Cost-based pragmatic implicatures in an artificial language experiment

Judith Degen, Michael Franke & Gerhard Jäger
Rochester/Stanford
Amsterdam
Tübingen

July 27, 2013

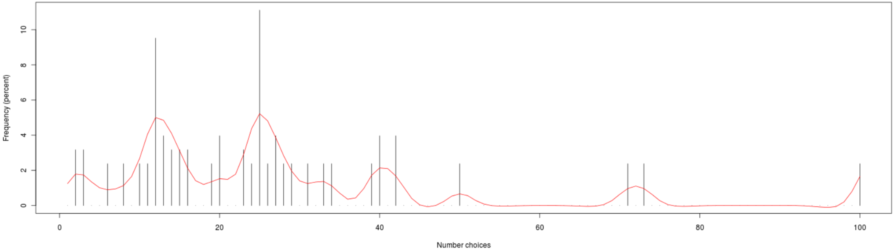
Workshop on Artificial Grammar Learning Tübingen

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

The Beauty Contest

- each participant has to write down a number between 0 and 100
- all numbers are collected
- the person whose guess is closest to 2/3 of the arithmetic mean of all numbers submitted is the winner

The Beauty Contest



(data from Camerer 2003, *Behavioral Game Theory*)

Signaling games

- sequential game:
  1. **nature** chooses a world w
     - out of a pool of possible worlds W
     - according to a certain probability distribution p^*
  2. nature shows w to sender **S**
  3. S chooses a message m out of a set of possible signals M
  4. S transmits m to the receiver **R**
  5. R chooses an action a, based on the sent message.
- Both S and R have preferences regarding R's action, depending on w.
- S might also have preferences regarding the choice of m (to minimize signaling costs).

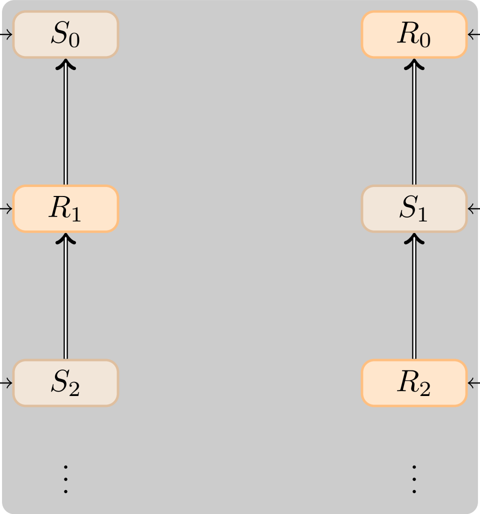
The Iterated Best Response sequence

sends any true message

best response to S_0

best response to R_1

⋮



interprets messages literally

best response to R_0

best response to S_1

⋮

Quantity implicatures

(1)
a. Who came to the party?
b. SOME: Some boys came to the party.
c. ALL: All boys came to the party.

Game construction

- $ct = \emptyset$
- $W = \{w_{\exists\neg\forall}, w_{\forall}\}$
- $w_{\exists\neg\forall} = \{\text{SOME}\}, w_{\forall} = \{\text{SOME}, \text{ALL}\}$
- $p^* = (1/2, 1/2)$

- interpretation function:

$$\begin{aligned} \|\text{SOME}\| &= \{w_{\exists\neg\forall}, w_{\forall}\} \\ \|\text{ALL}\| &= \{w_{\forall}\} \end{aligned}$$

- utilities:

	$a_{\exists\neg\forall}$	$a_{\forall}$
$w_{\exists\neg\forall}$	$1, 1$	$0, 0$
$w_{\forall}$	$0, 0$	$1, 1$

Truth conditions

	SOME	ALL
$w_{\exists\neg\forall}$	1	0
$w_{\forall}$	1	1

Example: Quantity implicatures

S_0	SOME	ALL
$w_{\exists\neg\forall}$	1	0
$w_{\forall}$	$1/2$	$1/2$

R_0	$w_{\exists\neg\forall}$	$w_{\forall}$
SOME	$1/2$	$1/2$
ALL	0	1

R_1	$w_{\exists\neg\forall}$	$w_{\forall}$
SOME	1	0
ALL	0	1

S_1	SOME	ALL
$w_{\exists\neg\forall}$	1	0
$w_{\forall}$	0	1

$$F = (R_1, S_1)$$

In the fixed point, SOME is interpreted as entailing $\neg$ALL, i.e. exhaustively.

Lifted games

1. a. Ann or Bert showed up. (= OR)
   b. Ann showed up. (= A)
   c. Bert showed up. (= B)
   d. Ann and Bert showed up. (= AND)

- w_a: Only Ann showed up.
- w_b: Only Bert showed up.
- w_{ab}: Both showed up.

Truth conditions

	OR	A	B	AND
$\{w_a\}$	1	1	0	0
$\{w_b\}$	1	0	1	0
$\{w_{ab}\}$	1	1	1	1
$\{w_a, w_b\}$	1	0	0	0
$\{w_a, w_{ab}\}$	1	1	0	0
$\{w_b, w_{ab}\}$	1	0	1	0
$\{w_a, w_b, w_{ab}\}$	1	0	0	0

Lifted games

IBR sequence: 1

S_0	OR	A	B	AND
$\{w_a\}$	$1/2$	$1/2$	0	0
$\{w_b\}$	$1/2$	0	$1/2$	0
$\{w_{ab}\}$	$1/4$	$1/4$	$1/4$	$1/4$
$\{w_a, w_b\}$	1	0	0	0
$\{w_a, w_{ab}\}$	$1/2$	$1/2$	0	0
$\{w_b, w_{ab}\}$	$1/2$	0	$1/2$	0
$\{w_a, w_b, w_{ab}\}$	1	0	0	0

Lifted games

IBR sequence: 2

R_1	$\{w_a\}$	$\{w_b\}$	$\{w_{ab}\}$	$\{w_a, w_b\}$	$\{w_a, w_{ab}\}$	$\{w_b, w_{ab}\}$	$\{w_a, w_b, w_{ab}\}$
OR	0	0	0	1	0	0	0
A	1	0	0	0	0	0	0
B	0	1	0	0	0	0	0
AND	0	0	1	0	0	0	0

Lifted games

IBR sequence: 3

S_2	OR	A	B	AND
$\{w_a\}$	0	1	0	0
$\{w_b\}$	0	0	1	0
$\{w_{ab}\}$	0	0	0	1
$\{w_a, w_b\}$	1	0	0	0
$\{w_a, w_{ab}\}$	$1/2$	$1/2$	0	0
$\{w_b, w_{ab}\}$	$1/2$	0	$1/2$	0
$\{w_a, w_b, w_{ab}\}$	1	0	0	0

Lifted games

- OR is only used in $\{w_a, w_b\}$ in the fixed point
- this means that it carries two implicatures:
  - exhaustivity: Ann and Bert did not both show up
  - ignorance: Sally does not know which one of the two disjuncts is true

Predicting behavioral data

- *Behavioral Game Theory*: predict what real people do (in experiments), rather what they ought to do if they were perfectly rational
- one implementation (Camerer, Ho & Chong, TechReport CalTech):
  - **stochastic choice:** people try to maximize their utility, but they make errors
  - **level-k thinking:** every agent performs a fixed number of best response iterations, and they assume that everybody else is less smart (i.e., has a lower strategic level)

Stochastic choice

- real people are not perfect utility maximizers
- they make mistakes $\rightsquigarrow$ sub-optimal choices
- still, high utility choices are more likely than low-utility ones

Rational choice: best response

$$P(a_i) = \begin{cases} \frac{1}{|\arg_j \max u_i|} & \text{if } u_i = \max_j u_j \\ 0 & \text{else} \end{cases}$$

Stochastic choice: (logit) quantal response

$$P(a_i) \propto e^{\lambda u_i}$$

Stochastic choice

- λ measures degree of rationality
- $\lambda = 0$:
  - completely irrational behavior
  - all actions are equally likely, regardless of expected utility
- $\lambda \to \infty$
  - convergence towards behavior of rational choice
  - probability mass of sub-optimal actions converges to 0

Iterated Quantal Response (IQR)

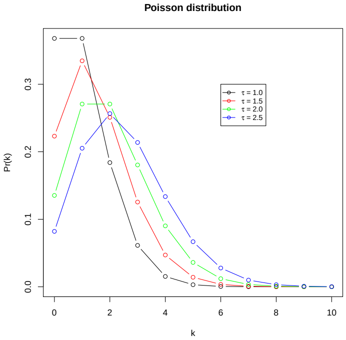
- variant of IBR model
- *best response* ist replaced by *quantal response*
- predictions now depend on value for λ
- no 0-probabilities
- IQR converges gradually

Level-k thinking

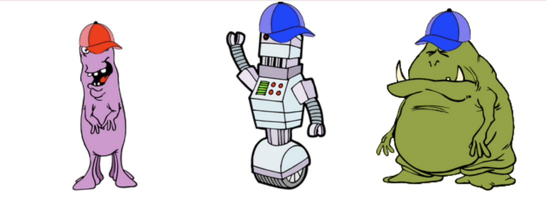
- every player:
  - performs iterated quantal response a limited number k of times (where k may differ between players),
  - assumes that the other players have a level $< k$, and
  - assumes that the strategic levels are distributed according to a **Poisson distribution**

$$P(k) \propto \tau^k / k!$$

- τ, a free parameter of the model, is the average/expected level of the other players

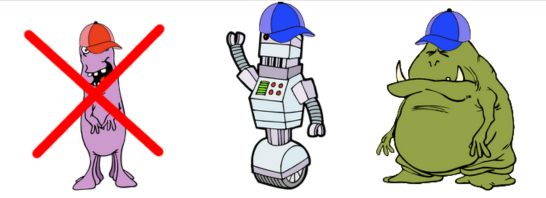


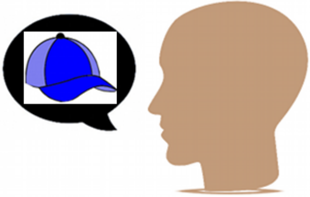
The experimental setup



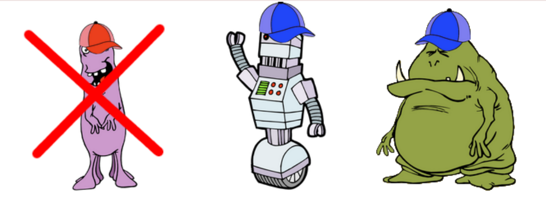
The experimental setup

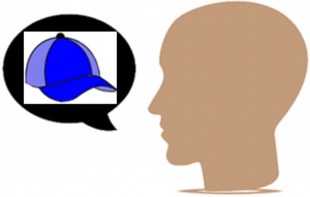
The experimental setup

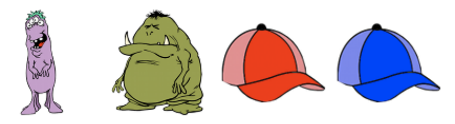




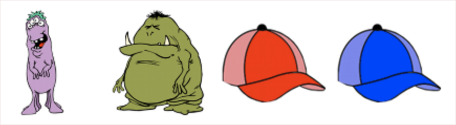
The experimental setup







The experimental setup

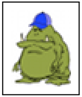


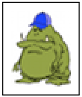
Simple condition: Literal meanings

S_0				
	1/2	0	0	1/2
	0	0	1	0
	0	1/2	1/2	0

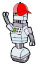
R_0			
	1	0	0
	0	0	1
	0	1/2	1/2
	1	0	0

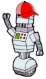
Simple condition: Iterated Best Response

R_1			
	1	0	0
	0	0	1
	0	1	0
	1	0	0

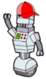
S_1				
	$1/2$	0	0	$1/2$
	0	0	1	0
	0	1	0	0

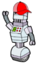
Complex condition: Literal meanings

S_0				
	0	$^1/_2$	0	$^1/_2$
	0	$^1/_2$	$^1/_2$	0
	0	0	0	1

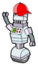
R_0			
	$^1/_3$	$^1/_3$	$^1/_3$
	$^1/_2$	$^1/_2$	0
	0	1	0
	$^1/_2$	0	$^1/_2$

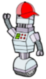
Complex condition: Iterated Best response

R_1			
	$1/3$	$1/3$	$1/3$
	$1/2$	$1/2$	0
	0	1	0
	0	0	1

S_1				
	0	$1/2$	0	$1/2$
	0	0	1	0
	0	0	0	1

Complex condition: Iterated Best response

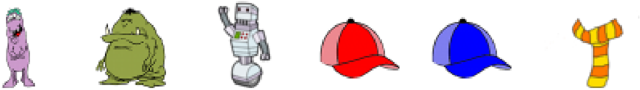
S_2				
	0	1	0	0
	0	0	1	0
	0	0	0	1

R_2			
	$1/3$	$1/3$	$1/3$
	1	0	0
	0	1	0
	0	0	1

Experiment 1 - comprehension

- test participants' behavior in a comprehension task implementing previously described signaling games
- 48 participants on Amazon's Mechanical Turk
- two stages:
  - language learning
  - inference
- 36 experimental trials
  - 6 *simple* (one-step) implicature trials
  - 6 *complex* (two-step) implicature trials
  - 24 filler trials (entirely unambiguous/ entirely ambiguous target)

Artificial language Zorx

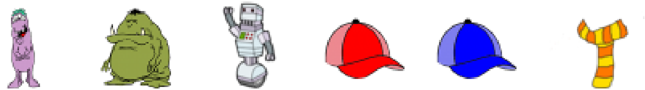


XEK RAV ∅ ZUB KOR ∅

Three stages of language learning:

1 2 3

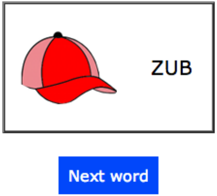
Artificial language Zorx



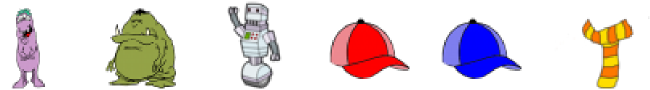
XEK RAV ∅ ZUB KOR ∅

Three stages of language learning:

1 2 3



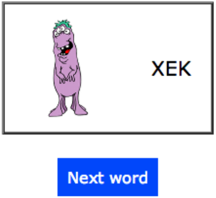
Artificial language Zorx



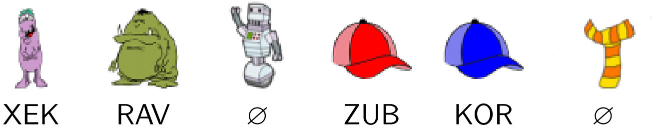
XEK RAV ∅ ZUB KOR ∅

Three stages of language learning:

1 2 3

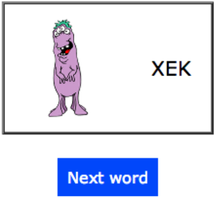


Artificial language Zorx

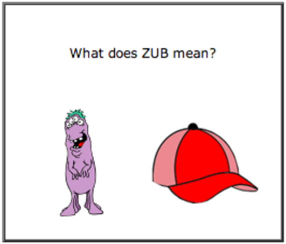


Three stages of language learning:

1

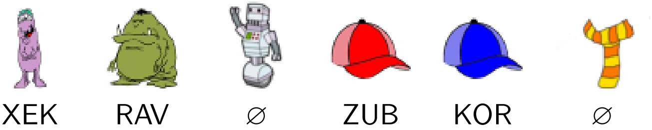


2



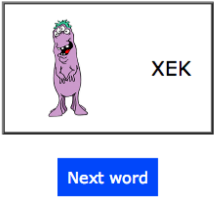
3

Artificial language Zorx

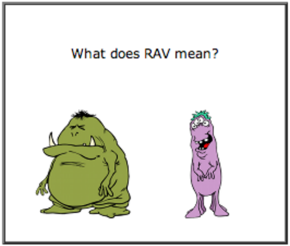


Three stages of language learning:

1

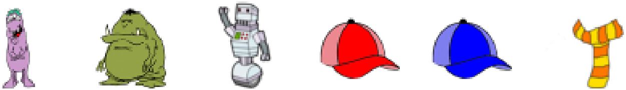


2



3

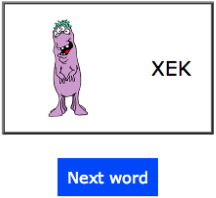
Artificial language Zorx



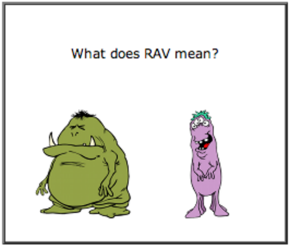
XEK RAV ∅ ZUB KOR ∅

Three stages of language learning:

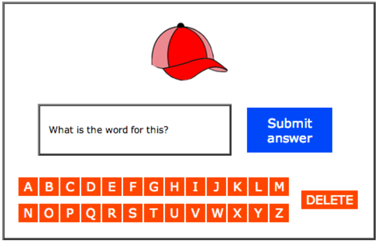
1



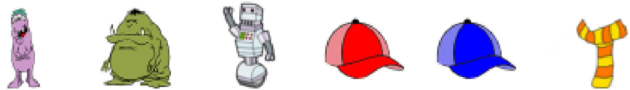
2



3



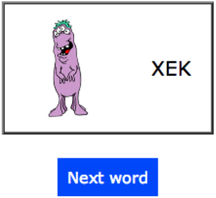
Artificial language Zorx



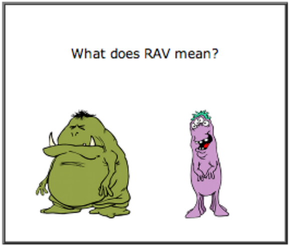
XEK RAV ∅ ZUB KOR ∅

Three stages of language learning:

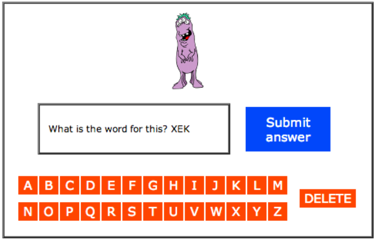
1



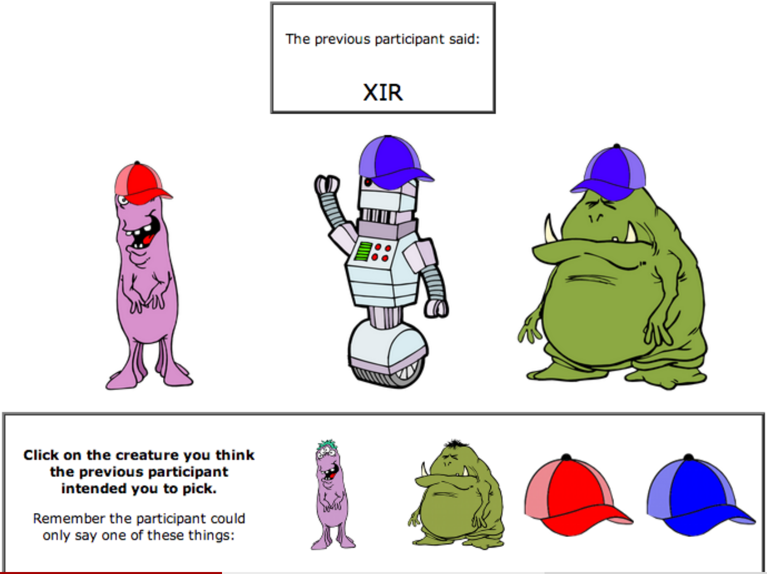
2



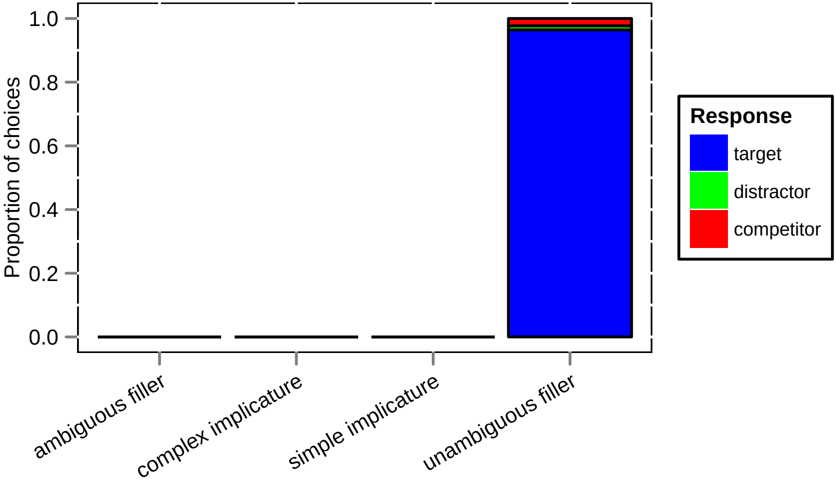
3



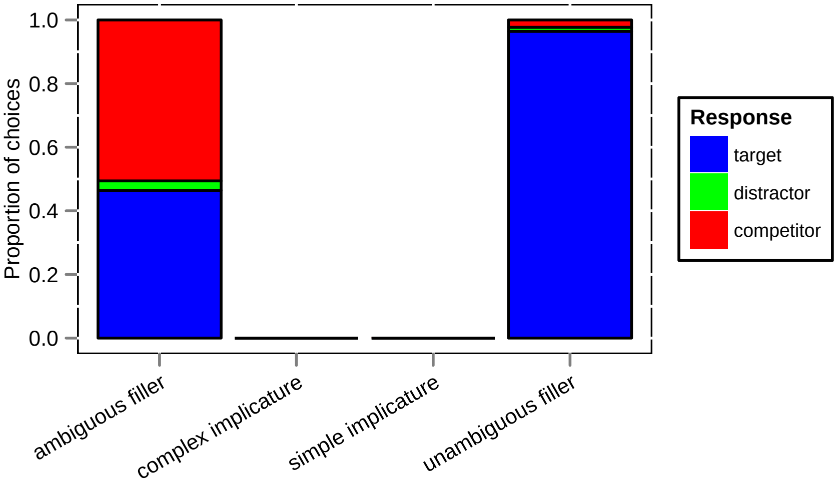
Inference trial



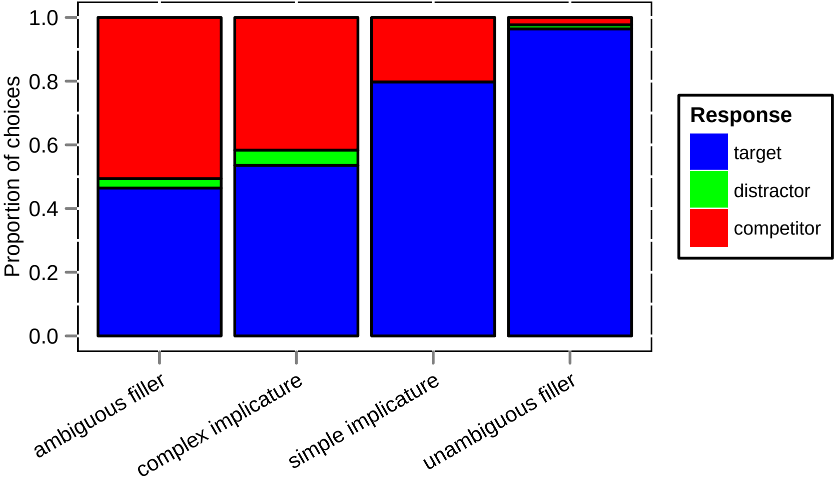
Results - proportion of responses by condition



Results - proportion of responses by condition



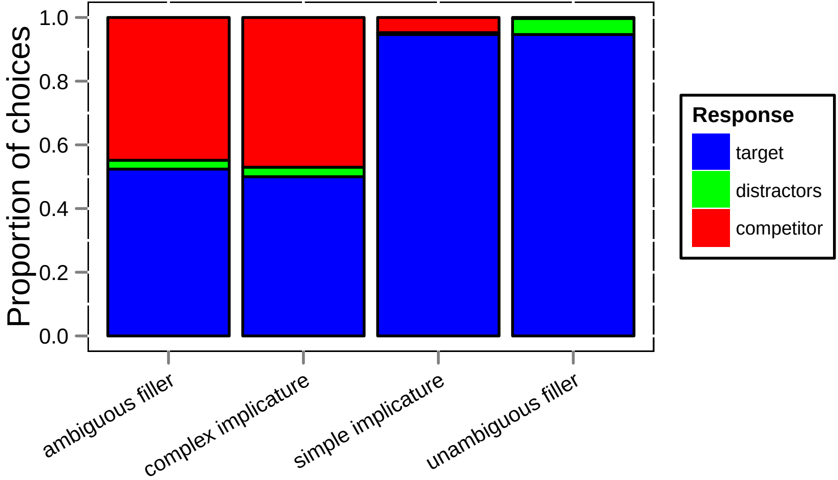
Results - proportion of responses by condition



Experiment 2 - production

- test participants' behavior in a production task implementing previously described signaling games
- 48 participants on Amazon's Mechanical Turk
- two stages:
  - language learning
  - inference
- 36 experimental trials
  - 6 *simple* (one-step) implicature trials
  - 6 *complex* (two-step) implicature trials
  - 24 filler trials (entirely unambiguous/ entirely ambiguous target)

Results - proportion of responses by condition



Experiment 3 - varying message costs

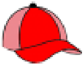
- Question 1: Are comprehenders aware of message costs?
- Question 2: If a cheap ambiguous message competes with a costly unambiguous one, do we find quantity implicatures, and if so, how does its likelihood depend on message costs?
- 240 participants on Amazon's Mechanical Turk
- three stages:
  - language learning
  - cost estimation
  - inference (18 trials, 6 inference and 12 filler trials)

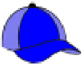
Extended Zorx

cheap messages

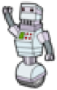

XEK


RAV


ZUB


KOR

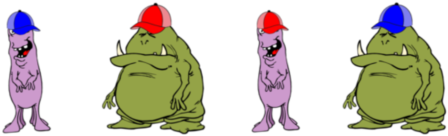
costly messages



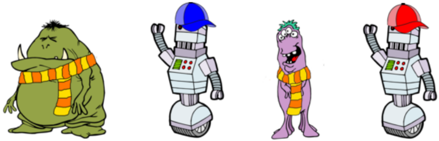
XAB	BAZ	no cost
BAZU	XABI	low cost
BAZUZE	XABIKO	high cost

Cost estimation

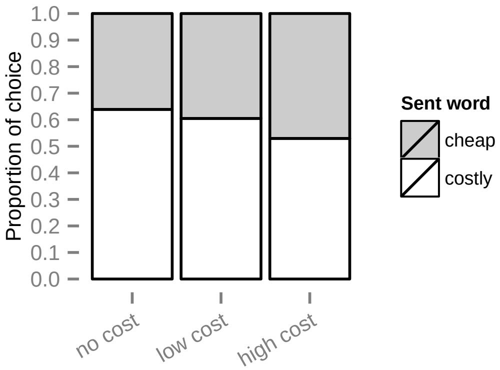
two cheap features



one cheap & one costly feature



Results - proportion of costly messages



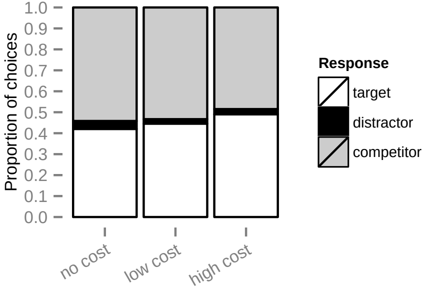
The use of costly messages decreases as the cost of that message increases.

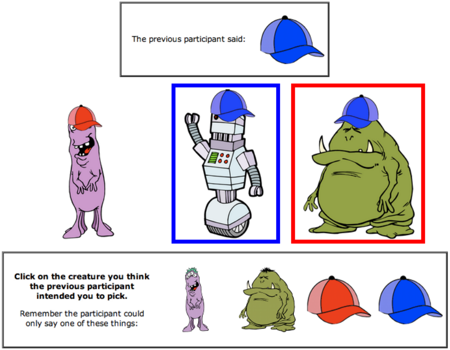
Simple condition: Literal meanings

S_0				
	$1/2$	0	0	$1/2$
	0	0	1	0
	0	$3/4$	$1/4$	0

R_0			
	1	0	0
	0	0	1
	0	$1/2$	$1/2$
	1	0	0

Inference results





The Quantity inference becomes more likely as the cost of the ambiguous message increases.

Model fitting

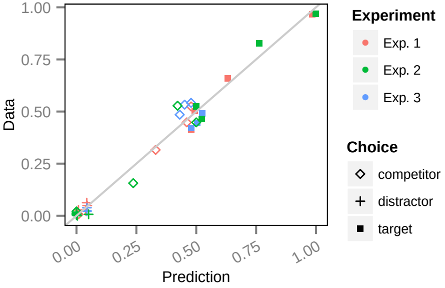
Fitted parameters

- cost estimation: mixed effects logistic regression on the data from experiment 3
- reasoning parameters fitted via least squares regression:
  - comprehension (experiments 1, 3)

    $$\lambda = 4.825, \tau = 0.625, r = 0.99$$

  - production (experiment 2)

    $$\lambda = 8.853, \tau = 0.818, r = 0.99$$



Conclusion

- proof of concept: game theoretic model captures experimental data quite well
- both speakers and listeners routinely perform simple inference steps
- likelihood of nested inferences is rather low
- speakers behave more strategically than listeners

Collaborators