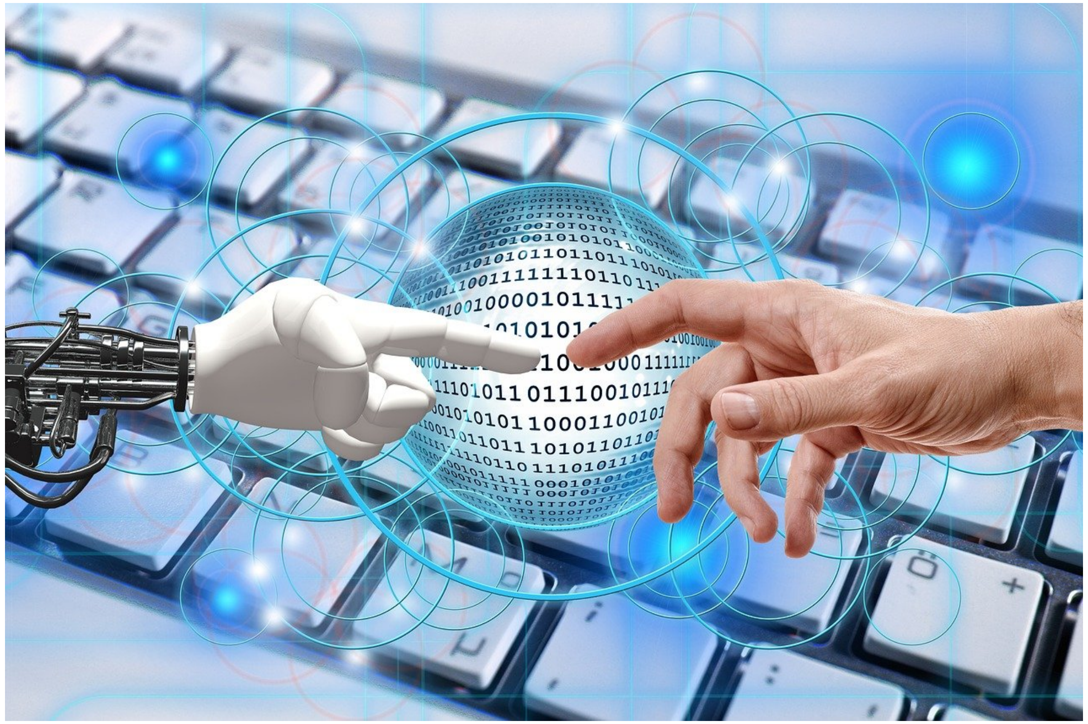
Building Socially Interactive Robots



Prof. Dr. Teena Hassan
teena.hassan@h-brs.de

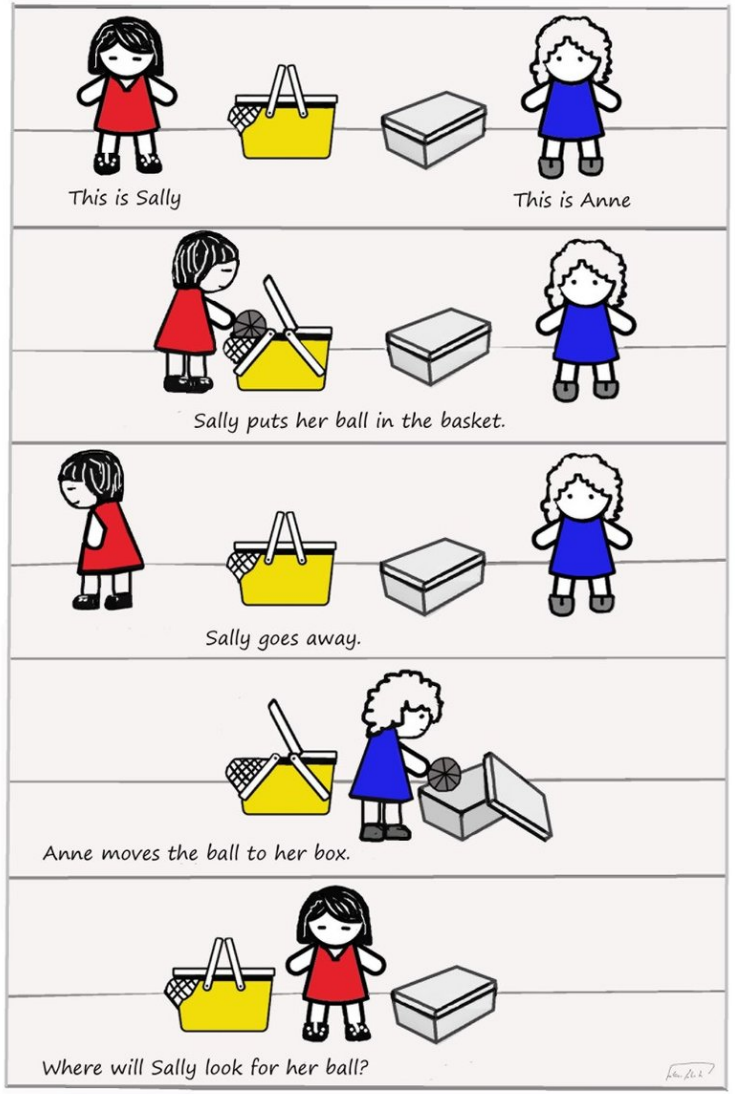
[Department of Computer Science]
Hochschule Bonn-Rhein-Sieg
Sankt Augustin

16th May 2024

Learning Goals

- At the end of today's lecture, you will be able to:
  1. Explain four key psychological processes associated with social intelligence.
  2. Categorize robots based on their social capabilities.
  3. Depict and explain a conceptually complete architecture for multimodal, social human-robot interaction.
  4. Describe the steps involved in designing a multimodal interaction architecture.
  5. Give examples for the integration of behavior explanation generation and metacommunication features in social robots.

Theory of Mind



- Children from the age of 5 can explain the causal reasons for false beliefs of others and understand how these will influence their actions.
  - That is, they "have a Theory of Mind (ToM)".
  - False beliefs --> Beliefs that differ from reality
  - Requires the ability to read other peoples' minds (beliefs and desires).

- Short summary: https://www.cell.com/current-biology/pdf/S0960-9822(05)00960-7.pdf

- More examples: https://www.theoryofmindinventory.com/task-battery/

Image source: Fig. 3 in (Felisberti et al., 2017).
Original implementation of Sally-Anne test: (Baron-Cohen et al., 1985).

A²S

Mind Reading or Mentalizing

- Mentalizing:
  - "The capacity to reflect on and interpret one's own behavior and that of others based on intentional internal mental states, such as beliefs, thoughts, and emotions (Fonagy et al., 1998, 2002)." [Rothschild-Yakar et al. 2019]

- Theory of Mind (ToM):
  - "Ability to form representations about others' intentional internal mental states such as thoughts, feelings, and beliefs (Heavey et al., 2000)." [Rothschild-Yakar et al. 2019]

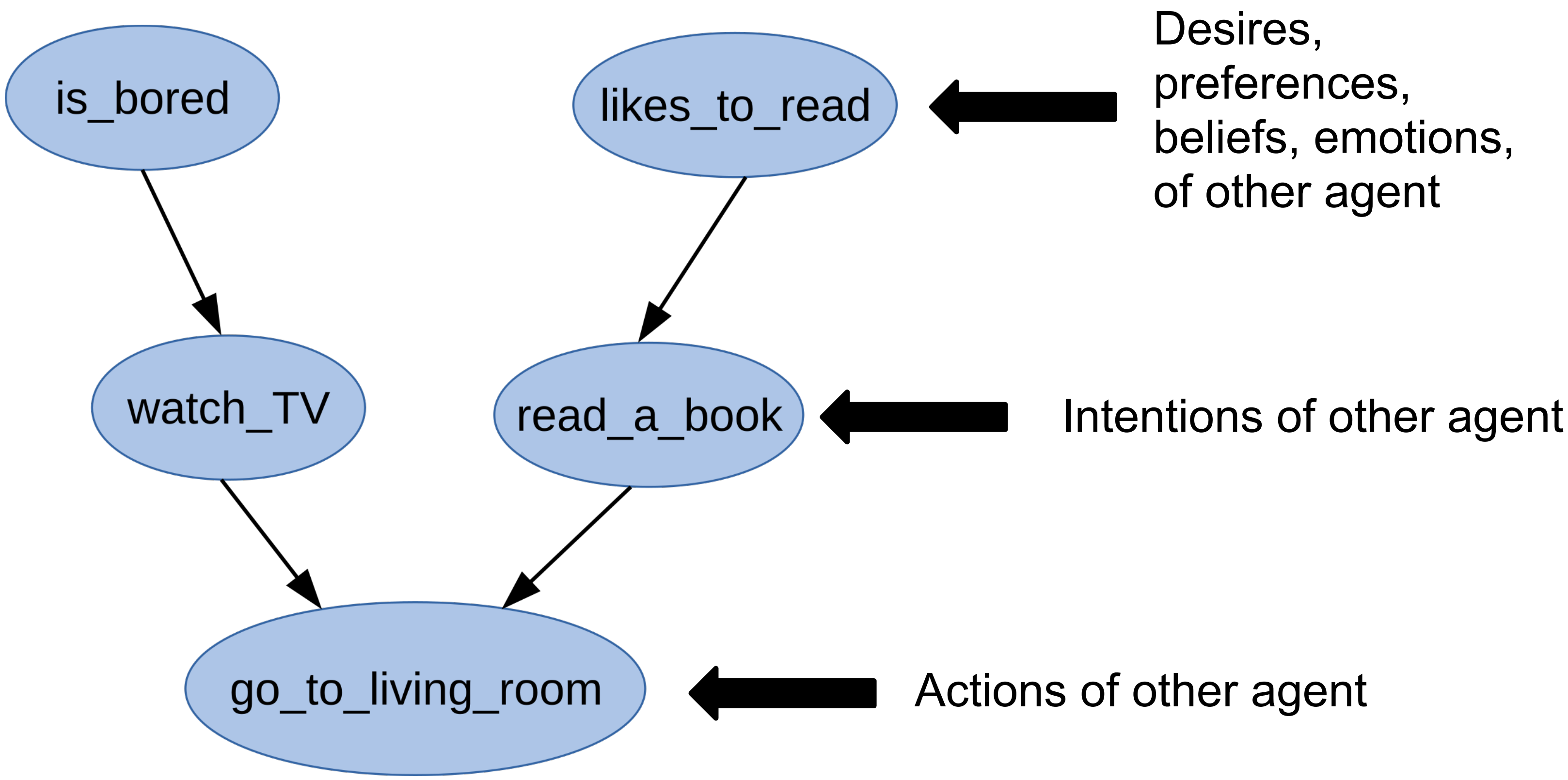
Why Should a Robot have ToM?

- To predict the intentions of humans or other robots.
- To predict the potential next actions of other agents.
- To plan own actions based on these predictions.

- Essential for safe and intuitive collaboration between humans and robots.

Bayesian Theory of Mind Framework

- Bayesian Theory of Mind framework (Pöppel & Kopp, 2018, Han & Pereira, 2013):
  - Uses Bayesian Network based approaches to model and reason about the minds of others based on the actions they perform.
- Models are user-specific and should be updated based on experience.



Joint Attention

https://youtu.be/1Ab4vLMMAbY

https://youtu.be/NTVtYMYwrGE

A²S

Joint Attention [Based on (Perez-Osorio, 2021)]

- ... "*takes place when two individuals coordinate their attentional processes to conjointly attend to the same object or situation in the environment.*" (Page 320, Sec. 9.2.2., Perez-Osorio, 2021)

- Helps to:
  - Identify the attentional focus of the interaction partner
  - Infer the intentions, mental state of the interaction partner
  - Predict the next action of the interaction partner

- Essential for perspective-taking, theory of mind, joint action, language acquisition (symbol emergence).

A²S

Important Definitions related to Affect & Emotions

- **Affect:** "*an umbrella term that refers to anything related to emotion, emotion processing, and emotion in social interaction.*"
  - Affective science: "*deals with the study of emotion in the broadest sense.*"

(Page 350, Section 10.1.1.1, Broekens, 2021)

- **Emotion:** "*an event-related affective reaction (it is about something) typically of short duration and relatively intense (one feels the emotion and is conscious of it).*" (Page 350, Section 10.1.1.1, Broekens, 2021)
  - The same event can elicit different emotions in different individuals.
  - **Appraisal**: "*The assessment of the personal relevance of a situation.*" (Page 349, Section 10.1.1, Broekens, 2021)

- **Mood:** "*the longer-term affective state an individual is in, is usually less intense, unrelated to a specific event, and less differentiated.*" (Page 351, Section 10.1.1, Broekens, 2021)

A²S

Emotions – Three Perspectives

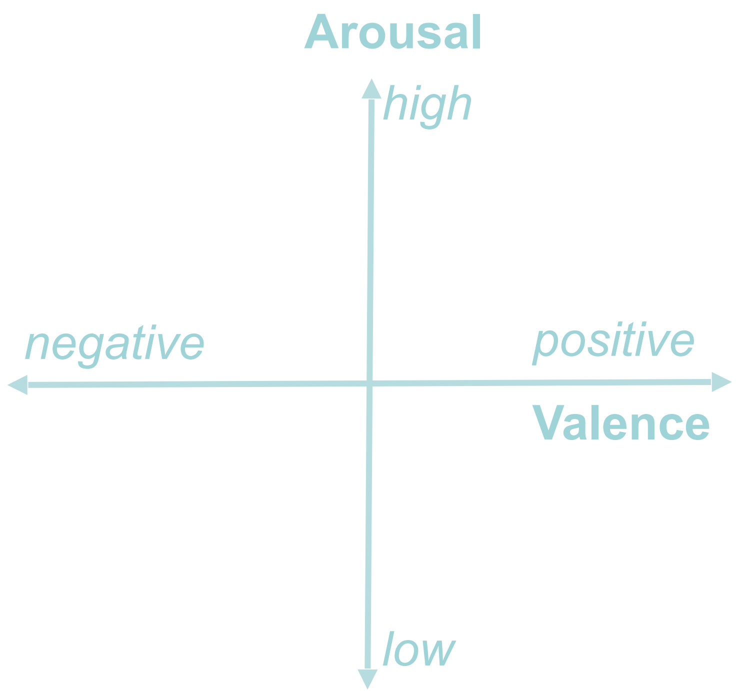
- Three psychological perspectives for studying and modeling emotions (Broekens, 2021):
  - Categorical view
  - Dimensional-constructionist view
  - Cognitive appraisal view

- Categorical view:
  - Emotions as "*specific multimodal responses*" to events or situations.
  - E.g. Ekman's six basic emotions (fear, disgust, anger, surprise, joy, sadness) (Ekman and Friesen, 1971)
  - In artificial agents: Useful to express, label or communicate emotions.

A²S

Dimensional-Constructionist View

- Emotions as our individual, subjective interpretation of affect, not necessarily connected to expressions or action tendencies.
  - E.g. Valence-arousal-dominance model
    - Valence: Degree of pleasantness
    - Arousal: Intensity of physiological responses
    - Dominance: Degree of control over the situation
- In artificial agents: Useful to
  - Express mood, emotions, etc. within a common framework.
  - Model continuous and dynamic changes in emotions.

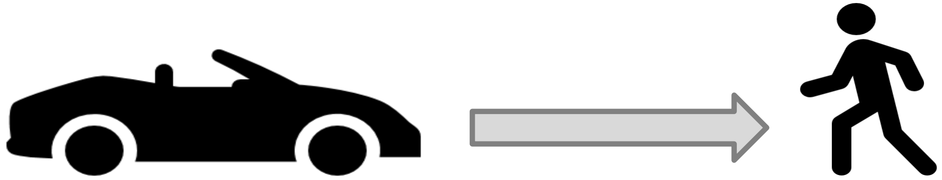


Reference: (Broekens, 2021)

A²S

Cognitive Appraisal View (Broekens, 2021)

- Emotion is the outcome of appraisal (assessment of the situation) at cognitive level.
  - Involves **"concern-based reasoning"** to evaluate to what extend the situation is of relevance to the individual.
  - **Motivate** the individual to perform appropriate actions in response to the situation.
- Simple versus complex appraisal
  - Simple: Based on properties of the stimulus
  - Complex: Based on causes and implications of the stimulus
- Links cognitive processing with emotion elicitation, but not always deliberative.
- Does not label the emotional response or resulting behavior.
- Useful to model the **emotion elicitation process**.



Fear
- Sudden
- High personal relevance
- Low goal congruence (e.g.goal to survive)
- Low control

A²S

Empathy (Paiva et al., 2021)

- **Empathy:** a psychological process leading a person to have "*feelings that are more congruent with another's situation than with their own situation.*" (Hoffman 2001)

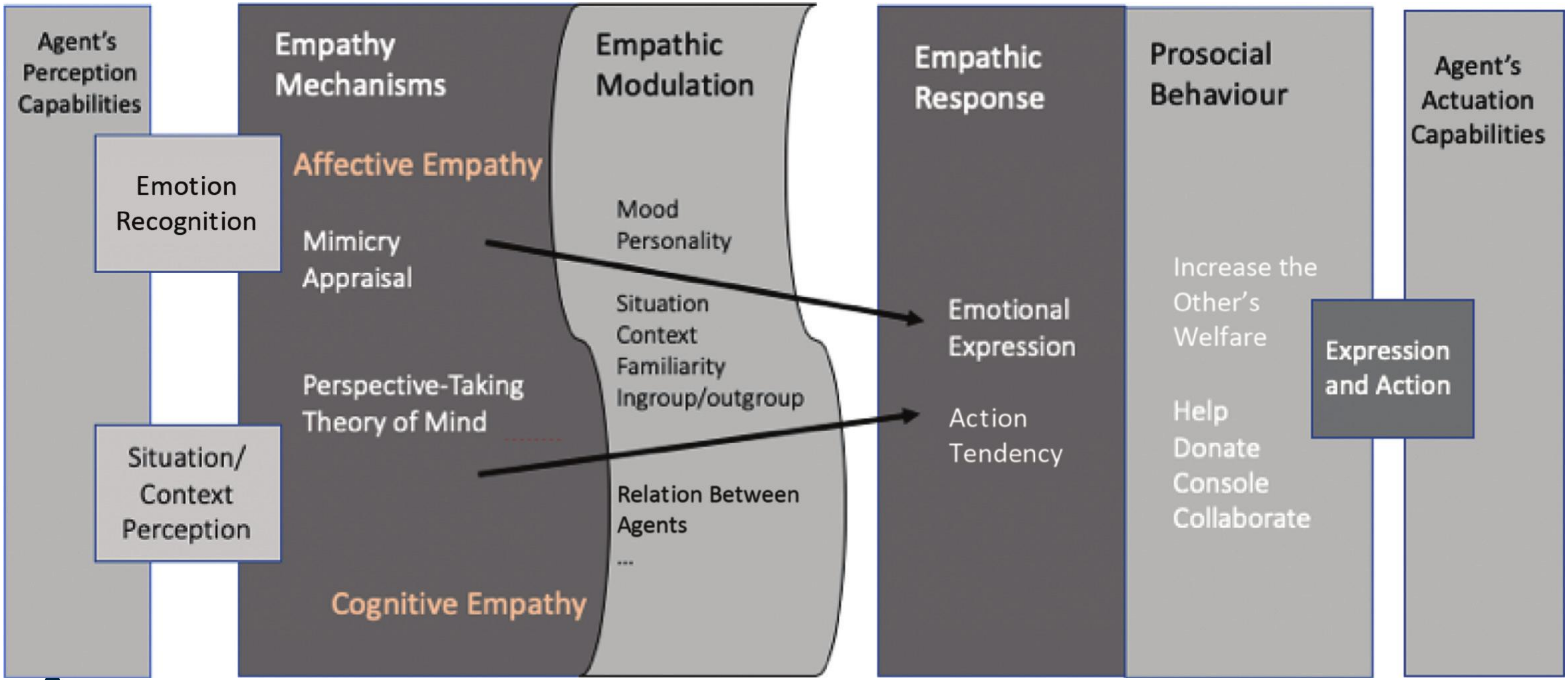
- **Cognitive empathy:** "*the capacity to put oneself in the other's position*" (perspective taking, a theory of mind of the other agent) (Page 390, Paiva et al., 2021)

- **Affective empathy:** the affective (emotion, mood) response to another's plight.

https://youtu.be/1Evwgu369Jw

A²S

Mechanisms for Empathy



© 2021 Ana Paiva et al.
Source: Fig. 11.3 in (Paiva et al,., 2021)

Group Activity — 20 Mins + 10 Mins

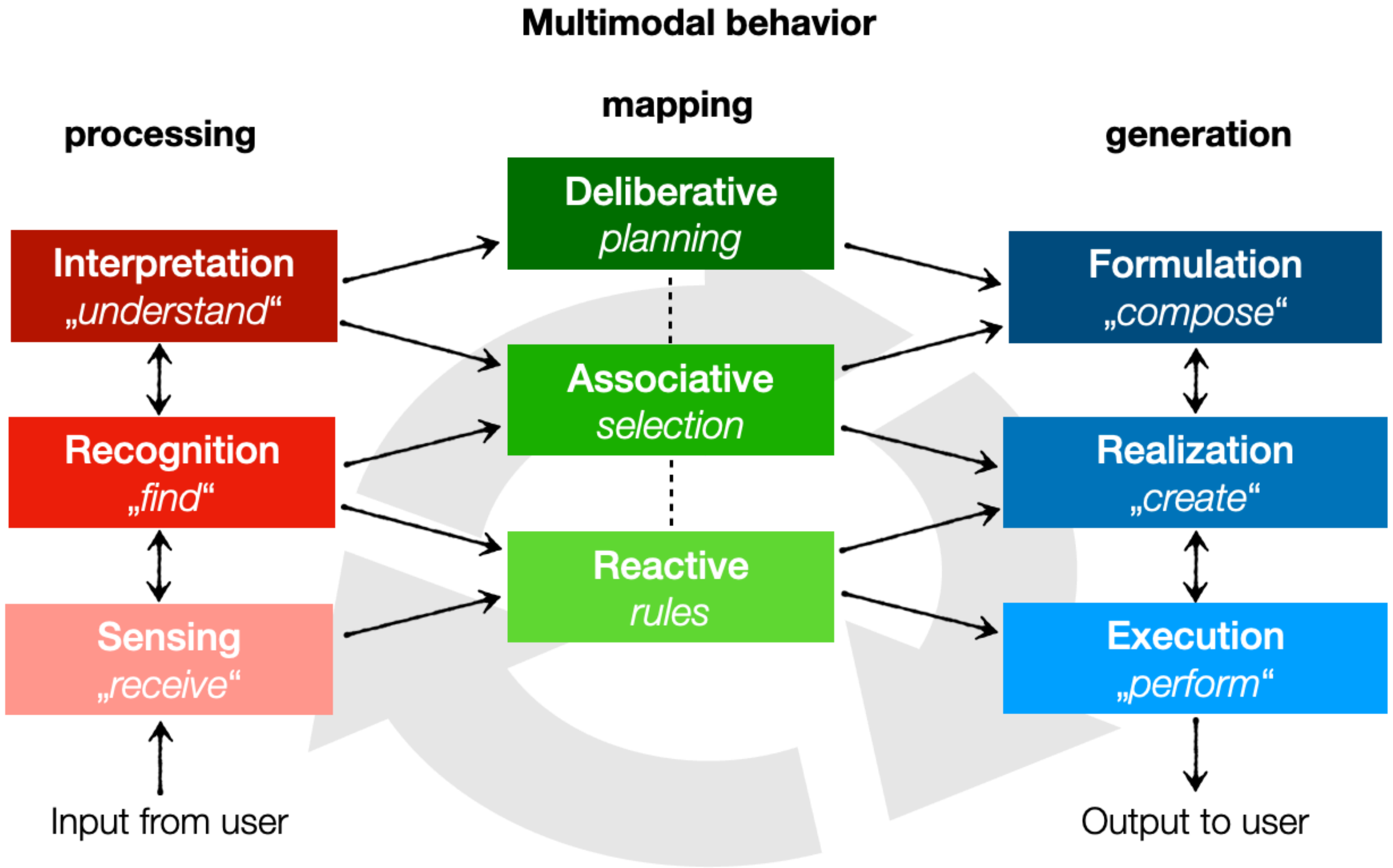
- You are the designer of a social robot which is a companion for older adults living alone. Imagine that the user is mourning the passing away of a close friend. The robot has to comfort the user.
  - Identify mechanisms that the robot would need to show empathy.
  - Illustrate how these mechanisms would enable empathetic responses from the robot during a conversation with the user.

Social Capabilities of a Robot

- Verbal and nonverbal communication
- Recognition and expression of emotion
- Exhibiting personality and traits
- Modeling and recognizing social aspects of humans (e.g. theory of mind)
- Learning and developing new social skills and competencies
- Establishing and maintaining social relationships

(Fong et al., 2002, Baraka et al., 2020)

A "Conceptually Complete" Architecture



Reference: (Kopp S & Hassan T., 2022.)

- Three columns
  - Processing multimodal input
  - Mapping responses
  - Generating multimodal output
- Multiple levels
  - Lower level: Sensing & motor behavior
  - Higher level: Conversional and socio-relational functions

A "Conceptually Complete" Architecture

(Kopp & Hassan, 2022)

- "receive" sensor data --> "recognize" features or patterns --> "interpret" semantics, interactional functions
- Determine the "composition" of behavior --> "create" the synchronized behaviors --> "act" to express the behavior
- Map input to output through different decision-making levels
  - Reactive level: Use "hard-wired rules"
  - Associative level: Select from a set of alternatives
  - Deliberative level: Plan new responses

A²S

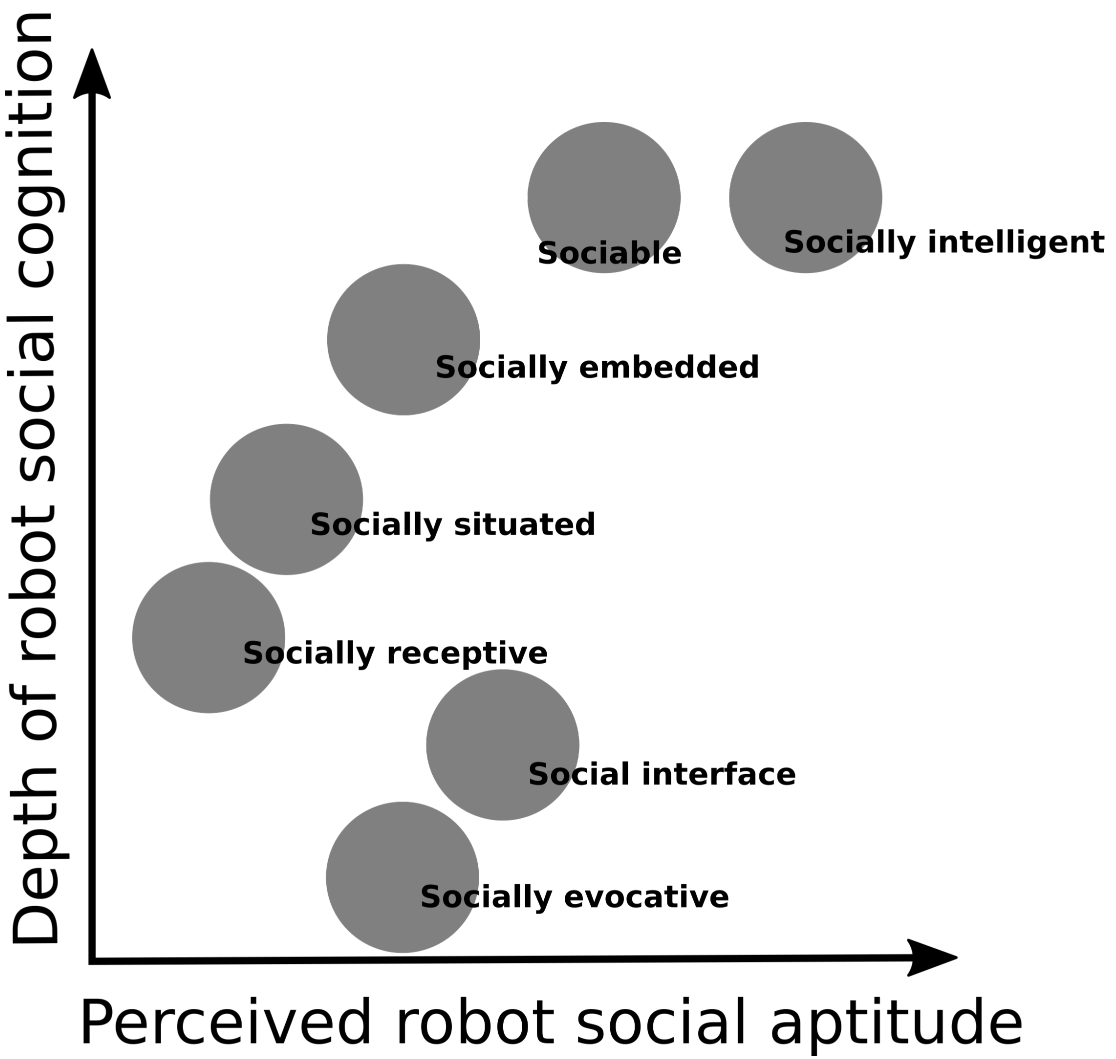
Different Processing Pathways (Kopp & Hassan, 2022)

- Multiple processing routes or pathways possible within interaction architectures
  - Single-route architectures
    - Sequential processing of information through one decision-making route.
    - E.g. for showing certain socially appropriate behaviors such as engagement, joint attention, etc.
  - Dual-route architectures
    - Involves two types of decision-making routes.
    - More complex interaction such as face-to-face conversations.
    - Behavior realizer should arbitrate the behavior generated through different routes.
  - Multidirectional, incremental architectures
    - Information flows in different directions (left-right, right-left, top-down, bottom-up)
      - To bias sensory processing, disambiguate sensor data interpretation, to adapt models, to dynamically update behavior plans.
    - Incremental information processing
    - Essential for fluent, naturalistic, full-fledged social interaction with humans.

A²S

Categories based on Social Capabilities of a Robot

- Two dimensions
  - Robot's actual social cognition capabilities
  - Human's perception of robot's social capabilities
- Shallow: Socially evocative, social interface, socially receptive
- Moderate: Socially situated, socially embedded
- Deep: Sociable, socially intelligent
- For definitions: Section 2.2 of (Baraka et al., 2020)



Recreation of Fig. 3 in (Baraka et al., 2020)

Shallow Social Capabilities

- Socially evocative (Breazeal, 2003)
  - Robot does not give social responses but evokes socioemotional responses in users through anthropomorphism.
- Social interface (Breazeal, 2003)
  - External communication interface (perception, action) is social (verbal, nonverbal), but does not process or model social behaviour deeply.
- Socially receptive (Breazeal, 2003)
  - Can process social signals (e.g. to learn from humans), but do not generate any social responses.

A²S

Moderate Social Capabilities

- Socially situated (Fong et al., 2002)
  - Exist in a social environment
  - Distinguishes social agents from other objects
  - Perceives and reacts to social agents
  - E.g. delivery robots at a hospital; asking humans to move...

- Socially embedded (Fong et al., 2002)
  - Socially situated and interacts with social agents in the environment
  - Is aware of complex social interaction structures like turn taking

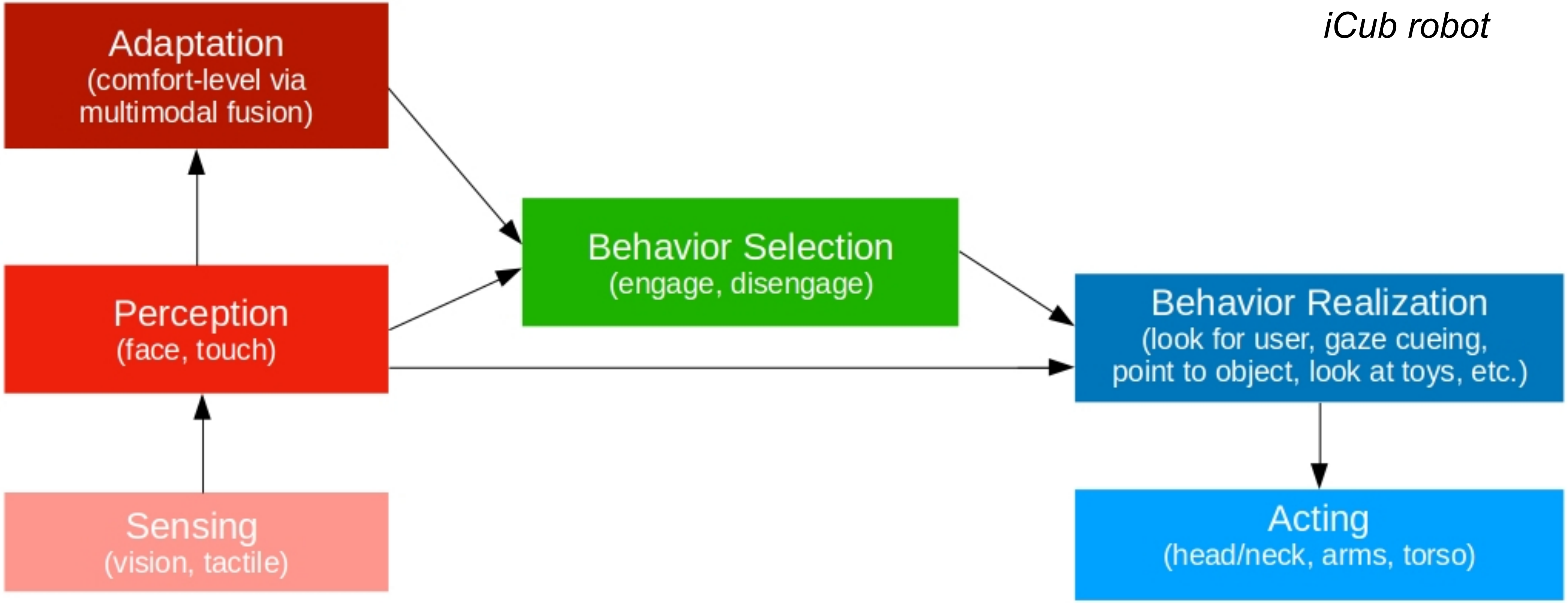
Deep Social Capabilities

- Sociable (Breazeal, 2003)
  - Has internal drives, emotions, needs, goals, etc.
  - Can self-initiate interaction with humans to satisfy own goals.
  - Deep social cognition including theory of mind.

- Socially intelligent (Fong et al., 2002)
  - Social intelligence similar to humans
  - Deep social cognition and competence (e.g. learning from and adapting to humans, establishing social relationships).

Single Route Architecture – Example
(Kopp & Hassan, 2022)

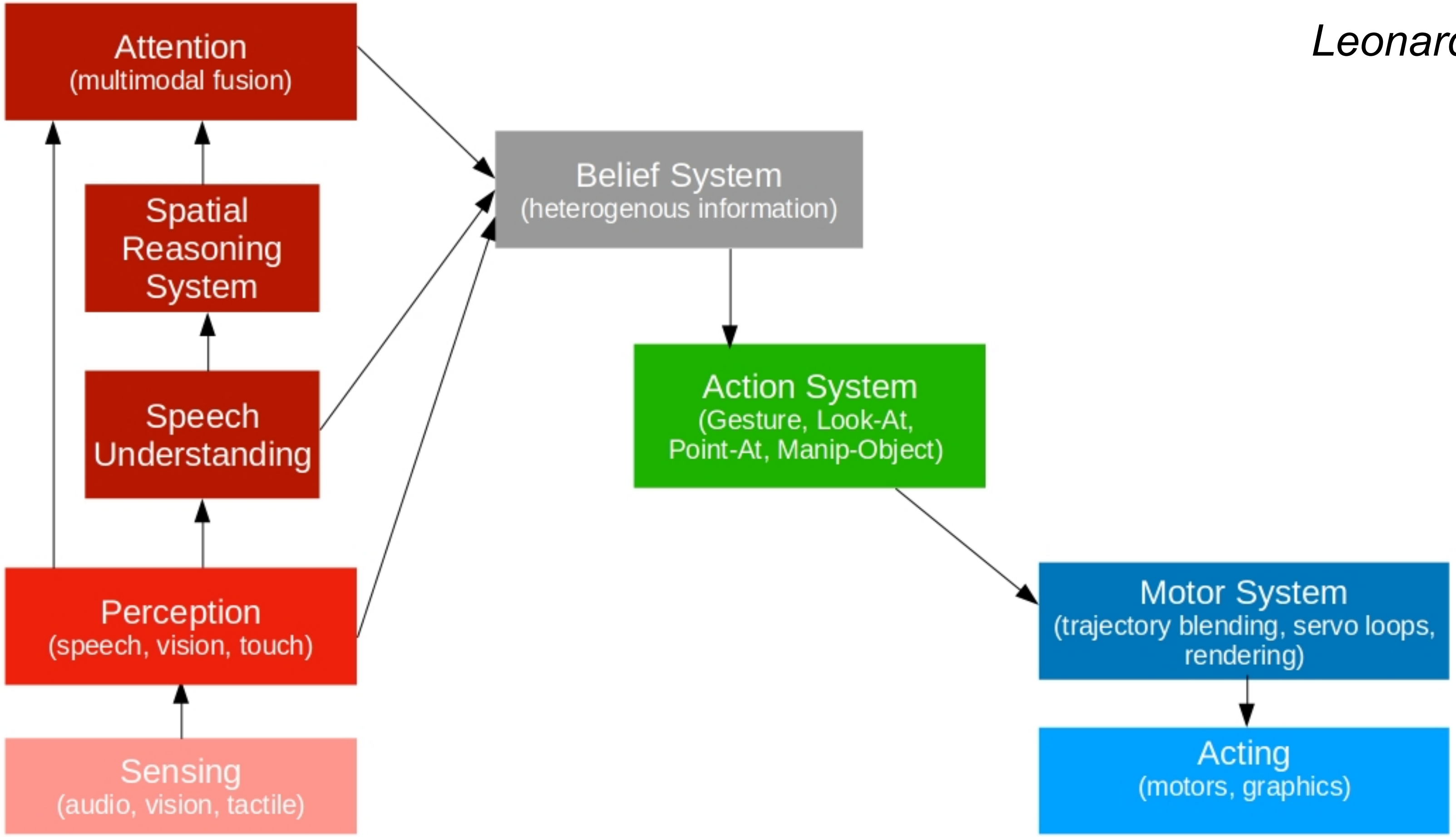
iCub robot



Based on (Tanevska et al., 2019)

Single Route Architecture – Example

(Kopp & Hassan, 2022)

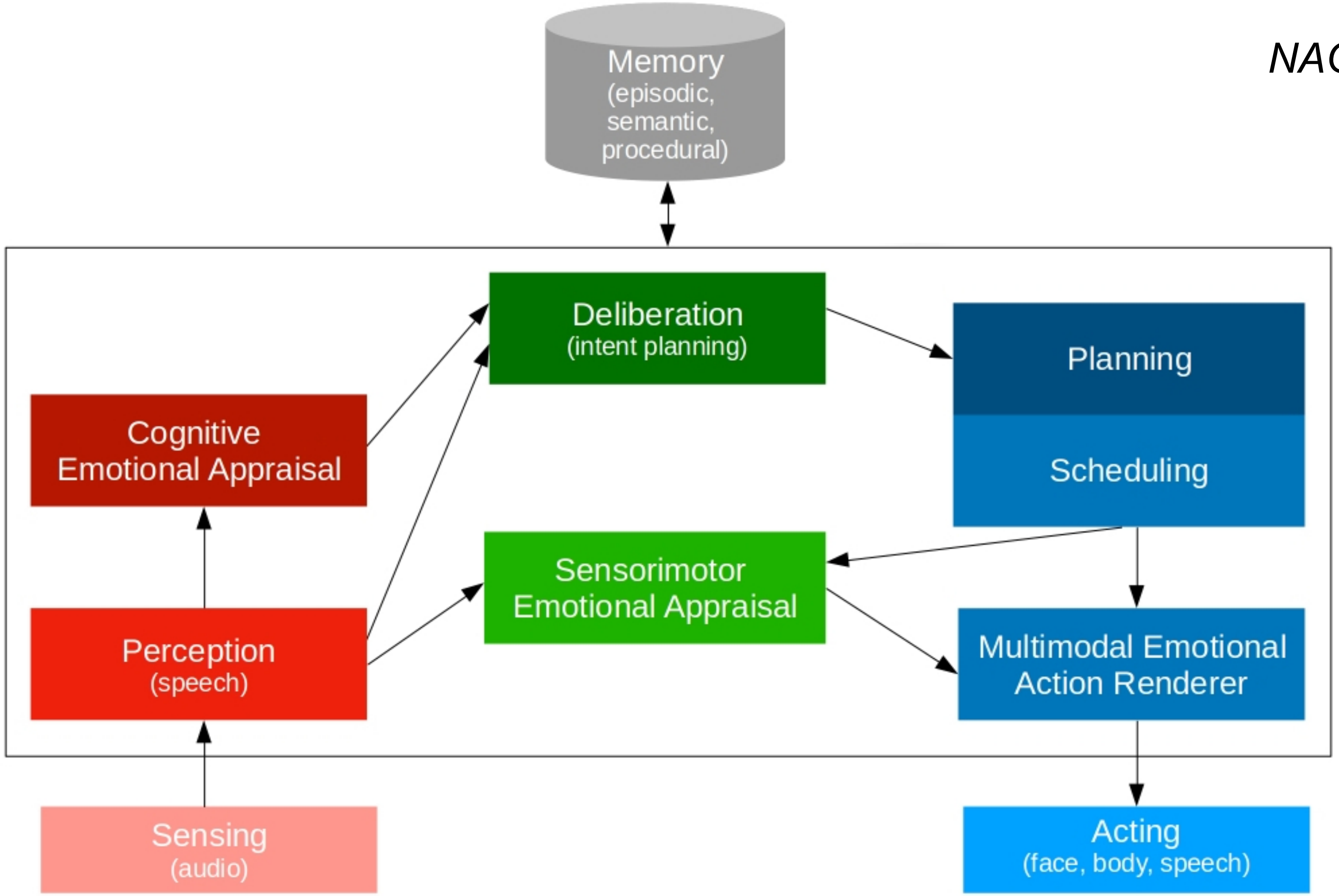


Leonardo robot

Based on (Breazeal et al., 2004)

Dual Route Architecture – Example (Kopp & Hassan, 2022)

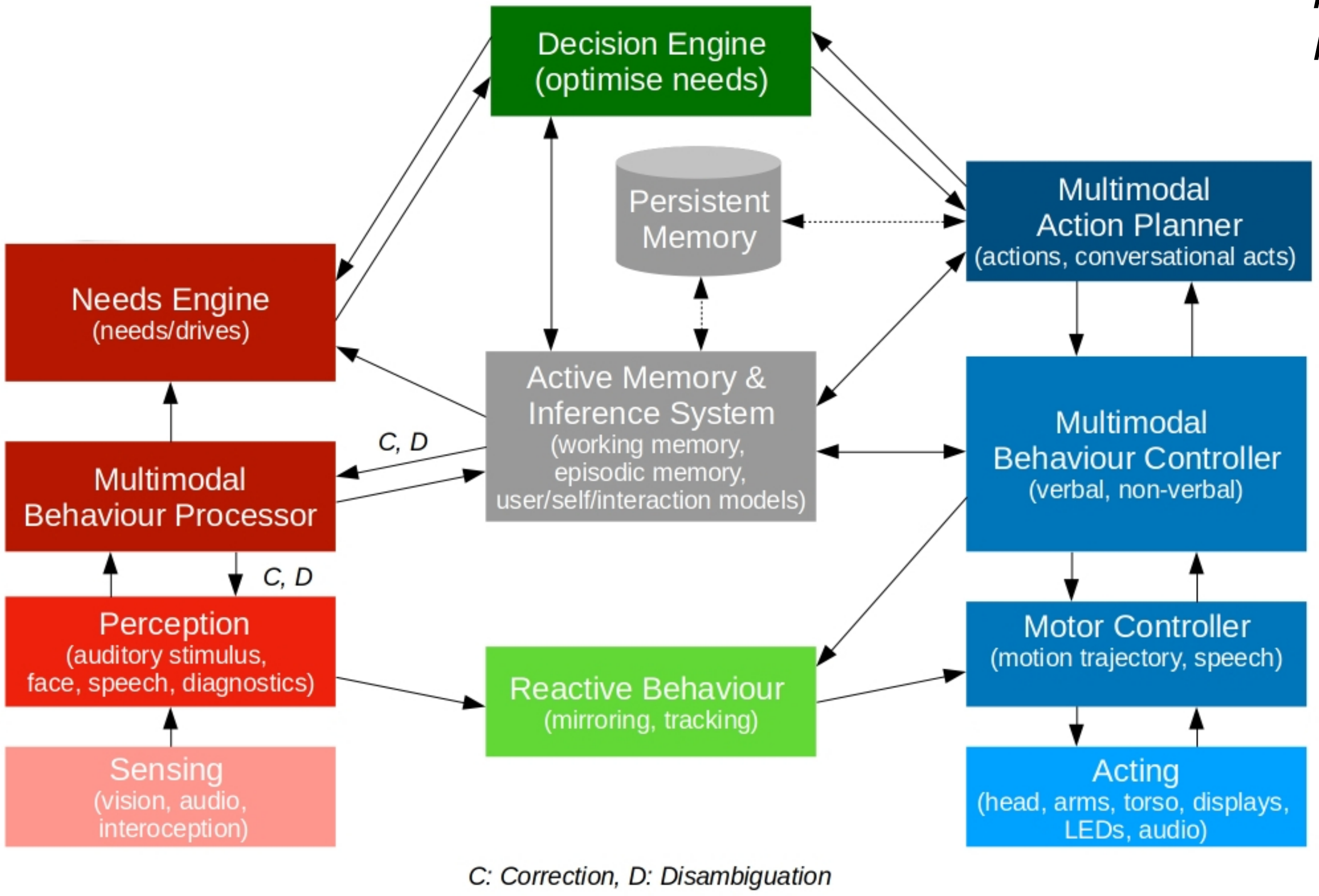
NAO robot



CAIO Architecture (based on [Adam et al., 2016])

Multidirectional, Incremental Architecture

Pepper robot
Navel robot



(Kopp & Hassan, 2022)

Designing an Interaction Architecture

- Given the requirements and user stories, we design the hardware-software architecture.
- Designing the conceptual framework:
  - Identify required components.
    - Separate embodiment-specific and embodiment-independent aspects.
  - Identify required communication channels between components.
  - Define the components and communication interfaces.
    - Names of components and interfaces
    - Direction of communication
    - Formats of message payloads
  - Choose the messaging protocol(s) to be used for inter-component communication.
    - ROS, MQTT, IPAACA, etc.
- Iteratively implement, evaluate and optimize the architecture.

Making Robots Understandable to Humans

- Due to the complexity of social interaction and the fact that users often do not have an accurate mental model of robots' internal workings, the robot's behavior might not be intuitive or understandable to the human interaction partner.

- Therefore, robots should be capable of explaining their behavior.
  - *"**Explainability** of embodied social agents is their ability to provide information about their inner workings using social cues such that an observer (user) can infer how/why the embodied agent behaves the way it does."* (Stange et al., 2022)

- The explanations should be (Stange et al., 2022):
  - Generated using empirically validated models.
  - Delivered in a user-centered way.

A²S

Behaviour Explanation Dialog Framework

- To autonomously construct and deliver behaviour explanations on-demand, the interaction architecture should:
  - Extract and consolidate information relevant for explanations simultaneous to behavior generation.
  - Save and retrieve explanation-relevant information to/from long-term memory (for past behavior).

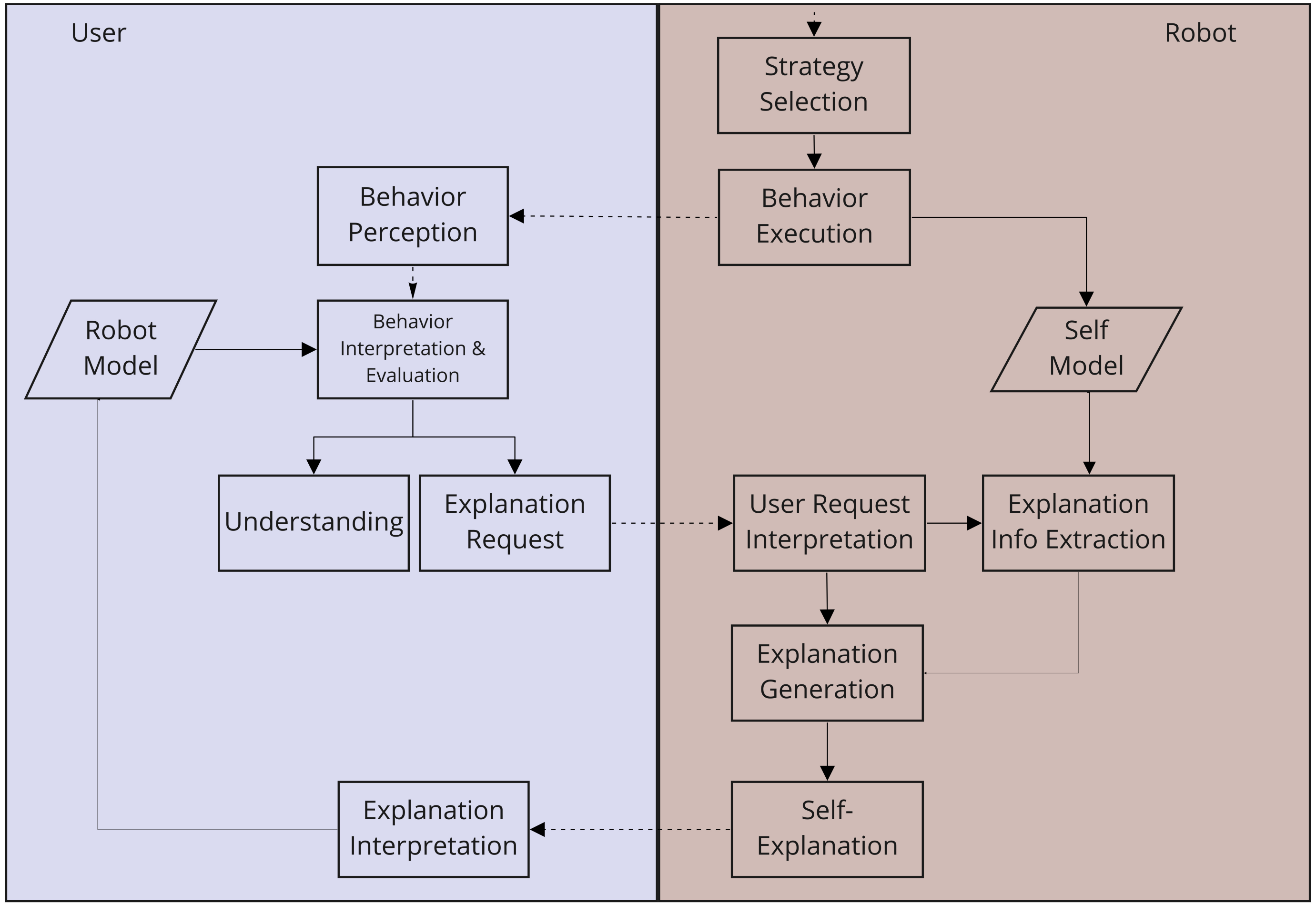


Image source: Fig. 2 in (Stange et al., 2022)

A²S

Verbal Explanations of Behaviour – An Example

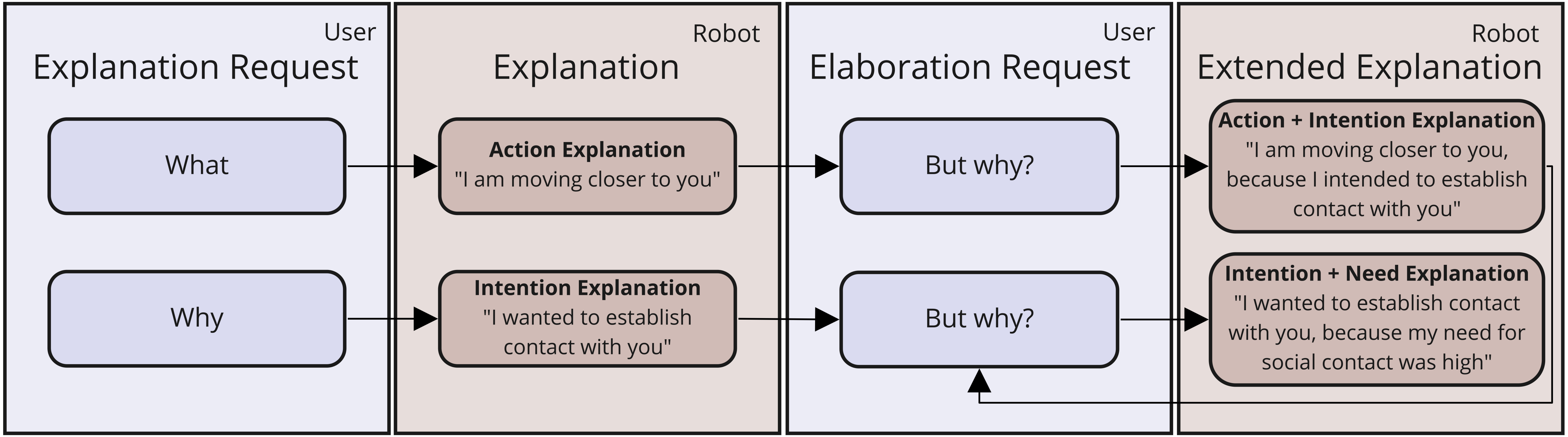
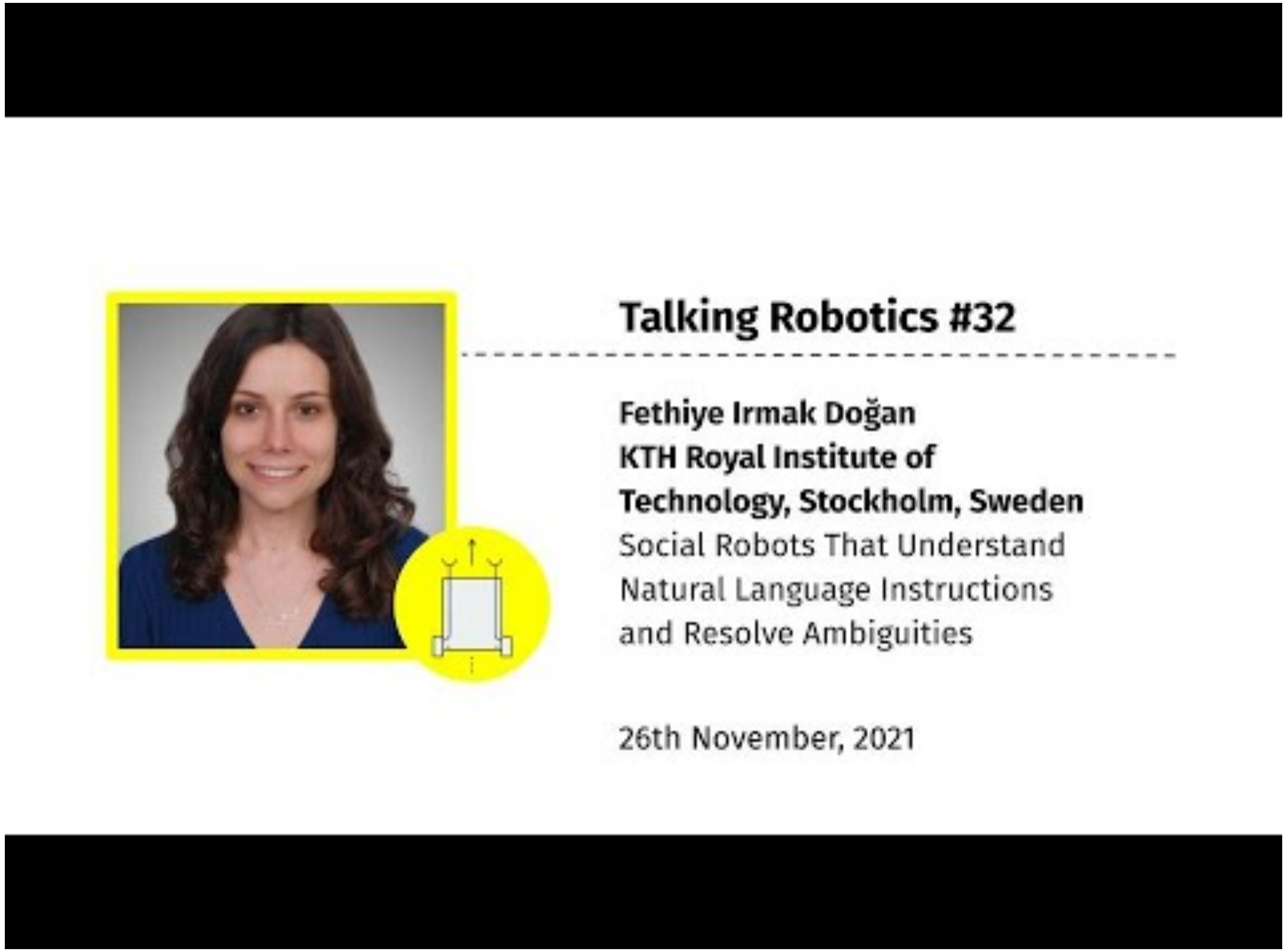


Image source: Fig. 3 in (Stange et al., 2022)

Helping Robots Understand Humans

- Metacommunication
  - Clarification requests

- E.g. Clarify ambiguities arising from user's referring expressions
  - Integrates:
    - Verbal communication
    - Visual-spatial perception
      - Scene understanding
    - Perspective taking
    - Symbol grounding



https://youtu.be/I21HicOzdI8

Conclusion

- In today's lecture, you learnt to:
  1. Explain four key psychological processes associated with social intelligence.
     - Theory of mind, joint attention, emotions, empathy
  2. Categorise robots based on their social capabilities.
     - Two dimensional framework: Depth of social cognition versus perceived social aptitude
  3. Depict and explain a conceptually complete architecture for multimodal, social human-robot interaction.
  4. Describe the steps involved in designing a multimodal interaction architecture.
  5. Give examples for the integration of behaviour explanation generation and meta-communication features in social robots.

References and Recommended Reading

[1] Theory of Mind, Joint Attention, Emotions, Empathy:

- Chapters 9, 10, 11 in Birgit Lugrin, Catherine Pelachaud, and David Traum (Eds.). 2021. ***The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition* (1st. ed.).** *ACM Books*, Vol. 37. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3477322 (https://sociallyinteractiveagents.org/HOME/)

[2] Multimodal interaction architectures:

- Kopp S, Hassan T. The Fabric of Socially Interactive Agents: Multimodal Interaction Architectures. In: Lugrin B, Pelachaud C, Traum D, eds. The Handbook on Socially Interactive Agents. ACM Books. Association for Computing Machinery; 2022. (https://sociallyinteractiveagents.org/HOME/)
- Stange S, Hassan T, Schröder F, Konkol J, Kopp S. Self-Explaining Social Robots: An Explainable Behavior Generation Architecture for Human-Robot Interaction . Frontiers in Artificial Intelligence. 2022;5: 866920.

[3] Social Robots

- Baraka, K., Alves-Oliveira, P., Ribeiro, T. (2020). An Extended Framework for Characterizing Social Robots. In: Jost, et al. (Eds.) Human-Robot Interaction. Springer Series on Bio- and Neurosystems, vol 12. Springer, Cham.

[The first reference in [2] would be important for the exam.]

References (1/2)

- C. Adam, W. Johal, D. Pellier, H. Fiorino, and S. Pesty. 2016. Social human-robot interaction: A new cognitive and affective interaction-oriented architecture. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He, eds., Social Robotics, pp. 253–263. Springer International Publishing, Cham. ISBN 978-3-319-47437-3.
- Baraka, K., Alves-Oliveira, P., Ribeiro, T. (2020). An Extended Framework for Characterizing Social Robots. In: Jost, et al. (Eds.) Human-Robot Interaction. Springer Series on Bio- and Neurosystems, vol 12. Springer, Cham.
- Simon Baron-Cohen, Alan M. Leslie, Uta Frith, Does the autistic child have a “theory of mind” ?, Cognition, Volume 21, Issue 1, 1985.
- Bratman M. Intention, Plans, and Practical Reason , Harvard University Press, Cambridge (1987).
- Breazeal, C.: Toward sociable robots. Robot. Auton. Syst. 42(3–4), 167–175 (2003).
- C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Chilongo. 2004. Tutelage and collaboration for humanoid robots. International Journal of Humanoid Robotics, 1(2). DOI: 10.1142/S0219843604000150.
- Joost Broekens. 2021. Emotion. The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods,
- Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System (FACS).*
- P. Ekman and W. V. Friesen. “Constants across cultures in the face and emotion.” In: Journal of Personality and Social Psychology 17.2 (1971).
- Felisberti, Fatima & King, Robert. (2017). Mind-Reading in Altruists and Psychopaths. 10.1007/978-3-319-68421-5_6.
- Fonagy, P., Gergely, G., Jurist, E., and Target, M. (2002). Affect Regulation, Mentalization, and the Development of the Self. New York, N.Y: Other Press.
- Fonagy, P., Target, M., Steele, H., and Steele, M. (1998). Reflective-Functioning Manual, Version 5.0, for Application to ADULT Attachment Interviews. London: University College London, 161–162.
- Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots: concepts, design and applications. Technical Report CMU-RI-TR-02-29, Robotics Institute, Carnegie Mellon University (2002).
- Han, The Anh and Pereira, Luís Moniz. ‘State-of-the-art of Intention Recognition and Its Use in Decision Making’. 1 Jan. 2013 : 237 – 246.

A²S

References (2/2)

- Heavey, L., Phillips, W., Baron-Cohen, S., and Rutter, M. (2000). The awkward moments test: a naturalistic measure of social understanding in autism. J. Autism Dev. Dis. 30, 225–236.
- M. L. Hoffman. 2001. Empathy and Moral Development: Implications for Caring and Justice. Cambridge University Press.
- Kopp S, Hassan T. The Fabric of Socially Interactive Agents: Multimodal Interaction Architectures. In: Lugrin B, Pelachaud C, Traum D, eds. The Handbook on Socially Interactive Agents. ACM Books. Association for Computing Machinery; In Press.
- Negre et al., "7 - Toward a Modeling of Population Behaviors in Crisis Situations," Editor(s): Florence Sèdes, How Information Systems Can Help in Alarm/Alert Detection, Elsevier, 2018.
- Ana Paiva, Filipa Correia, Raquel Oliveira, Fernando Santos, and Patrícia Arriaga. 2021. Empathy and Prosociality in Social Agents. The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition (1st ed.). Association for Computing Machinery, New York, NY, USA, 385–432.
- Jairo Perez-Osorio, Eva Wiese, and Agnieszka Wykowska. 2021. Theory of Mind and Joint Attention. The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition (1st ed.). Association for Computing Machinery, New York, NY, USA, 311–348.
- Jan Pöppel and Stefan Kopp. 2018. Satisficing Models of Bayesian Theory of Mind for Explaining Behavior of Differently Uncertain Agents: Socially Interactive Agents Track. In AAMAS '18.
- Behavior, Cognition (1st ed.). Association for Computing Machinery, New York, NY, USA, 349–384.
- Rothschild-Yakar L, Stein D, Goshen D, Shoval G, Yacobi A, Eger G, Kartin B and Gur E (2019) Mentalizing Self and Other and Affect Regulation Patterns in Anorexia and Depression. Front. Psychol. 10:2223. doi: 10.3389/fpsyg.2019.02223
- Stange S, Hassan T, Schröder F, Konkol J, Kopp S. Self-Explaining Social Robots: An Explainable Behavior Generation Architecture for Human-Robot Interaction . Frontiers in Artificial Intelligence. 2022;5: 866920.
- A. Tanevska, F. Rea, G. Sandini, L. Cañamero, and A. Sciutti. 2019. Eager to learn vs. quick to complain? how a socially adaptive robot architecture performs with different robot personalities. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 365–371. DOI: 10.1109/SMC.2019.8913903.

Next Session

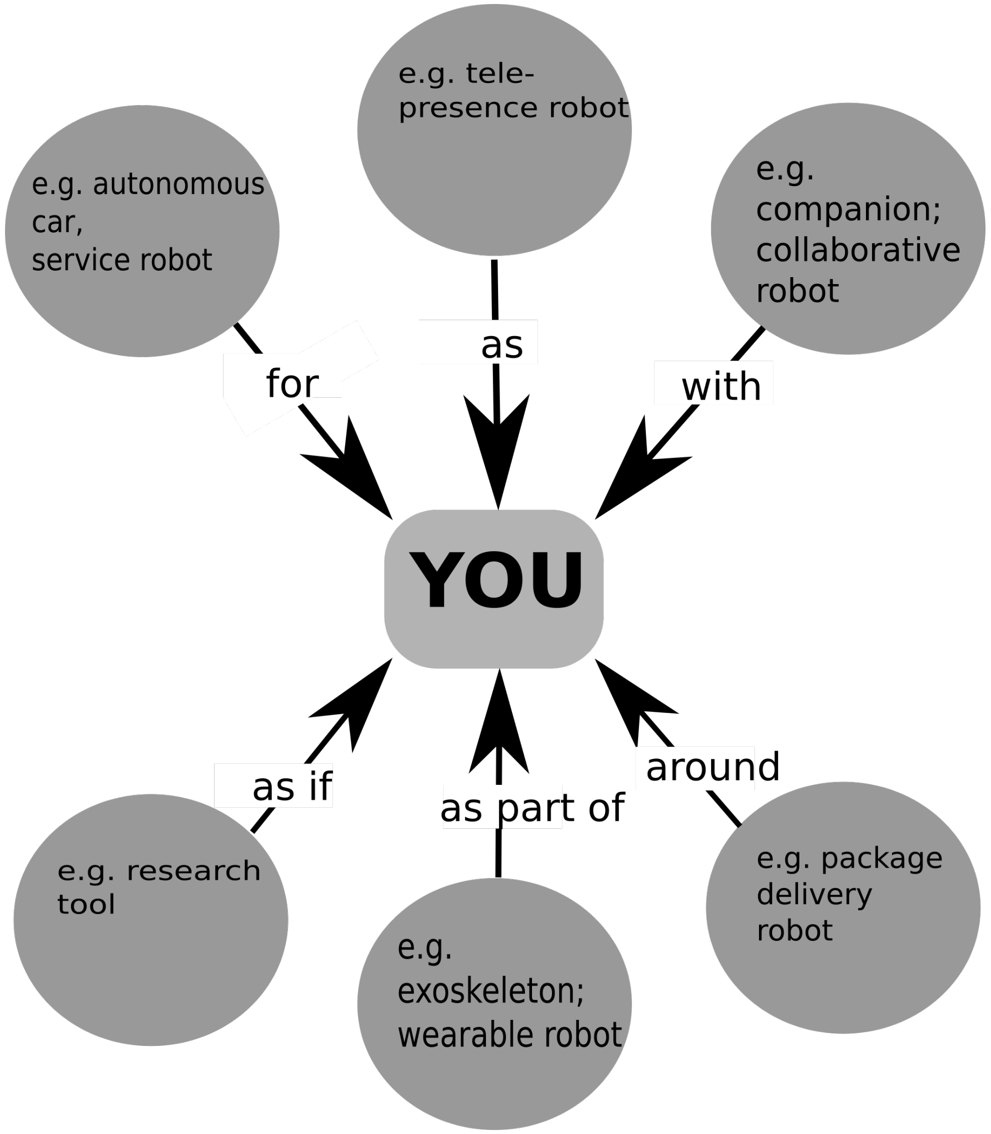
- Tutorial on "Bayesian networks and pyAgrum package"
  - At 10:45 today

- Lecture on “Physiological Computing" by Jordan Schneider.
  - Thursday, 23.05.2024 at 9 am
  - In: B060, Grantham Allee 20, 53757 Sankt Augustin

A²S

Relational Role of Robot in Interaction (Baraka et al., 2020)

- The role of the robot in relation to the human within an interaction.
- Independent of application domain or activities performed by the robot.
  1. "for-you": as a tool or servant
  2. "as-you": as a proxy for the human
  3. "with-you": as part of a team (collaboration)
  4. "as-if-you": as a research tool in social scientific research
  5. "around-you": co-existence, co-presence, bystander
  6. "as-part-of-you": augmenting human body and its physical capabilities



Re-representation of Fig. 5 in (Baraka et al., 2020)