Trust is good, strategic reasoning is better

Gerhard Jäger
gerhard.jaeger@uni-tuebingen.de

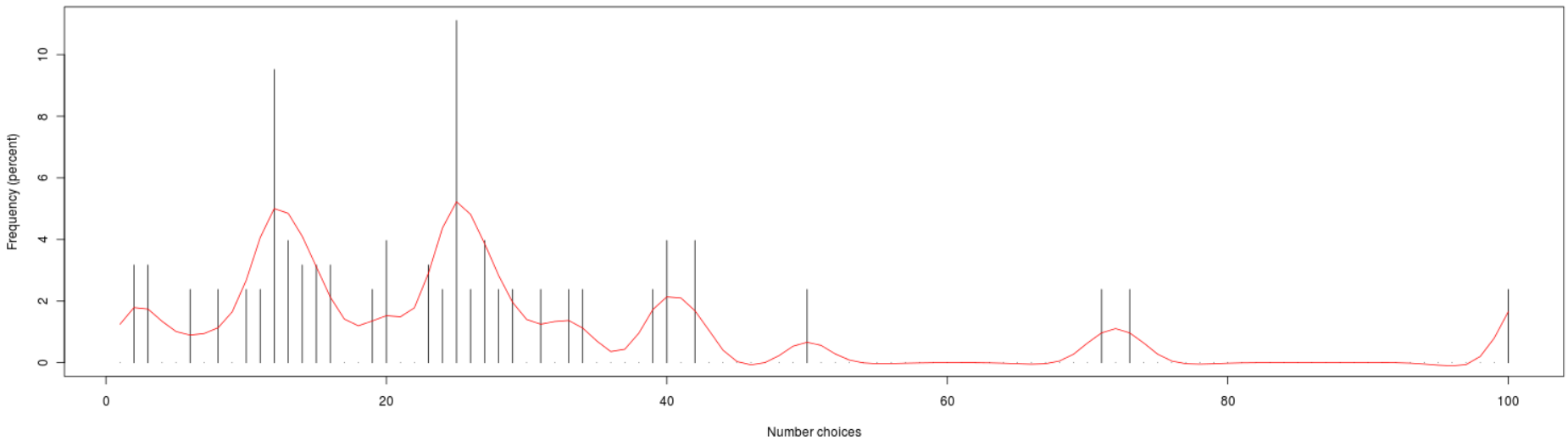
May 12, 2012

MIT

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).

Tea or coffee?

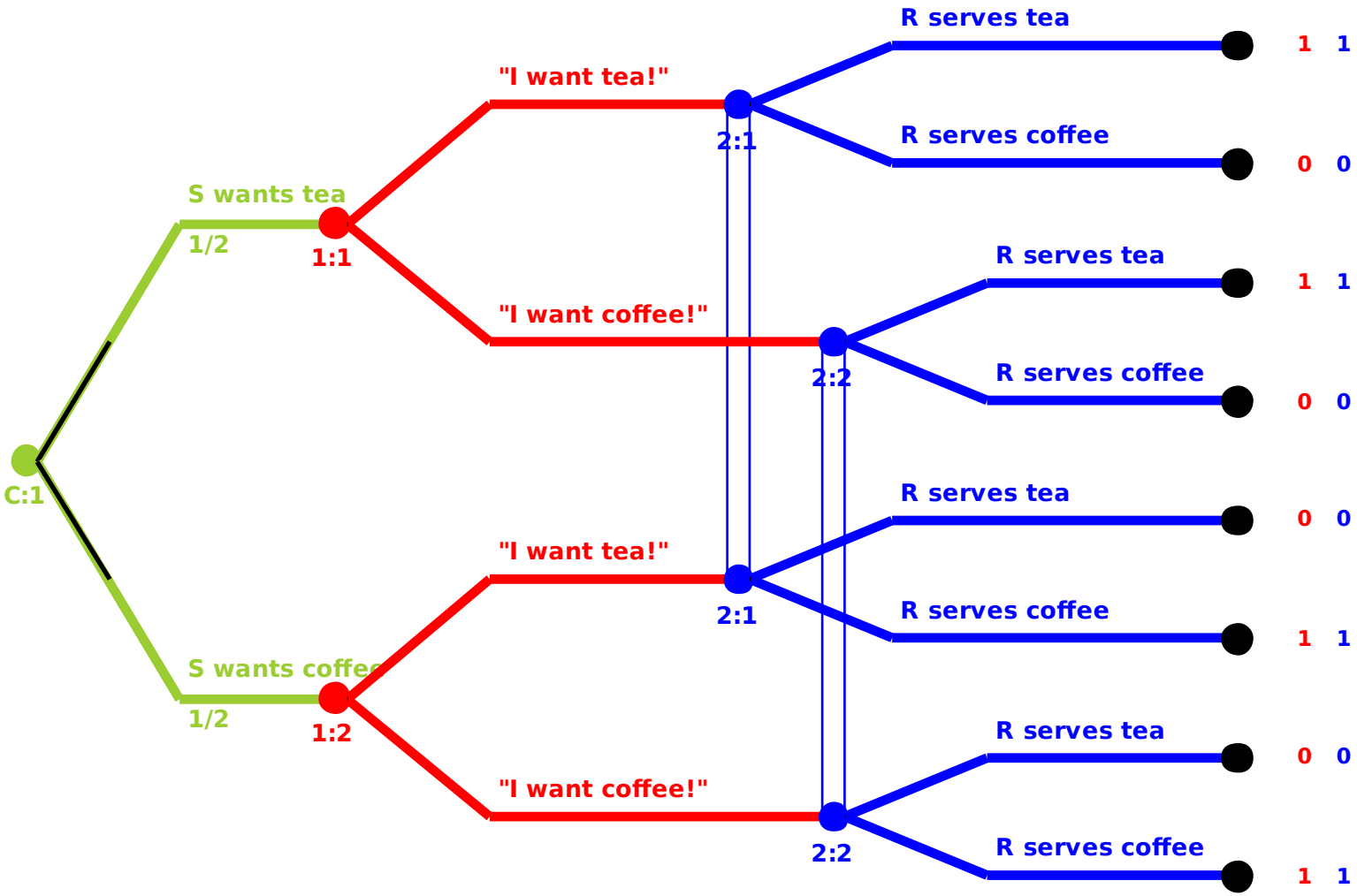
An example

- Sally either prefers tea (w_1) or coffee (w_2), with $p^*(w_1) = p^*(w_2) = \frac{1}{2}$.
- 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.

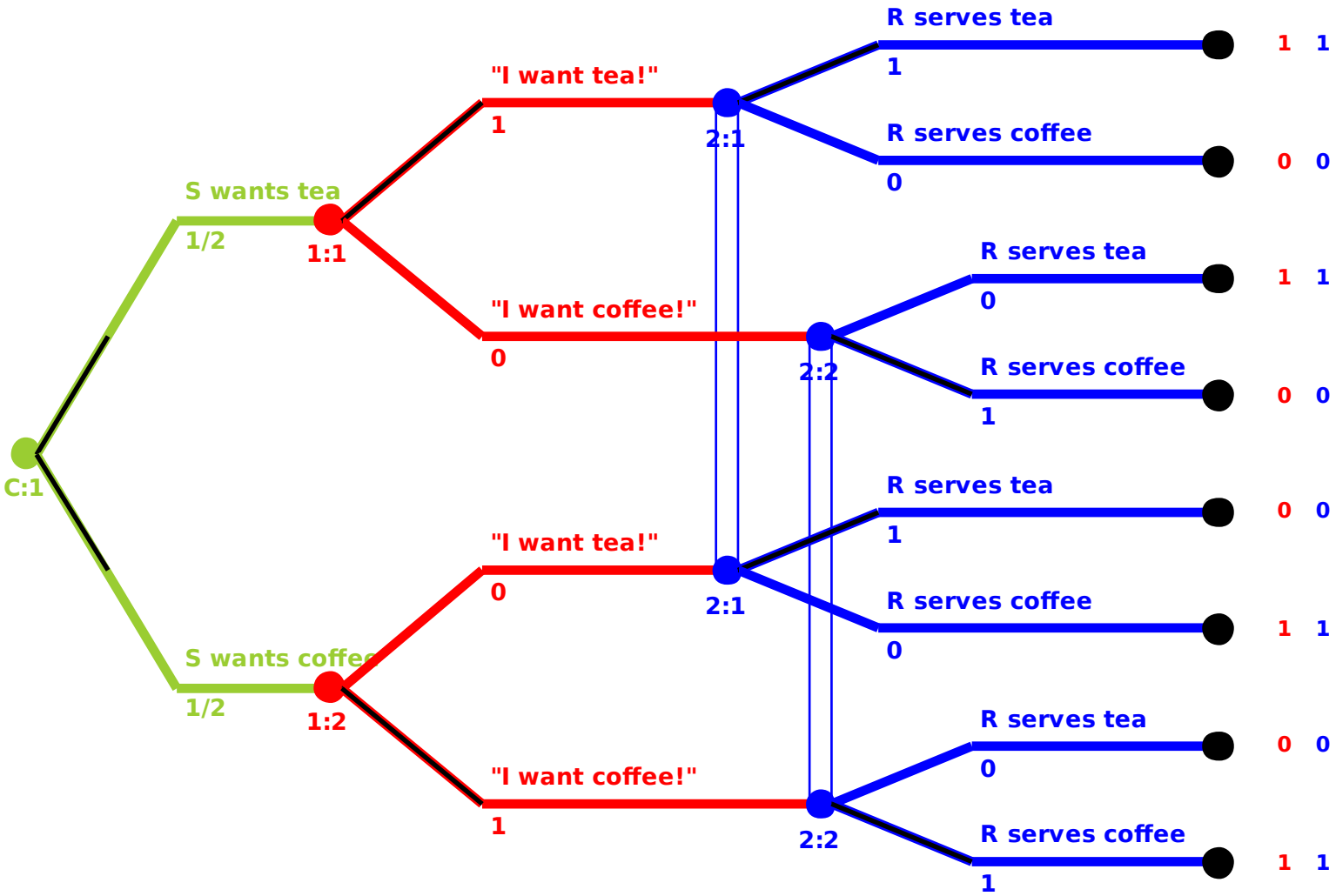
	a_1	a_2
w_1	$1, 1$	$0, 0$
w_2	$0, 0$	$1, 1$

Table: utility matrix

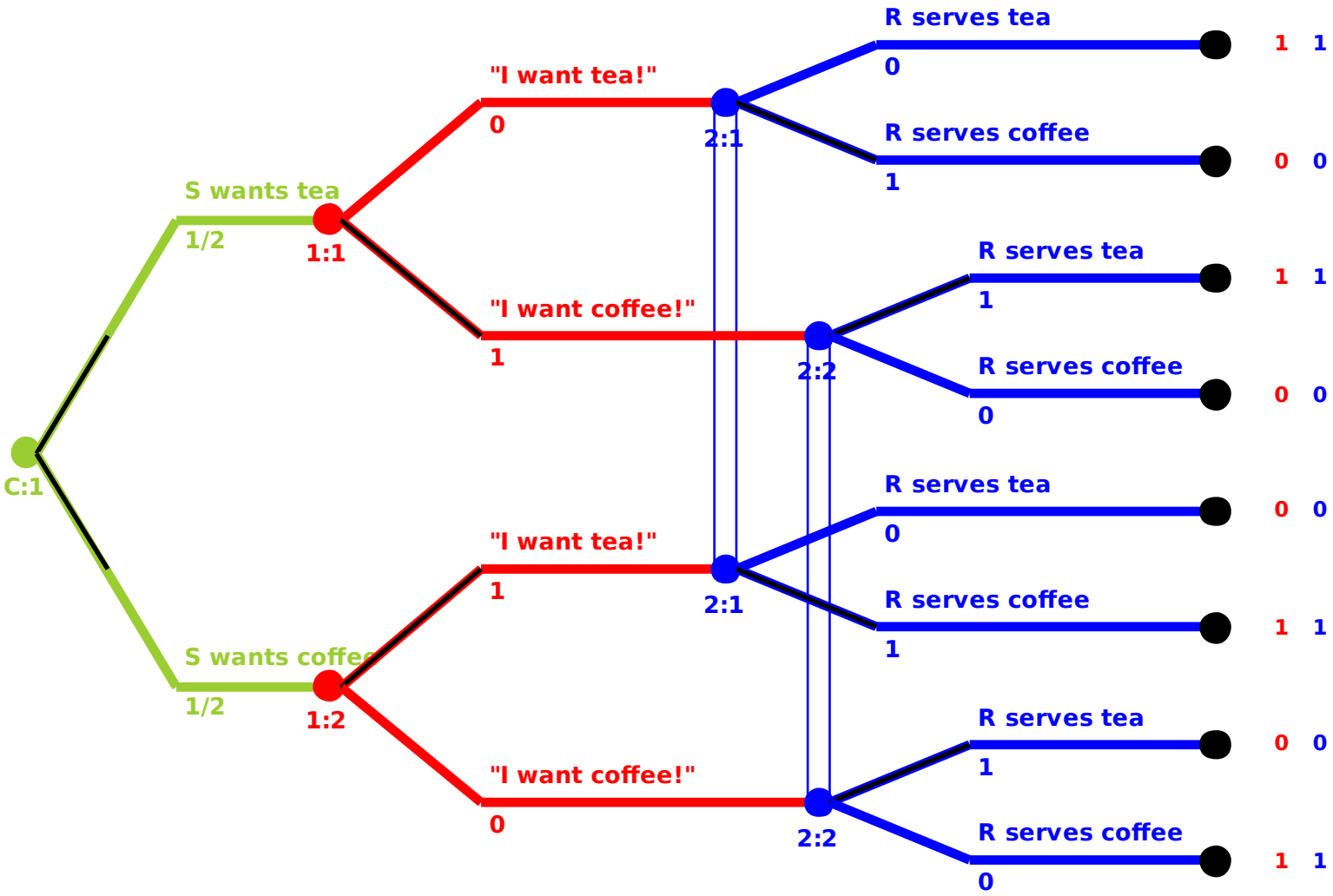
Extensive form



Extensive form



Extensive form



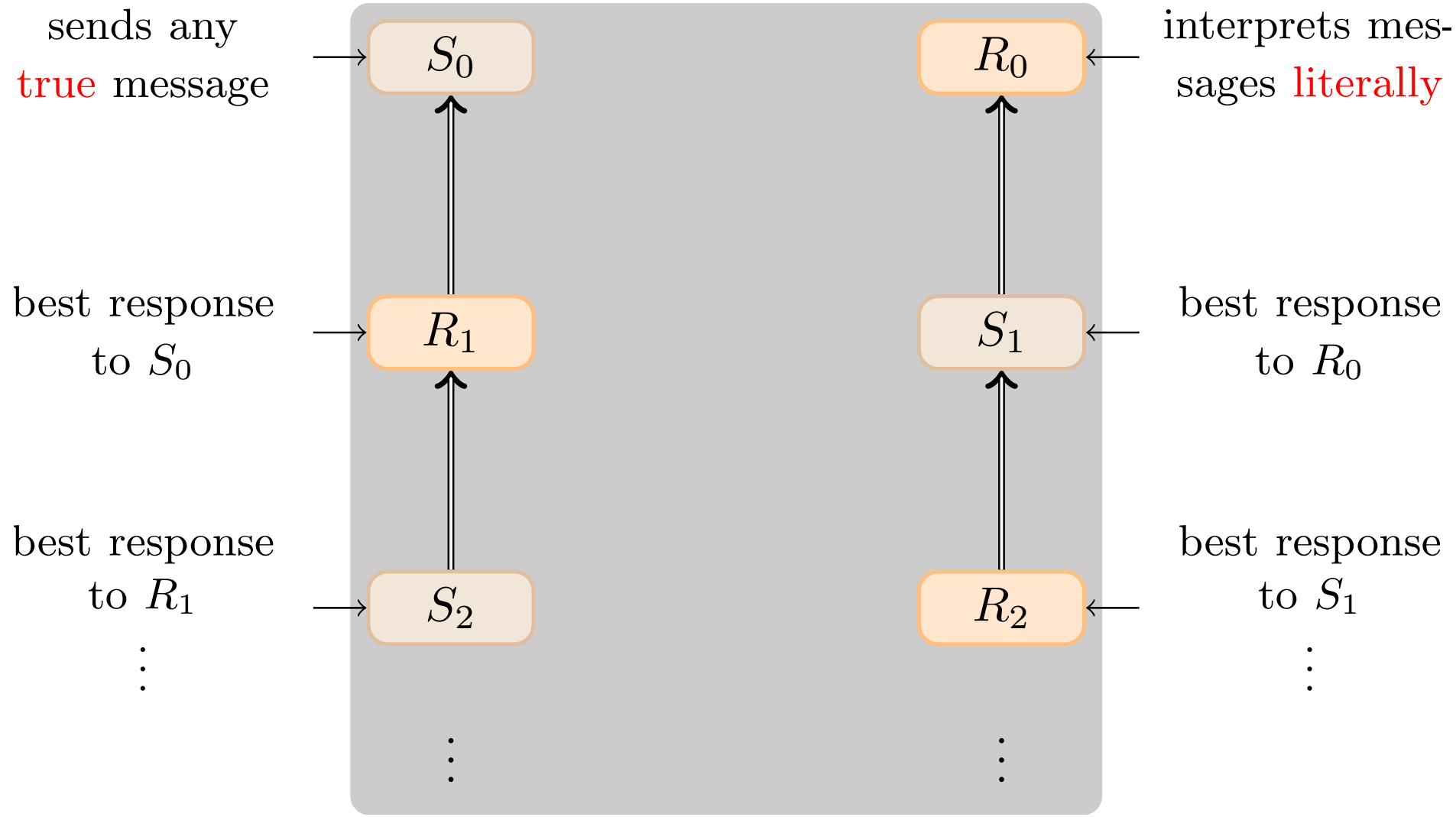
A coordination problem

- 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. $w \cup ct \nvdash \bot$
    - 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

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

- 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	$\frac{1}{2}$	$\frac{1}{2}$

σ_2	NO	SOME	ALL
$w_{\neg\exists}$	1	0	0
$w_{\exists\neg\forall}$	0	1	0
$w_{\forall}$	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

ρ_3	$w_{\neg\exists}$	$w_{\exists\neg\forall}$	$w_{\forall}$
NO	1	0	0
SOME	0	1	0
ALL	0	0	1

$$F \quad = \quad (\rho_1, \sigma_2)$$

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	$\frac{1}{2}$	$\frac{1}{2}$	0	0
w_b	$\frac{1}{2}$	0	$\frac{1}{2}$	0
w_{ab}	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

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_1	w_a	w_b	w_{ab}
OR	$\frac{1}{2}$	$\frac{1}{2}$	0
A	1	0	0
B	0	1	0
AND	0	0	1

ρ_3	w_a	w_b	w_{ab}
OR	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
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**:

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

- 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:

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

$$\begin{aligned}
c_r(\{w_a\}, \cdot) &= \tfrac{1}{49} & c_r(\{w_a, w_{ab}\}, \cdot) &= \tfrac{2}{49} \\
c_r(\{w_b\}, \cdot) &= \tfrac{1}{49} & c_r(\{w_b, w_{ab}\}, \cdot) &= \tfrac{2}{49} \\
c_r(\{w_{ab}\}, \cdot) &= \tfrac{1}{49} & c_r(\{w_a, w_b, w_{ab}\}, \cdot) &= \tfrac{3}{49} \\
c_r(\{w_a, w_b\}, \cdot) &= \tfrac{2}{49}
\end{aligned}$$

Lifted games

IBR sequence: 1

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

Sender costs

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 $c_s(\cdot, \text{AB-OR}) = \frac{1}{50}, c_s(\cdot, \text{A-OR}) = c_s(\cdot, \text{B-OR}) = \frac{1}{75}, c_s(\cdot, \text{OR}) = c_s(\cdot, \text{AND}) = \frac{1}{100}$), 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
$\{w_a\}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{4}$	0
$\{w_b\}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$\{w_{ab}\}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$
$\{w_a, w_b\}$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0	0
$\{w_a, w_{ab}\}$	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{4}$	$\frac{1}{4}$	0
$\{w_b, w_{ab}\}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$\{w_a, w_b, w_{ab}\}$	$\frac{1}{2}$	0	0	0	$\frac{1}{2}$	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	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$
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	$\frac{1}{3}$	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0	0
B-OR	0	$\frac{1}{3}$	$\frac{1}{3}$	0	0	$\frac{1}{3}$	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. $(= \mathrm{B} \wedge \forall x \mathrm{P}x)$

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	$\exists x \mathrm{P}x$	$\forall x \mathrm{P}x$	$\mathrm{B} \vee \exists x \mathrm{P}x$	$\mathrm{B} \wedge \exists x \mathrm{P}x$	$\mathrm{B} \vee \forall x \mathrm{P}x$	$\mathrm{B} \wedge \forall x \mathrm{P}x$
$\{w_{B\neg\exists}\}$	$\frac{1}{3}$	0	0	$\frac{1}{3}$	0	$\frac{1}{3}$	0
$\{w_{\neg B\exists\neg\forall}\}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	0	0	0
$\{w_{\neg B\forall}\}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0
$\{w_{B\exists\neg\forall}\}$	$\frac{1}{5}$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0
$\{w_{B\forall}\}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$
$\{w_{B\neg\exists}, w_{\neg B\exists\neg\forall}\}$	0	0	0	1	0	0	0
$\{w_{B\neg\exists}, w_{\neg B\forall}\}$	0	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	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	$\frac{1}{2}$	0	$\frac{1}{2}$
w_2	0	1	0
w_3	$\frac{1}{2}$	0	$\frac{1}{2}$
w_4	0	1	0

ρ_1	a_1	a_2	a_3	a_4
m_{100}	$\frac{1}{2}$	0	$\frac{1}{2}$	0
m_{101}	0	$\frac{1}{2}$	0	$\frac{1}{2}$
m_{ex100}	$\frac{1}{2}$	0	$\frac{1}{2}$	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}	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{3}$
m_{101}	0	1	0	0
m_{ex100}	$\frac{1}{2}$	0	$\frac{1}{2}$	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	$\frac{1}{2}$	$\frac{1}{2}$
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

$\vdots$ $\qquad$ $\vdots$

No fixed point; no stable information transmission.

Partially aligned interests

Crawford and Sobel's example (slightly simplified)

Sally is applying for a job. She has to rate her skills on a scale from 1 to 10. Robin is the employer and wants to assign Sally a job according to her skills. Higher skills means higher responsibility, higher demands, and a higher wage. Sally has an interest to exaggerate her skills a bit to increase her income, but she does not want to overdo it because she may end up being out of her depth. Robin wants to employ Sally exactly according to her actual skills.

Partially aligned interests

- possible worlds: $w_1 \cdots w_{10}$ (Sally's skill level)
- signals: $m_1 \ldots m_{10}$ (where Sally checks the application form)
- actions: $a_1 \ldots a_{10}$ (Sally's position in the job hierarchy after being hired)
- utilities: b, the **bias**, is the amount Sally wants to be overestimated

$$\begin{aligned} u_s(w_i, a_j) &= -(i + b - j)^2 \\ u_r(w_i, a_j) &= -(i - j)^2 \end{aligned}$$

Partially aligned interests

- outcome of IBR-sequence depends on b
  - $b \leq 0.5$: perfectly aligned interests; Sally will always send m_i in w_i, and Robin will react with a_i
  - $b > 2.5$: too large divergence of interests; Robin will ignore Sally's self-assessment because it is not to be trusted anyway
  - $0.5 < b \leq 2.5$: small number of informative but not maximally specific messages
- for $b = 1$:
  - Sally checks
    - m_9 or m_9 if she is in w_1,
    - m_9 if she is in w_2, w_3, or w_4, and
    - m_{10} if she is in $w_5 \ldots w_{10}$

Partially aligned interests

When a diplomat says yes, he means perhaps;
When he says perhaps, he means no;
When he says no, he is not a diplomat.

- (supposedly) Voltaire -

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

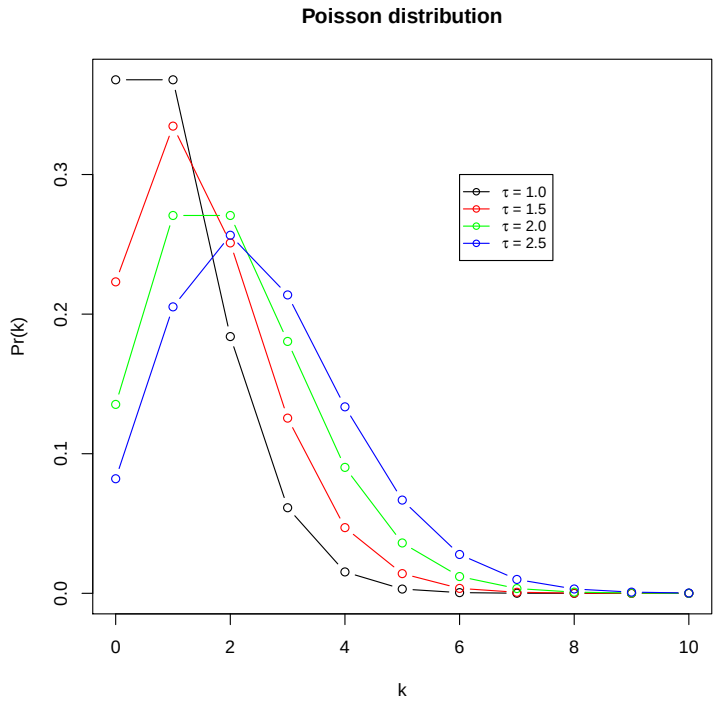
- λ measures degree of rationality
- $\lambda = 0$:
  - completely irrational behavior
  - all actions are equally likely, regardless of expected utility
- $\lambda \to \infty$
  - convergence towards behavior of rational choice
  - probability mass of sub-optimal actions converges to 0
- if everybody plays a quantal response (for fixed λ), play is in **quantal response equilibrium** (QRE)
- as $\lambda \to \infty$, QREs converge towards Nash equilibria

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**

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

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



Level-k thinking

- model has two parameters, λ and τ
- makes predictions about relative frequencies of signal use/interpretation that can be fitted and tested against experimental data