The Evolution of Linguistic Structure

L&G Project 2017

Thomas Brochhagen & Robert van Rooij

ILLC, Amsterdam, 2017.01.18

Yesterday

- Conventions as (more or less) stable object of analysis
- Synchronic perspective

Today

- Language as complex adaptive system
- Vertical and horizontal transmission
- Linguistic structure shaped by interaction, biases, environment

Linguistic structure is shaped by cultural transmission

- Learning & transmission perturbations
- Expressivity

Biological and social aspects play an important role as well

· M. A. Nowak and D. C. Krakauer. The evolution of language.
***Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999**

· Anton Benz, Gerhard Jäger, and Robert van Rooij. *An Introduction to Game Theory for Linguists*, pages 1–82.
Palgrave, 2005

· Luc Steels. Modeling the cultural evolution of language.
***Physics of Life Reviews*, 8(4):339–356, 2011**

· Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure.
***Cognition*, 141:87–102, 2015**

Motivations

- How do languages come to have the structure/features they have?
- What properties do we expect in a communication system?
  Generalization across linguistic systems & universals (given...)

For (adaptive) artificial languages

- Adopt robust features of
  natural language / biological signaling

Two complementary perspectives

- Individual level development

  solution concepts, rationalization, reinforcement learning, deep(er) learning, ...

- Population level development

  solution concepts, ESS, replicator dynamics, mutator dynamics, ...

Compositionality

It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a terrestrial being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thoughts corresponding to parts of a sentence, so that the structure of the sentence serves as the image of the structure of the thought.

Frege (1923)

Compositionality

"The meaning of a complex expression is a function of its parts and the way they are combined"

$$[\![\text{red house}]\!] = f_\alpha([\![\text{red}]\!], [\![\text{house}]\!])$$

$$[\![\text{book shelf}]\!] = f_\beta([\![\text{book}]\!], [\![\text{shelf}]\!]) \neq f_\beta([\![\text{shelf}]\!], [\![\text{book}]\!])$$

Uniquely and universally human?

- What

  Systematic association between simplex elements and the complex elements they are constituents of;

- Why

  Productivity

- How

  Players need to be biased towards (these) structural regularities

Why compositionality?

To be able to parse and generate new and unwitnessed expressions in a reasonable & practical way.

- Compression
  - Reduction of grammar & vocabulary complexity
  - Ease of transmission within and across generations (transmission bottleneck)
- Expression
  - Productivity

Note

These advantages need not obtain even if the system is compositional. Multiple factors involved: users need to be sensitive to regularities, no vacuous systematicity, etc.

Adaptive learning dynamics

How can (structured linguistic) conventions arise through interaction

- Without appeal to salience (of a strategy pairing)
- ... or iconicity
- ... or hyperrationality (or extra-linguistic agreement)
- but with minimal rationality assumptions

· David Catteeuw and Bernard Manderick. The limits and robustness of reinforcement learning in Lewis signaling games.
***Connection Science*, 26(2):161–177, 2014**

· Matthew Spike, Kevin Stadler, Simon Kirby, and Kenny Smith. Minimal requirements for the emergence of learned signaling.
***Cognitive Science*, pages 1–36, 2016**

Roth-Erev Reinforcement learning

"If an action was successful in the past given a state of affairs, perform it again if in the state of affairs"

· Richard J Herrnstein. On the law of effect.
***Journal of the experimental analysis of behavior*, 13(2):243–266, 1970**

· Justin Bruner, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger. David Lewis in the lab: experimental results on the emergence of meaning.
***Synthese*, 2014**

· Alvin E. Roth and Ido Erev. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term.
***Games and Economic Behavior*, 8(1):164–212, 1995**

· Ido Erev and Alvin E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria.
***American economic review*, 88(4):848–881, 1998**

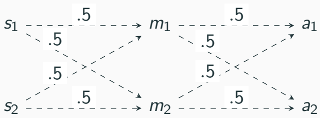
Reinforcement learning

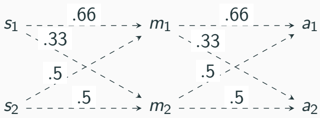
- Accumulated rewards for state-action pairs; $ar(p, q)$;
- $p(q|p) = \frac{ar(p,q)}{\sum_{q' \in Q} ar(p,q')}$;
- Discount factor λ;
- Initial value $ar_0(p, q) = n$
  (possibility to plug in prior)

If σ was in state s_i, then:

$$\uparrow_\sigma (ar(s_i, m_j)_{t+1}) = \begin{cases} \lambda ar(s_i, m_j)_t + u(s_i, m_k, a_l) & \text{if } j = k \\ \lambda ar(s_i, m_j)_t & \text{otherwise} \end{cases}$$

In 2-2-2-SGs





Some results for basic RL in SGs

Convergence to signaling system if

- $ar_0(s_i, s_j) = 1$ and $\lambda = 1$ and two equiprobable states
- learning strategy is win-stay/lose-randomize or $0 < \lambda \ll 1$ (faster than former)

Number of expected iterations to converge to a signaling system under win-stay/lose-randomize distributed according to probability of learning a new type-signal-act path p_l, N is the number of types and l the number of learned paths:

$$p_l = \frac{N-l}{N}\frac{N-l}{N}\frac{1}{N} = \frac{(N-l)^2}{N^3}$$

$$ExpctIter_N = \sum_{l=0}^{N-1} \frac{1}{p_l}$$

Some results for humans playing SGs

1. Subjects in unbiased 2-2-2 SGs approximate signaling systems ($\approx$ theoretical prediction)

2. Subjects in biased 2-2-2 or 3-3-3 SGs as above or pooling equilibrium
   ($\approx$ theoretical prediction, albeit more noise and a lot less pooling)

· Justin Bruner, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger. David Lewis in the lab: experimental results on the emergence of meaning.
***Synthese*, 2014**

Other learning algorithms and extensions

- Q-learning (combined with e.g. ϵ-greedy or softmax selection)
- Learning automata
- Declarative memory (boost/decay activation patterns)
- Lateral inhibition
- Spill-over reinforcement learning
- Noisy signaling
- Costly signaling
- ...

· Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*.
MIT Press, Cambridge, MA, USA, 1998

· David Reitter and Christian Lebiere. Towards explaining the evolution of domain languages with cognitive simulation.
***Cognitive Systems Research*, 2009**

· David Catteeuw and Bernard Manderick. The limits and robustness of reinforcement learning in Lewis signaling games.
***Connection Science*, 26(2):161–177, 2014**

Recent approaches

- Syntactic signaling games
- Signaling games with spill-over learning (& generalization)
- Signaling games with function words

Syntactic games

- Two or more independent senders
- One receiver who knows who sent what
- $|S| = |A| > |M|$

Here: $|S| = |M|^{\#\text{senders}}$

· Jeffrey A. Barrett. The evolution of coding in signaling games.
***Theory and Decision*, 67(2):223–237, 2009**

Syntactic games: some results

Signal success rate interval	16-state/16-term/ 1-sender proportion of runs	16-state/4-term/ 2-sender proportion of runs	16-state/2-term/ 4-sender proportion of runs
[0.0, 0.75)	0.003	0.014	0.017
[0.75, 0.80)	0.067	0.050	0.048
[0.80, 0.85)	0.110	0.131	0.173
[0.85, 0.90)	0.379	0.416	0.439
[0.90, 0.95)	0.369	0.335	0.290
[0.95, 1.0]	0.072	0.054	0.033
Mean signal success rate	**0.8829**	**0.8781**	**0.8704**

- Simple adaptive dynamics, but
- ... only ‘compositional’ structure for learned combinations
- ... no relation between parts and whole

Compositional structure in SGs I

Desired: Constituent-preserving f and g

$$f : S \to M$$
$$g : M \to A$$

Learning algorithm:

"If x and y are similar, then you should treat them in a similar fashion"

or

"If uncertain whether x and y are the same, then treat them in a similar fashion"

or

"Use similar actions for similar states"

Spill-over RL

Reinforce not only selected actions in a given state, but also similar state-action pairs in proportion to their similarity

$$\uparrow_\sigma (ar(s,m)_{t+1}) = \lambda ar(s,s)_t + U(s',m',a') \cdot (sim_S(s',s) \cdot sim_M(m',m))^\pi$$

$$\uparrow_\rho (ar(m,a)_{t+1}) = \lambda ar(m,a)_t + U(s',m',a') \cdot (sim_M(m',m) \cdot sim_A(a',a))^\pi$$

Makes constituent-based similarity more likely to conventionalize

E.g., $\langle s_0, m_a, a_0 \rangle$ increases a player's propensity for m_{ab} in s_{01}

Compositional structure in SGs II

Desired: Generalization over structurally similar elements

Learning algorithm:

"If $R(x, y)$ is comparable to $R(z, w)$, then what you learn about y with respect to x can be transfered to z with respect to w."

or

"Use comparable actions for comparable states"

(Structural) generalization

- $Sim_p = \{p' | 0 < sim_P(p, p') \leq 1\}$
- p and q have a similar structure iff $\forall p' \in Sim_p . \exists q' \in Sim_q . |d_X(p, p') - d_X(q, q')| < \epsilon$ and vice-versa
- Two state-action pairs have the same pattern, $P(p, q) = P(i, j)$, iff they have a similar pointwise structure

If s' was selected:

$$\uparrow_\sigma (ar(s, m)) = \begin{cases} \lambda ar(s, m) + U(s', m', a) \cdot \_^{\pi} & \text{if } (sim_S(s, s') \cdot sim_M(m, m')) > 0 \\ \lambda ar(s, m) + U(s', m', a) \cdot \gamma & \text{if } (sim_S(s, s') \cdot sim_M(m, m')) = 0 \\ & \text{and } P(s, m) = P(s', m') \\ \lambda ar(s, m) & \text{otherwise} \end{cases}$$

Summary of results I

For small π and γ:

- Spill and Spill+Gen reach more compositional signaling conventions than plain Roth-Erev RL
- Spill+Gen reach more than Spill-signalers;

Summary of results II

Generally:

- Iterations needed with no prior conventions:
  plain Roth-Erev RL $<$ Spill $<$ Spill+Gen
- Risk of pooling with no prior conventions:
  Spill+Gen $>$ Spill $>$ plain Roth-Erev RL
- Iterations needed with prior conventions:
  Spill+gen $<$ Spill $<$ plain Roth-Erev RL
- Risk of pooling with prior conventions:
  plain Roth-Erev RL $>$ Spill $>$ Spill+Gen

· Michael Franke. The evolution of compositionality and proto-syntax in signaling games. ***Journal of Logic, Language and Information*, 2015**

· Cailin O'Connor. Evolving to generalize: Trading precision for speed. ***British Journal for the Philosophy of Science*, forthcoming**

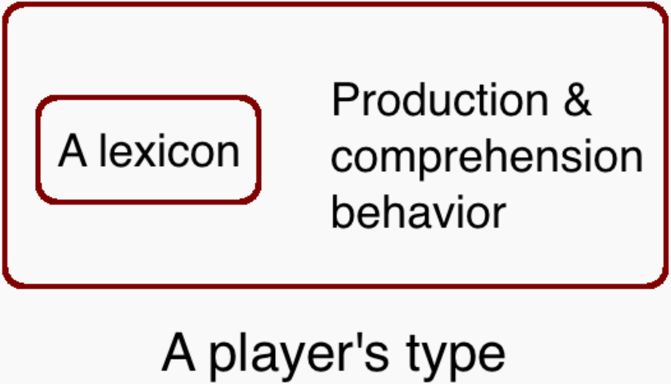
· Thomas Brochhagen. Minimal requirements for productive compositional signaling. **In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 2015**

- Little is needed to bootstrap compositional structure.
  - "Use similar actions for similar states"
  - "Use comparable actions for comparable states"
- This requires players to be sensible to these (structural) similarities in the first place
- Time to structure trade-off in learning between types of signalers fits well with comparable studies on generalization and claims on the advantages of compositionality

Population level dynamics

- Abstract away from individual-level interactions
- Population of strategies
- Either static (e.g. ESS) or dynamic (e.g. RMD)

· S. M. Huttegger and K. J. S. Zollman. Methodology in biological game theory.
***The British Journal for the Philosophy of Science*, 64(3):637–658, 2013**


A lexicon
Production &
comprehension
behavior
A player's type

RMD as cultural transmission

Two competing pressures: Expressivity & learnability

1. Imitation

   ... as replicator dynamics; $\dot{x}$

2. Transmission fidelity

   ... as mutator dynamics; Q

$\Bigg\}$ Replicator-mutator dynamics

$$\hat{x} = \dot{x} \cdot Q$$

· Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with bayesian agents. ***Cognitive Science*, 31(3):441–480, 2007**

· M. A. Nowak and D. C. Krakauer. The evolution of language. ***Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999**

Functional pressure (replicator dynamics); $\dot{x}_i = \frac{x_i f_i}{\Phi}$

- Population of types x

  x_i is the proportion of t_i in x

- Fitness of type i

  $f_i = \sum_j x_j U(x_i, x_j)$

- Average fitness in the population

  $\Phi = \sum_i x_i f_i$

· Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. ***Mathematical Bioscience*, 40(1–2):145–156, 1978**

Transmission matrix Q

- Noisy perception of data
- (Iterated) Bayesian learning
- Systematic transmission noise
- ...

· Martin A. Nowak, Natalia L. Komarova, and Partha Niyog. Evolution of universal grammar. ***Science*, 291(5501):114–118, jan 2001**

· Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with bayesian agents. ***Cognitive Science*, 31(3):441–480, 2007**

Iterated learning

- Parent produces data d based on
  - her language L
  - production and comprehension behavior
- Child infers L' from $\mathcal{L}$ using D based on
  - learning mechanism
  - $P(d|L')$ and $P(L')$
- Child turns parent and produces data d to be inferred by next generation

$\Rightarrow$ Transmission bottleneck amplifies prior effect
modulo learning mechanism

· Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with bayesian agents. ***Cognitive Science*, 31(3):441–480, 2007**

· Monica Tamariz and Simon Kirby. The cultural evolution of language. ***Current Opinion in Psychology*, 8:37–43, 2016**

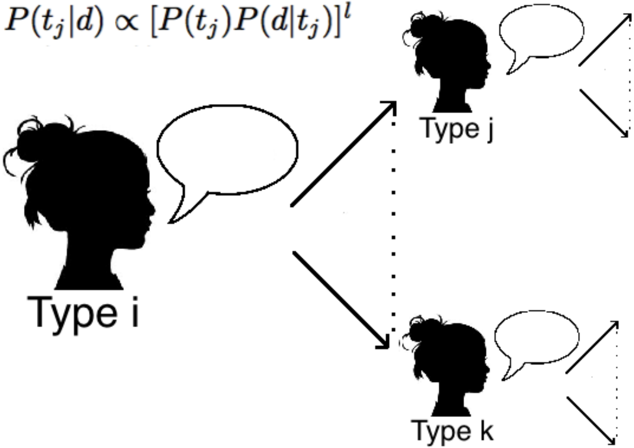
Learning with Bayes

- Data $d \in D$
- Hypotheses H
- $P(h|d) = \frac{P(d|h)P(h)}{P(d)}$ Posterior sampling
- $P(h|d) \propto [P(d|h)P(h)]^l$ $l > 1$ MAP-like
- $P(d) = \sum_{h \in H} P(d|h)P(h)$

Iterated learning as mutator dynamics

$$Q_{ij} \propto \sum_d P(d|t_i)\ P(t_j|d)$$

$$P(t_j|d) \propto [P(t_j)P(d|t_j)]^l$$



(Discrete) replicator-mutator dynamics

$$x_i' = \sum_j Q_{ji} \frac{x_j f_j}{\Phi} = (M(RD(x)))_i$$

$$(RD(x))_i = \frac{x_i f_i}{\Phi}$$

$$(M(x))_i = (x \cdot Q)_i = (\sum_j x_j Q_{ji})_i$$

Compositionality & IL: A simple setup

- $S = \{00, 01, 10, 11\}$
- $M = \{aa, ab, ba, bb\}$
- $|\mathcal{L}| = 4^4 = 256$ languages (four compositional)
- $P^*(s) = \frac{1}{|S|}$

· Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with bayesian agents. ***Cognitive Science*, 31(3):441–480, 2007**

Compositionality & IL: A simple setup

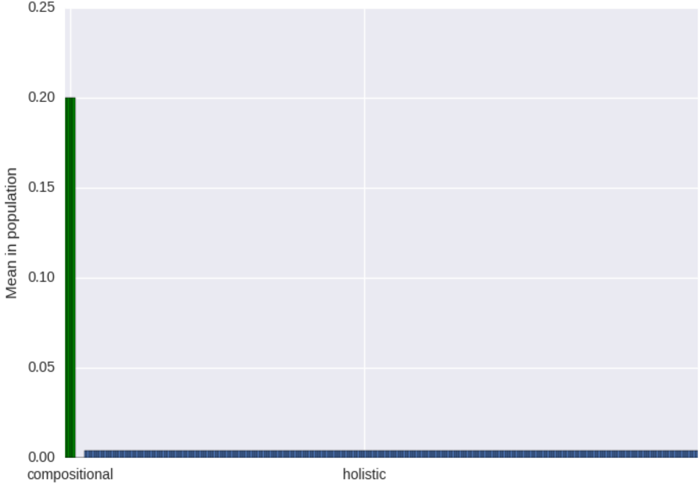
- Hierarchical $P \in \Delta(\mathcal{L})$

$$P(L) = \begin{cases} \frac{\alpha}{4} & \text{if L is compositional} \\ \frac{1-\alpha}{252} & \text{otherwise} \end{cases}$$

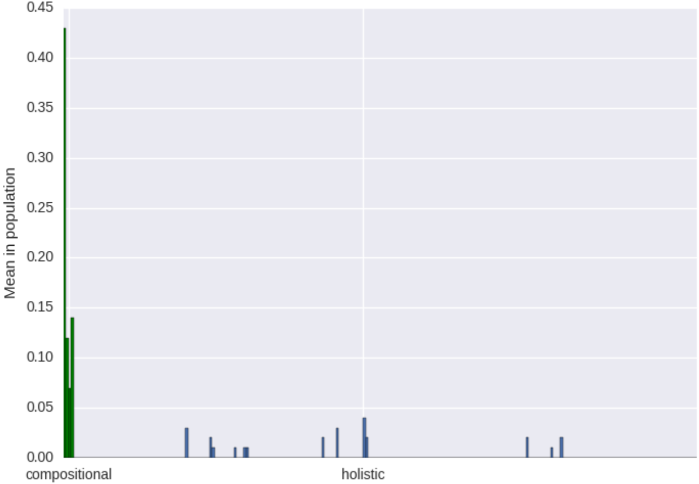
- Minor errors in production

$$P(m|s; L) \propto \begin{cases} 1-\epsilon & \text{if m is true of s in L} \\ \epsilon & \text{otherwise} \end{cases}$$

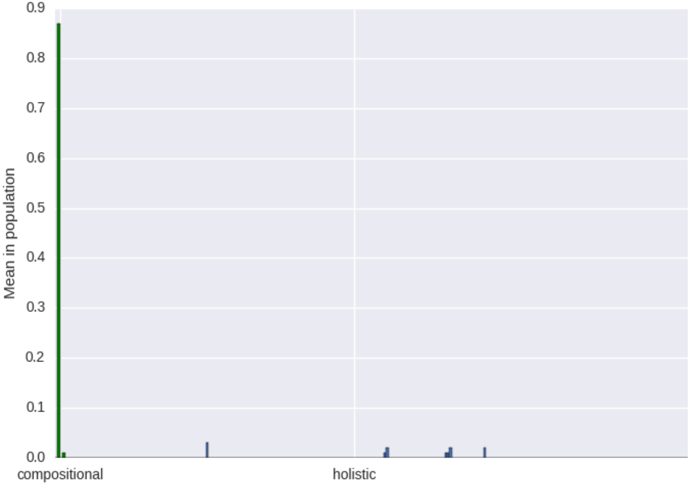
Compositionality & IL: Posterior sampling only ($\alpha = 0.8$)



Compositionality & IL: MAP-like ($\alpha = 0.8, l = 5$)



Compositionality & IL: Replication + MAP-like ($\alpha = 0.8, l = 5$)



Recent applications of vertical and horizontal dynamics

- Compositionality
- Grammaticalization cycles (e.g. aspect)
- Vagueness
- Categorization (e.g. color)
- Ambiguity
- Semantic-Pragmatics distinction
- Human experiments in artificial language learning with IL

Raffaele Argiento, Robin Pemantle, Brian Skyrms, and Stanislav Volkov.
Learning to signal: Analysis of a micro-level reinforcement model.
Stochastic Processes and their Applications, 119(2):373–390, 2009.

Jeffrey A. Barrett.
The evolution of coding in signaling games.
Theory and Decision, 67(2):223–237, 2009.

Alan W. Beggs.
On the convergence of reinforcement learning.
Journal of Economic Theory, 122(1):1–36, 2005.

Anton Benz, Gerhard Jäger, and Robert van Rooij.
***An Introduction to Game Theory for Linguists*, pages 1–82.**
Palgrave, 2005.

Justin Bruner, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger.
David Lewis in the lab: experimental results on the emergence of meaning.
Synthese, 2014.

Thomas Brochhagen.
Minimal requirements for productive compositional signaling.
In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 2015.

David Catteeuw and Bernard Manderick.
The limits and robustness of reinforcement learning in Lewis signaling games.
Connection Science, 26(2):161–177, 2014.

Ido Erev and Alvin E. Roth.
Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria.
American economic review, 88(4):848–881, 1998.

Michael Franke.
The evolution of compositionality and proto-syntax in signaling games.

Journal of Logic, Language and Information, 2015.

Thomas L. Griffiths and Michael L. Kalish.
Language evolution by iterated learning with bayesian agents.
Cognitive Science, 31(3):441–480, 2007.

Richard J Herrnstein.
On the law of effect.
Journal of the experimental analysis of behavior, 13(2):243–266, 1970.

S. M. Huttegger and K. J. S. Zollman.
Methodology in biological game theory.
The British Journal for the Philosophy of Science, 64(3):637–658, 2013.

Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith.
Compression and communication in the cultural evolution of linguistic structure.
Cognition, 141:87–102, 2015.

M. A. Nowak and D. C. Krakauer.
The evolution of language.
Proceedings of the National Academy of Sciences, 96(14):8028–8033, 1999.

Martin A. Nowak, Natalia L. Komarova, and Partha Niyog.
Evolution of universal grammar.
Science, 291(5501):114–118, jan 2001.

Cailin O'Connor.
Evolving to generalize: Trading precision for speed.
British Journal for the Philosophy of Science, forthcoming.

Alvin E. Roth and Ido Erev.
Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term.
Games and Economic Behavior, 8(1):164–212, 1995.

David Reitter and Christian Lebiere.
Towards explaining the evolution of domain languages with cognitive simulation.
Cognitive Systems Research, 2009.

Richard S. Sutton and Andrew G. Barto.
***Introduction to Reinforcement Learning*.**
MIT Press, Cambridge, MA, USA, 1998.

Matthew Spike, Kevin Stadler, Simon Kirby, and Kenny Smith.
Minimal requirements for the emergence of learned signaling.
Cognitive Science, pages 1–36, 2016.

Luc Steels.
Modeling the cultural evolution of language.
Physics of Life Reviews, 8(4):339–356, 2011.

Peter D. Taylor and Leo B. Jonker.
Evolutionary stable strategies and game dynamics.
Mathematical Bioscience, 40(1–2):145–156, 1978.

Monica Tamariz and Simon Kirby.
The cultural evolution of language.
Current Opinion in Psychology, 8:37–43, 2016.