Models of Language Evolution

Agent-Based Models

Michael Franke

Goals for today

1. look at 3 case studies of agent-based models for meaning evolution
   1. cellular automata
   2. naming game
   3. category game
2. see what's good and bad about each of these

Conway's Game of Life

- grid of cells
- each cell x:
  - has 8 neighbors
  - is alive or dead at any given time
- simultaneously update all cells:
  1. any live cell stays alive iff it has exactly 2 or exactly 3 live neighbors
  2. any dead cell becomes alive iff it has exactly 3 neighbors

http://web.mit.edu/jb16/www/6170/gameoflife/gol.html

Meaning Evolution in Cellular Automata

- finite grid of agents, with 8 neighbors each
- there are randomly walking *predators* and *food sources*
- each round each agent has a choice whether or not to do any of the following (coded as a bitvector $\vec{x} \in \{0, 1\}^4$):
  (i) open mouth
  (ii) hide
  (iii) emit sound 1
  (iv) emit sound 2
- each action incurs some (non-positive) cost $\vec{c} \in \mathbb{R}^4$
- agents receive positive payoffs f for opening the mouth when in a cloud of food
- agents receive negative payoffs b when *not* hiding in a cloud of predators

(Grim et al., 2004)

Meaning Evolution in Cellular Automata

- each agent i can condition her choice on whether or not any of the following happend in the previous round (coded as a bitvector $\vec{y}$):
  - (i) i's been fed
  - (ii) i's been hurt
  - (iii) i heard sound 1
  - (iv) i heard sound 2
- agents have 265 strategies in total (all functions from $\vec{y}$ to $\vec{x}$)
- each agent i gets a reward for each round t depending on his actions $\vec{x}$:

$$\mathrm{R}(i,t) = b + f + \vec{x} \cdot \vec{c}$$

- we consider the accumulated rewards (ARs) between round t and t':

$$\mathrm{AR}(i,t,t') = \sum_{t \leq \tau \leq t'} \mathrm{R}(i,\tau)$$

- every 100 rounds each agent compares the ARs of her neighbors and adopts the *strategy* of the most successful neighbor ("imitate-the-best dynamics")

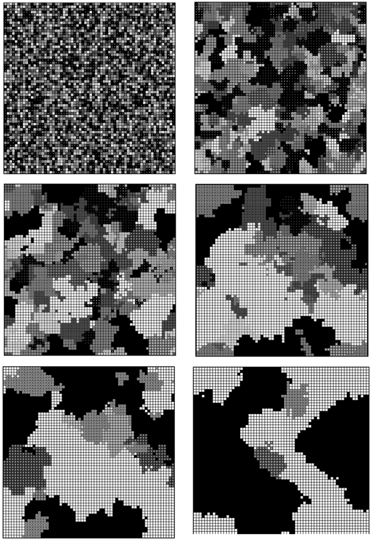
→ What's going to happen? ←

Result of Simulation

- starting from a random population
- regions of "perfect communicators" emerge:
  - perfect communicators use one signal for food, one for predators

Reflection

- is this a good / plausible model of meaning evolution?
- anything we would like to know further about the model?



(Minimal) Naming Game

- population of n agents looking for word for one object/meaning
- at each point in time each agent has a vocabulary of words
- initially all agents have one random word
- asynchronous update with actual play

(Loreto et al., 2010)

(Minimal) Naming Game: Play & Update Rule

let

V_S: vocabulary of speaker

V_H: vocabulary of hearer

select $w \in V_S$ uniformly at random

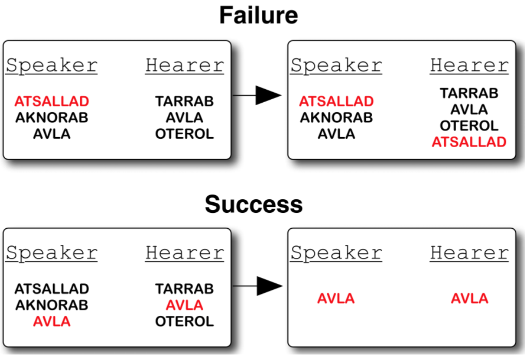
if $w \in V_H$: (play is a success)

$V_S \leftarrow \{w\}$

$V_H \leftarrow \{w\}$

otherwise: (play is a failure)

$V_H \leftarrow V_H \cup \{w\}$



(Minimal) Naming Game: Results

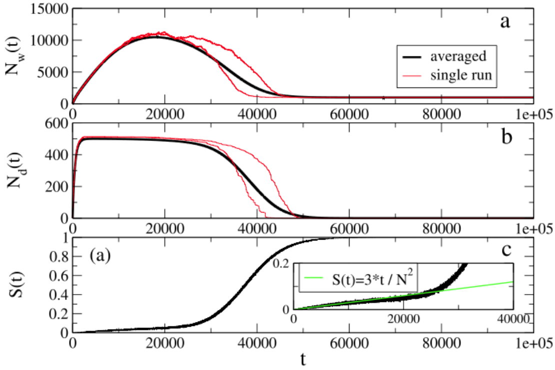


Fig. 2 Naming Game. (**a**) Total number of words present in the system, $N_w(t)$; (**b**) Number of different words, $N_d(t)$; (**c**) Success rate $S(t)$, i.e., probability of observing a successful interaction at time t. The *inset* shows the linear behavior of $S(t)$ at small times. The system reaches the final absorbing state, described by $N_w(t) = N$, $N_d(t) = 1$ and $S(t) = 1$, in which a global agreement has been reached. From Baronchelli et al. (2006b)

AB Model

(minimal) naming game with only two possible words A & B

rate of change can be calculated:

$$\begin{aligned}
\dot{n}_A &= -n_A n_B + n_{AB}^2 + n_A n_{AB} \\
\dot{n}_B &= -n_A n_B + n_{AB}^2 + n_B n_{AB} \\
\dot{n}_{AB} &= +2n_A n_B - 2n_{AB}^2 - (n_A + n_B) n_{AB}
\end{aligned}$$

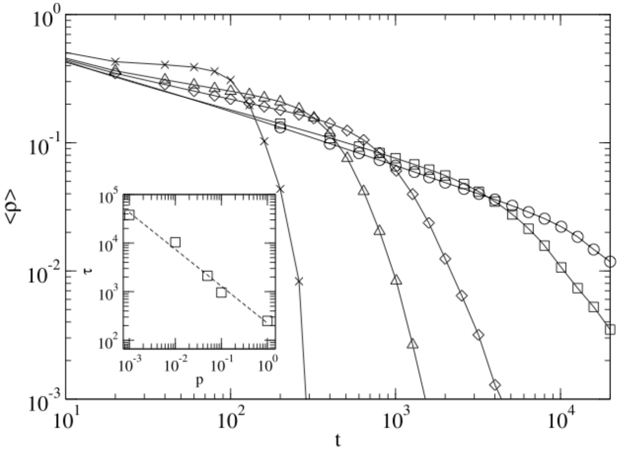
fixed point solutions:

1. $n_A = 1$
2. $n_B = 1$
3. $n_A = n_B = 2n_{AB}$

n_X is the proportion of agents with vocabulary X

AB Model on SW-Networks

Fig. 3 AB-model. Time evolution of the average density $\langle \rho \rangle$ of bilingual individuals in small-world networks for different values of the rewiring parameter p. From *left* to *right*: $p = 1.0$, 0.1, 0.05, 0.01, 0.0. The *inset* shows the dependence of the characteristic lifetime τ on the rewiring parameter p. The *dashed line* corresponds to a power law fit $\tau \sim p^{-0.76}$. From Castelló et al. (2006)



Category Game

co-evolution of perceptual & linguistic categories

(for a continuous 1-dim perceptual space $[0, 1]$)

- population of n agents
- each agent i has:
  - a set of **categories** C_i (think: partition of $[0, 1]$)
  - for each $c_j \in C_i$ a **vocabulary** V_{ij} (set of words for c_j)
  - for some $c_j \in C_i$ a **designated word** d_{ij}
    - last successful word, if exists
    - else the last one introduced, if exists
    - else none
- initially:
  - all $C_i = \{[0; 1)\}$
  - all $V_{ij} = \emptyset$
- asynchronous update with actual play (heterogeneous population)

Category Game: Play & Update Rule

Preliminaries

- $C_i \subseteq [0;1]$ (represent intervals by upper-bound)
- $V_i : C_i \to \mathcal{P}(\mathbb{N})$ (integers as words)
- $C_i(a) = \min(\{z \in C_i \mid z > a\})$ (category of $a \in [0;1)$)

Category Game: Play (1)

$i, j \leftarrow$ random speaker and hearer

$a, b \leftarrow$ random pair of perceptions from $[0; 1)$ s.t. $|a - b| > d_{\text{min}}$

a is the "topic" the speaker wants to talk about

sender distinguishes stimuli if necessary

if $C_i(a) = C_i(b)$: (*i*'s categories don't distinguish *a* and *b*)

add $a+b/2$ to C_i (introduce new category boundary)

add $\langle a+b/2, V_i(C_i(\max(a, b)))\rangle$ to V_i

(new category inherits old vocabulary)

$w_1, w_2 \leftarrow$ random new words

add w_1 to $V_i(a+b/2)$

add w_2 to $V_i(C_i(\max(a, b)))$ (add new random words)

$D_i(C_i(\frac{a+b}{2})) \leftarrow w_1$

$D_i(C_i(\max(a + b))) \leftarrow w_2$ (new words are distinguished)

Category Game: Play (2)

speaker chooses word w^* to send

if $D_i(C_i(a))$ defined: (if i has a distinguished word)

$\quad w^* \leftarrow D_i(C_i(a))$ (choose distinguished word)

else:

$\quad w^* \leftarrow$ uniform random from $V_i(C_i(a))$ (choose random word)

hearer collects possible interpretations

$I \leftarrow \{x \in \{a, b\} \mid w^* \in V_j(C_j(x))\}$

hearer guesses intended referent

if $I \neq \emptyset$:

$\quad o^* \leftarrow$ uniform random from I

determine success or failure

if $I = \emptyset$ or $o^* \neq a$: failure

else: success

Category Game: Update Rule

hearer distinguishes objects if necessary

if $C_j(a) = C_j(b)$: (j's categories don't distinguish a and b)

add ${}^{a+b}/_2$ to C_j (introduce new category boundary)

add $\left\langle {}^{a+b}/_2, V_j(C_j(\max(a,b))) \right\rangle$ to V_j

(new category inherits old vocabulary)

updating agents' vocabularies

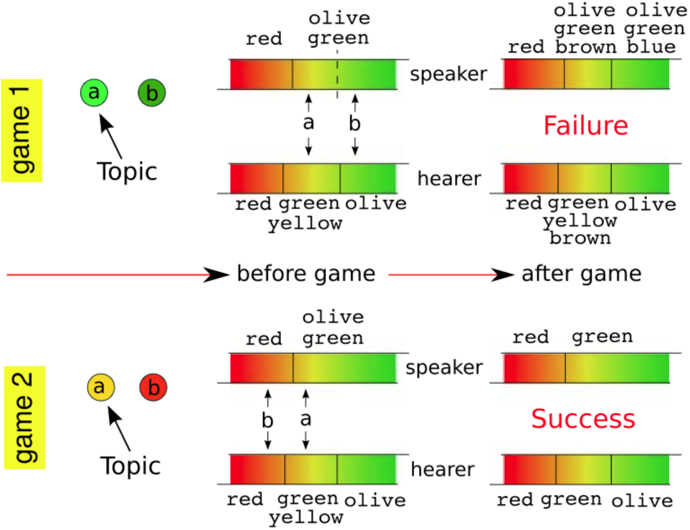
if success: (keep only w^*)

$$V_i(C_i(a)) \leftarrow \{w^*\} \qquad D_i(C_i(a)) \leftarrow w^*$$
$$V_j(C_j(a)) \leftarrow \{w^*\} \qquad D_j(C_j(a)) \leftarrow w^*$$

else:

add w^* to $V_j(C_j(a))$

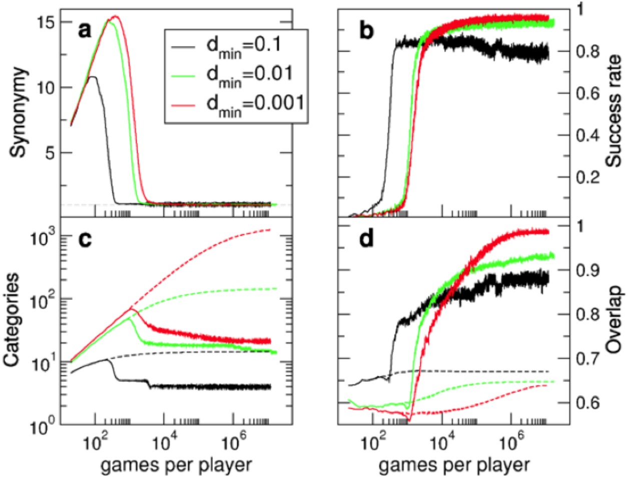
Category Game: Example



(Loreto et al., 2010)

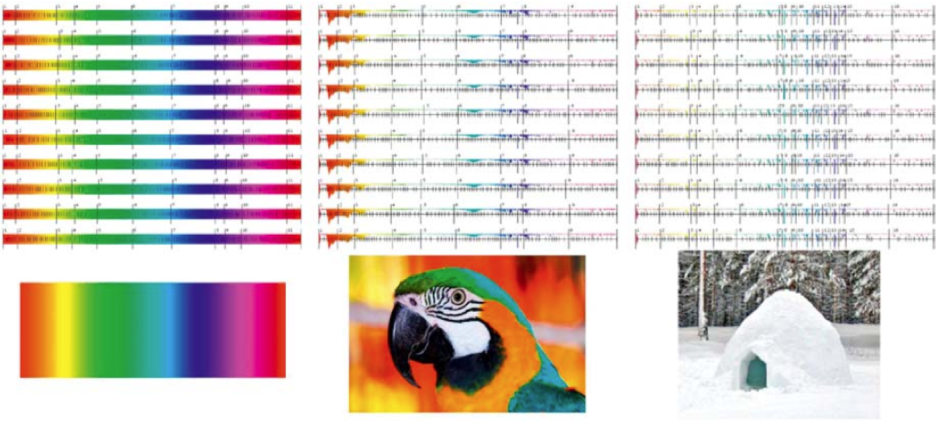
Category Game: Results

Fig. 5 Simulations results with $N = 100$ and different values of d_{min}: (**a**) Synonymy, i.e. average number of words per category; (**b**) Success rate measured as the fraction of successful games in a sliding time windows games long; (**c**) Average number of perceptual (*dashed lines*) and linguistic (*solid lines*) categories per individual; (**d**) Averaged overlap, i.e., alignment among players, for perceptual (*dashed curves*) and linguistic (*solid curves*) categories



(Loreto et al., 2010)

Category Game: Results



(Loreto et al., 2010)

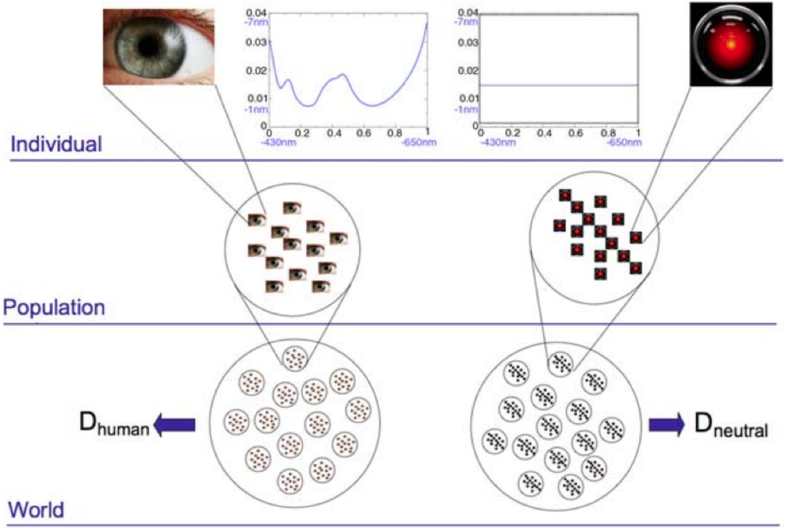
Numerical World Color Survey

- compare two worlds: one with a uniform & one with a variable d_{min}
  - variable d_{min} implements human JND for hue
  - uniform d_{min} is set to .0143, the average of human JND
- run 50 populations (50 agents each) in each world & look at resulting languages
- compare simulation data against (subset of) data from **world color survey**
  - 110 languages (without writing systems; small-scale, non-industrialized societies)
  - basic color term for each of 330 color chips for each language
  - ca. 24 speakers per language
- dispersion as a measure of common clustering (Kay and Regier, 2003):

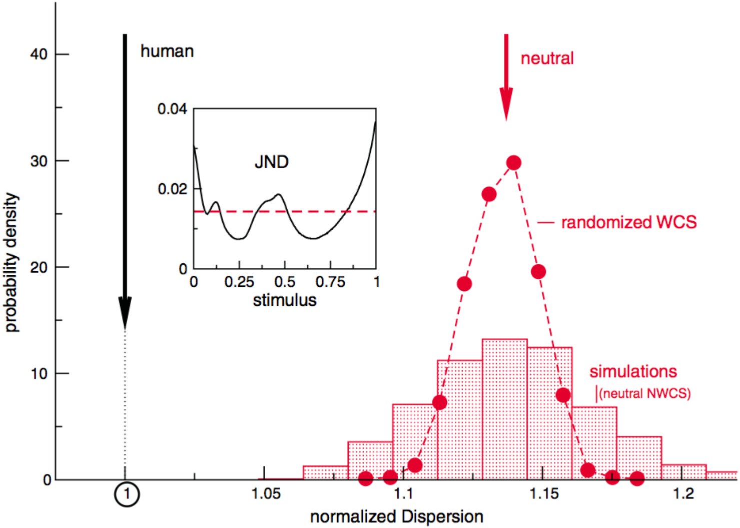
$$D = \sum_{L,L'} \sum_{c \in L} \min_{c^* \in L} \text{distance}(c, c^*)$$

(Baronchelli et al., 2010)

NWCS: Set-Up


0.04
-7nm
0.03
0.02
0.01
-1nm
0
0 0.2 0.4 0.6 0.8 1
-430nm -650nm
0.04
-7nm
0.03
0.02
0.01
-1nm
0
0 0.2 0.4 0.6 0.8 1
-430nm -650nm
Individual
Population
World
D_{human}
$D_{neutral}$

NWCS: Results



Reading for Next Class

Michael Franke & Elliott Wagner (2014). "Game Theory and the Evolution of Meaning"
Language and Linguistics Compass 8/9, 359–372

References

Baronchelli, Andrea et al. (2010). "Modeling the Emergence of Universality in Color Naming Patterns". In: *PNAS* 107.6, pp. 2403–2407.

Grim, Patrick et al. (2004). "Making Meaning Happen". In: *Journal for Experimental and Theoretical Artificial Intelligence* 16, pp. 209–244.

Kay, Paul and Terry Regier (2003). "Resolving the question of color naming universals". In: *PNAS* 100.15, pp. 9085–9089.

Loreto, Vittorio et al. (2010). "Mathematical Modeling of Language Games". In: *Evolution of Communication and Language in Embodied Agents*. Ed. by Stefano Nolfi and Marco Mirolli. Springer-Verlag. Chap. 15, pp. 263–281.