Logic and higher-order social cognition

The facts matter, and so do computational models

Rineke Verbrugge
Institute of Artficial Intelligence
Faculty of Mathematics and Natural Sciences
University of Groningen

ESSLLI Workshop Logical Methods for Social Concepts
Bordeaux, 22 July 2009

Overview

- What is higher-order social cognition?
- The challenge: mixed multi-agent environments
- Some current cognitive science perspectives on social cognition
  - Pilot 1: Mastersminds
  - Pilot 2: Backward induction versus story-tasks
- Logical and computational models of higher-order social cognition
  - Vici project preview

Theory of mind

- Understand and predict external behavior by attributing internal mental states:
  - knowledge, beliefs, intentions, plans

Other people’s minds

- In daily life it is important to reason about others’ knowledge, beliefs, intentions.
- Cooperation:
  - Does he know that I intend to pass the ball to him, and not to Kluivert?
- Natural language interpretation and common knowledge
  - Can I felicitously refer to “the movie showing at the Roxy tonight”?
  - I did see him noticing the announcement in the afternoon paper, but maybe he does not know that I saw it, so maybe he does not know that I know that he knows that “the movie showing at the Roxy tonight” is “Monkey Business”. [Clark & Marshall]

Other people's minds

- Competition in card games:
  - I show her a card from which I believe that she can deduce as little new knowledge as possible.
  - Does she know that I know that she's bluffing (trying to make me believe she has more valuable cards than she in fact possesses)?

Other people’s minds

- The initial situation of **negotiation** is a conflict of interests, together with a need for cooperation.
- Main goal: to make a deal.

- Negotiation has elements of
  - cooperation: joint problem solving to find mutual gains, ‘enlarging the pie’
  - competition: dividing the pie

- Reasoning about knowledge and ignorance:
  - “I do not want the buyer to know that I am in a hurry with the sale because I already bought a new house”

Overview

- What is higher-order social cognition?
- The challenge: mixed multi-agent environments
- Some current cognitive science perspectives on social cognition
  - Pilot 1: Mastersminds
  - Pilot 2: Backward induction versus story-tasks
- Logical and computational models of higher-order social cognition
  - Project preview

Current multi-agent systems

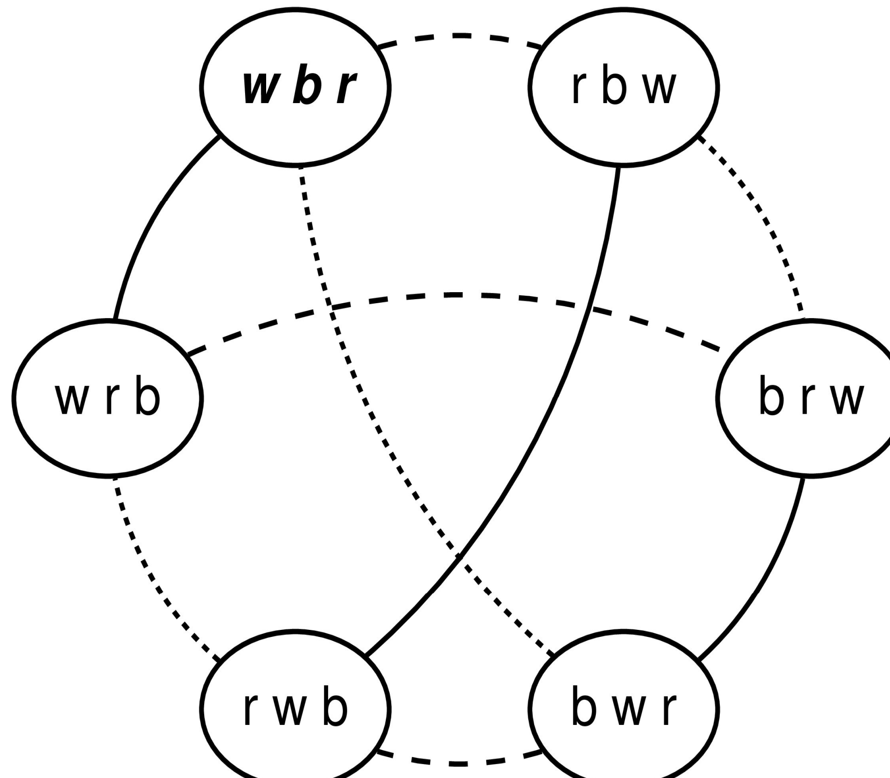
- Multi-agent system
  - cooperating computational systems
  - solve complex problems beyond expertise of individuals
- Applications
  - air traffic control
  - flexible car manufacturing control (Daimler-Chrysler)

Future multi-agent environments

- Trend
  - Mixed teams: robots, persons and software agents
  - Example: rescue systems after disasters
- Challenge
  - Current formal models of ‘ideal’ intelligent interaction
  - But human participants have bounded rationality
- Aim
  - Design improved intelligent interaction
  - Use strengths and weaknesses of different agent types
  - Investigate how agents learn complex interactions

Modal logics for multi-agent environments?

- Many modal logics for intelligent interaction place unrealistic assumptions on human reasoning
  - logical omniscience
  - positive and negative introspection
  - unbounded recursion



Logical omniscience

- If is A is true in all possible worlds, then everyone knows A
- Agents know all logical consequences of their knowledge

Introspection and transparency

$$K_i\varphi \rightarrow K_iK_i\varphi \text{ (positive introspection)}$$

$$\neg K_i\varphi \rightarrow K_i\neg K_i\varphi \text{ (negative introspection)}$$

$$K_iKw\varphi \rightarrow K_i\varphi \text{ (transparency)}$$

Unbounded recursion

- “I do not know whether Emiliano knows whether Andreas knows that I know that TARK will be held in the Netherlands in 2011.”

A puzzle: Sum and Product

The following is common knowledge:

- $x,y \in \mathrm{N}$ with $2 \leq x \leq y \leq 99$
- S and P are perfect at epistemic logic and arithmetic
- S knows the sum of x,y
- P knows the product of x,y

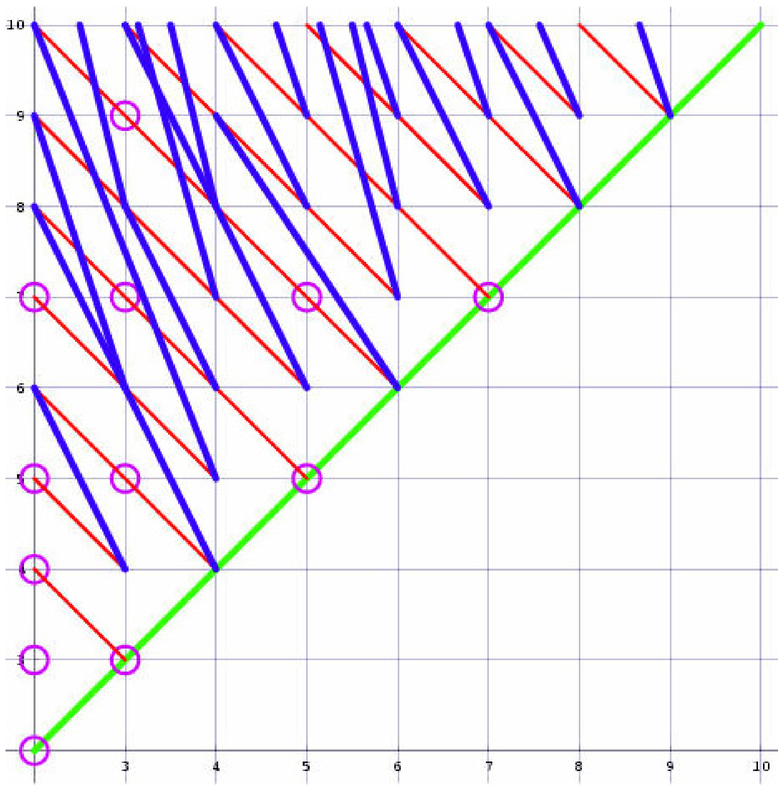
The following dialogue takes place:

1. P: I don't know the numbers.
2. S: I know you didn't know.
3. P: Now I know the numbers.
4. S: Now I know them, too.

Compute x and y!

Freudenthal, 1968 / McCarthy / Plaza / Panti

Sum & Product puzzle



1. P: I don't know the numbers

The Kripke model after 1: all product-isolated states can be deleted

Social cognition: how difficult?

- Introspection suffices to know that, at least sometimes, some people (logicians) can reason correctly at various orders of social cognition.   No amount of experimentation can deny this.
- Empirical research is needed for more general questions:
  - Under what circumstances do people engage in higher-order social cognition?
  - Do they apply it correctly?
  - Can they learn to apply it in unusual contexts?
- Some experimental findings indicate that the degree to which people correctly apply social cognition is rather less than is often assumed.

Overview

- What is higher-order social cognition?
- The challenge: mixed multi-agent environments
- Some current cognitive science perspectives on higher-order social cognition
  - Pilot 1: Mastersminds
  - Pilot 2: Backward induction versus story-tasks
- Logical and computational models of higher-order social cognition
  - Vici project preview

Logic and reasoning: do the facts matter? (1)

- "Advertising 'mismatches' between inferential predictions of logical systems, usually without proper attention to the modeling phase, and what is observed in experiments with human subjects seems entirely the wrong focus to me - not to mention the fact that it is silly and boring. The much more interesting issue is to avail ourselves of what is involved in how people really reason"

(Johan van Benthem, Logic and reasoning: do the facts matter? Studia Logics 88 (2008) pp. 67-84)

Cognitive aspects: first-order theory of mind

- Small infants reason about the other’s behavior and intentions
  (earlier than about other’s beliefs)
  - And even some apes and crows seem to be able do this (Call and Tomasello; Clayton)
  - But do they explicitly represent mental states, or do they simply follow ‘behavioral rules’? (Van der Vaart)
    - “If the other bird does not know that I’m storing food because it cannot see me, I can safely store my food here” OR
    - “If I am situated far from other birds, I can safely store my food here”

Cognitive aspects: first-order theory of mind

- By age 4, the ability to distinguish between one's own and others' beliefs is firmly in place.
- Experiment with reflective 'false-belief' task [Wimmer & Perner, *Cognition*, 1983]
  - "Maxi left chocolate in blue cupboard, then left the room. In Maxi's absence, his mother moved the chocolate to the green cupboard."
  - "Where will Maxi look for the chocolate first?"
  - 3 year old thought Maxi would later look for the chocolate in the green cupboard (confusing Max's belief with her own).
  - 5 year old thought Max would follow his own, false, belief.

Cognitive aspects: first-order theory of mind

- Keysar, Lin & Barr [*Cognition*, 2003]: Though normally-developed adults can reflectively distinguish their own beliefs from others', this ability does not always allow spontaneous, non-reflective use.
- Experiment with non-reflective task
  - Grid of objects, among which a cassette tape, visible to both director and participant.
  - Participant has one object hidden in a bag: a roll of tape.
  - When director said "Move the tape":
    - 46% moved the bag in most cases;
    - 71% attempted to move the bag at least once;
    - 82% were delayed in identifying the intended object (eye-fixation on "wrong" object) .

Research questions

- Is proficiency in social cognition in a strategic game related to proficiency in social cognition as evidenced in language use (pragmatic/logical)?
- Do people learn to apply higher-order social cognition when this is profitable during strategic game playing?

Pragmatic inference

- Grice's maxim of quantity: Make your contribution as informative as required, but not more so.
- Scalar implicatures involve expressions that can be ordered on a scale of informativity determined by entailment relations, such as <some, most, all>.
    - "Some Indian logicians like Kulfi" does not logically imply "Not all Indian logicians like Kulfi".
    - Apparently, a speaker's use of *some* indicates that he had reasons not to use the more informative terms from the same scale, so *some* pragmatically implicates *not all*.
    - Such reasoning about interlocutor requires 2^{nd} order social cognition

Pragmatic inference and social cognition

Non-cooperative situations require truth-functional, non-pragmatic productions or interpretations

- Suppose in happy families ('kwartetten') you ask Max for "Gödel" of the family "famous logicians". Max replies "No, I don't have it".
- Because of his desire to win, Max does not want you to know which cards he has. Thus, you can not infer that Max does not have any member of the famous logicians family.

The Masterminds experiment

- Participants play a symmetric version of Mastermind:

-A and B both take a secret code (4 different, ordered colors), e.g.:

1: red, 2: blue, 3: grey, 4: black

-Both take turns guessing the opponent's secret code.

-When A guesses, B gives feedback on A's guess w.r.t. B's code, and A gives feedback on his own guess w.r.t his own code, selecting e.g.:

- 1,2,3,4 a / some / most / all colors are right
- 1,2,3,4 a / some / most / all color(s) are in the right place

-They give the experimenters their interpretations of the other's feedback, in terms of situations they consider possible.

-They answer questions about their strategy, and complete a questionnaire afterwards.

Verbrugge & Mol: Learning to apply theory of mind. *JoLLI*, 2008.

Mastersminds experiment: results

- 4 out of 12 participants used 2-nd order social cognition (none higher); These
  - used a strategy of being as uninformative as possible;
  - used a strictly logical interpretation of the sentences.
- The other 8 used 1-st order social cognition. Of these players:
  - 2 used a strategy of being as uninformative as possible, and a fairly logical interpretation;
  - 6 used a strategy of being informative or did not consider the information being revealed, and used a mostly pragmatic interpretation of the sentences.
- Changes over time:
  - All 4 participants using 2-nd order social cognition did so from the start;
  - 4 other players shifted between pragmatic and logical interpretation.
  - 1 player shifted from being uninformative to being informative, 1 (to give the opponent a better chance of winning!)

Mastersminds experiment: results

- Hypothesis was: In an uncooperative conversation, people will shift their interpretation and production from pragmatic (Grice’s quantity maxim) to non-pragmatic use.
- Falsified: none of the participants developed a more truth-functional language use.

Design of a new experiment: separating the tasks

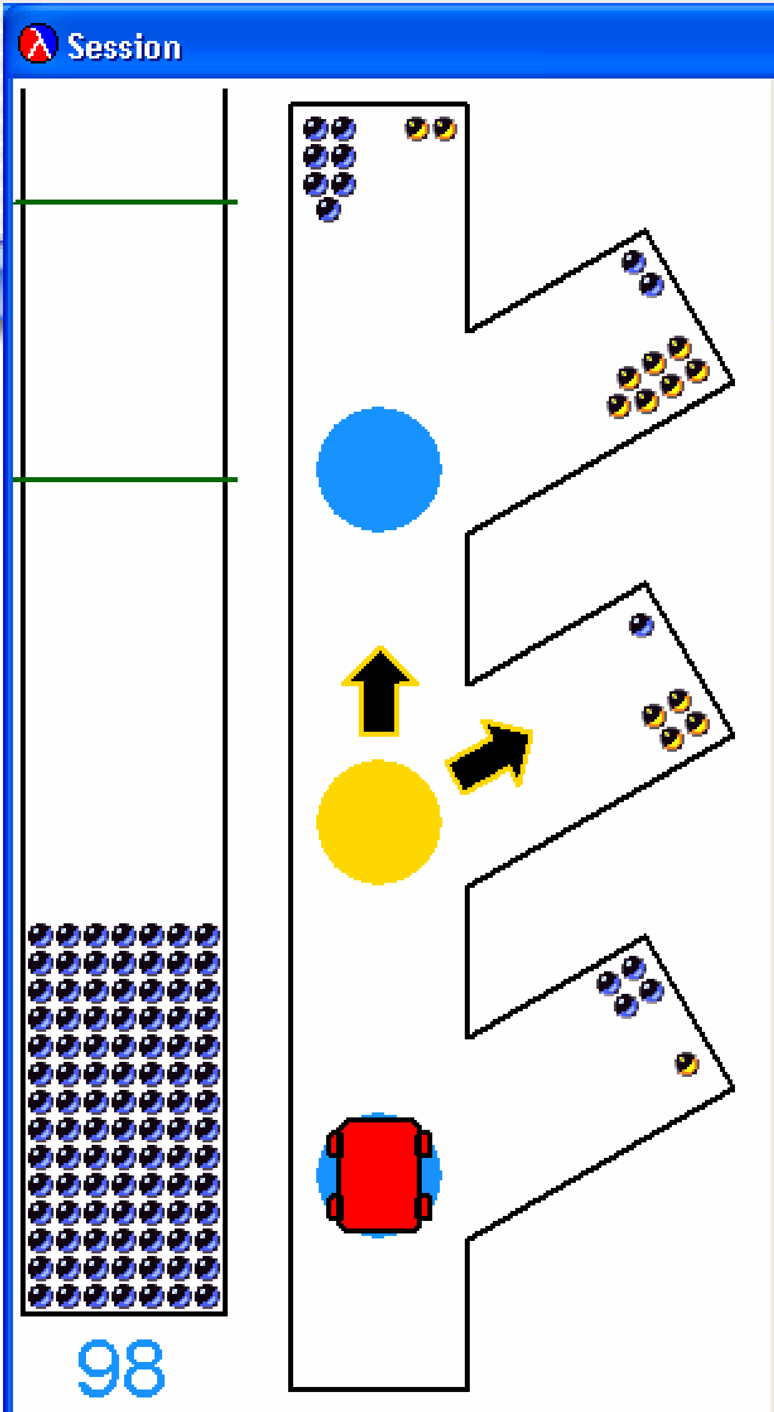
- Two groups: adults (n = 27) and 8-10 year old children (n = 40)
- Two tests (within subjects):
  - A strategic game, based on Hedden & Zhang's backward induction experiment (2002)
  - A verbal second-order false belief task
  - Flobbe, Verbrugge, Hendriks & Krämer: Children's application of theory of mind in reasoning and language. *JoLLI*, 2008.

The backward induction game

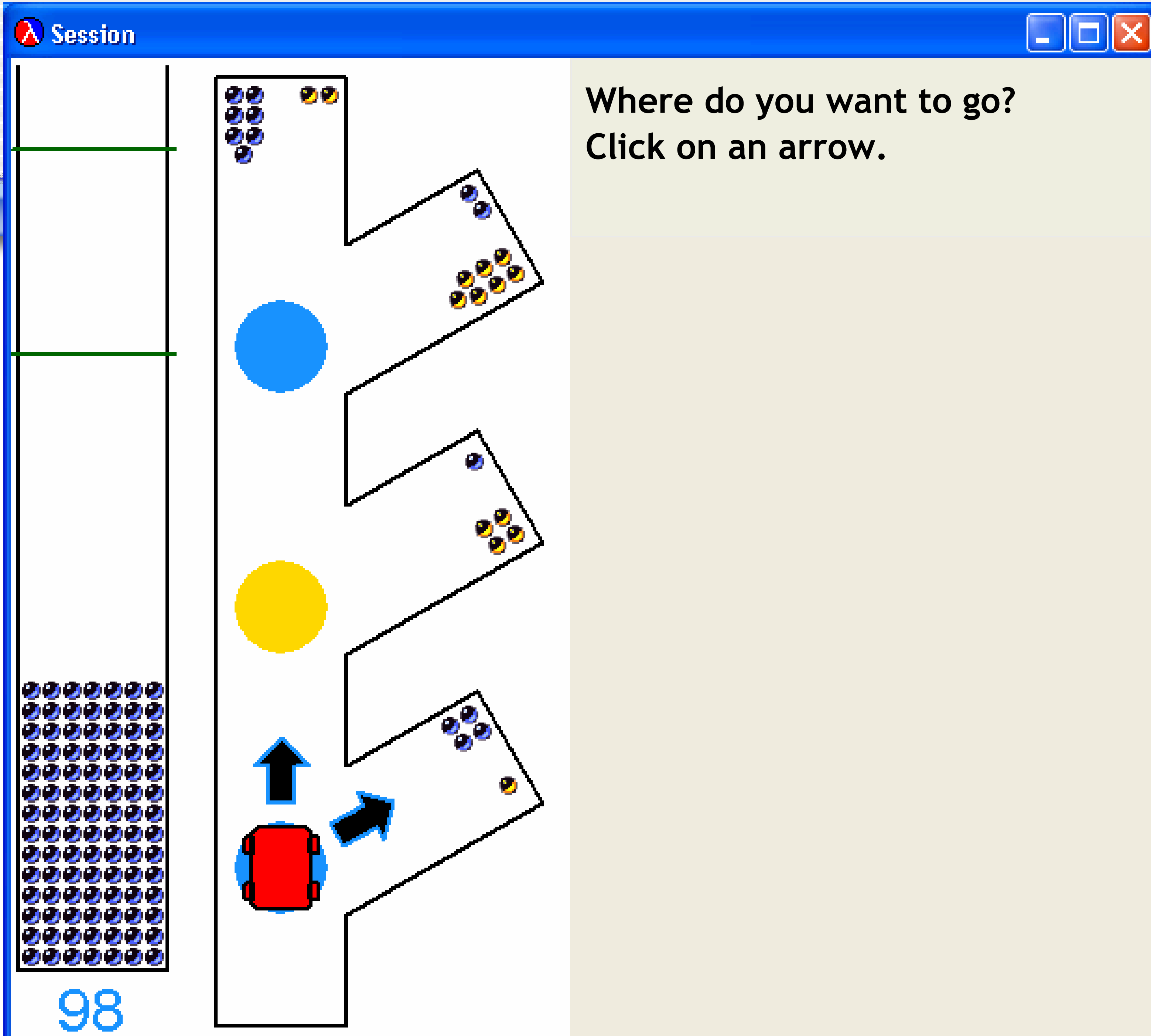
- Non-cooperative sequential game with two players
- The game can end in one of 4 cells
- Each cell contains a payoff for each player
- Each player tries to maximize his own payoff
- Optimal decisions require second order reasoning

Game design

- The structure and rules of the game should be understandable to children
- Payoffs (rewards) should be ‘real’
- Children are encouraged to be
  - Egoistic (get as many marbles as they can),
  - not competitive (get more marbles than opponent)



What do you think YELLOW will do if the car reaches the yellow t-section? Click on an arrow.


Session
98
Where do you want to go?
Click on an arrow.

Order of reasoning in the game

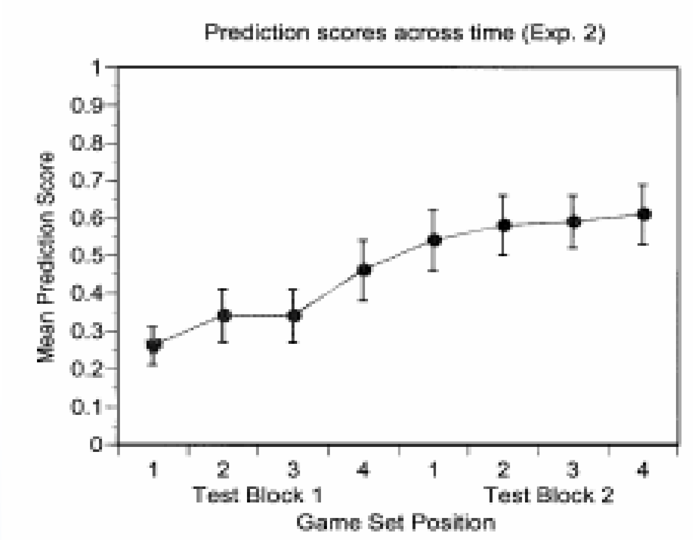
- **Last intersection** - the child only needs to consider his own interests: **no social cognition**
- **Middle intersection** - the opponent considers the child’s interest at the last intersection: **first-order social cognition**
- **First intersection** – the child must consider the opponent’s interests and the opponent’s model of his own interests: **second-order social cognition**

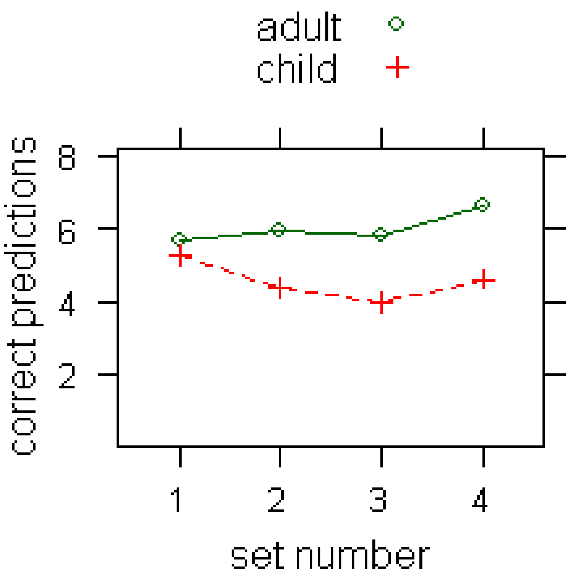
2nd order reasoning in the game

H&Z: adults start with a default first-order strategy and gradually adopt a second order strategy.

Our results: people use second-order reasoning from the start.

Little effect of length of exposure to the game.





2nd order reasoning in the game

- Adults score better than children.
- Our adults do better than in Hedden & Zhang, but far from perfect.
- Both groups score above chance level.
- Both groups reliably use their predictions in choosing their strategy: after a correct prediction only 4.8% of adult moves and 15.5% of child moves are incorrect.

Conclusion on backward induction experiment

- Most 8-10 year old Dutch children ‘pass’ a second-order false belief story task.
- Adults do better at the game than children, so applied social cognition-reasoning continues to develop after ‘passing’ the false belief story task.
- Performance does not improve much during the game. People who use second-order reasoning, do so from the start of the game.
- Even adult performance is far from perfect.
- Considerable differences between individuals.

Conclusions on experiments

- We did *not* find evidence that people learn to apply higher-order social cognition during a strategic game, even if it is profitable.
- Possibly, proficiency in social cognition for developing a successful game strategy and for effective language use are related.

Overview

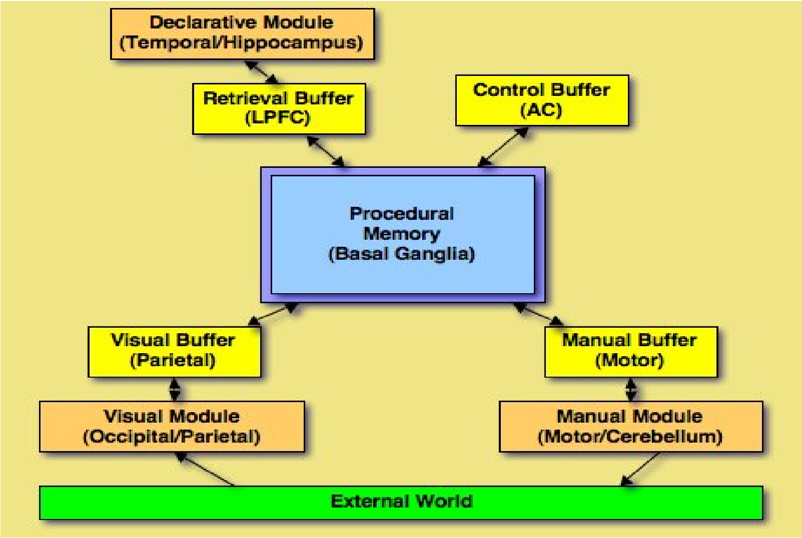
- What is higher-order social cognition?
- The challenge: mixed multi-agent environments
- Some current cognitive science perspectives on higher-order social cognition
  - Pilot 1: Mastersminds
  - Pilot 2: Backward induction versus story-tasks
- Logical and computational models of higher-order social cognition
  - Vici project preview

Project
Cognitive systems in interaction

- To investigate human powers and limitations
  - Empirical research: does proficiency in social cognition transfer between tasks?
  - Where are the bottlenecks from 1 to 2-order (and further)?
  - Computational cognitive modeling: how do we learn higher-order social cognition?
  - What mechanisms in the brain correspond to higher-order social cognition?
- Computational simulation: how did social cognition evolve?
- Design logics appropriate for resource-bounded agents
- Design computational systems supporting people in tasks that require higher-order social cognition (such as negotiation)

Research method: Computational cognitive models

- Integrated cognitive model of social cognition
  - Why is higher-order social cognition so hard to learn and apply?
    - Working memory (fMRI)
  - What kinds of support would help?



ACT-R: computational cognitive architecture

Develop new type of knowledge rules for social cognition

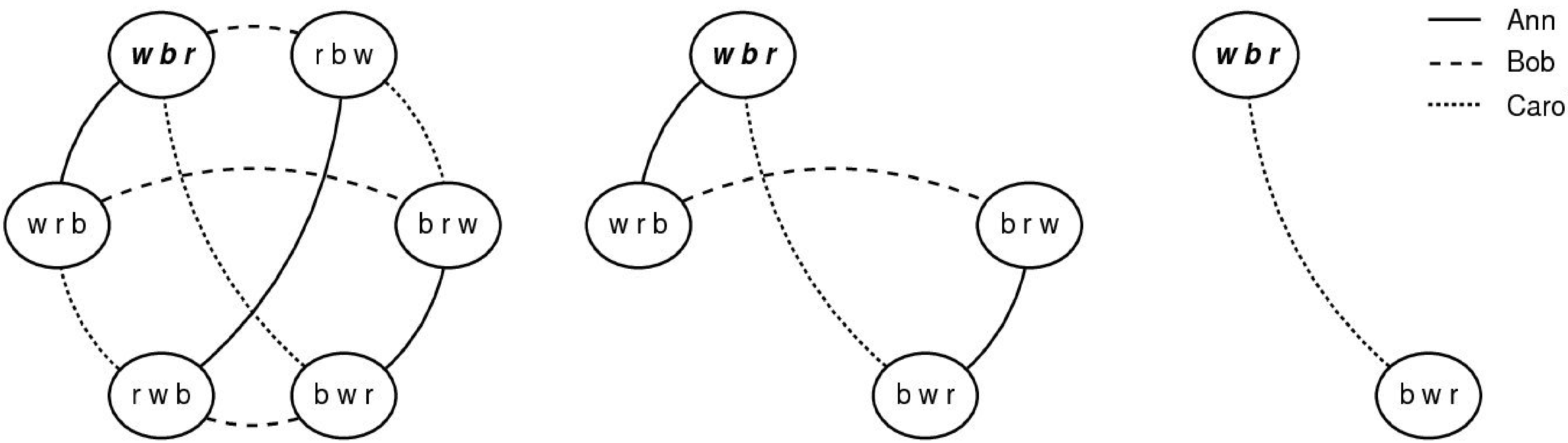
Research method: Logical models

- Formalize higher-order cognition for **different agent** types
  - Parametrize resource-bounded logics by capabilities
  - Inference, reflection, recursion, revision

- Extend with realistic component for **group reasoning**
  - Common belief
  - Common knowledge
  - Collective intention
  - Collective commitment

Research method: Logical models

Construct **dynamic** logic model of changing states in mixed teams



Ann: "Bob, do you have red?"

Bob:"No"

Research method: Agent-based models

- Construct realistic agent-based simulations of **evolution** of higher-order social cognition
  - Choose logical representation
  - Extend agent-based modeling by evolving rule forms
    - Genetic algorithm selects strategies: highest expected pay-off
  - For which environments & tasks is higher-order reasoning adaptive?

At least one *VisibilityClause:*

if [Object x] [is / isn't] visible

Zero or more *ConditionClauses:*

and [Object x] [distance-to / angle-to] [Object y] is [more-than / less-than] [value]

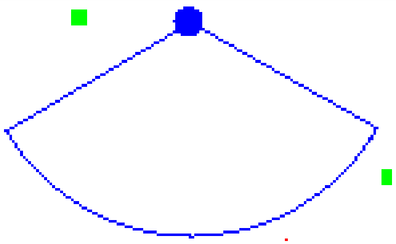
Exactly one *ActionClause:*

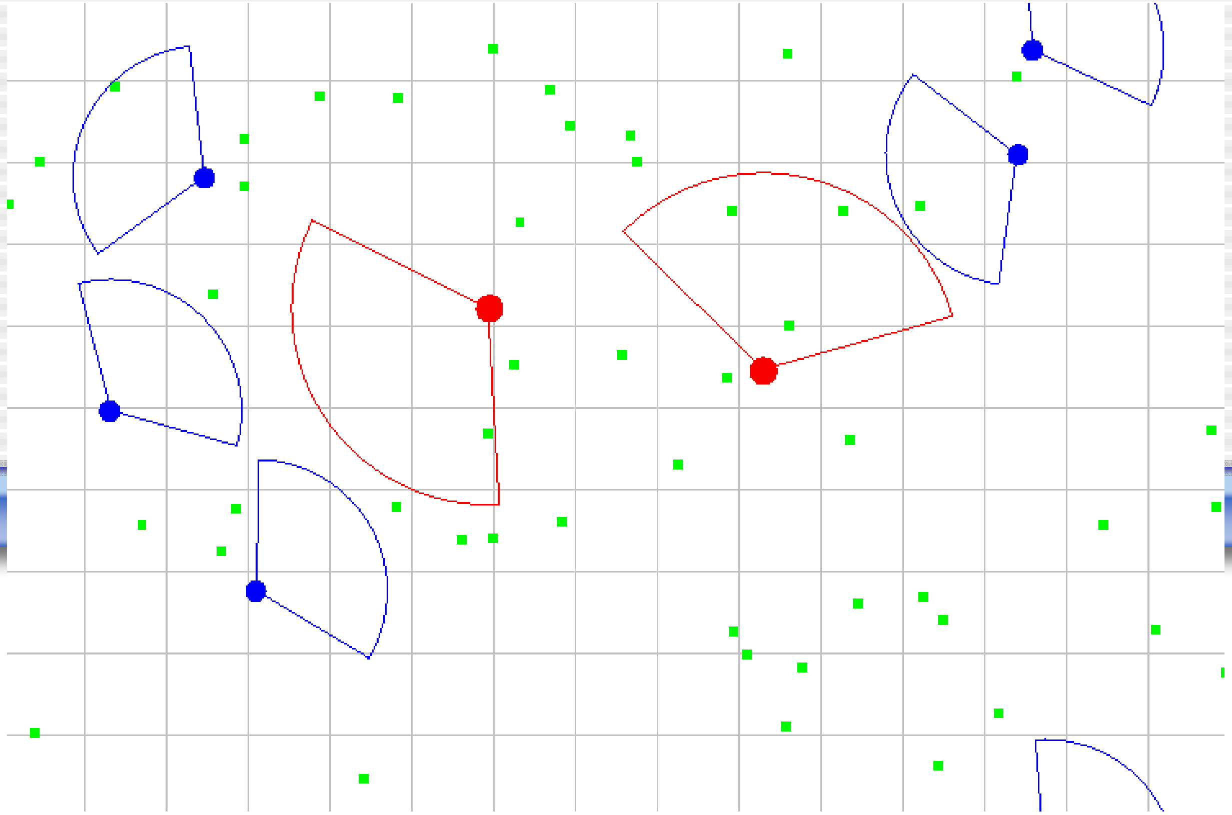
then [hide / search] *or* [move / turn] [value] *or* [move-to / turn-to] [Object x]

IF PREDATOR VISIBLE
THEN HIDE

IF FOOD [1] VISIBLE
THEN GO-TO FOOD [1]

IF FOOD ! VISIBLE
THEN SEARCH





Combining the research methods: Mixed human-computer teams

- Intelligent agents that support mixed human-computer teams
  - For negotiations during teamwork
  - Based on a combination of the developed
    - logics
    - cognitive models
    - agent-based models

- Roles of logic in the cognitive science experiments:
  - Precise definitions: (levels of knowledge and group action, Parikh 2003)
  - Form of rules for social cognition in computational cognitive models and agent-based models

Conclusions

- Current formal models of intelligent interaction not suited for modeling cooperation in mixed teams of people with software agents and /or robots
- We aim to develop **logical theory** and **computer models** for
  - improved understanding of human social reasoning
  - implementation in computer systems that support mixed teams

Advertisement

- 5 researchers will be part of an interdisciplinary project team:
  - A cognitive science PhD students and postdoc start Summer 2009
    - (behavioral experiments with children and adults, fMRI scans, ACT-R models)
  - One postdoc logician (plm. 4 years) will be hired in 2010
  - Two artificial intelligence / computer science PhD students will be hired in about a year
    - (agent-based models; support software for teamwork)

Logic and reasoning: do the facts matter? (2)

- "Indeed, the logical theories of inference, update, and interaction all suggest interesting testable hypotheses about human behaviour, and one could easily imagine a world where a logician who has created a new logical system does two things instead of one: like now, submit to a logic conference, usually far abroad, but also: telephone the psychologist next door to see if some nice new experiment can be done.

  And finally, going a bit further, I would think that logic can also contribute to a better understanding of how humans form and maintain representations of scenarios and their relevant information, the stage prior to any significant processing. What this would involve is a broadening of current 'model theory' to a 'theory of modeling'.

(Johan van Benthem, Logic and reasoning: do the facts matter? Studia Logics 88 (2008) pp. 67-84)