Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

HRI: Cognitive Models and The Theory of Mind

Nikoletta Xirakia

T|A
M|S

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik

Technische Aspekte Multimodaler Systeme

28. November 2016

Outline

Motivation

Achieving a successful Human-Robot Interaction requires both partners to have sufficient perception of each other's actions.

For Humans:

- Appearance of robot
- Communication

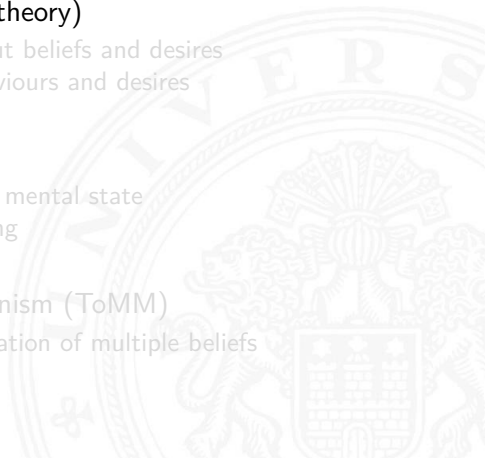For Robots:

- Attribute mental states (ToM)
- Understand humans

[6]

Theory of Mind

The Theory of Mind (ToM) is the ability to attribute mental states, to self or others: [14]

- ▶ Beliefs
- ▶ Intentions and desires
- ▶ Thoughts
- ▶ Emotions

Key capability for:

- ▶ Cognitive development
- ▶ Social Interaction

[7]

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
  - ▶ Set of laws, theories about beliefs and desires
  - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
  - ▶ Representation of others' mental state
  - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
  - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
    - ▶ Set of laws, theories about beliefs and desires
    - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
    - ▶ Representation of others' mental state
    - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
    - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
    - ▶ Set of laws, theories about beliefs and desires
    - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
    - ▶ Representation of others' mental state
    - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
    - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
    - ▶ Set of laws, theories about beliefs and desires
    - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
    - ▶ Representation of others' mental state
    - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
    - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
    - ▶ Set of laws, theories about beliefs and desires
    - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
    - ▶ Representation of others' mental state
    - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
    - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
  - ▶ Set of laws, theories about beliefs and desires
  - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
  - ▶ Representation of others' mental state
  - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
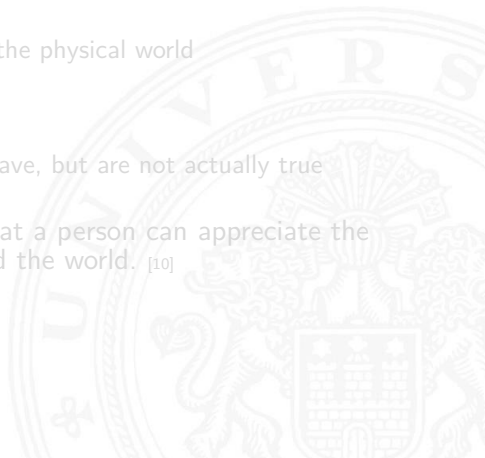 - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
    - ▶ Set of laws, theories about beliefs and desires
    - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
    - ▶ Representation of others' mental state
    - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
    - ▶ Generation and representation of multiple beliefs

Model of ToM

Competing views in the context of "belief and desire" reasoning:

- ▶ Conceptual change (theory-theory)
    - ▶ Set of laws, theories about beliefs and desires
    - ▶ Explain and predict behaviours and desires

- ▶ Simulation Theory
    - ▶ Representation of others' mental state
    - ▶ Using own decision-making

- ▶ The Theory of Mind Mechanism (ToMM)
    - ▶ Generation and representation of multiple beliefs

Ability to distinct between true and false beliefs:

▶ True-beliefs (TB)

  ▶ Beliefs which are true in the physical world

▶ False-beliefs (FB)

  ▶ Beliefs that others may have, but are not actually true

This ability indicates evidence that a person can appreciate the distinction between the mind and the world. [10]
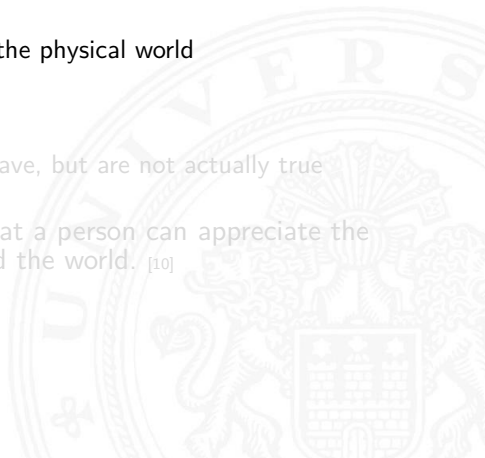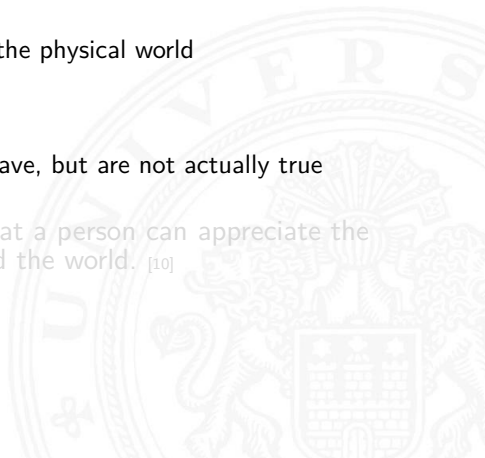
True & False beliefs

Ability to distinct between true and false beliefs:

▶ True-beliefs (TB)
  ▶ Beliefs which are true in the physical world

▶ False-beliefs (FB)
  ▶ Beliefs that others may have, but are not actually true

This ability indicates evidence that a person can appreciate the distinction between the mind and the world. [10]
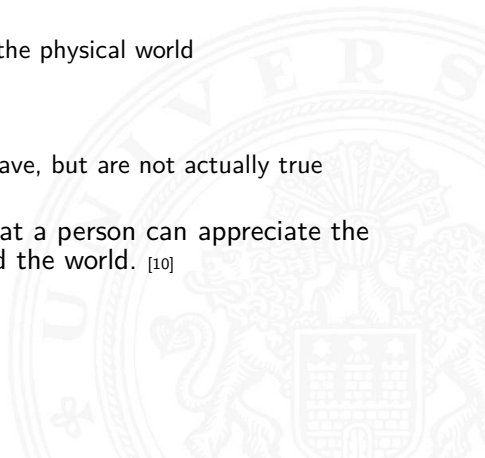
True & False beliefs

Ability to distinct between true and false beliefs:

- ▶ True-beliefs (TB)
    - ▶ Beliefs which are true in the physical world

- ▶ False-beliefs (FB)
    - ▶ Beliefs that others may have, but are not actually true

This ability indicates evidence that a person can appreciate the distinction between the mind and the world. [10]

True & False beliefs

Ability to distinct between true and false beliefs:

- ▶ True-beliefs (TB)
    - ▶ Beliefs which are true in the physical world

- ▶ False-beliefs (FB)
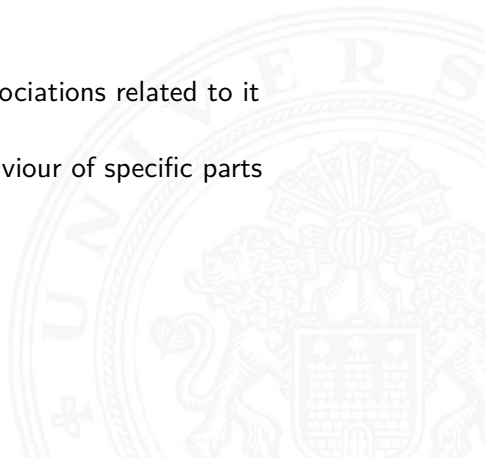   - ▶ Beliefs that others may have, but are not actually true

This ability indicates evidence that a person can appreciate the distinction between the mind and the world. [10]

True & False beliefs

Ability to distinct between true and false beliefs:

- ► True-beliefs (TB)
    - ► Beliefs which are true in the physical world

- ► False-beliefs (FB)
    - ► Beliefs that others may have, but are not actually true

This ability indicates evidence that a person can appreciate the distinction between the mind and the world. [10]

Humans' perception

The form of an object conveys information:

- ▶ Use and functionalities

- ▶ Symbolic information to associations related to it

- ▶ Aesthetics indicate the behaviour of specific parts

Anthropomorphism

Anthropomorphism, is understood as the attribution of humanlike properties or characteristics to real or imagined non-human agents and objects. [15]

[1]

Factors affecting humanlikeness of robots:
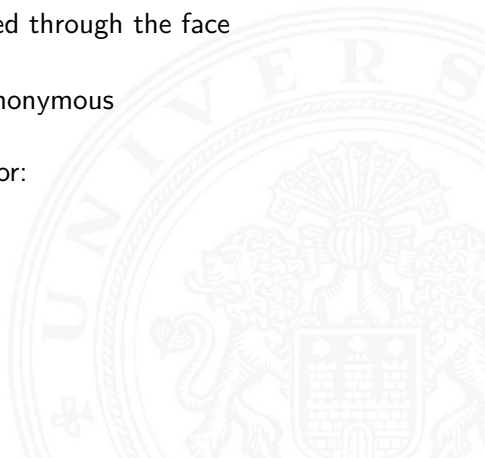
- ▶ Embodiment

- ▶ Verbal communication

- ▶ Emotions

- ▶ Gestures

eMuu [2]

Anthropomorphism

The importance of the robot's head in HRI:

▶ Non-verbal cues are mediated through the face

▶ Without a face a robot is anonymous

▶ Physiognomy is important for:

     ▶ Humanlikeness

     ▶ Knowledge

     ▶ Sociability

Uncanny Valley

The terror from something externally alien or un-known, but also strangely familiar. [16]

- ▶ Perception of an object with human characteristics

- ▶ Rationalisation of its actions and appearance

- ▶ Development of empathy

[13]

The Prisoner's Dilemma Experiment

Evoke ToM and monitor brain activation while playing the
Prisoners' Dilemma with different partners.

Participants:

▶ Computer (CP)

▶ Functional Robot (FR)

▶ Anthropomorphic robot (AR)

▶ Human confederate (HP)
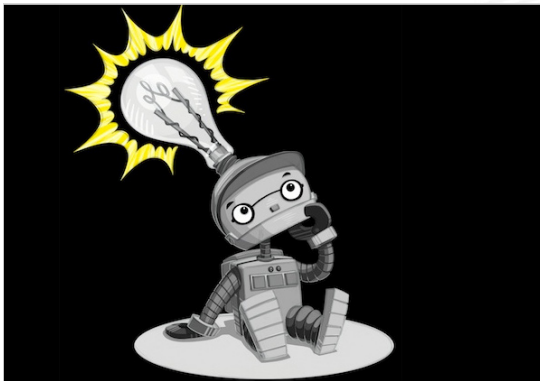

[12]

[3]

Anthropomorphic Robot

Functional Robot

Briefing

[3]
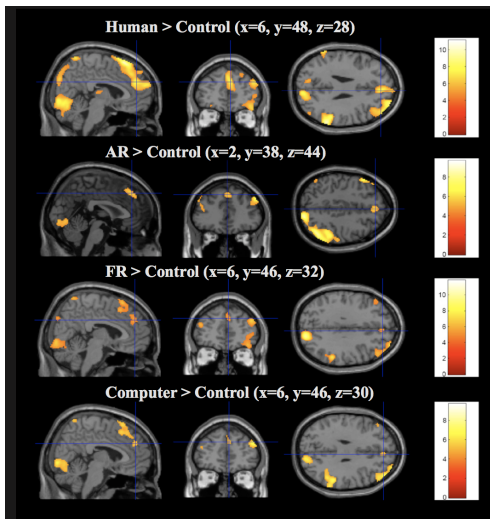
One of the most important milestones in the development of ToM, is gaining the ability to attribute a false belief task.
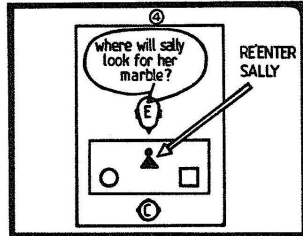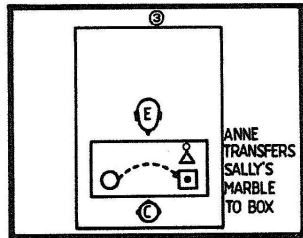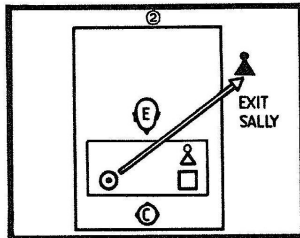
But can a cognitive agent do that ?

[8]
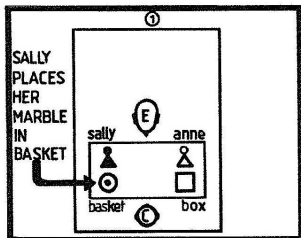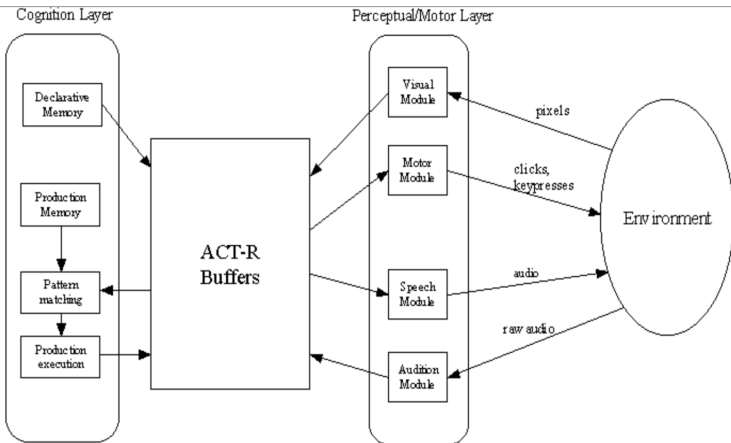
[4]

Cognitive agent model

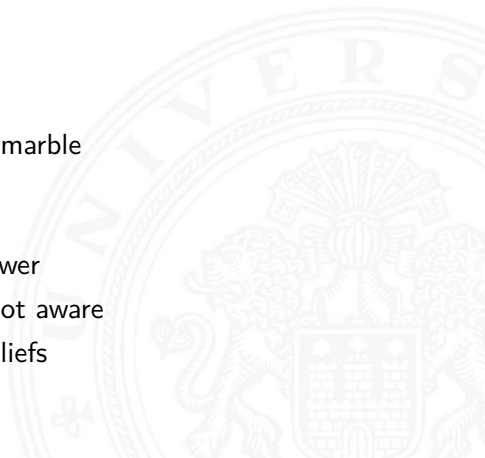ACT-R as core cognitive architecture:

[9]

Stage simulates the Sally-Anne task, which feeds the model with visual information.

The agent records:

- ▶ What happened?
- ▶ Who saw it happen?
- ▶ The current location of the marble

False-belief question:

- ▶ Highest activation - TB answer
- ▶ Answer incorrect - Sally is not aware
- ▶ Consideration of possible beliefs

Developmental Mechanisms

Learning mechanism:

- ▶ Beginning: Answer based on the highest activation chunk

- ▶ Learning: Considers if Sally knows about the belief

- ▶ Maturation parameter:
    - ▶ Gradual
    - ▶ Guideline on the strength of model's abilities

- ▶ Selection parameter:
    - ▶ Determines availability of productions (beliefs)
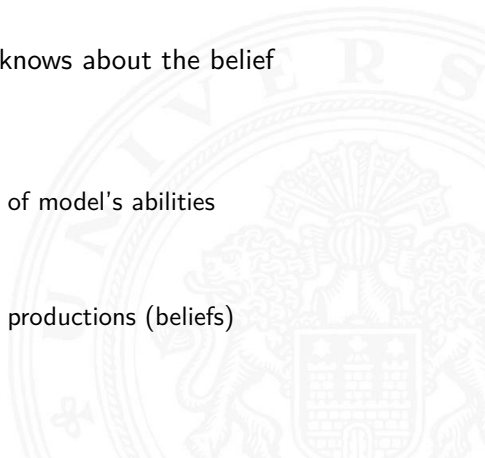
▶ Model begins to generate multiple beliefs at the age of 2 (approximately)
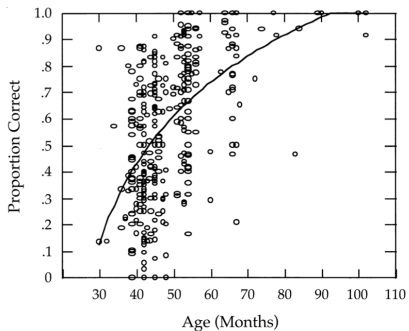
▶ For selection parameter of 0.5 , does not know to do the selection

▶ By the age 3.7 years , the selection parameter is up to 0.8

▶ By the age 5.7 years , the selection parameter equals 0.95

▶ As the selection parameter increases, so does the efficacy of learning

Results

Children [10]

Agent [11]

Conclusion
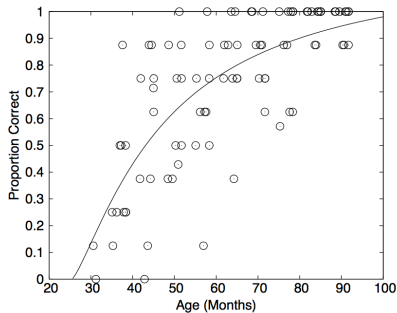
▶ Anthropomorphism and embodiment have significant effect on the human interaction, as partners expect a human like behaviour

▶ Humans evoke stronger emotional responses to embodied agents, as they prove to be able to physically manipulate the environment

▶ ToM is developed by concurrent learning and maturation and it's achievable in cognitive agents

▶ The model proved to be a good match to existing data from developing children, which enables such data to be used for further research on other aspects of ToM and HRI

References

1. Zlotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction. International Journal of Social Robotics, 7(3), 347-360

2. Bartneck, C., Okada, M. (2001), eMuu An Emotional Robotic User Interface, Philips User Interface Conference (UI2001), Eindhoven

3. F. Hegel, S. Krach, T. Kircher, B. Wrede and G. Sagerer, "Theory of Mind (ToM) on robots: A functional neuroimaging study,"2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Amsterdam, 2008, pp. 335-342.

4. Baron-Cohen, S., Leslie, A. M., Frith, U. (1985). Does the autistic child have a "theory of mind"?. Cognition 21:37-46, p.42

5. Zlotowski, J., Strasser, E., & Bartneck, C. (2014). Dimensions of Anthropomorphism - From Humanness to Humanlikeness. Proceedings of the ACM / IEEE International Conference on Human-Robot Interaction, Bielefeld pp. 66-73. | DOI: 10.1145/2559636.2559679

References

6. http://www.redorbit.com/news/technology/1112892978/robotics-perception-key-to-interaction-070913/

7. http://theautismblog.seattlechildrens.org/autism-theory-mind/

8. https://laughingsquid.com/2nd-annual-robot-film-festival-in-new-york-city/

9. https://www.cs.umd.edu/class/fall2002/cmsc838s/tichi/actr.html

10. Wellman, H. W., Cross, D., & Watson, J. (2001). Meta- analysis of theory-of-mind development: The truth about false belief. Child Development, 72(3), 655-684

11. L. M. Hiatt and J. G. Trafton. A cognitive model of theory of mind. In Proceedings of ICCM, 2010.

12. https://www.youtube.com/watch?v=t9Lo2fgxWHw

13. http://www.animatorisland.com/the-uncanny-valley/

14. Dennett, D.C., 1980. The milk of human intentionality (commentary on Searle). Behav. Brain Sci. 3, 428?430.

15. N. Epley, A. Waytz, and J. T. Cacioppo. On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4):864?886, Oct. 2007.

16. S. Freud (1953). ?The Uncanny?. The Standard Edition of the Complete Psychological Works of Sigmund Freud, ed & trs. James Strachey, vol. XVII (London: Hogarth), pp. 219-252.

Further reading

1. Sandoval, E. B., Brandstetter, J., Yacil, U., & Bartneck, C. (2016). Can a Robot Bribe a Human? The Measurement of the Dark Side of Reciprocity in Human Robot Interaction. Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction, Christchurch pp. 117 - 124. | DOI: 10.1109/HRI.2016.7451742

2. Zlotowski, J., Sumioka, H., Nishio, S., Glas, D., Bartneck, C., & Ishiguro, H. (2015). Persistence of the Uncanny Valley: the Influence of Repeated Interactions and a Robot?s Attitude on Its Perception. Frontiers in Cognitive Science, 6(883). | DOI: 10.3389/fpsyg.2015.00883

3. Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2015). Reciprocity in Human Robot Interaction ? A Quantitative Approach Through The Prisoner?s Dilemma And The Ultimatum Game. International Journal on Social Robotics 8(2), pp 303-317. | DOI: 10.1007/s12369-015-0323-x

4. Bartneck, C., Hoek, M. v. d., Mubin, O., & Mahmud, A. A. (2007). ?Daisy, Daisy, Give me your answer do!? - Switching off a robot. Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction, Washington DC pp. 217 - 222.

14