanguage” to modern language with complex grammars

Overview

1. "The Mimesis Hierarchy": a conceptual tool for comparing mimetic capacities in human beings and (other) animals (Zlatev 2008a)
2. Support from "social neuroscience" (Zlatev 2008b)
3. Recent experimental and theoretical work on the transition from bodily mimesis to speech
4. Grammaticalization as evidence for the cultural evolution of grammar (Heine & Kuteva 2007)

1. Levels and kinds of bodily mimesis

Evidence from comparative psychology

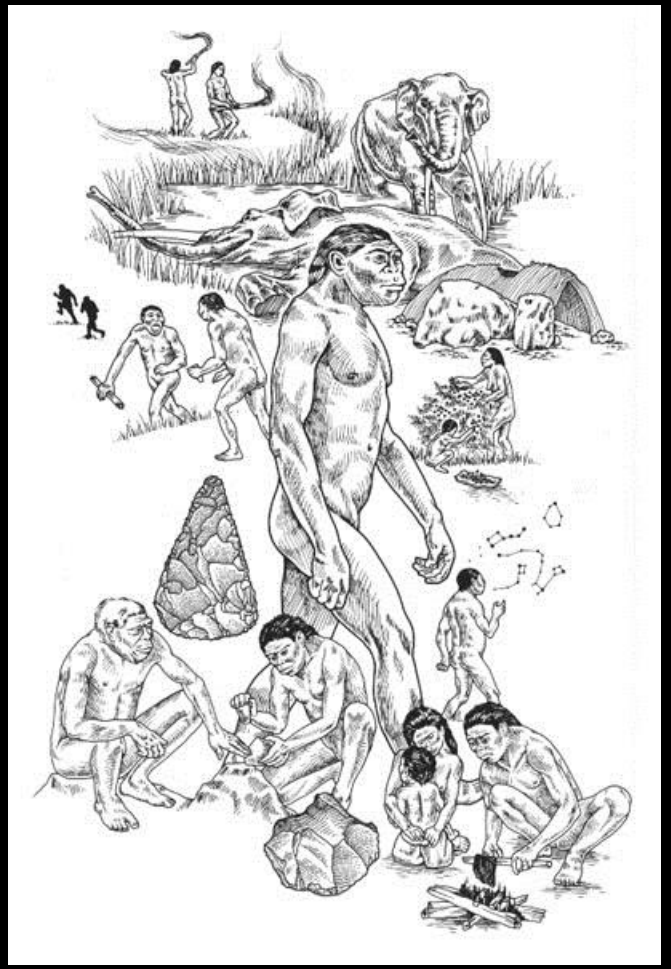
?

Stage	Species/period	Novel forms of representation	Manifest change	Cognitive governance
EPISODIC	primate	complex episodic event-perceptions	improved self-awareness and event-sensitivity	episodic and reactive; limited voluntary expressive morphology
MIMETIC (1st transition)	early hominids, peaking in *H. erectus;* 4M-0.4 Mya	nonverbal action-modelling	revolution in skill, gesture (including vocal), nonverbal communication, shared attention	mimetic; increased variability of custom, cultural "archetypes"
MYTHIC (2nd transition)	sapient humans, peaking in *H. sapiens sapiens;* 0.5 Mya - present	linguistic modelling	high-speed phonology, oral language, oral social record	lexical invention, narrative thought, mythic framework of governance
THEORETIC (3rd transition)	recent sapient cultures	extensive external symbolization, both verbal and nonverbal	formalisms, large scale theoretic artifacts and massive external memory storage	institutionalized paradigmatic thought and invention

(Donald 2007: 218)

Mimesis as "the missing link"

- *"Mimetic skills or mimesis rests on the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic."* (Donald 1991: 168)
- Mime, gesture, imitation, skill, mimetic imagination
- A domain-general adaptation, possibly initially for tool use (in early *Homo*)



The generality of the concept raises issues…

- Present day apes have at least some mimetic skills. Why not the "common ancestor"?
- If originally non-communicative, how did it get recruited for intentional communication?
- What were its own precursors, and how did the transition to a conventional (though still "grammarless") protolanguage come about?

Bodily mimesis (Zlatev 2005, 2007, 2008a, 2008b)

An act of cognition or communication is an act of bodily mimesis if and only if:

1. It involves a *cross-modal mapping* between exteroception (e.g. vision) and proprioception (e.g. kinesthesia).
2. It is *under conscious control* and *corresponds to* some action, object or event.
3. The subject *intends* the act to stand for some action, object or event *for an addressee* (and for the addressee to recognize this intention).
4. It is *not fully conventional* (and normative).
5. It does *not divide (semi)compositionally* into meaningful sub-acts that *systematically relate* to other similar acts (as in grammar).

The Mimesis Hierarchy (Zlatev 2005, 2007, 2008a, 2008b)

Level	Acts
Post-mimesis 2 (= Language)	… dividing (semi)compositionally into meaningful sub-acts that systematically relate to other similar acts (as in grammar).
Post-mimesis 1 (= Protolanguage)	… that are conventional-normative.
Triadic mimesis (Explicitly communicative)	… intended to stand for some action, object or event for an addressee (and for the addressee to recognize this intention).
Dyadic mimesis	… under conscious control and corresponding to to some action, object or event.
Proto-mimesis	… involving cross-modal mapping between exteroception (e.g. vision) and proprioception (e.g. kinesthesia).

Level	Key novel feature	Relevant capacities
Post-mimesis 2 (= Language)	• Systematicity	- Spoken or signed language
Post-mimesis 1 (= Protolanguage)	• Normativity (strong conventionality)	- One-word utterances - Holophrases - Emblematic gestures
Triadic mimesis (Explicitly communicative)	• Communicative signs	- Declarative pointing - Iconic signs - Full joint attention
Dyadic mimesis	• Volition and Perspective-taking	- True imitation - Cognitive empathy - Mirror self-recognition - Shared attention
Proto-mimesis	• Exteroception-proprioception mapping	- Emotional and attentional contagion - Neonatal mirroring - Mutual gaze

- Method: Compare capacities between pre-linguistic children and non-human primates

1. Proto-mimesis: *Neonatal mirroring*

In new-born children
(Meltzoff & Moore 1977, 1983)

In resus macaques
(Ferrari et al. 2006)

1. Proto-mimesis: *Mutual gaze*

- Rates of MG between infants and their mothers nearly the same in 3-month old children (18-20) and chimpanzees (17)

- A group ("cultural") difference between the apes at Primate Research Institute, Japan and those at Yerkes National Primate Research Center, USA

  - USA: 12 episodes/h. + 71% cradling of infant
  - Japan: 22 episodes/h. + 40% cradling of infant

  Bard et al. (2005). Group differences in the mutual gaze of chimpanzees (Pan troglodytes). *Developmental Psychology 41*, 616-624.

LARGE species differences in mutual gaze: somewhat older infants (apes in Leipzig, children in Lund, 3 h/subject)

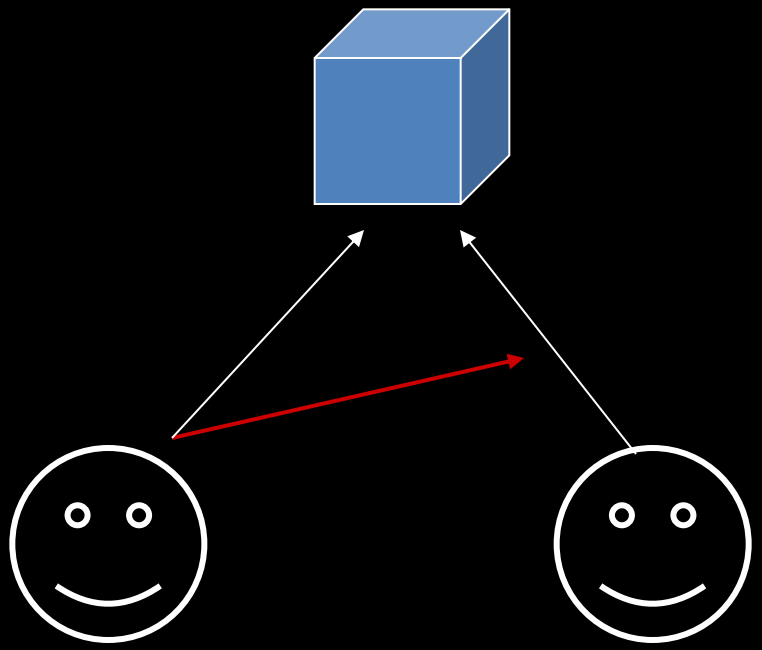
Apes	MG/hour	Length (sec)	Children (fictive names)	MG/hour	Length (sec)
Luiza (bonobo, 7m)	4.38	1.02	**Kate (5;0m)**	33.96	3.77
Kara (chimp, 5m)	1.45	0.61	**Mårten (5;18m)**	68.93	2.89
Kofi (chimp, 4.5m)	2.23	0.75	**Walt (5;22m)**	34.74	4.93
Lobo (chimp, 19m)	1	0.85	**Anna (7;6m)**	14.77	2.29
Sammi (gorilla, 9m)	0.67	0.62	**Eric (8;11m)**	22.05	2.11
Mean: 8.9m	**1.94**	**0.82**	**Mean : 6;11m**	**34.89**	**3.33**

(Zlatev et al. forthcoming)

1. Proto-mimesis: *contagious yawning*

- chimpanzees (Anderson et al. 2004; Campbell et al. 2009; Campbell and de Waal 2011),
- gelada baboons (Palagi et al. 2009)
- domestic dogs (Joly-Mascheroni et al. 2008; Madsen & Persson 2012)
- children < 4 years (Piaget 1962; Zlatev et al., in preparation)

2. Dyadic mimesis



<u>Second</u>-order mentality

- Cognitive empathy, e.g. consolation (Preston & de Waal 2002)
- Understanding others' intentions (e.g. Hare, Call & Tomasello 2001)
- Perceptual intersubjectivity, "joint attention" (Zlatev, Brinck and Andrén 2008)

Perceptual intersubjectivty

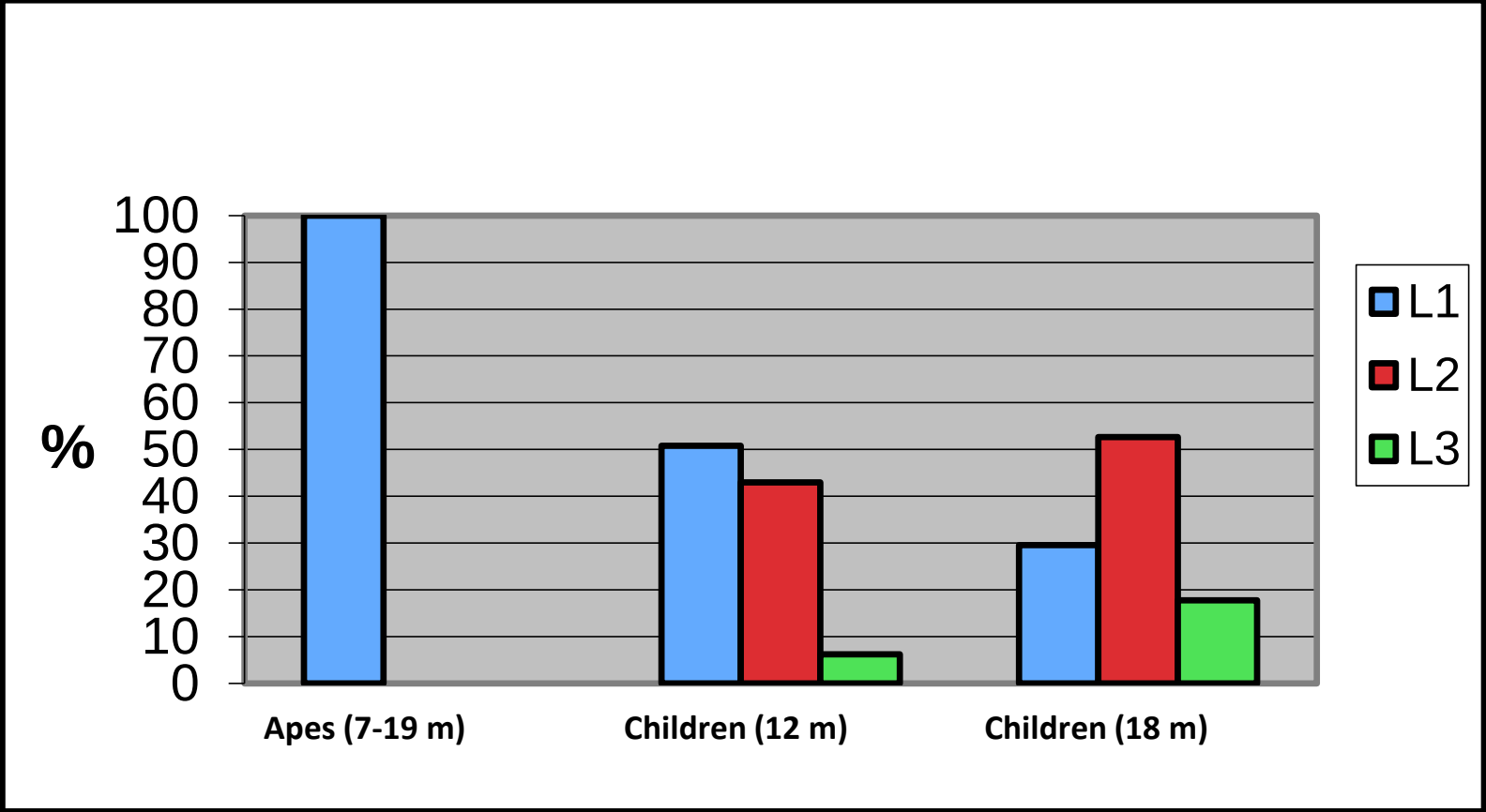
(Zlatev, Brinck & Andrén 2008)

- the process in which two or more subjects focus their attention on the same perceptually given target (more commonly known as "joint attention")

Level	Capacities
(I) Synchronous PI	A and B **synchronize** their (not intentionally communicative) actions in time and space
(II) Coordinated PI	A and B **coordinate** their (intentionally communicative) actions in time and space
(III) Reciprocal PI	A and B perform their (intentionally communicative) actions **in acknowledgement** to those performed by the other

Level	Term	Operational definition
1	Synchronous PI	INFANT: GAZE TO NEW T (+ NON-COMM. REF BEHAVIOUR TO T) ADULT: GAZE TO T
2.1	Coordinated PI - simple	INFANT: GAZE TO NEW T + COMM. REFERENTIAL BEHAVIOUR TO T ADULT: GAZE TO T
2.2	Coordinated PI - complex	INFANT: GAZE TO NEW T + COMM. REF BEHAVIOUR TO T + GAZE-TURNING TO ADULT ADULT: GAZE TO T
3	Reciprocal PI	INFANT: GAZE TO NEW T + COMM. REF BEHAVIOUR TO T + GAZE-TURNING TO ADULT + MUTUAL GAZE WITH ADULT (WITHIN IN OWN TURN) ADULT: MUTUAL GAZE WITH INFANT + GAZE TO T

Episodes of perceptual intersubjectivity, N = 190



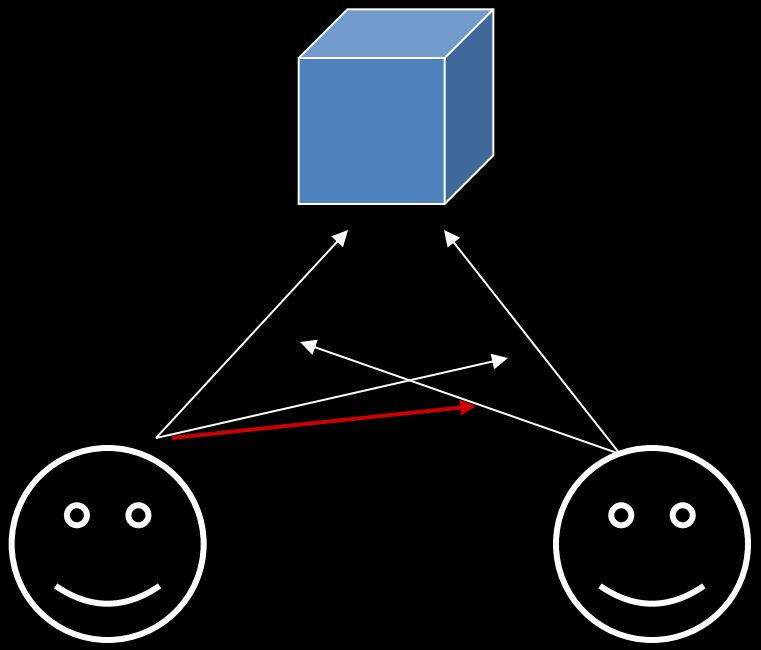
(Zlatev, Brinck & Andrén 2008)

2. Dyadic mimesis: *do-as-I-do imitation*

Novel actions	
10. Shake hand:	the right hand was shaken loosely from the wrist
11. Raise one arm:	the right arm was put into the air
12. Stamp foot:	the right foot stamped the floor several times
13. Pat stomach:	the stomach was patted alternately with the palm of each hand several times
14. Raise two arms	both arms were put into the air simultaneously
15. Touch chin:	the right index finger was placed on the chin
16. Praying hands:	both palms were touching each other
17. Wipe face:	the palm of one hand was wiped down over the face several times
18. Slap floor:	the floor was slapped several times with the right palm
19. Raise foot:	the right foot was raised from the floor
20. Wipe floor:	the right palm was wiped from side to side across the floor several times
21. Touch armpit:	the left arm was raised and the right index finger was placed on the left armpit
22. Grab wrist:	the left wrist was grasped by the right hand
23. Swing arm:	the right arm was swing back and forth several times
24. Wipe hands:	the palms were wiped together several times

Male juvenile chimpanzee trained in 92 sessions for 7 months, 17% transfer
(Hribar, Call & Sonesson 2011)

3. Triadic mimesis



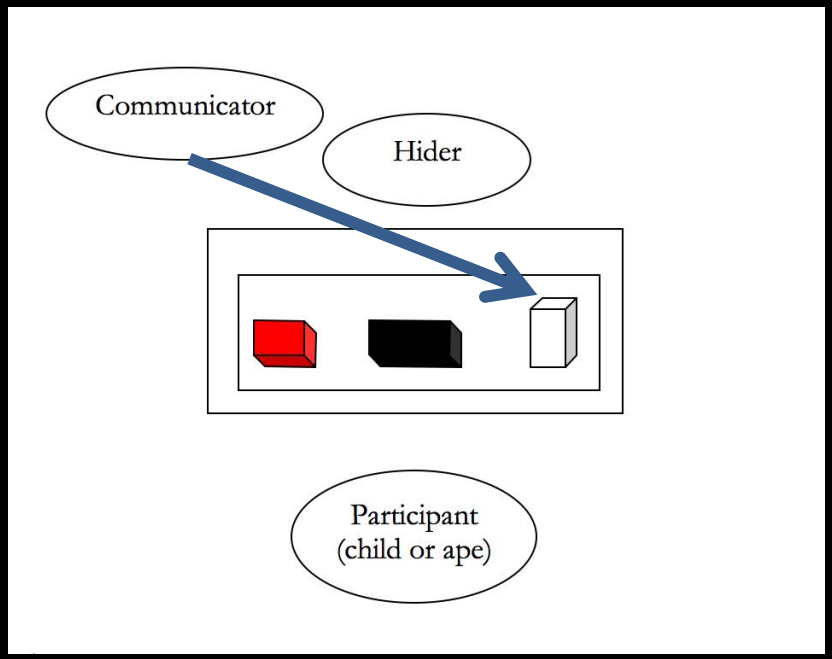
Third-order mentality

- Full joint attention: "I see that you see what I see" (and vice versa) – beyond "shared attention" (second-order mentality)
- Communicative intentions: "I want you to do/understand X by means of recognizing my intention." (cf. Grice 1957)

“Object choice task”

Location of reward is communicated by different type of cues:

1) Pointing to X
2) Placing a marker on X
3) Showing a replica of X
4) Showing a picture of X



"Object choice task"

… no ape was able to do this for any of the communicative signs that they did not know before the experiment. One explanation of these results is that the apes were not able to understand that the human beings had intentions toward their own attentional states.

The children, in contrast, treated each communicative attempt as an expression of the adult's intention to direct their attention in ways relevant to the current situation.

(Tomasello 1999: 102)

Remaining Issues: Children

- Age
  - Tomasello *et al.* (1997) was based on 30/36 month old children – why not younger?
  - Behne *et al.* (2006): why only pointing and ostensive gaze?
  - DeLoache (2000): pictures at 30 months, Replicas at 36 – but in a more difficult task
- Role of language
  - A "language independent" explanation cannot be *assumed* for children (almost) 3 years old

Remaining Issues: Apes

- Rearing history and familiarity
  - Herrmann *et al.* (2006) – no understanding of pictures in an object-choice task, but Hribar *et al.* (in press): an imitation trained chimpanzee does understand.
  - The objection of Leavens *et al.* (2008): is it fair to compare emotionally deprived apes with middle-class children?
  - Lyn *et al.* (2010): Language-trained (enculturated) chimpanzees produce and understand declarative points.

Remaining Issues: Method

- Types of cues used
  - Cues are used rather unsystematically in the various studies (and perhaps negative results unreported)

- Three or two boxes?
  - Why was the 3-box design from Tomasello *et al.* (1997) substituted for a 2 box-design, without discussion?
  - Distance between the boxes matters for non-human subjects: why?

Research Questions

1. *Would the children's performance on pointing be different from markers?*
2. *Would there be a correlation between the children's linguistic skills and their comprehension of communicative intentions/signs?*
3. *Would there be evidence of a developmental progression: Pointing > Marker > Picture > Replica*
4. *Would the chimpanzees perform better with a familiar than unfamiliar communicator, and if so: for which cues?*
5. *Would there still be a clear difference between the children and the chimpanzees in understanding communicative intentions, and if so: for which cues and in which ages?*

Hiding-Finding Game (adapted from Tomasello *et al.* 1997)

<u>Participants</u>

Children: 72 : 18 m, 24 m, 30 m (24 per group), at Humanities Lab, Lund

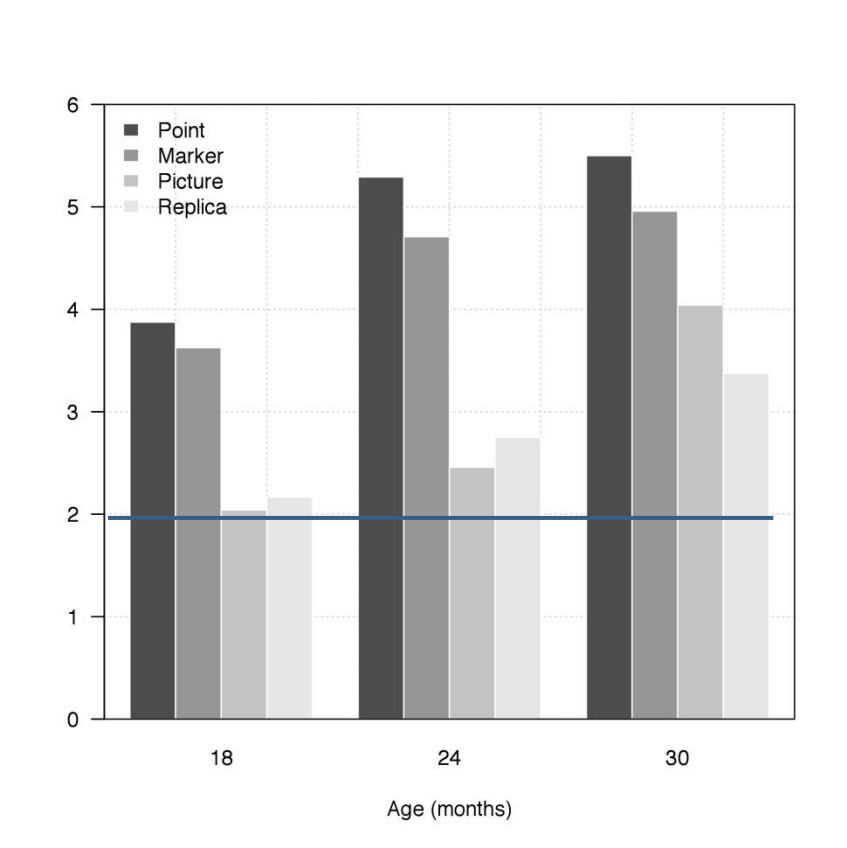
Chimps: 4 (3 adults, 1 juvenile): housed at Lund University Primate Research Station Furuvik (LUPRSF)

The communicator

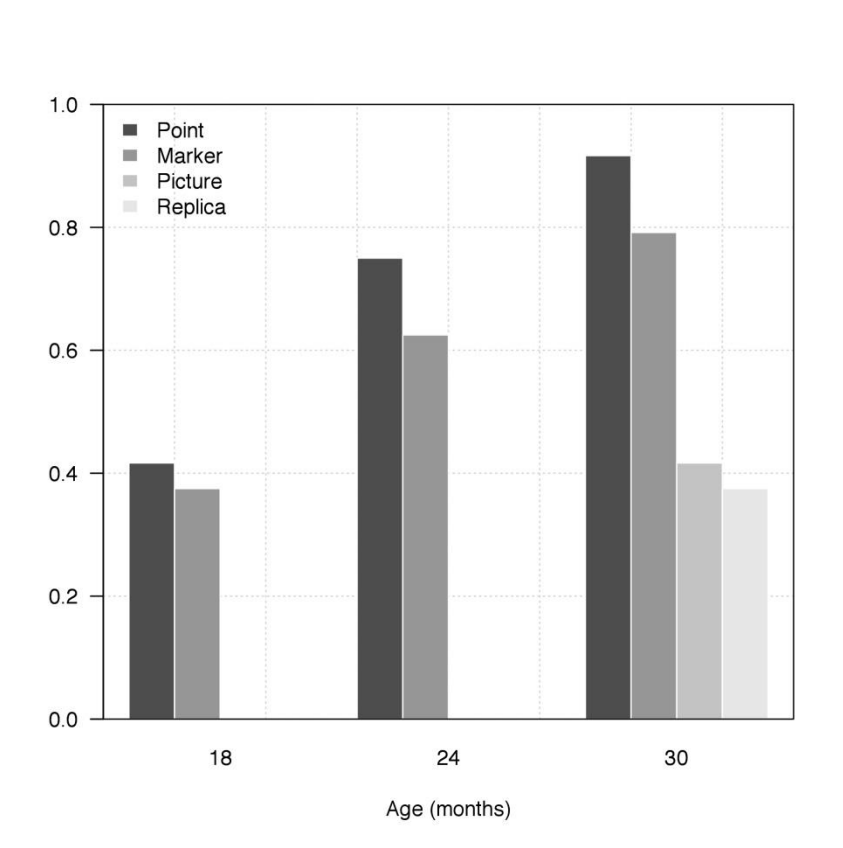
- Obtains the subject's attention
- Expresses "helpfulness" by facial gestures
- Produces one of the following cues:
  - **Point** (Proximal, dynamic, index finger point to baited box)
    Gaze: BOX-CHILD
  - **Marker** (Places a yellow "post-it" note on top of baited box)
    Gaze: BOX-CHILD
  - **Picture** (Holds up photo of the baited box in mid position)
    Gaze: PHOTO-BOX-CHILD
  - **Replica** (Holds up an identical replica of baited box in mid position
    Gaze: REPLICA-BOX-CHILD

Design for chimpanzees

Results: children

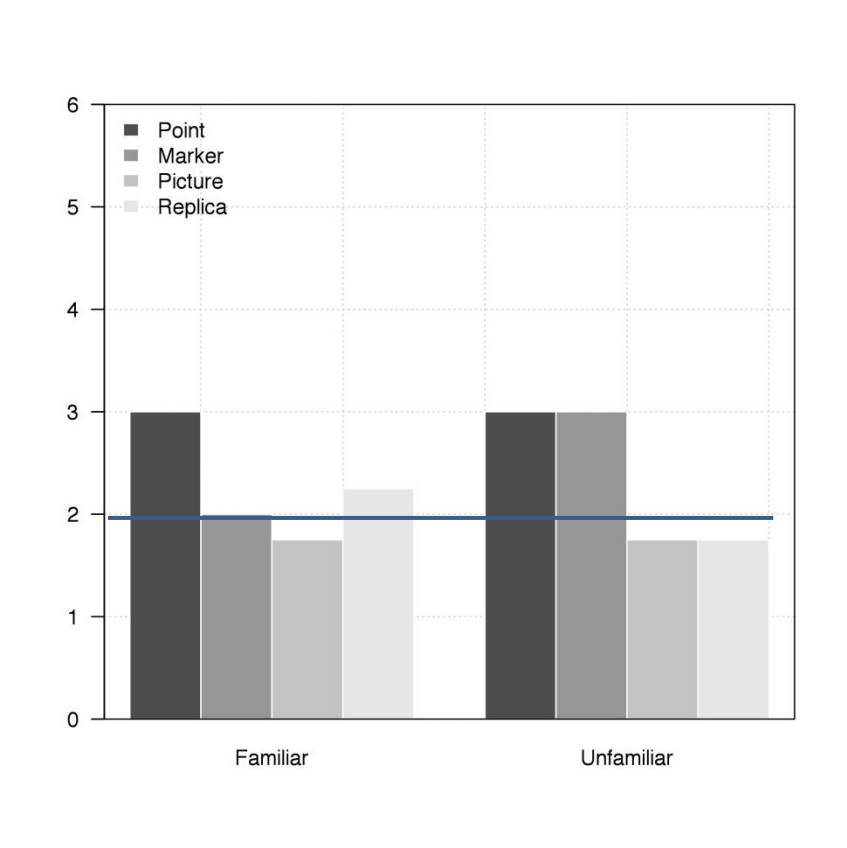


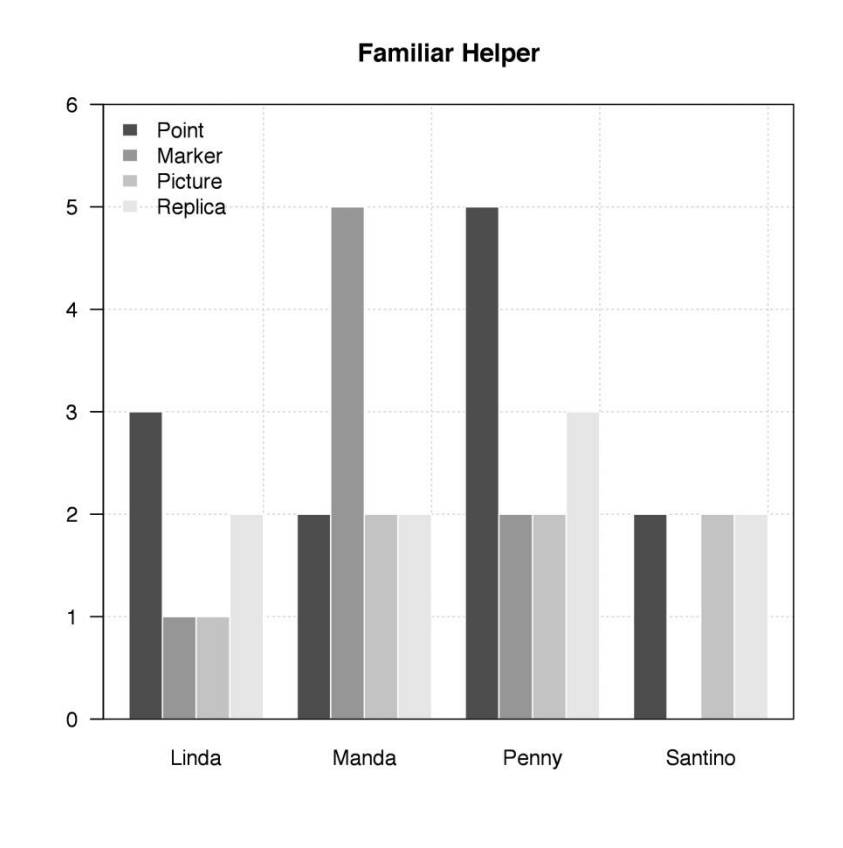
Mean number of trials correct (Max = 6)



% children, at least 5 trials correct

Results: chimpanzees





Mean number of trials correct (Max = 6)

Zlatev et al. (submitted)

Results

- Children
  - 18 < 24 m (Pointing, Marker)
  - 24 < 30 m (Picture, Replica)
  - Pointing = (?) Marker > Picture = (?) Replica
  - No correlation with Vocabulary scores
- Chimps
  - Familiar = Unfamiliar Communicator
  - Indexical cues > Iconic cues (tendency)
- Comparative
  - Children > Chimps (not only due to language)

Summary: fully developed capacity for bodily mimesis is uniquely human

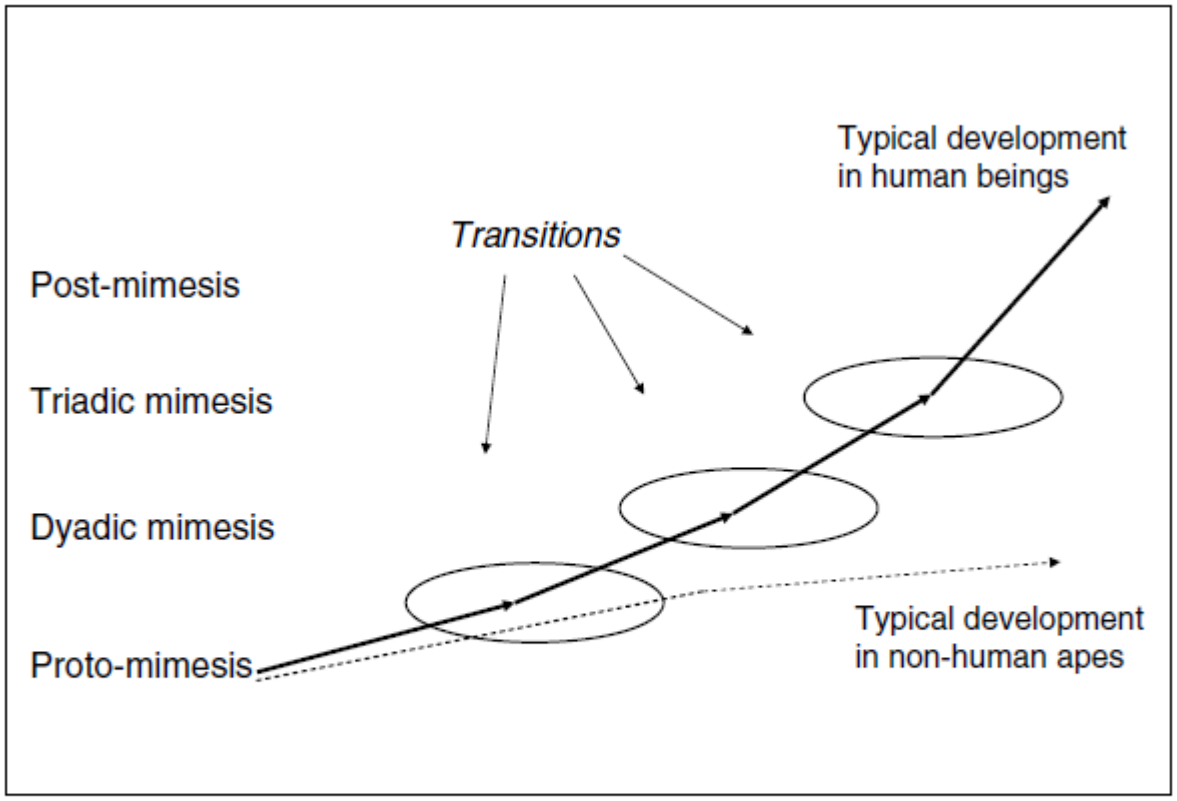
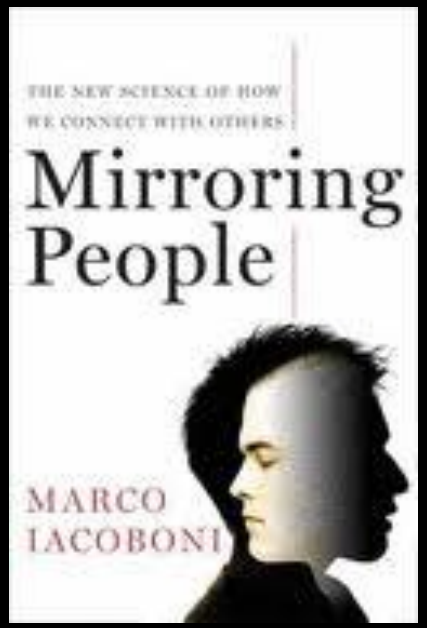


Fig. 1. A schematic representation of the divergent developmental trajectories of (typically developing) children and apes in terms of the stages of the Mimesis Hierarchy model (the two levels of Post-mimesis are not distinguished).

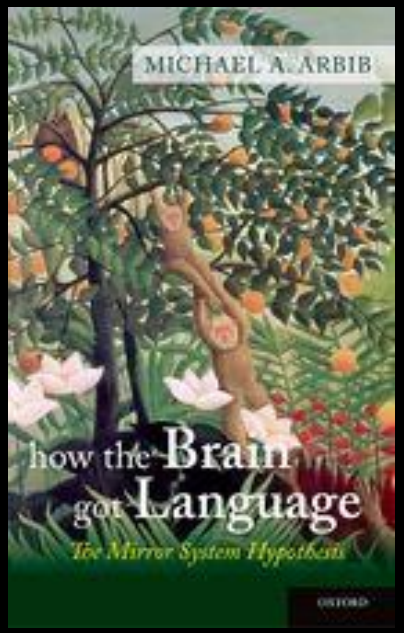
2. Evidence from social neuroscience

From “mirror neurons” to the neural substrates of gesture and speech

Two (different) recent books



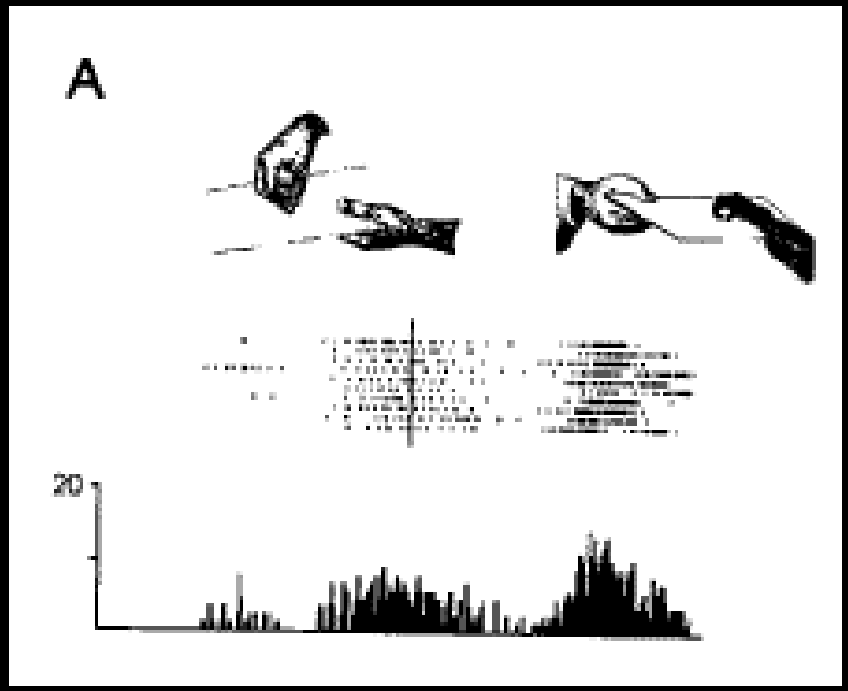
Iacoboni (2008)



Arbib (2012)

“Mirror” and “canonical” neurons in macaque premotor cortex

- Canonical neurons: active during observing (static) scenes that afford action
- Mirror neurons: active both during execution and observation of similar movements/actions



(Rizzolatti et al. 1996)

Great expectations…

- It is amazing how these cells have been proposed as a solution to just about every mystery in the *human* mind: from empathy to imitation, "mind-reading", language (evolution), autism, homo-sexuality….

- It is not surprising that their role has been regarded as much over-rated by some researchers in the field (Preston & de Waal 2002; Hurford 2004; Donald 2005; Csirba 2007).

Problems

- Directly recorded only in monkey brains, with only indirect evidence (PET, fMRI, TMS, EEG) for ape and human brains.
- Not sufficient for “simulation”, representation or signification: X stands for Y for subject S
- Present in macaques, while monkeys can neither imitate, gesture, nor use language…

Basic proposal

- The MNS serves as a likely neural basis for the evolution of bodily mimesis and language

- Under the condition: from "mirror neurons" to increasingly expanded neural circuits, involving most of the brain!

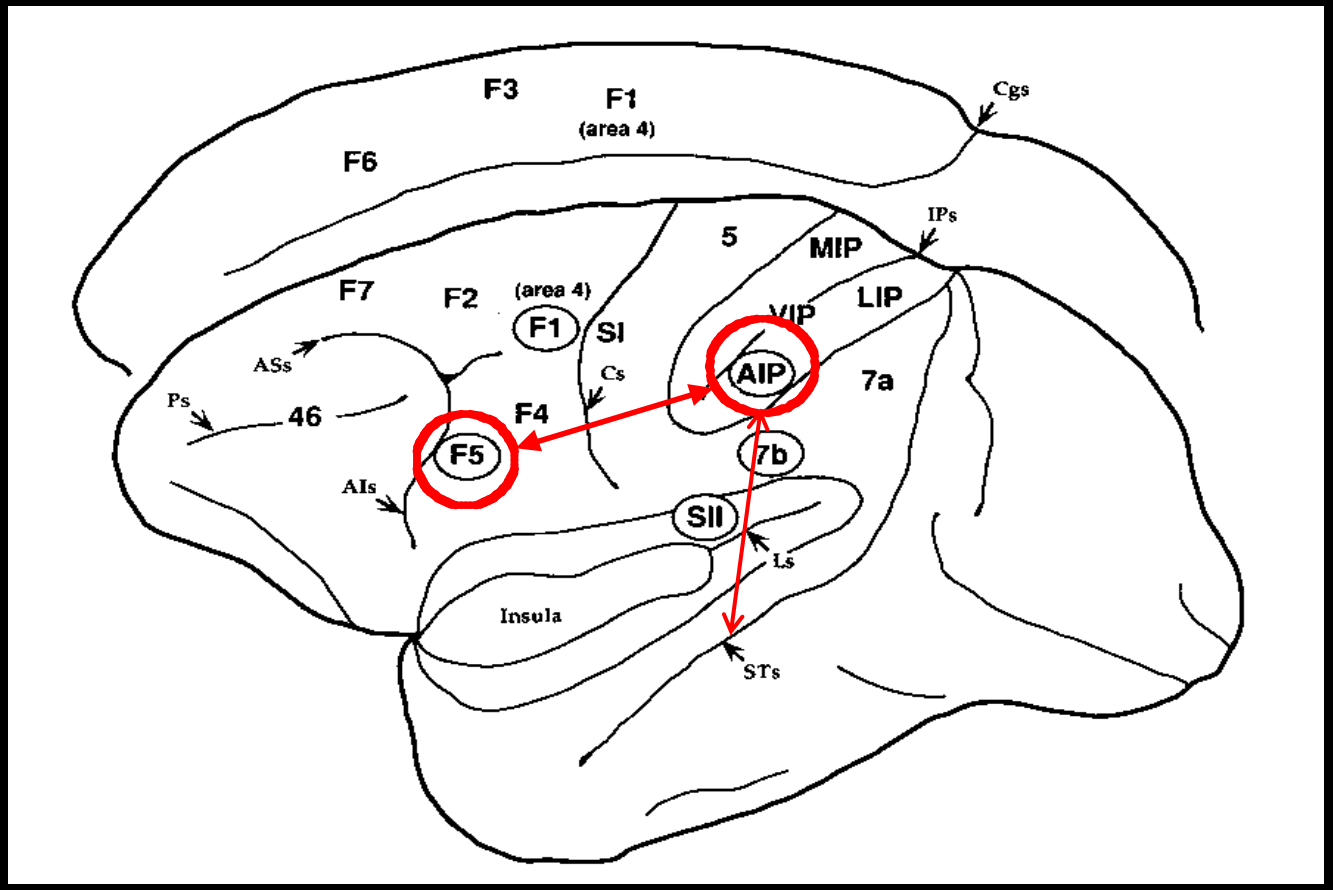
- Basic idea: gradual (step-wise) evolution of the monkey mirror neural system for manual actions; similar to Arbib (2005, 2012), but with a few important differences

1. Proto-mimesis (monkeys)

F5: area in premotor cortex

AIP: anterior intraparietal area

ST: Superior Temporal

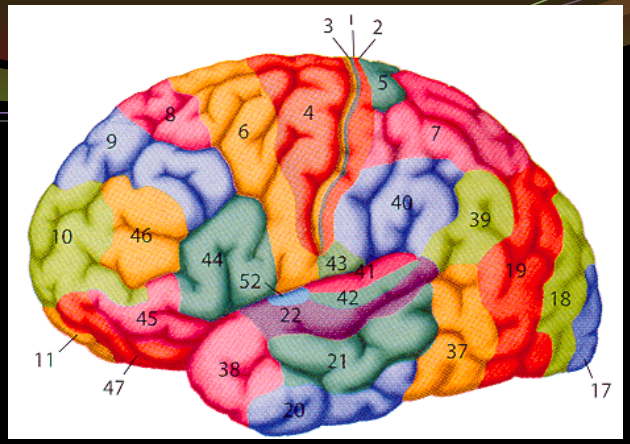


Arbib (2005)

1.Proto-mimesis: *resonance with the other*

- "Mirror neurons" are part of a frontal-parietal-temporal system in the monkey brain, which
- Responds to an open-ended set of manual actions, of both self and other
- Provides a basis for
  - recognizing and anticipating the results of others' actions (without "theory of mind")
  - affective empathy and contagion (Preston & deWaal 2002)

2. Dyadic mimesis:
(Great Apes, *Homo sapiens*)

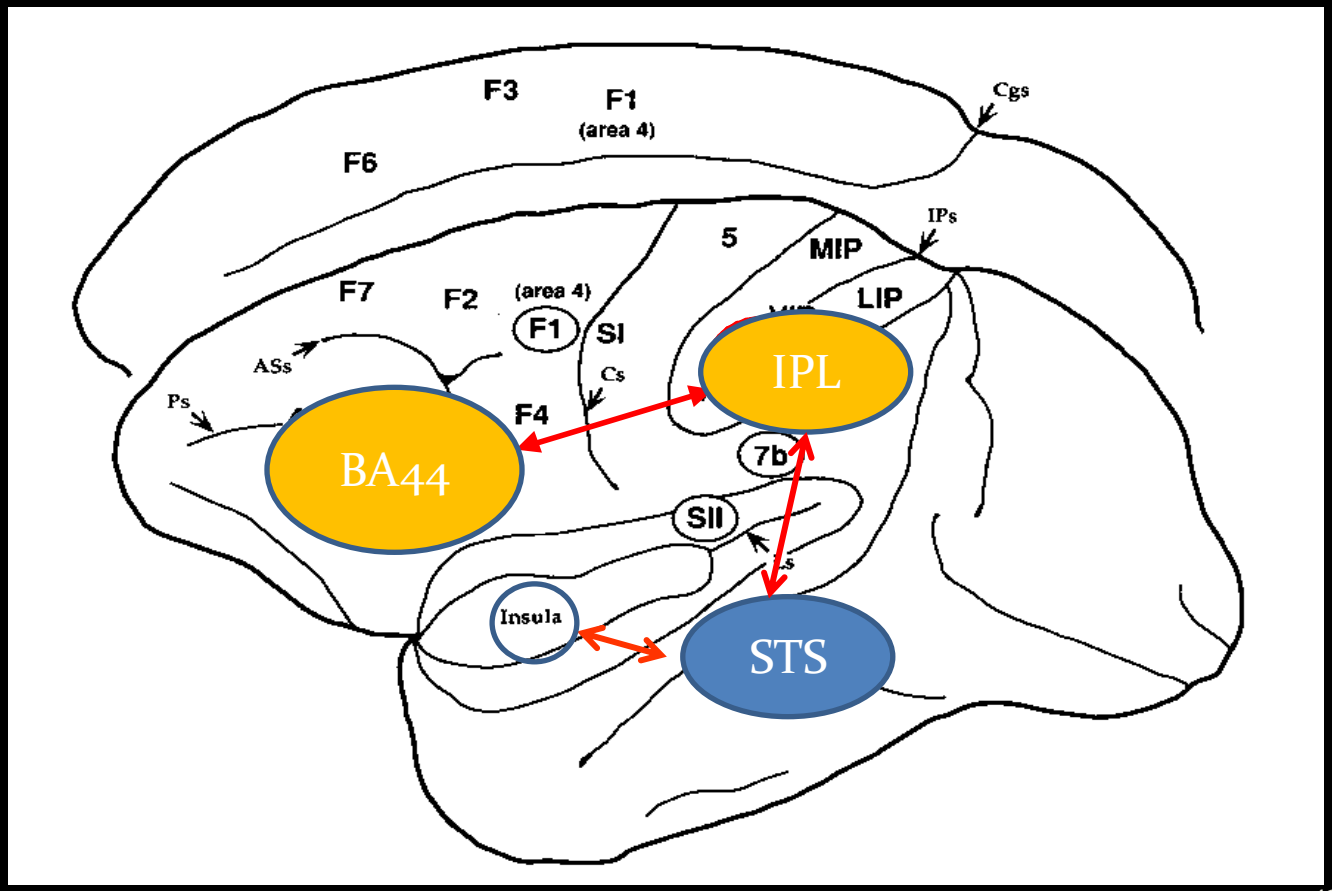


BA 44: *pars opercularis* of the inferior frontal gyrus

IPL: inferior pariatal lobule (body image)

STS: superior temporal sulcus (biological motion)

Insula: project to limbic system



2. Dyadic mimesis:

imitation and self-recognition

- The perisylvian cortex and the prefrontal cortex have expanded the most in the human brain compared to apes (Deacon 1997)
- On the basis of behavioral and anatomical data - also in apes compared to monkeys (Arbib 2005)
- Segregation of BA44 into a dorsal part (active during action) and a ventral part active only during imitation (Iacoboni 2005)
- Brain lateralization
  - Right-HS IPL active when imagining the motion of others
    Left-HS IPL when imagining self-motion
  - Right-HS STS: "analysis of other's action in relation to the intention of the self" (Decety and Chaminade 2005)
    Left-HS STS: analysis of other's motion

3.Triadic mimesis:

pointing and iconic gestures (*Homo erectus?*)

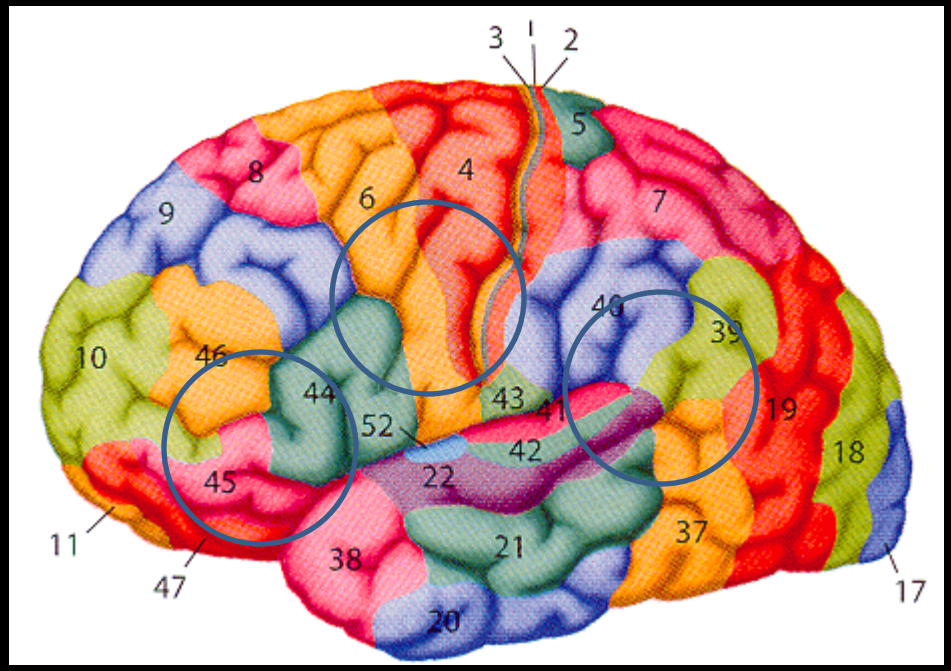
- The monkey F5 mirror neuron circuits do not respond to "intransitive actions" (i.e. pantomime)
- McNeill (2005): the content (imagery) of gestures is based on RH-activity, their "orchestration" mainly on LH ("Broca's area")
- Conjecture: Further extending, and differentiating between expression and content along with lateralization: iconic gestures (pantomime) and pointing: the first true signs!
- Combining iconic gesture and pointing: gestural predication: (Point-X, Iconic-gesture)

4.Proto-language:
Multimodal (gestural-vocal)

Additional adaptations of the system

- Segregation of "Broca's area": BA45 ("heteromodal") and BA 44 (primarily for speech)
- BA4a and BA6 in ventral pre-motor cortex (PMv) – active during both production and perception of meaningless syllables (Wilson et al 2004)
- Wernicke's area (Superior and middle temporal gyrus): "combining capabilities for recognizing proto-sign and proto-speech to support a language-ready brain that is capable of learning signed languages as readily as spoken languages" (Arbib 2005)

"The language-ready brain"



BA 44, 45 = "Broca"

BA 22, 39, 40 = "Wernicke"

Overlap extensively with the MNS (Arbib 2005; Iacoboni 2005; Decety & Chaminande 2005): in tasks of perception-action matching, imitation, imagination, pantomime…

BA 4, 6 = perception-production of meaningless syllables (Wilson et al. 2004)

No evidence anywhere for a "syntax module", "recursion", "faculty of language (narrow) etc…

3. From bodily mimesis to speech

The transition from gestural-vocal to vocal-gestural

Brown (2012)

"A major step in the evolutionary process by which human communication could have emerged has been proposed in the bodily mimesis hypothesis. … This ability provides a foundation from which symbolic communication can arise, but how such a transition would have taken place has not been fully examined. This thesis examines the gap between bodily mimesis and symbolic communication…" (: 1)

The Evolution of Symbolic Communication: An Embodied Perspective, PhD Thesis, University of Edinburgh

Three kinds of "gestural primacy"

- "Switch": (from gesture to speech): Deacon (1997); Corballis (2002); Arbib (2005): why not signed languages? (cf. Fitch 2010)
- "Immediate multimodal": McNeil et al. (2005; 2008) ("multimodal referential communication was a combination of arbitrary and non-arbitrary representation from inception": 116): underestimate the degree of non-arbitrariness of speech today
- "Gradual multimodal": Mithen (2004); Kita (2008); Zlatev (2008) (conventionalization and partial loss of iconicity in both modalities): "do not provide a reason why one modality is now predominantly symbolic and not the other" (:120)

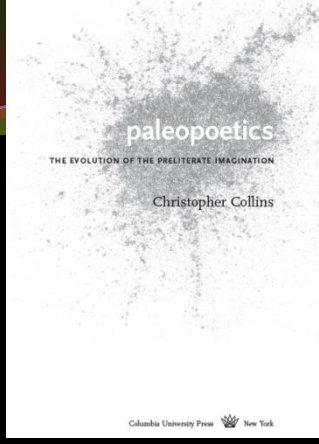
Brown's proposal

- "the conventionalization process requires a rich and supportive communicative infrastructure in … so that the intended form-meaning relationships could be correctly interpreted" (: 81)
- "the vocal modality would have become predominantly symbolic because its lower non-arbitrary capacity" (: 134)
- "symbolic re-interpretations … do not arise via intentional creation by individuals, but instead by non-arbitrary signs becoming symbolic when they are transmitted to others who did not participate in their creation and development" (: 191)

Evidence

- Review of emerging signed languages (Nicaraguan Sign Language, Bedouin Sign Language)
- Experimental sign research: the gestural modality carries more "communicative load" than the vocal modality when communication is restricted to non-conventional signaling (e.g. Fay & Lim 2010)
- Computational models of language evolution: the stabilization of a conventional semiotic code across a greater number of speakers requires extensive feedback, identical context, or – support from parallel non-arbitrary signals

Collins (2013)



"While human primates must have been at first better at transmitting information through gesture than through voice, at some point voice became the preferred vehicle. But what if this "point" was a transitional period of over half a million years, say, from the appearance of *Homo erectus* to that of archaic *Homo sapiens*? And what if, during all this time, humans regularly communicated bi-modally, only gradually shifting from a code that foregrounded gesture to one that foregrounded voice…?" (: 136)

Paleopoetics: The Evolution of the Preliterate Imagination,
New York: Columbia University Press

Collins' proposal

"Assuming that gesture was the earliest medium of human communication, it would have been complemented by what I have termed vocal "paragesture." [...] gesture and vocal paragesture were integrated in a bimodal communicative code, one in which gesture received focal attention, while its vocal accompaniment received subsidiary attention. At first these sounds must have resembled the barks and hoots of latter-day chimps and bonobos.

Only gradually did the human voice attain the articulatory control necessary, first, to mimic environmental sounds, such as animal and bird calls … and later, to utter a range of phonemes that could be serially linked and conventionally assigned particular meanings…" (: 139)

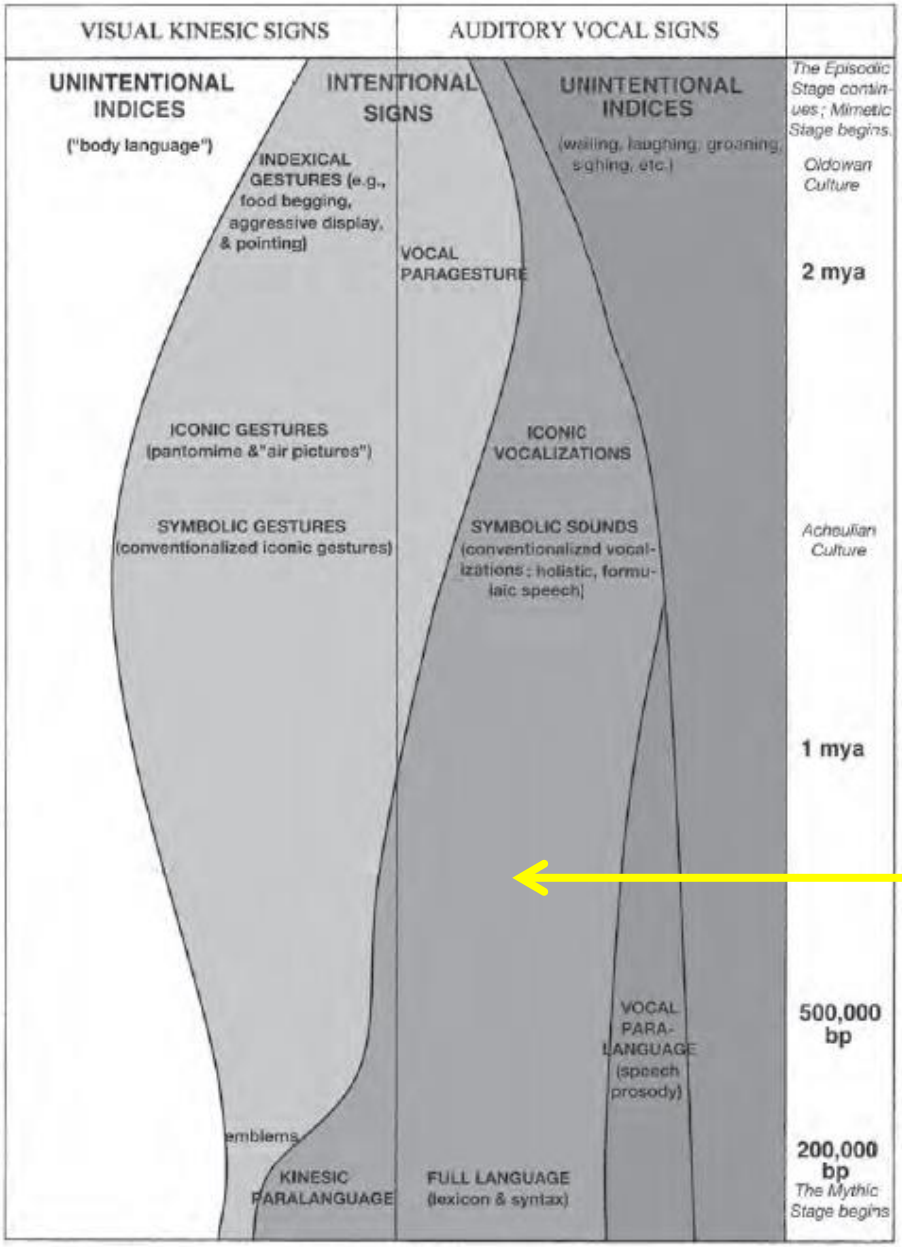


Figure 5.2 The evolution of semiotic skills from gesture to speech

Vocal language

Evaluation: a compromise

- Collins' proposal is more consistent with the essentially gradual nature of evolution and the archeological evidence (e.g. a long period of non-cumulative cultural evolution).
- It also squares in better with the increasing evidence for non-arbitrariness in speech, i.e. "sound symbolism" (Ahlner & Zlatev 2010).
- But Brown makes and important point that if the transition "from gesture to speech" involved an intermediary stage of "symbolic gestures" (cf. Arbib's "proto-sign"), that would have minimized support for the stabilization of a conventional code.

4. From speech to grammar

Exaptations and cultural evolution (with grammaticalization)

Possible exaptations (pre-adaptations)

- "where" (**x**) and "what" **P** pathways in the visual system, giving rise to **P**(**x**): "proto-predication" (Hurford 2003)
- Social scripts: "structured generalized patterns of social behaviour" (Johansson 2005), Aiello (1998)
- Tool-making scripts (Johansson 2005: 231)
- Hierarchical structure in action sequences (Greenfield 1991)
- Hierarchical structure of social cognition (Harder 2004)

Stages in the evolution of grammar

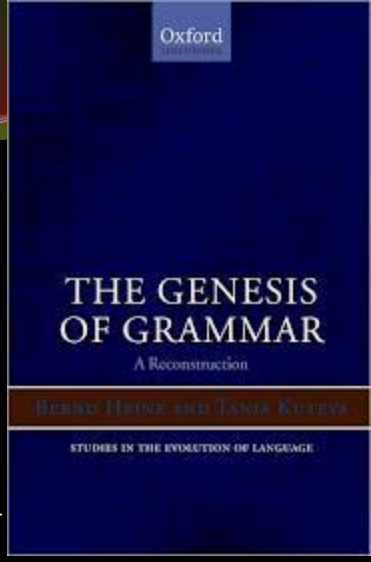
1. Utterance structure ("two word stage")
2. Hierarchical structure, but no recursivity (e.g. no subordination)
3. Flexible structure: "different ways to express the 'same' meaning" (: 234)
4. Recursive structure: "flows naturally from the ability to handle nested predications" (: 237)

Johansson, S. (2005) Origins of Language: Constraints on Hypotheses. Amsterdam: Benjamins.

Stages in the evolution of grammar

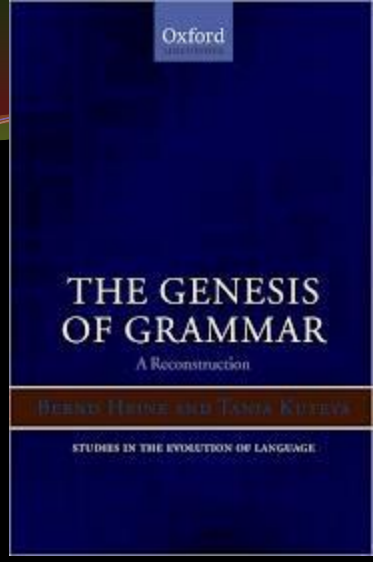
- It is far from clear that there are any biological adaptations for stages 2-4
- Hence, the evolution of grammar, at least from the "two word stage" can have proceeded through processes of cultural-historical, rather then biological evolution, during the last 100,000 years since the spread of our common ancestors out of (and throughout) Africa.
- Grammaticalization theory (Heine & Kuteva 2002, 2007) can help explain, and chart, this process of cultural evolution.

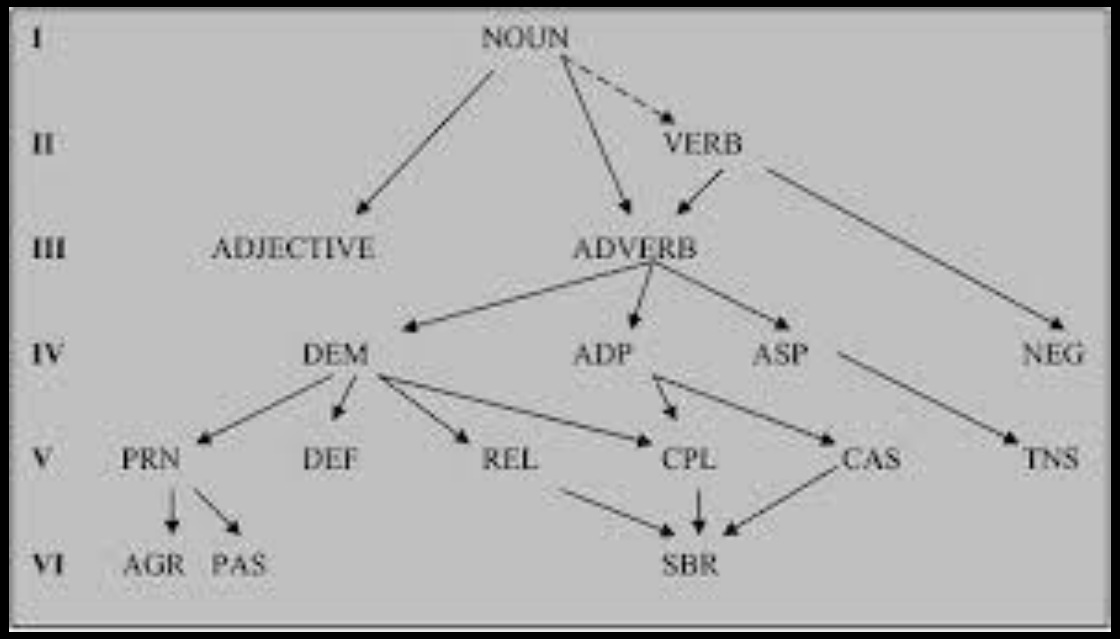
The cultural evolution of grammar



- Languages spoken 10,000 years ago, typologically not much different from present languages
- Push back in historical time, using generalizations concerning processes of grammaticalization (e.g. 'want' > FUT)
  - From lexical to grammatical, and "even more grammatical"
  - Basically unidirectional
  - Processes of grammaticalization have been similar in the past: process - not structure - "uniformitarianism"

Reconstucting “early language”





- Thing-words + Process-words
- No morphology or “grams”
- Only word order for argument structure
- Concepts for location, possession – but not grammatically expressed
- No personal pronouns, but names, roles

5.Language (vocal-gestural, signed)

- Built "atop" bodily mimesis, supported by a an expanding and differentiating MNS
- Exaptations for hierarchical structure for action and imitation (involving not only the extended MNS, but basal ganglia, pre-SMA and cerebellum)
- Grammar: from protolanguage (over the last 100 000 years) on the basis of historical "post-biological" (Arbib 2005)
- Writing and literacy: structural complexity along with morphological simplification (Evans 2009, *Dying Words*)

Summary: Stages in the evolution of human communication

YA	Species/kind	Dominant form of communication
6,000,000	Last common ancestor	vocal signals and "flexible", but not human-like gestures
4,000,000	*Ardipethcus ramides/ austalopithecines*	vocal grooming, mutual gaze
2,000,000	*Homo ergaster/erectus*	Triadic mimesis (pantomime and pointing) "vocomimesis"
500,000	*Homo hielderbegensis /neanderthalensis*	gestural-vocal proto-language
100,000	*Homo sapiens*	vocal-gestural "early language"
5000	Modern human beings	complex, grammatical, and multimodal language
0	Technological human beings	Writing, internet…

But remember: Evolution ≠ Progress