Optimal Reasoning About Referential Expressions

Judith Degen[1] Michael Franke[2] Gerhard Jäger[3]

[1]Department of Brain and Cognitive Sciences
University of Rochester

[2]Institute for Logic, Language and Computation
Universiteit van Amsterdam

[3]Seminar für Sprachwissenschaft
Universität Tübingen

June 8, 2012

Reference to objects

Reference to objects

Reference to objects

Reference to objects

A hard problem

Production (audience design)

Clark & Murphy, 1982; Horton & Keysar, 1996; Brown-Schmidt et al., 2008

Choose a message to convey a given intended meaning with sufficiently high probability.

Comprehension (perspective-taking)

Keysar et al., 2000; Hanna et al., 2003; Heller et al., 2008

Infer the most likely intended interpretation upon observing an utterance.

Today

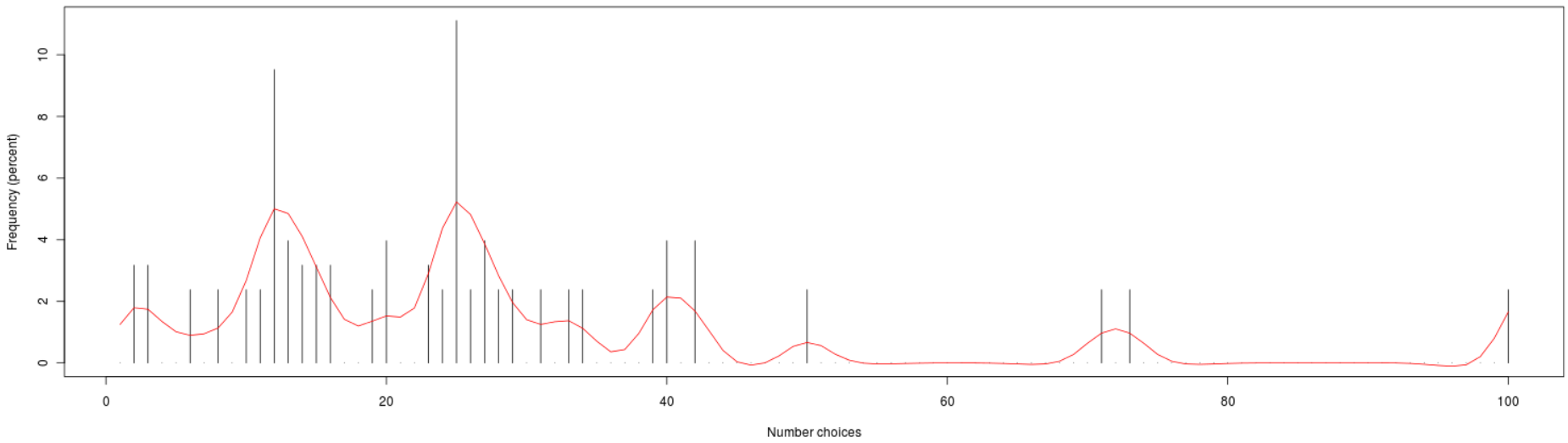
- Provide a game-theoretic model of the inferences involved in production and comprehension of referential expression that provides a benchmark model of rationality.
- Provide experimental evidence from two experiments that language users' choices are boundedly rational.
- Provide a sketch of how to update the standard model that better captures participants' probabilistic choices.

Outline

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



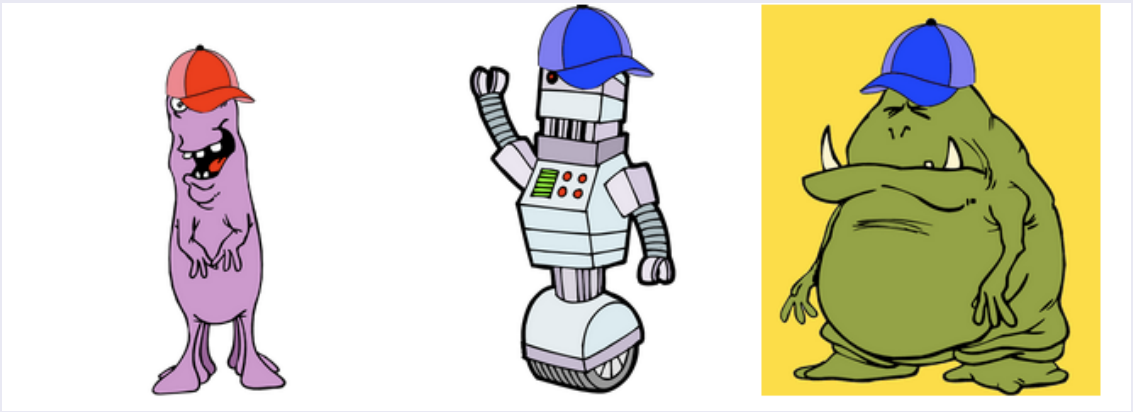
(data from Camerer 2003, *Behavioral Game Theory*)

Signaling games

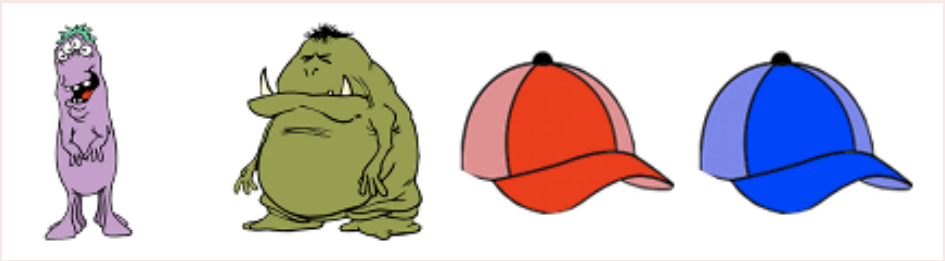
- sequential game:
  1. **nature** chooses a type t
     - out of a pool of possible types T
     - according to a certain probability distribution p^*
  2. nature shows t to sender **S**
  3. S chooses a message m out of a set of possible signals M
  4. S transmits m to the receiver **R**
  5. R guesses a type t', based on the sent message.
- if $t = t'$, both players score a point

An example

Types

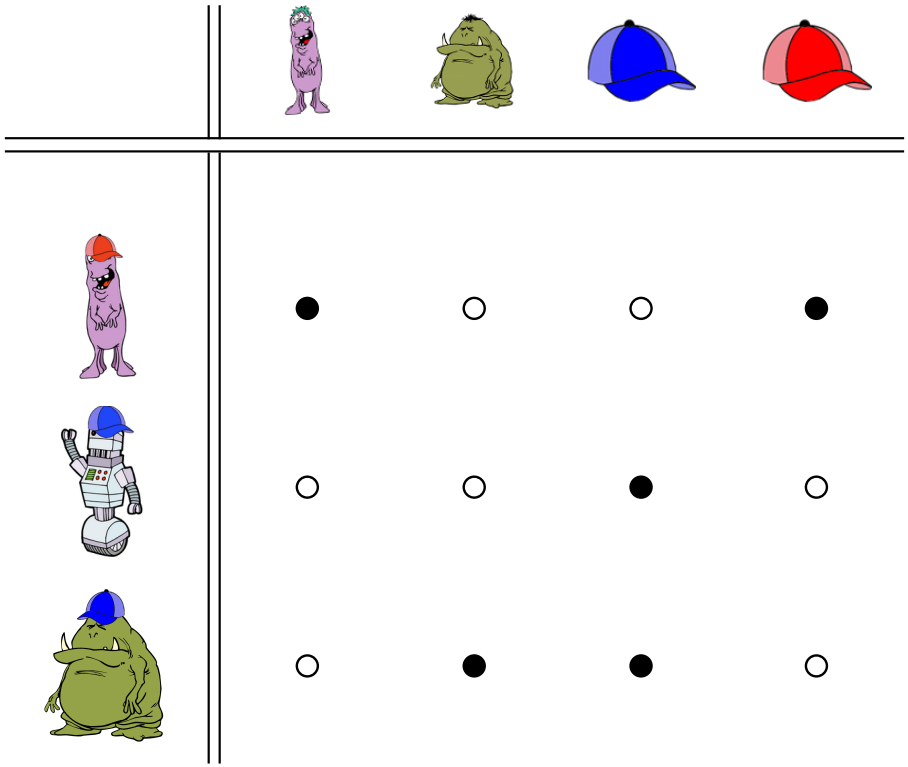


Messages

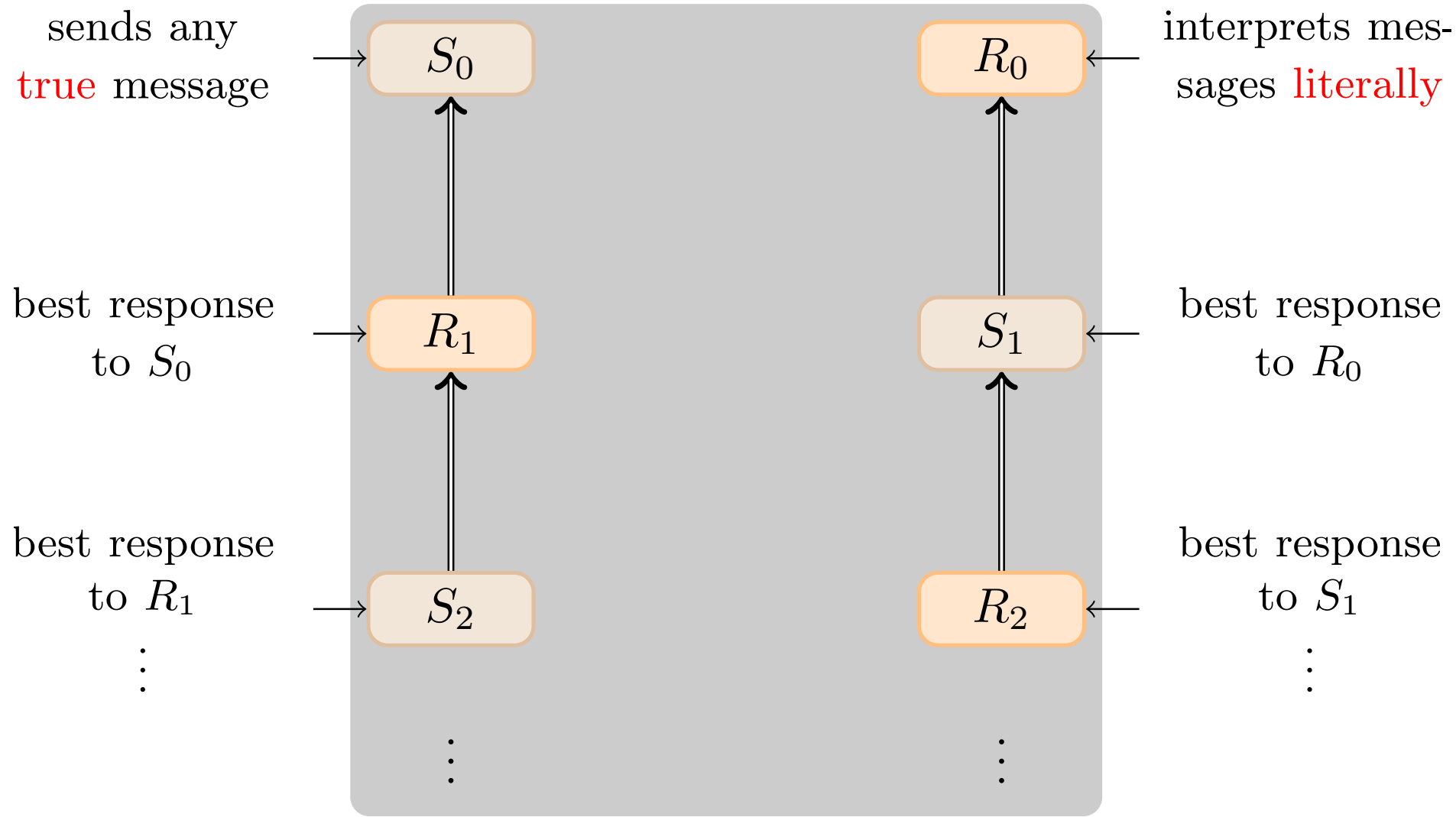


Exogeneous meaning

- Messages may have conventional or iconic meaning (which is common knowledge among the players)
- in our example:

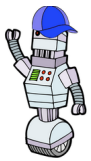


The Iterated Best Response sequence



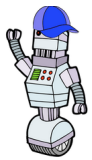
Sender

- Sender strategy S_k gives probabilistic function from types to messages
- if several options are equally good, they are chosen with the same probability
- if $k > 0$, only messages are chosen that maximize the expected utility of S, given R_{k-1}

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

Receiver

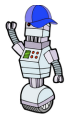
- Receiver strategy R_k gives stochastic function from messages to types
- if several options are equally good, they are chosen with the same probability
- if $k > 0$, only messages are chosen that maximize the expected utility of R, given S_{k-1}

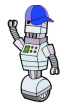
R_0			
	1	0	0
	0	0	1
	0	1/2	1/2
	1	0	0

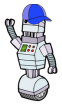
Computing best responses

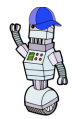
- to compute the best response to a matrix A:
  - transpose A
  - put a 1 in each cell that is maximal within its row, and a 0 everywhere else
  - normalize row-wise

Iterated Best Response

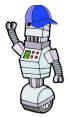
S_0				
	1/2	0	0	1/2
	0	0	1	0
	0	1/2	1/2	0

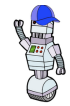
R_0			
	1	0	0
	0	0	1
	0	1/2	1/2
	1	0	0

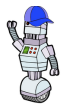
R_1			
	1	0	0
	0	0	1
	0	1	0
	1	0	0

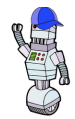
S_1				
	1/2	0	0	1/2
	0	0	1	0
	0	1	0	0

Iterated Best Response (cont.)

S_2				
	1/2	0	0	1/2
	0	0	1	0
	0	1	0	0

R_2			
	1	0	0
	0	0	1
	0	1	0
	1	0	0

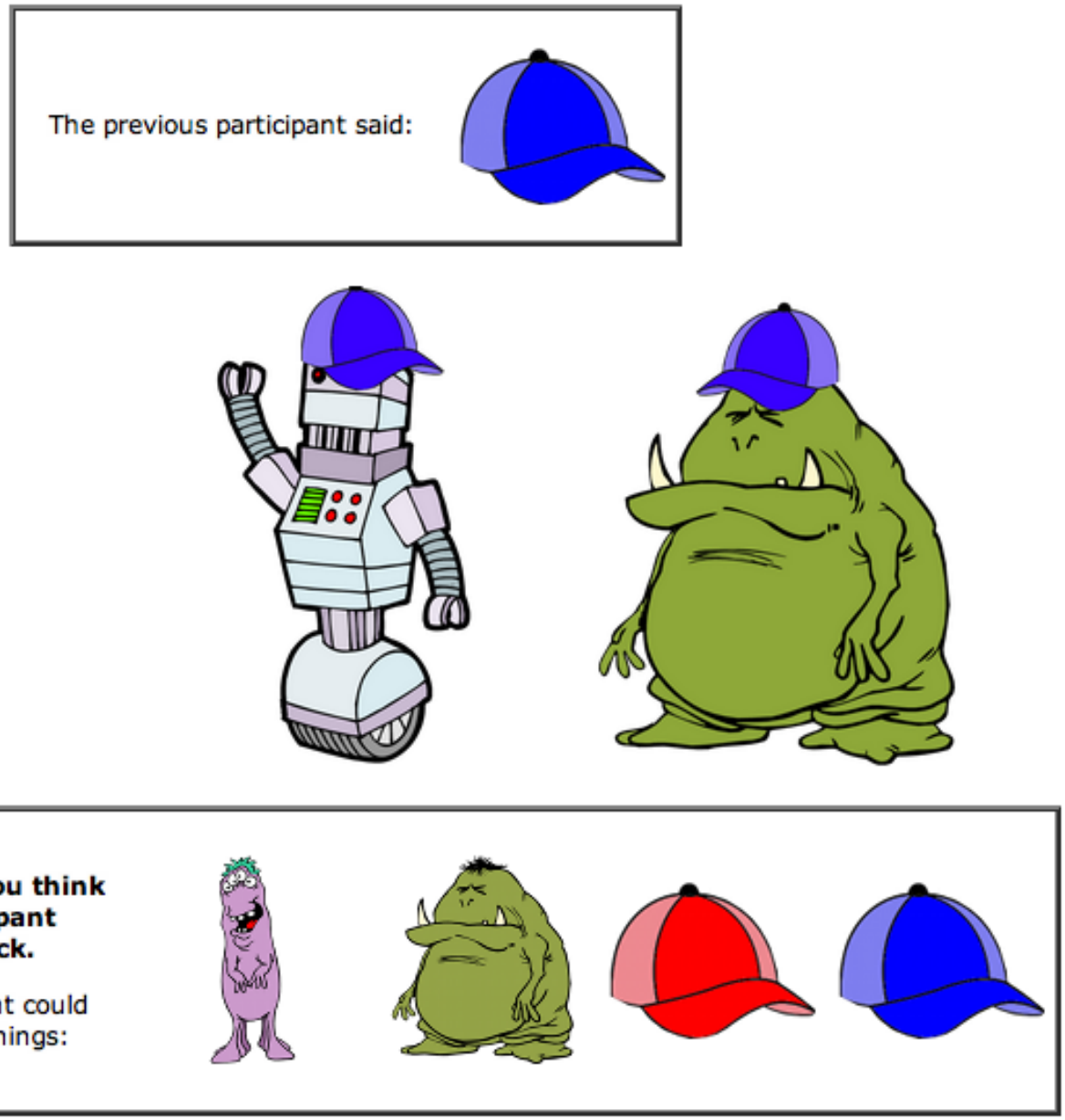
R_3			
	1	0	0
	0	0	1
	0	1	0
	1	0	0

S_3				
	1/2	0	0	1/2
	0	0	1	0
	0	1	0	0

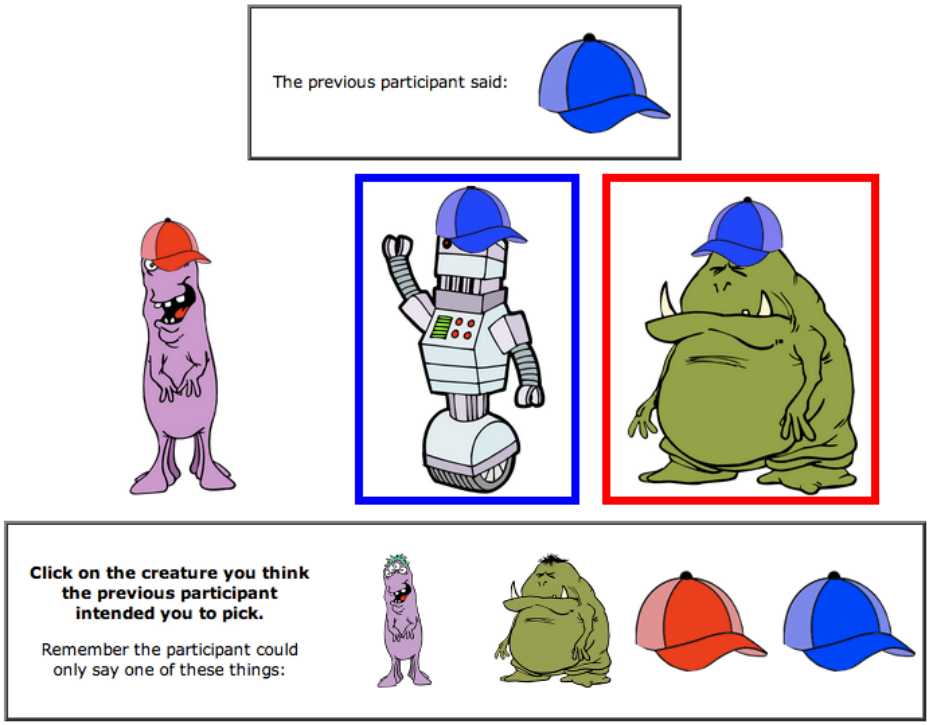
Experiment 1 - comprehension

- test participants' behavior in a comprehension task implementing previously described signaling games
- 30 participants on Amazon's Mechanical Turk
- initially 4 trials as senders
- 36 experimental trials
  - 6 *simple* (one-step) implicature trials
  - 6 *complex* (two-step) implicature trials
  - 24 filler trials (entirely unambiguous/ entirely ambiguous target)

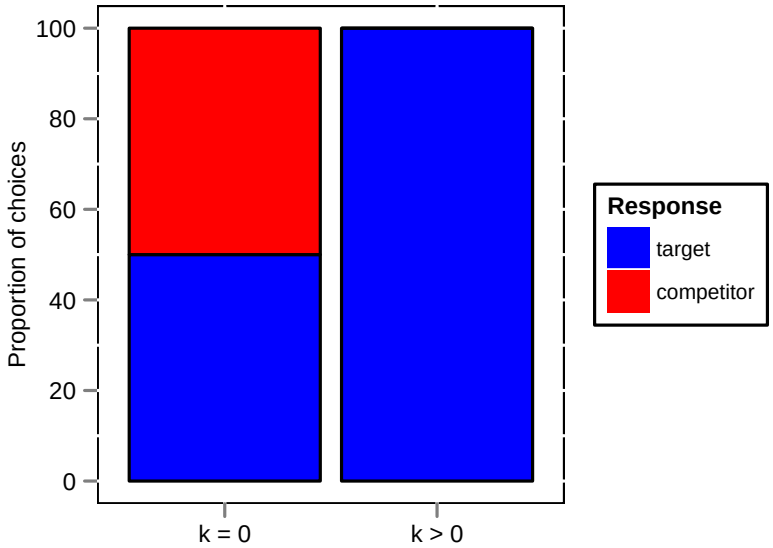
Simple implicature trial



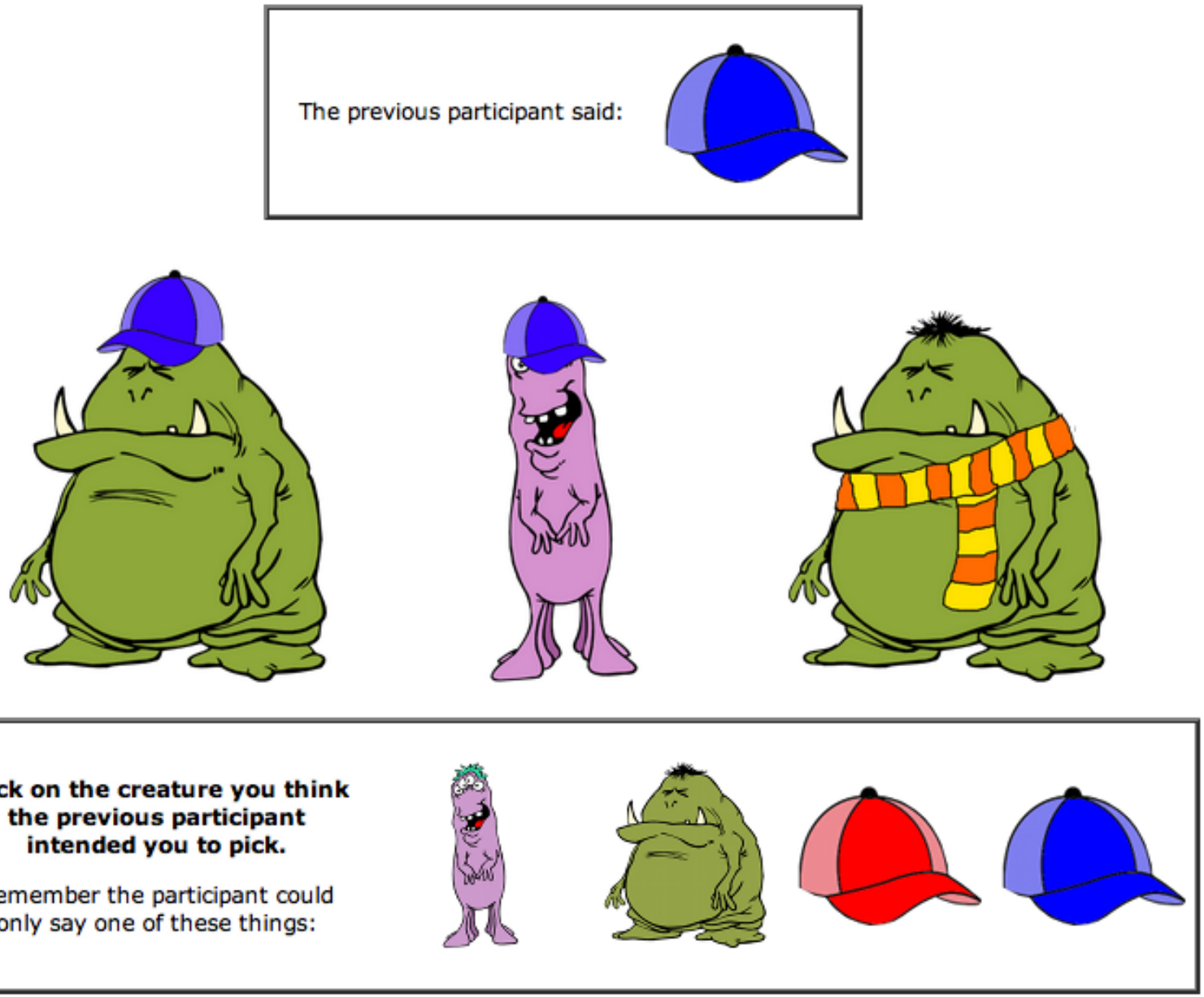
Simple implicature trial - predictions



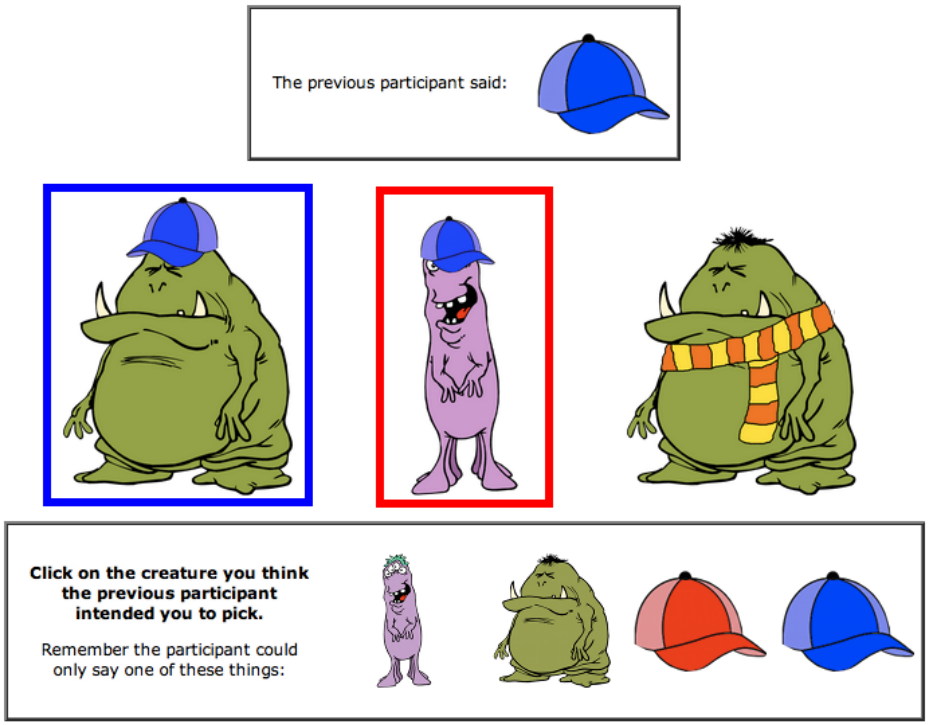
- IBR predictions for distribution of responses over target and competitor:



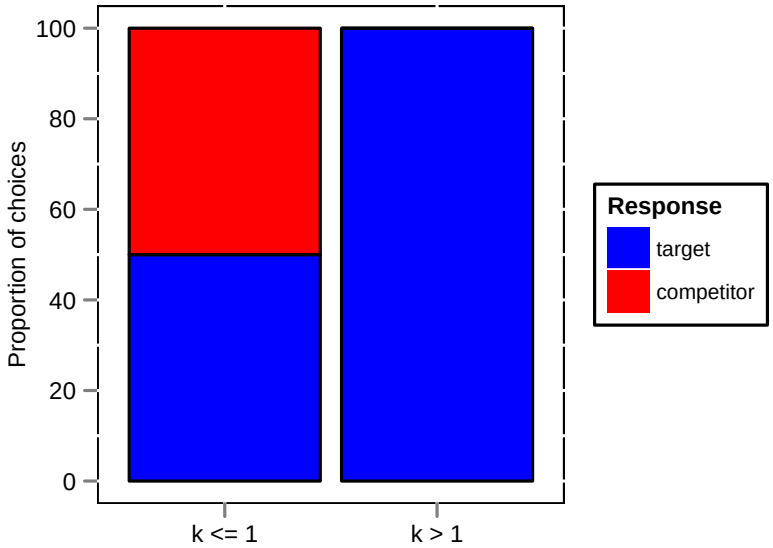
Complex implicature trial



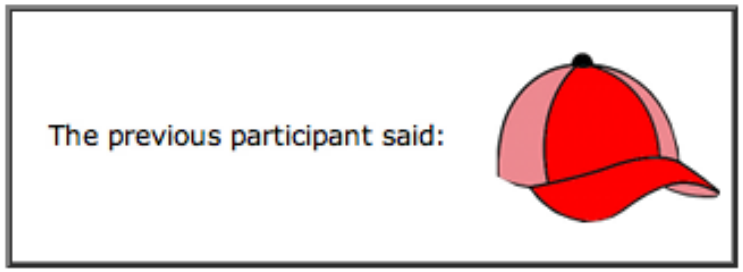
Complex implicature trial - predictions

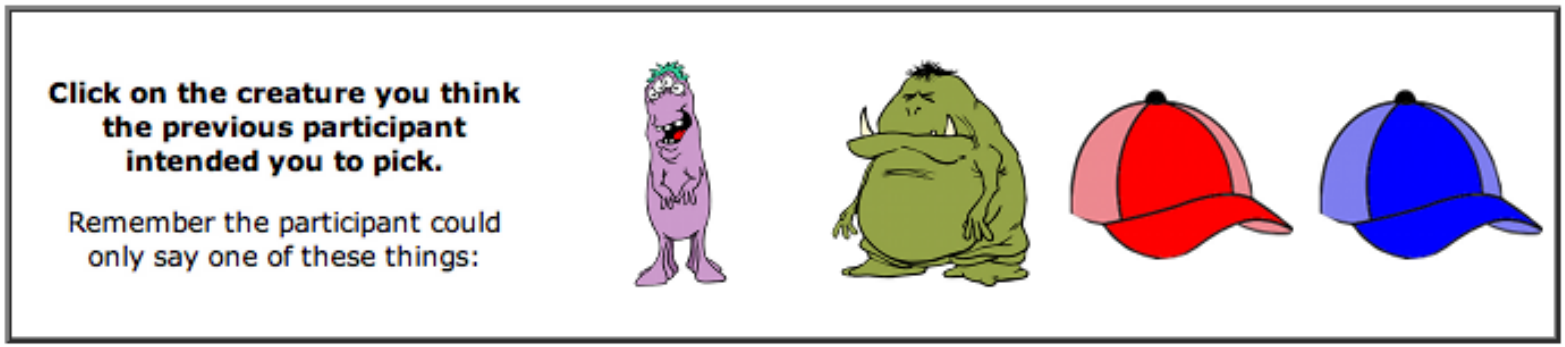


- IBR predictions for distribution of responses over target and competitor:

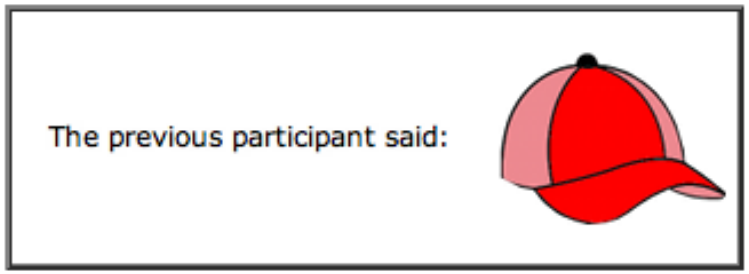


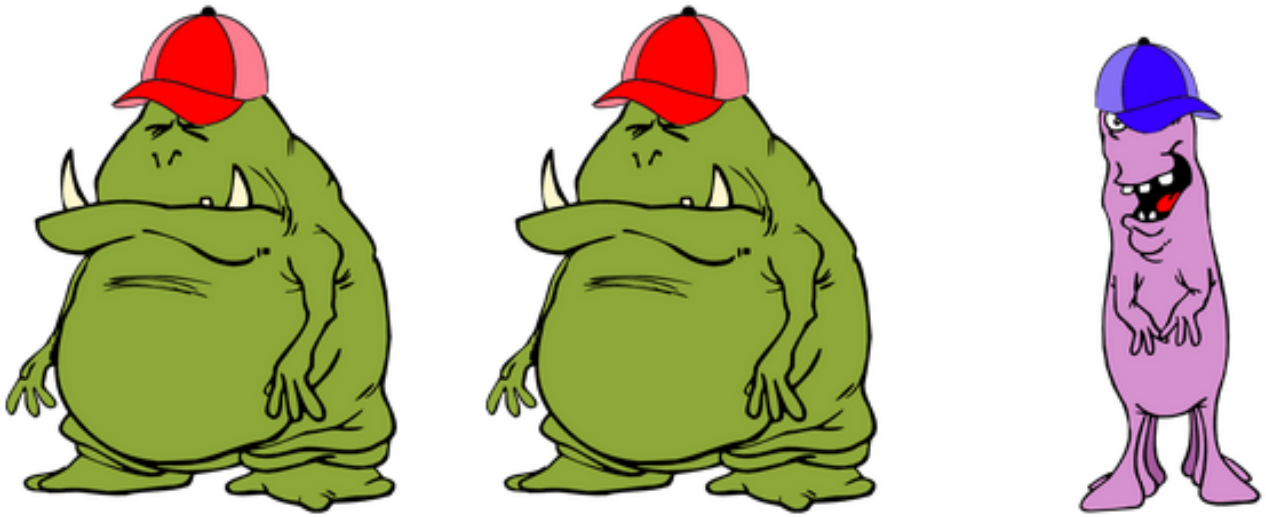
Unambiguous filler

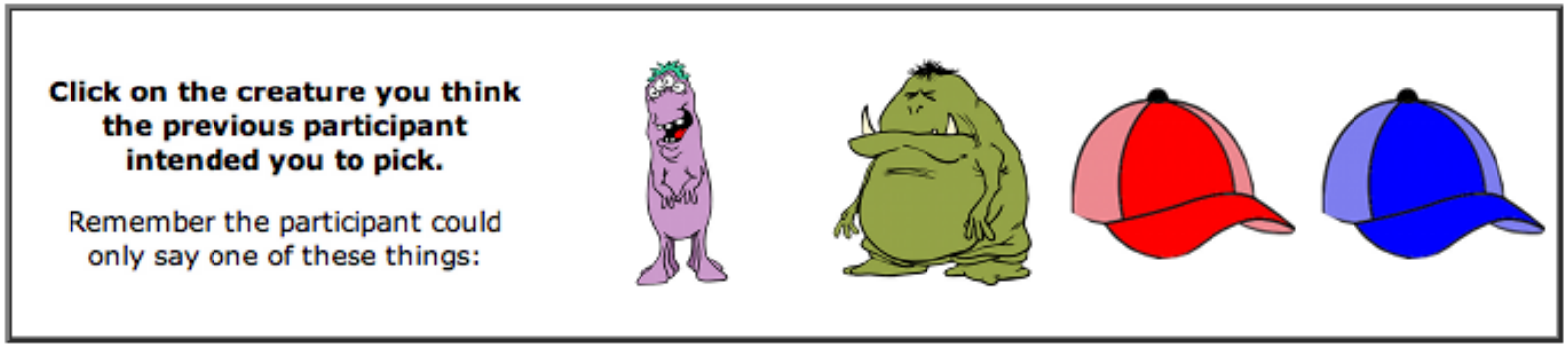




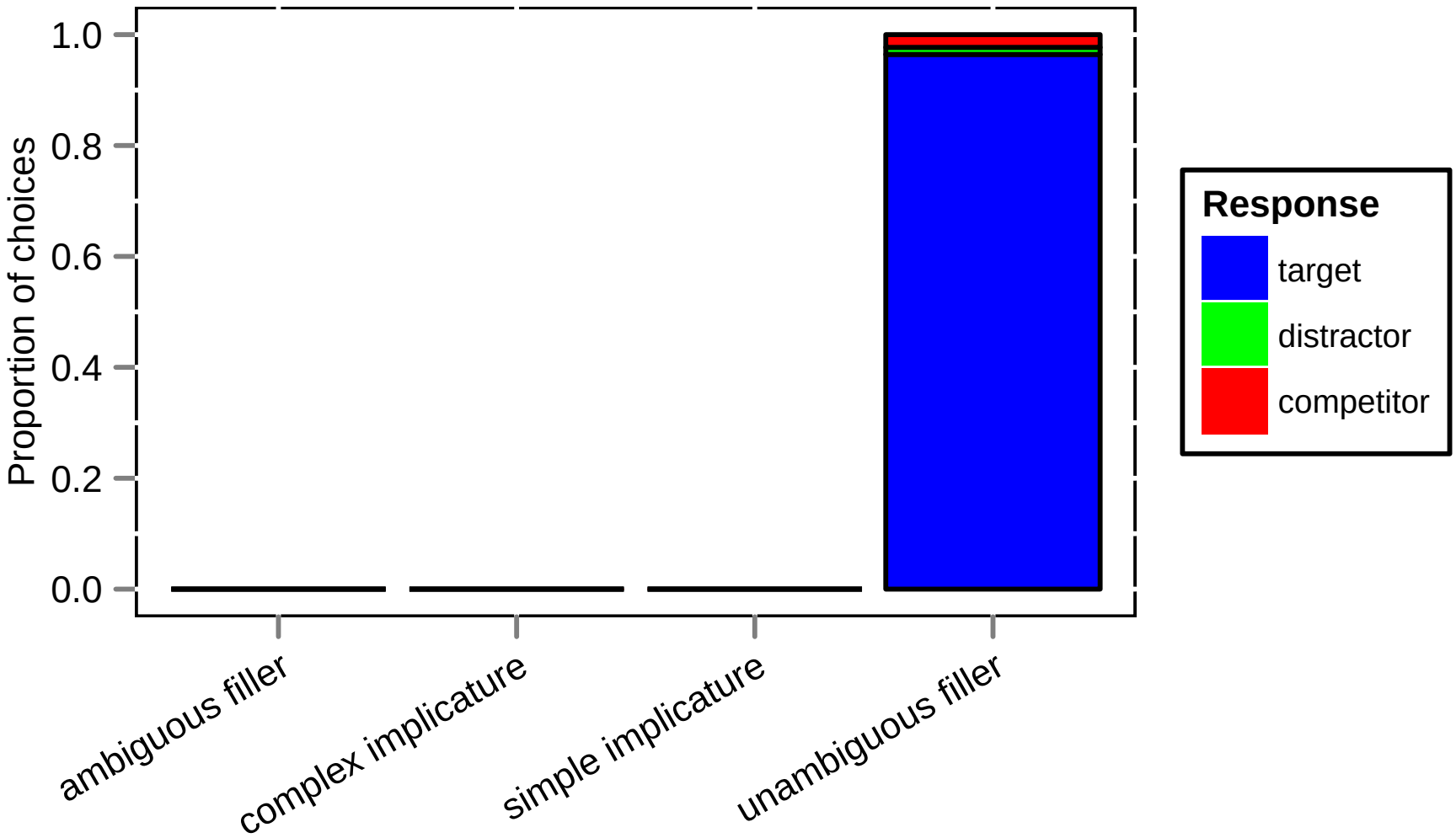
Ambiguous filler



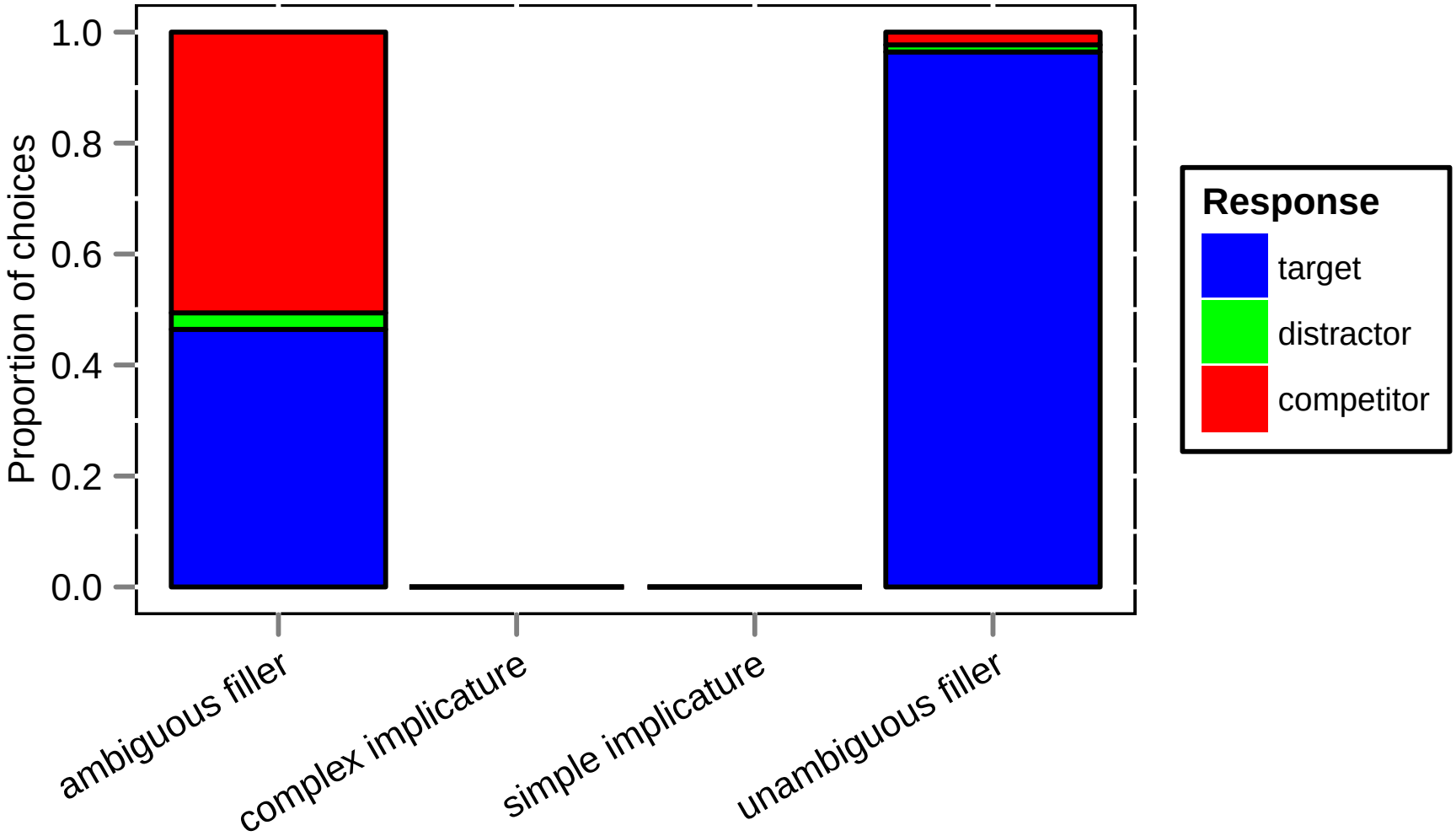




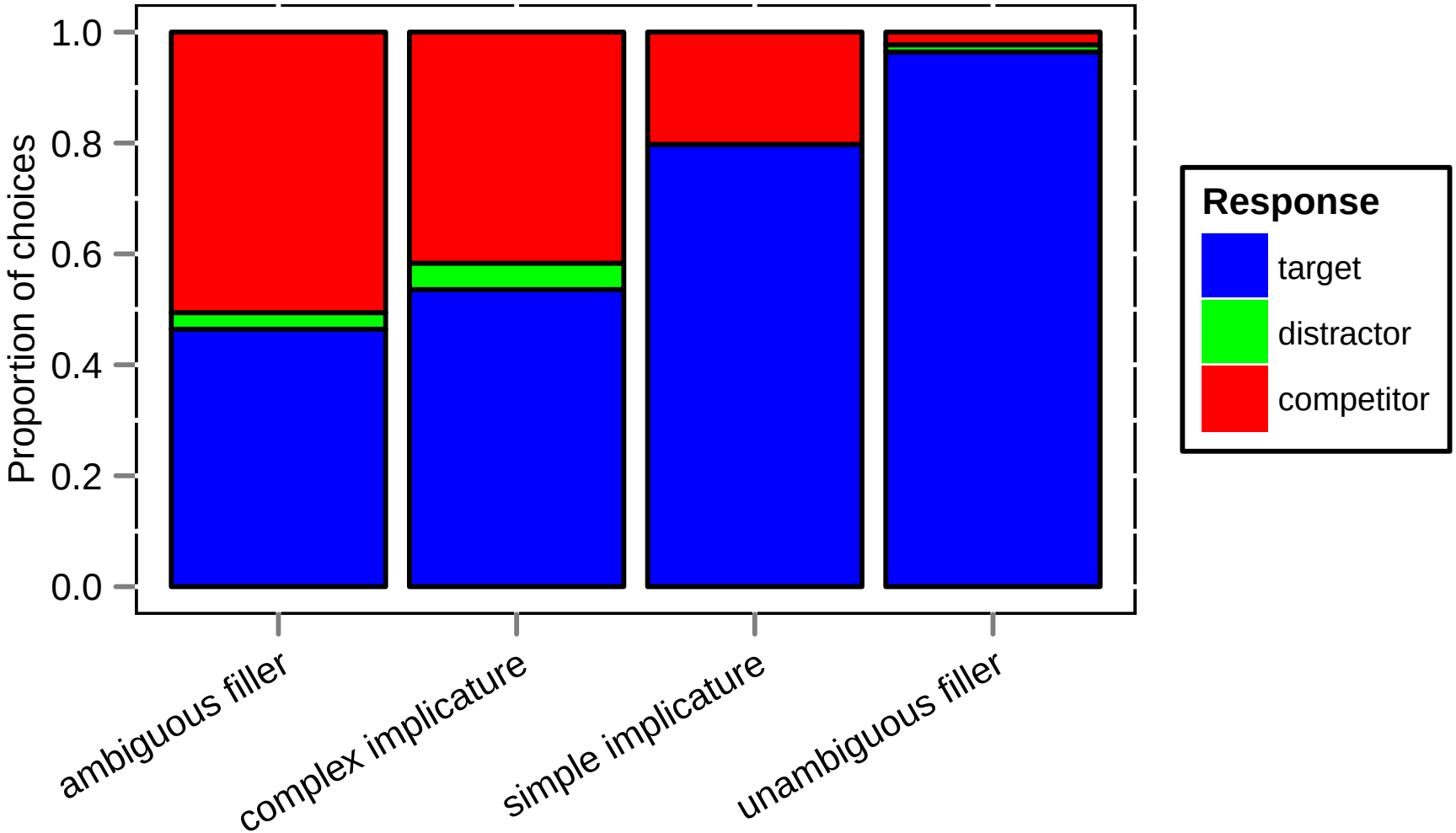
Results - proportion of responses by condition



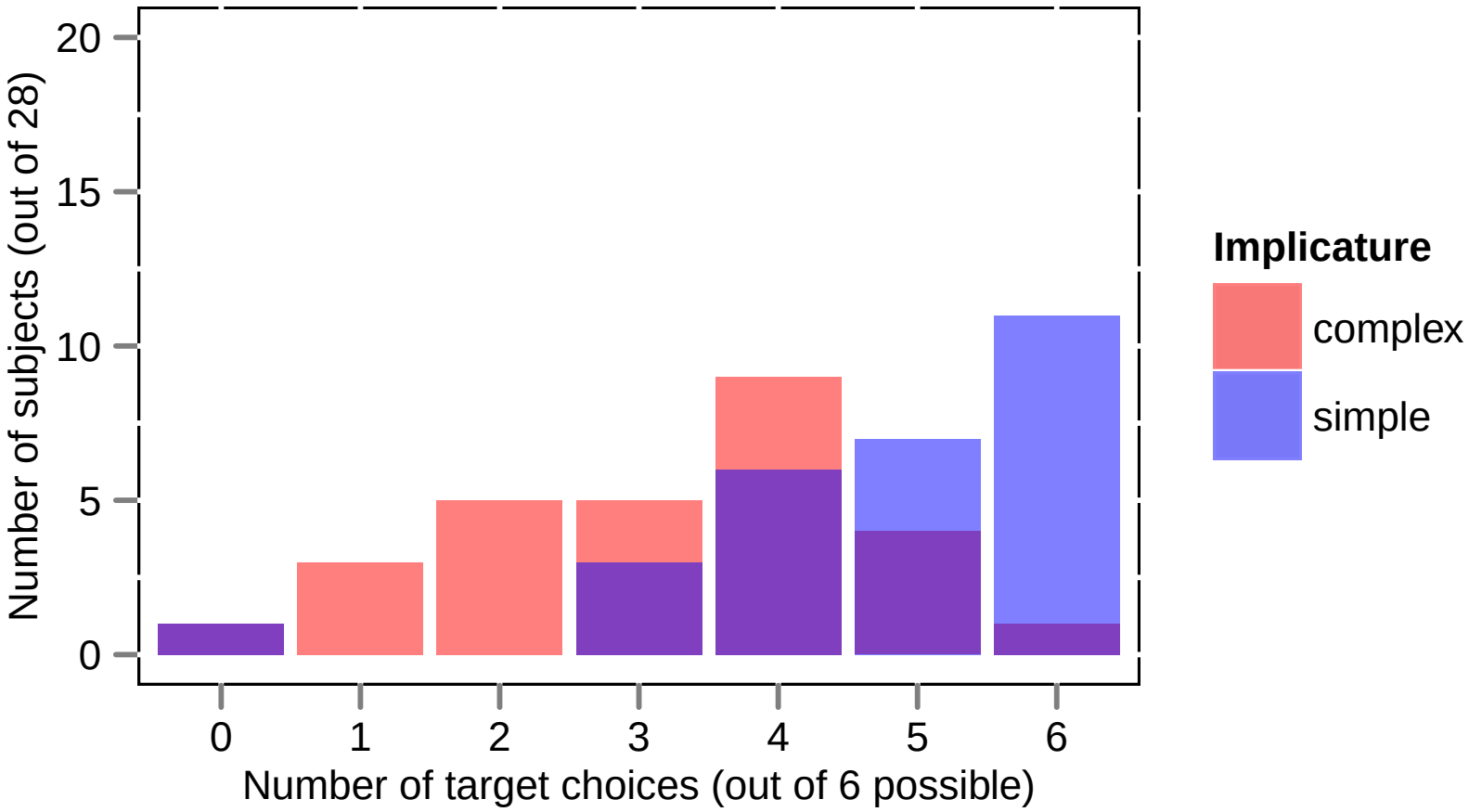
Results - proportion of responses by condition



Results - proportion of responses by condition

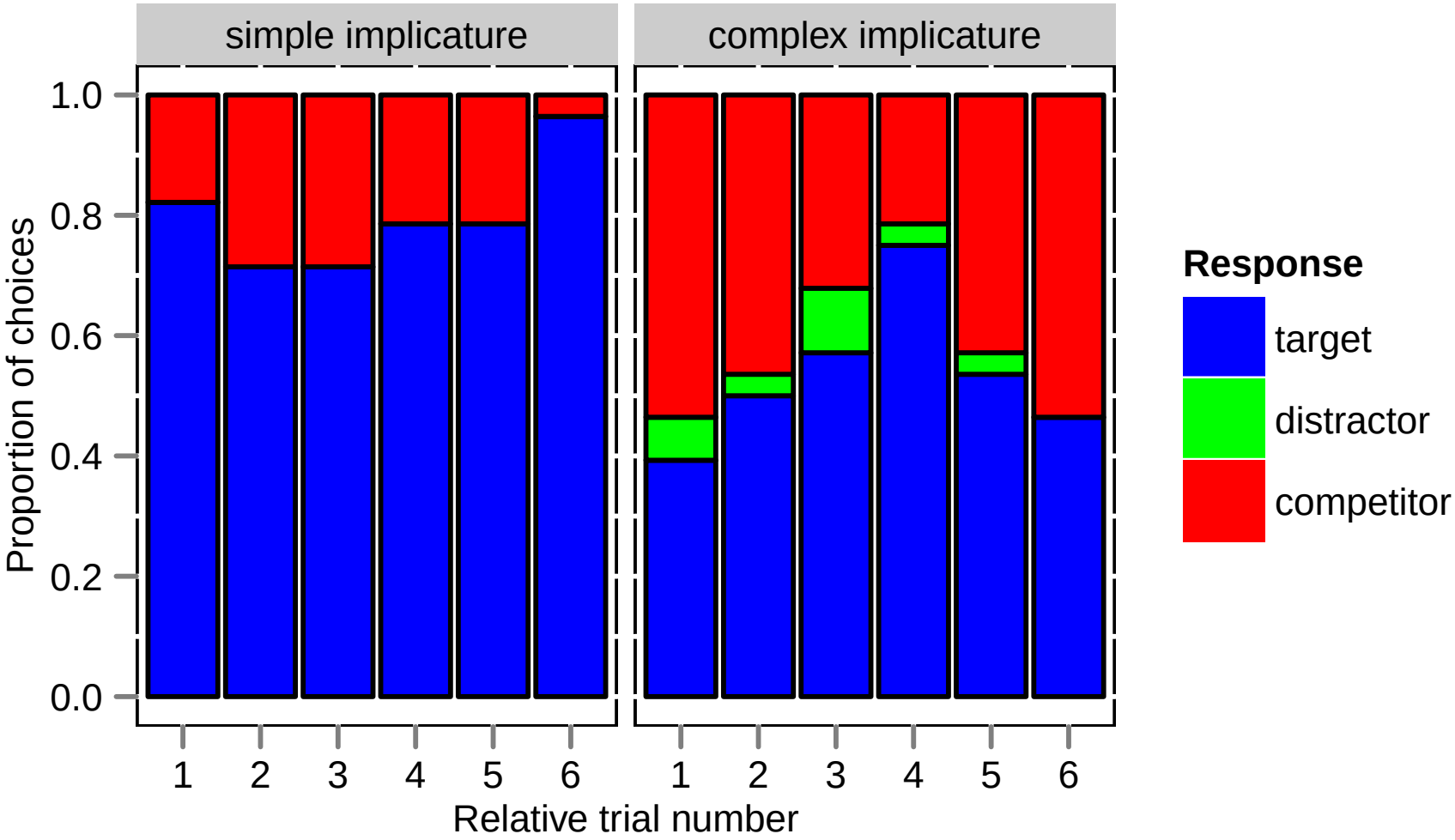


Results - distribution of subjects over target choices



→ not predicted by standard IBR

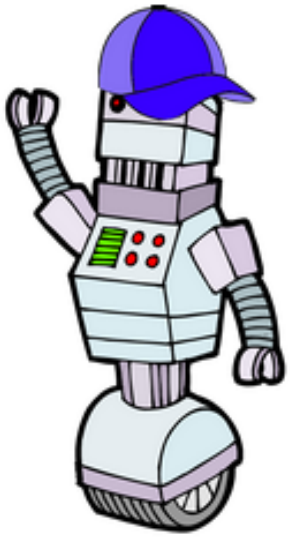
Results - learning effects

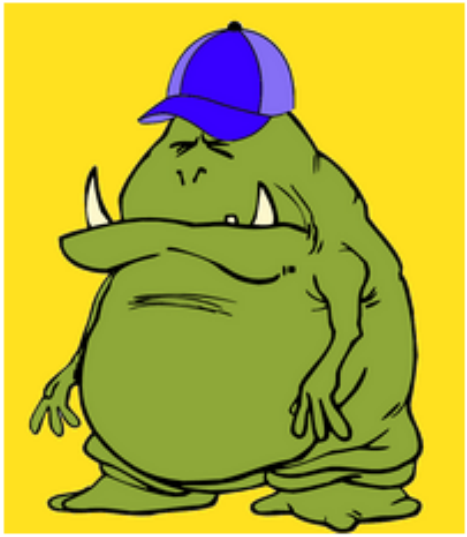


Experiment 2 - production

- test participants' behavior in the analogous production task
- 30 participants on Amazon's Mechanical Turk
- 36 experimental trials
  - 6 *simple* (one-step) implicature trials
  - 6 *complex* (two-step) implicature trials
  - 24 filler trials (entirely unambiguous/ entirely ambiguous target)

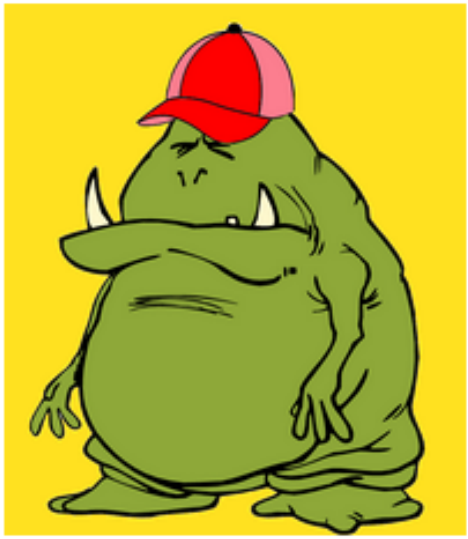
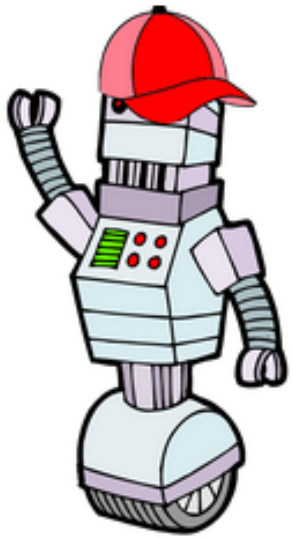
Simple implicature trial





Your task is to get another worker to pick out the highlighted creature. It's not highlighted on their display. **Click on one of the following four messages to send it to the other worker and get them to pick out the right creature.** The other worker knows you can only send these messages.

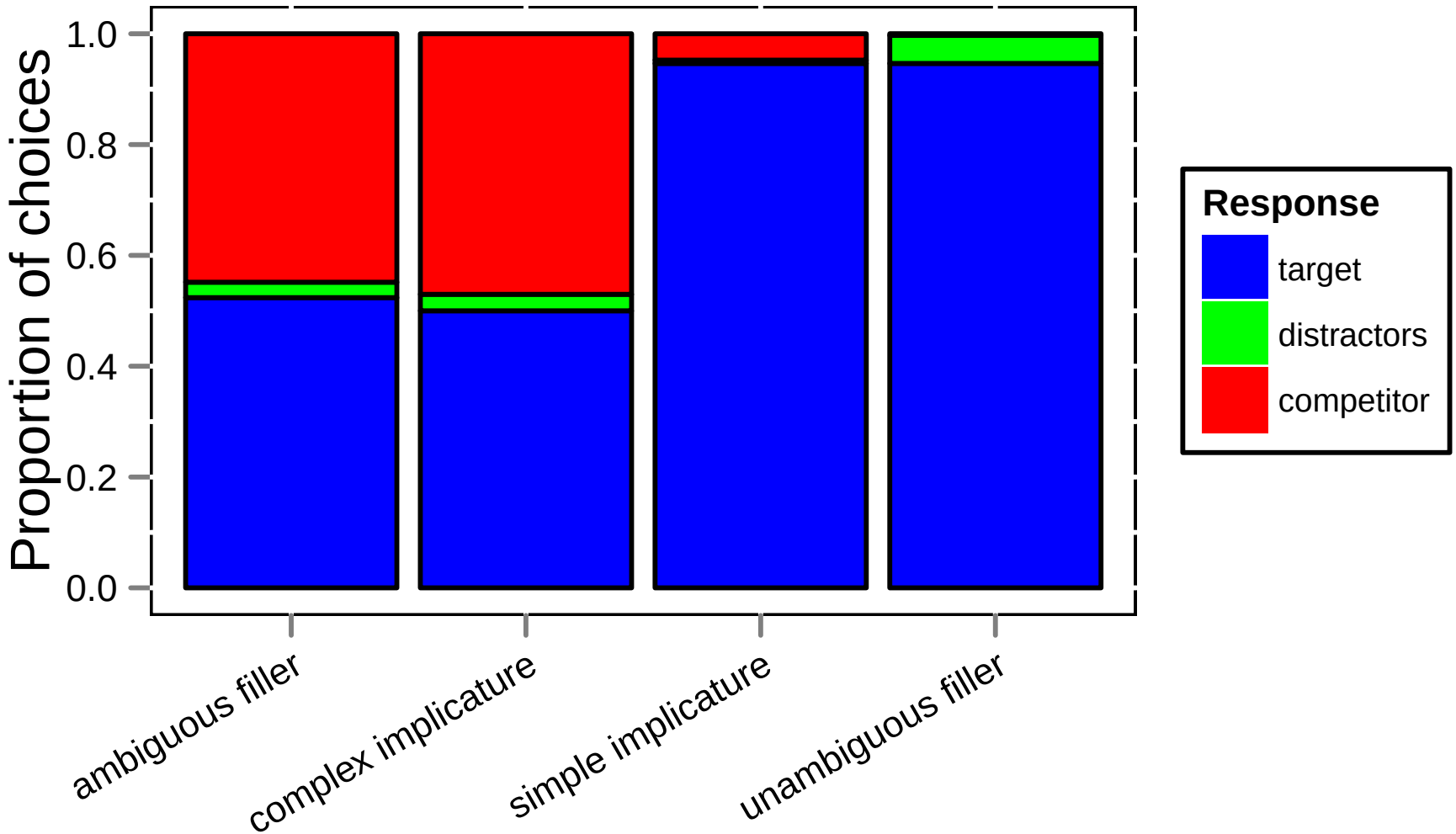
Complex implicature trial


Your task is to get another worker to pick out the highlighted creature. It's not highlighted on their display. **Click on one of the following four messages to send it to the other worker and get them to pick out the right creature.** The other worker knows you can only send these messages.

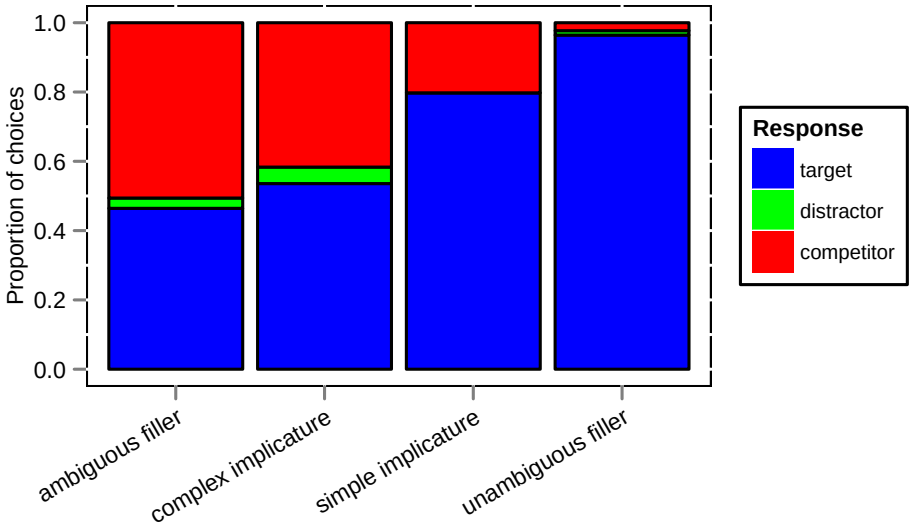



Results - proportion of responses by condition

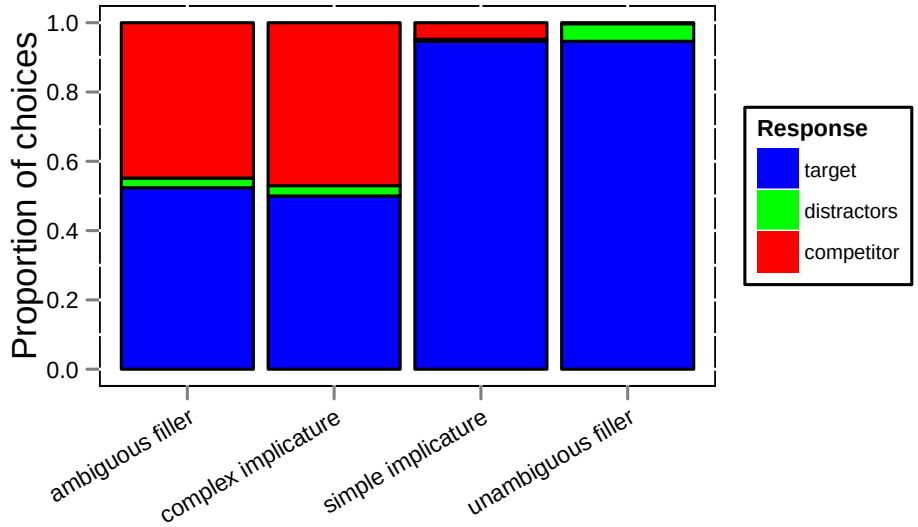


Results - proportion of responses by condition

Experiment 1
(comprehension)

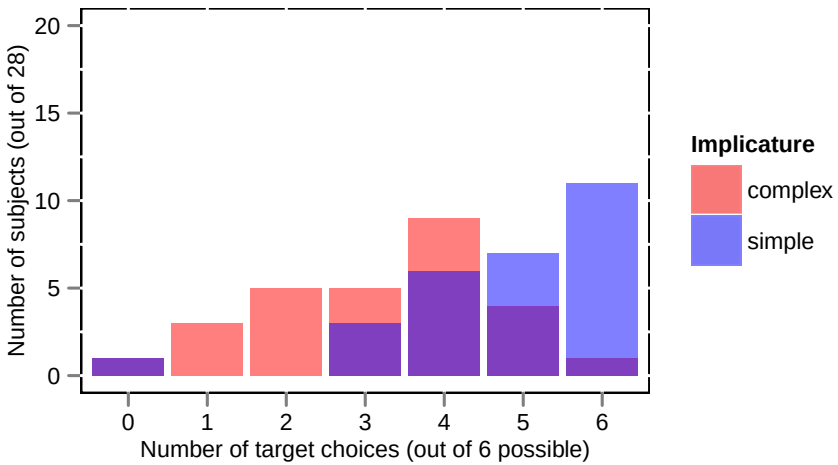


Experiment 2
(production)

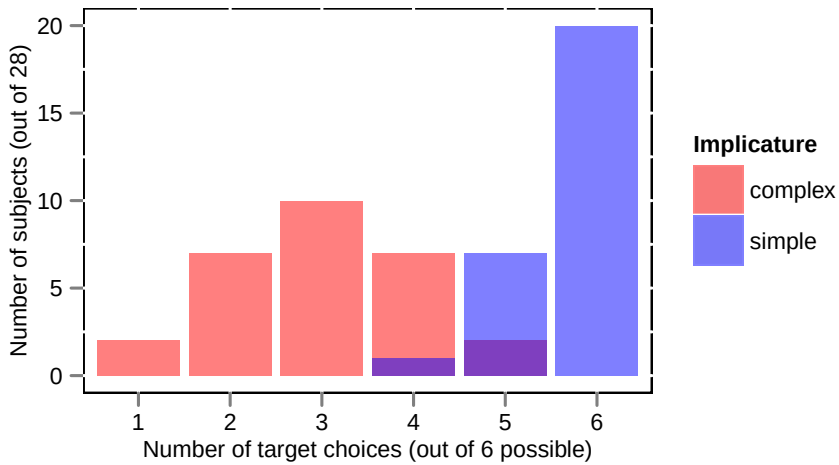


Results - distribution of subjects over target choices

Experiment 1
(comprehension)



Experiment 2
(production)



Interim summary

- asymmetry in production and comprehension: simple implicatures easier in production than comprehension and vice versa for complex implicatures
- not predicted by standard IBR

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

$$P(a_i) \propto \exp(\lambda u_i)$$

Stochastic choice

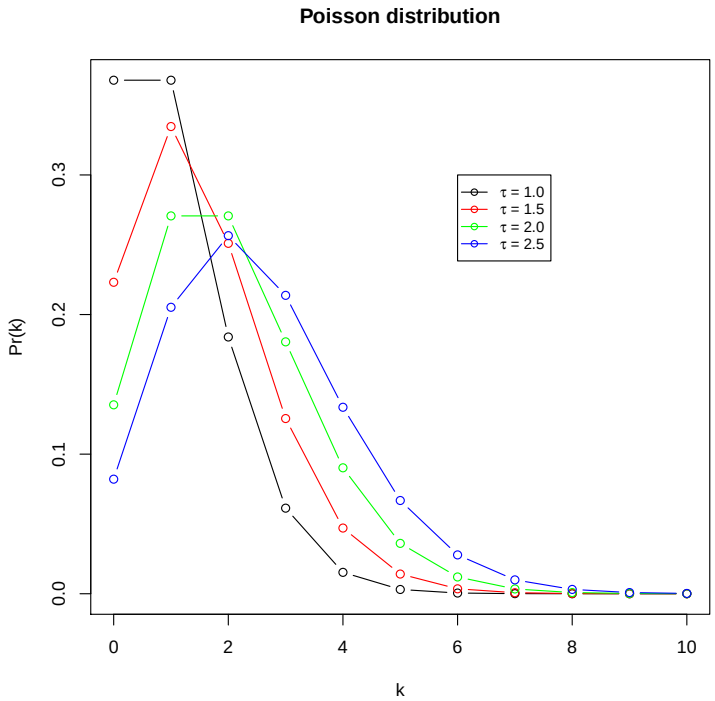
- λ measures degree of rationality
- $\lambda = 0$:
  - completely irrational behavior
  - all actions are equally likely, regardless of expected utility
- $\lambda \to \infty$
  - convergence towards behavior of rational choice
  - probability mass of sub-optimal actions converges to 0

Level-k thinking

- every player:
  - performs iterated best 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**

$$P(k) \propto \frac{\tau^k}{k!}$$

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



Fitting the data

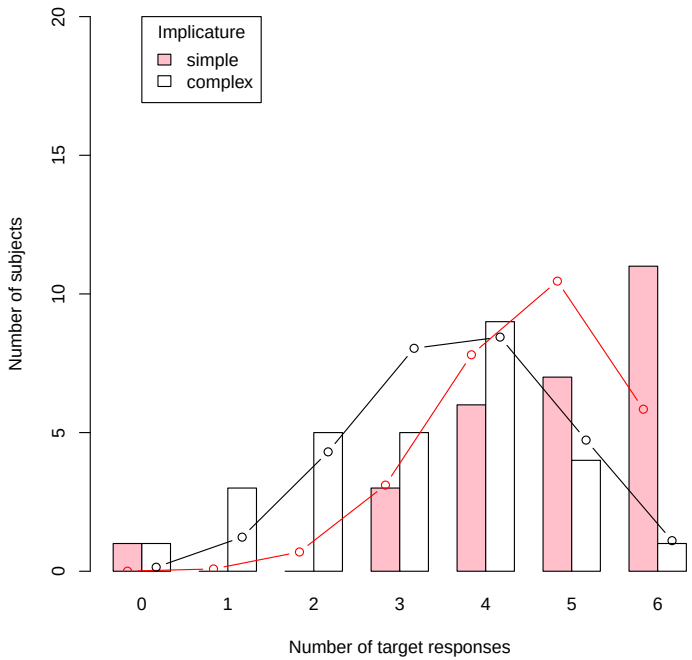
- maximum likelihood estimation of λ and τ on the basis of our experiments:

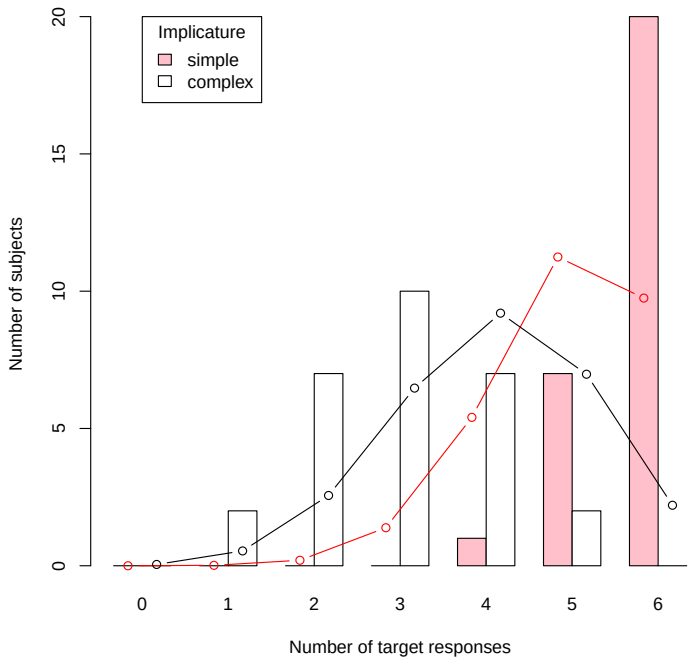
Experiment 1 (comprehension):

- $\lambda_1 = 6.33$
- $\tau_1 = 0.87$

Experiment 2 (production):

- $\lambda_2 = 6.52$
- $\tau_2 = 1.25$

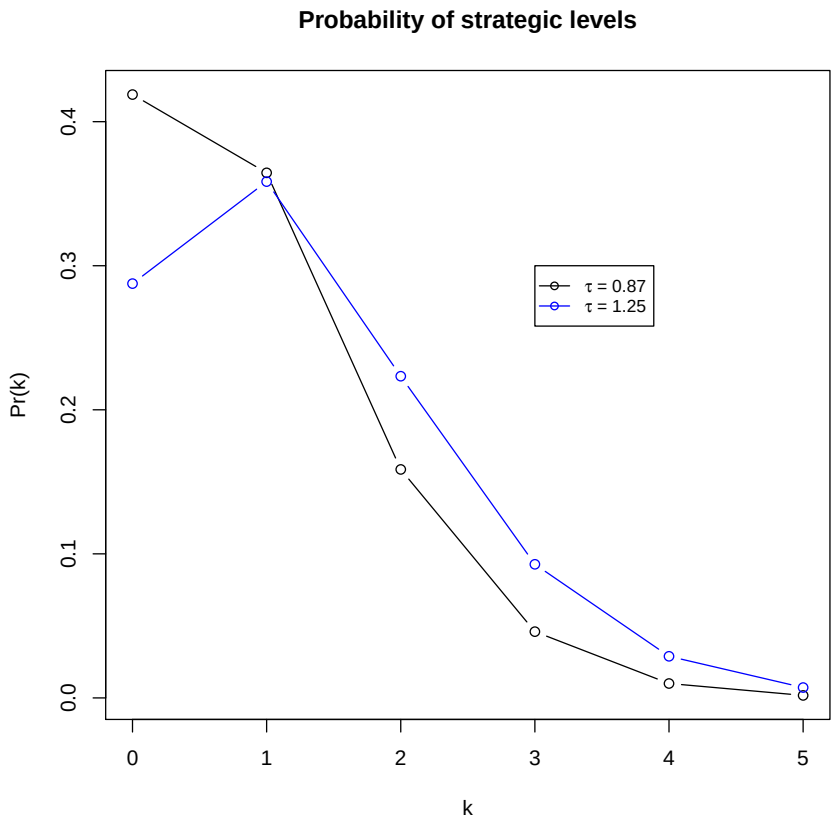




Tentative interpretation

- production/comprehension asymmetry:

Speakers are more strategic than listeners!



Alternative hypothesis

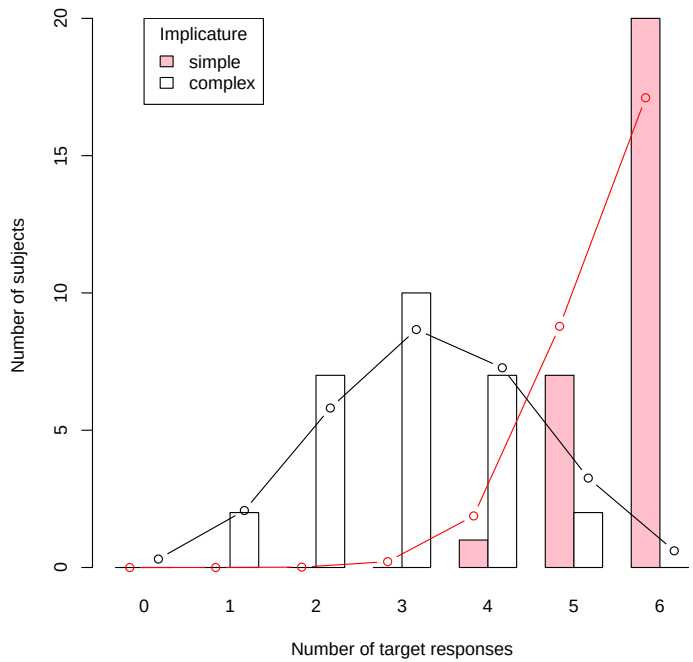
- This model took it for granted that non-strategic senders simply pick a true message at random.
- Results of experiment 2 suggest that this is not true; virtually everybody chooses the message that is most **informative**.
- Alternative hypothesis: S_0 uses the following utility function:

$$u_{S_0}(m|t) = \begin{cases} \frac{1}{|\{t'|t' \in [\![m]\!] \}|} & \text{if } t \in [\![m]\!] \\ 0 & \text{else} \end{cases}$$

Fitting the data, # 2

Experiment 2 (production):

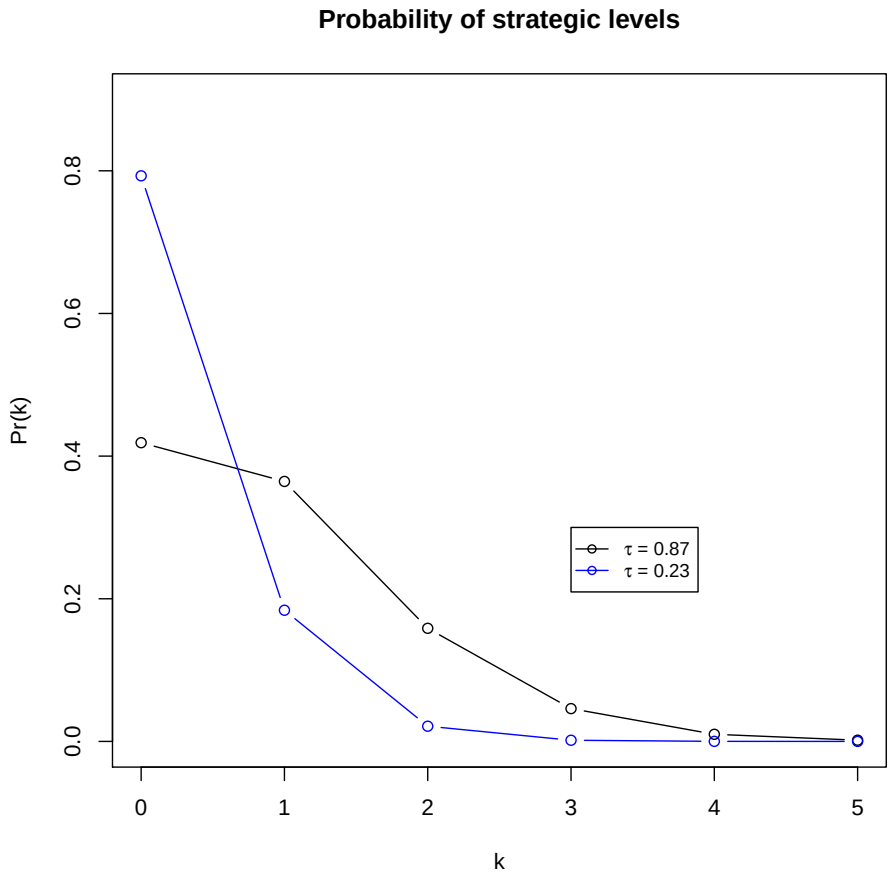
- $\lambda_2' = 5.35$
- $\tau_2' = 0.23$



Tentative interpretation # 2

- production/comprehension asymmetry:

Speakers barely reason at all, they just have a useful heuristics!



Conclusion

- interlocutors do take perspective and simulate each others' beliefs
  - but not always optimally
  - and less so as the number of required reasoning steps increases

Future directions

- utility manipulation
- message cost manipulation - moving into the realm of actual language
- interactive experiments with feedback $\overset{?}{\rightsquigarrow}$ learning

Thanks to

- EURO-XPRAG
- Tanenhaus lab
- Mike Tanenhaus & the NIH
- Florian Jaeger

References I

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008, June). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*(3), 1122–34.

Camerer, C., Ho, T., & Chong, K. (2002). *Behavioral game theory: Thinking, learning and teaching* (Tech. Rep.). Pasadena: CalTech.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton: Princeton University Press.

Clark, H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. LeNy & W. Kintsch (Eds.), *Language and comprehension.* Amsterdam: North-Holland.

Hanna, J., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43-61.

References II

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition*, *108*, 831-836.

Horton, W., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Keysar, B., Barr, D. J., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32-37.