The Iterated Best Response Model of game theoretic pragmatics and its relatives

Gerhard Jäger
joint work with Judith Degen and Michael Franke
gerhard.jaeger@uni-tuebingen.de

March 16, 2013

Beckman Center, Irvine

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

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).

Tea or coffee?

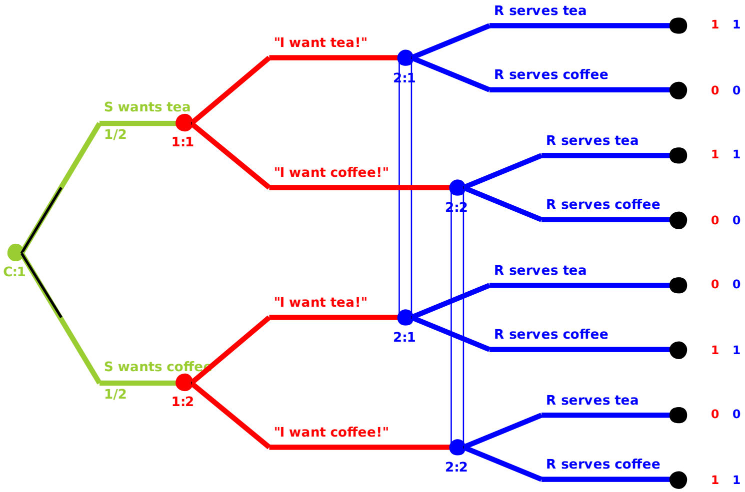
An example

- Sally either prefers tea (w_1) or coffee (w_2), with $p^*(w_1) = p^*(w_2) = 1/2$.
- Robin either serves tea (a_1) or coffee (a_2).
- Sally can send either of two messages:
  - m_1: *I prefer tea.*
  - m_2: *I prefer coffee.*
- Both messages are costless.

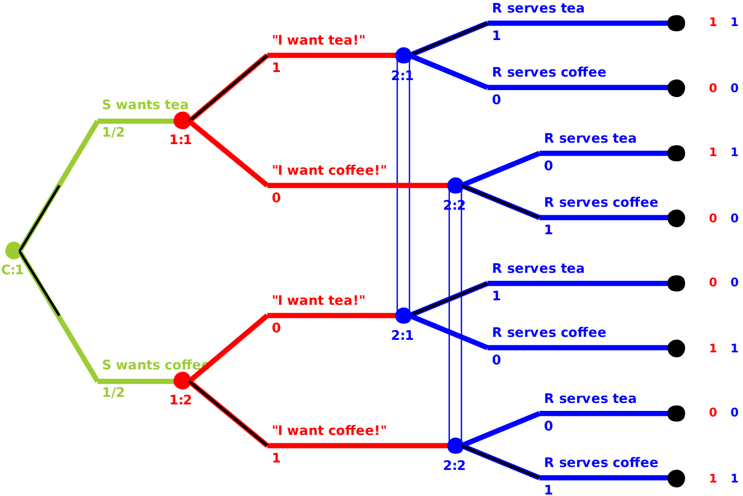
	a_1	a_2
w_1	$1, 1$	$0, 0$
w_2	$0, 0$	$1, 1$

Table: utility matrix

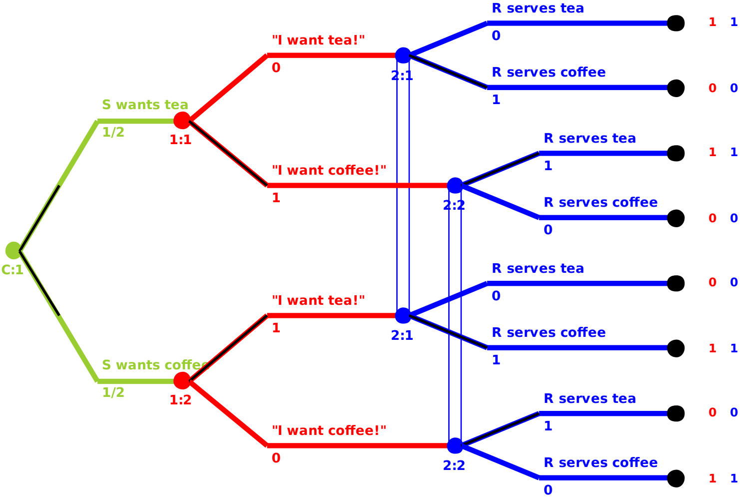
Extensive form



Extensive form



Extensive form



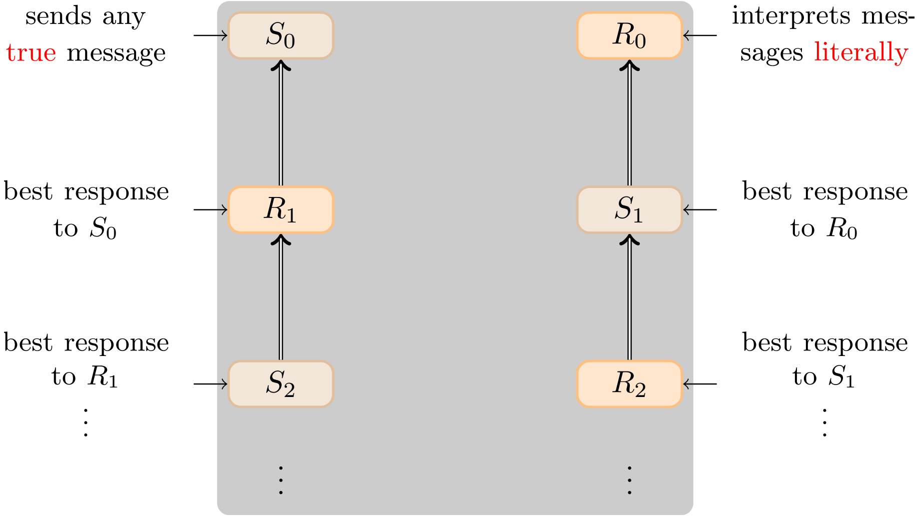
A coordination problem

- two strict Nash equilibria
  - S always says the truth and R always believes her.
  - S always says the opposite of the truth and R interprets everything ironically.
- Both equilibria are equally rational.
- Still, first equilibrium is more reasonable because it employs exogenous meanings of messages for equilibrium selection.
- Criterion for equilibrium selection:

> *As a default, S and R use/interpret signals according to their literal meaning. They only deviate from this if there self-interest dictates them to do so.*

- What exactly does this mean?

The Iterated Best Response sequence



Interpretation games

- How does this relate to linguistic examples?
- There is a quasi-algorithmic procedure (due to Franke 2009) how to construct a game from an example sentence.

What is given?

- example sentence
- set of expression alternatives
- jointly form set of messages
- question under discussion QUD
- set of complete answers to QUD is the set of possible worlds

What do we need?

- interpretation function $\|\cdot\|$
- prior probability distribution p^*
- set of actions
- utility functions

Interpretation games

QUD

- often QUD is not given explicitly
- procedure to construct QUD from expression m and its alternatives $ALT(m)$:
  - Let ct be the context of utterances, i.e. the maximal set of statements that is common knowledge between Sally and Robin.
  - any subset w of $ALT(m) \cup \{\neg m' | m' \in ALT(m)\}$ is a possible world iff
    - w and ct are consistent, i.e. $w \cup ct \not\vdash \bot$
    - for any set $X : w \subset X \subseteq ALT(m) \cup \{\neg m' | m' \in ALT(m)\}$, $ct \cup X$ is inconsistent

Interpretation games

Game construction

- interpretation function:

$$\|m'\| = \{w|w \vdash m\}$$

- p^* is uniform distribution over W
- justified by principle of insufficient reason
- set of actions is W
- intuitive idea: Robin's task is to figure out which world Sally is in
- utility functions:

$$u_{s/r}(w,a) = \begin{cases} 1 & \text{iff } w = a \\ 0 & \text{else} \end{cases}$$

- both players want Robin to succeed

Quantity implicatures

(1)
 a. Who came to the party?
 b. SOME: Some boys came to the party.
 c. NO: No boys came to the party.
 d. ALL: All boys came to the party.

Game construction

- $ct = \emptyset$
- $W = \{w_{\neg\exists}, w_{\exists\neg\forall}, w_{\forall}\}$
- $w_{\neg\exists} = \{\text{NO}\}, w_{\exists\neg\forall} = \{\text{SOME}\}, w_{\forall} = \{\text{SOME}, \text{ALL}\}$
- $p^* = (1/3, 1/3, 1/3)$

- interpretation function:

$$\begin{aligned} \|\text{SOME}\| &= \{w_{\exists\neg\forall}, w_{\forall}\} \\ \|\text{NO}\| &= \{w_{\neg\exists}\} \\ \|\text{ALL}\| &= \{w_{\forall}\} \end{aligned}$$

- utilities:

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

Interpretation games

- utility functions are identity matrices
- therefore the step *multiply with utility matrix* can be omitted in best response computation
- also, restriction to uniform priors makes simplifies computation of posterior distribution
- simplified IBR computation:

Interpretation games

Sally

1. flip ρ along diagonal
2. place a 0 in each cell that is non-maximal within its row
3. normalize each row

Robin

1. flip σ along diagonal
2. if a row contains only 0s, fill in a 1 in each cell corresponding to a true world-message association
3. place a 0 in each cell that is non-maximal within its row
4. normalize each row

Example: Quantity implicatures

σ_0	NO	SOME	ALL
$w_{\neg\exists}$	1	0	0
$w_{\exists\neg\forall}$	0	1	0
$w_{\forall}$	0	$1/2$	$1/2$

ρ_0	$w_{\neg\exists}$	$w_{\exists\neg\forall}$	$w_{\forall}$
NO	1	0	0
SOME	0	$1/2$	$1/2$
ALL	0	0	1

ρ_1	$w_{\neg\exists}$	$w_{\exists\neg\forall}$	$w_{\forall}$
NO	1	0	0
SOME	0	1	0
ALL	0	0	1

σ_1	NO	SOME	ALL
$w_{\neg\exists}$	1	0	0
$w_{\exists\neg\forall}$	0	1	0
$w_{\forall}$	0	0	1

$$F = (\rho_1, \sigma_1)$$

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

Lifted games

- So far, it is hard-wired in the model that Sally has complete knowledge (or, rather, complete belief — whether or not she is right is inessential for IBR) about the world she is in.
- corresponds to strong version of **competence assumption**
- Sometimes this assumption is too strong:

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.

Utility matrix

	a_a	a_b	a_{ab}
w_a	1	0	0
w_b	0	1	0
w_{ab}	0	0	1

Lifted games

IBR sequence

σ_0	OR	A	B	AND
w_a	$1/2$	$1/2$	0	0
w_b	$1/2$	0	$1/2$	0
w_{ab}	$1/4$	$1/4$	$1/4$	$1/4$

ρ_0	w_a	w_b	w_{ab}
OR	$1/3$	$1/3$	$1/3$
A	$1/2$	0	$1/2$
B	0	$1/2$	$1/2$
AND	0	0	1

ρ_1	w_a	w_b	w_{ab}
OR	$1/2$	$1/2$	0
A	1	0	0
B	0	1	0
AND	0	0	1

σ_1	OR	A	B	AND
w_a	0	1	0	0
w_b	0	0	1	0
w_{ab}	0	0	0	1

Lifted games

IBR sequence

σ_2	OR	A	B	AND
w_a	0	1	0	0
w_b	0	0	1	0
w_{ab}	0	0	0	1

ρ_2	w_a	w_b	w_{ab}
OR	$1/3$	$1/3$	$1/3$
A	1	0	0
B	0	1	0
AND	0	0	1

OR comes out as a message that would never be used!

Lifted games

- full competence assumption is arguably too strong
- weaker assumption (Franke 2009):
  - Sally's information states are **partial answers to QUD**, ie. **sets** of possible worlds
  - Robin's task is to figure out which information state Sally is in.
  - *ceteris paribus*, Robin receives slightly higher utility for smaller (more informative) states

Costs

- Preferences that are independent from correct information transmission are captured via *cost functions* for sender and receiver.
- For the sender this might be, *inter alia*, a preference for simpler expressions.
- For the receiver, the *Strongest Meaning Hypothesis* is a good candiate.

Lifted games

Formally

- cost functions c_s, c_r: $c_s : (POW(W) - \{\emptyset\}) \times M \mapsto \mathbb{R}^+$
- costs are **nominal**:

$$0 \leq c_s(i,m), c_r(i,m) < \min(1/|POW(W)-\emptyset|^2, 1/|ALT(m)|^2)$$

- guarantees that cost considerations never get in the way of information transmission considerations
- new utility functions:

$$u_s(i,m,a) = -c_s(i,m) + \begin{cases} 1 & \text{if } i = a, \\ 0 & \text{else,} \end{cases}$$

$$u_r(i,m,a) = -c_r(a,m) + \begin{cases} 1 & \text{if } i = a, \\ 0 & \text{else.} \end{cases}$$

Modified IBR procecure

Sally

- flip ρ along the diagonal
- subtract c_s
- place a 0 in each cell that is non-maximal within its row
- normalize each row

Robin

- flip σ along diagonal
- if a row contains only 0s,
  - fill in a 1 in each cell corresponding to a true world-message association
- else
  - subtract c_r^T
- place a 0 in each cell that is non-maximal within its row
- normalize each row

The Strongest Meaning Hypothesis

- if in doubt, Robin will assume that Sally is competent
- captured in following cost function:

$$c_r(a, m) = {|a|}/{\max(|M|, 2^{|W|})^2}$$

$$c_r(\{w_a\}, \cdot) = 1/49 \qquad c_r(\{w_a, w_{ab}\}, \cdot) = 2/49$$

$$c_r(\{w_b\}, \cdot) = 1/49 \qquad c_r(\{w_b, w_{ab}\}, \cdot) = 2/49$$

$$c_r(\{w_{ab}\}, \cdot) = 1/49 \qquad c_r(\{w_a, w_b, w_{ab}\}, \cdot) = 3/49$$

$$c_r(\{w_a, w_b\}, \cdot) = 2/49$$

Lifted games

IBR sequence: 1

σ_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: flipping and subtracting costs

	$\{w_a\}$	$\{w_b\}$	$\{w_{ab}\}$	$\{w_a, w_b\}$	$\{w_a, w_{ab}\}$	$\{w_b, w_{ab}\}$	$\{w_a, w_b, w_{ab}\}$
OR	0.48	0.48	0.23	**0.96**	0.46	0.46	0.94
A	**0.48**	−0.02	0.23	−0.04	0.46	−0.04	−0.06
B	−0.02	**0.48**	0.23	−0.04	−0.04	0.46	−0.06
AND	−0.02	−0.02	**0.23**	−0.04	−0.04	−0.04	−0.06

Lifted games

IBR sequence: 2

ρ_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

σ_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

More ignorance implicatures

2. a. Ann or Bert or both showed up. (= AB-OR)
   b. Ann showed up. (= A)
   c. Bert showed up. (= B)
   d. Ann and Bert showed up. (= AND)
   e. Ann or Bert showed up. (= OR)
   f. Ann or both showed up. (= A-OR)
   g. Bert or both showed up. (= B-OR)

- Message (e) is arguably more efficient for Sally than (a)
- Let us say that $c_s(\cdot, \text{AB-OR}) = 0.006, c_s(\cdot, \text{A-OR}) = c_s(\cdot, \text{B-OR}) = 0.004, c_s(\cdot, \text{OR}) = c_s(\cdot, \text{AND}) = 0.003$, and $c_s(\cdot, \text{A}) = c_s(\cdot, \text{B}) = 0$.

More ignorance implicatures

IBR sequence: 1

σ_0	AB-OR	A	B	AND	OR	A-OR	B-OR
$\{w_a\}$	$1/4$	$1/4$	0	0	$1/4$	$1/4$	0
$\{w_b\}$	$1/4$	0	$1/4$	0	$1/4$	0	$1/4$
$\{w_{ab}\}$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$
$\{w_a, w_b\}$	$1/2$	0	0	0	$1/2$	0	0
$\{w_a, w_{ab}\}$	$1/4$	$1/4$	0	0	$1/4$	$1/4$	0
$\{w_b, w_{ab}\}$	$1/4$	0	$1/4$	0	$1/4$	0	$1/4$
$\{w_a, w_b, w_{ab}\}$	$1/2$	0	0	0	$1/2$	0	0

More ignorance implicatures

IBR sequence: 1

ρ_1	$\{w_a\}$	$\{w_b\}$	$\{w_{ab}\}$	$\{w_a, w_b\}$	$\{w_a, w_{ab}\}$	$\{w_b, w_{ab}\}$	$\{w_a, w_b, w_{ab}\}$
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
OR	0	0	0	1	0	0	0
A-OR	1	0	0	0	0	0	0
B-OR	0	1	0	0	0	0	0

More ignorance implicatures

IBR sequence: 2

σ_2	AB-OR	A	B	AND	OR	A-OR	B-OR
$\{w_a\}$	0	1	0	0	0	0	0
$\{w_b\}$	0	0	1	0	0	0	0
$\{w_{ab}\}$	0	0	0	1	0	0	0
$\{w_a, w_b\}$	0	0	0	0	1	0	0
$\{w_a, w_{ab}\}$	0	1	0	0	0	0	0
$\{w_b, w_{ab}\}$	0	0	1	0	0	0	0
$\{w_a, w_b, w_{ab}\}$	0	0	0	0	1	0	0

More ignorance implicatures

IBR sequence: 2

ρ_2	$\{w_a\}$	$\{w_b\}$	$\{w_{ab}\}$	$\{w_a, w_b\}$	$\{w_a, w_{ab}\}$	$\{w_b, w_{ab}\}$	$\{w_a, w_b, w_{ab}\}$
AB-OR	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$	$1/7$
A	1	0	0	0	0	0	0
B	0	1	0	0	0	0	0
AND	0	0	1	0	0	0	0
OR	0	0	0	1	0	0	0
A-OR	$1/3$	0	$1/3$	0	$1/3$	0	0
B-OR	0	$1/3$	$1/3$	0	0	$1/3$	0

More ignorance implicatures

IBR sequence: 3

σ_3	AB-OR	A	B	AND	OR	A-OR	B-OR
$\{w_a\}$	0	1	0	0	0	0	0
$\{w_b\}$	0	0	1	0	0	0	0
$\{w_{ab}\}$	0	0	0	1	0	0	0
$\{w_a, w_b\}$	0	0	0	0	1	0	0
$\{w_a, w_{ab}\}$	0	0	0	0	0	1	0
$\{w_b, w_{ab}\}$	0	0	0	0	0	0	1
$\{w_a, w_b, w_{ab}\}$	1	0	0	0	0	0	0

More ignorance implicatures

IBR sequence: 3

ρ_4	$\{w_a\}$	$\{w_b\}$	$\{w_{ab}\}$	$\{w_a, w_b\}$	$\{w_a, w_{ab}\}$	$\{w_b, w_{ab}\}$	$\{w_a, w_b, w_{ab}\}$
AB-OR	0	0	0	0	0	0	1
A	1	0	0	0	0	0	0
B	0	1	0	0	0	0	0
AND	0	0	1	0	0	0	0
OR	0	0	0	1	0	0	0
A-OR	0	0	0	0	1	0	0
B-OR	0	0	0	0	0	1	0

Embedded implicatures

3. **a.** Kai had broccoli or some of the peas. $(\mathrm{B} \vee \exists x \mathrm{P} x)$
   b. Kai had broccoli or all of the peas. $(\mathrm{B} \vee \forall x \mathrm{P} x)$

Alternatives:

4. **a.** Kai had broccoli. $(= \mathrm{B})$
   b. Kai had some of the peas. $(= \exists x \mathrm{P} x)$
   c. Kai had all of the peas. $(= \forall x \mathrm{P} x)$
   d. Kai had broccoli and some of the peas. $(= \mathrm{B} \wedge \exists x \mathrm{P} x)$
   e. Kai had broccoli and all of the peas. $(= \mathrm{B} \wedge \forall x \mathrm{P} x)$

- Messages (2a,b) and (3d,e) ar arguably more costly for Sally than the simple ones
- let us say that complex messages incur a cost of 0.001 for Sally

Embedded implicatures

Possible worlds:

- $w_{B\neg\exists} = \{\mathrm{B}, \mathrm{B} \vee \exists x \mathrm{P}x, \mathrm{B} \vee \forall x \mathrm{P}x\}$,
- $w_{\neg B\exists\neg\forall} = \{\exists x \mathrm{P}x, \mathrm{B} \vee \exists x \mathrm{P}x\}$,
- $w_{\neg B\forall} = \{\exists x \mathrm{P}x, \forall x \mathrm{P}x, \mathrm{B} \vee \exists x \mathrm{P}x, \mathrm{B} \vee \forall x \mathrm{P}x\}$,
- $w_{B\exists\neg\forall} = \{\mathrm{B}, \exists x \mathrm{P}x, \mathrm{B} \vee \exists x \mathrm{P}x, \mathrm{B} \vee \forall x \mathrm{P}x, \mathrm{B} \wedge \exists x \mathrm{P}x\}$,
- $w_{B\forall} = \{\mathrm{B}, \exists x \mathrm{P}x, \mathrm{B} \vee \exists x \mathrm{P}x, \mathrm{B} \vee \forall x \mathrm{P}x, \mathrm{B} \wedge \exists x \mathrm{P}x, \mathrm{B} \wedge \forall x \mathrm{P}x\}$.

Embedded implicatures

σ_0	B	$\exists x \mathrm{P} x$	$\forall x \mathrm{P} x$	$\mathrm{B} \vee \exists x \mathrm{P} x$	$\mathrm{B} \wedge \exists x \mathrm{P} x$	$\mathrm{B} \vee \forall x \mathrm{P} x$	$\mathrm{B} \wedge \forall x \mathrm{P} x$
$\{w_{B\neg\exists}\}$	$^1/_3$	0	0	$^1/_3$	0	$^1/_3$	0
$\{w_{\neg B\exists\neg\forall}\}$	0	$^1/_2$	0	$^1/_2$	0	0	0
$\{w_{\neg B\forall}\}$	0	$^1/_4$	$^1/_4$	$^1/_4$	0	$^1/_4$	0
$\{w_{B\exists\neg\forall}\}$	$^1/_5$	$^1/_5$	0	$^1/_5$	$^1/_5$	$^1/_5$	0
$\{w_{B\forall}\}$	$^1/_7$	$^1/_7$	$^1/_7$	$^1/_7$	$^1/_7$	$^1/_7$	$^1/_7$
$\{w_{B\neg\exists}, w_{\neg B\exists\neg\forall}\}$	0	0	0	1	0	0	0
$\{w_{B\neg\exists}, w_{\neg B\forall}\}$	0	0	0	$^1/_2$	0	$^1/_2$	0

Embedded implicatures

ρ_1	$\{w_{B\neg\exists}\}$	$\{w_{\neg B\exists\neg\forall}\}$	$\{w_{\neg B\forall}\}$	$\{w_{B\exists\neg\forall}\}$	$\{w_{B\forall}\}$	$\{w_{B\neg\exists}, w_{\neg B\exists\neg\forall}\}$	$\{w_{B\neg\exists}, w_{\neg B\forall}\}$
B	1	0	0	0	0	0	0
$\exists x\mathrm{P}x$	0	1	0	0	0	0	0
$\forall x\mathrm{P}x$	0	0	1	0	0	0	0
$\mathrm{B} \vee \exists x\mathrm{P}x$	0	0	0	0	0	1	0
$\mathrm{B} \wedge \exists x\mathrm{P}x$	0	0	0	1	0	0	0
$\mathrm{B} \vee \forall x\mathrm{P}x$	0	0	0	0	0	0	1
$\mathrm{B} \wedge \forall x\mathrm{P}x$	0	0	0	0	1	0	0

Embedded implicatures

σ_2	B	$\exists x\mathrm{P}x$	$\forall x\mathrm{P}x$	$\mathrm{B}\vee\exists x\mathrm{P}x$	$\mathrm{B}\wedge\exists x\mathrm{P}x$	$\mathrm{B}\vee\forall x\mathrm{P}x$	$\mathrm{B}\wedge\forall x\mathrm{P}x$
$\{w_{B\neg\exists}\}$	1	0	0	0	0	0	0
$\{w_{\neg B\exists\neg\forall}\}$	0	1	0	0	0	0	0
$\{w_{\neg B\forall}\}$	0	0	1	0	0	0	0
$\{w_{B\exists\neg\forall}\}$	0	0	0	0	1	0	0
$\{w_{B\forall}\}$	0	0	0	0	0	0	1
$\{w_{B\neg\exists}, w_{\neg B\exists\neg\forall}\}$	0	0	0	1	0	0	0
$\{w_{B\neg\exists}, w_{\neg B\forall}\}$	0	0	0	0	0	1	0

- (σ_2, ρ_1) form fixed point
- critical example is interpreted as *Kay had broccoli and no peas, or he had broccoli and some but not all of the peas, but not both.*

Measure terms

Krifka (2002,2007) notes that measure terms can be used in a precise or in a vague way, and that more complex expressions are less likely to be used in a vague way. Here is a schematic analysis:

- w_1, w_3: 100 meter, w_2, w_4: 101 meter
- m_{100}: "one hundred meter"
  m_{101}: "one hundred and one meter"
  m_{ex100}: "exactly one hundred meter"
- $\|m_{100}\| = \|m_{ex100}\| = \{w_1, w_3\}$,
  $\|m_{101}\| = \{w_2, w_4\}$
- $c(m_{100}) = 0$,
  $c(m_{101}) = c(m_{ex100}) = 0.15$
- a_1, a_3: 100, a_2, a_4: 101

- in w_1, w_2 precision is important
- in w_3, w_4 precision is not important

	a_1	a_2	a_3	a_4
w_1	1	0.5	1	0.5
w_2	0.5	1	0.5	1
w_3	1	0.9	1	0.9
w_4	0.9	1	0.9	1

Measure terms

σ_0	m_{100}	m_{101}	m_{ex100}
w_1	$1/2$	0	$1/2$
w_2	0	1	0
w_3	$1/2$	0	$1/2$
w_4	0	1	0

ρ_1	a_1	a_2	a_3	a_4
m_{100}	$1/2$	0	$1/2$	0
m_{101}	0	$1/2$	0	$1/2$
m_{ex100}	$1/2$	0	$1/2$	0

σ_2	m_{100}	m_{101}	m_{ex100}
w_1	1	0	0
w_2	0	1	0
w_3	1	0	0
w_4	1	0	0

ρ_3	a_1	a_2	a_3	a_4
m_{100}	$1/3$	0	$1/3$	$1/3$
m_{101}	0	1	0	0
m_{ex100}	$1/2$	0	$1/2$	0

σ_4	m_{100}	m_{101}	m_{ex100}
w_1	0	0	1
w_2	0	1	0
w_3	1	0	0
w_4	1	0	0

ρ_5	a_1	wa_2	a_3	a_4
m_{100}	0	0	$1/2$	$1/2$
m_{101}	0	1	0	0
m_{ex100}	1	0	0	0

Conflicting interests

Poker

5. Do you have the ace of hearts?

- m_1: *Yes.*
- m_2: *No.*

	$a_{\heartsuit}$	$a_{\not\heartsuit}$
$w_{\heartsuit}$	$0, 1$	$1, 0$
$w_{\not\heartsuit}$	$1, 0$	$0, 1$

Table: utility matrix

Conflicting interests

σ_0	YES	NO
$w_{\heartsuit}$	1	0
$w_{\not\heartsuit}$	0	1

ρ_0	$a_{\heartsuit}$	$a_{\not\heartsuit}$
YES	1	0
NO	0	1

σ_1	YES	NO
$w_{\heartsuit}$	0	1
$w_{\not\heartsuit}$	1	0

ρ_1	$a_{\heartsuit}$	$a_{\not\heartsuit}$
YES	0	1
NO	1	0

$\vdots$ $\vdots$

No fixed point; no stable information transmission.

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

$$P(a_i) = \frac{\exp(\lambda u_i)}{\sum_j (\lambda \exp u_j)}$$

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
- if everybody plays a quantal response (for fixed λ), play is in **quantal response equilibrium** (QRE)
- as $\lambda \to \infty$, QREs converge towards Nash equilibria

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

Some/All-game

$$\lambda = 2.0$$

```
[1] "k=0"
[[1]]
  some all
e  1.0 0.0
a  0.5 0.5

[[2]]
       e   a
some 0.5 0.5
all  0.0 1.0

[1] "k=1"
[[1]]
       some       all
e 0.7310586 0.2689414
a 0.2689414 0.7310586

[[2]]
             e         a
some 0.6607564 0.3392436
all  0.1192029 0.8807971
```

```
[1] "k=2"
[[1]]
       some       all
e 0.7470815 0.2529185
a 0.2529185 0.7470815

[[2]]
             e         a
some 0.7159041 0.2840959
all  0.2840959 0.7159041

[1] "k=3"
[[1]]
       some       all
e 0.7034157 0.2965843
a 0.2965843 0.7034157

[[2]]
             e         a
some 0.7287571 0.2712429
all  0.2712429 0.7287571
```

```
[1] "k=100"
[[1]]
       some       all
e 0.5598841 0.4401159
a 0.4401159 0.5598841

[[2]]
             e         a
some 0.5593575 0.4406425
all  0.4406425 0.5593575

[1] "k=10000"
[[1]]
       some       all
e 0.5061221 0.4938779
a 0.4938779 0.5061221

[[2]]
             e         a
some 0.5061215 0.4938785
all  0.4938785 0.5061215
```

Some/All-game

$$\lambda = 20$$

```
[1] "k=0"
[[1]]
  some all
e  1.0 0.0
a  0.5 0.5

[[2]]
       e   a
some 0.5 0.5
all  0.0 1.0

[1] "k=1"
[[1]]
          some          all
e 9.999546e-01 4.539787e-05
a 4.539787e-05 9.999546e-01

[[2]]
                e           a
some 9.987290e-01 0.001271016
all  2.061154e-09 0.999999998
```

```
[1] "k=2"
[[1]]
          some          all
e 1.000000e+00 2.114221e-09
a 2.114221e-09 1.000000e+00

[[2]]
                e          a
some 1.0000e+00 2.0649e-09
all  2.0649e-09 1.0000e+00

[1] "k=3"
[[1]]
          some          all
e 1.000000e+00 2.061154e-09
a 2.061154e-09 1.000000e+00

[[2]]
                e            a
some 1.000000e+00 2.061154e-09
all  2.061154e-09 1.000000e+00
```

```
[1] "k=4"
[[1]]
          some          all
e 1.000000e+00 2.061154e-09
a 2.061154e-09 1.000000e+00

[[2]]
                e            a
some 1.000000e+00 2.061154e-09
all  2.061154e-09 1.000000e+00

[1] "k=5"
[[1]]
          some          all
e 1.000000e+00 2.061154e-09
a 2.061154e-09 1.000000e+00

[[2]]
                e            a
some 1.000000e+00 2.061154e-09
all  2.061154e-09 1.000000e+00
```

IQR

- for all games discussed so far, as $\lambda \to \infty$, IQR fixed point converges towards IBR fixed point
- there are difference though

Some but not all

(1)
- **a.** Who came to the party?
- **b.** SOME: Some boys came to the party.
- **c.** ALL: All boys came to the party.
- **d.** SBNA: Some but not all boys came to the party.

- Let us suppose that the costs of (1d) is 0.1, and the other messages are costless

Some but not all

σ_0	SOME	ALL	SBNA
$w_{\exists\neg\forall}$	$1/2$	0	$1/2$
$w_\forall$	$1/2$	$1/2$	0

ρ_0	$w_{\exists\neg\forall}$	$w_\forall$
SOME	$1/2$	$1/2$
ALL	0	1
SBNA	1	0

ρ_1	$w_{\exists\neg\forall}$	$w_\forall$
SOME	$1/2$	$1/2$
ALL	0	1
SBNA	1	0

σ_1	SOME	ALL	SBNA
$w_{\exists\neg\forall}$	0	0	1
$w_\forall$	0	1	0

- (σ_1, ρ_1) is the fixed point
- prediction: no scalar implicature if there is a competing expression with exhaustive truth conditions, even if the latter is more expensive

Some but not all — the IQR sequence(s)

$$\lambda = 7$$

σ_7	SOME	ALL	SBNA
$w_{\exists\neg\forall}$	0.668	0.001	0.332
$w_{\forall}$	0.001	0.999	0.000

ρ_7	$w_{\exists\neg\forall}$	$w_{\forall}$
SOME	0.999	0.001
ALL	0.001	0.999
SBNA	0.999	0.001

$$\lambda = 100$$

σ_2	SOME	ALL	SBNA
$w_{\exists\neg\forall}$	1	0	4.5×10^{-5}
$w_{\forall}$	0	1	1.7×10^{-48}

ρ_2	$w_{\exists\neg\forall}$	$w_{\forall}$
SOME	1	0
ALL	0	1
SBNA	1	0

Horn's division of pragmatic labor

6. **a.** John stopped the car. (= STOP)
   b. John made the car stop. (= MAKE-STOP)

- w_1: John used the foot brake.
- w_2: John drove the car against a wall.
- $\|\text{STOP}\| = \|\text{MAKE-STOP}\| = \{w_1, w_2\}$
- $c(\text{STOP}) = 0$; $c(\text{MAKE-STOP} = 0.1$
- $p^*(w_1) = .6$; $p^*(w_2) = .4$.

Utility matrix

	a_1	a_2
w_1	1	0
w_2	0	1

Horn's division of pragmatic labor: IBR

IBR sequence

σ_0	STOP	MAKE-STOP
w_1	1/2	1/2
w_2	1/2	1/2

ρ_0	a_1	a_2
STOP	1/2	1/2
MAKE-STOP	1/2	1/2

ρ_1	a_1	a_2
STOP	1	0
MAKE-STOP	1	0

σ_1	STOP	MAKE-STOP
w_1	1	0
w_2	1	0

σ_2	STOP	MAKE-STOP
w_1	1	0
w_2	1	0

ρ_2	a_1	a_2
STOP	1	0
MAKE-STOP	1/2	1/2

ρ_3	a_1	a_2
STOP	1	0
MAKE-STOP	1/2	1/2

σ_3	STOP	MAKE-STOP
w_1	1	0
w_2	0	1

σ_4	STOP	MAKE-STOP
w_1	1	0
w_2	0	1

ρ_4	a_1	a_2
STOP	1	0
MAKE-STOP	0	1

Horn's division of pragmatic labor: IQR

- generalizes two higher number of types and signals
- crucially rests on the way surprise messages are treated in IBR
- in IQR, there are no surprise messages — every message has positive probability for each type
- how does IQR deal with Horn games?

Horn's division of pragmatic labor: IQR

$$\lambda = 20$$

```
[1] "k=0"
[[1]]
    f1  f2
t1 0.5 0.5
t2 0.5 0.5

[[2]]
    t1  t2
f1 0.5 0.5
f2 0.5 0.5

[1] "k=1"
[[1]]
          f1        f2
t1 0.8807971 0.1192029
t2 0.8807971 0.1192029

[[2]]
          t1         t2
f1 0.9820138 0.01798621
f2 0.9820138 0.01798621
```

```
[1] "k=2"
[[1]]
          f1        f2
t1 0.8807971 0.1192029
t2 0.8807971 0.1192029

[[2]]
          t1         t2
f1 0.9820138 0.01798621
f2 0.9820138 0.01798621

[1] "k=3"
[[1]]
          f1        f2
t1 0.8807971 0.1192029
t2 0.8807971 0.1192029

[[2]]
          t1         t2
f1 0.9820138 0.01798621
f2 0.9820138 0.01798621
```

Horn's division of pragmatic labor: IQR

- the difference in cost between $f1$ and $f2$ is the same
- due to the exponential decision rule, constant differences in utility translates into constant proportions in probabilities, regardless of the type
- possible solution (Michael Franke, p.c.):
  - convex monotonic transformation of utilities before they are fed into quantal decision rule
  - for instance:

Modified stochastic choice

$$P(a_i) = \frac{\exp(\lambda(u_i - \min_k u_k)^{1.1})}{\sum_j(\lambda \exp(u_j - \min_k u_k)^{1.1})}$$

Horn's division of pragmatic labor: IQR

$$\lambda = 20$$

```
[1] "k=0"
[[1]]
    f1  f2
t1 0.5 0.5
t2 0.5 0.5

[[2]]
    t1  t2
f1 0.5 0.5
f2 0.5 0.5

[1] "k=1"
[[1]]
          f1        f2
t1 0.8839913 0.1160087
t2 0.8839913 0.1160087

[[2]]
          t1         t2
f1 0.9678716 0.03212838
f2 0.9678716 0.03212838
```

```
[1] "k=2"
[[1]]
          f1        f2
t1 0.8985421 0.1014579
t2 0.8304270 0.1695730

[[2]]
          t1         t2
f1 0.9678716 0.03212838
f2 0.9678716 0.03212838

[1] "k=3"
[[1]]
          f1        f2
t1 0.8985421 0.1014579
t2 0.8304270 0.1695730

[[2]]
          t1         t2
f1 0.9839273 0.01607265
f2 0.3084919 0.69150806
```

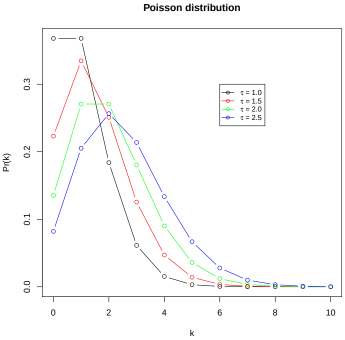
```
[1] "k=6"
[[1]]
             f1           f2
t1 1.000000e+00 2.259551e-10
t2 1.839206e-08 1.000000e+00

[[2]]
             t1           t2
f1 1.000000e+00 2.062151e-09
f2 2.061168e-09 1.000000e+00

[1] "k=7"
[[1]]
             f1           f2
t1 1.000000e+00 2.259551e-10
t2 1.839206e-08 1.000000e+00

[[2]]
             t1           t2
f1 1.000000e+00 2.061155e-09
f2 2.061154e-09 1.000000e+00
```

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**

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

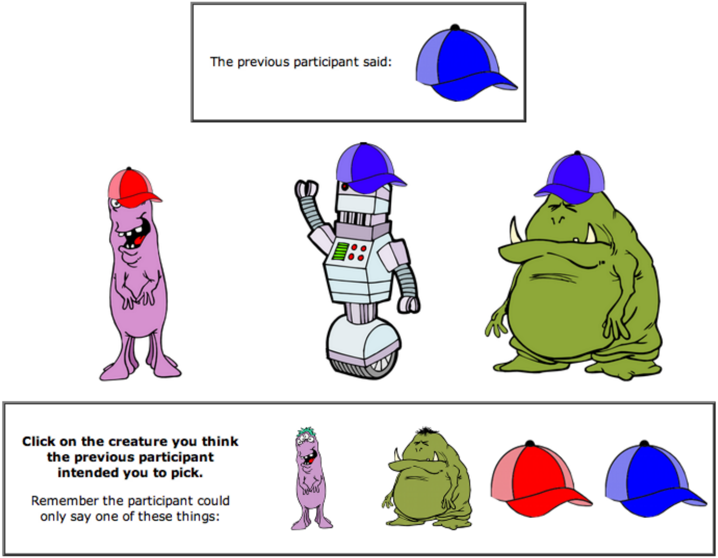
- τ, a free parameter of the model, is the average/expected level of the other players



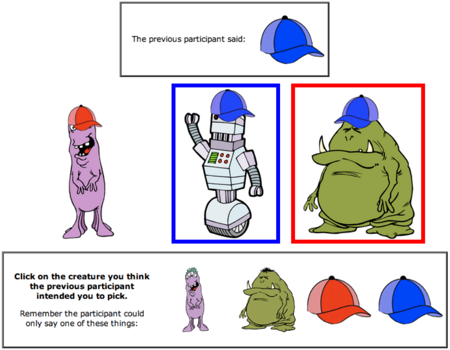
Experiment 1 - comprehension

- 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)

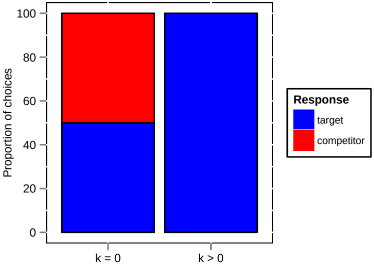
Simple implicature trial



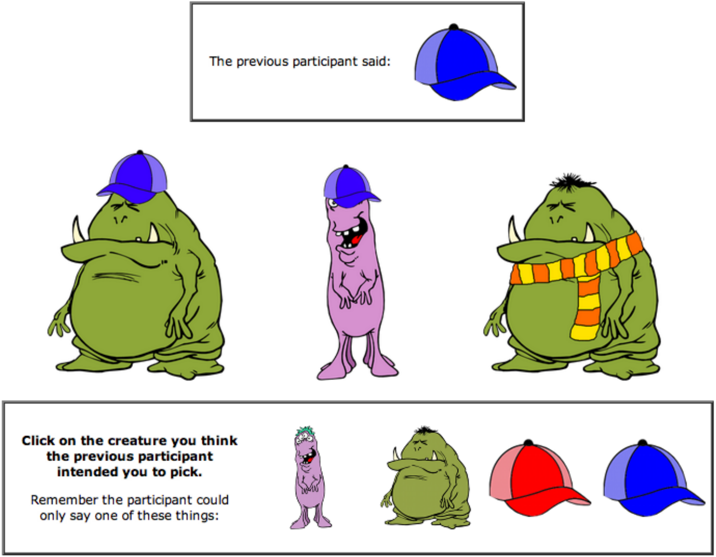
Simple implicature trial - predictions



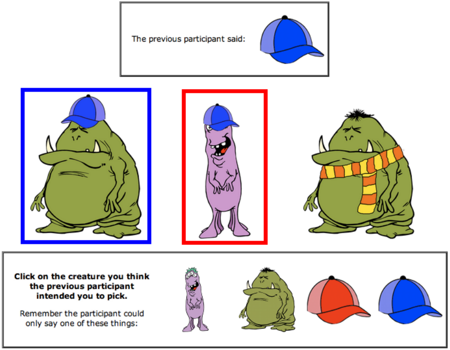
- IBR predictions for distribution of responses over target and competitor:



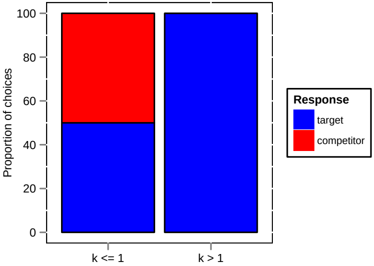
Complex implicature trial



Complex implicature trial - predictions

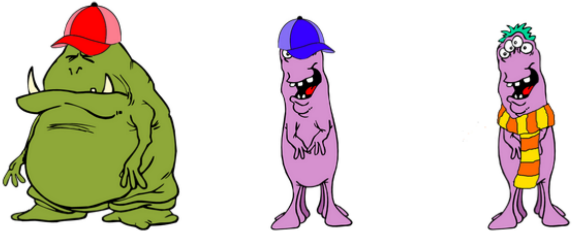


- IBR predictions for distribution of responses over target and competitor:



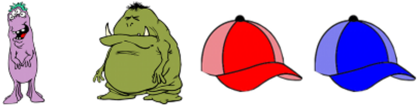
Unambiguous filler

The previous participant said:

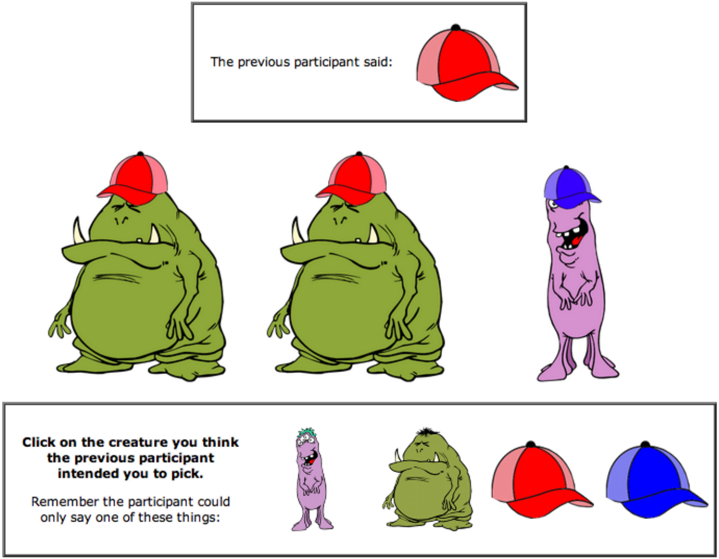


Click on the creature you think the previous participant intended you to pick.

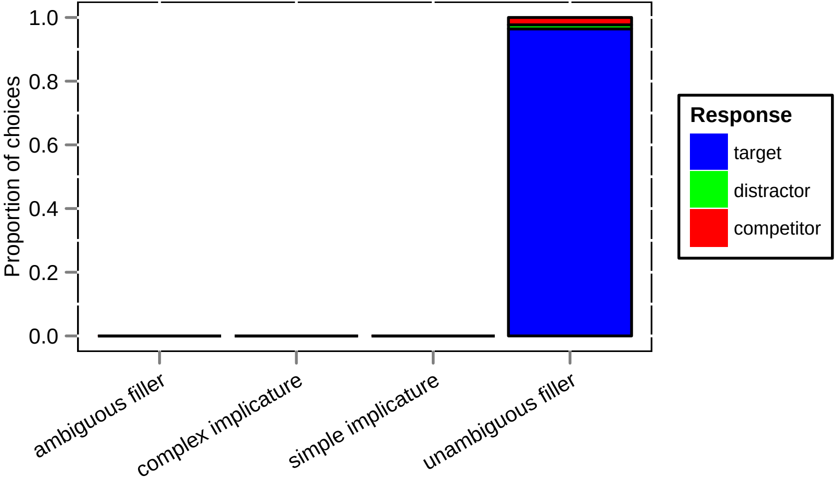
Remember the participant could only say one of these things:



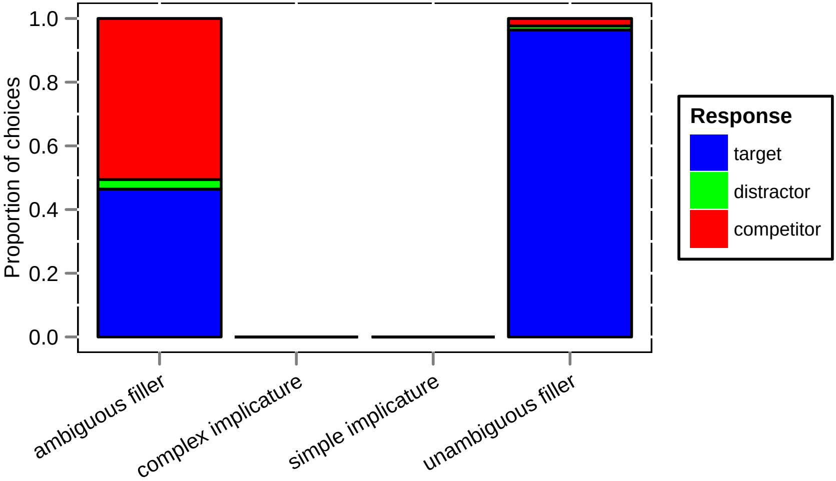
Ambiguous filler



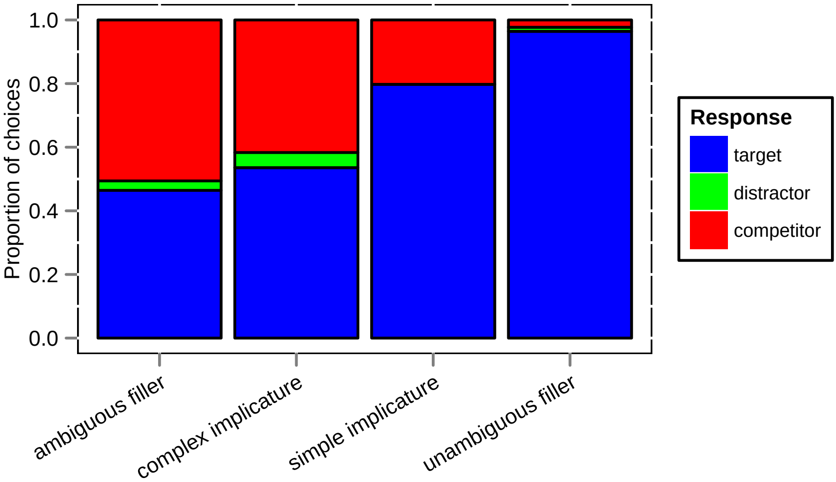
Results - proportion of responses by condition



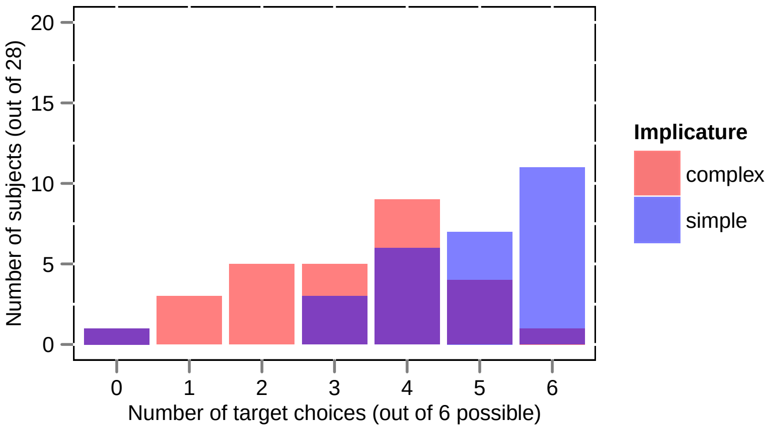
Results - proportion of responses by condition



Results - proportion of responses by condition

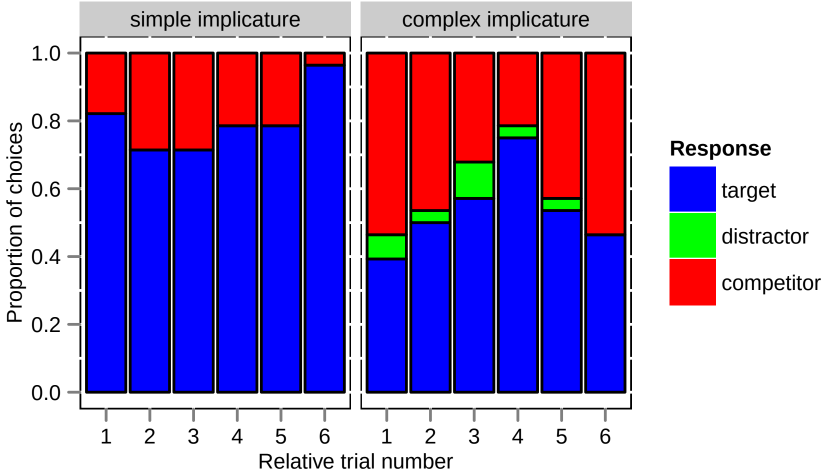


Results - distribution of subjects over target choices



→ not predicted by standard IBR

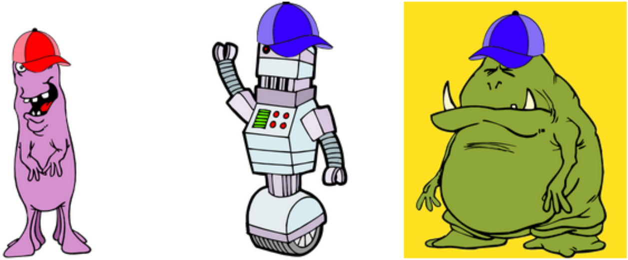
Results - learning effects



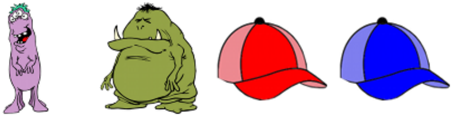
Experiment 2 - production

- 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)

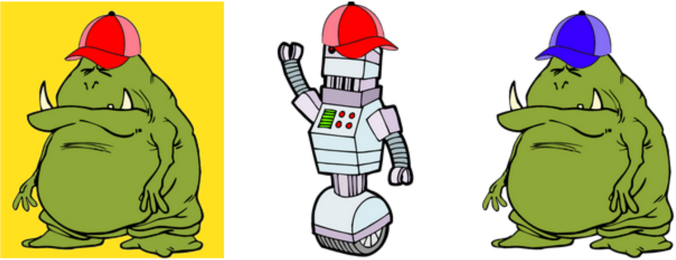
Simple implicature trial



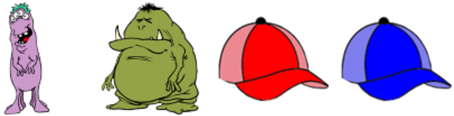
Your task is to get another worker to pick out the highlighted creature. It's not highlighted on their display. **Click on one of the following four messages to send it to the other worker and get them to pick out the right creature.** The other worker knows you can only send these messages.



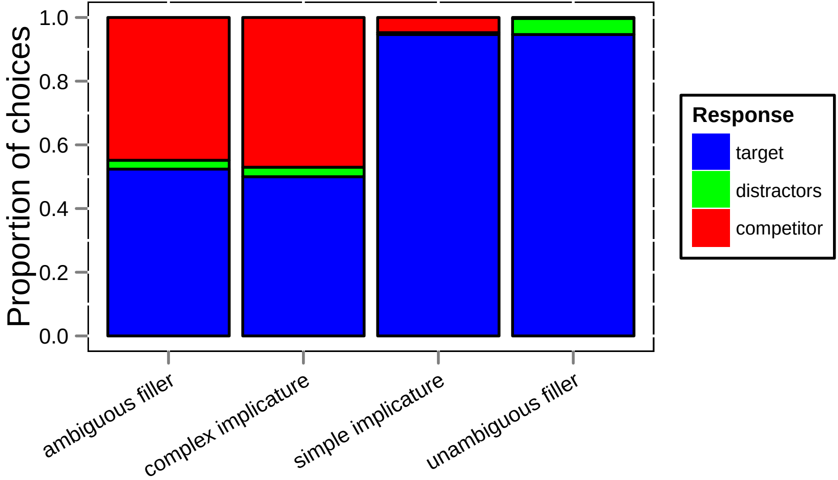
Complex implicature trial



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.

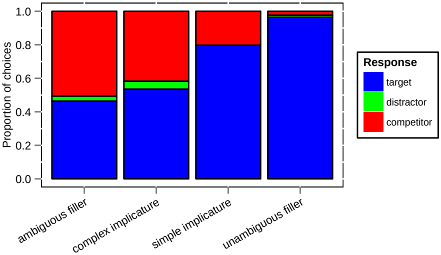


Results - proportion of responses by condition

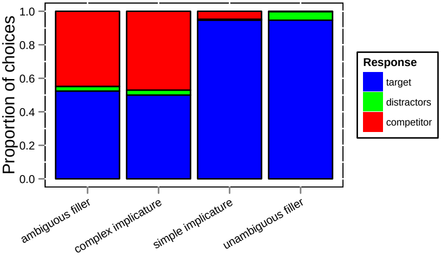


Results - proportion of responses by condition

Experiment 1
(comprehension)

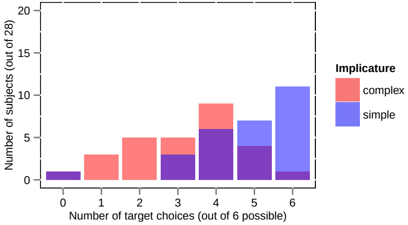


Experiment 2
(production)

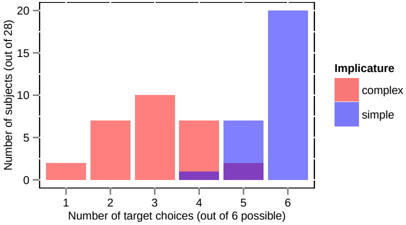


Results - distribution of subjects over target choices

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

Fitting the data

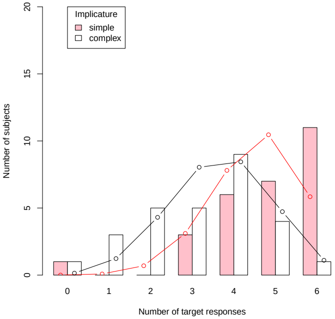
- maximum likelihood estimation of λ and τ on the basis of our experiments:

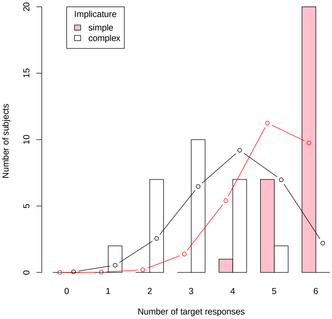
Experiment 1 (comprehension):

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

Experiment 2 (production):

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

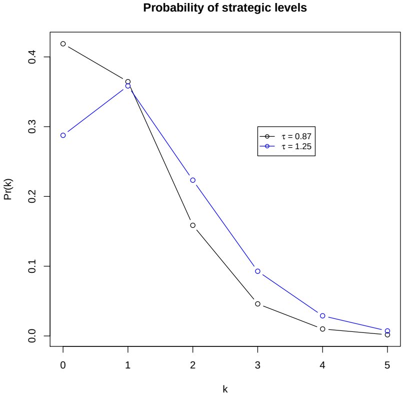




Tentative interpretation

- production/comprehension asymmetry:

Speakers are more strategic than listeners!



Alternative hypothesis

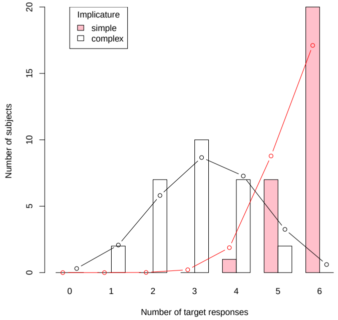
- This model took it for granted that non-strategic senders simply pick a true message at random.
- Results of experiment 2 suggest that this is not true; virtually everybody chooses the message that is most **informative**.
- Alternative hypothesis: S_0 uses the following utility function:

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

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!