Learning biases may prevent the lexicalization of pragmatic inferences

A case study combining iterated learning and replicator dynamics

Thomas Brochhagen
ILLC, University of Amsterdam

Joint work with Michael Franke (Tübingen) & Robert van Rooij (Amsterdam)

Centre for Language Evolution seminar series, Edinburgh, 2016.07.19

The semantics-pragmatics distinction

Semantics

Literal meaning (truth-conditional)

Pragmatics

Information beyond literal meaning (e.g. defeasible inferences)

Scalar inferences

(1) $\langle$*few*, *some*, *many*, *most*, *all*$\rangle$

a. All students came to class
$\rightarrow$ Some students came to class

b. Some students came to class
$\rightsquigarrow$ Not all students came to class

(2) $\langle$*may*, *should*, *must*$\rangle$

(3) $\langle$*one*, *two*, *three*, ...$\rangle$

(4) $\langle$*or*, *and*$\rangle$

(5) ...

The use of a less informative expression when a more informative one could have been used* can license a defeasible inference that stronger alternatives do not hold

*The hearer assumes the speaker to be knowledgeable and cooperative

· Laurence R. Horn. *On the Semantic Properties of Logical Operators in English*.
Indiana University Linguistics Club, Bloomington, IN, 1972

· Gerald Gazdar. *Pragmatics, Implicature, Presupposition and Logical Form*.
Academic Press, New York, 1979

· Paul Grice. Logic and conversation.
In *Studies in the Ways of Words*, chapter 2, pages 22–40. Harvard University Press, Cambridge, MA, 1975

1. Why are (pragmatically inferred) upper-bounds of weak(er) alternatives not part of semantics?
2. What justifies semantic structure in light of pragmatic enrichment?

Today's talk:

- Propose a model to address (2) and analyze dynamics of linguistic pressures more broadly
- Use (1) as a case study for (2)

Explanations for a lack of semantic upper-bounds I

(1) Signal no commitment to stronger alternatives when knowledge/cooperativity are not given

Cf. lexicalizing ‘some’ and ‘sbna’

Conditions that may pressure for English-like semantics:

- Contextual cues are very reliable
- Morphosyntactic disambiguation is not frequently necessary
- Morphosyntactic disambiguation is not very costly
- Cost of larger lexica higher than morphosyntactic disambiguation

Explanations for a lack of semantic upper-bounds II

(2) Lack of upper-bounds provides learnability advantage

Model

Components I: Probabilistic (pragmatic) language users

- Varied lexica
- Varied production and comprehension behavior (here: parametrized *literal* or *pragmatic*)

$\Bigg\}$ A player's type

· Anton Benz, Gerhard Jäger, Robert Van Rooij, and Robert Van Rooij, editors. *Game theory and pragmatics.* **Springer, 2005**

· Leon Bergen, Roger Levy, and Noah D Goodman. Pragmatic reasoning through semantic inference. ***Semantics and Pragmatics*, 2016**

· M. C. Frank and N. D. Goodman. Predicting pragmatic reasoning in language games. ***Science*, 336(6084):998–998, 2012**

· Michael Franke and Gerhard Jäger. Pragmatic back-and-forth reasoning. ***Semantics, Pragmatics and the Case of Scalar Implicatures.*, pages 170–200, 2014**

Lexica

- s_1: Bill read some but not all books
- s_2: Bill read all books

$$L_a = \begin{array}{c c} & \begin{array}{cc} m_{\text{all}} & m_{\text{some}} \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \end{array}$$

$$L_b = \begin{array}{c c} & \begin{array}{cc} m_{\text{all}} & m_{\text{some}} \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

Literal behavior

$$R_0(s|m;L) \propto P^*(s)L_{sm} \tag{1}$$

$$S_0(m|s;L) \propto \exp(\lambda L_{sm}) \tag{2}$$

Example: L_a with $\lambda = 10$

$$\begin{array}{cc} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} m_{\text{all}} \\ m_{\text{some}} \end{array} & \begin{pmatrix} 0 & 1 \\ .5 & .5 \end{pmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{cc} m_{\text{all}} & m_{\text{some}} \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{pmatrix} \exp(0) & \exp(\lambda) \\ \exp(\lambda) & \exp(\lambda) \end{pmatrix} \end{array}$$

Pragmatic behavior

$$R_1(s|m;L) \propto P^*(s) S_0(m|s;L) \tag{3}$$

$$S_1(m|s;L) \propto \exp(\lambda R_0(s|m;L)^\alpha) \tag{4}$$

<u>Example:</u> L_a with $\lambda = 10$ and $\epsilon = 1/e^\lambda$

$$\begin{array}{cc} & \begin{array}{cc} s_1 & s_2 \end{array} \\ \begin{array}{c} m_{\text{all}} \\ m_{\text{some}} \end{array} & \begin{pmatrix} \epsilon & .5 \\ 0.99 & .5 \end{pmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{cc} m_{\text{all}} & m_{\text{some}} \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{pmatrix} \exp(0) & \exp(\lambda .5^\alpha) \\ \exp(\lambda 1^\alpha) & \exp(\lambda .5^\alpha) \end{pmatrix} \end{array}$$

Components II: Cultural transmission

Two competing pressures:

$$\left.\begin{array}{l}\text{1. Communicative efficiency} \\ \quad\text{... as replicator dynamics} \\ \text{2. Learnability} \\ \quad\text{... iterated Bayesian learning} \\ \quad\quad\text{as mutator dynamics}\end{array}\right\}\ \text{Replicator-mutator dynamics}$$

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

· Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics.
***Bulletin of the American Mathematical Society*, 40(04):479–520, 2003**

· 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)

- 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

  $\Phi = \sum_i x_i f_i$

Iterated learning (mutator dynamics)

- $Q_{ji} \propto \sum_d P(d|t_j)P(t_i|d)$
- $d = \langle\langle s_h, m_n\rangle, ..., \langle s_l, m_o\rangle\rangle$ of length k
- $P(d|t_j)$
- $P(t_i|d) \propto [P(t_i)P(d|t_i)]^l$, $l \geq 1$
- Prior encodes learning biases of players prior to data exposure

Replicator-mutator dynamics

$$\hat{x}_i = \sum_j Q_{ji} \frac{x_j f_j}{\Phi}$$

Analysis

Lexica, signaling behavior & types

Lexica

$$L_1 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \quad L_2 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad L_3 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$L_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad L_5 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \quad L_6 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

Signaling behavior

Literal or *pragmatic*

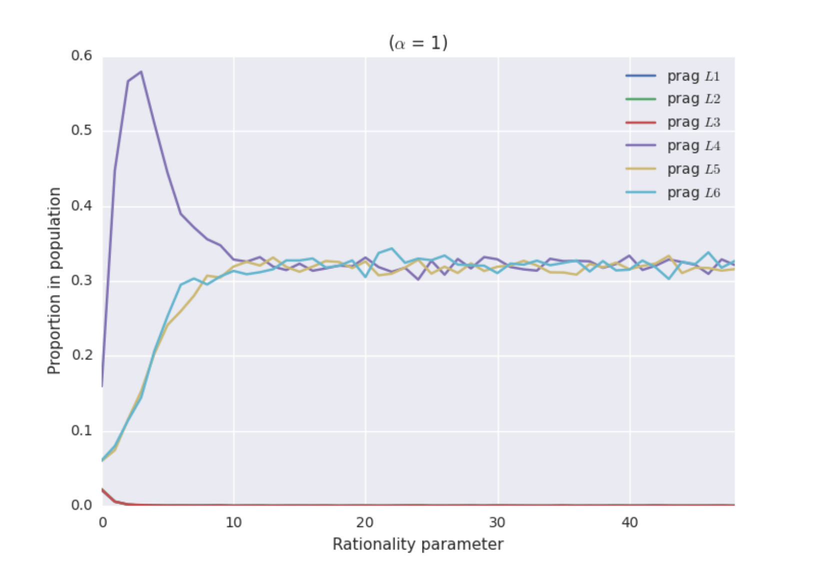
Types

12 types (2 behaviors $\times$ 6 lexica)

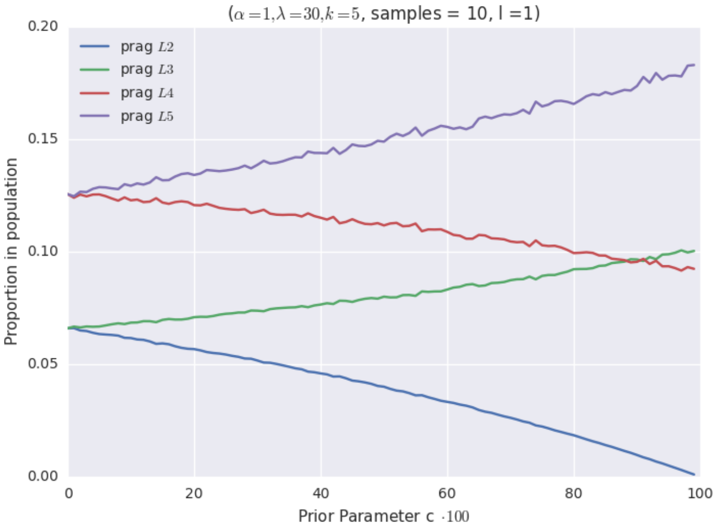
What factors lead to the selection of L_5-like semantics?

- $P(t_i)$ managed by parameter $c \in [0, 1]$
- Signaling behavior: λ, α
- Learning c, k, $\#$ of sequences of length k, l

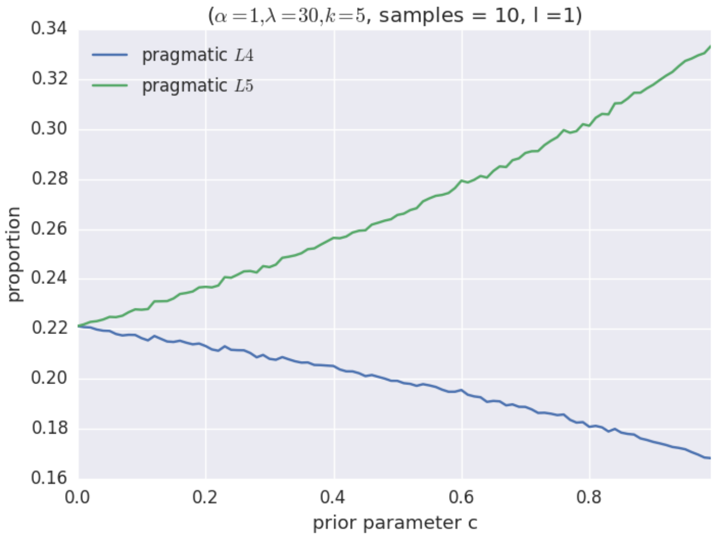
Fitness only



Learning only



Fitness and learning



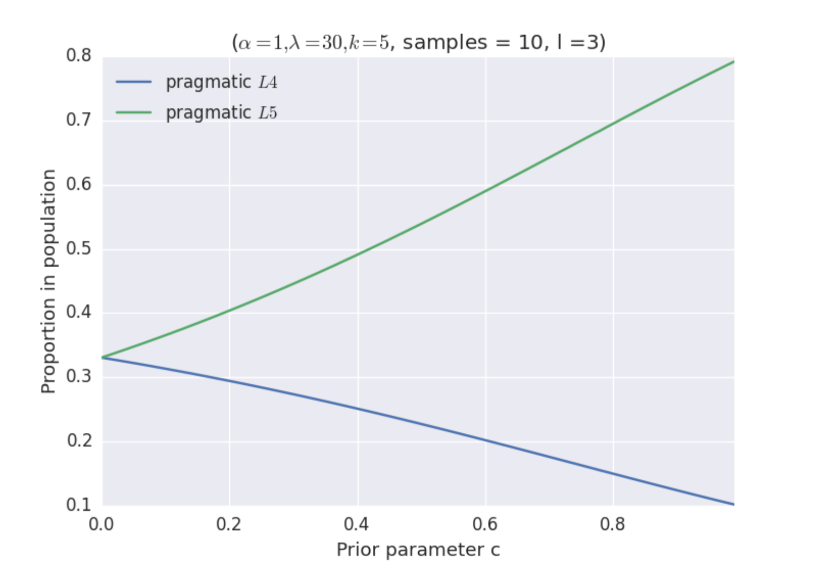
Interim remarks

- Lack of semantic-upper bounds can overcome pressures and stabilize in a population provided...
    - Bias for simpler representations
    - Pragmatics to compensate lack of upper-bounds in use

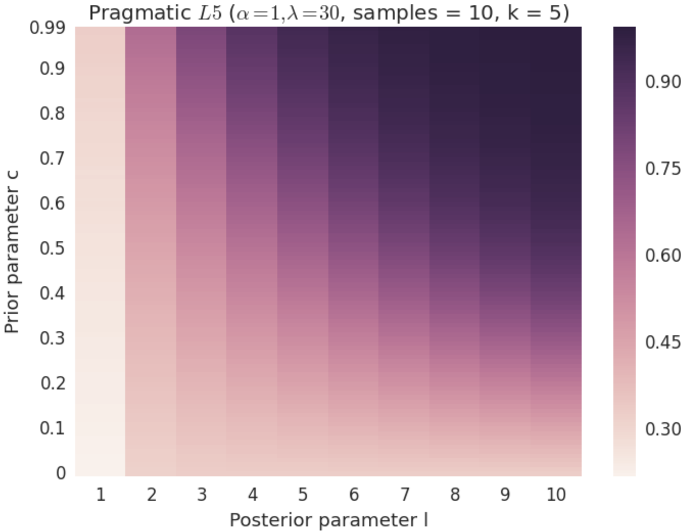
But

- Highly polymorphic populations even for high c
- Role of learning mechanisms, rationality, and learning observations unexplored

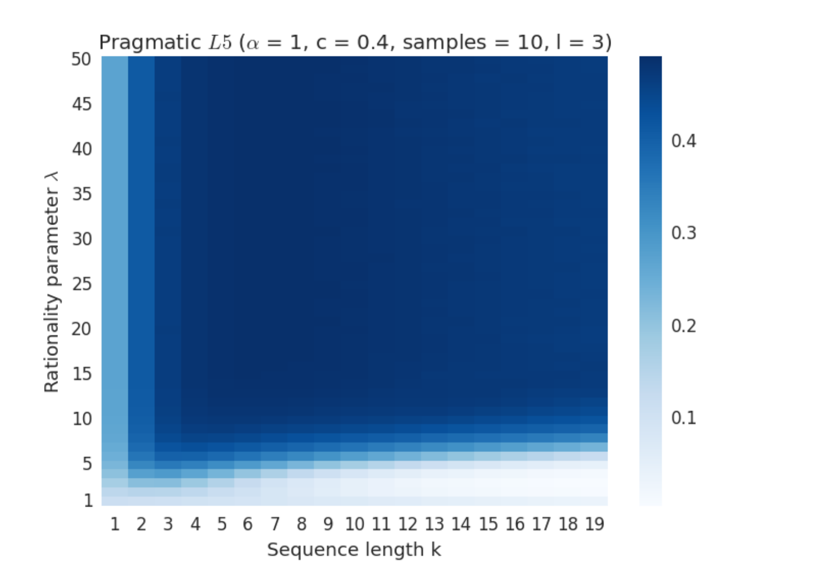
Effect of prior with more posterior maximization



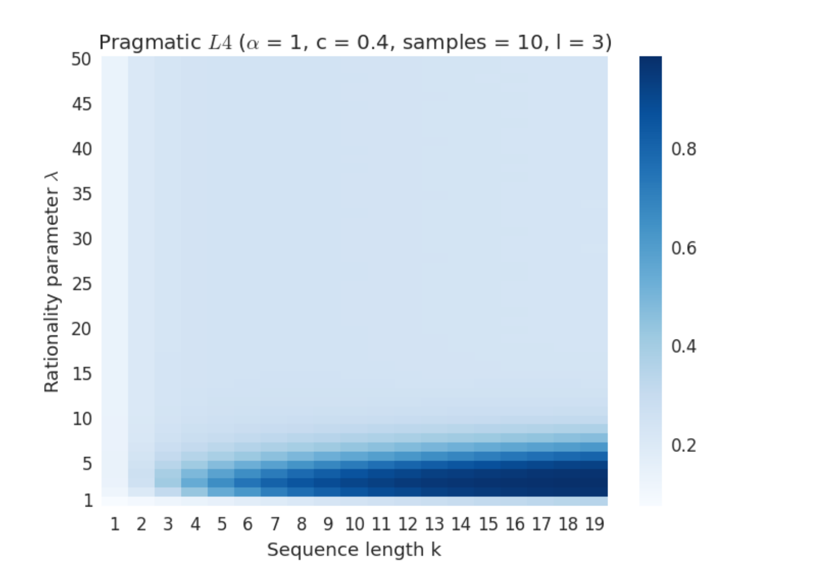
Prior and posterior



Rationality and sequence length I



Rationality and sequence length II



Concluding remarks: Application

- Learnability steers language towards simpler semantics
- Pragmatics compensates for potential loss in expressivity

Lack of semantic upper-bounds

- Functional advantages
    - Reliability of contextual cues to cancel implicatures
    - Lexicon size
    - ...

Provided

- Some degree of rationality in learning & choice
- Relation between bounds and simplicity

Concluding remarks: Model

- Combines
  - Functional pressure
  - Learning pressure
  - (Probabilistic) hearer & speaker models
  - Distinct languages

Future directions

- Generalization
- Frequency effects
- Larger lexica & uncertainty
- Further applications

References I

[BJVRVR05] Anton Benz, Gerhard Jäger, Robert Van Rooij, and Robert Van Rooij, editors.
Game theory and pragmatics.
Springer, 2005.

[BLG16] Leon Bergen, Roger Levy, and Noah D Goodman.
Pragmatic reasoning through semantic inference.
Semantics and Pragmatics, 2016.

[FG12] M. C. Frank and N. D. Goodman.
Predicting pragmatic reasoning in language games.
Science, 336(6084):998–998, 2012.

[FJ14] Michael Franke and Gerhard Jäger.
Pragmatic back-and-forth reasoning.
Semantics, Pragmatics and the Case of Scalar Implicatures., pages 170–200, 2014.

[Gaz79] Gerald Gazdar.
Pragmatics, Implicature, Presuposition and Logical Form.
Academic Press, New York, 1979.

[GK07] Thomas L. Griffiths and Michael L. Kalish.
Language evolution by iterated learning with bayesian agents.
Cognitive Science, 31(3):441–480, 2007.

[Gri75] Paul Grice.
Logic and conversation.
In *Studies in the Ways of Words*, chapter 2, pages 22–40. Harvard University Press, Cambridge, MA, 1975.

References II

[Hor72] Laurence R. Horn.
***On the Semantic Properties of Logical Operators in English*.**
Indiana University Linguistics Club, Bloomington, IN, 1972.

[HS03] Josef Hofbauer and Karl Sigmund.
Evolutionary game dynamics.
Bulletin of the American Mathematical Society, 40(04):479–520, 2003.

[NK99] M. A. Nowak and D. C. Krakauer.
The evolution of language.
Proceedings of the National Academy of Sciences, 96(14):8028–8033, 1999.