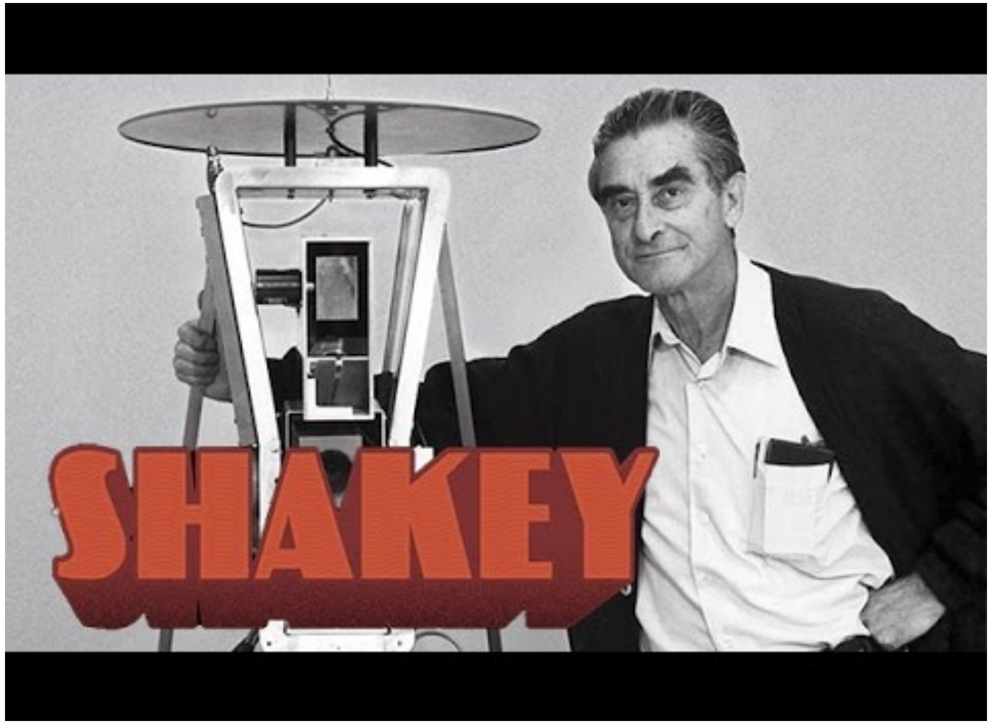
The Beginning of Modern-Day Robotics...

https://youtu.be/xyj6N-i6asQ



https://youtu.be/7bsEN8mwUB8
https://youtu.be/GmU7SimFkpU

Unimate:

- Created by George C. Devol in 1954 and deployed at the General Motors in 1961.
- First programmable manipulator for the manufacturing industry for specialized tasks.

Shakey:

- Created by Charles Rosen and team at Stanford Research Institute during 1966-1972.
- First AI-enabled mobile intelligent robot.
  - Sense, reason, act
- IEEE Milestone Achievement Award!

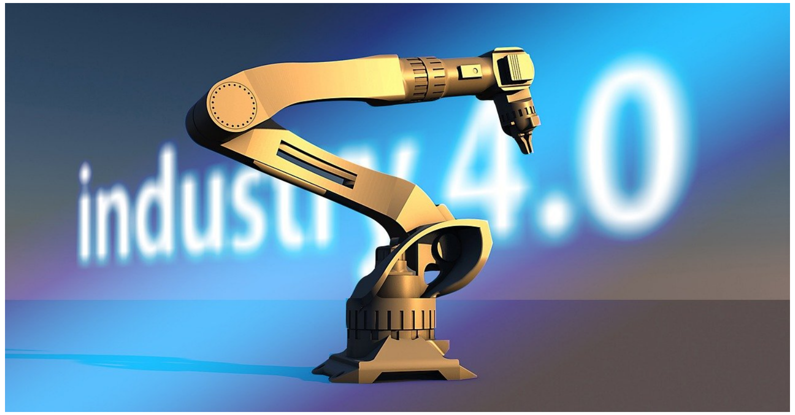
A²S

What is a Robot?

- A robot is ..
  - ... a machine able to extract information from its environment and use knowledge about its world to act safely in a meaningful and purposeful manner (Ron Arkin, Behavior-based Robotics, 1998).
  - … *an atuonomous system which* ***exists*** *in the* ***physical world****, can* ***sense*** *its* ***environment****, and can* ***act*** *on it to* ***achieve*** *some* ***goals****.* (Maja J. Matarić, The Robotics Primer, 2007).
- Robotics...
  - … is the study of robots and addresses **perception** and **action** in the physical world (loosely based on (Maja J. Matarić, The Robotics Primer, 2007)).

Robots Today – Diverse Forms and Tasks

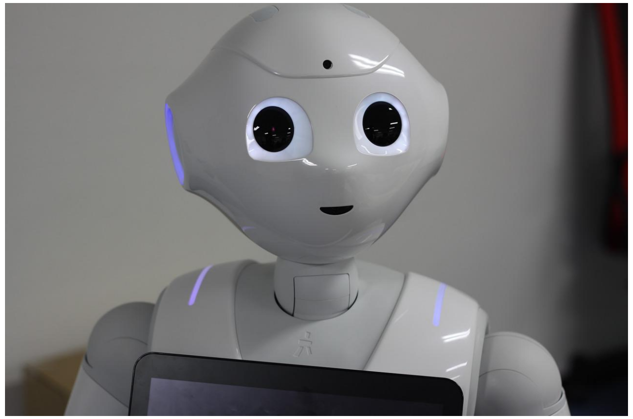
- Robots come in different **shapes** and **sizes**.
- They **interact** with humans through different **channels** to fulfil different **goals**.



Robotic arms in factories

Autonomous drones in warehouses

Vacuum cleaner robots

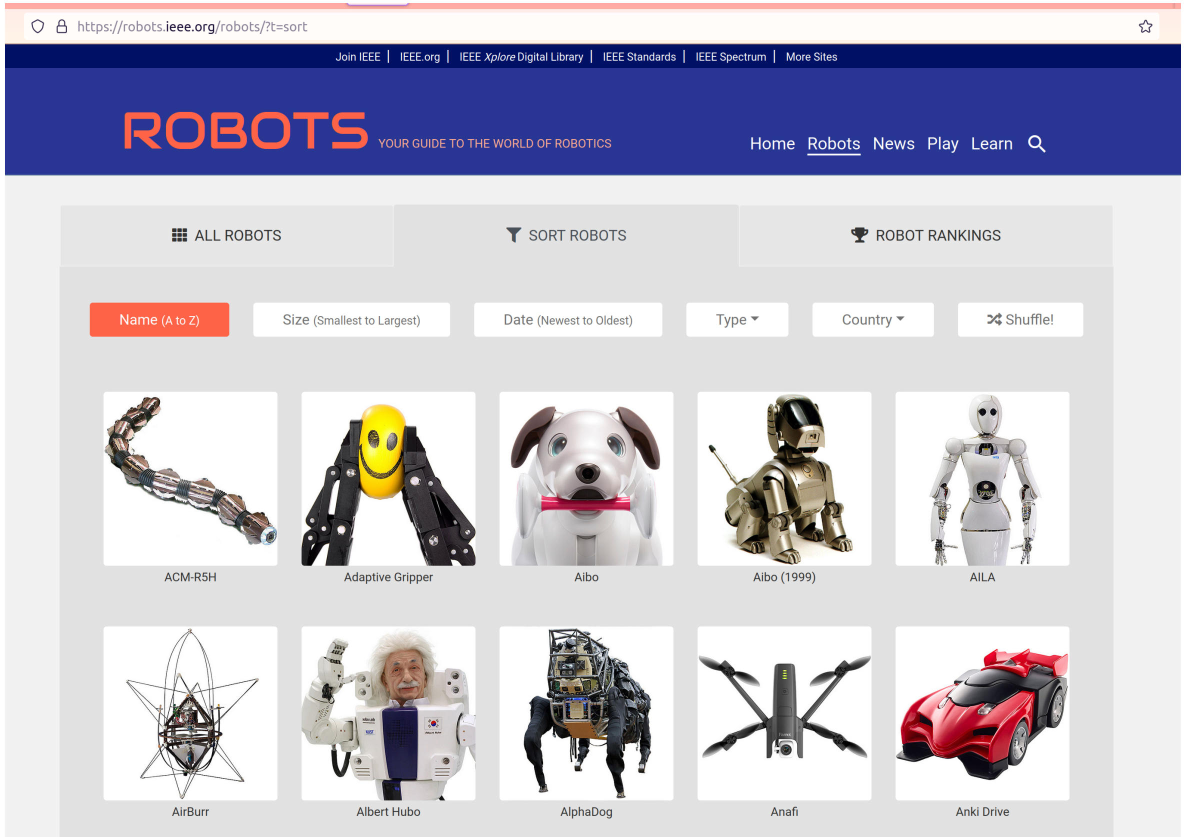


Social robots at homes

A²S

Robotics Application Domains

- Based on purpose, we have robots for (Baraka et al., 2020):
  - Healthcare and therapy
  - Education
  - Entertainment
  - Search and rescue
  - Telepresence
  - Military and security
  - Industry
  - Public service
  - Home and workplace
  - Research
  - ...



https://robots.ieee.org/robots/?t=sort

Baraka, K., Alves-Oliveira, P., Ribeiro, T. (2020). An Extended Framework for Characterizing Social Robots. In: *Jost, et al. (Eds.)* Human Robot Interaction – Evaluation Methods and Their Standardization. Springer, Cham. https://doi.org/10.1007/978-3-030-42307-0_2

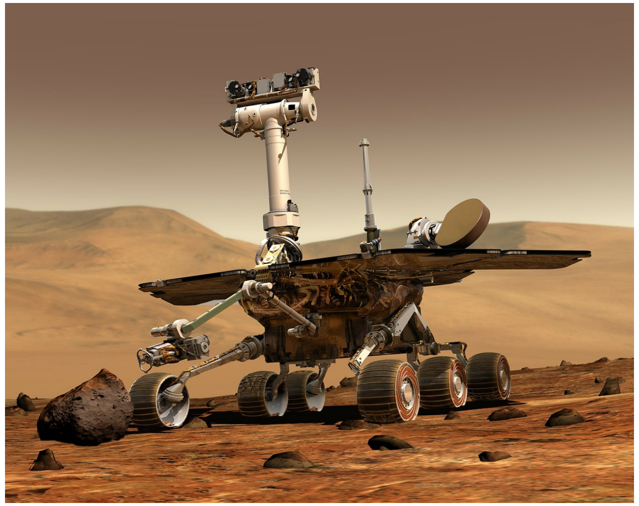
A²S

Autonomous Robots

- Traditionally, the focus has been on making robots **more autonomous**, i.e. to make them **less dependent on human supervision** or **control** to fulfil their tasks.

Teleoperation

Completely controlled by a human
(not truly a robot)



Fully Autonomous

Makes decisions on its own and
operates without any human control

(Maja J. Mataric. 2007. The Robotics Primer (Intelligent Robotics and Autonomous Agents). The MIT Press.)

Levels of Autonomy – One-Dimensional View

- Sheridan and Verplank (1978) suggested 10 levels of autonomy (summarized as follows by (Parasuraman et al., 2000).

Level	Description (The computer...)
10 (High)	… **decides everything** and acts autonomously, **ignoring** the **human**.
9	… **informs** the human **only if it**, the computer, **decides to**.
8	… **informs** the human **only if asked**, or
7	… **executes automatically**, then **necessarily informs** the **human**, and
6	… **allows** the **human** a restricted **time to veto** before automatic execution, or
5	… **executes** that suggestion **if the human approves**, or
4	… **suggests** one **alternative**, or
3	… **narrows** the **selection** down to a few, or
2	… **offers** a complete set of decision/action **alternatives**, or
1 (Low)	… offers no assistance; the **human** must **take all decisions** and **actions**

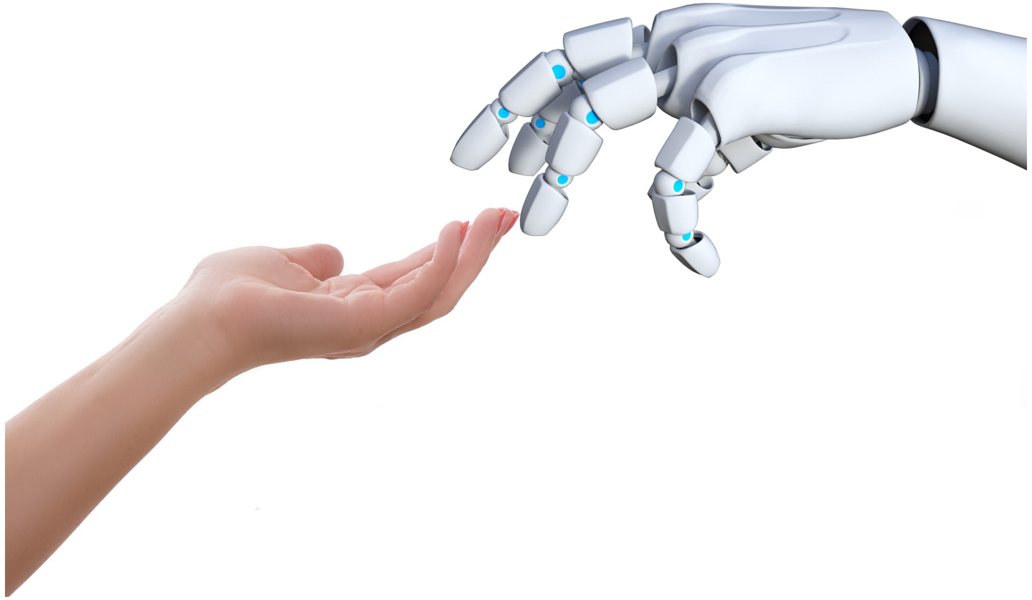
- Sheridan, T. B., & Verplank, W. L. (1978).Human and computer control of undersea teleoperators. Massachusetts Institute of TechnologyCambridge Man-Machine Systems Lab.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation.IEEETransactions on Systems, Man and Cybernetics-Part A: Systems andHumans,30(3), 286–297.

A²S

Humans and Autonomous Robots

- **Nowadays, robots no longer exist only in isolation.**
- **Robots and humans are moving closer to each other** in a wide range of applications.

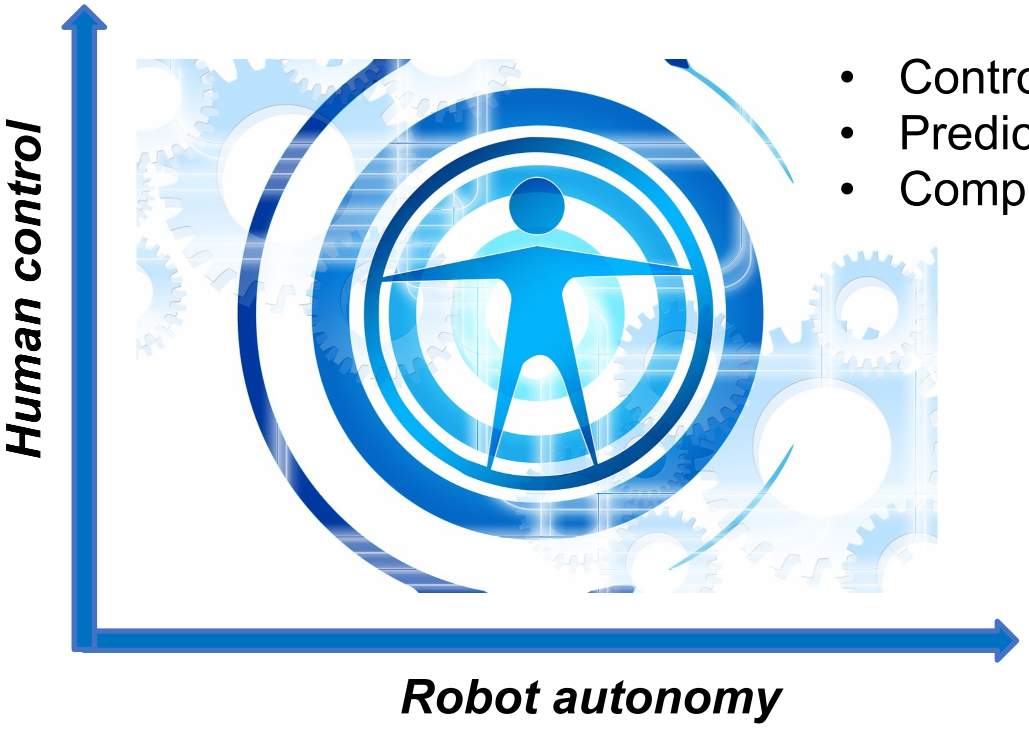
  - Industrial robotic manipulators and humans <u>share the same workspace</u> and collaborate to complete certain tasks.
  - Social robots <u>coinhabit private or public spaces</u> and provide services to people at homes, in shopping malls, or museums.
  - Space robots are being built to <u>assist astronauts</u> on future planetary missions.



A²S

Two-Dimensional, Human-Centered View

- No longer a question of whether robots need humans to act (autonomy).
- But, a question of how robots can **act together with** humans in a safe, reliable and trustworthy manner to **empower** and **enhance** human performance.



- Controllable
- Predictable
- Comprehensible

Source: Ben Shneiderman (2020) Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, International Journal of Human–Computer Interaction, 36:6, 495-504, DOI: 10.1080/10447318.2020.1741118

The focus is now shifting towards the human interaction partner.
The human is at the center of human-robot interaction!

A²S

Industrial Human-Robot Collaboration -- Levels

Coexistence < Cooperation < Collaboration

Coexistence

- Human and robot **share (partially or fully)** the same **physical workspace**.
- But, **activities** of the human and the robot are **unrelated (no shared goals)**.

Cooperation

- Human and robot **coexist**, and have a **shared goal**.
- Robot is aware of the human's involvement in task completion.
- Some coordination of activities.

Collaboration

- Human and robot **coexist**, have a **shared goal** and work on a **shared object**.
- Highest level of interaction.
- Higher joint effort and higher joint situation awareness.

Source: Iina Aaltonen, Timo Salmi, Ilari Marstio, Refining levels of collaboration to support the design and evaluation of human-robot interaction in the manufacturing industry, Procedia CIRP, Volume 72, 2018, Pages 93-98, ISSN 2212-8271, https://doi.org/10.1016/j.procir.2018.03.214.

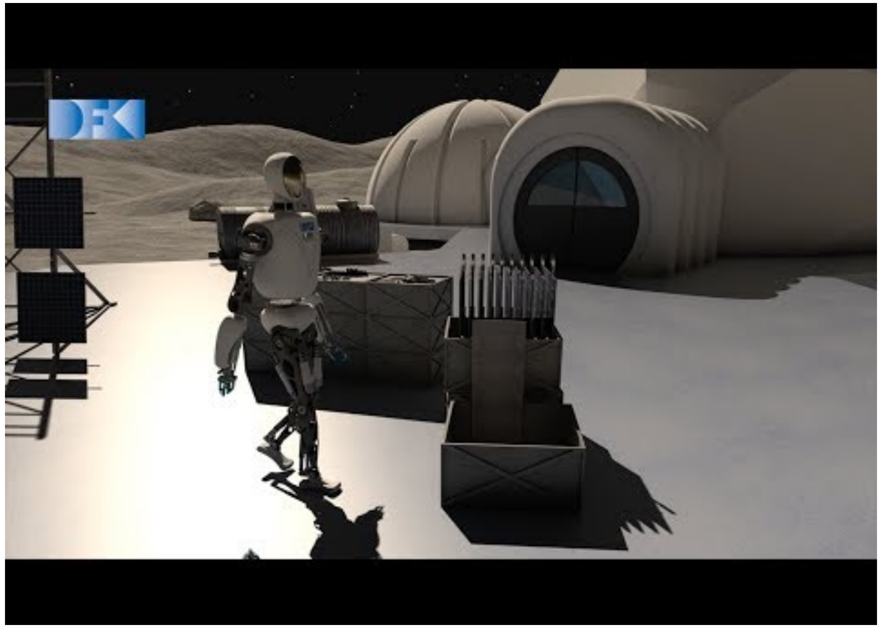
Industrial Human-Robot Collaboration -- Examples

- **No coexistence** => No interaction
  - E.g. robot stays and operates on its own behind a fence.
- **Coexistence** => Minimal or indirect interaction
  - Human enters robot's workspace to fetch an object and the robot slows down to ensure human safety.
- **Cooperation**
  - Robot stops when human enters its workspace to turn a welded object and continues welding when human leaves.
- **Collaboration**
  - Robot holds a machine part in place and the human tightens it.

Source: Iina Aaltonen, Timo Salmi, Ilari Marstio, Refining levels of collaboration to support the design and evaluation of human-robot interaction in the manufacturing industry, Procedia CIRP, Volume 72, 2018, Pages 93-98, ISSN 2212-8271, https://doi.org/10.1016/j.procir.2018.03.214.

A²S

Human-Robot Collaboration in Space



https://youtu.be/UwI3XeXvAjo

- Varying levels of autonomy in robot's actions:
  - Teleoperation
  - Full autonomy

- Robot assistants in space are expected to closely collaborate with human astronauts.

- Shared physical space
  - e.g. a base station on a far away planet
- Shared goals
  - e.g. To set up solar panels
- Shared objects
  - e.g. solar panel

- **Safety** requirements change with the level of collaboration.

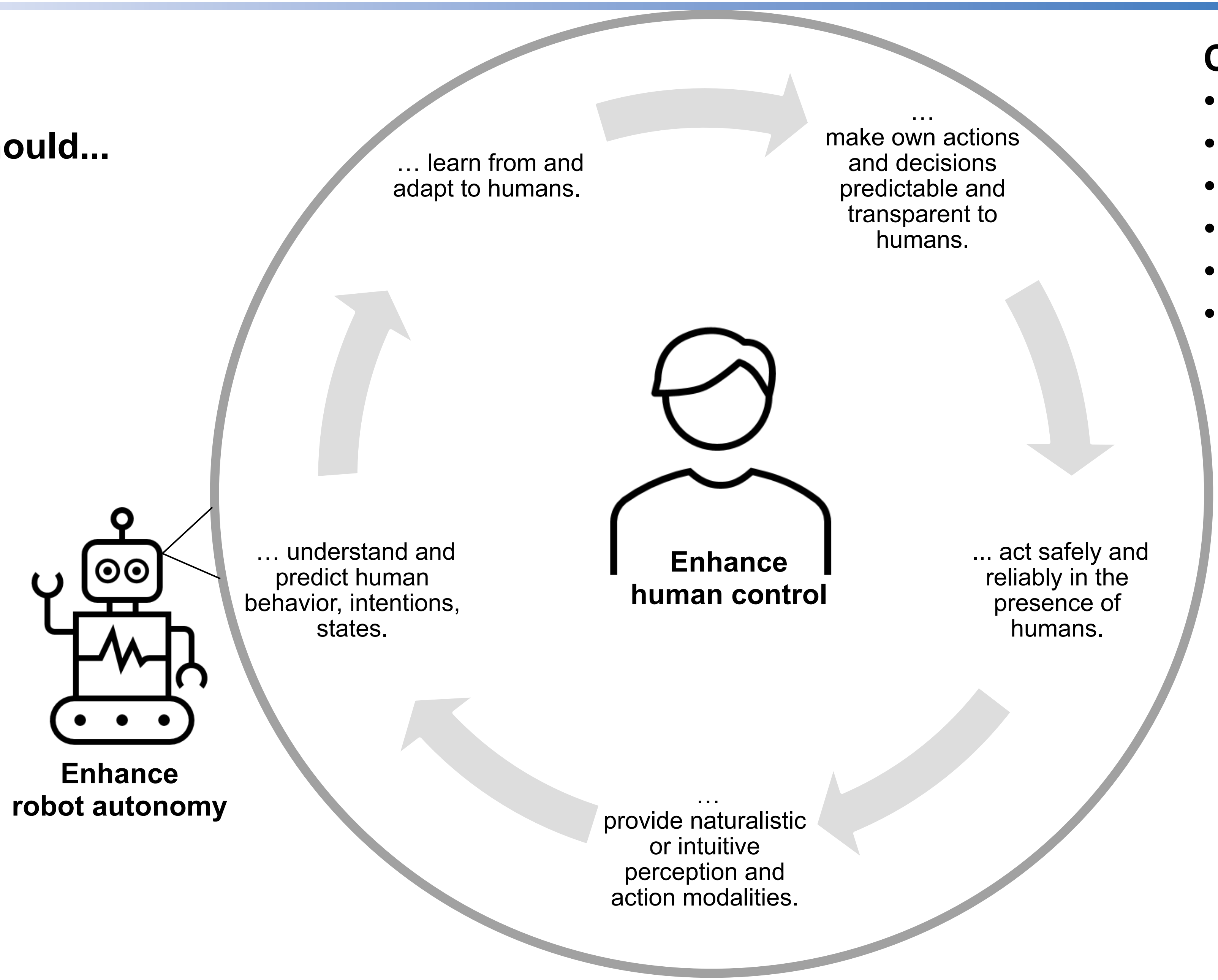
Interaction with Social Robot Companions

- Social companion robots: Engage with human users in social interaction in every day contexts.
- Social interaction: Less structured and more complex than task-based interaction in industrial contexts.
- Humans and robots: Coexistence or copresence in same unstructured and highly dynamic physical space.
- Robot goals: Conflicting goals possible, implying higher levels of robot autonomy and a higher need for **transparency** towards humans.

A²S

Human-Centered Interaction in Robotics

Robots should...



Consider ethical aspects:
- **Safety**
- **Privacy**
- **Fairness**
- **Transparency**
- **Trust**
- **...**

Leading to empowerment of humans!

Human-Centered Interaction in Robotics

1. Humans should be able **interact naturally** and **intuitively** with robots.
2. **Robots** should be able to **understand humans** (behavior, intentions, states).
3. **Robots** should be able to **learn from humans** and **adapt** to them.
4. **Humans** should be able to **understand robots** (behavior, intentions, states).
5. **Interaction** with robots should be **safe**, **reliable**, and **trustworthy** for humans.
6. Enhanced robot autonomy and **enhanced human control** should go hand-in-hand.
7. **Ethical issues** should be identified and **resolved** early.

1. Verbal and Nonverbal Interaction

- Natural interaction modalities
  - Verbal: Speech
  - Nonverbal: Overt versus Covert
    - Gestures, facial expressions, body movements, touch, brain signals, etc.
- Alternate sensory and action modalities (e.g. touchscreens, LED displays, etc.) should be intuitive enough to require less effort from humans.

- One-way communication
  - From human to robot: Give verbal commands to the robot, point to something, etc.
  - From robot to human: Answer a question, ask for help, smile at the human, etc.
- Two-way communication
  - Engage in a conversation

Doncieux, S., Chatila, R., Straube, S. *et al.* Human-centered AI and robotics. *AI Perspect* 4, 1 (2022). https://doi.org/10.1186/s42467-021-00014-x

A²S

2. Understanding Human Behavior

- Humans' behavior, beliefs, desires, intentions, emotions, etc.
  - **Interpersonal** differences in expression.
    - Each person is different at physical, physiological and cognitive levels.
    - Behavior and communication are influenced by several factors, including sociocultural ones.
  - **Intrapersonal** difference in expression.
    - Same person might show different behavior at different times and in different contexts.
  - **Multimodal** expression
    - Not one, but different types of signals (verbal, nonverbal) should be analyzed simultaneously to make good inferences.
  - **Context-based** semantic differences
    - Same signals might carry a different message or meaning depending on the context.

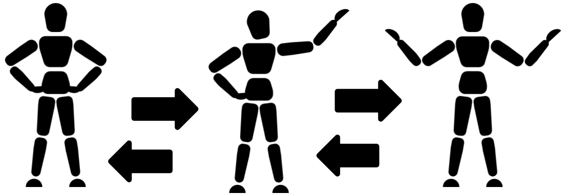
A²S

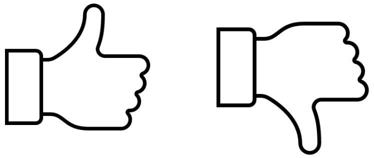
3. Learning from Humans

Interactive Reinforcement Learning

Human input could be used to guide and accelerate the robot’s learning, or to change its optimal behaviour (personalisation).

- What types of human input can be given?
  - Demonstration
    - Inverse RL (learn reward function from demonstrations)
    - Imitation learning (learn the policy from demonstrations)
  - Advice or instruction (often verbally)
    - Converted into rules and added to knowledge-base.
    - Used for reward shaping and to learn reward functions.
  - Evaluative feedback -- How good or bad was the robot's action?
    - Explicit versus implicit feedback
    - Used as reward together with other reward functions.





A²S

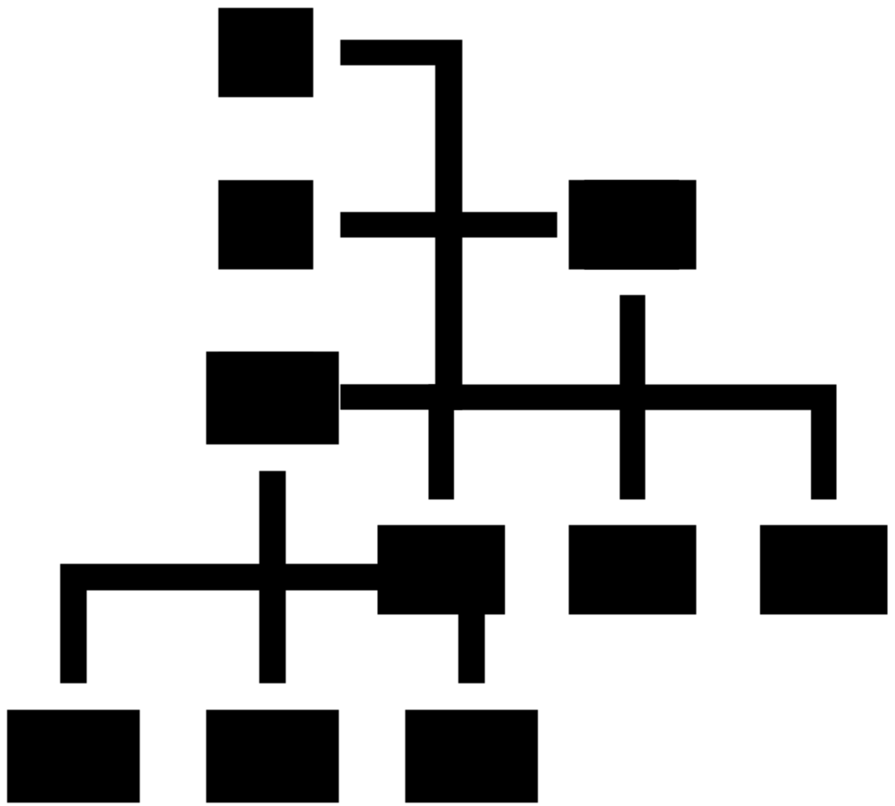
4. Understanding Robot's Capabilities and Intentions...

- Robot's capabilities should be transparent to the user:
    - **What can the robot do? What can it not do?**

- Robot's actions, intentions and decisions should be predictable, transparent and comprehensible to humans.
    - **What is the robot doing? Why is it doing it?**

- Robot should be capable of explaining its own behavior and decision, if needed.
    - **Proactive** explanation versus **on-request** explanation
    - **Verbal** and/or **nonverbal** (e.g. gaze, gestures, etc.)
    - <u>Eventually, robot behavior becomes understandable without requiring explicit explanation.</u>

Explainable human-robot interaction architectures



A²S

5. Safe, Human-Aware Physical Interaction

Design levels for safety:

Doncieux, S., Chatila, R., Straube, S. *et al.* Human-centered AI and robotics. *AI Perspect* 4, 1 (2022). https://doi.org/10.1186/s42467-021-00014-x

- Compliant materials and elements (e.g. springs)
  - Primarily to ensure safety of humans interacting physically with the robot.
- Compliant control based on haptic feedback
  - To reduce human effort in guiding or controlling robot movements, e.g.
    in exoskeletons for rehabilitation.

- Perception-based adaptive control
  - Adaptive control based on exteroceptive outputs, e.g. slow down robot's speed progressively when human approaches.
- Human-aware planning and execution
  - Understand beliefs, intentions, emotions, etc. of humans to jointly plan, coordinate and execute actions, e.g. hand-over of a tool.

https://youtu.be/mDODMNMC5zc

A²S

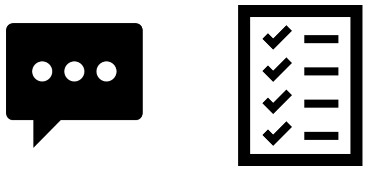
6+7. User-Centered Design and Development

- User-centeredness in the design of robots and interaction:
  - Gather **user expectations** and **requirements** early on (from mechanical design onwards).
  - Iteratively **evaluate, analyze, and improve** the design.
  - Design **intuitive interfaces** that support ease and transparency during interaction.
  - Reduce the amount of **adaptation** needed on the side of the humans.
  - Identify, analyse and address **ethical issues** from the beginning of the process.
    - Violation of privacy
    - Potential for physical harm or exhaustion
    - Discrimination or violation of human dignity
    - Development of affective bonding with the robot
    - ...

7. Interaction Experiments

- Conduct qualitative and quantitative evaluations of interaction in well-defined and reproducible experiments.

- What are evaluated?
  - Technical efficiency and usage
  - Influence or impact on human interaction partner
    1. How do humans perceive the robotic interaction partner?
       - Likeable, trustworthy, predictable, human-like, etc.
    2. How do humans experience the interaction with the robot?
       - Engagement, effort, predictability, acceptability, etc.
    3. What impact does the interaction have on human skills or socioemotional state?
       - E.g. therapeutic effect, sociocognitive skills, motor skills, etc.
    - Immediate / short-term / long-term studies
  - Require clearance from an ethics committee!



A²S

Conclusion

- In today's lecture, you learnt to:
  1. Define a robot.
  2. List some of the application domains of robots.
  3. Describe Sheridan and Verplank's 10 levels of robot autonomy.
  4. Distinguish between co-existence, cooperation and collaboration.
  5. Give examples of human-robot interaction from different application domains.
  6. List and explain the central aspects of human-centered interaction in robotics.
     1. Intuitive interfaces for interaction
     2. Robots understanding humans
     3. Robots learning from and adapting to humans
     4. Humans understanding robots
     5. Safe, reliable, trustworthy interaction
     6. Enhanced human control
     7. Ethical considerations

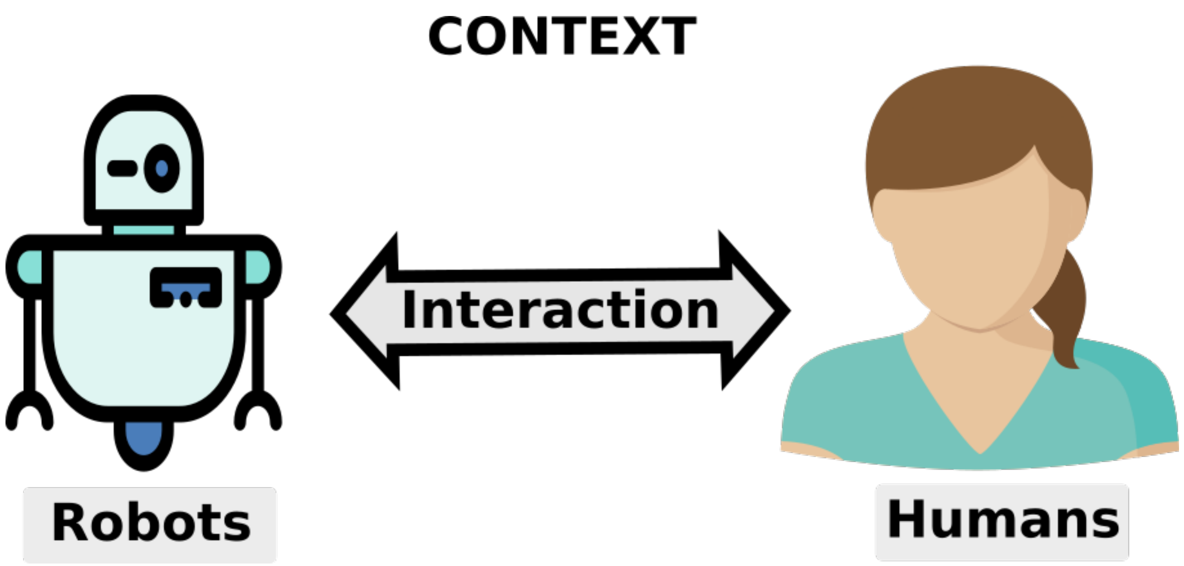
Variability in Human-Robot Interaction Scenarios

https://youtu.be/2ZUn9qtG8ow

https://youtu.be/hEgJOmRkAKg

What differences do you observe between the two scenarios?

Human-Robot Interaction: Elements



- **Interaction partners**
  - Humans and robots
  - Their respective roles
- **Interaction:** Type and level, frequency, time duration, modalities, etc.
  - Types: Physical, verbal, nonverbal
  - Coexistence, cooperation, collaboration
- **Interaction context**
  - Application domain
  - Purpose / goal of interaction
  - Interaction setting
    - Controlled versus uncontrolled
    - Active versus passive

Reproducibility and Comparability

- Wide range of robots and application fields.
- Therefore, a lot of variability is possible in human-robot interaction scenarios.
- Common practice: Experiments are based on individual use cases.
- Challenges:
  - Results from a single experiment (use case) cannot be generalized.
    - Experiments and scenarios should be **reproducible**.
  - Conflicting results from experiments evaluating similar hypotheses are difficult to explain.
    - Experiments and scenarios should be **comparable**.
- Reproducibility and comparability of experiments are necessary to advance the knowledge about the field of human-robot interaction.

A²S

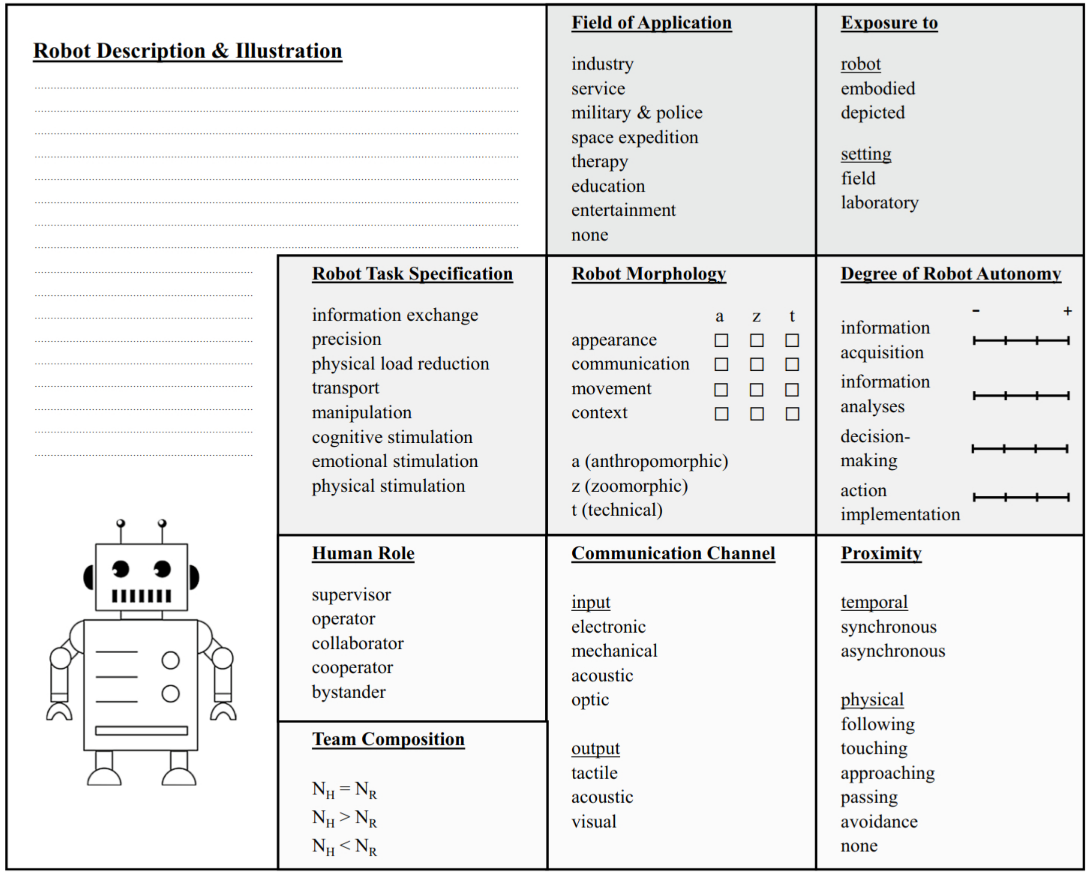
The Problem and The Solution

- Why is reproducibility and comparability difficult in human-robot interaction?
  - Omission of important details describing the interaction scenario and experimental s etting.
  - Use of inconsistent terminology with different semantics.
- Solution:
  - We need **a common language** to **completely describe** any human-robot interaction scenario.
    - In other words, all elements and aspects of a human-robot interaction scenario should be describable using standard, commonly agreed terminologies.
    - That is to say, we need a **taxonomy**!

HRI Taxonomy: Overview

- Onnasch and Roesler (2021) proposed a three-level taxonomy to describe the different elements of a human-robot interaction scenario.
- This taxonomy takes inspiration from, combines and extends several existing frameworks proposed in literature.
- Hierarchical framework
  - First layer (top): Classifies interaction context
  - Second layer: Classifies the robot
  - Third layer: Classifies the human-robot team

Fig. 1 of Onnasch, L., Roesler, E. A Taxonomy to Structure and Analyze Human–Robot Interaction. *Int J of Soc Robotics* 13, 833–849 (2021). https://doi.org/10.1007/s12369-020-00666-5



This canvas is downloadable from https://www.psychologie.hu-berlin.de/de/prof/ingpsy/hri-taxonomy/psychofinalll.pdf

Work with the HRI Canvas

- Download the canvas from supplementary materials of Homework #1.
- Can you edit the PDF digitally?
- Create two copies – one for each example shown in the videos on previous slide.
- Work in pairs.
- For each quiz, discuss and fill in the corresponding fields in both sheets.

Layer 1: Interaction Context Classification

Field of Application

- What is the field of application of the interaction with robots?

Field of Application	Exposure to
industry	robot
service	embodied
military & police	depicted
space expedition	
therapy	setting
education	field
entertainment	laboratory
none	

Exposure to

- What form of representation of the robot is the human subject exposed to?
  - Embodied
    - A physical robot whose body can be touched
  - Depicted
    - Virtual agents (or, computer animations)
    - Videos or images of robots
- Where is the interaction taking place?
  - Laboratory
    - A controlled setting
  - Field
    - Outside a laboratory, "in the wild"

A²S

Fields of Application

- Industry – e.g. assistance at assembly lines
- Service – lawn mowing, vacuum cleaning, etc.
- Military and police – bomb disposal, wildfire control, search and rescue, ...
- Space expedition – exploring astronomical bodies
- Therapy – behavioral therapy, rehabilitation
- Education – language learning, solving mathematical problems
- Entertainment – for use as pets or toys
- None – when no information is provided about the field of application in the study description.

- Based on the field of application, we can predict the general characteristic of both the human users as well as the environment.

Layer 2: Robot Classification

- Robot task specification
  - What task does the robot perform in the field of application?
    - An abstract goal or purpose of the interaction.
- Robot morphology
  - How does the robot look?
  - What does it resemble when it communicates or moves?
  - How is it framed in the interaction context?
- Degree of robot autonomy
  - How much human intervention does the robot need in order to interact?

Robot Task Specification

information exchange
precision
physical load reduction
transport
manipulation
cognitive stimulation
emotional stimulation
physical stimulation

Robot Morphology

	a	z	t
appearance	☐	☐	☐
communication	☐	☐	☐
movement	☐	☐	☐
context	☐	☐	☐

a (anthropomorphic)
z (zoomorphic)
t (technical)

Degree of Robot Autonomy

\- +

information acquisition
information analyses
decision-making
action implementation

A²S

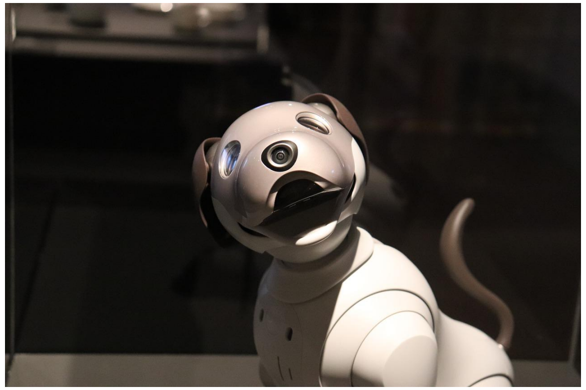
Robot Task Specification

- **Information exchange**
  - Gather, analyse and transfer information to humans
    - E.g. Space / underwater exploration missions, search and rescue
- **Precision**
  - Tasks like minimally invasive surgeries that require high precision
- **Physical load reduction**
  - Reduce the physical load or physical effort of humans to do a specific task
    - E.g. to walk, to lift heavy objects, carry objects over a distance, fix objects, etc.
- **Transport**
  - Transport objects from one location to another.
    - E.g. in warehouses, hospitals
- **Manipulation**
  - Physically modifies the environment
    - E.g. welding, pick and place
- **Cognitive stimulation**
  - Engage user at cognitive level through verbal and nonverbal communication, e.g. to educate, train.
- **Emotional stimulation**
  - Stimulate emotional expressions and reactions during interaction.
    - Verbal, nonverbal communication
    - Through companionship, entertainment
- **Physical stimulation**
  - Rehabilitation support to regain control or stability of physical body functions

Robot Morphology

- Why is it important for interaction?
  - Creates expectations about functional and communicational capabilities of the robot.
    - If a robot looks like a human, then the user would expect to interact with it verbally.
    - Resemblance with known objects enables an intuitive interaction.
- Anthropomorphic, zoomorphic or technical
  - In: appearance, communication, movement, context (framing)

Anthropomorphic



Zoomorphic

Technical

A²S

Framing

- In the social sciences:
  - *"To frame is to* ***select some aspects*** *of a perceived reality and* ***make them more salient*** *in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."* -- *Framing*: Toward *clarification of a fractured paradigm*. *Entman, Robert* M. Journal of Communication; Autumn *1993*; 43, 4; ABI/INFORM Global.

- Through personified or zoomorphic narratives, even a technical-appearing robot can be made to be perceived as a child or a pet.

- Framing is done in human-robot interaction experiments or studies in the form of a narrative (a story).
  - Provides info about the context of the interaction.

Degree of Robot Autonomy

- Autonomy:
  - *"The extend to which a robot can operate in the tasks it was designed for (or that it creates for itself) without external intervention."* -- Baraka et al. (2020).

- Four stages of doing a task:
  1. Information acquisition
  2. Information analyses
  3. Decision-making / Action selection
  4. Action implementation

- Autonomy in doing each stage of a task
  - For practical convenience, three levels:
    - low/none, medium, high/complete

Layer 3: Team Classification

- **Human role**
  - What role does the human play in the interaction?
- **Team composition**
  - How many humans and robots are involved in the interaction?
    - Exact numbers can be specified.
- **Proximity**
  - Spatial
    - Physical distance between the workspaces of the human and the robot.
  - Temporal
    - Do the humans and robots work at the same time?
      - Yes ==> Synchronous
      - No ==> Asynchronous

- **Communication channel**
  - Input:
    - How can the robot perceive the human?
    - How can information flow from human to robot?
  - Output:
    - Which of the human senses are involved in receiving information from the robot?

Human Role	Communication Channel	Proximity
supervisor	input	temporal
operator	electronic	synchronous
collaborator	mechanical	asynchronous
cooperator	acoustic	
bystander	optic	physical
		following
Team Composition	output	touching
	tactile	approaching
$N_H = N_R$	acoustic	passing
$N_H > N_R$	visual	avoidance
$N_H < N_R$		none

A²S

Human Role

- **Supervisor**
  - Monitors and gives instructions.
- **Operator**
  - Controls the robot
    - Direct control of robot's actions.
    - Changes internal implementation, when behaviour not acceptable.
- **Collaborator**
  - Humans and robots share same goals and subgoals and work jointly.
- **Cooperator**
  - Humans and robots do different tasks to fulfil the shared overall goal.
- **Bystander**
  - No interaction, no shared goals, but share the same space.
    - Mental models of the robot and its actions are still necessary to avoid collisions.
    - E.g. Passers-by who try to avoid collisions with a mobile robot.

Spatial Proximity

- In decreasing order of physical proximity:
  - Following
    - Stable physical contact for longer time. E.g. using joystick or manipulating same objects.
  - Touching
    - Share the same workspace and interact directly with physical contact.
  - Approaching
    - Work closely in the same space but no physical contact.
  - Passing
    - Workspaces overlap partially or fully, but contact is prevented.
  - Avoiding
    - Workspaces not close to each other, and contact is avoided.
  - None
    - Work in different environments (e.g. different planets, watching videos of a robot).

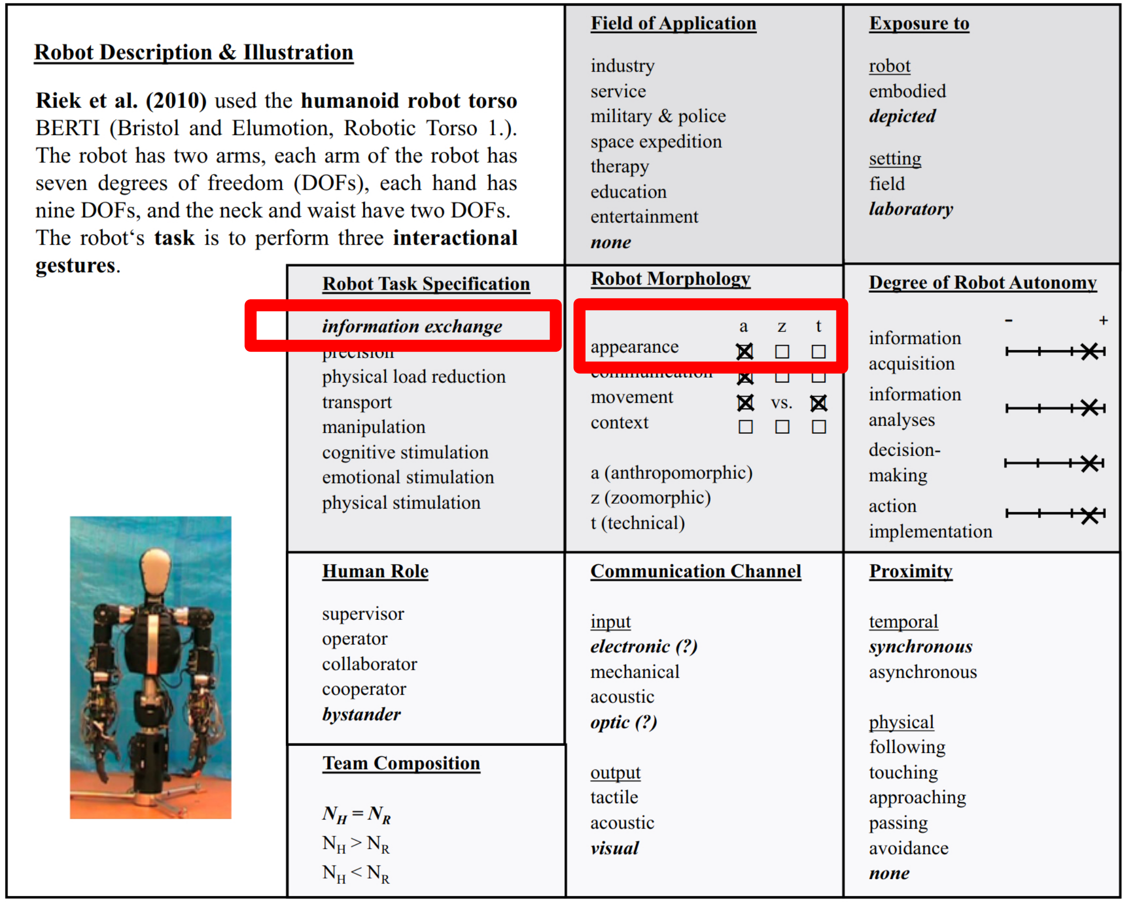
A²S

Communication channels

- **Input to robot from human**
  - Electronic
    - E.g. remote control, joystick, touch screen, etc.
  - Mechanical
    - E.g. movement of the robot's arm
  - Acoustical
    - E.g. verbal instructions
  - Optical
    - E.g. gestures
- **Output from robot to human**
  - Tactile
    - E.g. vibrations
    - Human sense of touch
  - Acoustical
    - E.g. tones, animal sounds
    - Human sense of sound
  - Visual
    - E.g. robot's body movements

Comparing Two HRI Studies

An error in Fig. in the paper



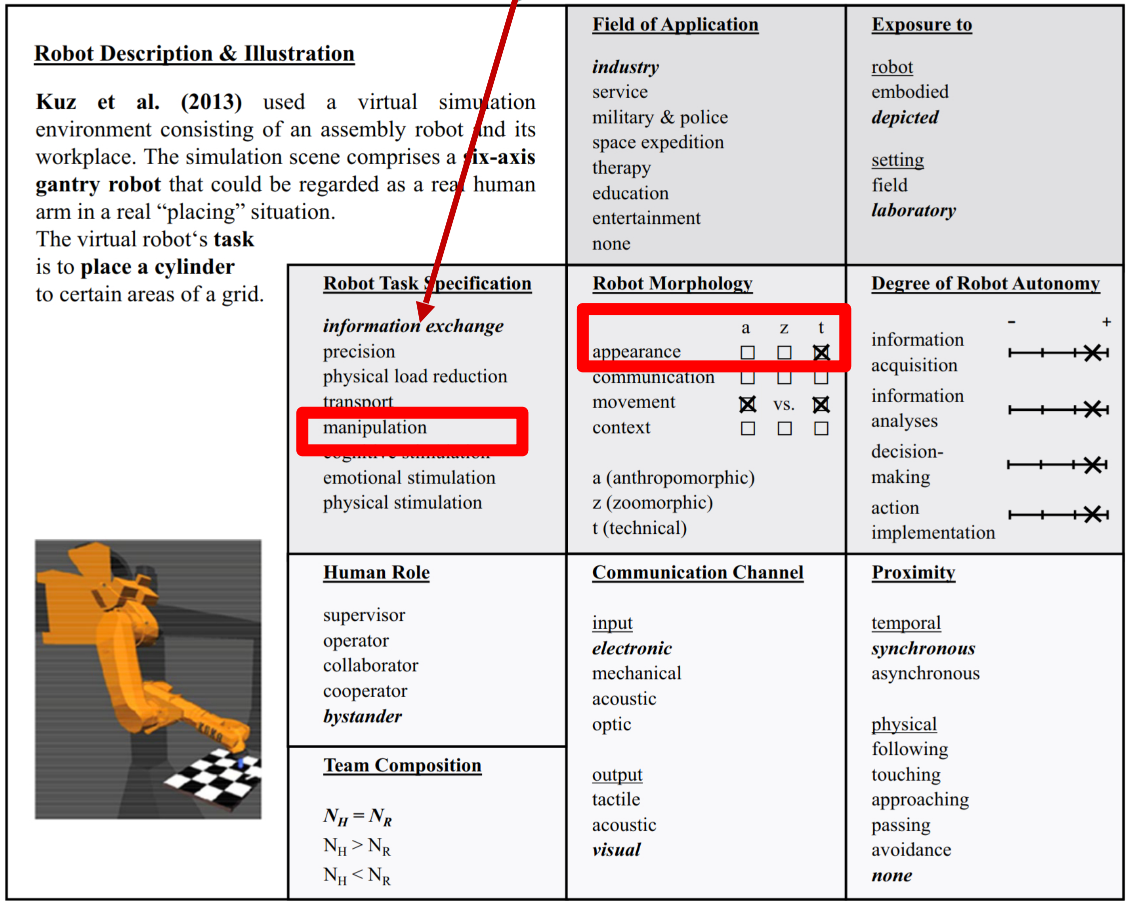


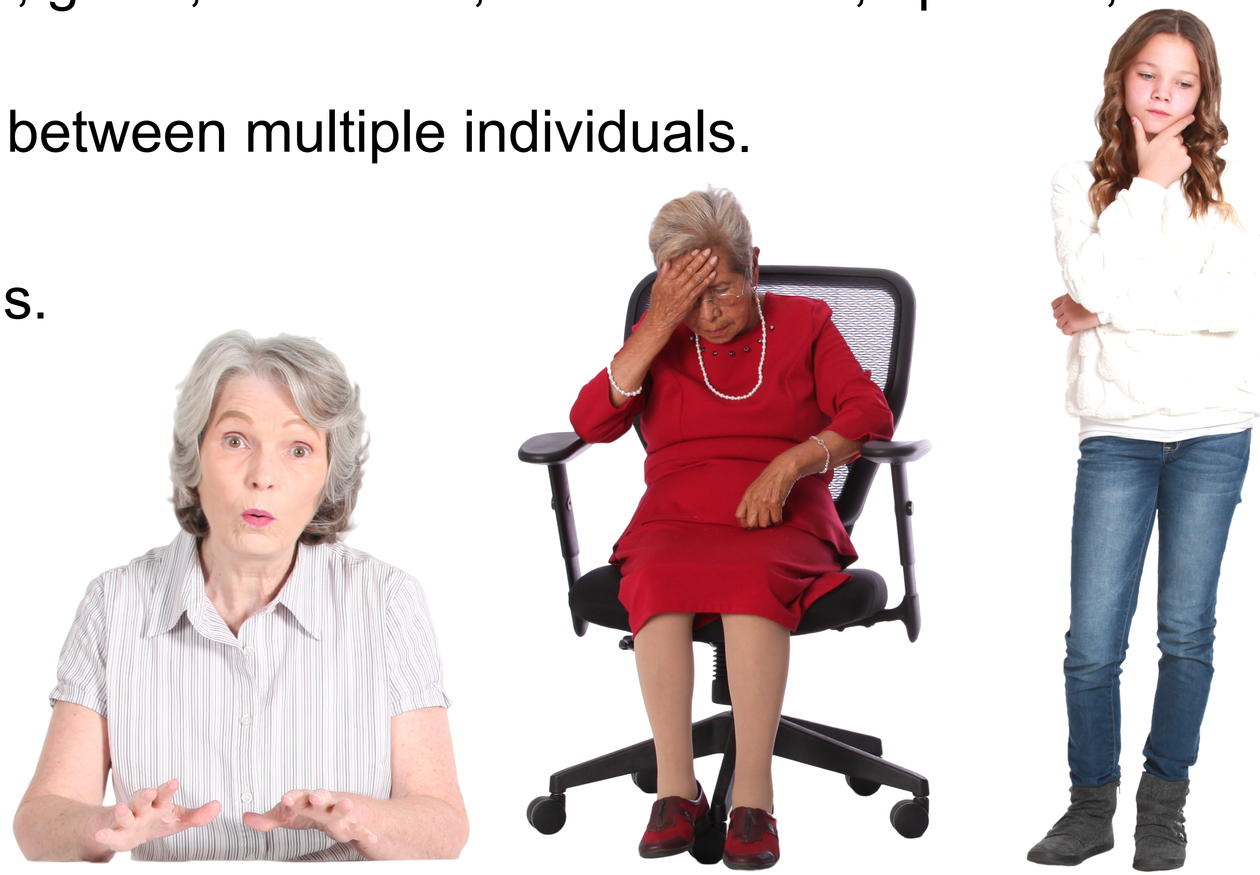
Fig. 4 of Onnasch, L., Roesler, E. A Taxonomy to Structure and Analyze Human–Robot Interaction. *Int J of Soc Robotics* 13, 833–849 (2021). https://doi.org/10.1007/s12369-020-00666-5

A²S

- Kuz S, Mayer MP, Müller S, Schlick CM (2013) Using anthro-pomorphism to improve the human-machine interaction in indus-trial environments (Part I). In: International conference on digital human modeling and applications in health, safety, ergonomics and risk management. Springer, Berlin, Heidelberg, pp 76–85.
- Riek LD, Rabinowitch TC, Bremner P, Pipe AG, Fraser M, Robinson P (2010) Cooperative gestures: Effective signaling for humanoid robots. Human-Robot Interaction (HRI). In: Pro-ceedings of the 5th ACM/IEEE international conference on human-robot interaction. IEEE Press, pp 61–68.

Why do We Communicate?

- Communication is about sending and receiving messages.
- We communicate:
  - To exchange information.
  - To understand each other's behaviors, goals, intentions, mental states, opinions, perspectives, preferences, etc.
  - To coordinate or synchronize actions between multiple individuals.
  - To operate cohesively within groups.
  - To satisfy own as well as shared goals.
  - To survive.
  - ...



A²S

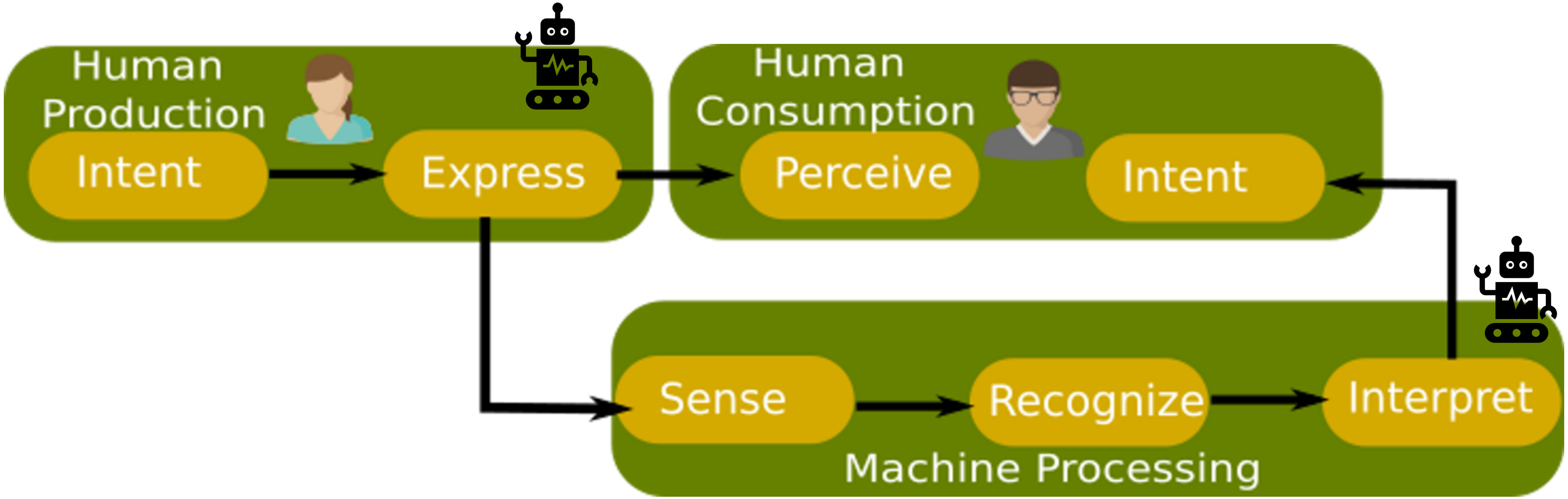
Is Communication Easy?

https://youtu.be/dBT6u0FyKnc

- Obviously not!
- Researchers have looked at how humans communicate, in order to identify:
  - The **elements** of communication
  - The **modes** of communication
  - The **processes** involved in communication
  - The **factors** that influence communication
- .. and to create **a model for interhuman communication and comprehension**...
- … which could be used to study communication problems as well as to design artificial communicative agents.

Human Communication (1/2)

- Involves two broad categories of processes:
  1. Perceiving and interpreting signs
  2. Generating and expressing signs



Recreation of Fig. 3 in (Narayanan and Georgiou, 2013)

Human Communication (2/2)

- Has two components:
  1. Verbal
     - Involves the use of a language
     - Discrete signals (phonemes, words)
     - E.g. spoken language, sign language
  2. Nonverbal
     - Involves movement of different parts of the body, often in sync with verbal component
     - Continuous signals
     - E.g. intonation, muscle or joint movement.

A²S

Nonverbal Communication

https://youtu.be/ilmDN2e_Flc

What nonverbal behaviours do you observe?

- Raising or dropping of ears
- Blinking of eyes
- Opening and closing of mouth
- Eye movements
- Head, hand, body movements
- ….

- Multiple modalities and all are synchronized to convey emotional responses to objects
  - Liking, happiness
  - Repulsion, fear

Nonverbal Communication Modes

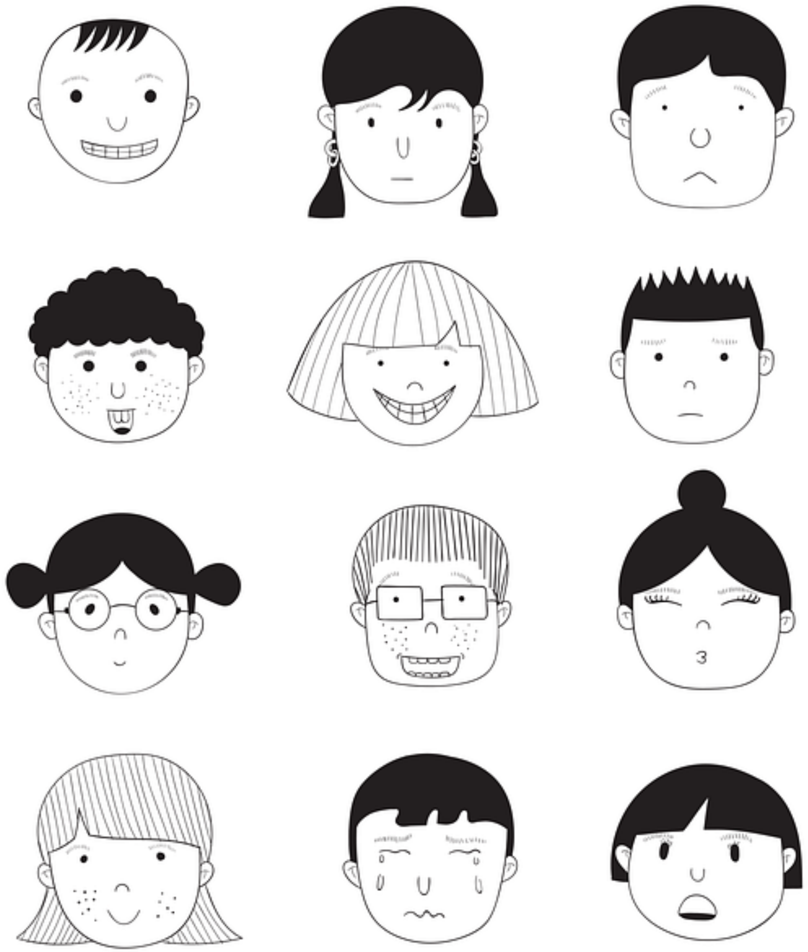
- Kinesics
  - Involves movements and positioning of the robot's body.
  - E.g. **Facial expressions**, hand and arm **gestures**, head and body movements, **eye gaze**
- Proxemics
  - The physical distance between agents influences communication.
    - Hall's four proxemic zones: **intimate < personal < social < public**. (Hall, 1966)
    - Also indicates the relationship between the agents.
- Haptics
  - Physical feedback to interaction partner; e.g. vibrations, touch
- Vocalics
  - Nonverbal parts of speech or other vocalizations; e.g. pitch, loudness, tempo, etc.
- Chronemics
  - How time-related aspects influence communication.

Source: https://open.lib.umn.edu/communication/chapter/4-2-types-of-nonverbal-communication/

A²S

Facial Expressions: Recognition and Generation

Facial Expressions



- Reflect our state of mind: emotions, pain, distraction, interest, …
  - (Darwin, 1872; Kunz and Lautenbacher, 2014; Gjoreski et al. 2020; Yeasin et al., 2006)
- Formed by facial muscle movements:
  - **Facial Action Units (AUs)** defined in the Facial Action Coding System (Ekman & Friesen, 1978)

A²S

Facial Action Coding System (FACS)

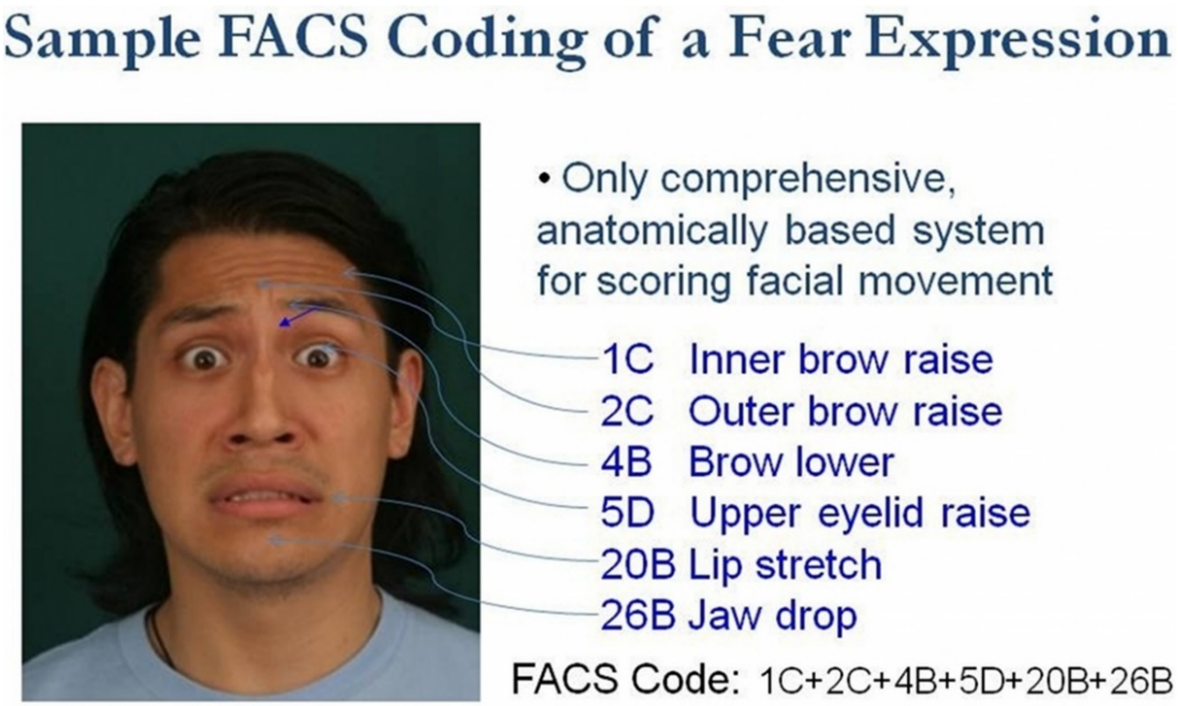
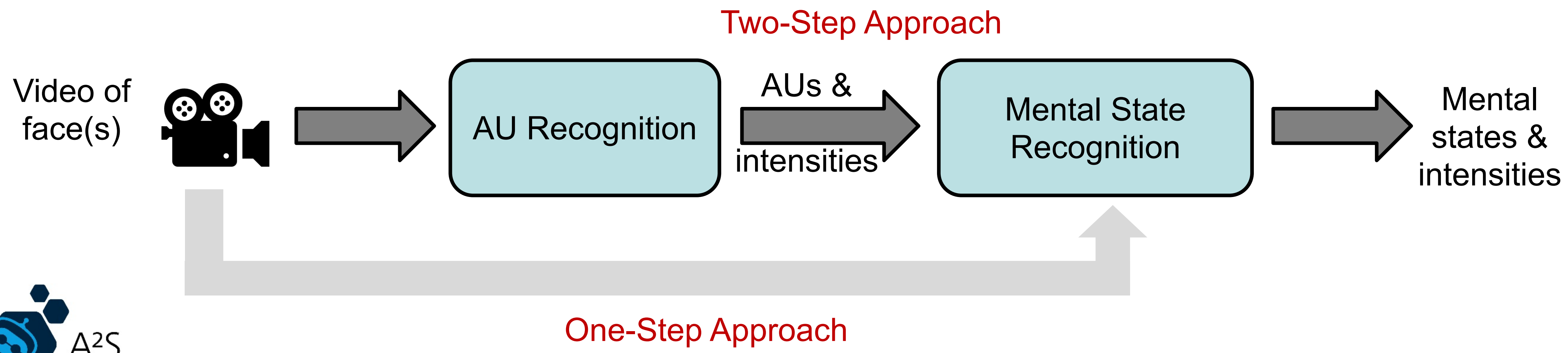


Image source: David Matsumoto and Paul Ekman (2008), Scholarpedia, 3(5):4237

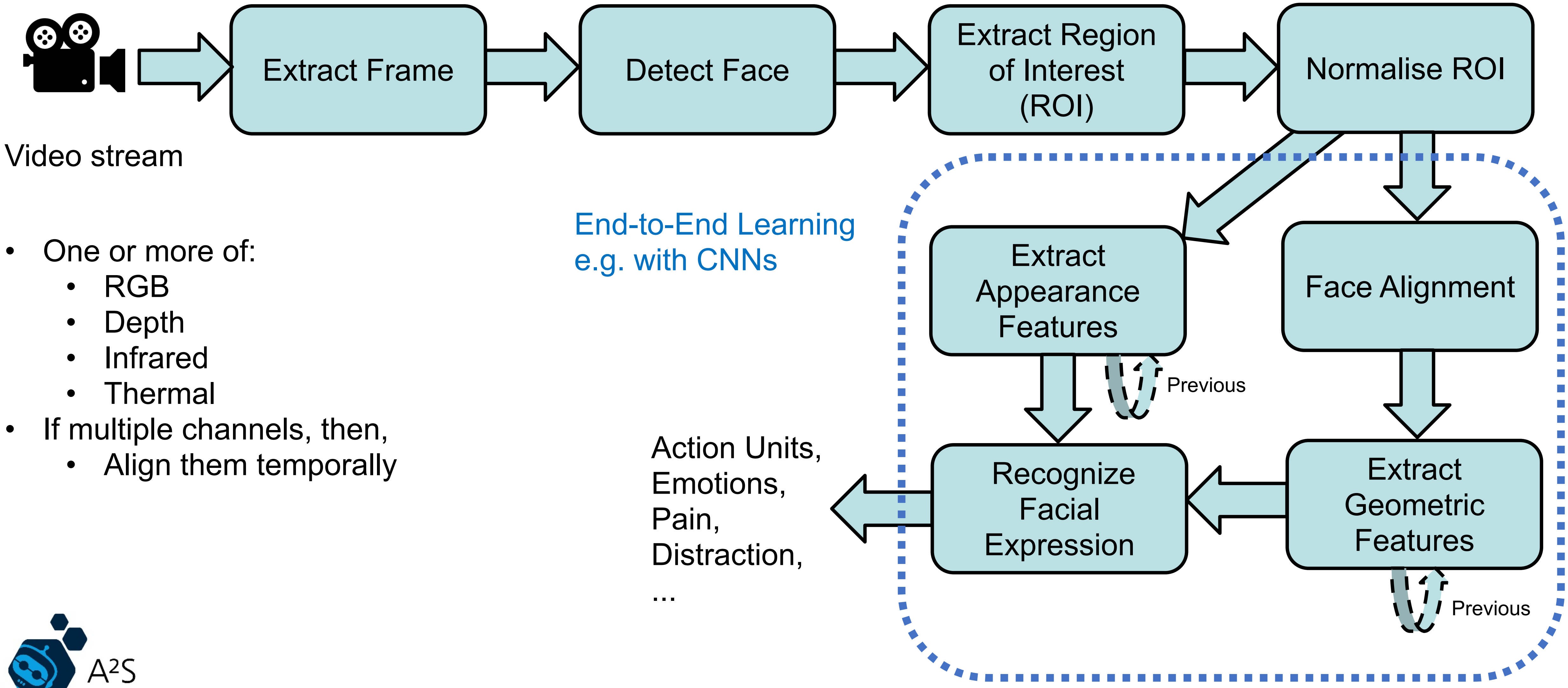
- 44 different AUs
- Produced by individual muscles or groups of muscles
- Can be visually differentiated from one another.
- Can describe any facial expression
- Coding of facial expression
  - AU label (1 to 44)
  - AU intensity (absent or [A, E])

A²S

- Sign judgment
  - Detect AUs and their intensities
- Message judgment
  - Recognize the mental state that is being conveyed by the facial expression
    - e.g. pain, emotions, distraction, engagement, etc.
  - One-step versus two-step approaches (Hassan et al., 2021)

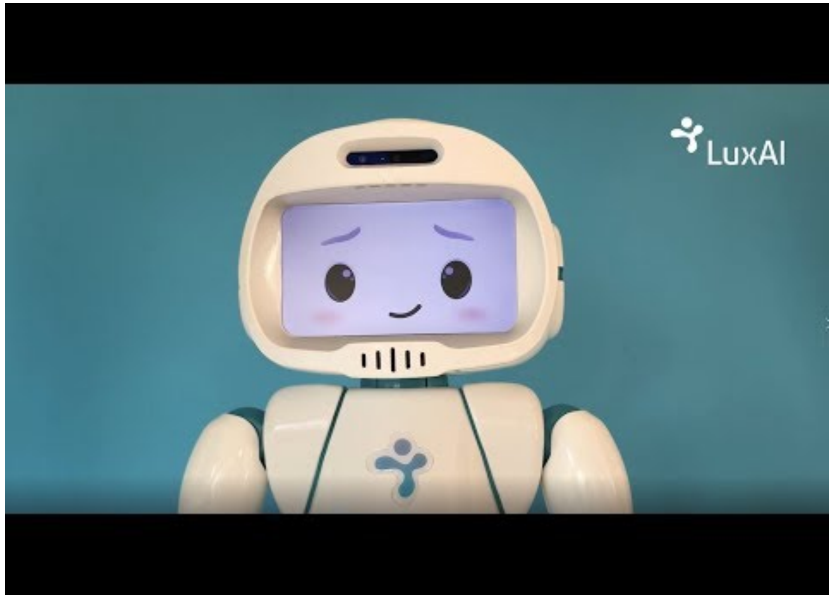


Features for Facial Expression Recognition



- One or more of:
  - RGB
  - Depth
  - Infrared
  - Thermal
- If multiple channels, then,
  - Align them temporally

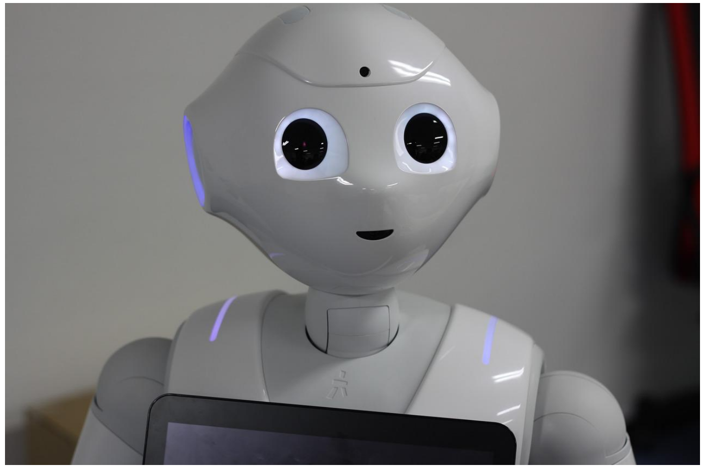
Facial Expression Generation



Virtual, human-like expressions
https://youtu.be/QnTbtbZupWE

Physical, human-like expressions
https://youtu.be/dG7kNhxrOG8

Ameca



Artificial / Technical modality – LEDs around eyes

Realization of facial expressions (e.g. smile) is embodiment-specific and differs from one robot to another.

- True also for other nonverbal cues such as gaze.

A²S

Facial Expression Generation

- Option 1:
  - Create templates for **facial expressions of emotions** based on the embodiment, with the option to vary the intention of the expression.
  - Limited no. of expressions possible.
- Option 2:
  - Create templates for **components of facial expressions** (similar to AUs), and combine several components at various intensities.
  - Greater variety of expressions possible.

Eye Gaze: Recognition and Generation

Role of Gaze in Communication

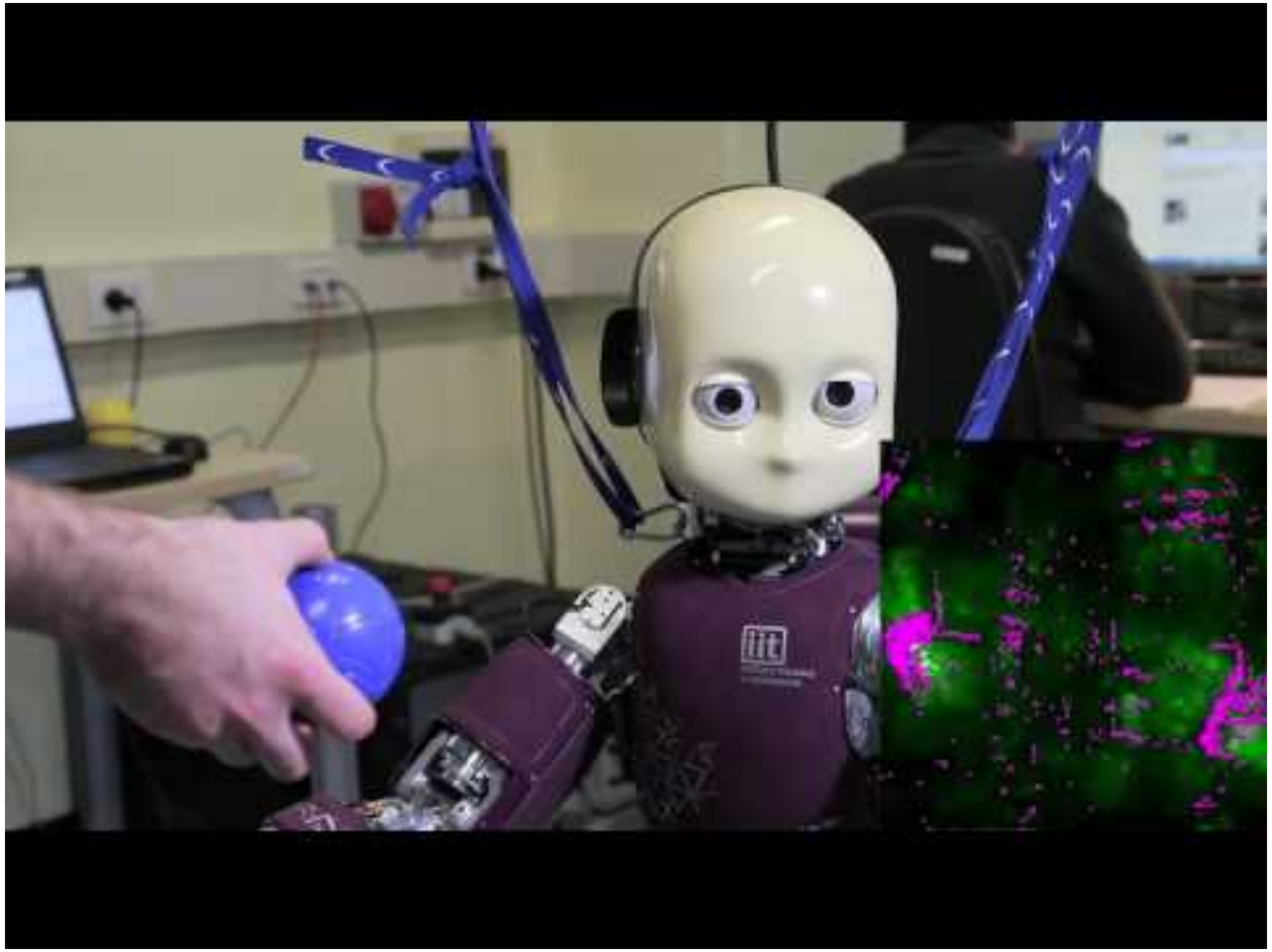
- Eyes play a crucial role in social communication.
- Eye gaze helps to communicate our intent, emotions, engagement, interest, attention, etc. to others. E.g.
  - What is a person looking at?
  - What might the person want to do?
    - E.g. pick up an object; seize conversation floor; cross a street, ...
  - How might the person be feeling?
- Gaze direction, gaze duration, gaze direction changes – all convey different messages during interaction.
  - Look at interaction partner now-and-then while speaking or listening.
    - But, avoid staring for a long time.
  - Change gaze to guide the attention of listener(s) to an object of interest (*joint attention*).

Eye Movements

Movement Type	Description	Functionality	Application in Interaction
Fixations	Eyes stationary	Information acquisition, attention, cognitive processing	Reading, scene perception
Saccades	Rapid eye movements between fixations	Switch attention targets	Visual search
Smooth pursuit	Eyes fixed on a moving object	Follow a moving target	Tracking a moving object, gaze-based steering
Scanpath	A sequence of alternating short fixations and saccades	Scanning	Analyse behavior of agent
Gaze duration	Sum of all fixations (and proportion of time spent) in an area of interest before the eye leaves that area	Convey intent, cognitive processing	Level of engagement or interest, object of interest
Blinks	Rapid closing and opening of eyelids	Indicate behavioral states	Indicate liveliness, stress, fatigue, etc.
Pupil diameter	Size of the pupil	Convey cognitive effort	Cognitive load, fatigue

Based on (Kar and Corcoran, 2017)

Gaze Control in Robots



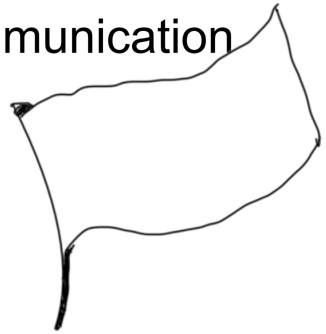
https://youtu.be/n6qTkw5U7YI

Which parts (i.e. joints) of iCub were used to indicate its gaze?

- Eyeballs
- Head / neck
- Body

- Gaze shift requires trajectory planning involving several joints.

- **Central question:** How to plan the trajectory such that it appears naturalistic and suits the interaction and communication context?

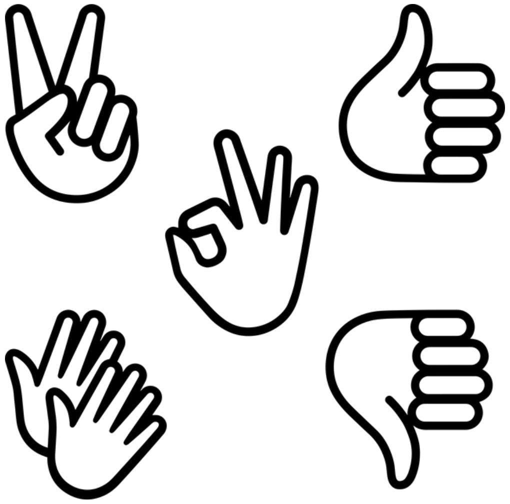


A²S

Hand Gestures: Recognition and Generation

Gestures: Types

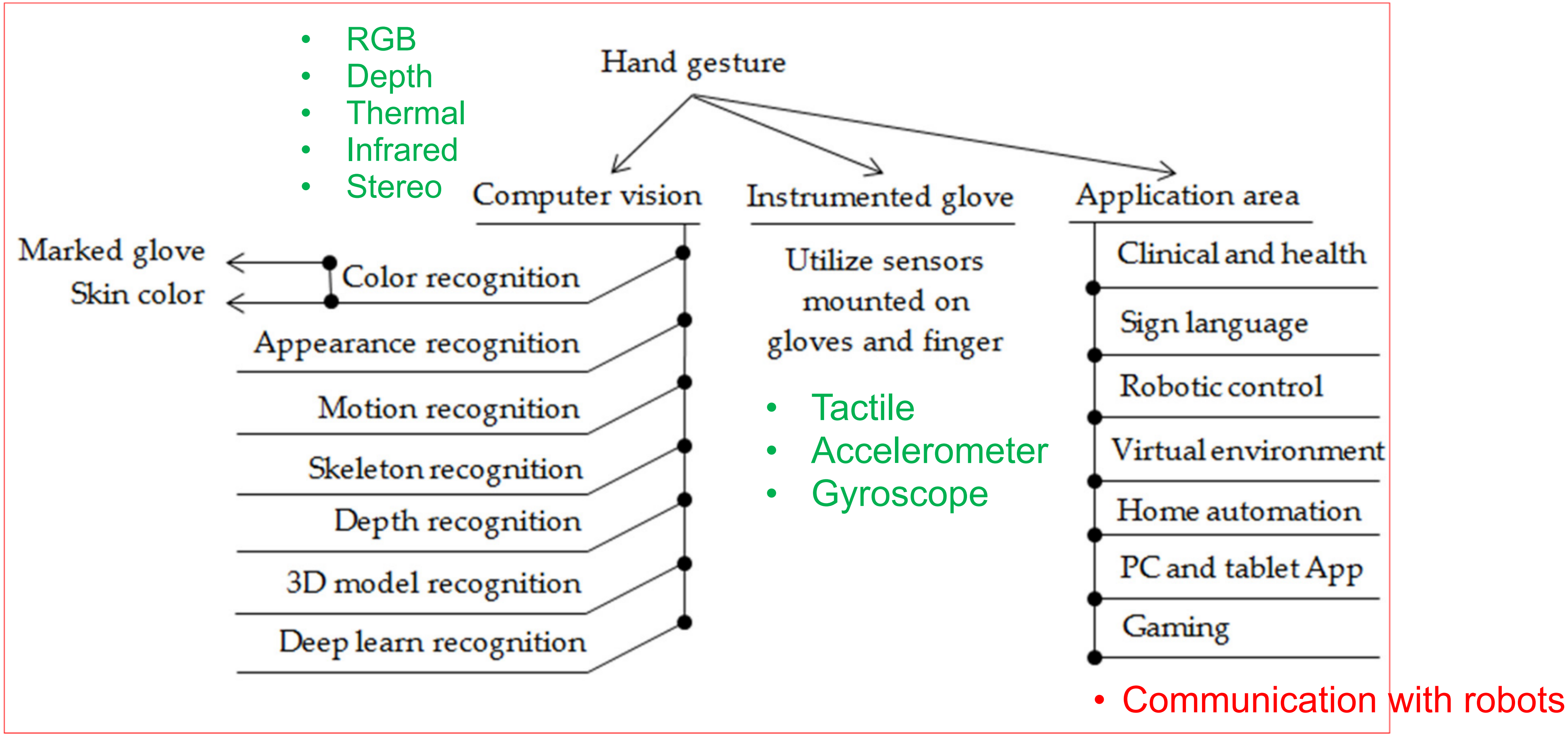
- Adaptors
  - Gestures that indicate mental states like anxiety, uneasiness, arousal, boredom, etc.
    - Fidgeting with fingers, clicking pens, shaking of legs, …
    - To release (excessive) energy.
- Emblems
  - Have a specific meaning that has been agreed-to by specific communities.
    - Thumbs-up, victory sign, claps, etc.
  - Culture-specific: Same sign can have different meanings in different cultures or societies --> Use carefully!
    - E.g. the V-sign in Britain
- Illustrators
  - Linked to speech and generated more subconsciously.



Source: https://open.lib.umn.edu/communication/chapter/4-2-types-of-nonverbal-communication/

A²S

Hand Gesture Recognition



Source: Figure 2 in (Oudah et al., 2020) extended with info from Figure 1.

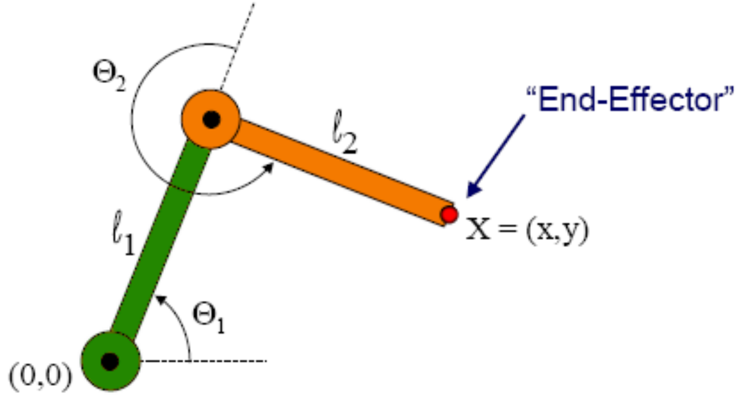
Gestures Accompanying Speech

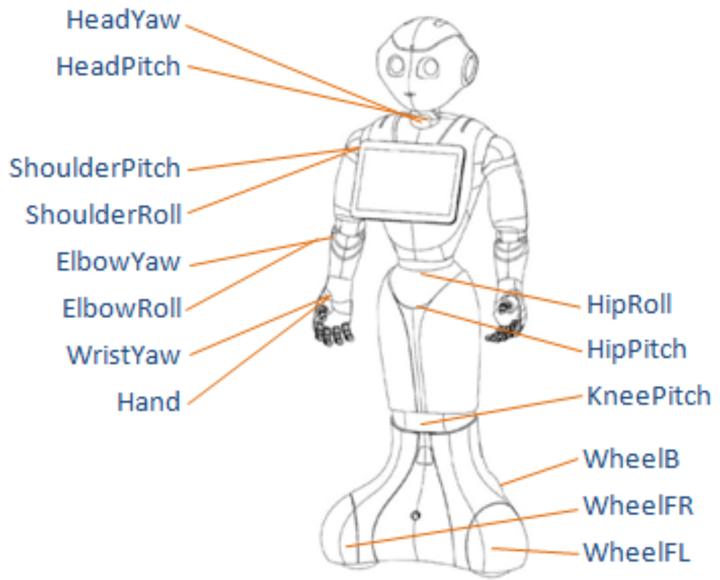
- "... Speakers gesture more when they talk about spatial topics than when they talk about abstract or verbal ones" (Alibali, 2005, p. 313).
- Bielefeld Speech and Gesture Alignment (SaGA) (Lücking et al., 2010)
  - Spatial communication tasks: give direction and describe a scene
  - Includes following gestures:
    - Indexing (pointing), placing, shaping, drawing, posturing, sizing, counting, hedging.
  - Morphology and movement:
    - Handshape: similar to American Sign Language
    - Palm (back of palm) orientation: Up, down, left, right, forward, backward
    - Wrist position and distance: Relative to the body of the gesturing person
    - Movement direction: Combinations and sequences of six cardinal directions
    - Movement trajectory: linear, curved

A²S

Gesture Generation in Robots

- Trajectories in joint space:
  - A sequence of waypoints describing (position, velocity and acceleration) for each (relevant) joint.
  - In joint space, positions are angles, and velocities and accelerations are time derivates of it.
- Could also be specified in Cartesian space:
  - Use inverse kinematics to compute trajectory in joint space.
- Pre-defined trajectories are tuned through trial and error & stored in a gesture library.
  - Adapted on-the-fly e.g. tempo.



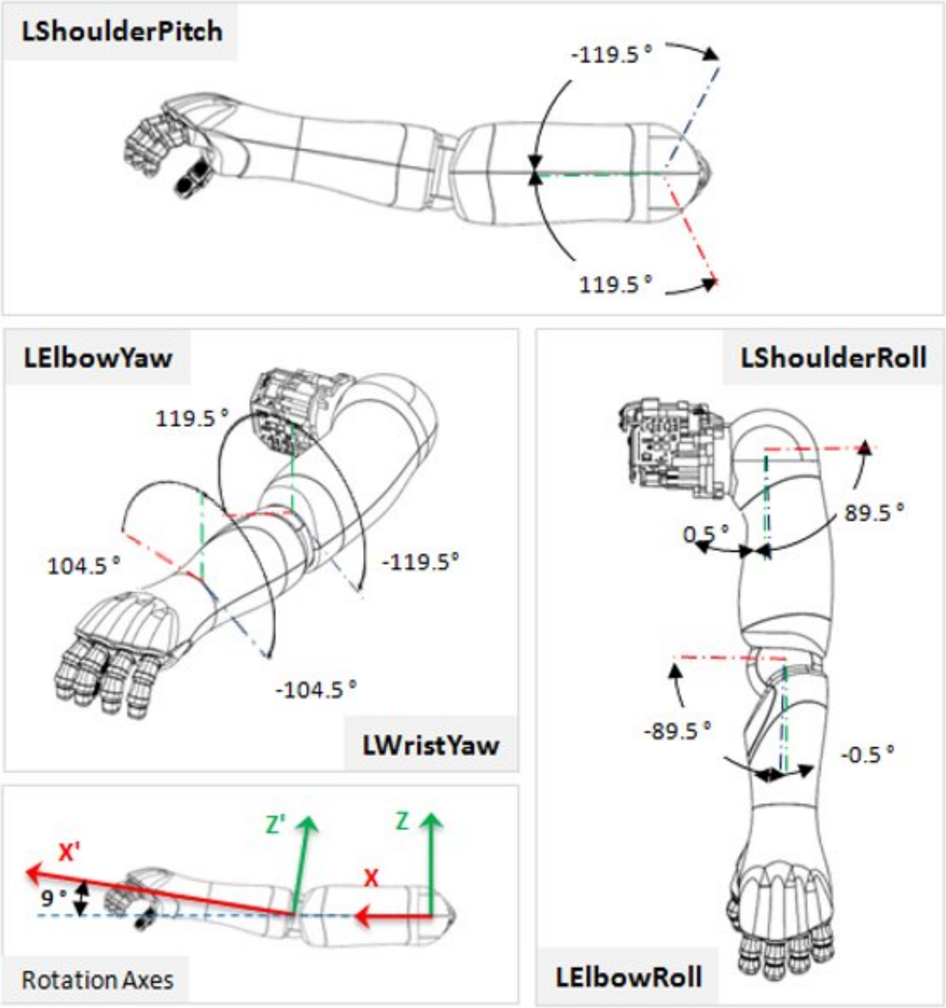


http://doc.aldebaran.com/2-4/family/pepper_technical/motors_pep.html

Gesture Generation in Robots

- Nowadays, machine learning methods are used to learn trajectories from human demonstrations.
  - Annotated videos
  - Human movement data

- But, direct 1-1 transfer to robots unlikely due to structural and kinematic differences.



http://doc.aldebaran.com/2-4/_images/joint_left_arm.png

A²S

Nonverbal Communication

- The nonverbal modalities that are available for communication depend on the physical embodiment.
  - For humans: Depends on the anatomical and muscular functioning.
  - For robots: What actuators and sensors are the robot built with?
- Expected nonverbal communication abilities depend on appearance and framing.
  - This affects how delays and errors in nonverbal behaviour influence human perception and acceptance of robots.
- How does it influence humans when robots show these nonverbal behaviors? (Saunderson and Nejat, 2019).
  1. Shift cognitive framing
  2. Elicit emotional responses
  3. Trigger specific behavioral responses
  4. Improve task performance

Recap Nonverbal Communication

- Nonverbal communication modes: Kinesics, proxemics, haptics, vocalics
- Nonverbal signals:
  - Facial expressions: Signs, messages
  - Eye movements: Fixations, gaze, saccades, smooth pursuits, scanpaths, blinks, pupil size
  - Gestures: Adaptors, emblems, illustrators
- Recognition of nonverbal signals
  - Mainly based on machine learning and mathematical models applied to sensor data
- Generation of nonverbal signals
  - Predefined trajectories or animations
  - Recently, machine learning models to predict robot poses and trajectories
  - Embodiment-specific adaptations essential

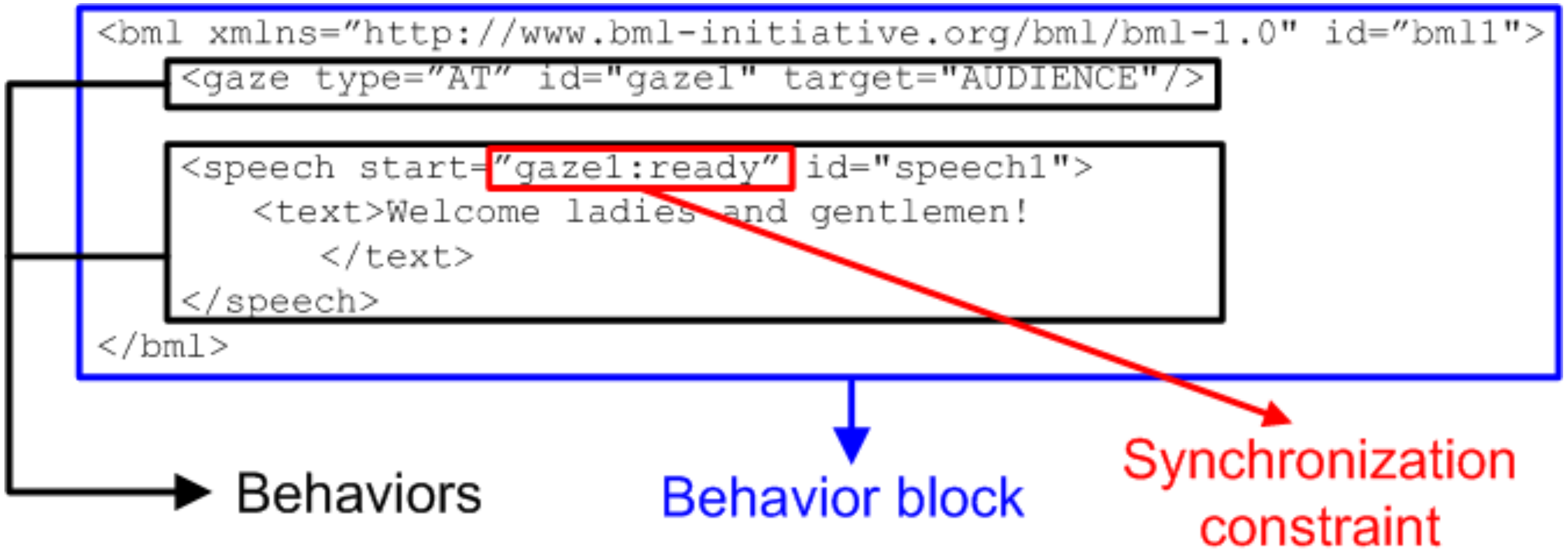
Multimodal Nonverbal Behaviour

Temporal Alignment

- Signals coming through different channels should be aligned in order to interpret its meaning correctly.
  - Speech
  - Hand gestures
- Alignment
  - Start
  - End
  - Duration
  - Rhythm
- Multimodal behaviour recognition
  - Communicative role of each modality
  - Use multiple information fusion strategies, different timescales
- Fluent generation of multimodal behaviour
  - What are needed?
    - Feedback on execution status
    - Predictions about execution
    - Behaviour planning
    - Status monitoring and scheduling
    - Incremental update
    - Failure handling

Behavior Markup Language (BML) (Vilhjálmsson et al., 2007)

- An XML description language for specifying multimodal (verbal and nonverbal) behaviours that should be expressed by an embodied conversational agent (a virtual agent or a physical robot).
- A BML Realizer understands BML requests and realises these behaviours on the agent.

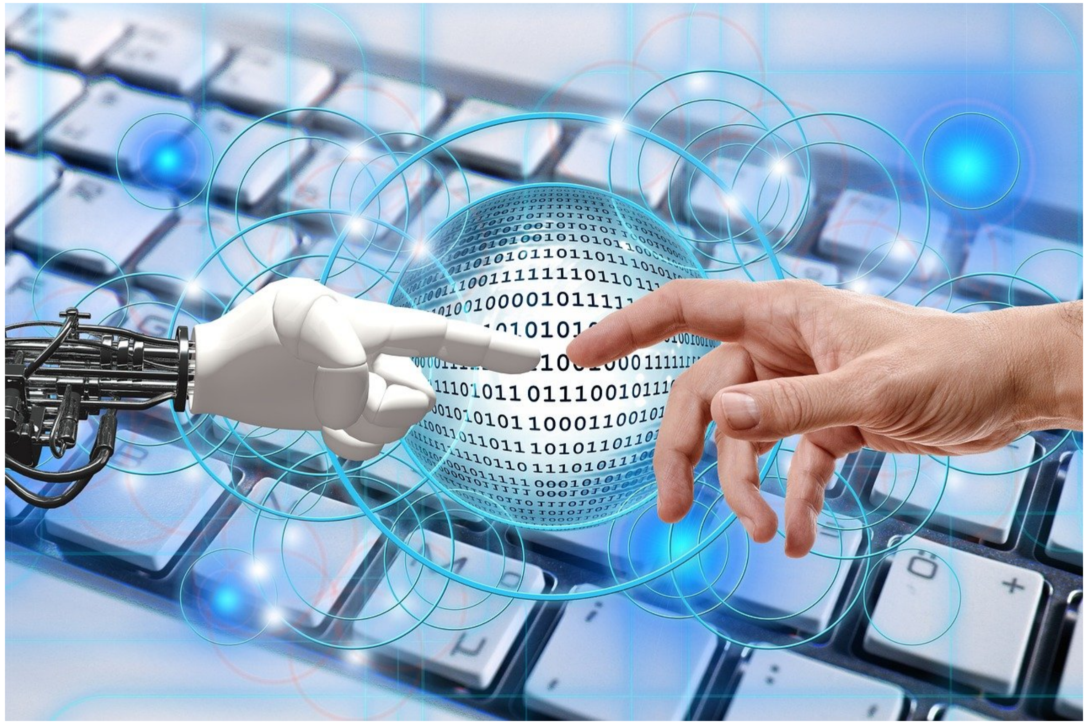


Example of a BML Request [Source: Figure 1 from https://projects.cs.ru.is/projects/behavior-markup-language/wiki]

Conclusion

- In today's lecture, you learnt to:
  1. List the communication modalities in humans and robots.
     - Verbal and nonverbal modalities
  2. Identify the key elements and concepts of nonverbal communication.
     - Kinesics, proxemics, haptics, vocalics
  3. Elucidate the technical aspects of realising multimodal communication between humans and robots.
     - Recognition and generation of nonverbal behaviours
       - Facial expressions, eye gaze, hand gestures
     - Synchronising verbal and nonverbal behaviours using BML

Lecture 04 - Verbal Communication in Humans and Robots



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

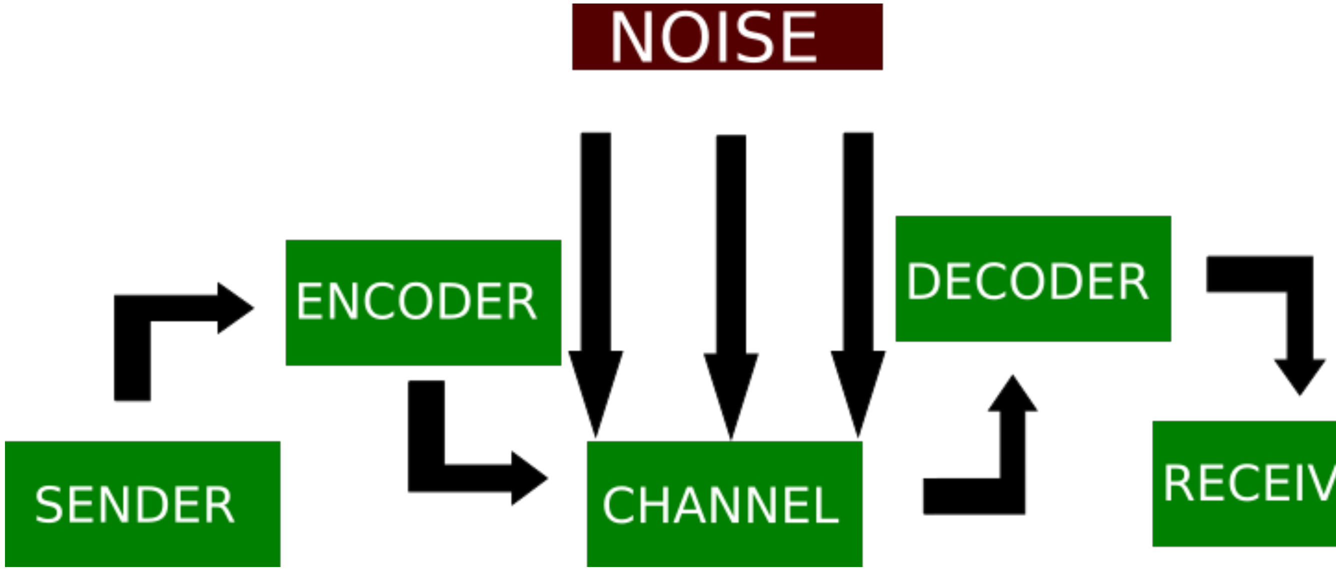
Department of Computer Science
Hochschule Bonn-Rhein-Sieg
Sankt Augustin

2nd May 2024

A²S

Model of Communication – Mathematical

- Transport the message from source to destination
- Information **sender** – source that produces the message.
- Information **receiver** – destination for whom/which the message is intended.
- **Channel** – medium that carries the message.
- **Encoder** – converts the message into a signal that can be transmitted through the channel.
- Channel introduces **noise** in the transmitted signal.
- **Decoder** – converts the received noisy signal back into the message.



Shannon, C.E., Weaver, W.: A Mathematical Model of Communication. University of Illinois Press, Urbana, IL (1949)

Models of Communication -- Social

- The linear model is not enough.
- **Context** of communication is important.
- Communication is **multimodal**.
  - Not only: What has been said?
  - But also: How was it said?
- **Feedback loops** between senders and receivers:
  - Communication is a cyclic process.
  - Reciprocity is an important part of social norms and polite behavior.
    - A greeting is expected to be returned; a question is expected to be answered, etc.
- Sender and receiver **co-construct** the meaning of communication.
  - Coding and decoding are dependent on each other.

We "cannot not communicate."
-- (Watzlawick et al., 1967)

Models of Communication – Ethno-Socio-Linguistic

- **Communication context**
  - Temporal (when)
  - Spatial (where)
  - Socio-cultural (with whom)
  - Others
    - Objects
    - Audience
- Same message has different meanings in different contexts.

Intentional or not, we are **continuously communicating**.

- Others are perceiving and interpreting us, even if we do not intentionally express signs.

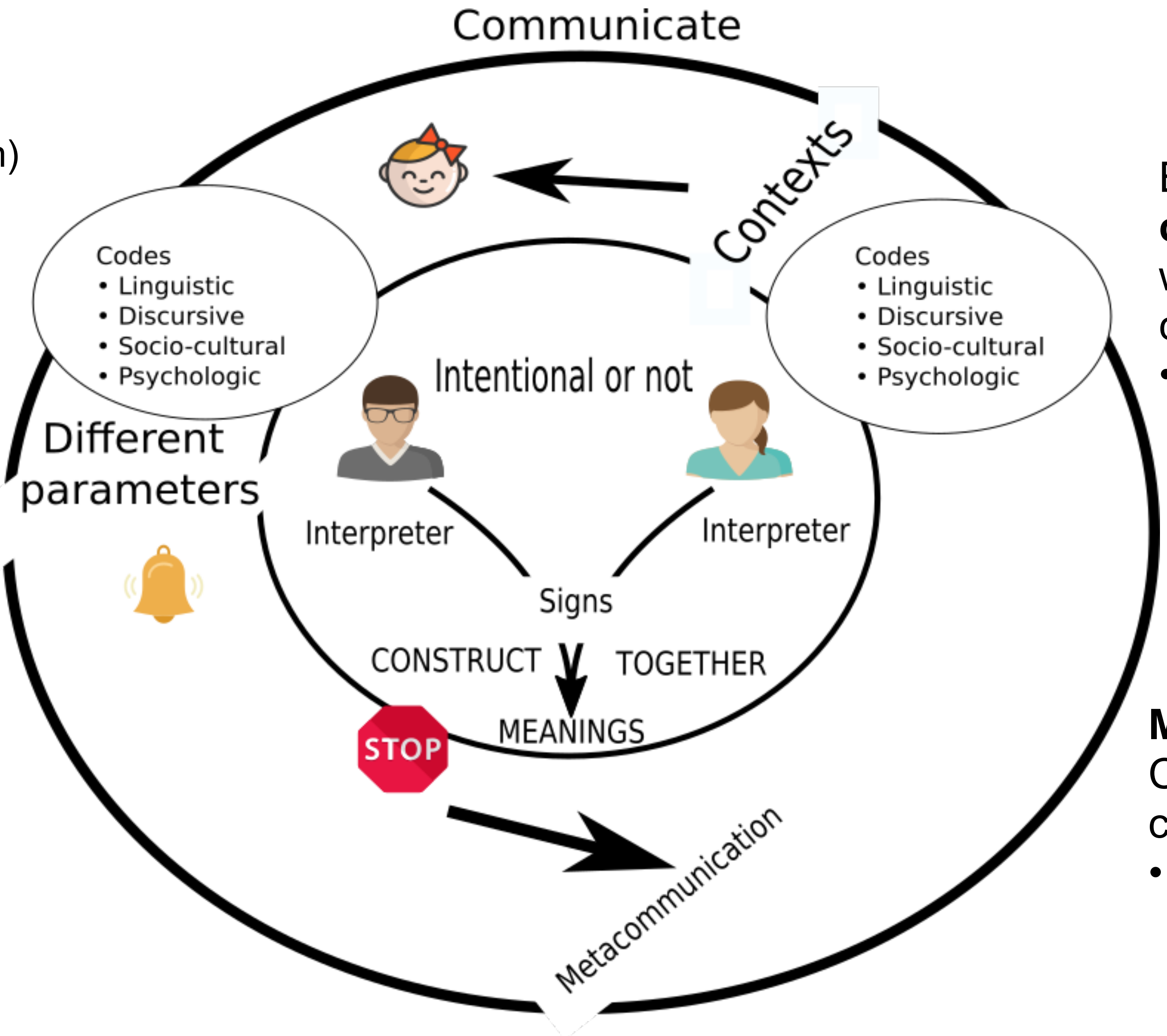


Image recreated from (Grandgeorge, 2020) and (Blanchet, 2000)

Each individual uses **codes** to express as well as interpret communicative signs.

- Our internal representation (model) of the world is key here.

Metacommunication: Communication about communication.

- e.g. clarification requests

Models of Communication – Ethno-Socio-Linguistic

- **Can be applied to human-robot communication.**

Robots should be aware that the human is constantly interpreting the signs sent or not sent by the robot and this could influence the subsequent interaction with the robot.

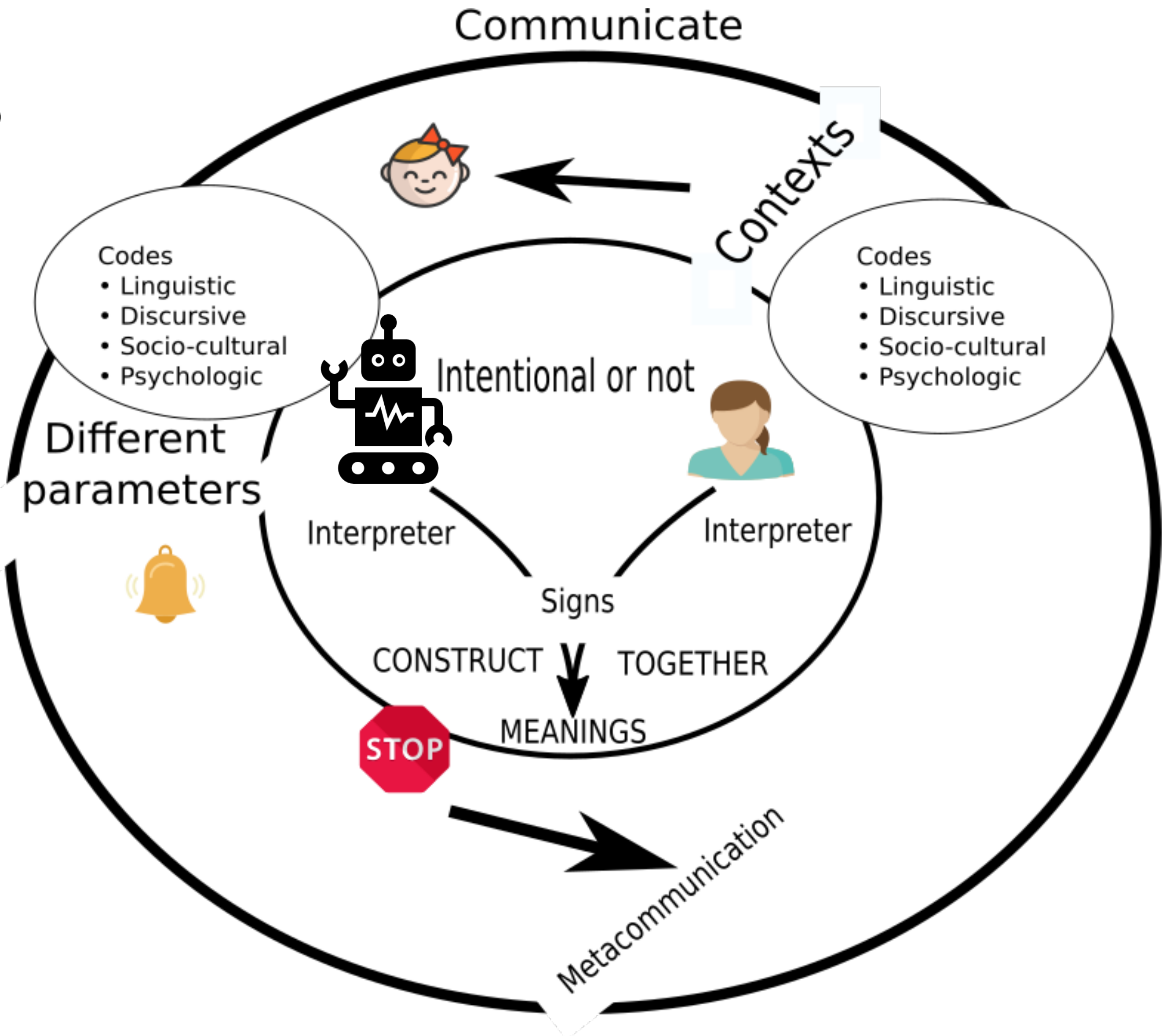


Image recreated from (Grandgeorge, 2020) and (Blanchet, 2000)

Verbal Communication

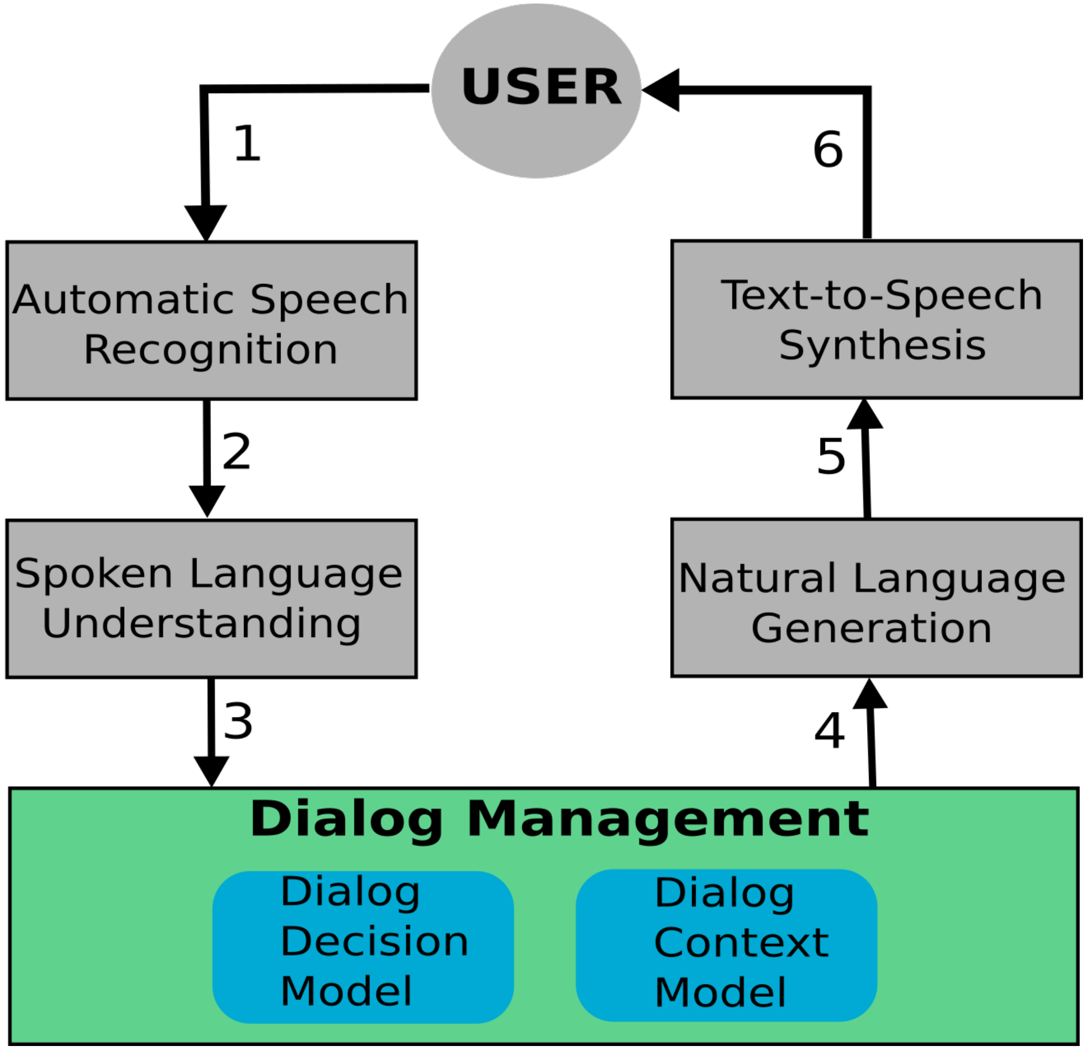
- **Verbal communication** *"is the ability of humans to express their thoughts and to communicate with each other through a system of vocal and/or graphic signs constituting a language."* -- CNRS, (Grandgeorge, 2020)
- **Language** is *"a system of communication by speaking, writing, or making signs in a way that can be understood, that allows the mention of (not only) present but also past and future situations."* -- (Grandgeorge, 2020)
  - A system which consists of a set of symbols (sentences) — realised phonetically by sounds — which are used in a regular order to convey a certain meaning. https://www.uni-due.de/ELE/LinguisticGlossary.html

- Spoken language:
  - Input: Auditory channel
  - Output: Vocal channel
  - Basic unit: Phonemes

- Sign language:
  - Input-Output: Visual-bodily channels
  - Formational units:
    - hand shapes
    - locations
    - movements
    - orientations
    - non-manual

Spoken Dialog System Architecture (Cañas et al., 2021)

- The processing pipeline and components needed for realizing spoken dialog systems.
  - Here, we are looking only at the verbal part of communication.
  - If the dialog system is used together with an embodiment (virtual or physical), then nonverbal communication is inevitable.
    - A full-fledged multimodal interaction architecture would be needed.



Recreation of Fig. 1 in (Cañas et al., 2021)

Spoken Dialog System Architecture (Cañas et al., 2021)

1. User makes a query, a command, etc. in the form of an **acoustic signal**.
2. **Automatic Speech Recognition (ASR)** analyses the acoustic signal to recognise the spoken words (along with an estimate of the agent's confidence in the recognised words).
3. **Spoken (Natural) Language Understanding (NLU / SLU)** analyses the recognized words and extracts the intents (instruct, query, reply, greet, etc.) and entities (name, place, time, thing, etc.) relevant for the application domain.
   - "Hello Robbie!" --> "Hello __name__" --> The user is greeting and Robbie is a name.
4. **Dialog Management (DM)** consists of two models:
   1. **Dialog Context Model (DCM)**
      - Tracks the dialog state and is updated by DM with information provided by NLU.
   2. **Dialog Policy Model (DPM)**
      - Decides the agent's next dialog act (intent, entities) given the information in DCM.
5. **Natural Language Generation (NLG)** converts the agent's dialog act into a sequence of words.
6. **Text to Speech (TTS)** converts the words into acoustic signals which are then heard by the user, and which could trigger another round of exchange.

A²S

Turn Taking in Verbal Communication

- Who controls the conversational floor?
  - That is, who would speak at a specific point in time. Consequently, others listen.

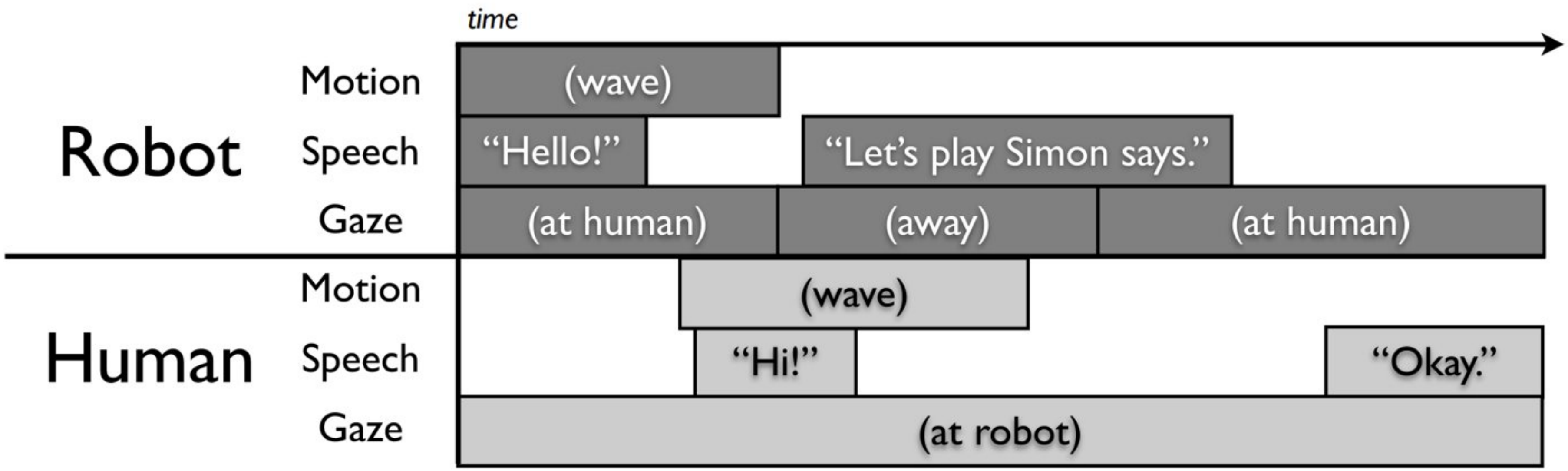


Figure 1. An example of event alignment for a multimodal interaction.

Source: Figure 1 from (Chao and Thomaz, 2012)

A²S

Turn Taking in Verbal Communication

- (Chao and Thomaz 2013) proposed Timed Petri Nets (TPN) for dynamic turn-taking during one-on-one interactions between a human and a robot.
  - CADENCE Architecture
  - Four parts of floor regulation during conversations:
    - seize the floor
    - yield the floor
    - hold the floor
    - audit the owner
  - **Backchanneling** from listener to speaker to confirm that the listener is paying attention to the speaker, is in agreement with the speaker, etc.
    - Through verbal (yeah, mhm, uh-huh, etc.) or nonverbal cues (head nodes, facial expressions, etc.).
    - Speaker holds the floor.
  - Backchanneling could also be used by listener to (eventually) seize the floor from the speaker.
    - E.g. "Yeah, but"
- Part of Dialog Management component of spoken dialog systems.

Turn Taking
BackChanneling
~~Meta Comm.~~
Symbol Grounding

A²S

Symbol Grounding

- At a dining table, PersonA says to PersonB: *"Could you please pass me **the salt**?"*
  - PersonB would know that PersonA is referring to **the salt on the table** next to PersonB.

- In **situated interaction**, an agent should have awareness about **spatial and temporal relations** from own perspective and other's perspective, and...
- … should be able to **connect symbols with** their **own sensorimotor flow** (i.e. with what they are currently sensing and doing).

See (Doncieux et al., 2022)

Symbol Grounding in Robots

- Some of the symbols (words) in the vocabulary will be linked to the physical environment.
- The robot should be able to distinguish whether the symbols referred to by the human (or the other agent) are abstract concepts or actually things in its current environment.
- It should then identify that thing in its environment and link the symbol to it *(symbol grounding)*.
  - Internal representation of the physical world.
    - Spatial and temporal relations.
    - Updated based on sensorimotor flow.
  - Perspective-taking

A²S

See (Doncieux et al., 2022)

Conclusion

- In today's lecture, you learnt to:
  1. Describe models for communication between humans.
     - Blanchet's ethno-sociolinguistic model for human communicaiton
  2. Identify the key elements and concepts of verbal communication.
     - Verbal: Spoken language, sign language, turn taking, symbol grounding
  3. Elucidate the technical aspects of realizing verbal communication between humans and robots.
     - Architecture of spoken dialog systems

Metacommunication (Grandgeorge, 2020)

- Each interpreter interprets the signs based on their own codes.
- When the received responses diverge from the intended outcome, then it indicates a discrepancy in the ongoing communication.
- The communication could be interrupted by at least one of the interpreters and they could now discuss about the communication itself.
- This is metacommunication.
- After the resolution of the misunderstandings, the original exchange can be resumed.
- Metacommunication involves high-level cooperation between the agents, where the interpreters are able to analyze and discuss about their own communication strategies.

A²S

Unintentional Communication

- "Meanings 'given off' (Goffman) by an individual's body language through nonverbal leakage, or unconsciously signified by their appearance, dress, or behaviour, including whatever may be noticeable by its absence in a particular context (*see also* analogic communication). As Watzlawick observed, 'one cannot *not* communicate', regardless of whether an observer's inferences are warranted."
  - Oxford Reference (https://www.oxfordreference.com/view/10.1093/oi/authority.20110803110707314)

A²S

Ethnography and Sociolinguistics

- **ethnography of communication** “The study of cultural differences in acts of communication. This is a comprehensive term which goes beyond simple differences in language to cover additional aspects such as formulaic use of language (e.g. in greeting or parting rituals), proxemics (the use of distance between partners in a conversation) and kinesics (the study of body movements used in communication).”

- **sociolinguistics** “The study of the use of language in society. Although some writers on language had recognised the importance of social factors in linguistic behaviour it was not until the 1960's with the seminal work of Labov that the attention of large numbers of linguists was focussed on language use in a social context. In particular the successful explanation of many instances of language change helped to establish sociolinguistics as an independent sub-discipline in linguistics and led to a great impetus for research in this area.”

- Source: https://www.uni-due.de/ELE/LinguisticGlossary.html#GlossF

Spoken Language

- **Phoneme:**
  - *“In traditional phonology the smallest unit in language which disinguishes meaning, e.g /k/ and /g/ as seen in coat and goat. Each phoneme has one or more realisations, called allophones.”*

- **Phonology:**
  - *"The study of the sound system of one or more languages. Phonology involves the classification of sounds and a description of the interrelationship of the elements on a systematic level.”*

Source: https://www.uni-due.de/ELE/LinguisticGlossary.html#GlossF

A²S

Speech and Language

- **Speech**
  - *“The production of sounds using the organs of speech; contrasts directly with writing which is a secondary medium for communication via language.”*

- **Prosody**
  - *“A term which refers to all the suprasegmental properties of language such as pitch, loudness, tempo and rhythm.”*
    - The nonverbal part of speech.

- **Suprasegmental**
  - *“A reference to phenomena which do not belong to the sound segments of language but which typically are spread over several segments, e.g. intonation, stress, tempo, etc.”*

Source: https://www.uni-due.de/ELE/LinguisticGlossary.html#GlossF

User Intent in Verbal Communication

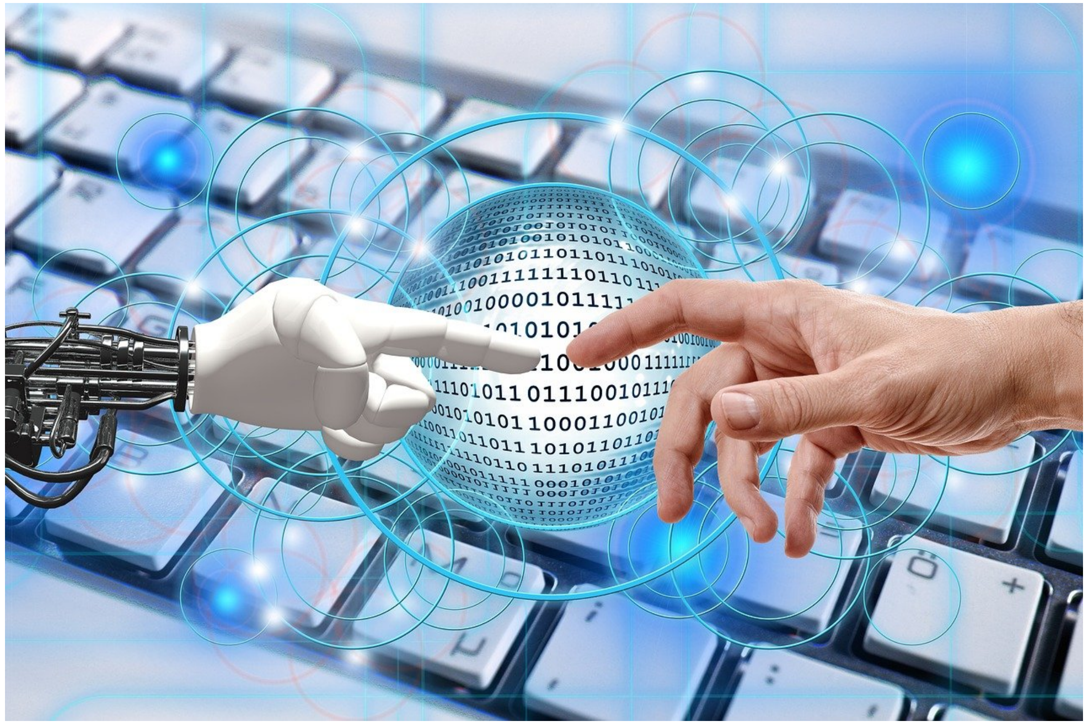
request	instruct query propose check	task, cancel, backtrack open, select, yn suggest, offer align, confirmation, repeat
assert	provide	elaboration, statement, opinion
respond	yes no reply notify	agree, accept disagree, reject open, select acknowledge, buy_time, success, failure
social	greetings politeness interpersonal	opening, closing apology, thanks, acknowledge_thanks feedback
other	other	other_intent, no_intent
add_on	add_on	goes_with, correct

Table 1: Dialog acts tags inventory.

Source: Table 1 in Silvia Pareti and Tatiana Lando. 2018. Dialog Intent Structure: A Hierarchical Schema of Linked Dialog Acts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

A²S

Building Socially Interactive Robots

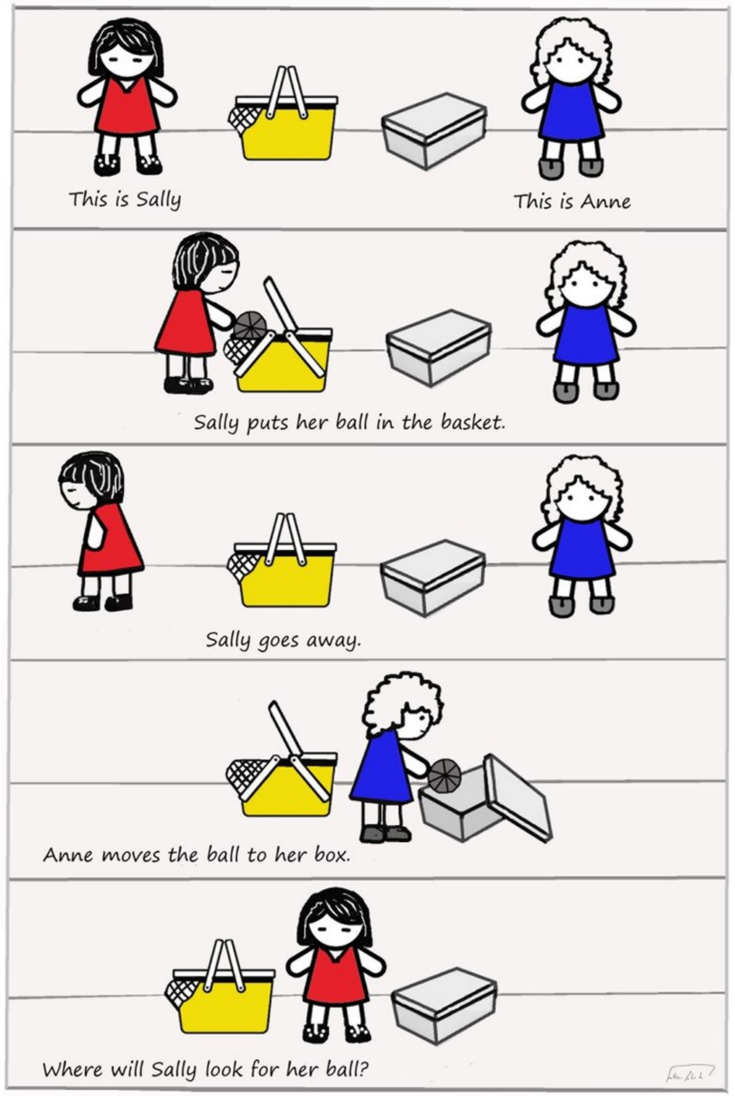


Prof. Dr. Teena Hassan
teena.hassan@h-brs.de

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

16th May 2024

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).

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]

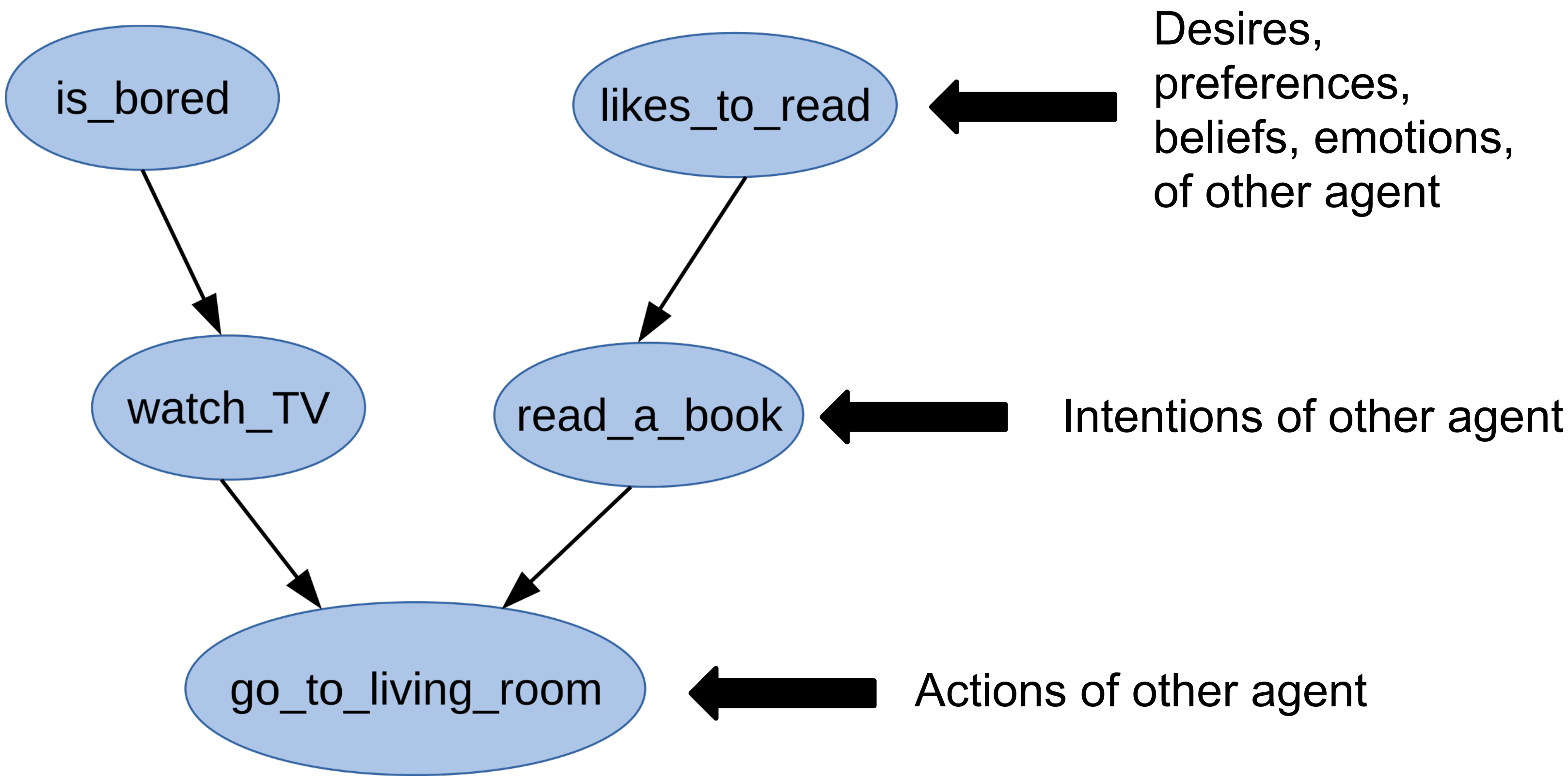
Why Should a Robot have ToM?

- 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.



A²S

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

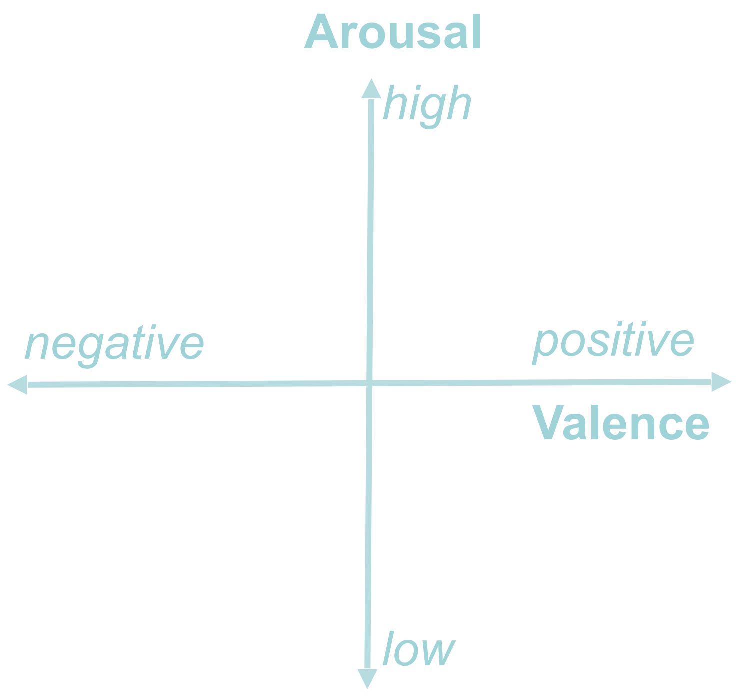
- Three psychological perspectives for studying and modeling emotions (Broekens, 2021):
  - Categorical view
  - Dimensional-constructionist view
  - Cognitive appraisal view

- 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.

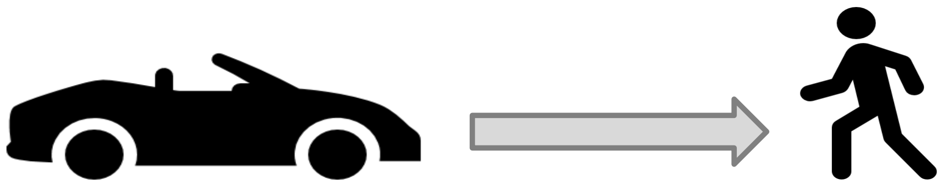


Reference: (Broekens, 2021)

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)

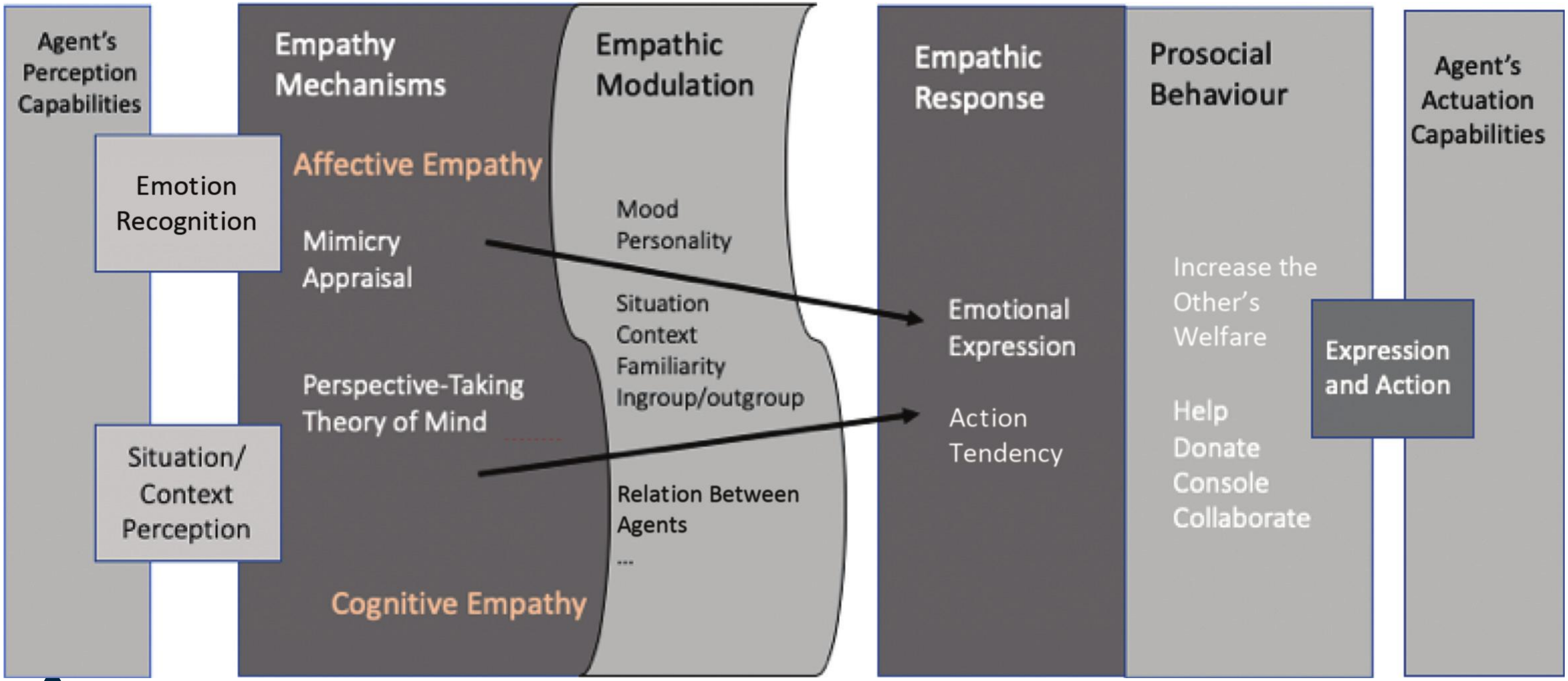
- **Empathy:** a psychological process leading a person to have "*feelings that are more congruent with another's situation than with their own situation.*" (Hoffman 2001)

- **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

Mechanisms for Empathy



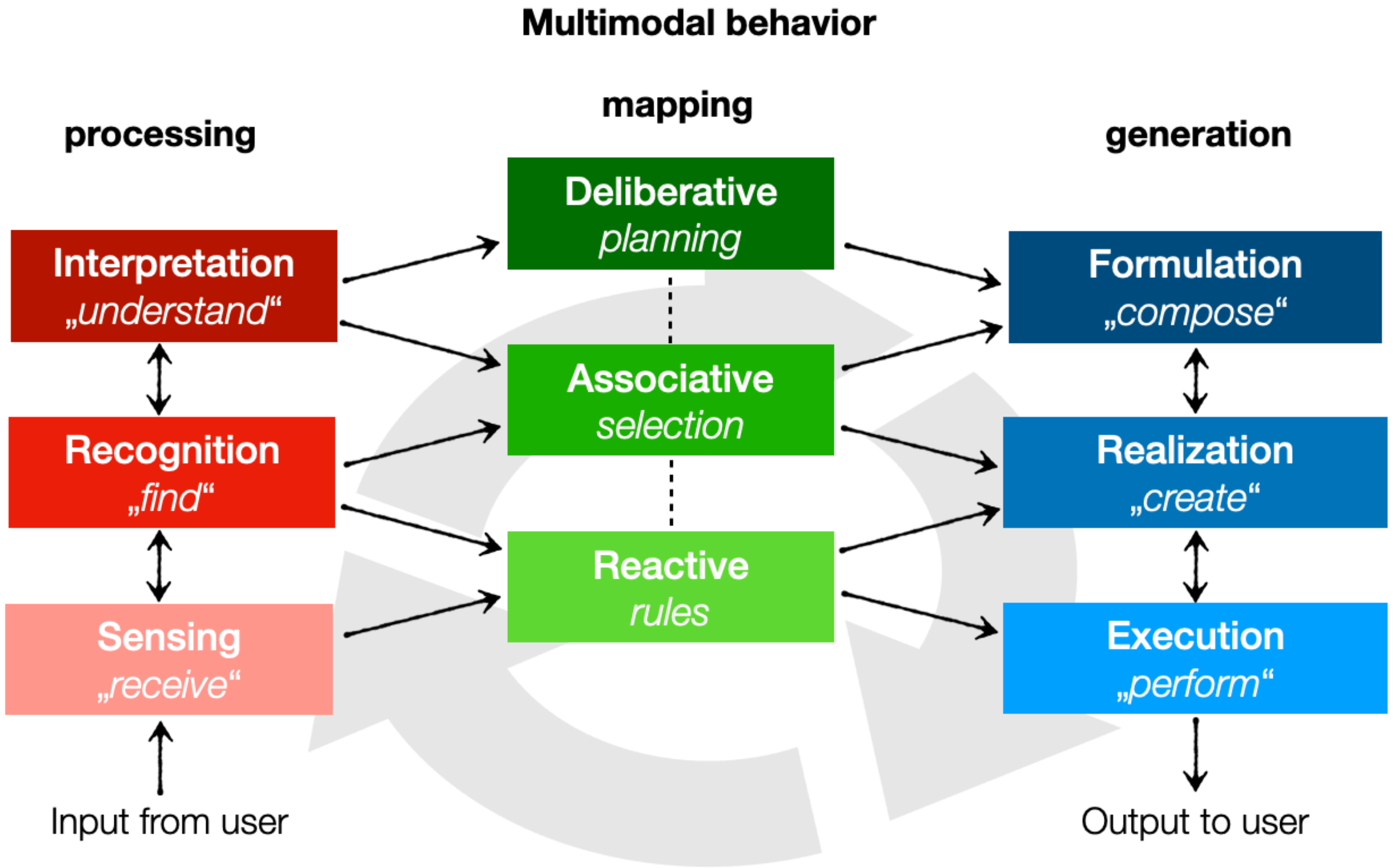
Source: Fig. 11.3 in (Paiva et al,., 2021)

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

A²S

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

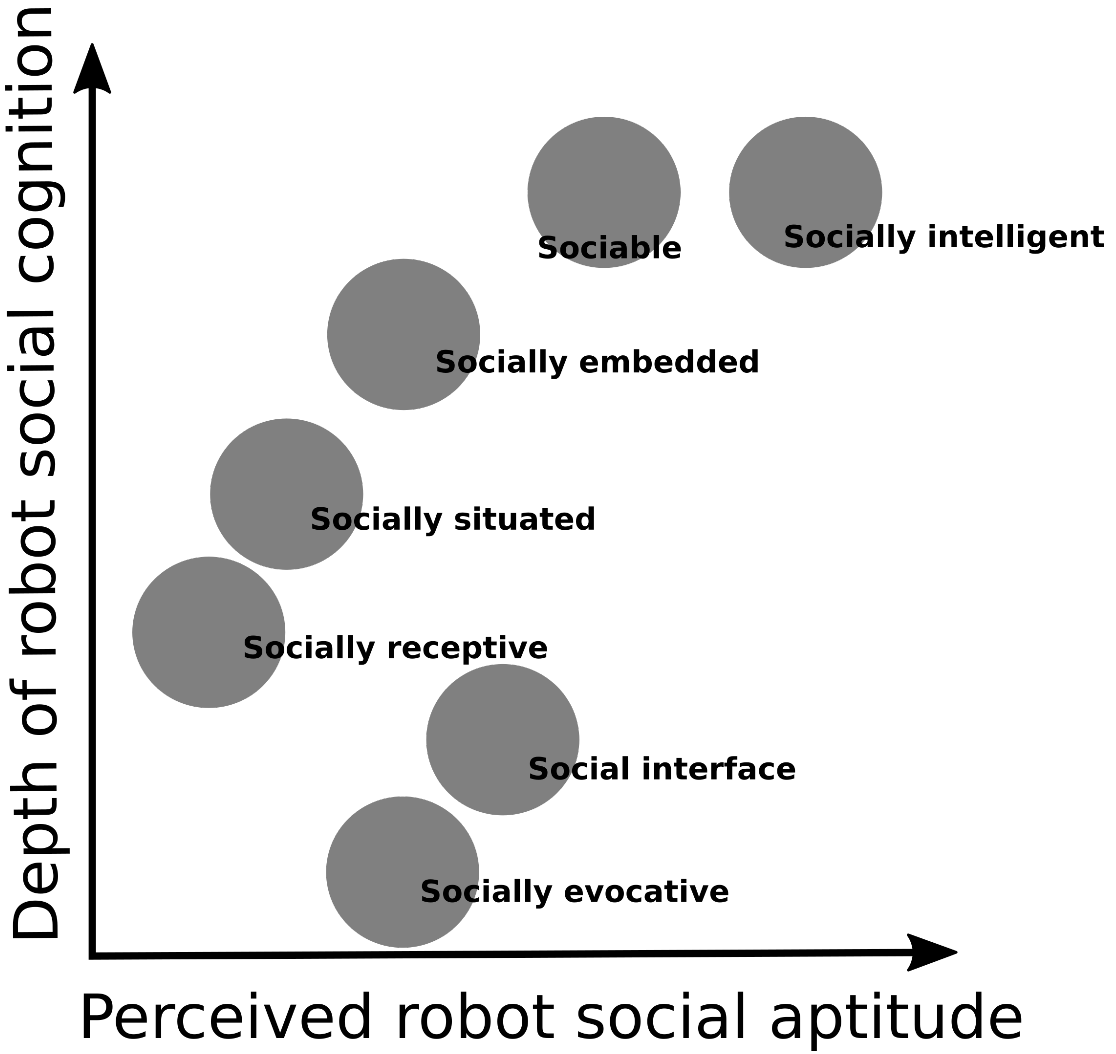
Different Processing Pathways (Kopp & Hassan, 2022)

- 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.

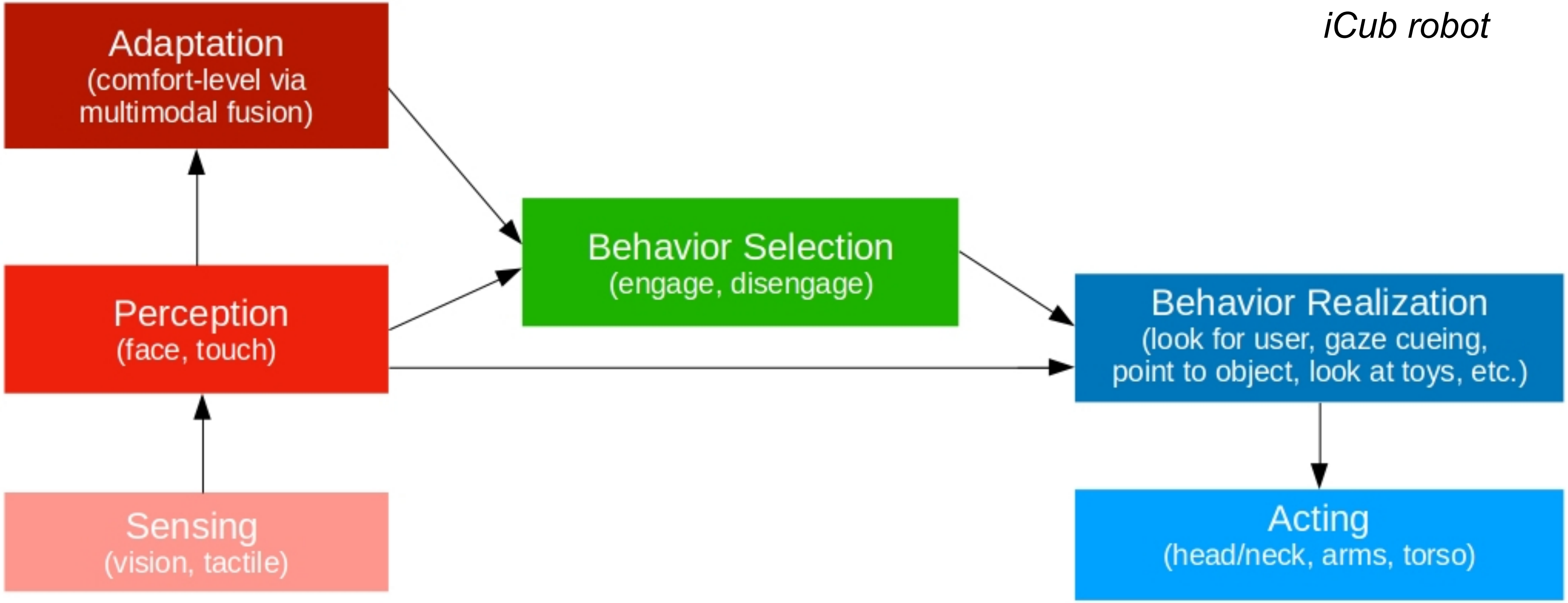
- Socially intelligent (Fong et al., 2002)
  - Social intelligence similar to humans
  - Deep social cognition and competence (e.g. learning from and adapting to humans, establishing social relationships).

A²S

Single Route Architecture – Example

(Kopp & Hassan, 2022)

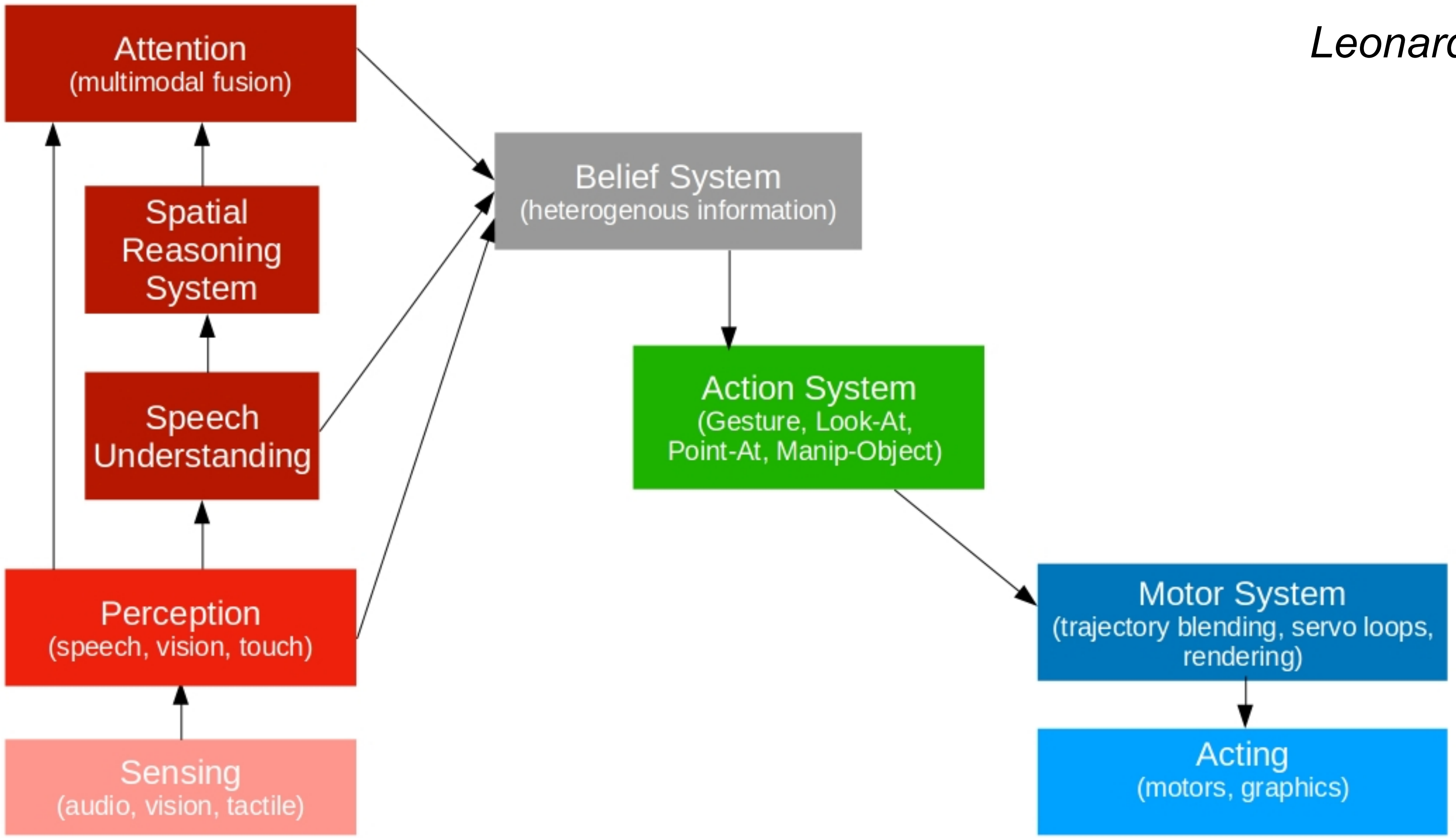
iCub robot



Based on (Tanevska et al., 2019)

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

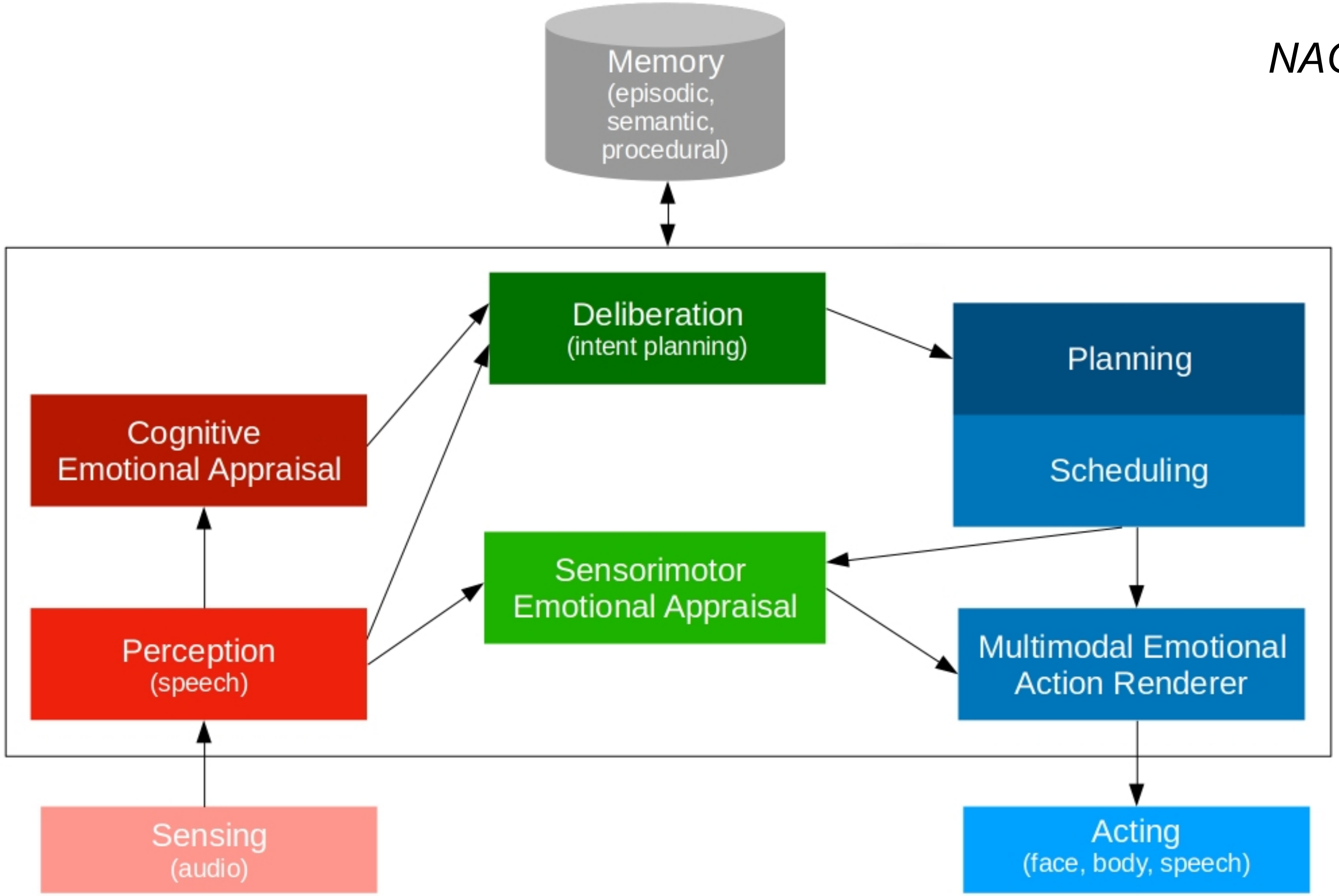
Leonardo robot



Based on (Breazeal et al., 2004)

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

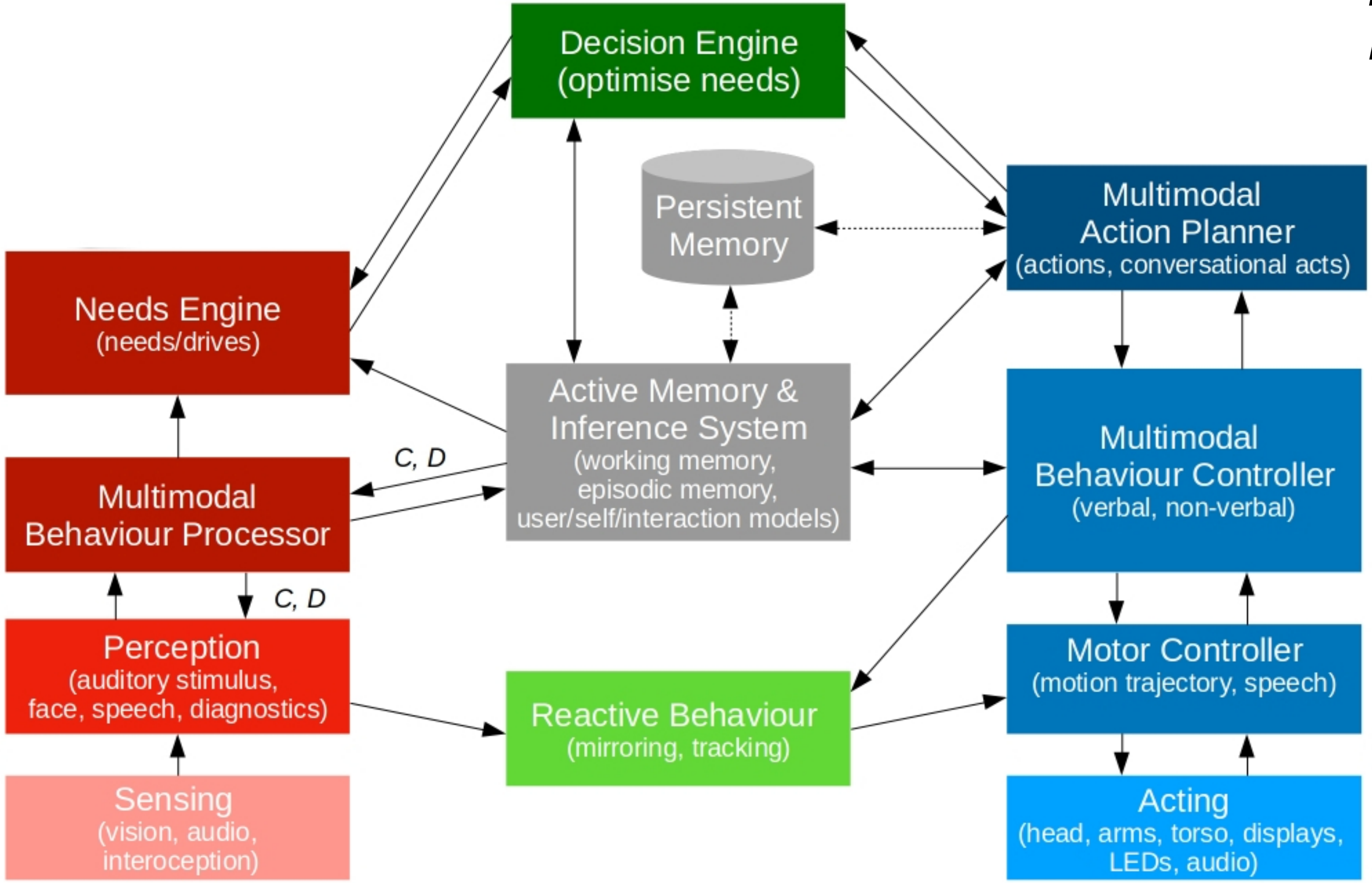
NAO robot



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

Multidirectional, Incremental Architecture

Pepper robot
Navel robot



C: Correction, D: Disambiguation

(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.

A²S

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.

Behaviour Explanation Dialog Framework

b-it

- 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).

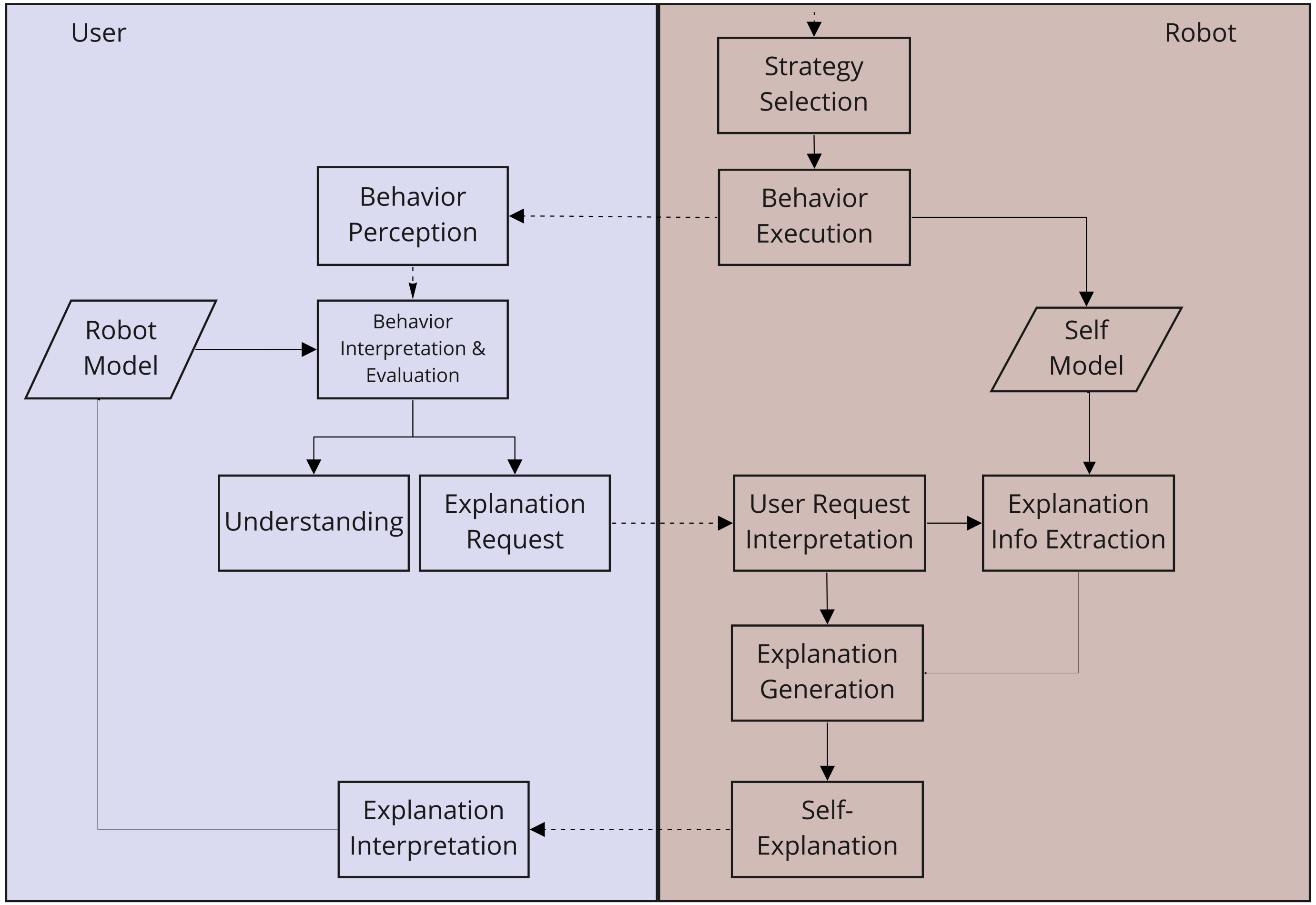


Image source: Fig. 2 in (Stange et al., 2022)

A²S

Verbal Explanations of Behaviour – An Example

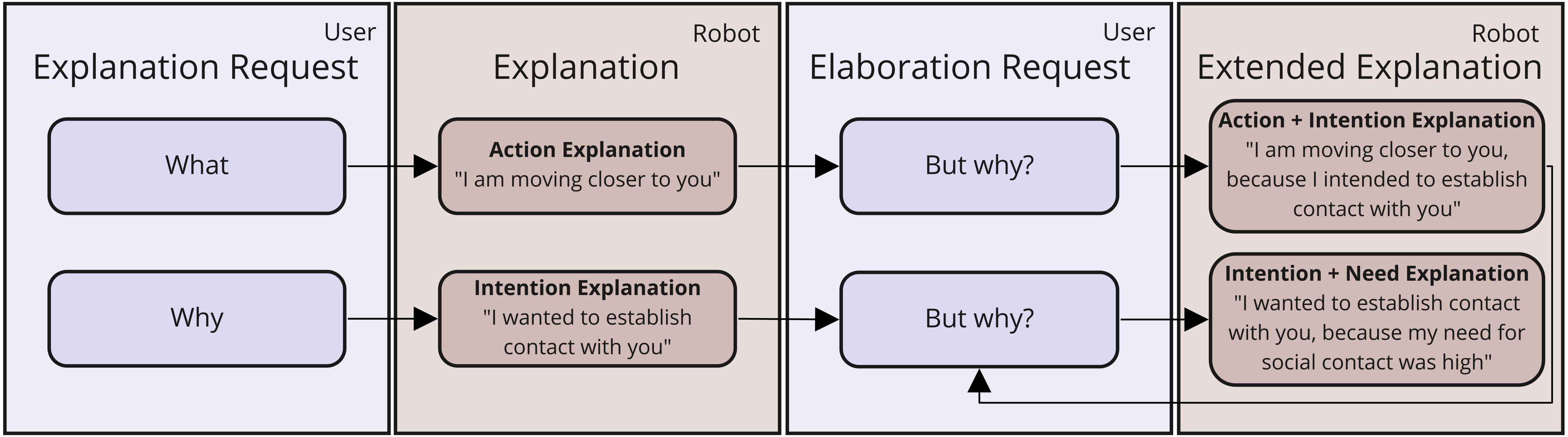


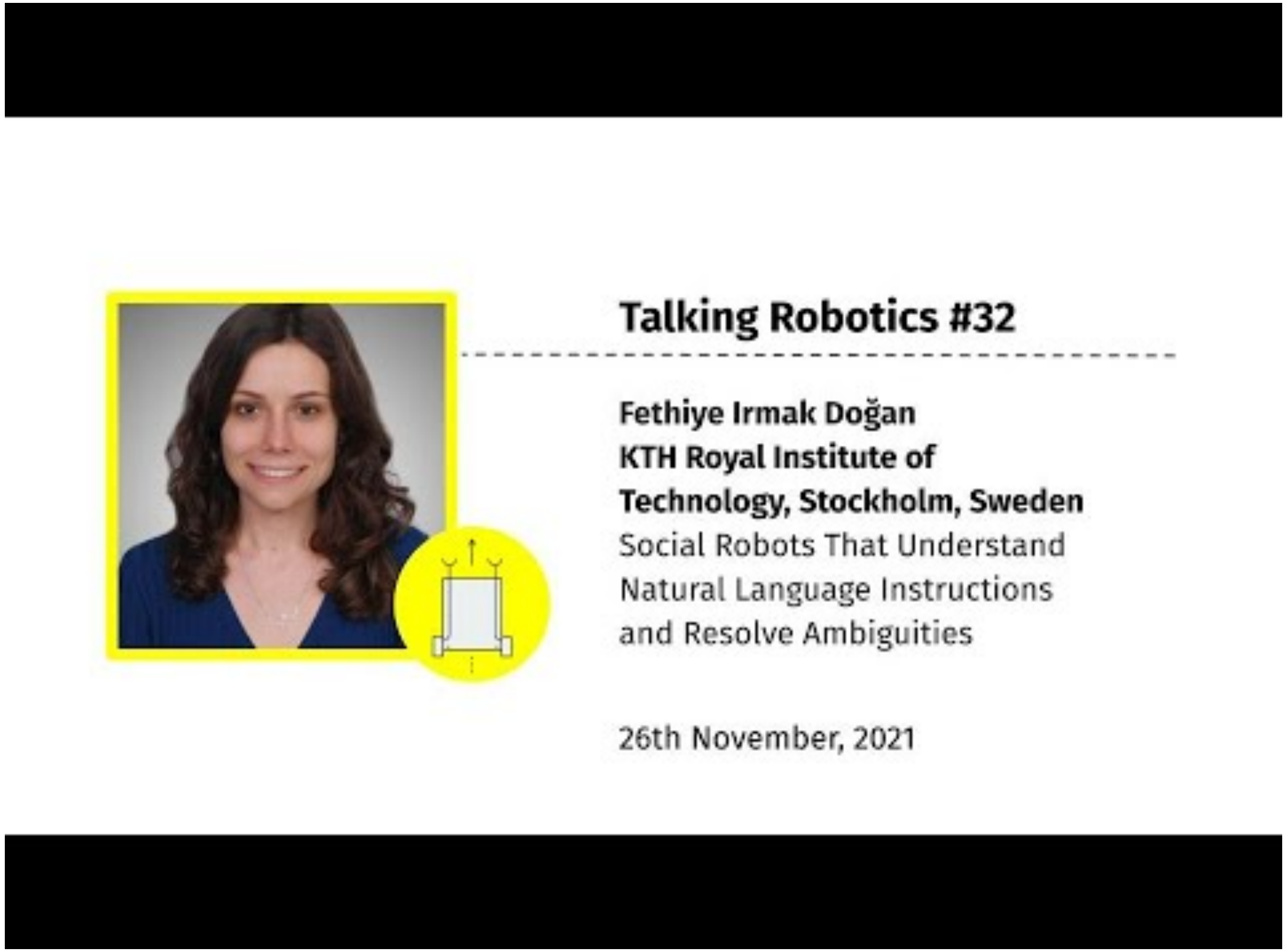
Image source: Fig. 3 in (Stange et al., 2022)

A²S

Helping Robots Understand Humans

- Metacommunication
  - Clarification requests

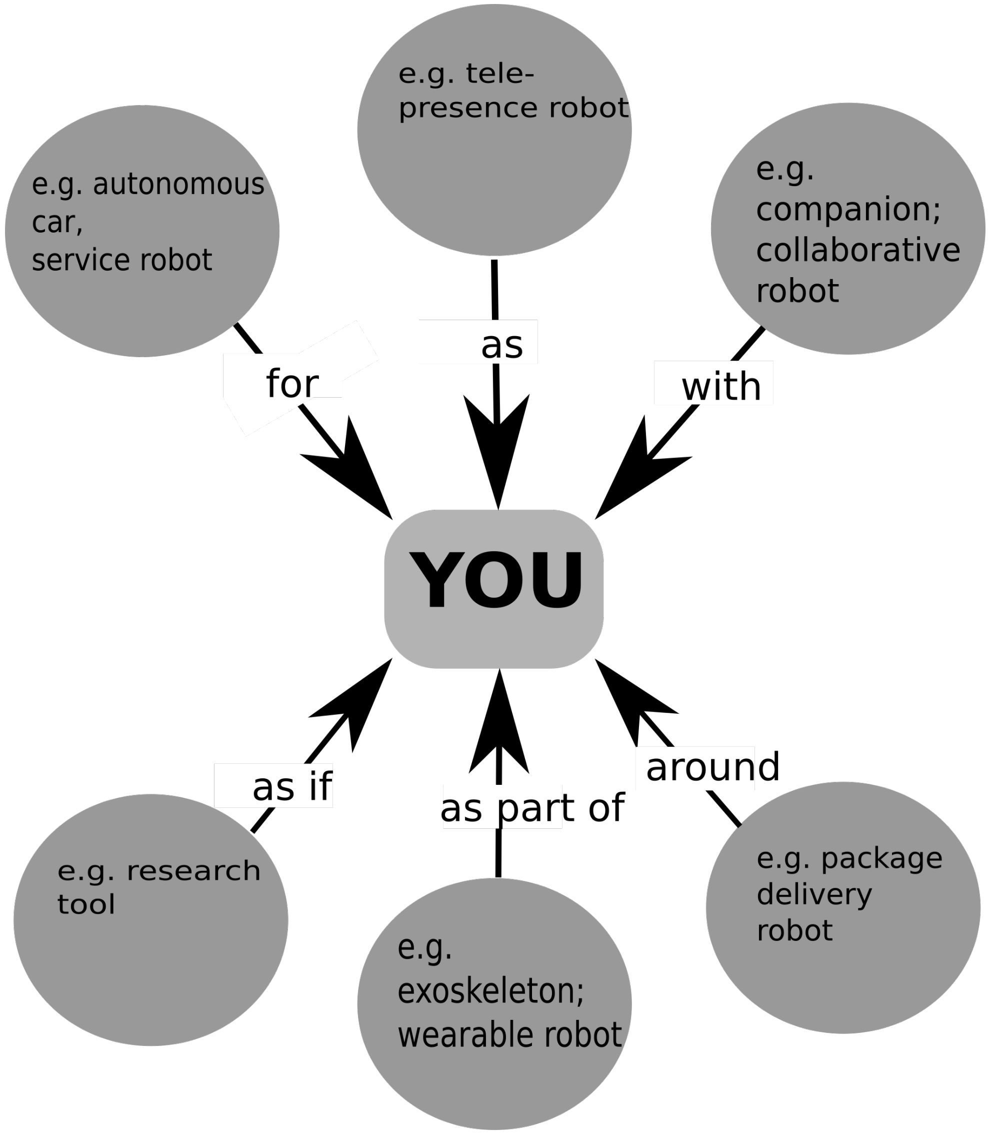
- E.g. Clarify ambiguities arising from user's referring expressions
  - Integrates:
    - Verbal communication
    - Visual-spatial perception
      - Scene understanding
    - Perspective taking
    - Symbol grounding



https://youtu.be/I21HicOzdI8

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)

Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bonn-Aachen
International Center for
Information Technology

Interactive Reinforcement Learning

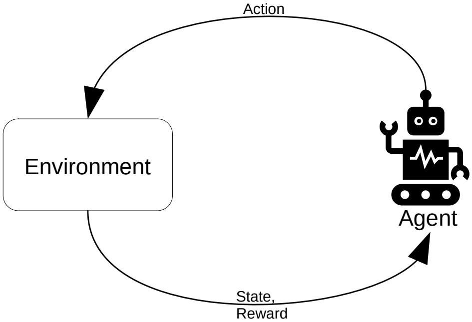
Guest Lecture

June 13th, 2024

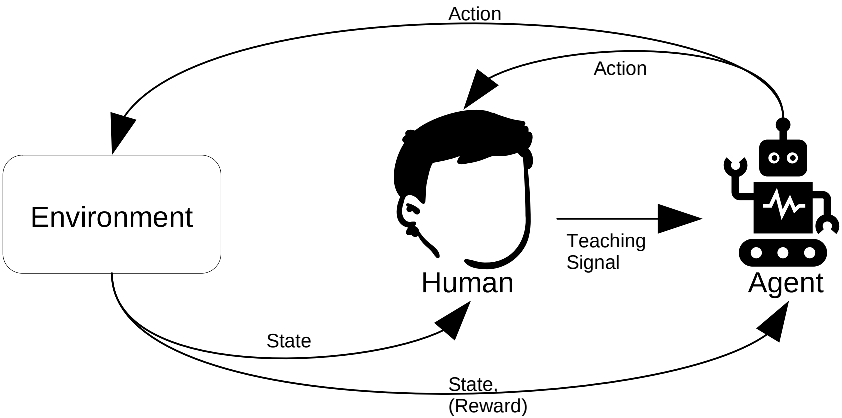
Michał Stolarz
Prof. Dr. Teena Chakkalayil Hassan

Reinforcement Learning (RL)

- RL allows an agent to learn an optimal policy that maximises the **cumulative reward** by interacting with an environment.
- In real-world tasks, the agent is faced with the **sample efficiency problem**, making the learning slow.



Interactive Reinforcement Learning (IRL)



Advantages of IRL

- Humans possess **knowledge** about the environment and have **experience** in acting in that environment.
- Human input could be used to guide and **accelerate the robot's learning**, or to change its optimal behaviour (personalisation).

1. What is Interactive Reinforcement Learning?
2. Teaching signals used in IRL
3. Human evaluative feedback
4. Learning from evaluative feedback
5. TAMER
6. SABL
7. COACH
8. Combining evaluative feedback and environmental reward
9. Real-life application
10. Summary

Teaching Signals

Feedback

✓ Human feedback is provided with the intention of **evaluating** the robot's action.

✓ The value of the **evaluative feedback** depends on the last action performed by the robot.

Demonstration

✓ Demonstration is produced with the intention of **showing** a **state-action sequence** to robot.

✗ Teaching with a demonstration strategy imposes a **significant burden** on the human teacher.

Instruction

✓ An instruction is produced with the intention of **communicating** the **action** to be performed in a given task state.

✓ Learning from instructions - mapping instructions (e.g. natural language) to a sequence of executable actions.

[1] M. Chetouani, "Interactive Robot Learning: An Overview," **ECCAI Advanced Course on Artificial Intelligence**, pp. 140–172, 2021

Teaching Signals

Teaching signals		Feedback	Demonstration	Instruction
Nature	Notation	$H(s,a)$	$D = \{(s_t, a_t^*), (s_{t+1}, a_{t+1}^*)\ldots\}$	$I_\pi(s) = a_t^*$
	Value	Binary/Scalar	State-Action pairs	Probability of an action
Time-step	t − 1		✓	✓
	t		✓	
	$t+1$	✓		
Human	Intention	Evaluating/Correcting	Showing	Telling
	Teaching cost	Low	High	Medium
Robot	Interpretation	State-Action evaluation Reward-/Value-like	Optimal actions Policy-like	Optimal action Policy-like
	Learning cost	High	Low	High

[1] M. Chetouani, “Interactive Robot Learning: An Overview,” **ECCAI Advanced Course on Artificial Intelligence**, pp. 140–172, 2021

How Can Humans Deliver Evaluative Feedback?

Two **ways** of delivering evaluative **feedback**:

- Hardware
- Natural interaction modalities

Two **types** of evaluative **feedback**:

- **Explicit** - user gives **direct feedback** to the robot e.g. through the graphical interface.
- **Implicit** - **spontaneous behaviour** of the human user is analysed and used as feedback. Feedback is estimated based on social signals such as valence, engagement, facial expressions.

Arity of Human Evaluative Feedback

- Unimodal $\rightarrow$ Only one feedback modality is used.
- Multimodal:
  - Multiple modalities are used either **disjunctively** (OR) or **conjunctively** (AND).
  - If **one** modality is unavailable, then the others can be used $\rightarrow$ **robustness**.
  - If **all** are available, then the feedback is more **reliable**.
  - Examples:
    - Speech and gesture
    - Laugh (audio) and smile (visual)
    - Facial expressions of emotions + task-related features
- Human evaluative feedback can also be combined with a pre-defined environmental reward function.

Challenges with Human Feedback

- Delayed feedback
  - E.g. Due to reaction times involved in evaluating the action and providing the feedback.
  - To which action should the feedback be mapped?
- Interpersonal variability
  - E.g. The reaction times differ from person to person.
  - The social signals used for feedback vary in modality, in expression, and in intensity.
  - The same teacher might change the feedback strategy or type over time (e.g. change from binary to categorical feedback).

Challenges with Human Feedback

- Decay in feedback frequency
  - Intense feedback at the beginning, sparse feedback later.
- Multiple feedback
  - Should the feedback be aggregated or should only one of those multiple feedback be used?
- Noise in feedback channel
  - Discrepancies between what the teacher intends to convey and what the agent actually observes.

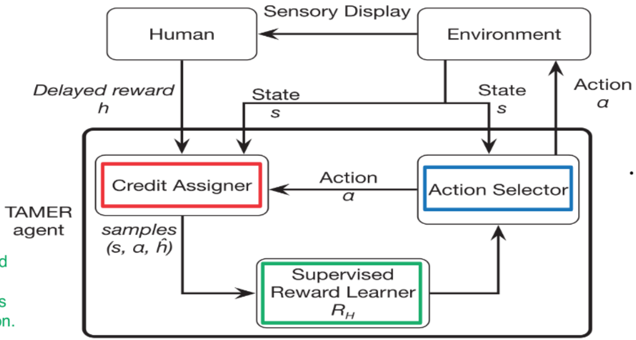
Human-centered RL

- TAMER [2]
  - Human feedback is mapped to numeric value.
- SABL [3]
  - Human feedback is mapped to categorical strategies.
- COACH [4]
  - Human feedback is mapped to agent's policy (selecting actions).

[2] W. B. Knox and P. Stone, "Interactively shaping agents via human reinforcement: The TAMER framework," in **Proc. of the Fifth Int. Conf. on Knowledge Capture**, 2009, pp. 9–16
[3] R. Loftin **et al.**, "A Strategy-Aware Technique for Learning Behaviors from Discrete Human Feedback," in **Proc. of the AAAI Conf. on Artificial Intelligence**, vol. 28, no. 1, 2014
[4] J. MacGlashan **et al.**, "Interactive Learning from Policy-Dependent Human Feedback," in **Proc. of the 34th Int. Conf. on Machine Learning**, ser. Proc. of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 2285–2294

TAMER

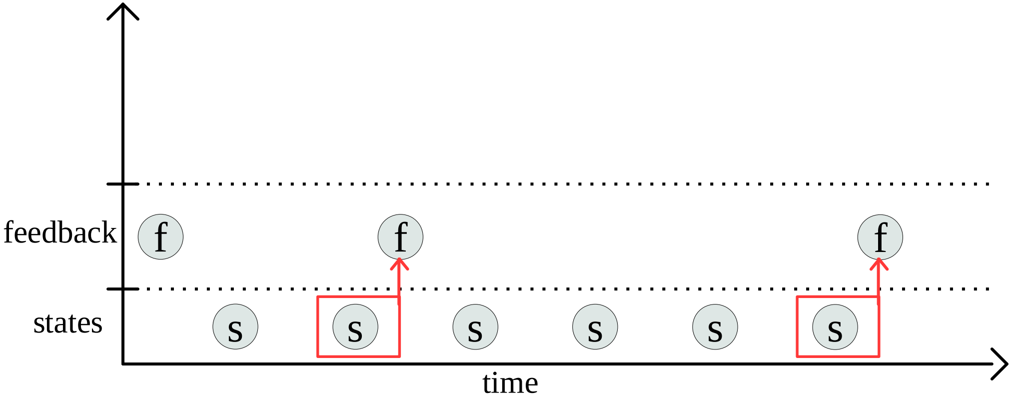
- Probabiliy density function to model the feedback delay.
- Learns a parameterised model for expected human reward, which is used for action selection.



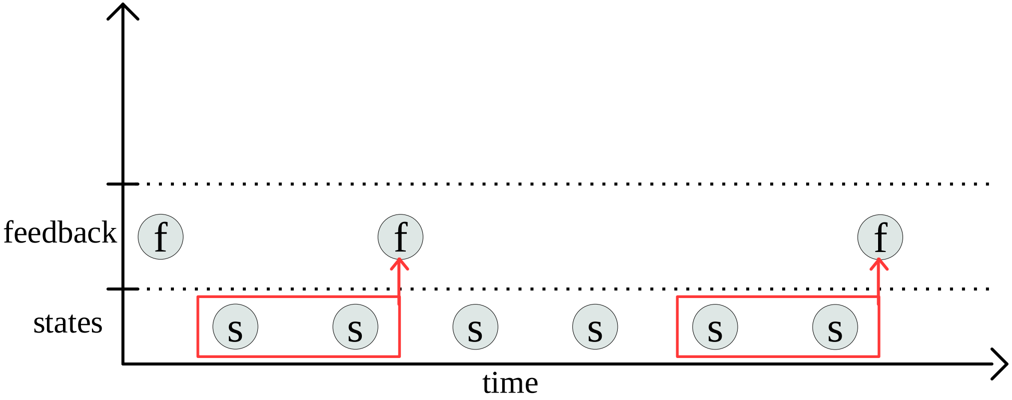
- Myopic rewards (γ=0) under the assumption that the human takes the long-term consequences into account in their evaluative feedback.

[5] W. B. Knox, “Learning from human-generated reward,” Ph.D. dissertation, University of Texas at Austin, 2012

Without Credit Assignment



With Credit Assignment



Practical Example

- Agent is deployed in a simple **Mountain Car** environment, with
  - **continuous state space** - position of the car along the x-axis and velocity of the car
  - **discrete action space** - accelerate to the left, don't accelerate, accelerate to the right
- The **user** can give positive/negative **feedback** using the keyboard or no feedback (by not giving any input)

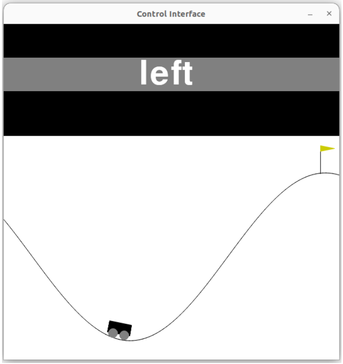
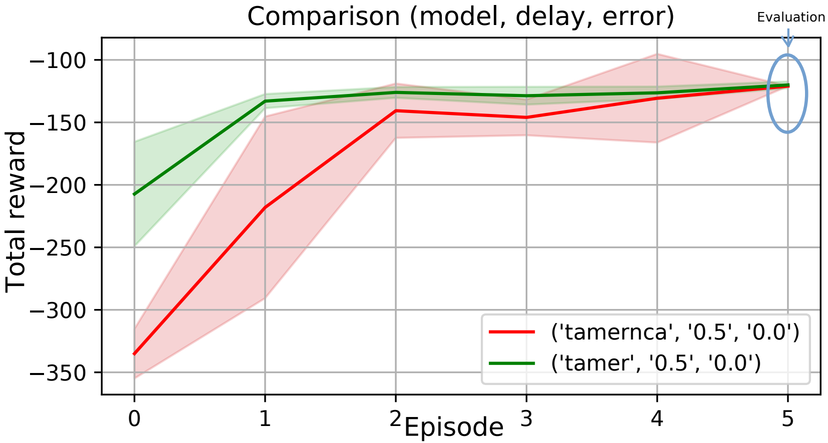


Figure 1: View of the deployment environment

Advantage of Credit Assignment



Going "deeper" ...

- One needs to find features (e.g. x car position, velocity) that can sufficiently define a state for TAMER.
- Would be easier if TAMER finds these features itself based on the image (like humans do).
- This problem is addressed by Deep TAMER [6].

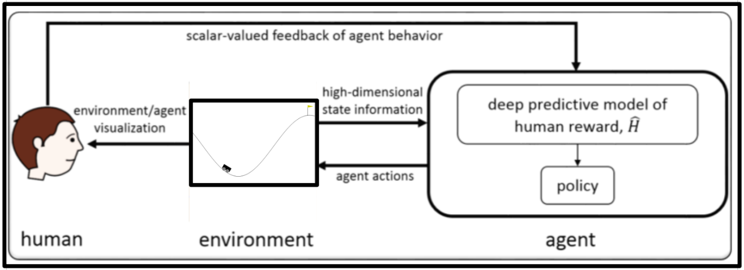


Figure 2: Based on [6]

[6] G. Warnell, N. Waytowich, V. Lawhern, and P. Stone, “Deep TAMER: Interactive Agent Shaping in High-Dimensional State Spaces,” in **Proc. of the AAAI Conf. on Artificial Intelligence**, vol. 32, no. 1, 2018

Is Deep TAMER a Solution?

- Needs pretraining of the encoder part (a lot of data necessary).
- Needs two input images for some environments (e.g. to estimate velocity).
- Hard to assess the quality of the extracted features.

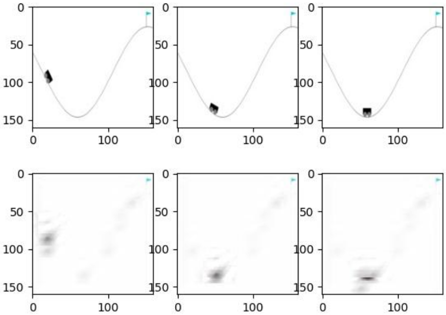


Figure 3: Reconstruction of the images from the features extracted by the encoder part of Deep TAMER

Categorical Feedback Strategies

A human teacher's feedback can be categorized into four types (inspired by behaviourism and animal training):

- Positive reward (R+)
  - Explicit feedback for correct behaviour.
- Negative reward (R-)
  - No feedback for correct behaviour.
- Positive punishment (P+)
  - Explicit feedback for wrong behaviour.
- Negative punishment (P-)
  - No feedback for wrong behaviour.

Categorical Feedback Strategies

- Different combinations of these feedback types are possible and it forms the teacher's feedback strategy.
  - Reward-focused $\rightarrow$ R+/P-
  - Punishment-focused $\rightarrow$ P+/R-
  - Balanced $\rightarrow$ R+/P+ (explicit reward and explicit punishment)
  - Inactive $\rightarrow$ R-/P- (rarely gives explicit feedback)

- The teacher can change the strategy during the course of training.
  - Teacher's feedback modelled probabilistically [7] and used with SABL algorithm.
  - Parameters:
    - » $\mu+$ $\rightarrow$ Probability that teacher will not give explicit feedback for correct behaviour
    - » $\mu-$ $\rightarrow$ Probability that teacher will not give explicit feedback for wrong behaviour
    - » ϵ $\rightarrow$ Probability that teacher misjudges the correctness of an action

[7] R. Loftin **et al.**, "Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning," **Autonomous Agents and Multi-Agent Systems**, vol. 30, pp. 30–59, 2016

Strategy-Aware Bayesian Learning (SABL)

- Bayessian inference.
- Assumes that teacher's strategy is **known**, i.e. $\mu+$ and $\mu-$ are known.
- Policy is updated based on the categorical probability of the given human feedback.
- Can be used only for low-dimensional discrete state space.
- Variant: **Inferring-SABL** or **I-SABL**
  - In reality, the teacher's strategy (i.e. $\mu+$ and $\mu-$) is **unknown**.
  - I-SABL **infers** the **teacher's strategy** by analyzing the feedback history.

Motivation for COACH

The following **characteristics** of training strategies used by humans:

- Correct actions are given **less positive feedback** progressively, as the **agent learns** to use that action succesfully.
- **Strength** of feedback varies depending on how much **improvement** or **deterioration** is observed in the agent's behaviour.
- Suboptimal actions may receive **positive** feedback if it **improves** the agent's behaviour; after the behaviour improves, the same suboptimal actions are given **negative feedback**.

[4] J. MacGlashan **et al.**, "Interactive Learning from Policy-Dependent Human Feedback," in **Proc. of the 34th Int. Conf. on Machine Learning**, ser. Proc. of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 2285–2294

COnvergent Actor-Critic by Humans (COACH)

- COACH [4] is actor-critic-based reinforcement learning algorithm where human feedback is used as an **advantage function**.
  - advantage function - function that describes an **advantage** of selecting a certain action **over** the agent's policy
- The **sparse feedback** (also delayed feedback) problem is faced with **eligibility traces** which can smooth observed human feedback over past transitions.
- **Deep COACH** [8], namely COACH for **high-dimensional** input.

[4] J. MacGlashan **et al.**, "Interactive Learning from Policy-Dependent Human Feedback," in **Proc. of the 34th Int. Conf. on Machine Learning**, ser. Proc. of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 2285–2294
[8] D. Arumugam, J. K. Lee, S. Saskin, and M. L. Littman, "Deep Reinforcement Learning from Policy-Dependent Human Feedback," **arXiv preprint arXiv:1902.04257**, 2019

Improving Learning from Evaluative Feedback

- **Human feedback** is usually:
  - **dense** (at the beginning of the interaction)
  - **flawed** (people generally make mistakes evaluating the agent's behaviour)
- On the other hand, **environmental reward** is usually:
  - **sparse**
  - **flawless** (determines optimal behaviour)
- Why not combine **human feedback** (HF) and **environmental reward** (ER) for agent learning?

Combining HF and ER

- **Reward shaping**
  - Modelled human feedback $\hat{H}$ is interpreted as a **reward**.
  - $r'(s, a) = r(s, a) + \beta * \hat{H}(s, a)$
- **Value shaping**
  - Modelled human feedback $\hat{H}$ is interpreted as **action-value function** (expected cumulative reward given that the agent starts with action a from s following policy π).
  - $Q'(s, a) = Q(s, a) + \beta * \hat{H}(s, a)$
- **Policy shaping**
  - Modelled human feedback $\hat{H}$ employed to directly influence the agent's **policy**.
  - e.g. $P(a = argmax(\hat{H}(s, a))) = min(\beta, 1)$

[1] M. Chetouani, "Interactive Robot Learning: An Overview," **ECCAI Advanced Course on Artificial Intelligence**, pp. 140–172, 2021

[9] W. B. Knox and P. Stone, "Combining manual feedback with subsequent MDP reward signals for reinforcement learning," in **Proc. of the 9th Int. Conf. on Autonomous Agents and Multiagent Systems: volume 1-Volume 1**. Citeseer, 2010, pp. 5–12

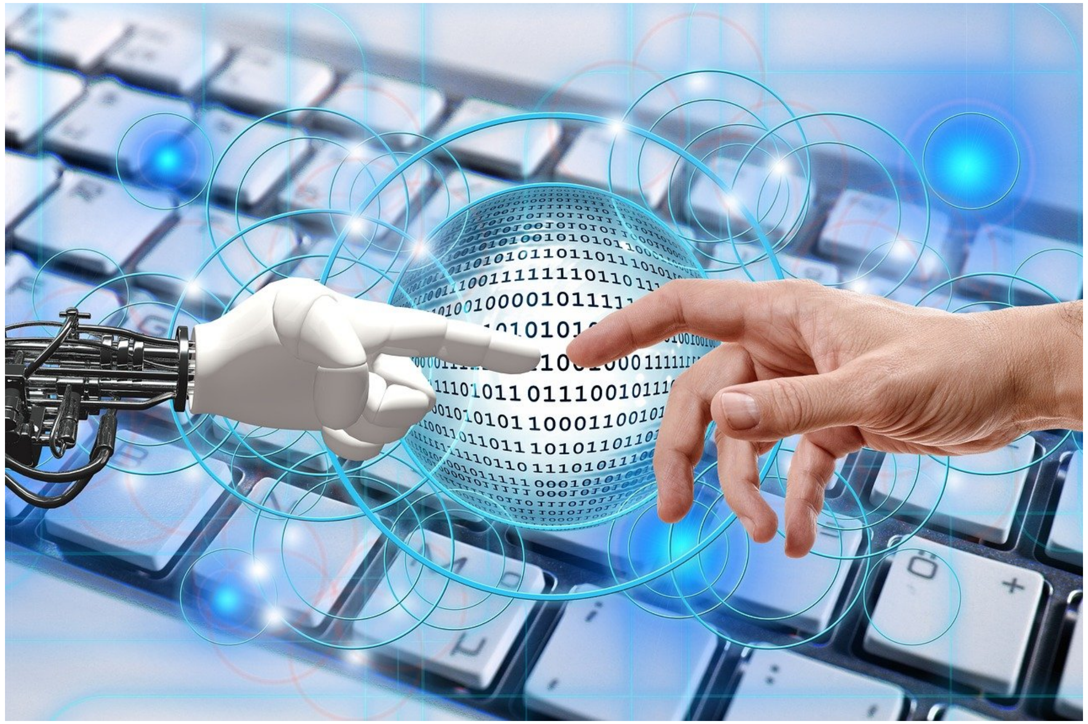
1. What is Interactive Reinforcement Learning?
2. Teaching signals used in IRL
3. Human evaluative feedback
4. Learning from evaluative feedback
5. TAMER
6. SABL
7. COACH
8. Combining evaluative feedback and environmental reward
9. **Real-life application**
10. Summary

1. What is Interactive Reinforcement Learning?
2. Teaching signals used in IRL
3. Human evaluative feedback
4. Learning from evaluative feedback
5. TAMER
6. SABL
7. COACH
8. Combining evaluative feedback and environmental reward
9. Real-life application
10. Summary

Summary

- Human input can be used to **speed-up robot learning** in real-world tasks.
- Human input can take the form of **demonstrations**, **instruction**, or **evaluative feedback**.
- Learning from human evaluative feedback is called **human-centered reinforcement learning**.
- There are several **challenges** associated with obtaining, interpreting and using human input.
- Frameworks and methods that use human evaluative feedback include **TAMER**, **SABL** and **COACH**, to name a few.
- There are methods to combine human evaluative feedback and environmental reward including **reward shaping**, **value shaping** and **policy shaping**.

Conducting Human-Robot Interaction Experiments



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

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

20th June 2024

- Investigate and make inferences about the impact of robots on humans.
  - Which design is perceived as **safer** and **more trusted**?
  - Is a robot tutor **more effective** than a human tutor?
  - Are the emotions expressed by the robot **believable**?
  - Do the explanations given by the robot **improve understandability**?

Technical design

Human factor

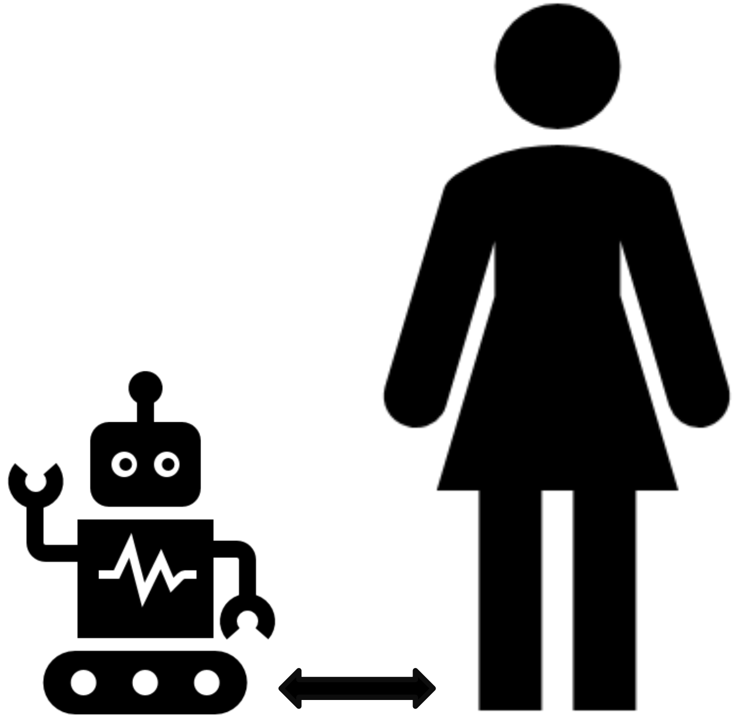
HRI – An Example Scenario

- You have a robot that can walk alongside humans.

- Motivated by findings on how humans walk alongside humans:

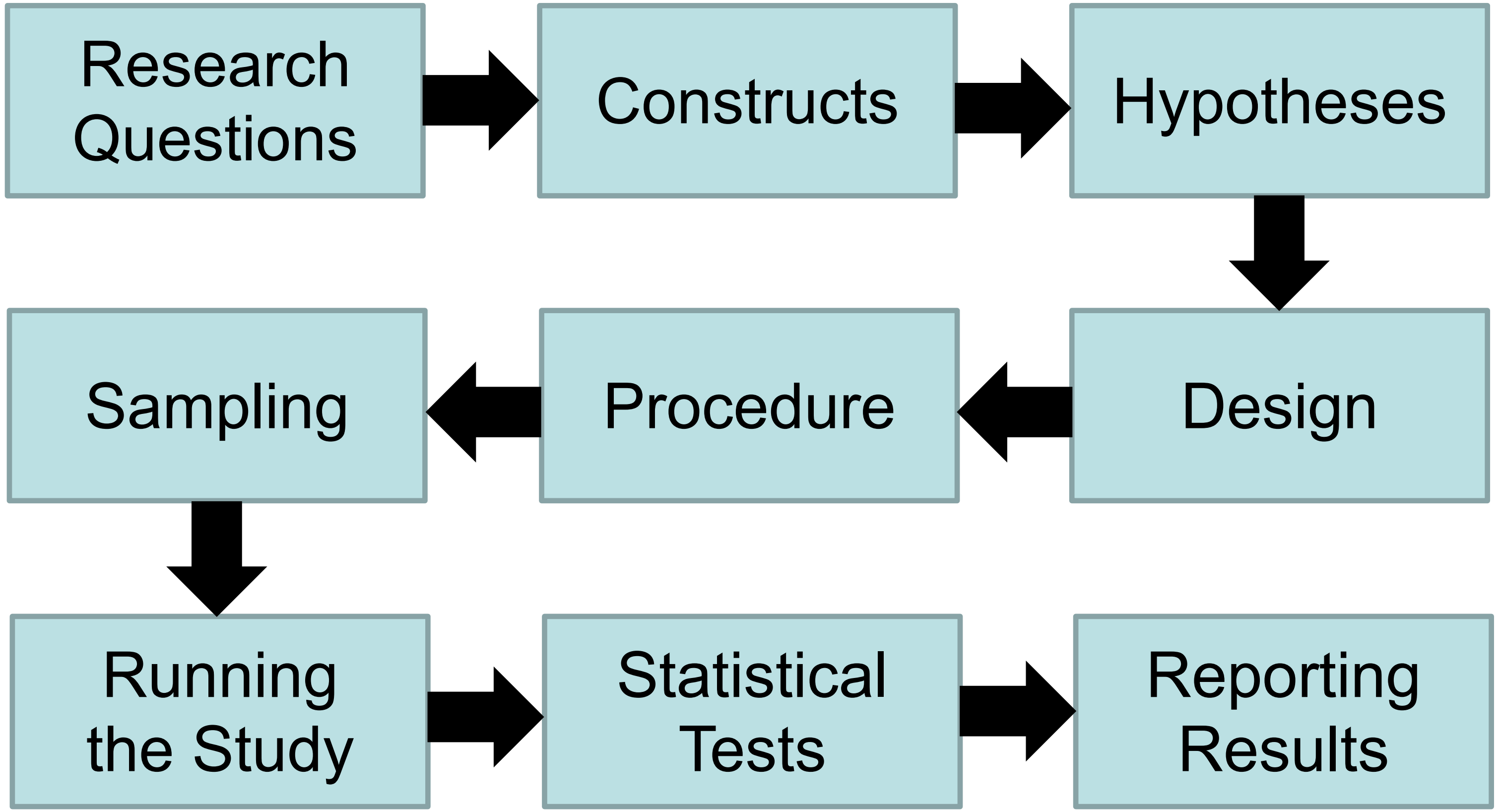
  - A novel, human-aware, navigation algorithm that can **adapt the robot's path according to the movement patterns of the human**.

Example from (Hoffman & Zhao, 2020)



A²S

Hypothesis-Driven Experimental Research



(Hoffman & Zhao, 2020)

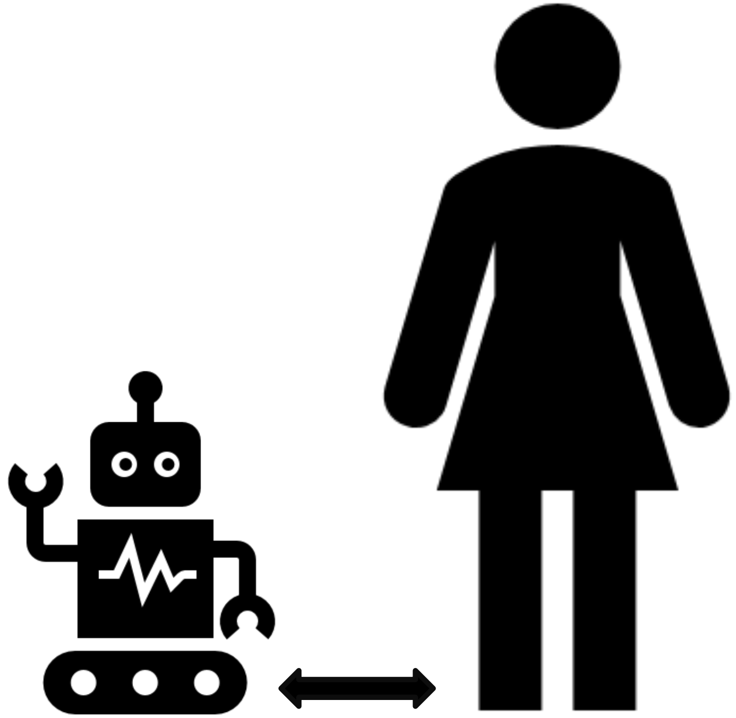
Identify Research Questions

- **Define research questions** based on theoretical foundations, previous experiments.

- Research questions:
  1. "To what extent, if any, will a human-adaptive path algorithm make people **trust** the robot to accompany them? "

  2. "Do people feel **safe** walking with a robot running the new algorithm?"

Example from (Hoffman & Zhao, 2020)

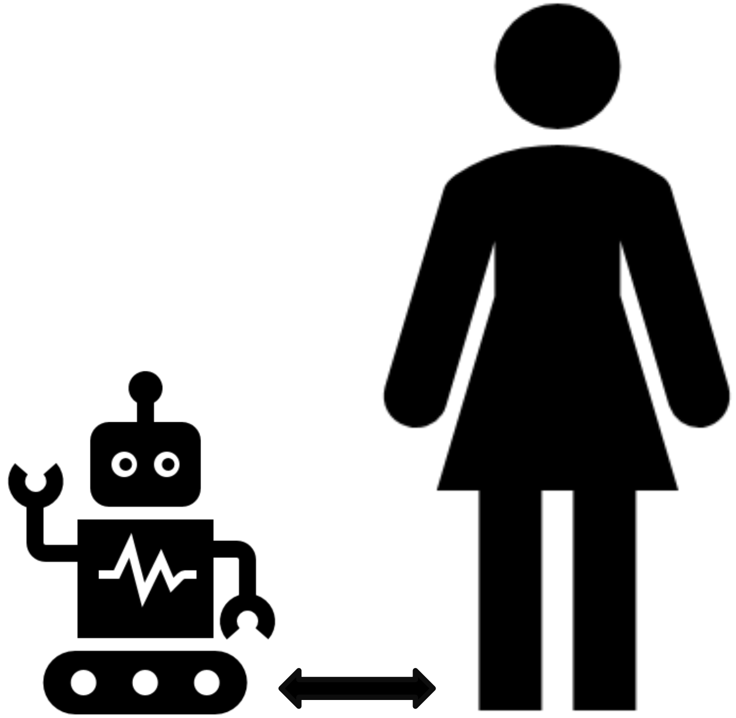


A²S

- **Identify the constructs** (theoretical / abstract concepts) that are central to the research questions.

1. Human-adaptive movement of robot (C1)
2. Trust of human in the robot (C2)
3. Human's feeling of safety (C3)

(Hoffman & Zhao, 2020)



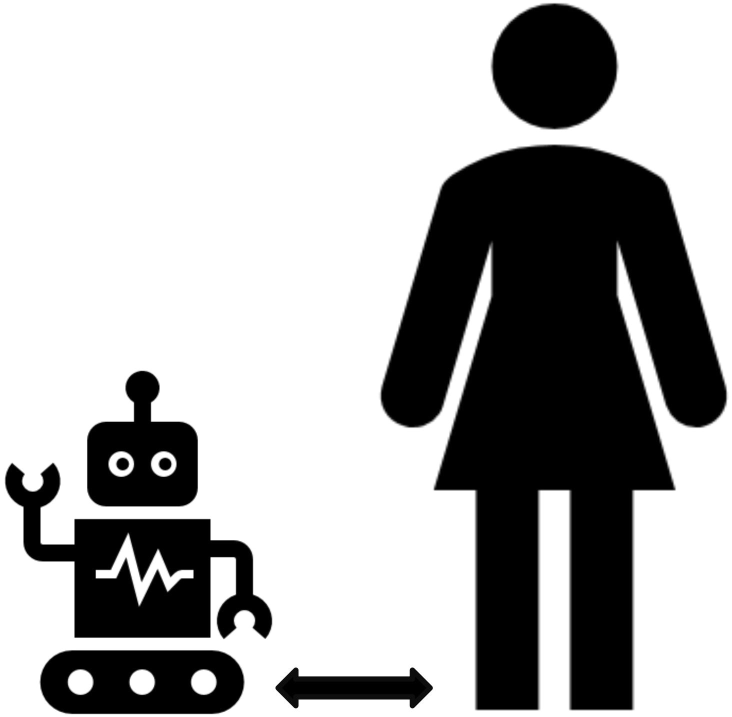
Formulate Hypotheses

- Formulate hypotheses about the **relationship between constructs**.
  - Causal versus correlation

- Formulate hypotheses before running the study.
- Otherwise it is an exploratory study.

- Always compare against a baseline.

(Hoffman & Zhao, 2020)



A²S

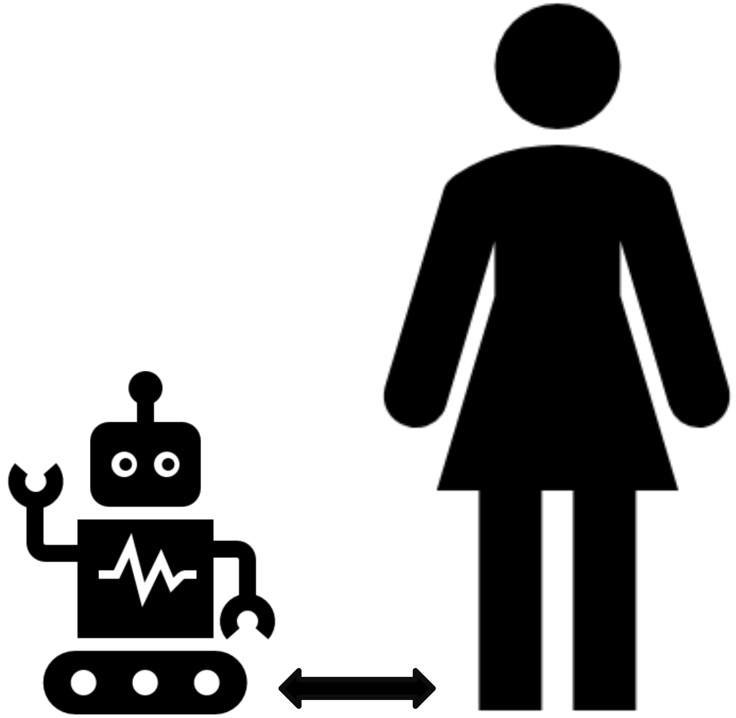
Formulate Hypotheses

- One example hypothesis:
  - H1: Links C1 and C2 causally.
    - Users trust (C2) the robot with the human-adaptive movement (C1) more than a robot that walks along a straight-line path to goal.

C1 is predictor (discrete).
C2 is outcome.

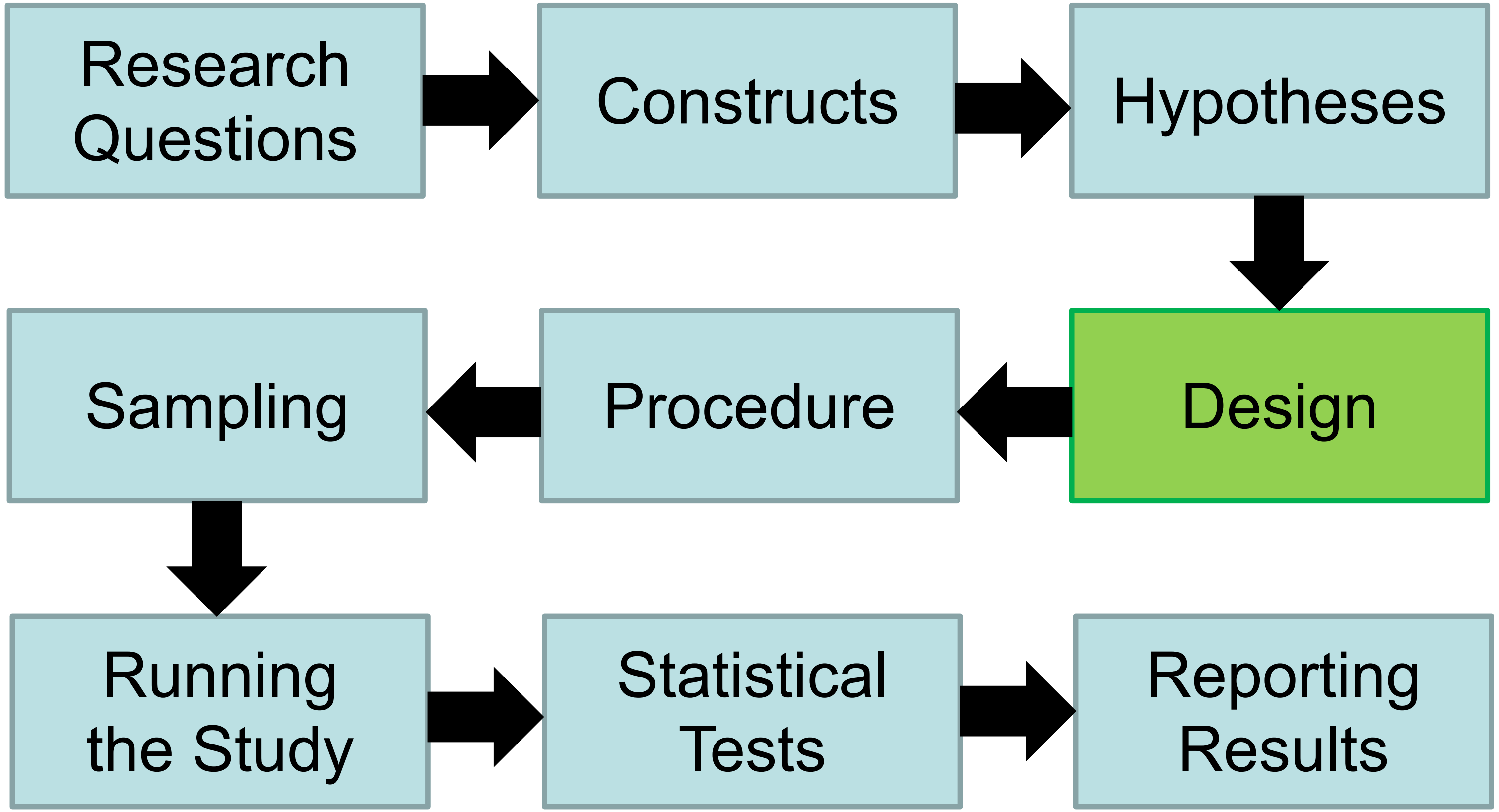
What is the baseline here?

(Hoffman & Zhao, 2020)



A²S

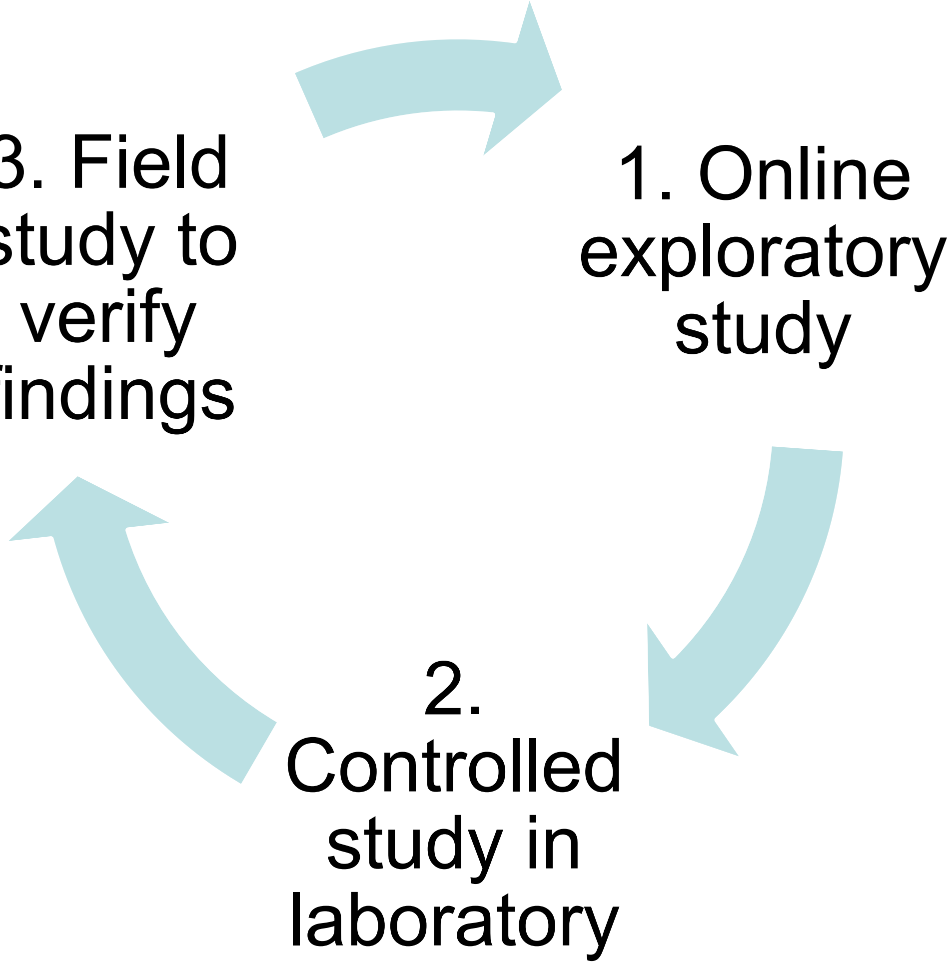
Hypothesis-Driven Experimental Research



(Hoffman & Zhao, 2020)

A²S

Designing Study: Full-Cycle Research



Chad R. Mortensen and Robert B. Cialdini. 2010. Full-cycle social psychology for theory and application. *Soc. Pers. Psychol. Compass* 4, 1 (2010), 53–63.

Designing Study

Field	Laboratory	Online
• Close to reality • Good external and ecological validity	• Easy to control the variables • Less influence of confounding variables	• Low cost • Easy to find participants • More diversity
• Confounding variables affect the internal validity	• Ecological and external validity could be poor.	• Difficult to control environment • Experience levels differ • No real interaction possible

A²S

Designing Study

Operationalise constructs into variables that can be manipulated or measured.

For H1:

1. Human-adaptive movement of robot (C1)
   - Conditions tested in the experiment

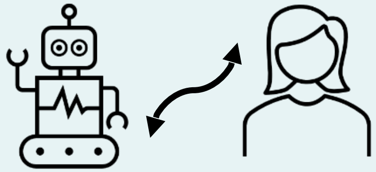
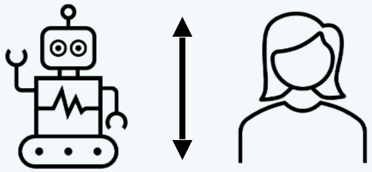
     → Independent variable

2. Trust of human in the robot (C2)
   - Subjective: Questionnaires
   - Objective: Behaviour

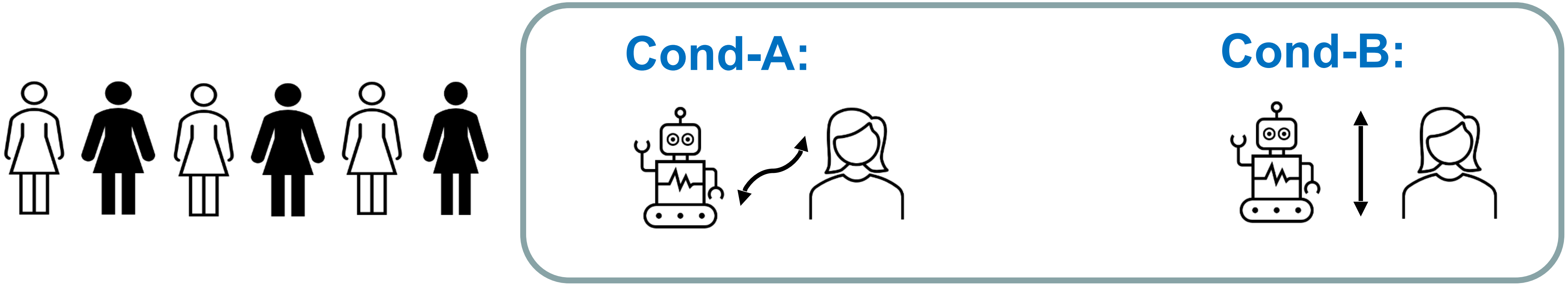
     → Dependent variables

Designing Study

- How many conditions should each subject (samples from a population) be exposed to?

Condition	Description	Icon
Cond-A	**Robot performs human-adaptive movement**	
Cond-B	**Robot moves along straight-line path**	

Within-Subjects Design



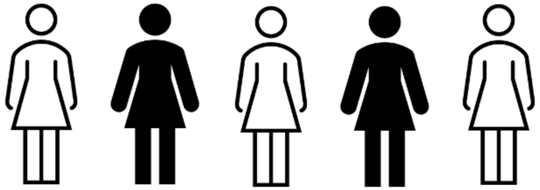
Between-Subjects Design

A²S

We Need…

Power Analysis

Simply put, a **statistically sound** method to determine **the sample size** and **power** of an experiment.



Sample size

Power

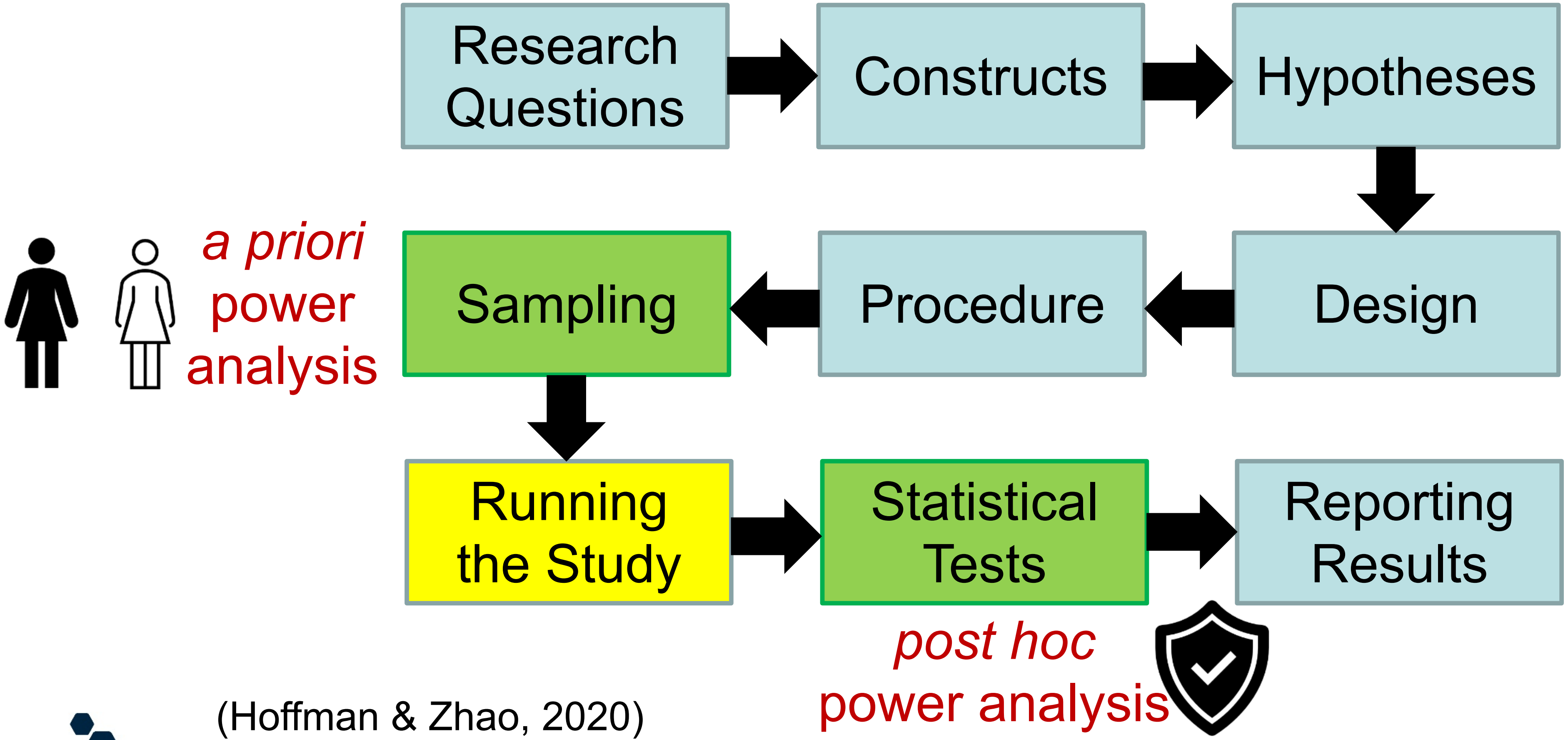
A²S

Why is Power Analysis Important in HRI?

- Humans are **unpredictable** and **exhibit variability** in their behavior.
  - Inter- and intrapersonal variability
- When we conduct an experiment:
  - How can we be sure that the **effects that we observed are meaningful** and not noise?
- **We should perform a priori and post hoc power analyses.**

(Bartlett et al., 2022)

Hypothesis-driven Experimental Research

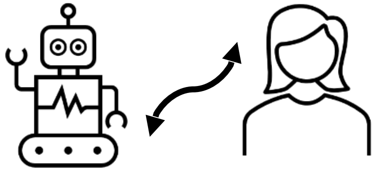


(Hoffman & Zhao, 2020)

A²S

Hypothesis H1 Revisited

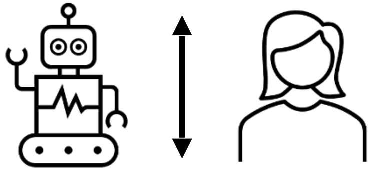
Level of users‘ trust in:



Cond-A

>

Level of users‘ trust in:

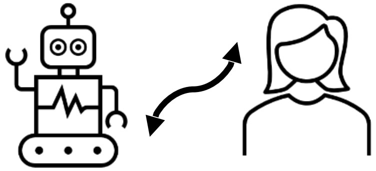


Cond-B

A²S

The Null Hypothesis H0

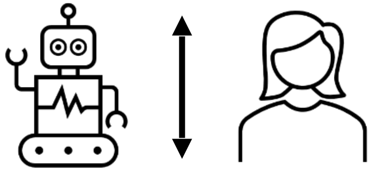
Level of users‘ trust in:



Cond-A

=

Level of users‘ trust in:



Cond-B

➢ **H0** posits there is no difference between users‘ level of trust in cond-A and cond-B.

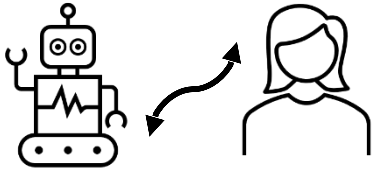
Null Hypothesis

A²S

Objective of the Experiment

- **Can we reject H0:**

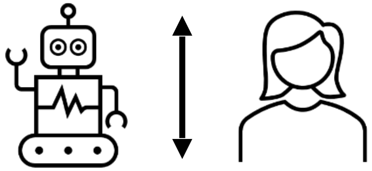
Level of users' trust in:



Cond-A

=

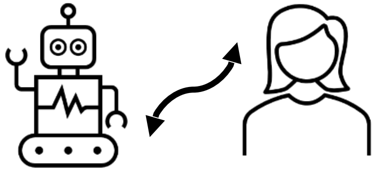
Level of users' trust in:



Cond-B

- **And accept H1:**

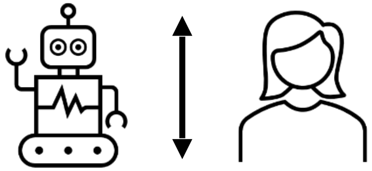
Level of users' trust in:



Cond-A

>

Level of users' trust in:



Cond-B

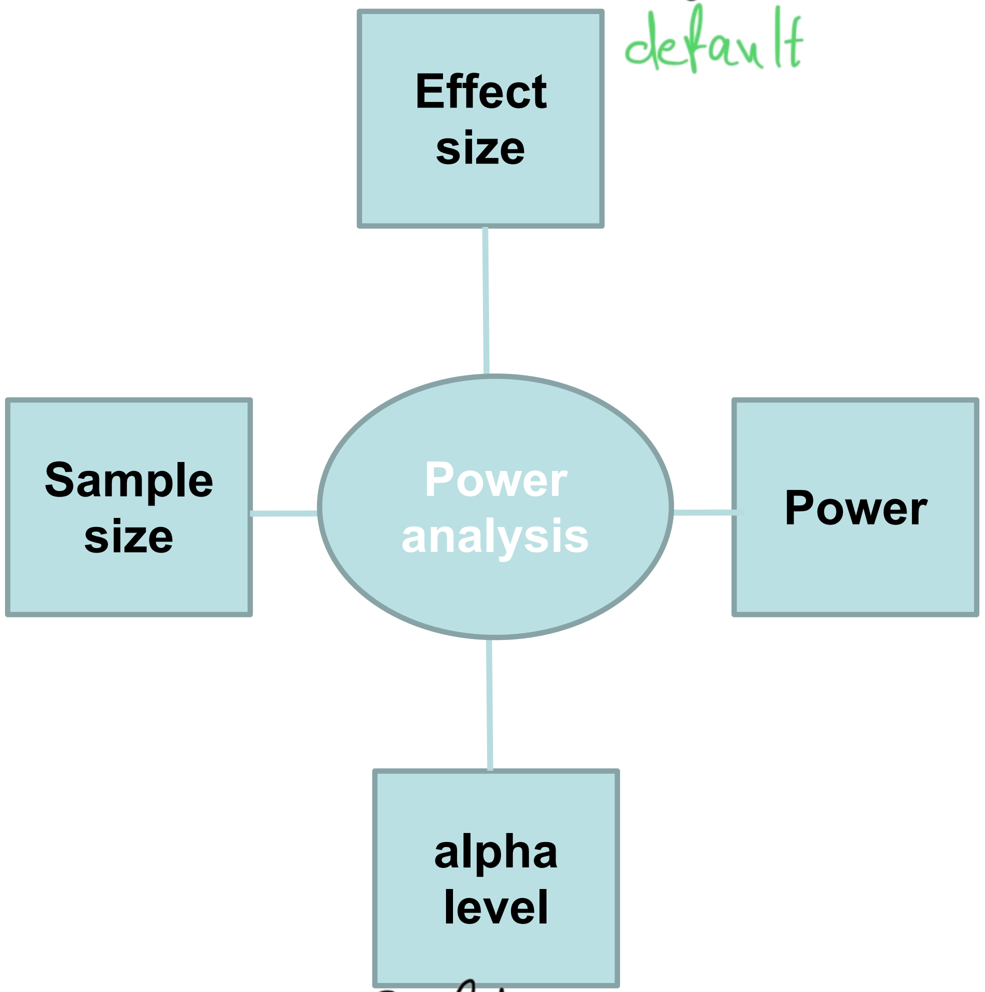
Power Analysis

- Examines the relationship between four parameters:
  1. alpha level
  2. Power
  3. Effect size
  4. Sample size

- Given any three of these parameters, we can predict the fourth.

Small Medium High

default



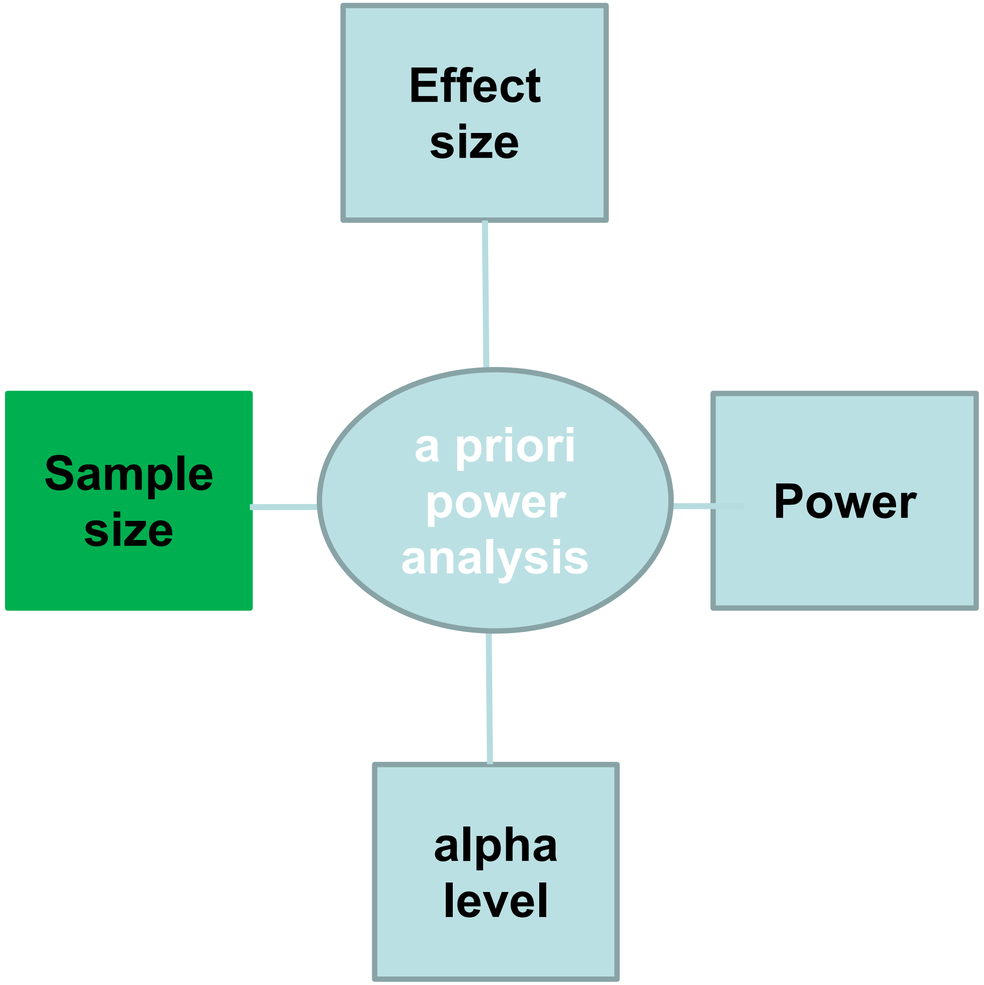
Confidence

Too small alpha can be High Samples

↓↓ Alpha ↑↑ Better

a priori Power Analysis

- Given:
  - Expected or desired values for:
    1. alpha level
    2. Power
    3. Effect size
- Predict, sample size needed for the experiment.

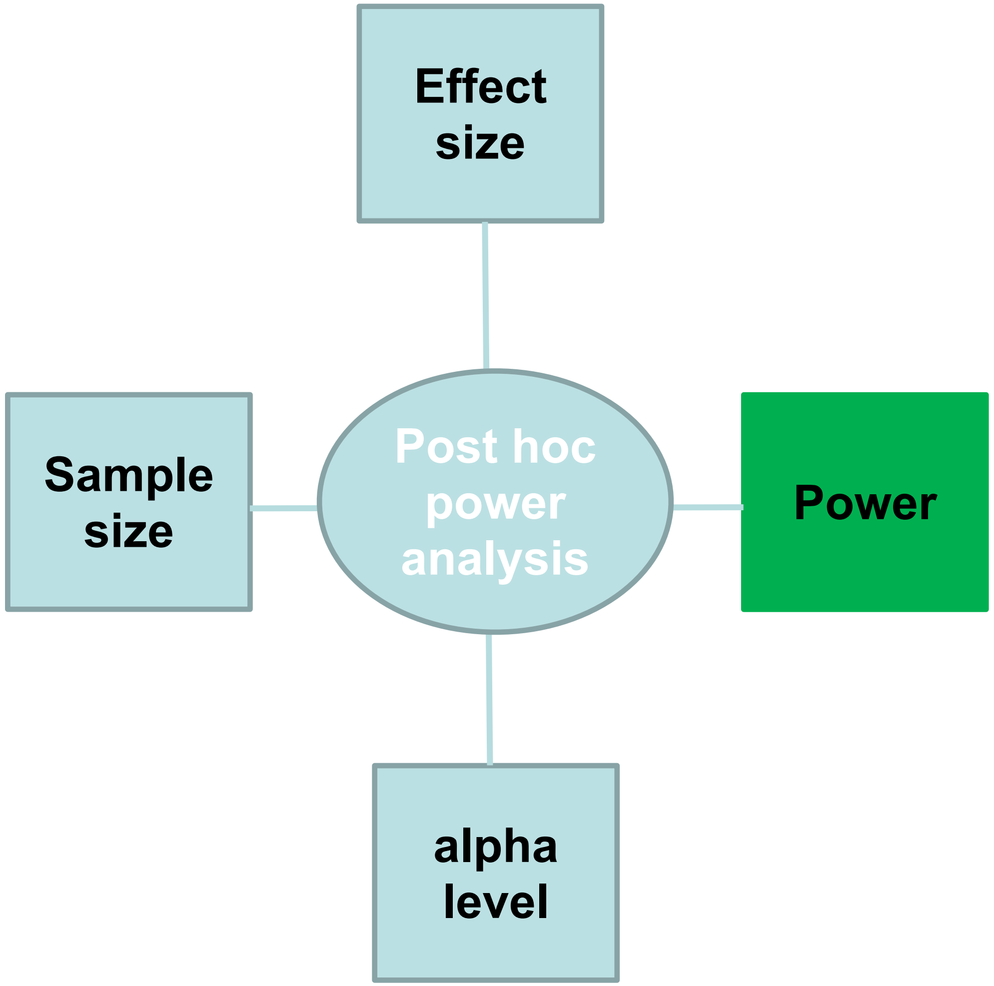


A²S

post hoc Power Analysis

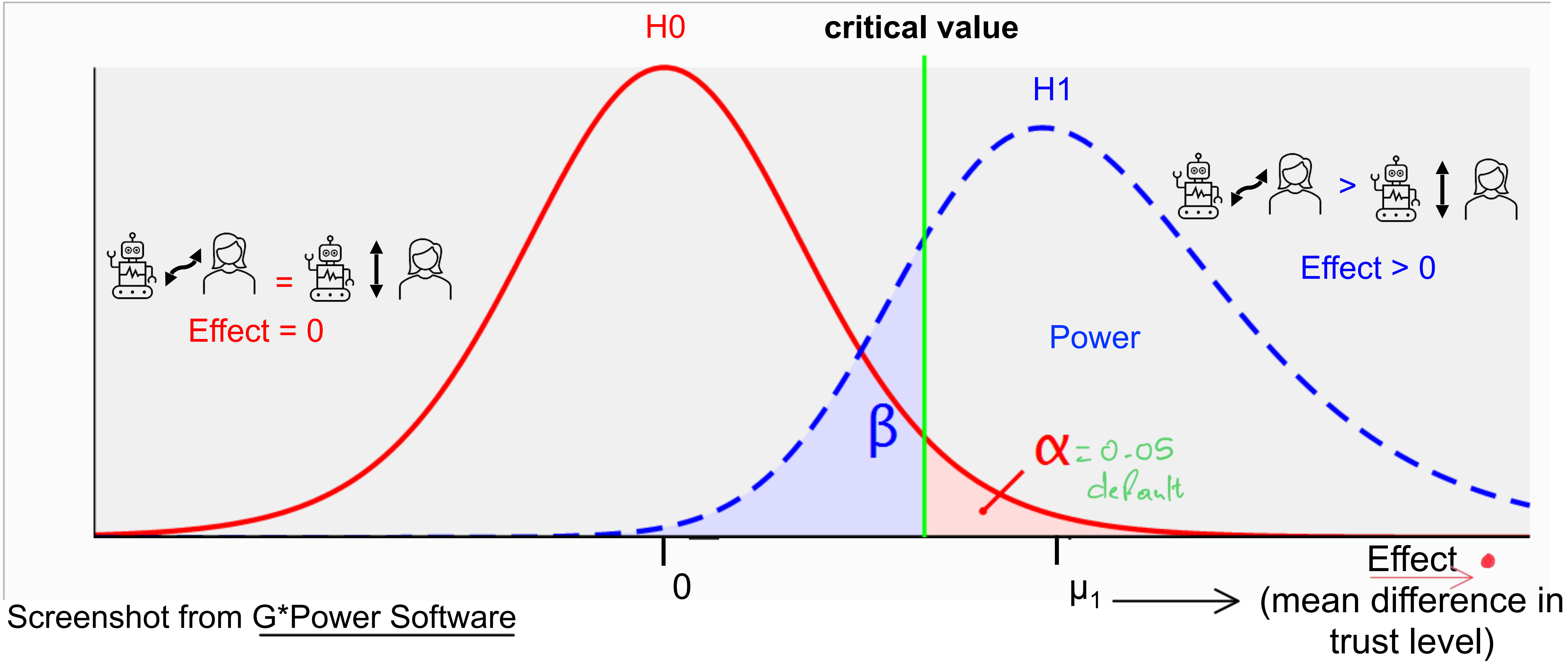
- Done after the experiment.
- Given:
  - Values computed from experimental data:
    1. Effect size
    2. Sample size
    3. alpha level

- Predict statistical power of the experiment.



A²S

Alpha Level, Power, Effect Size



Screenshot from G*Power Software

To get the same power after reducing alpha, we should increase the sample size.

Errors Induced by Sampling – Type I, Type II

	Reject H0 Accept H1	Accept H0 Reject H1
H0 is True H1 is False	α False positive rate (Type I error)	1- α True negative rate
H0 is False H1 is True	1- β True positive rate **(POWER)**	β False negative rate (Type II error)

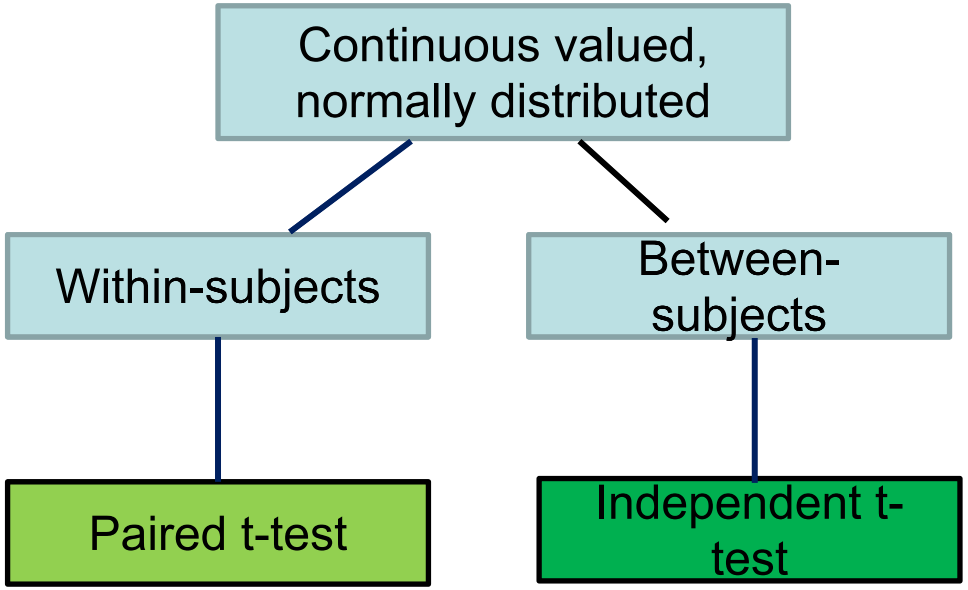
Holds in the general population

Holds in the sampled population

A²S

Estimating Effect Size

- **Effect size is the difference between outcomes measured in different conditions.**
- Effect size computation depends on the method chosen for the statistical test.



A²S

Computing Effects Using t-tests

- Independent t-test
  - Between subject
    - Two separate groups of subjects A and B
    - Each group exposed to only one condition
    - Two sampling distributions
  - Computes means of dependent variable in either distribution: μ_A, μ_B
  - Difference between means: $\Delta = \mu_A - \mu_B$
  - Effect: Compares Δ to zero.
  - Effect size (Cohen's d): Δ / σpooled

- Paired t-test
  - Within subject
    - One group of subjects
    - Each subject exposed to both conditions A and B
  - One sampling distribution S based on difference Δ_s in response of each subject s in condition A and condition B.
    - $\Delta_s = a_s - b_s$
    - μ_S= Mean Δ_s over all s
  - Effect: Compares μ_S to zero.
  - Effect size (Cohen's d): μ_S / σ_S

$$\sigma_{pooled} = \sqrt{\left(\sum(a - \mu_A)^2 + \sum(b - \mu_B)^2\right) / (n_A + n_B - 2)}$$

A²S

a priori Power Analysis

- Commonly used values:
  - alpha: 0.05 or 0.01
  - power: 0.8 (HRI), 0.95 (clinical)
  - effect-size (for t-tests): 0.5

	Small effect size	**Medium effect size**	**Large effect size**
Cohen's d	0.2	0.5	0.8

Jacob Cohen. 1992. A power primer. Psychol. Bull. 112, 1 (1992).