Bootstrapping for Interactions between **Humans and Robots**

Kerstin Fischer
University of Southern Denmark @ Sonderborg

Overview

- bootstrapping in humans (and Martians)
  - what: the uses of categories
  - how:
    - the nature of word meaning
    - child language acquisition
    - grammar as surface generalization
- bootstrapping in robots
  - and the role of social guidance
- bootstrapping for interaction
  - how does interaction work?
  - consequences for HRI design

INVOICE

Sonic.net, Inc. 1
2260 Apollo Way
Santa Rosa, CA 95407
707.522.1000
support@sonic.net

Today's Date: 2003-10-22 2
Service from 2003-11-01 to 2003-12-01

For:
John Q. Customer <login@sonic.net> 3
123 Any Street
Santa Rosa, CA 95401

Description	Monthly Price	Total	
Basic Internet Service, login login	18.95	18.95	4
Email account mbox1	0.00	0.00	
Domain Registration Service domain1.tld	0.00	0.00	5
DNS domain1.tld	1.95	1.95	
Multihomed webspace domain1.tld	19.00	19.00	
SSL Key Management login	19.00	19.00	
MySQL Database - login	0.00	0.00	
Shell Services Enabled	0.00	0.00	
SBC-ASI DSL, IP 64.142.0.0 / PVC 42/49	0.00	0.00	
IP address space 64.142.0.0/30	0.00	0.00	
SBC-ASI DSL - Basic 384kbps-1.5Mbps/128kbps service	0.00	0.00	
Credit DSL Introductory Pricing 2003-09-26 to 2003-10-17	n/a	16.10cr	8
Credit DSL Introductory Pricing 2003-10-17 to 2003-11-17	n/a	23.00cr	
Charge Additional Disk Space Usage for 10/02	n/a	5.00	
Charge SBC-ASI DSL passthrough billing -- Basic Service 2003-10-17 to 2003-11-17	n/a	39.00	9
Charge SBC-ASI DSL passthrough billing -- Pro-rated Service 2003-09-26 to 2003-10-17	n/a	27.30	
	Total:	91.10	10

6 7

If you have questions, you can contact Sonic.net support by sending email to support@sonic.net, or by calling 707-547-3400. 11

INVOICE

some text

some text

some text

Description Monthly Price Total

some list

some numbers

some text

Invoices

- layout provides clues to
  - types of information
  - relevance

- layout result of
  - recurrent problems/tasks
  - evolution and conventionalization over decades
  - readers' perceptual capabilities

Invoices

- ‘knowledge’ about invoices not explicit
- never taught
- learned via exposure to exemplars
- as adults/adolescents

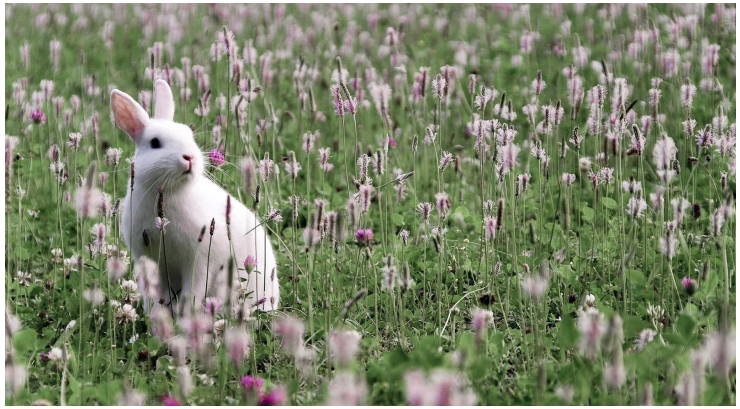
Obvious advantages:

- easy access: categorization of parts of the document
- speeds up recognition (short cut to meaning)
- makes invoices predictable

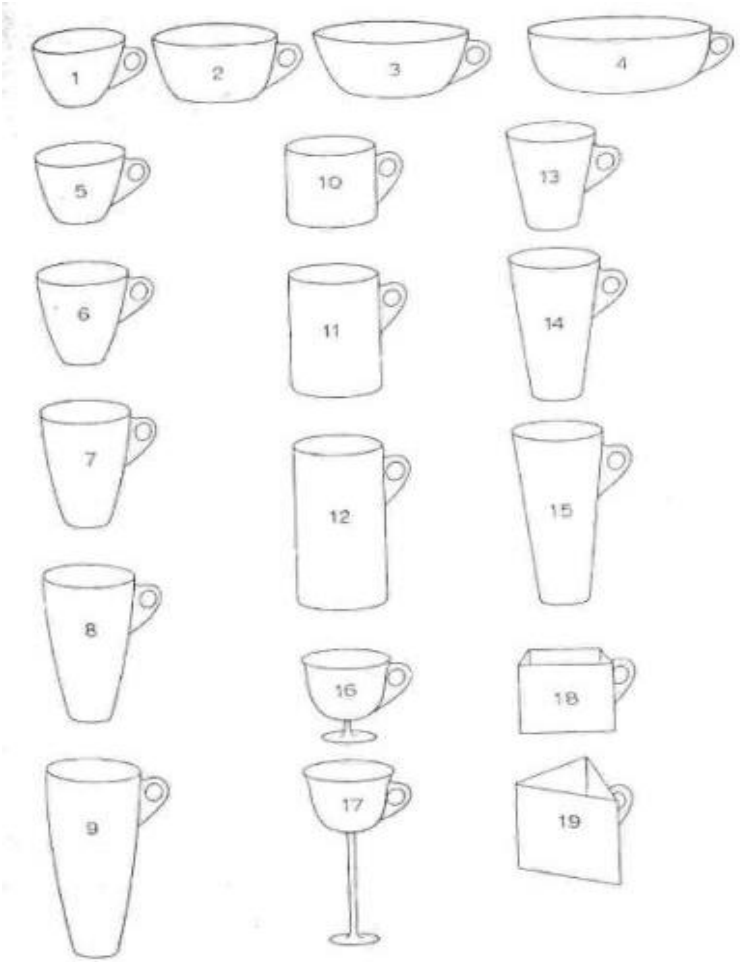
Bootstrapping in Humans (and Martians): the 'Gavagai' problem

The Gavagai Problem

- identify the ‘essential’ characteristics
- identify distributional regularities
- identify central exemplars
- ignore accidental properties



What is meaning?



what is the meaning of ‘cup’?

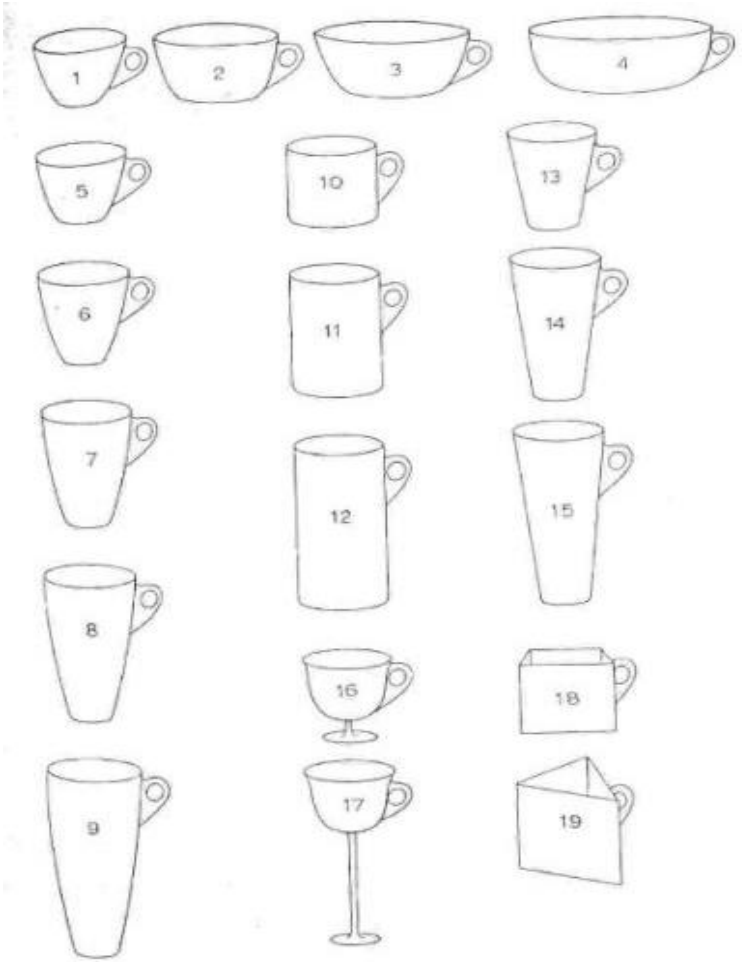
What is the meaning of 'cup'?

- formal semantics:
  - the meaning of the expression '*cup*' is CUP
- structuralists:
  - not a mug, not a bowl, not a vase, not an X…
- Labov (1972): objective features, yet fuzzy boundaries
- cognitive linguistics:
  - embodiment
  - functionality
  - perspective
  - prototype effects

What is the meaning of 'cup'?

Labov (1972)

- width-depth ratio
- with one handle
- made of opaque vitreous material
- with a saucer
- circular in cross-section
- for consumption of hot liquid food

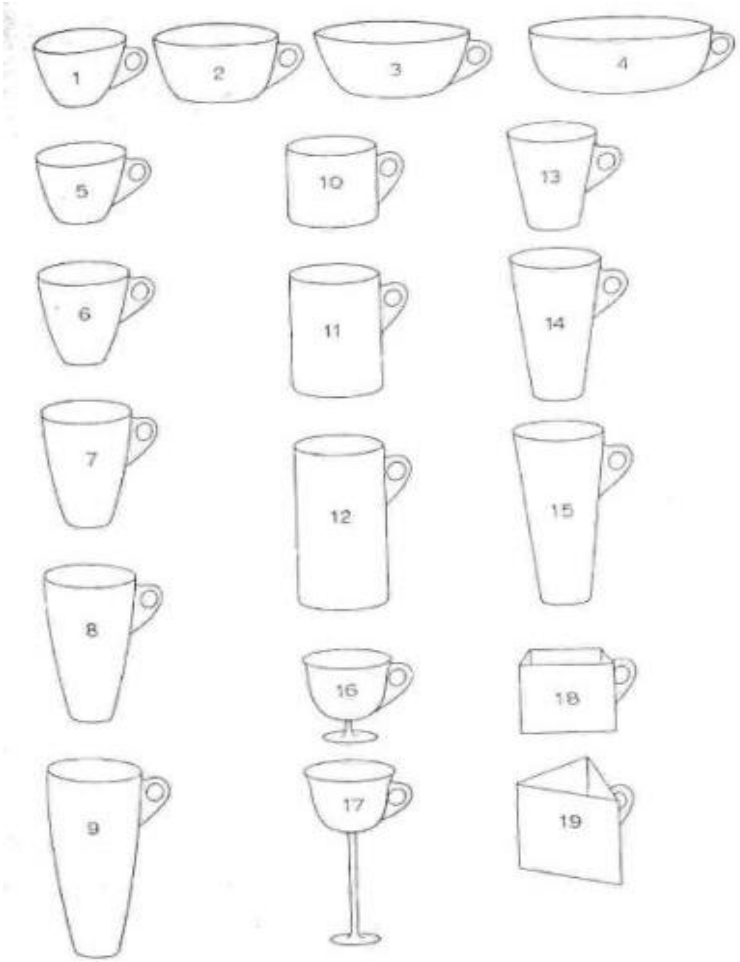


- ➢ prototype structure of categories
- ➢ fuzzy boundaries
- ➢ gradience

What is the meaning of 'cup'?

Wierzbicka (1985)

- an artefact
- to lift hot liquids to your mouth
- with one hand
- for drinking
- while sitting at a table
- …



- language structure reflects
  - human embodiment
  - human categorization

Language reflects human categorization

- language = direct clue to categorization
  - nouns = entities, concepts, 'bounded regions'
  - grammar = regularly occurring experience (syntax as frozen discourse)

- f.ex.:

Who does what to whom?

➢ argument structure constructions

e.g. *she crutched him the ball*

e.g. *she crutched him*

Language reflects human categorization

- language = direct clue to categorization
  - nouns = entities, concepts, 'bounded regions'
  - grammar = regularly occurring experience (syntax as frozen discourse)

- f.ex.:

Who does what to whom?

➢ argument structure constructions

➢ also robots evolve them (Steels 2005, 2006…)

Language reflects human categorization

- language = direct clue to categorization
  - nouns = entities, concepts, 'bounded regions'
  - grammar = regularly occurring experience (syntax as frozen discourse)

- f.ex.:

here is something I want to stress:

- prosody: I like MARY.
- extraposition: Mary, I like; it is Mary I like; who I like is Mary

  - context-dependent

Language reflects human categorization

- language = direct clue to categorization
  - nouns = entities, concepts, 'bounded regions'
  - grammar = regularly occurring experience (syntax as frozen discourse)

- f.ex.:

semantic category

➢ word order:

touch the	**Color-Adj**	**Shape-N**
	ø	ø
	red	ball
	blue	block
	pink	octopus

Language as an Inventory of Constructions

- form-meaning pairs
- language can be exhaustively described as a system of signs, i.e. constructions

  - "they are abstract templates obtained by reinforcing the commonality inherent in a set of instances" (Langacker 2008: 23)
  - "abstracted from usage events" (2008: 458)
  - "entrenched patterns of processing activity we can evoke and execute as needed" (2008: 459):

Bootstrapping in Humans

- human (language) categories are shaped by human capabilities, needs, embodiment
- natural language categories are not defined by objective criteria
- natural language categories exhibit prototype effects and fuzzy boundaries
- natural language is a system of pairings between objectively identifiable forms and 'humane' meanings

So How Do Humans Learn it?

- special adjustments to the language (and cognition) learner

- increase contingency
  - distributional regularities
  - increased coordination of cues
  - connection words – activities

So How Do Humans Learn it?

- special adjustments to the language (and cognition) learner

- increase contingency
- support comprehension
  - adjust MLU (mean length of utterance)
  - isolated words
  - conceptual decomposition
  - reformulation
  - embedding in recurrent action

So How Do Humans Learn it?

- special adjustments to the language (and cognition) learner

- increase contingency
- support comprehension
- highlight distinctions
  - provision of relevant contrasts
  - increase transparency
  - isolated words
  - variation sets

So How Do Humans Learn it?

- special adjustments to the language (and cognition) learner

- increase contingency
- support comprehension
- highlight distinctions

- ➢ form-meaning pairs co-develop
- ➢ social guidance crucial during acquisition

Bootstrapping in Robots

- motor babbling
- exploration
- kinesthetic guidance, imitation, demonstration
- supervised learning (high level feedback/evaluation)
- …

Bootstrapping in Robots

Role of language in bootstrapping categories

- Mirolli & Parisi (2009, 2012):
  - category learning & formation
  - abstraction (ignore irrelevant & highlight relevant dimensions)
  - memory
  - control & mental life
- Leugger & Nolfi (2011): labels & self talk facilitate learning
- Bhorghi & Cinatti (2012): abstract concepts learned on the basis of language as a starting point

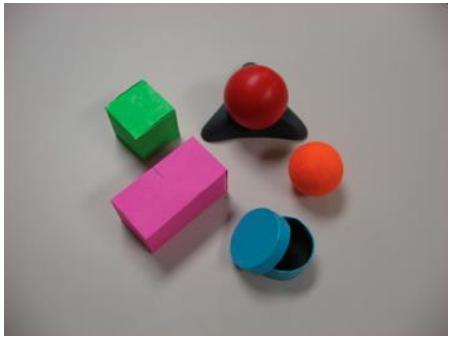
Bootstrapping in Robots

Role of language in bootstrapping categories

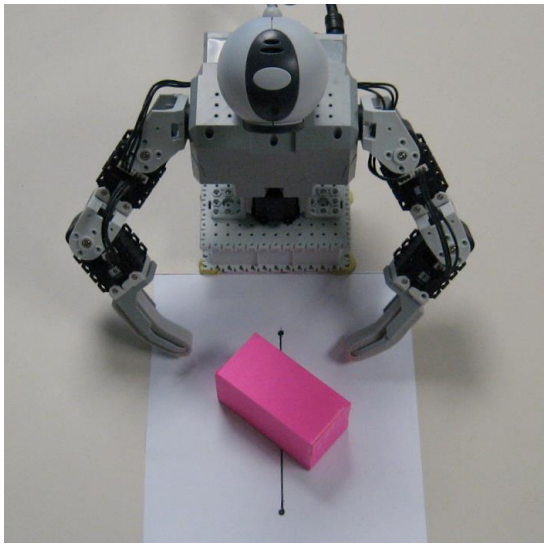
- Mirolli & Parisi (2009, 2012):
  - category learning & formation
  - abstraction (ignore irrelevant & highlight relevant dimensions)
  - memory
  - control & mental life
- Leugger & Nolfi (2011): labels & self talk facilitate learning
- Bhorghi & Cinatti (2012): abstract concepts learned on the basis of language as a starting point

➢ inspite of natural language being so 'humane'

Bootstrapping in Robots



- socially guided machine learning
  - Thomaz & Cakmak (2009):
    - balance of positive & negative examples
    - number of examples provided corresponds to number of affordances of object
    - example quality:
      - people start simple
      - pointing out rare affordances
      - help in parsing action goals
    - learning results for social versus systematic data sets equally good or better, especially for rare affordances



Intuitive Tutoring: Experimental Set-up

- Participants
  - 30 naïve participants
  - simulated robot ‘Babyface’
    - behavior = gaze
    - driven by attention model

- Procedure
  - task: explain sentences to the robot
  - analyze people’s utterances according to the participants in the scene described

Conclusions: Intuitive Tutoring

- naïve participants decompose intuitively grammatical meanings for the robotic learner
  - unexpected since linguistic knowledge is merely tacit

- users produce tutoring strategies
  - based on their understanding of the complexities of language
  - based on the robot's contingent feedback

- the socially guided learning paradigm is thus also highly promising for grounded grammar learning

Conclusion: Bootstrapping in Robots

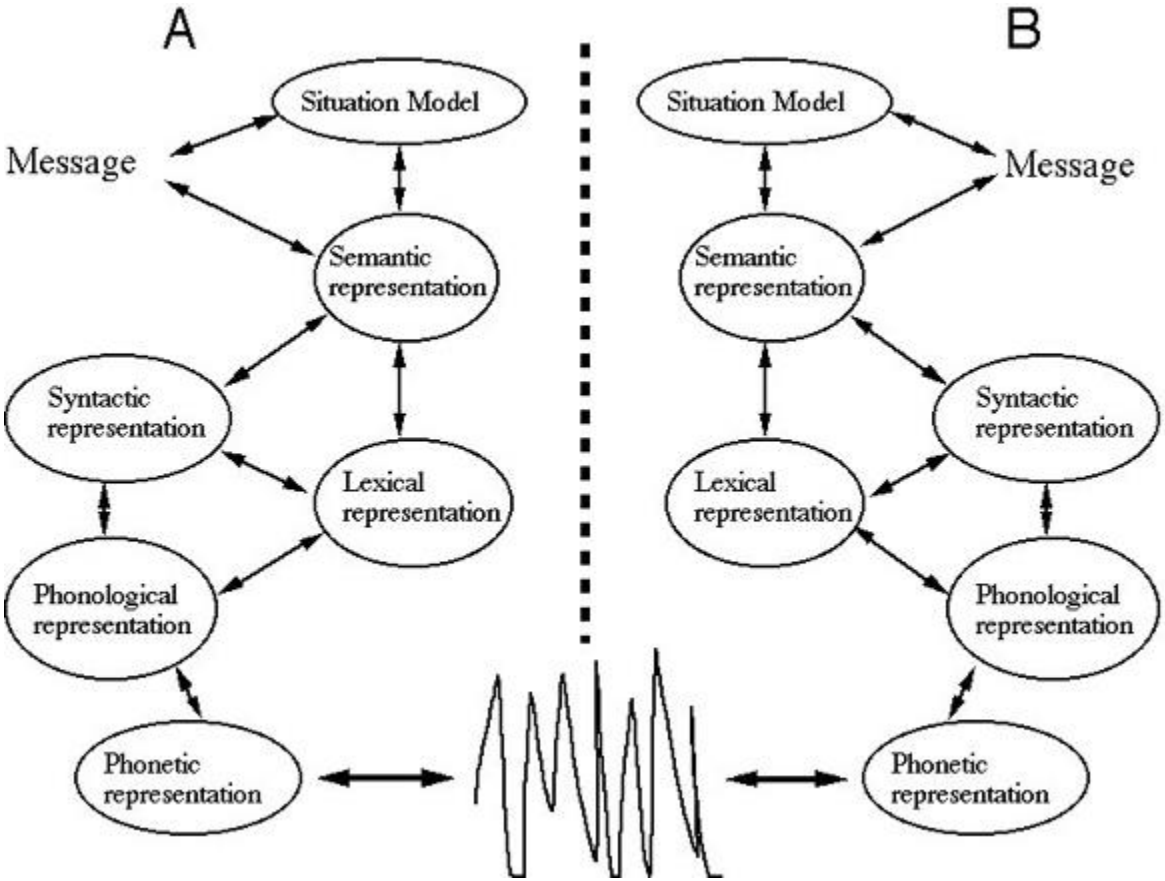
- bootstrapping in robots profits from human guidance,
  - linguistic
  - and other

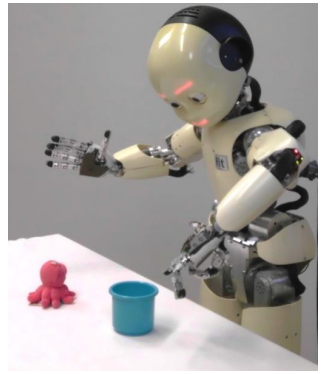
Conclusion: Bootstrapping in Robots

- bootstrapping in robots profits from human guidance,
  - linguistic
  - and other

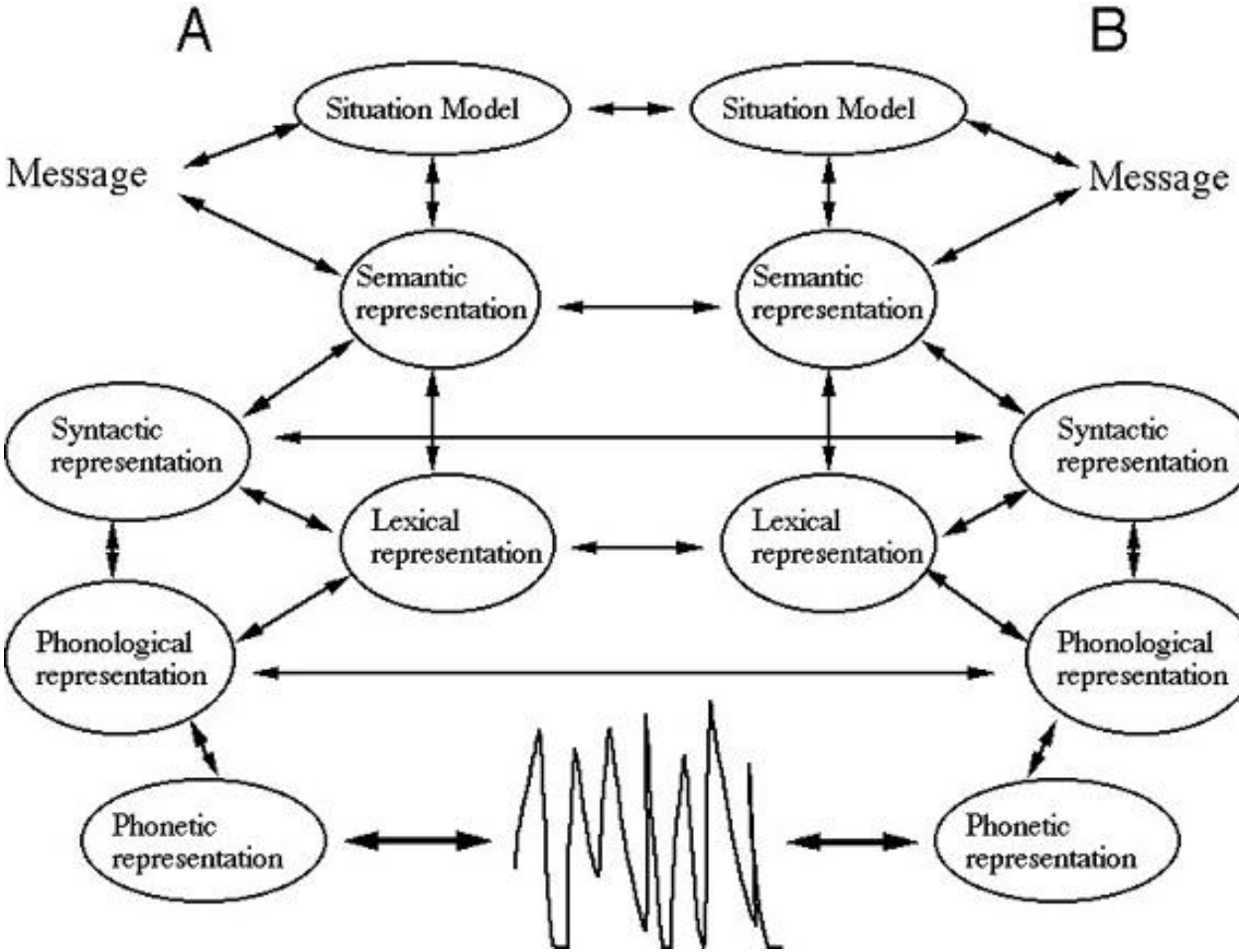
- using natural language categories furthermore facilitates HRI

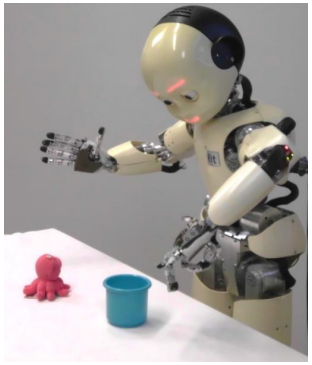
Communication (LevelI)





The Interactive Alignment Model





Shared Background

- spatial instruction dialogs
- real-time, sophisticated dialog system
- complex computational spatial model, e.g.

"*go to the block in the middle*"
(out of the group of objects on your left)

Shared Background

Command1: drive straight ahead to the right

Command2: turn 45 degrees to the right

Command3: turn to the right

Command4: drive 10 cm ahead

Command5: cme on

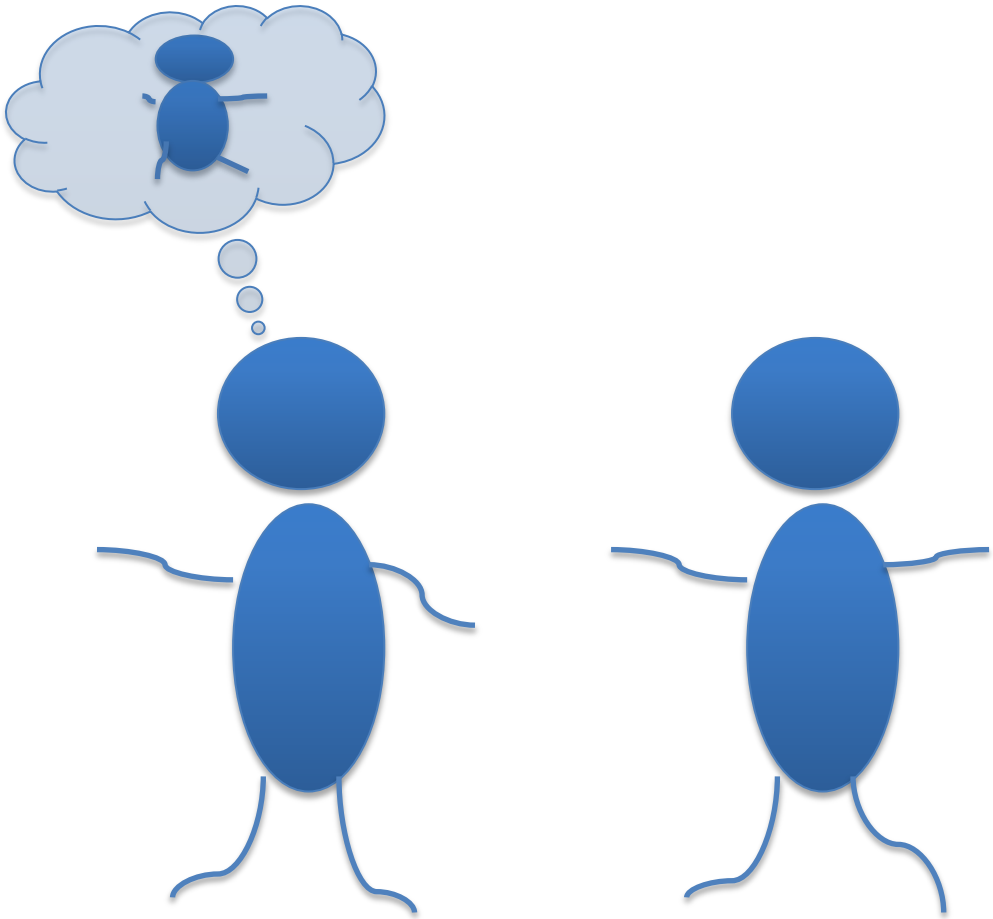
Command6: come on

Command7: start driving

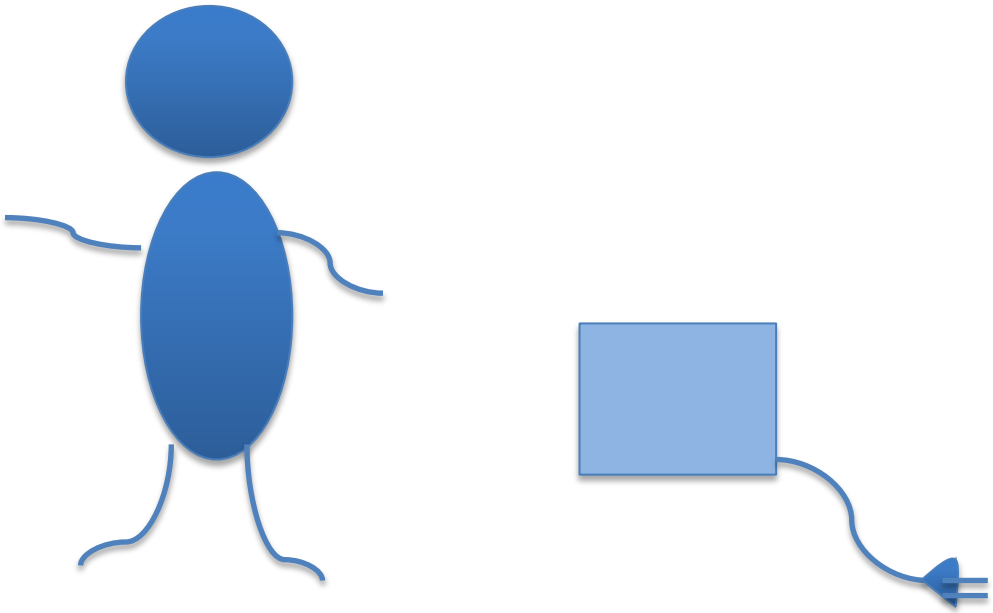
Command8: engine on

goal > path > movement > instrumental action

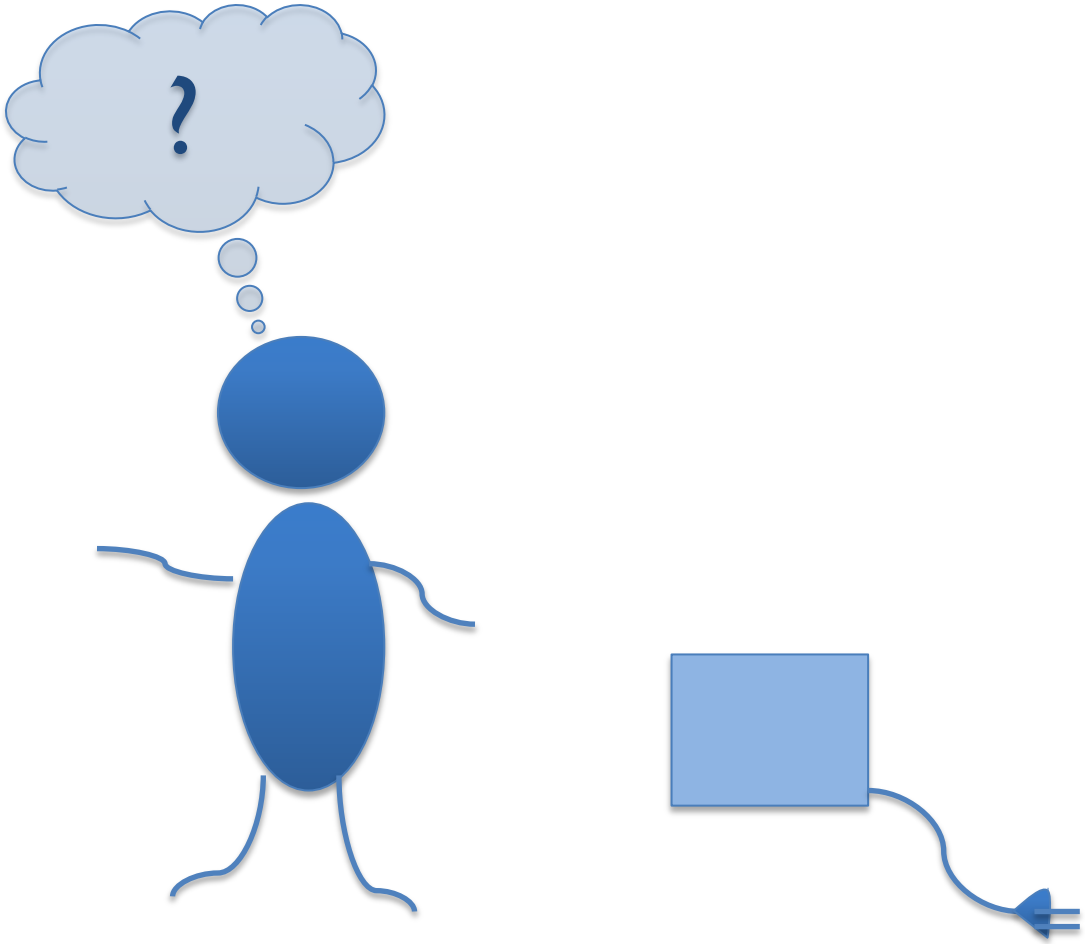
Interaction



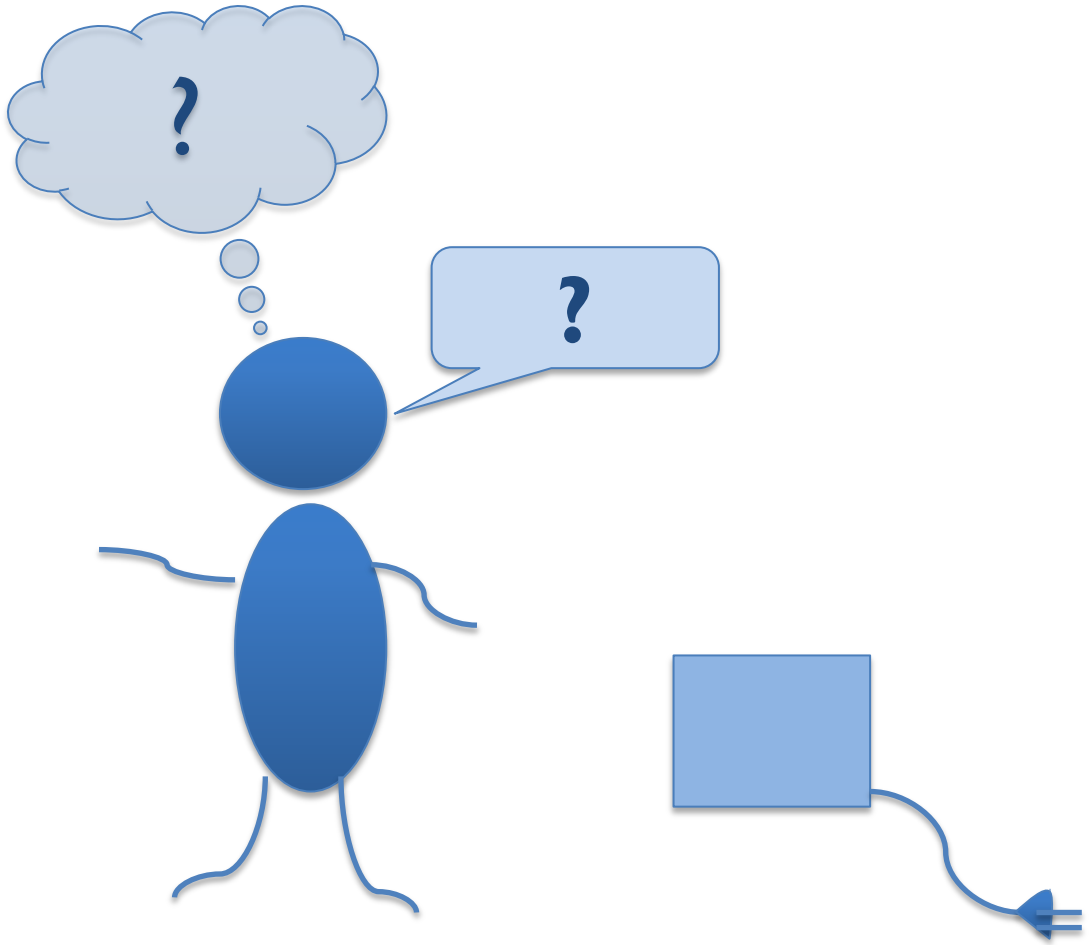
Partner Models



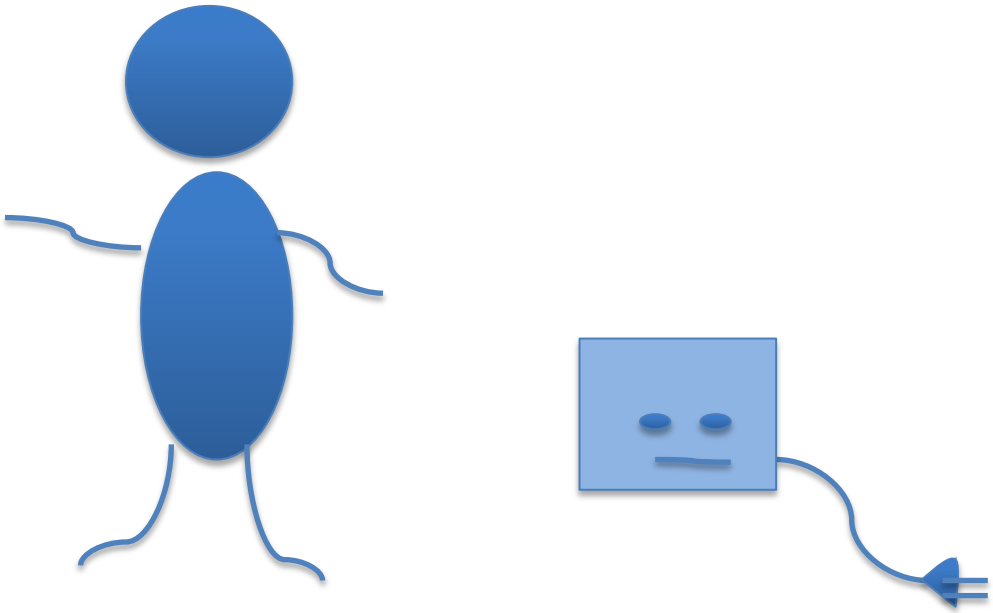
Partner Models



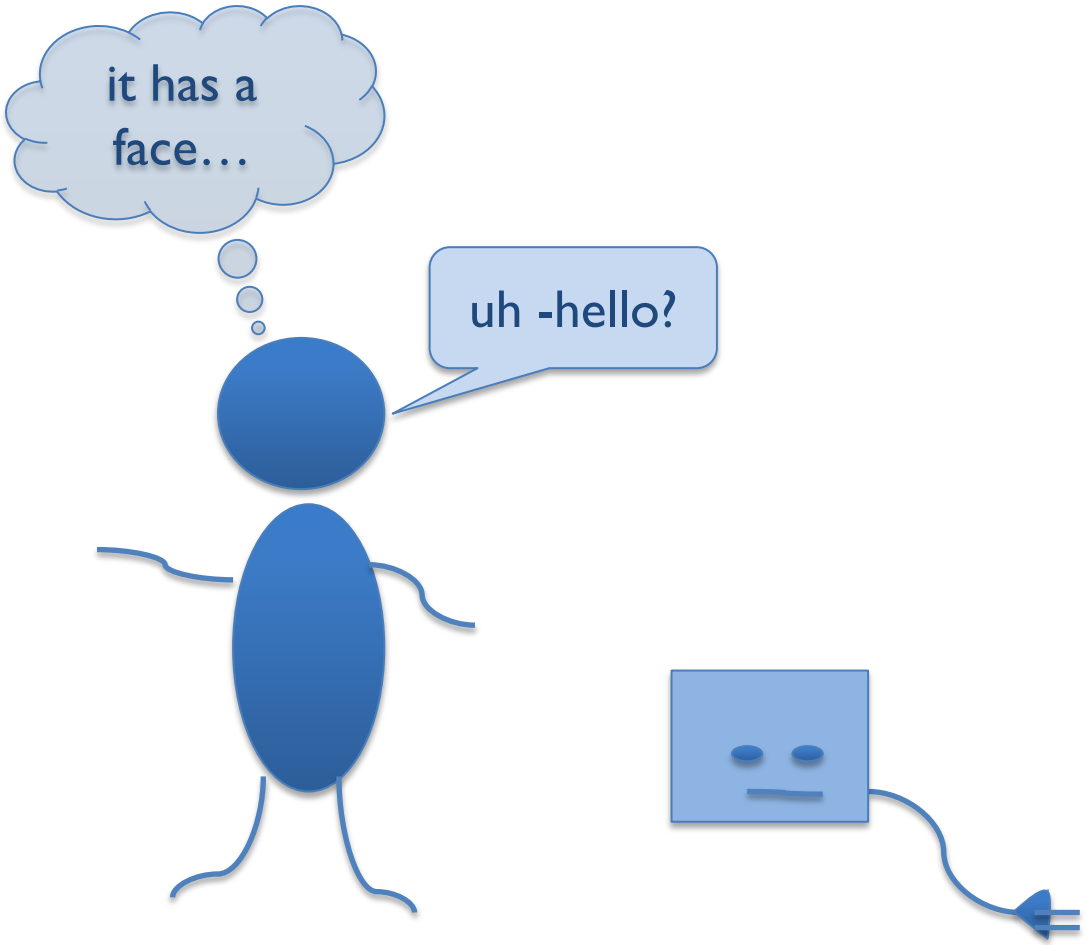
Partner Models



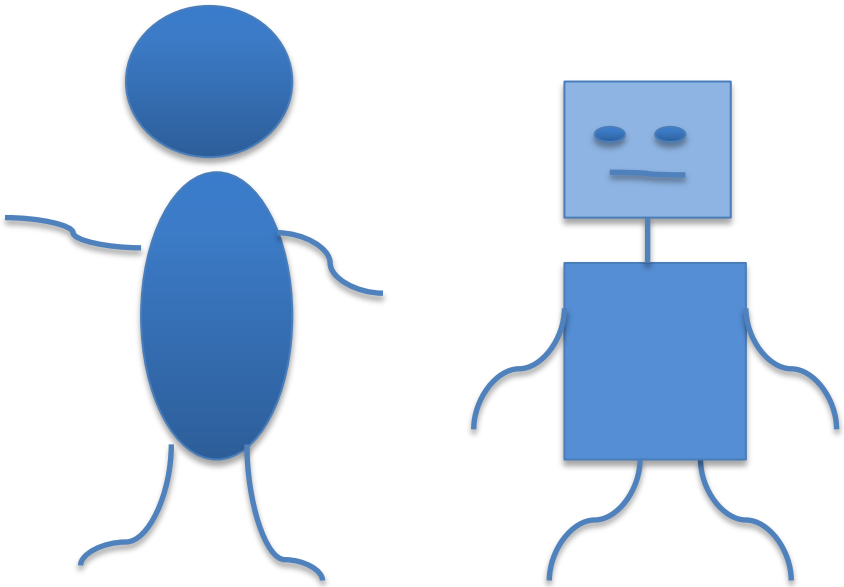
Partner Models



Partner Models



Communication with Robots



- appearance
- functionality
- whether it talks
- where the voice comes from
- what it says
- how it says what it says
- what it does
- when it does what it does
- …

Human-Robot Interaction

- in order to coordinate with each other, people build up partner models
- people have only vague ideas about robots and thus rely on every cue they can get
- How do we get people to interact with robots in ways that facilitate
  a) bootstrapping?
  b) communication?

Getting Attention

Getting Help

Getting Attention & Help

Social Framing vs Beep: Results

	‘speech’ condition N=8	‘beep’ condition N=11
average number of glances per participant (number of participants glancing towards the robot in %)	1.25 (85.7)	0.91 (45.5)
average length of longer looks to robot in seconds (number of participants looking longer towards the robot in %)	2.13 (100)	1.64 (81)
average total looking time to robot	51.2	29.2

Social Framing vs Beep: Results

- in the speech condition, people find the robot significantly 'more approachable'
- 100% participants react to the robot's speech, yet only 18.2% to the beep
- women are significantly more likely than men to use speech to the robot
- people look at the robot much longer if it is speaking

- ➢ whether or not the robot uses speech influences the degree with which people perceive the robot as a social interaction partner
- ➢ the robot's behavior contributes to people's partner models

Infant-Directed Speech Data

- 28 interactions between parents and infants
- 28 parents (i.e. mothers and fathers) and their 8-11 months old children (average: 10.25)

Robot-Directed Speech Data

- 30 human-robot interactions
- simulated robot
  - resembles baby
  - eye movement
  - attention to
    - movements
    - colours
    - skin

Tasks

- explain lamp
- show bell
- show putting on salt
- block world task
- putting rings in box
- stacking cups

Summary of Results

linguistic verbosity	CDS	=	HRI
linguistic diversity	CDS	<	HRI
linguistic complexity	CDS	<	HRI
linguistic interactivity	CDS	>	HRI
gesture	CDS	<	HRI

Summary of Results

linguistic verbosity	CDS	=	HRI
linguistic diversity	CDS	<	HRI
linguistic complexity	CDS	<	HRI
linguistic interactivity	CDS	>	HRI
gesture	CDS	<	HRI

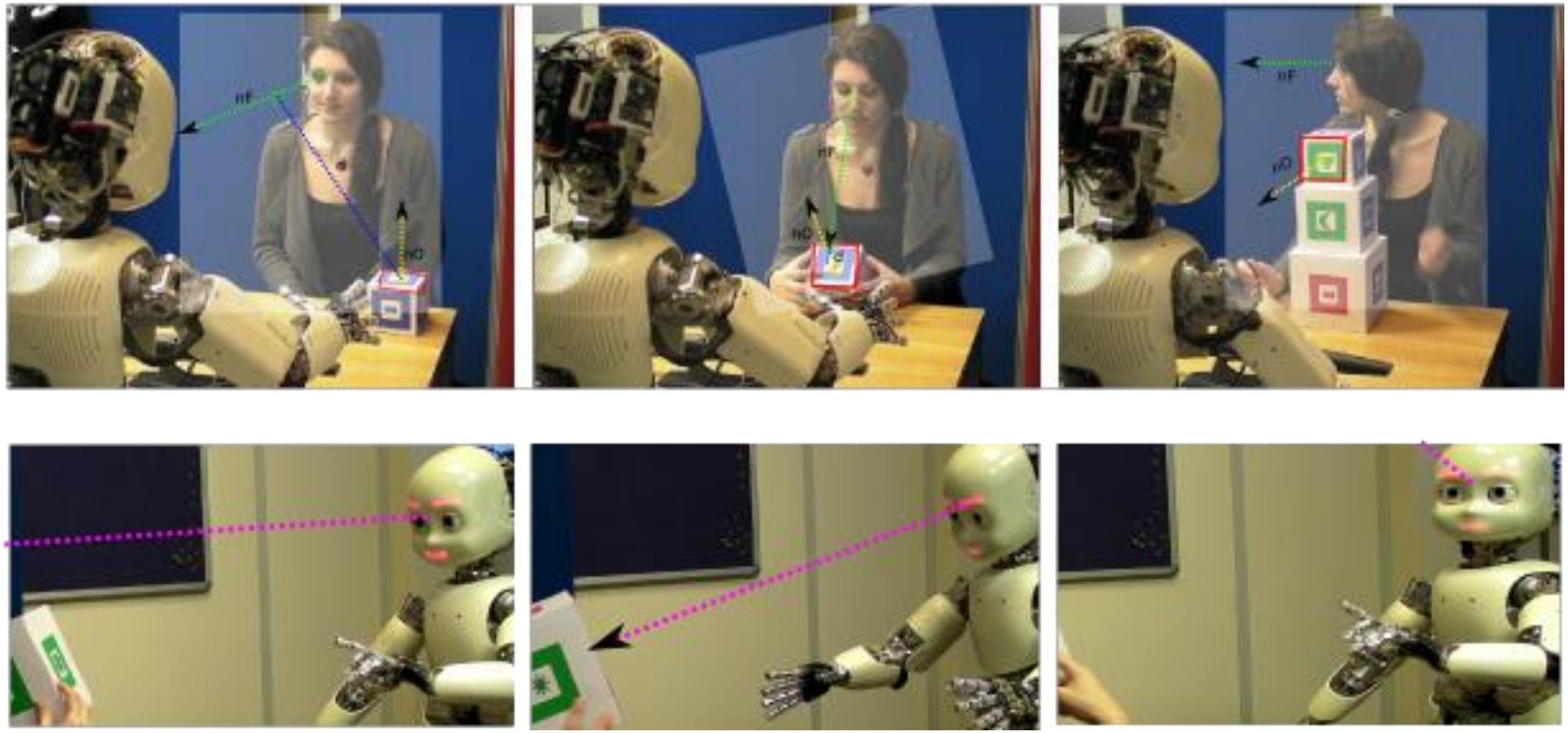
- people do not speak to a robot like to a child
- yet they adapt to what they have contingent feedback for!

The Role of Contingency

- Contingency:
  - socially contingent robot response
  - robot response contingent to object movement
- Robot feedback:
  - in sessions 2 and 3, the robot repeats the words it had learned in between

Contingency

- socially contingent robot response

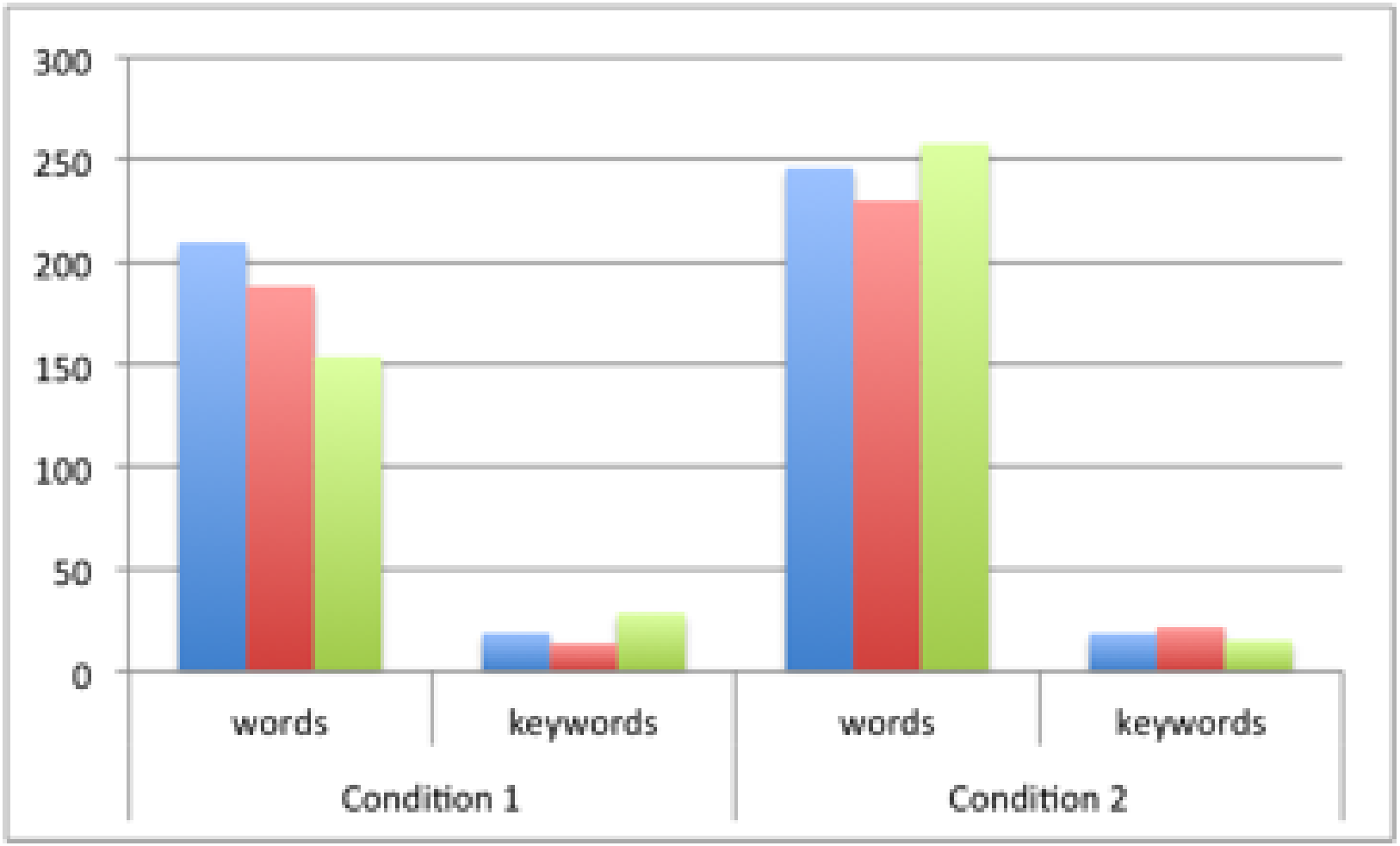


Example interactions

Contingency: Results

- contingent robot:
  - more turns
  - more structuring cues
  - more shorter utterances per turn
  - less diversity
  - fewer ‘I’, more ‘let’s’
  - more repetition, fewer references to past

  ✓ more tutoring for the socially contingent robot



- people adjust only to the socially contingent robot!

Conclusions

- in the human child, language and cognition co-develop
- caregivers highlight the connection between linguistic and cognitive distinctions
  - increasing transparency
  - increasing contingency between form and meaning
  - highlighting distinctions
- in order to interact, robots and humans have to co-ordinate their categories
- learning categories from humans is helpful
  - language carves out certain slices of experience, facilitating learning, category formation, memory, etc.

Conclusions

- human tutors intuitively provide useful cues to robots
- social framing increases people's attention to a robot gesturing for help
- while people do not automatically provide the robot with all helpful clues from child-directed interactions, contingent robot response makes people adjust to the robot considerably

One word on terminology…

- in the language acquisition literature, 'bootstrapping' is associated with
  - one particular approach proposed in the 1980ies
    - that relied exclusively on distributional learning over linguistic forms
- 'bootstrapping' here is used in the wide sense, as learning categories from sensorimotor or social experience
- the approaches to language acquisition presented here are usually treated under the labels of
  - construction grammar/constructionist approaches to language acquisition

Thank you!