Communicating Agents in a Shared World

Natalia Komarova (IAS & Rutgers)
Partha Niyogi (Chicago)
* Two linguistic agents in a shared world.

Q: How can we quantify the rate of success in information transfer?

Q: Given the language of the first agent, what language makes the best communicator?
A population of communicating agents.

Q: Can they develop a common communication system?

Q: What are the properties of this system?
Plan

Part I.
• Formalize the notion of “language”
• Introduce the “mutual intelligibility” function for two communicating agents
• What is the optimal communicator?

Part II.
• Implications for evolution: populations of communicating agents
• Nash equilibria and evolutionarily stable strategies
Linguistic system

- A set of associations between **form** and **meaning**
- Let $S$ be the set of all possible **linguistic forms** (words, codes, signals, sentences, …)
- Let $M$ be the set of all possible **semantic objects** (meanings, messages, events, referents, …)
- Language, $L$, is a probability measure over $M \times S$
<table>
<thead>
<tr>
<th>All signals</th>
<th>cat</th>
<th>dog</th>
<th>hippo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m11</td>
<td>m12</td>
<td>m13</td>
</tr>
<tr>
<td></td>
<td>m21</td>
<td>m22</td>
<td>m23</td>
</tr>
<tr>
<td></td>
<td>m31</td>
<td>m32</td>
<td>m33</td>
</tr>
</tbody>
</table>
The sets $S$ and $M$:

- Can be finite (animal signals)
- Or infinite (human sentences)
- We assume that they are countable...
- The measure on $S \times M$ is called an association matrix
Two modes of communication

Production mode  Comprehension mode
Production mode

Given a meaning, what is the probability to use each signal?

Normalize the association matrix by column…
Given a signal, what is the probability to interpret it as each possible meaning?

Normalize the association matrix by row…
Each agent has a language, that is,
Each agent comes equipped with two matrices:
P matrix for production and
C matrix for comprehension

\[
\begin{align*}
P &= \begin{pmatrix}
1 & 0.2 & 0.01 \\
0 & 0.8 & 0.01 \\
0 & 0 & 0.98
\end{pmatrix} \\
C &= \begin{pmatrix}
1 & 0 & 0 \\
0 & 0.9 & 0.1 \\
0 & 0.01 & 0.99
\end{pmatrix}
\end{align*}
\]
Production and comprehension

• The production and comprehension modes are connected (in humans and animals).
• They arise out of highly structured systems in the brain and cannot change independently.
• We assume that they arise out of a common association matrix
A **fully-coordinated** (non-ambiguous) language:

\[ P = C \]

P and C are binary permutation matrices

<table>
<thead>
<tr>
<th>P =</th>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C =</th>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P =</th>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C =</th>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Intelligibility function

\[ L_1 : P_1, C_1 \]

\[ L_2 : P_2, C_2 \]

Probability of successful communication when agent 1 speaks:

\[
P(1 \rightarrow 2) = \sum \sigma_i \sum_{i} P(1 \text{ uses signal } j \text{ to express meaning } i) P(2 \text{ interprets signal } j \text{ as meaning } i)
\]
Intelligibility function

Agent 1

\[ L_1 : \quad P_1, C_1 \]

Agent 2

\[ L_2 : \quad P_2, C_2 \]

Probability of successful communication when agent 2 speaks:

\[
P(2 \rightarrow 1) = \sum \sigma_i \sum \text{P}(2 \text{ uses signal } j \text{ to express meaning } i) \text{P}(1 \text{ interprets signal } j \text{ as meaning } i)
\]
Intelligibility function

\[ F(L_1, L_2) = \frac{1}{2} [P(1 \rightarrow 2) + P(2 \rightarrow 1)] \]

\[ F(L_1, L_2) = \frac{1}{2} [\text{tr}(P_1 \sigma C_2^T) + \text{tr}(P_2 \sigma C_1^T)] \]
Given the language of agent 1, what language of agent 2 maximizes the mutual intelligibility?

$L_1 : P_1, C_1$

$L_2 : ?, ?$
A language is fully coordinated if it defines a one-to-one correspondence between signals and meanings (binary permutation matrices).

For fully coordinated languages, the production and comprehension matrices are the same.

The best communicator with a fully-coordinated language is the language itself.
What happens if the language is not fully coordinated?
Given P1

<table>
<thead>
<tr>
<th>Agent 1</th>
<th>Agent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
</tr>
</tbody>
</table>

P1 = C2 =

\[ P(1 \rightarrow 2) \propto \sum_{i,j} P^{(1)}_{i,j} C^{(2)}_{i,j} \]
Given $P_1$ What is the best $C_2$?

$P_1 = C_2 = \begin{pmatrix}
0.1 & 0.3 & 0.1 \\
0.2 & 0.4 & 0.6 \\
0.7 & 0.3 & 0.3 \\
\end{pmatrix}$

$C_2 = \begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
\end{pmatrix}$

\[
P(1 \rightarrow 2) \propto \sum_{i,j} P_{i,j}^{(1)} C_{i,j}^{(2)}
\]
Given C1

<table>
<thead>
<tr>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>*</td>
</tr>
<tr>
<td>0.6</td>
<td>*</td>
</tr>
<tr>
<td>0.4</td>
<td>*</td>
</tr>
</tbody>
</table>

C1 =

<table>
<thead>
<tr>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>*</td>
</tr>
<tr>
<td>0.4</td>
<td>*</td>
</tr>
<tr>
<td>0.3</td>
<td>*</td>
</tr>
</tbody>
</table>

What is the best P2?

P(2 → 1) ∝ \sum_{i,j} P_{i,j}^{(2)} C_{i,j}^{(1)}
Given $C_1$ What is the best $P_2$?

$C_1 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \\ 0.6 & 0.4 & 0 \\ 0.4 & 0.3 & 0.3 \end{bmatrix}$

$P_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$$P(2 \rightarrow 1) \propto \sum_{i,j} P_{i,j}^{(2)} C_{i,j}^{(1)}$$
Constructing the best communicator

• Given the production matrix, can construct the optimal decoder
Constructing the best communicator

• Given the production matrix, can construct the optimal decoder
• Given the comprehension matrix, can construct the optimal encoder
Constructing the best communicator

• Given the production matrix, can construct the optimal decoder
• Given the comprehension matrix, can construct the optimal encoder
• Recall: the production and comprehension mode are connected.
This puts a constraint on possible pairs of production and comprehension matrices.

If the best encoder and decoder are constructed, do they correspond to a self-consistent language?
Constructing the best communicator

Given a measure on the space $S \times M$, construct the production and comprehension matrices

$P =$

$C =$
For P construct the best decoder.

\[
P = \begin{bmatrix}
0.004 & 0.302 & 0.008 & 0.094 & 0.324 \\
0.337 & 0.038 & 0.176 & 0.332 & 0.151 \\
0.194 & 0.363 & 0.252 & 0.008 & 0.180 \\
0.322 & 0.071 & 0.286 & 0.299 & 0.212 \\
0.143 & 0.226 & 0.277 & 0.266 & 0.133 \\
\end{bmatrix}
\]

\[
00001
\]

\[
00000
\]

\[
10000
\]

\[
10000
\]

\[
00100
\]

\[
00000
\]

\[
00010
\]

\[
00001
\]

\[
00001
\]

Best P =

For C construct the best encoder.

\[
C = \begin{bmatrix}
0.006 & 0.356 & 0.011 & 0.128 & 0.500 \\
0.340 & 0.030 & 0.158 & 0.306 & 0.158 \\
0.219 & 0.318 & 0.248 & 0.008 & 0.207 \\
0.290 & 0.050 & 0.224 & 0.241 & 0.195 \\
0.152 & 0.188 & 0.259 & 0.254 & 0.145 \\
\end{bmatrix}
\]

\[
01100
\]

\[
00000
\]

\[
10000
\]

\[
10000
\]

\[
00010
\]

\[
00001
\]

\[
00001
\]

Best C =

\[
00001
\]
Can we identify a measure on the space $\mathcal{S} \times \mathcal{M}$ such that it gives rise to these production and comprehension matrices?

Best $C =$

\[
\begin{array}{ccccc}
0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
\end{array}
\]

Best $P =$

\[
\begin{array}{ccccc}
0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 \\
\end{array}
\]
Can we identify a measure on the space $S \times M$ such that it gives rise to these production and comprehension matrices?

Do they come from a self-consistent language?
Can we identify a measure on the space $S \times M$ such that it gives rise to these production and comprehension matrices?

Do they come from a self-consistent language?

In general, NO.
Constructing the best communicator

**Theorem.** Let language $L$ be self-consistent. That is, its production and comprehension matrices come from some measure on the space $S \times M$. Let $C^*$ and $P^*$ be the best encoder and decoder constructed as above.

It is always possible to find a family of measures on the space $S \times M$, parameterized by a parameter, $\mathcal{E}$, such that it gives rise to production and comprehension matrices arbitrarily close to $C^*$ and $P^*$.

These languages tend to the best communicator as $\mathcal{E} \to 0$. 
Step 1. Build an auxiliary matrix

Best C =

Best P =
Step 2. Connect vertices that share a row (column)

Best C =

<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Best P =

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Step 3. Assign direction

Best C =

Best P =
Step 4. Assign powers of \( \epsilon \) so that the entries increase in the direction of the arrows.
Constructing the best communicator

Given language

\[
\frac{1}{1245} \times \begin{bmatrix}
1 & 64 & 2 & 23 & 90 \\
92 & 8 & 42 & 81 & 42 \\
53 & 77 & 60 & 2 & 50 \\
88 & 15 & 68 & 73 & 59 \\
39 & 48 & 66 & 65 & 37
\end{bmatrix}
\]

The constructed family of languages which approaches the best communicator as

\[
\varepsilon \to 0
\]
Constructing the best communicator

- For fully coordinated languages, the language is the best communicator with itself
- For general languages, the best communicator is different from the language itself
- In general, the best communicator cannot be constructed exactly, but it can be approached arbitrarily close
- This theory can be extended to noisy communication, infinite languages etc.
Evolutionary dynamics of communicating agents.
Evolutionary dynamics

- Consider a population of communicating agents
- Each agent has a language (an association matrix matrices $P$ and $C$)
- Apply a “genetic algorithm”: agents proliferate and die
- “Children” inherit/learn the language of their “parent”
- Keep the score for each agent on how successful it communicates with others…
Evolutionary dynamics
Evolutionary dynamics

• The score is computed by means of mutual intelligibility:

\[ F_i = \sum_j F(L_i, L_j) \]

• Agents with a higher score have a higher chance of reproduction (lower chance of death)
Evolutionary dynamics

Hurford 1989  Oliphant 1997
Steels 1996  Kirby 1999
Steels & Vogt 1997  Briscoe 2000
Steels & Kaplan 1998  Smith 2002
Oliphant & Batali 1996  Smith 2003

Numerical studies of evolutionary dynamics of communicating agents
A typical outcome…

A common, highly coordinated communication system

Random initial condition
Evolutionary dynamics

• We attempt an analytical description
Evolutionary dynamics

- We attempt an analytical description
- Use properties of the “score” function $F(L_i, L_j)$
Evolutionary dynamics

- We attempt an analytical description
- Use properties of the “score” function $F(L_i, L_j)$
- Identify and classify important states of the system
Navigating in the fitness landscape

\[ F(L_1, L_2) \]
Definitions of Game Theory

- Language $L$ is strict Nash equilibrium if for all $L' \neq L$, $F(L,L) > F(L,L')$
Definitions of Game Theory

- Language $L$ is *Nash equilibrium* if for all $L' \neq L$, $F(L, L) \geq F(L, L')$
Definitions of Game Theory

- Language $L$ is an *evolutionary stable state* (ESS) if $L$ is Nash and for every $L'$ with $F(L, L) = F(L, L')$ we have $F(L, L) > F(L', L')$
Definitions of Game Theory

- Language $L$ is a weak ESS if $L$ is Nash and for every $L'$ with $F(L, L) = F(L, L')$ we have $F(L, L) \geq F(L', L')$
Definitions of game theory

- Strict Nash
- ESS
- Nash
- Weak ESS

Diagram showing the relationships between these concepts.
Definitions of game theory

Strict Nash \rightarrow Nash

ESS \rightarrow Weak ESS

This is what we observe in computer simulations
Evolutionarily Stable States

- A language is an ESS if it is an extended permutation matrix

\[
\begin{array}{cccccc}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]
Evolutionarily stable states

- ESS are important in the dynamics
- However, much more common outcomes of evolutionary dynamics are weak ESS. The system will converge to a weak ESS and then slowly drift to the nearest ESS without change in communicability function.
Evolutionarily stable states

- ESS are important in the dynamics.
- However, much more common outcomes of evolutionary dynamics are weak ESS. The system will converge to a weak ESS and then slowly drift to the nearest ESS without change in communicability function.
- Weak ESS are harder to characterize…
A language has synonyms (homonyms) if some column (row) of the association matrix has more than one positive elements.

### Synonyms
- Bank – bank
- Spring – spring
- Fall – to fall
- Beautiful – fair
- Medicine – drug
- To phone – to call

### Homonyms
- Bank
- Spring
- Fall
- Beautiful
- Medicine
- To phone

<table>
<thead>
<tr>
<th>Meanings</th>
<th>Signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Isolated Synonyms

- A language has *isolated synonyms* if the corresponding rows of the association matrix have no other positive elements.
Weak ESS

**Theorem.** The language \( L \) is a weak ESS if and only if the only kind of synonyms it has are isolated synonyms. It cannot have homonyms.

\[
\frac{1}{a + b + c + d + e + f}
\]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>a</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>b</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>f</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>e</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>c</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>d</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Definitions of game theory

Strict Nash \rightarrow\text{Nash} \rightarrow\text{Weak ESS} \rightarrow\text{ESS}

Fully coordinated languages

Languages that can have isolated synonyms but no homonyms

This is exactly what is observed in computer simulations
Paradox

• Studies of natural languages suggest that full synonyms are extremely rare, and homonyms are quite common

• Our results (together with many numerical simulations) indicate that synonyms are evolutionarily stable, and homonyms are not
Paradox

• Studies of natural languages suggest that full synonyms are extremely rare, and homonyms are quite common
• Our results (together with many numerical simulations) indicate that synonyms are evolutionarily stable, and homonyms are not
• What is wrong?
Paradox: resolution

- The columns of our matrices refer to meanings; the rows refer to signals
- Let us add context in the picture
- Intelligibility should be defined *per utterance*, rather than *per word*
- Now the results can be re-evaluated
Paradox: resolution

- At the new level of description, the mathematical model correctly describes the following observation:

  There are no true homonyms that remain homonyms in the presence of contextual clues. Context-dependent synonyms are rather common.
Example: the level of words

- **Fall** (autumn) and **Fall** (down) are homonyms on the level of words.
- **Beautiful** and **Fair** are not complete synonyms on the level of words (fair price?)
Example: the level of utterances

- **Fall** (autumn) and **Fall** (down) are homonyms on the level of words. On the level of utterances they are not!
- **Beautiful** and **Fair** are not complete synonyms on the level of words. They are synonymous on the level of utterances (beautiful lady = fair lady)
Context

• The presence of context destroys homonymy and creates synonymy
• This is exactly what is observed in the model.
• Synonyms are stable because they do not reduce coherence. Homonyms make the language ambiguous and thus reduce coherence. They are unstable.
Conclusions

• Given the language L, there can be found a self-consistent language, L’, which approximates the best communicator arbitrarily well.

• The best communicator is not equal to the language itself, unless L is fully coordinated.
Conclusions

- In a population of learning agents, the mutual intelligibility will increase
- The system will reach a weak ESS state characterized by a common language with no homonyms. This language may contain isolated synonyms
- Slow random drift may then bring the system to a fully coordinated language without change in intelligibility