Mirror Systems: Evolving Imitation
& the Bridge from Praxis To Language

Michael A Arbib
arbib@usc.edu
University of Southern California

Part I

What Should a Theory of Language Evolution Explain?

A Diversity of Divides
in Devising System Models for Language Evolution
**Divide O: Universal Grammar**

What evolved was a Chomskian Universal Grammar and the accompanying Language Acquisition Device

-but if so, which Universal Grammar?

**Versus**

Diverse mechanisms evolved – both through natural selection and cultural evolution (and co-evolution) – to support language performance

To this end, different studies of language evolution must address:

- Data on language use in communication and thought, including use in conversation
- Data on language acquisition
- Data from historical linguistics
- Archeological data which seek to infer clues to the language of a culture from the material traces of that culture
- Animal communication systems in general or for primates in particular, seeking commonalities with and differences from language
- Animal behavior in general or primate behavior in particular, seeking commonalities with and differences from communication systems
- Data on brain function

---

**Biological: How did biological evolution endow Homo sapiens with brains and bodies that can acquire and use language?**

- **Divide 1:**
  - Our ancestors evolved a capability for protolanguage – which had an open lexicon but little if any syntax – before they developed language versus no intermediary was involved.

- **Divide 2:**
  - equating speech with language versus
  - stressing the shared mechanisms that support signing & speech

In either case, the importance of spoken language asks us to understand the evolution of the vocal apparatus which made it able to produce speech, and the related mechanisms for perception, and for the neural control of perception and production.

- **Divide 3:**
  - Biological evolution of language must make crucial use of data on the brain versus
  - the primary data are those of linguistics alone
**Cultural: What aspects of language are innate, and what are the fruits of historical change?**

**Divide 4:**
- Biological evolution endowed us with a Universal Grammar *versus*
- it gave us mechanisms which made the eventual *invention* of language possible

In either case, one needs to understand
- what the child learns during *language acquisition*, and what biological evolution provided to make such learning possible.
- what happens during *historical language change*, and what biological evolution provided to make such processes possible.

**Divide 5:**
- Initially, much of protolanguage was *holophrastic* (with a protoword describing a frequently occurring or significant situation) *versus*
- protolanguage started with words akin in scope to modern words (such as nouns and verbs)

**Divide 6:**
- Language evolution is to be understood solely in terms of adaptive pressures for communication or thought *versus*
- language evolution rests in part on the exaptation of adaptations that are not directly related to communication.

---

**64 Classes of Frameworks**

These six divides (and, of course, there are others) define 64 overall approaches to language evolution.

Any *general* framework must justify (at least) which side of each of the six divides it lies on.

By contrast, more focused models may ignore many of these issues to address the possible evolution of mechanisms responsible for some specific set of data, such as control of the human speech apparatus to learn to produce the observed sounds of a human languages.

Our choice of relevant data determines whether we look at
- *Performance Systems*
- *Developmental Systems*
- *Historical Systems*, and/or
- *Evolutionary Systems*
Relating this Lecture to Other Lectures:
A Combination of Three Themes (& Modeling Too)

Emphasis on Neuroscience
- Brain evolution Terrence Deacon
- Anatomical determinants of speech and language David Poeppel
- Biological constraints and language development Lucie Ménard
- Brain lateralization and the emergence of language Nathalie Tzourio-Mazoyer
- Linguistic Theory and the Origin of Language Denis Bouchard

Emphasis on Animal Models
- Animal communication Stephanie White
- Primate communication Klaus Zuberbühler
- What has ape language research taught us about human language? Duane M. Rumbaugh and E. Sue Savage-Rumbaugh
- Mice, chimpanzees and the molecular basis of speech Wolfgang Enard

Taking Gestural Communication Seriously
- Origins of Human Communication Michael Tomasello
- Gestural theory Michael Corballis

Part II

The Mirror System Hypothesis
The choices made in developing the Mirror System Hypothesis

MD1: Our ancestors evolved a capability for protolanguage before they developed language

MD2: We seek to understand the shared mechanisms that support signed languages as well as spoken languages

MD3: We make crucial use of data on the brain

MD4: Biological evolution gave us mechanisms which made the eventual invention of language possible

MD5: Initially, much of protolanguage was holophrastic

MD6: Language evolution rests in part on the exaptation of adaptations that are not directly related to communication

---

Mirror Neurons

The effective observed movement

The effective executed movement

Rizzolatti, Fadiga, Gallese, and Fogassi, 1995
Key data:
- Monkey F5 (with its mirror system for grasping) is homologous to human Broca’s area.
- Imaging studies show activation for both grasping and observation of grasping in or near Broca’s area.

Language Within Our Grasp

... and this is where the story really starts

Mirror System Hypothesis:
the evolutionary basis for language parity is provided by the mirror system for grasping, rooting speech in communication based on manual gesture

From “praxis” to communication: a neural basis for a gestural origins view of the evolution of language
**Simple and Complex Imitation in MSH**

The mirror system shared by the common ancestor of monkeys, apes, and humans evolved in human ancestors as part of successively larger, more competent systems.

* An enlarged system to support *simple imitation* evolved in the common ancestor of humans and apes: acquiring on a limited basis some novel actions by extensive observation and repetition,

* *Complex imitation* evolved in the human line since the divergence from the great apes: based on the ability to observe a novel performance and see, to a first approximation, its key subgoals and the actions which appear to achieve them.

The ability to imitate praxic skills conferred selective advantage for those who can learn from the successful goal achievements of others.

**What about communication?**

* The vocal repertoire of nonhuman primates is relatively fixed

* But *simple imitation* allows apes (and, presumably, the common ancestor of apes and humans) to acquire a small but open repertoire of communicative manual gestures

* ontogenetic ritualization + social learning

---

**Imitation: From Praxis to Communication**

```
Simple imitation  →  Ape Gesture
                  ↓
Complex imitation  →  Protosign
                     ↓
                          Protolanguage
                              ↓
                                Language
```

* The vocal repertoire of nonhuman primates is relatively fixed

* But *simple imitation* allows apes (and, presumably, the common ancestor of apes and humans) to acquire a small but open repertoire of communicative manual gestures

* ontogenetic ritualization + social learning
Pantomime is Transitional from
Use of Complex Imitation for Communication to Protosign

Two key parts of the Hypothesis:

Pantomime exploited complex imitation to create an open semantic space for communication:
- The ability to create an open-ended set of complex messages exploiting the primates’ open-ended manual dexterity

This leads to discovery of the use of abstract gesture:
- As a pantomime becomes familiar to a group, it may become ritualized and thence become a symbol recognized only by members of the group, but not by a general ability to interpret
- Once a group has acquired the understanding that new symbols can provide non-iconic messages, the difficulty of separating certain meanings by pantomime encourages creation of further new signs. This yields a protosign system.

Dissociating pantomime from signed language

Pantomime and signing dissociate with left hemisphere damage.

But there is no difference between “pantomimic” and non-pantomimic signs

Note: ASL is a full human language, not a protosign system.
### Extending the Mirror System Hypothesis (Arbib, 2002, 2005 and many more since then …)

<table>
<thead>
<tr>
<th>S1) grasping</th>
<th>S2) a mirror system, matching action observation and execution for grasping</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shared with common ancestor of human and monkey</td>
</tr>
<tr>
<td>S3) a simple imitation system for grasping</td>
<td>Shared with common ancestor of human and great apes</td>
</tr>
</tbody>
</table>

#### Pre-Hominid

<table>
<thead>
<tr>
<th>S4) a complex imitation system: complex imitation combines</th>
</tr>
</thead>
<tbody>
<tr>
<td>- the ability to recognize another's performance as a set of familiar movements with</td>
</tr>
<tr>
<td>- the ability to use this recognition to repeat the performance, and</td>
</tr>
<tr>
<td>- (more generally) to recognize that another’s performance combines actions which can be imitated at least crudely be by variants of actions already in the repertoire, with increasing practice yielding increasing skill.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S5) protosign: a manual-based communication system, breaking through the fixed repertoire of primate vocalizations to yield an open repertoire</th>
</tr>
</thead>
<tbody>
<tr>
<td>S6) protospeech: resting on the “invasion” of the vocal apparatus by collaterals from the communication system based on F5/Broca's area</td>
</tr>
</tbody>
</table>

#### Cultural Evolution in Homo Sapiens

| S7) language: from action-object frames to verb-argument structures to syntax and a compositional semantics: Co-evolution of cognitive & linguistic complexity |

### Duality of Patterning

In speech

* meaningful units are composed from a smaller set of meaningless units as when discrete sounds combine to form words
* words combine to form phrases which combine to form sentences

Definition: The **phonology** of a language comprises a fixed set of meaningless units together with the “rules” whereby they can be constructed into meaningful units

Stokoe (1960) demonstrated that a sign language also has duality of patterning – meaningless handshapes, locations, and movements combine to form a large set of lexical items. This provides the basis for **sign phonology**
**The Emergence of Phonology**

Pantomime does not need phonology

- we may vary the pantomime of opening a door in many many ways
- Conventionalization of such a pantomime will capture aspects of one of the many possible performances rather than being built from constituents.

Early utterances of protospeech might

- echo the movements of a protosign; or
- come closer to the vocalization of a cat than the “meow” that invokes the phonology of English.

But as Hockett (1987) says:

- “If a vocal-auditory [or gestural-visual] system comes to have a larger and larger number of distinct meaningful elements, those elements inevitably come to be more and more similar to one another in sound [or appearance, respectively]”
- and this would provide the pressure for segmenting protowords into pieces which could then be replaced by an increasingly conventionalized system of “meaningless units”

---

**Variation in Al-Sayyid Bedouin Sign Language (ABSL) Signs “close” to pantomime**

Aronoff et al. (2008) find an unexpectedly high degree of inter-signer variation in Al-Sayyid Bedouin Sign Language

- e.g., “tree” “dog” and “banana” remain close to pantomime though the signs within a family may be similar.

suggesting that linguistic proficiency can occur without duality of patterning

- a (sign) language can occur without phonology
Part IV

A Different Source of Empirical Data

From hand to mouth: An experimental simulation of language origin

Nicolas Fay, Stephanie Lim & Sue-Lynn Teo

www.psy.uwa.edu.au/comlab/

---

**Task**

“Director” communicates 18 items to “Matcher”
- who chooses from the list of options
- they play the game 6 times

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Action</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>tired</td>
<td>fleeing</td>
<td>rock</td>
</tr>
<tr>
<td>pain</td>
<td>sleeping</td>
<td>fruit</td>
</tr>
<tr>
<td>angry</td>
<td>fighting</td>
<td>predator</td>
</tr>
<tr>
<td>happy</td>
<td>throwing</td>
<td>water</td>
</tr>
<tr>
<td>disgust</td>
<td>chasing</td>
<td>tree</td>
</tr>
<tr>
<td>danger</td>
<td>eating</td>
<td>hole</td>
</tr>
</tbody>
</table>

---

vocalization ONLY

[Matcher does not see Director]

gesture + vocalization

---

Arbib: From Praxis to Language

Montreal, June 2010
Example Results

**Gesture**
- **EMOTION:** Pain ✓
- **ACTION:** Fleeing ✓
- **OBJECT:** Fruit ✓

**Vocalization**
- **EMOTION:** Tired ✓
- **ACTION:** Fleeing [Response: Predator] ✗
- **OBJECT:** Fruit [Response: Eating] ✗

---

Results: Identification Accuracy

![Graph showing performance (%) for Emotion, Action, and Object with Vocal, Gesture, and Combined categories over six games.](chart.png)

Performance (%) vs. GAME
The Charms of Iconicity

Among “modern day humans” gesture is more effective than vocalization in establishing a communication system where none exists … even though the subjects use spoken language, not sign language.

This suggests – in concert with the data on human co-speech gestures and data on ape gesture – an important role for gesture in the origin of language though it does not rule out a role for vocalization

Gesture lends itself to iconic communication & this helps ground shared sign-meaning mappings

Part IV

A Different Approach to Modeling

Developing computational models of brain systems which correspond to posited stages of evolution

but primarily conceptual models of the evolutionary processes which link such stages
**What the Macaque Brain Tells the Human Mind**

- Develop more data on macaque brain regions which are possible homologues of human brain areas relevant to language, and add data on the connectivity of these areas in both human and macaque, to yield improved estimates of degrees of homology.
- Develop models, rooted in detailed neurophysiology and neuroanatomy, of the mirror neuron system and other brain regions involved in generation and recognition of sequential behavior in macaque.
- Extend these to models of human circuitry to be tested by, e.g., Synthetic PET & fMRI.
- And Synthetic ERPs and MEG?

![Diagram showing homology and connectivity between macaque and human brains](image)

**What’ versus 'How’ (not just ‘Where’) in Human**

- **DF:** Jeannerod et al.
  Lesion here: Inability to Preshape
  (except for objects with size “in the semantics”)

- **AT:** Goodale and Milner
  Lesion here: Inability to verbalize or pantomime size or orientation

- **Visual Cortex**
- **Inferotemporal Cortex**
- **Parietal Cortex**

![Diagram showing connectivity between different brain regions](image)
Introducing AIP and revisiting F5 in Monkey
IP = IntraParietal Sulcus

A key theme of visuomotor coordination: *parietal affordances* drive frontal motor schemas

**F5** - grasp commands in premotor cortex
Giacomo Rizzolatti

**AIP** - grasp affordances in parietal cortex
Hideo Sakata

LIP – “where” for eye movements

---

**FARS (Fagg-Arbib-Rizzolatti-Sakata) Model**
*(Fagg & Arbib, 1998 plus Rizzolatti & Luppino, 2003)*

Establishing a trajectory to yield a desired *binary* relationship between hand and affordance – with prefrontal selection of the affordance appropriate to the task

**Dorsal Stream: Affordances**

**Visual Input**

**Motor Schema**

**Premotor F5 (Canonical)**

**PFC**

- Working Memory (46)
- Instruction Stimuli (F2)

**Go Signal**

**"It's a mug"**

**clPS**

**AIP**

**Modulating Choice of Affordances**

**"It's a mug"**

**"Go Signal"**

**"I'm hungry"**

**Motor F1**

**Hand Control**

**Visual Stream: Recognition**
Modelling Mirror Systems in Action Recognition

**Main idea:** The mirror system is trained during observation of self-recognized actions.

- The representation is such that observation of another individual’s action evokes the same pattern of input.

**Issues:**

- Recognition before completion (prediction)
- Inference over partially hidden states
- Multimodal action recognition

MNS2 – Partially Hidden Grasps

Working memory and dynamic remapping of hand working memory allows hidden grasps to be recognized.

No response if object is not visible and not in working memory.
Schema Theory: A general framework for the study of action and perception which has been extended to the beginnings of a theory of neurolinguistics.
VISIONS: From Visual Scenes to Schema Assemblages

Vision itself is “inferential” and schemas carry much that is not verbalized

Segmentation
Low-Level Vision
- Competition and Cooperation at the level of local image features
- grow edges and regions to yield a
- first-pass subdivision of the image
to ground semantic analysis

Recognition
High-Level Vision
- Sky: Data driven
- Roof: Data driven
  (but with context)
- Wall: Hypothesis driven
- Schema instances compete and cooperate
to interpret different regions

Hanson and Riseman 1978

A Mirror for Words as Phonological Actions

The phonological form gains its meaning by being linked to an assemblage of perceptual and motor schemas which support both perception and the planning of action.

**Putting Words Together**

The key data for MSH place a mirror system for grasping, and thus (inferentially) for language, in Broca’s area, suggesting that the prime effect of damage to Broca’s area would be the inability to pronounce words.

However, a major factor in Broca’s aphasia is *agrammatism*, a pattern of syntactically defective speech that may range in severity from production only of one-word utterances to mildly ‘telegraphic’ speech.

Thus, the transition from

- *Actions to Compound Actions*
- and from
- *Words to Constructions*
  
provides key challenges for research.

---

**From Idioms to Construction grammar**

Consider *idiomatic expressions* like *kick the bucket*

- An idiom’s meaning cannot be (entirely) predicted on the basis of its parts. But this suggests that the meaning of each idiom must thus be stored in each speaker’s mind.

Fillmore, Kay & O’Connor (1988) suggested that the tools they used in analyzing idioms could form the basis for *construction grammar* as a new model of grammatical organization.

Each construction combines form and meaning.

Contrast this with a *generative description* [Chomsky] which seeks to explain language structure in terms of general, autonomous syntactic rules, with meaning “added later”, and any idiosyncratic properties derived from the lexicon.
Linking Schema Assemblages to Language: SemRep & Construction Grammar
Work with Jinyong Lee

<table>
<thead>
<tr>
<th>Auditory System</th>
<th>Our Emphases</th>
<th>Our Relevant Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting the Baseline: Modeling Basic Capabilities Shared with Macaques</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grasping: manual dexterity</td>
<td>Primates</td>
<td>The shared capability of primates is assumed not modeled.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FARS model: based on macaque data, shown consistent with human brain imaging</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KLGM: Model of how grasps may be acquired through reinforcement learning</td>
</tr>
<tr>
<td>a mirror system, matching action observation and execution for grasping</td>
<td>Shared with common ancestor of human and monkey – will provide the basis for parity</td>
<td>MNS, MNS2: learning models of how neurons may become mirror neurons for certain actions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACQ: learning model for action scheduling that posits new role for mirror neurons in self-actions</td>
</tr>
<tr>
<td>Infrate vocalization system</td>
<td>Not modeled. Used for data points for comparative analysis of communication across primates.</td>
<td></td>
</tr>
</tbody>
</table>
**Initiation: An evolving capacity first in the manual domain to support transfer of practical skills**

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A simple imitation system for grasping ability to recognize some goals</td>
<td>Shared with common ancestor of human and great apes</td>
<td>Initial semi-computational models based on (inverse model, forward model) pairs. Current modeling target: Data on skill learning and imitation in apes. Key concern: The model must not be powerful enough to also provide human capability.</td>
</tr>
<tr>
<td>Simple gestural systems with limited learning</td>
<td>Exploits simple imitation for limited communication. Shared with common ancestor of human and great apes</td>
<td>An conceptual model of how simple manual imitation could support the level</td>
</tr>
<tr>
<td>a complex imitation system</td>
<td>Unique to the hominid line: resting on the ability to recognize another's performance as a set of familiar movements structured to achieve subgoals. (Apes have good recognition, to some extent.)</td>
<td>Conceptual modeling should soon yield computational models which can address data from human brain imaging and from studies of apraxia.</td>
</tr>
<tr>
<td>Pedagogy</td>
<td>Humans evolve to be caregivers as well as learners</td>
<td>Conceptual model linking mirror neurons to data on caregivers helping children acquire new skills</td>
</tr>
<tr>
<td>Pantomime as creator of open-ended semantics</td>
<td>Distinguished from complex imitation</td>
<td>Conceptual model only</td>
</tr>
<tr>
<td>Protosign</td>
<td>Initial stages precede protospeech</td>
<td>Conceptual model only</td>
</tr>
<tr>
<td>Protospeech</td>
<td>Protospeech is separated from the call system, scaffolded by protosign</td>
<td>Conceptual model only</td>
</tr>
</tbody>
</table>

---

**Elaboration of protolanguage: Fractionation & Construction Formation**

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protolanguage: Initially holophrastic; semantics first, phonology later</td>
<td>Wray-Kirby model of fractionation. New emphasis: complementary formulation of constructions</td>
<td></td>
</tr>
<tr>
<td>Prosody</td>
<td>Not considered</td>
<td></td>
</tr>
<tr>
<td>Phonology and the Articulatory system evolve in response to increasing richness of protolanguage</td>
<td>Based on the &quot;semantic&quot; employment of the fractionation mechanism</td>
<td>Conceptual model linked to ABSL data</td>
</tr>
<tr>
<td>Memory Advances</td>
<td>Focus on procedural memory to support complex imitation</td>
<td></td>
</tr>
<tr>
<td>Increasing syntactic and semantic complexity to yield language as distinct from protolanguage: Cultural Evolution in Homo Sapiens</td>
<td>&quot;Modern&quot; application of Fractionation &amp; Construction Formation: Jane Hill's computational model of language acquisition showed how similar mechanisms may apply in language acquisition. Conceptual model: Role of these processes in emergence of new sign languages, NSL and ABSL</td>
<td></td>
</tr>
</tbody>
</table>
Observation of lipreading activated the left pars opercularis of the inferior frontal gyrus. Observation of lip-smacking activated a small focus in the pars opercularis bilaterally, and the observation of barking did not produce any activation in the frontal lobe. Observation of all types of mouth actions induced activation of extrastriate occipital areas.

**Hypothesis:** Actions belonging to the motor repertoire of the observer (e.g., biting and speech reading) are mapped on the observer's motor system. Actions that do not belong to this repertoire (e.g., barking) are recognized based without such mapping.

---

**Refining the Motor Theory of Speech Perception/Linking to Hickok & Poeppel**

- **Mirror neurons for native phonemes**
  - Articulators for native phonemes
  - Corollary discharge P
  - Articulation: nonwords

- **Mirror neurons for words**
  - Articulators for words
  - Corollary discharge W
  - Articulation: familiar words

- **A?**
  - Auditory Input
  - DORSAL:
    - P
    - native phoneme candidates
  - VENTRAL:
    - X
    - Top-down word bias
    - A
    - Training feedback

- **Competing HMMs**
  - Recognized word
  - Semantics