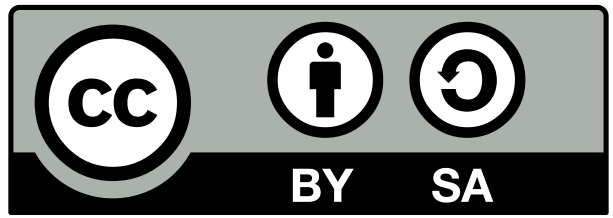
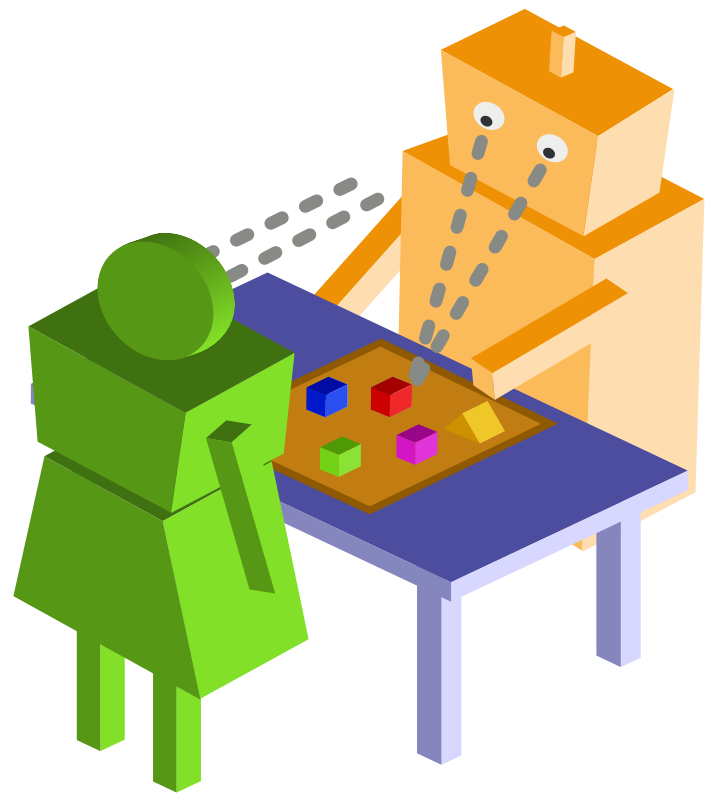


You can download the sources of this presentation here:

github.com/severin-lemaignan/presentation-codesign

Co-designed head to toe
Towards end-to-end participatory design

LAAS-CNRS | 13 Dec 2021



Séverin Lemaignan

PAL Robotics Senior Scientist AI & Social Interactions

situation assessment

symbolic grounding

symbolic reasoning

SYMBOLIC SOCIAL COGNITION FOR ROBOTS

ontologies

perspective taking

cognitive architectures

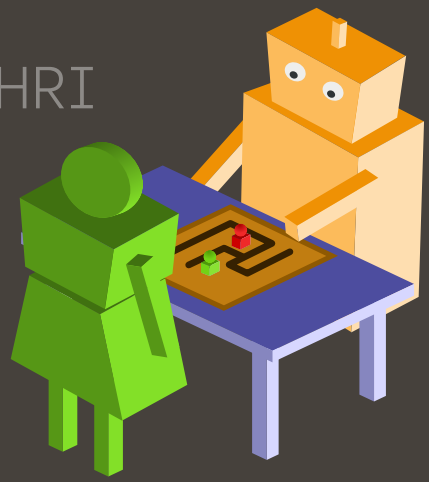
social situation assessment

joint action

ROS4HRI

natural language processing

REAL-WORLD SOCIAL AUTONOMY



DATA-DRIVEN HRI

learning of social policies

large datasets

theory of mind

group dynamics

human-in-the-loop ML

experimental robotics

robotics for learning

CHILD-ROBOT INTERACTION

trust

HUMAN FACTORS

engagement

responsible AI

anthropomorphism

social robotics

participatory design

persuasion

SOCIAL ROBOTICS

Creating interactive robots that are **embedded and understand their (human) social context**; **generate and adopt appropriate social behaviours**; have a **positive impact on human society**.

⇒ designing and implementing the **assistant and companion robots** for tomorrow.

⇒ direct impact on ageing society, education, customer service; **major socio-economic challenge**

SOCIAL ROBOTICS

Major scientific challenges:

- Model open-ended, underspecified situations; rich semantics; complex social dynamics;

SOCIAL ROBOTICS

Major scientific challenges:

- Model open-ended, underspecified situations; rich semantics; complex social dynamics;
- Close the interaction loop;

SOCIAL ROBOTICS

Major scientific challenges:

- Model open-ended, underspecified situations; rich semantics; complex social dynamics;
- Close the interaction loop;
- Understand and sustain long-term autonomous social interactions;

SOCIAL ROBOTICS

Major scientific challenges:

- Model open-ended, underspecified situations; rich semantics; complex social dynamics;
- Close the interaction loop;
- Understand and sustain long-term autonomous social interactions;
- Real-world algorithmic robustness;

SOCIAL ROBOTICS

Major scientific challenges:

- Model open-ended, underspecified situations; rich semantics; complex social dynamics;
- Close the interaction loop;
- Understand and sustain long-term autonomous social interactions;
- Real-world algorithmic robustness;
- Complex ethical landscape;

SOCIAL ROBOTICS

Major scientific challenges:

- Model open-ended, underspecified situations; rich semantics; complex social dynamics;
- Close the interaction loop;
- Understand and sustain long-term autonomous social interactions;
- Real-world algorithmic robustness;
- Complex ethical landscape;
- $\Rightarrow$ cross-disciplinary & holistic approach required
- $\Rightarrow$ involve all the stakeholders; participatory approach

LET SET OURSELVES A CHALLENGE

Design & run a study with:

- a real robot

LET SET OURSELVES A CHALLENGE

Design & run a study with:

- a real robot
- a real interaction (...with a **human**!)

LET SET OURSELVES A CHALLENGE

Design & run a study with:

- a real robot
- a real interaction (...with a **human**!)
- a **continuous** interaction

LET SET OURSELVES A CHALLENGE

Design & run a study with:

- a real robot
- a real interaction (...with a **human**!)
- a **continuous** interaction
- a realistic task (**large state vector & action space**)

LET SET OURSELVES A CHALLENGE

Design & run a study with:

- a real robot
- a real interaction (...with a **human**!)
- a **continuous** interaction
- a realistic task (**large state vector & action space**)
- **in the wild**

LET SET OURSELVES A CHALLENGE

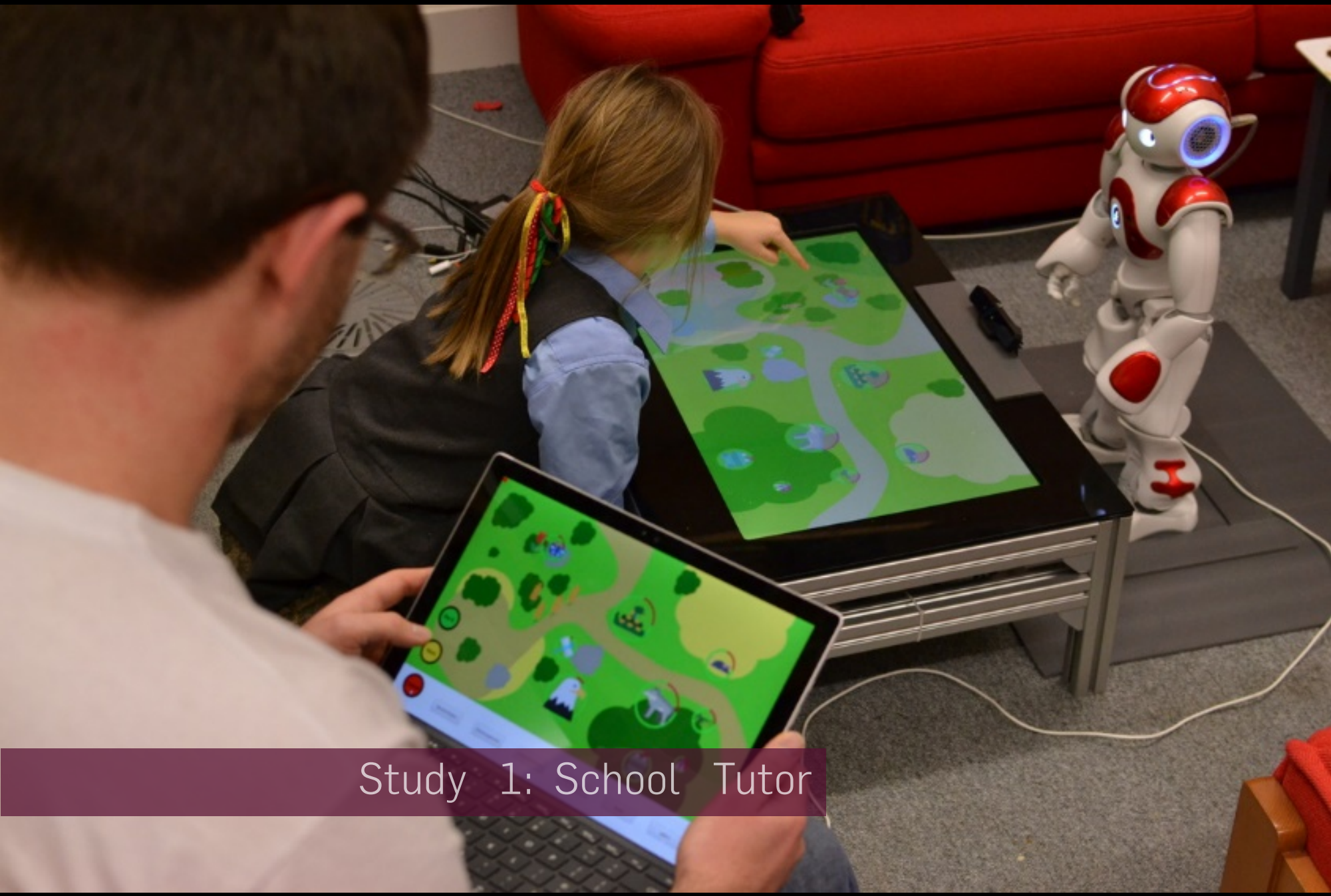
Design & run a study with:

- a real robot
- a real interaction (...with a **human**!)
- a **continuous** interaction
- a realistic task (**large state vector & action space**)
- **in the wild**
- also including **social behaviours & social dynamics**

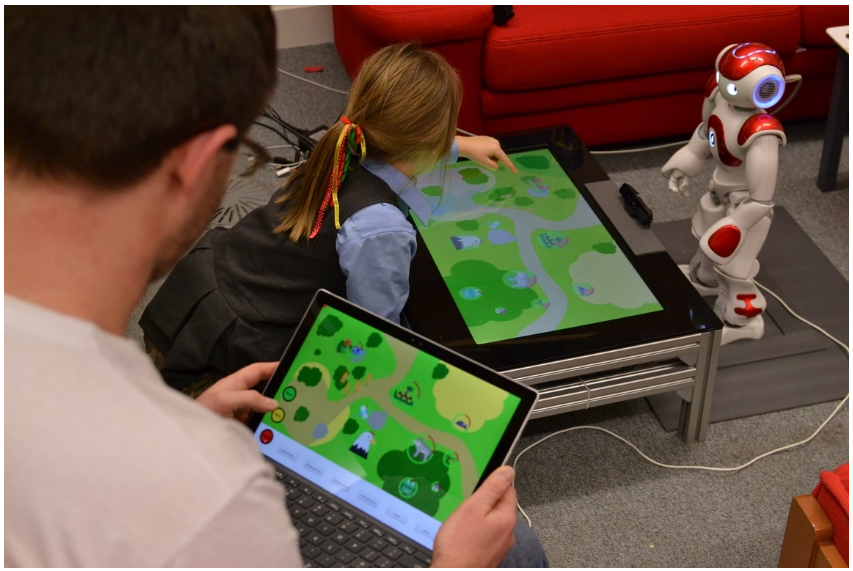
LET SET OURSELVES A CHALLENGE

Design & run a study with:

- a real robot
- a real interaction (...with a **human**!)
- a **continuous** interaction
- a realistic task (**large state vector & action space**)
- **in the wild**
- also including **social behaviours & social dynamics**
- ...and of course, the robot should be **autonomous**


Study 1: School Tutor

IRL APPLIED TO SOCIAL ROBOTICS



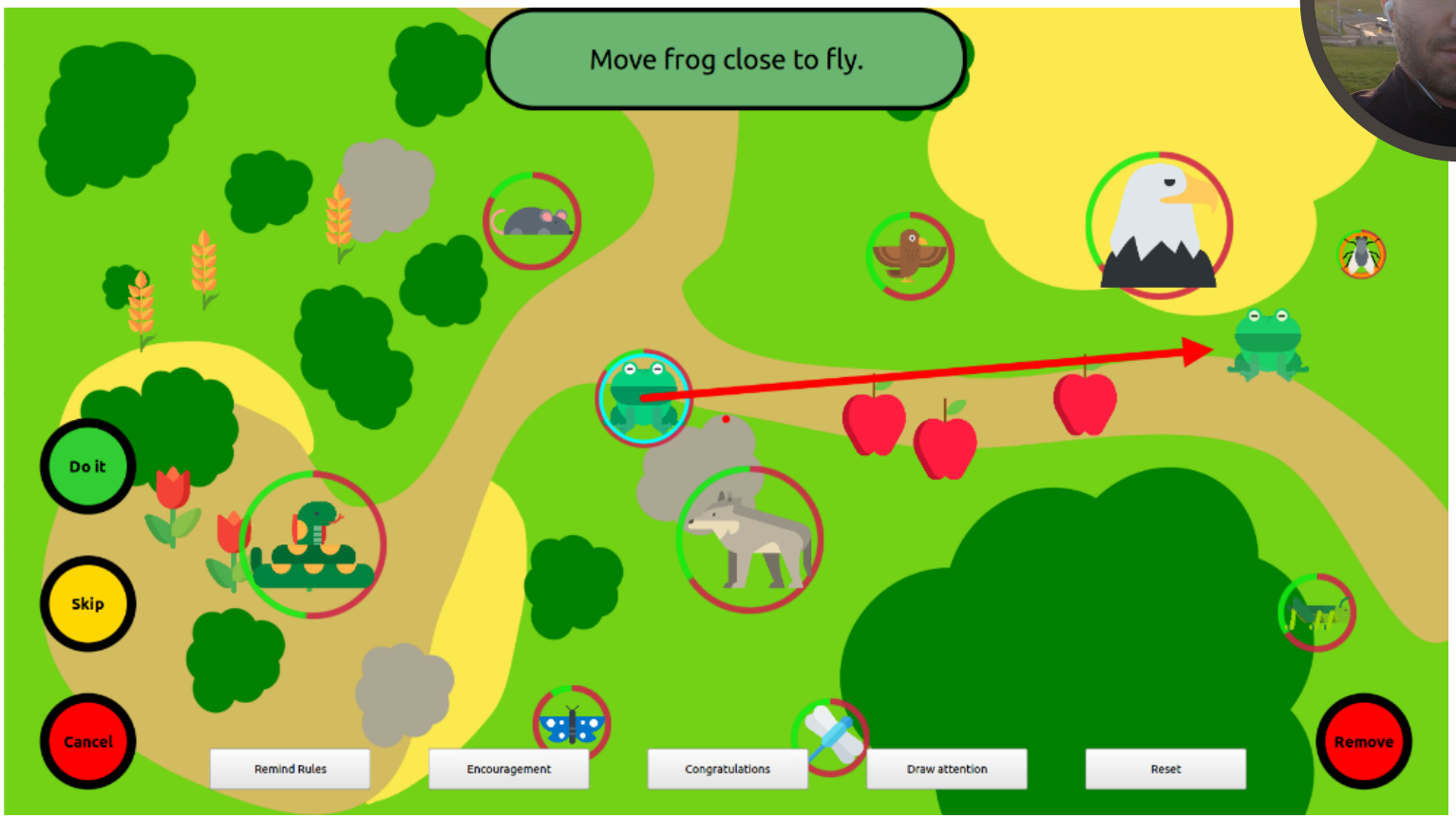
The children plays a game about food chains; the robot learns to guide them (*task-specific action policy*) and encourage them (*social action policy*)

Interactive Machine Learning (IRL) to teach the robot.

$|state| = 210$ $|action\_space| = 655$

[Senft et al. **Teaching robots social autonomy from in situ human guidance** Science Robotics 2019]

TEACHER'S INTERFACE



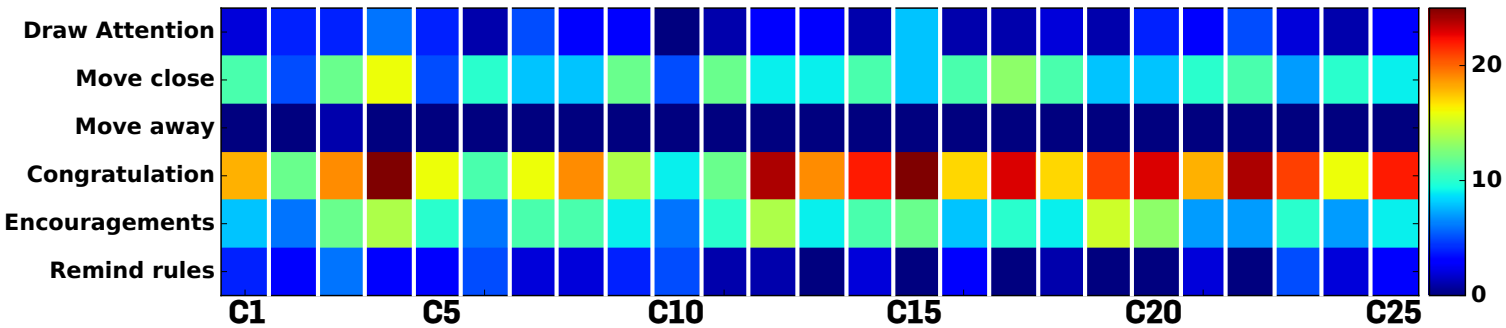
The robot's teacher (an end-user: might be the actual child's teacher) has a tablet interface that mirror's the child one, and add robot's teleoperation and rewards.

[Senft et al. **Teaching robots social autonomy from in situ human guidance** Science Robotics 2019]

LEARNT ROBOT'S BEHAVIOUR

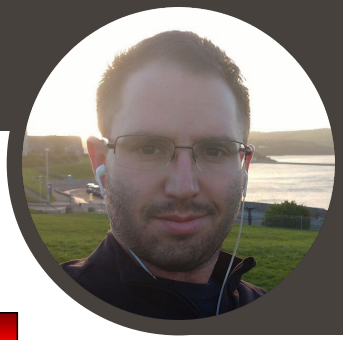
Distribution of actions for the 25 children participants:

Supervised



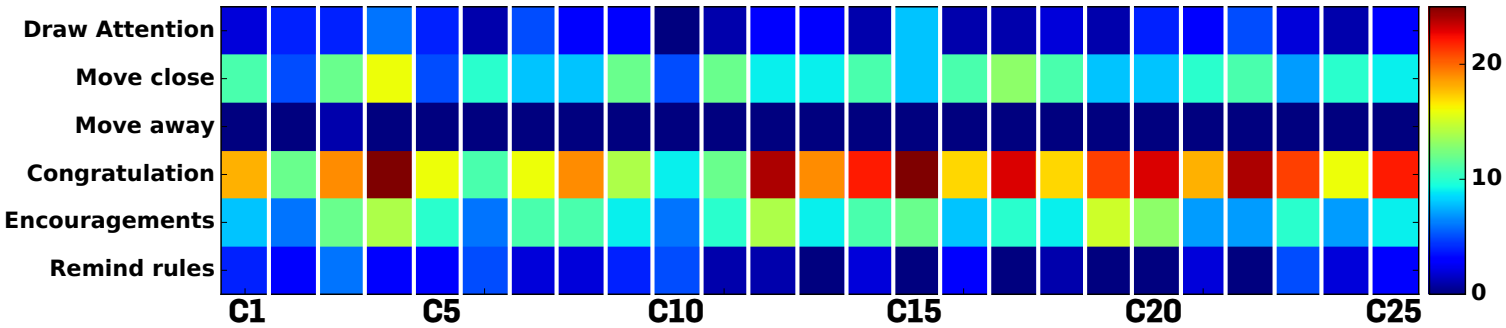
[Senft et al. **Teaching robots social autonomy from in situ human guidance** Science Robotics 2019]

LEARNT ROBOT'S BEHAVIOUR

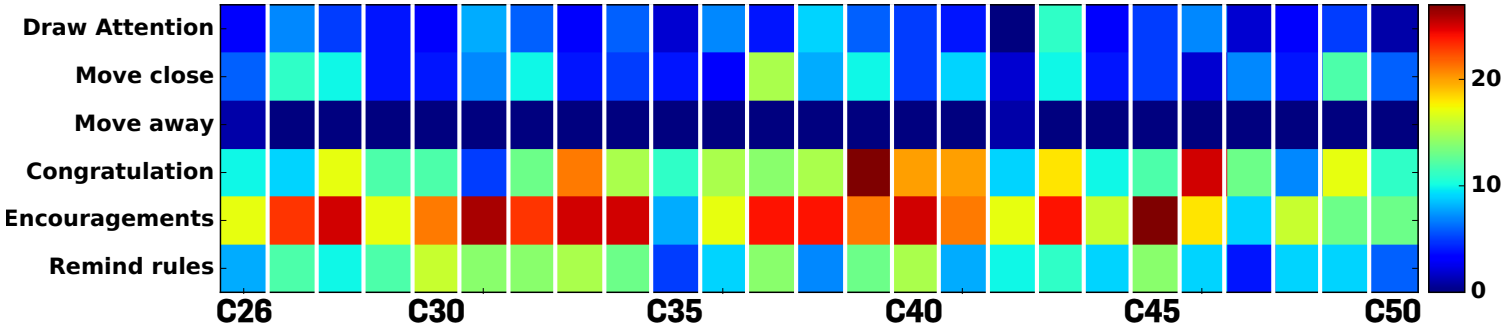


Distribution of actions for the 25 children participants:

Supervised



Autonomous

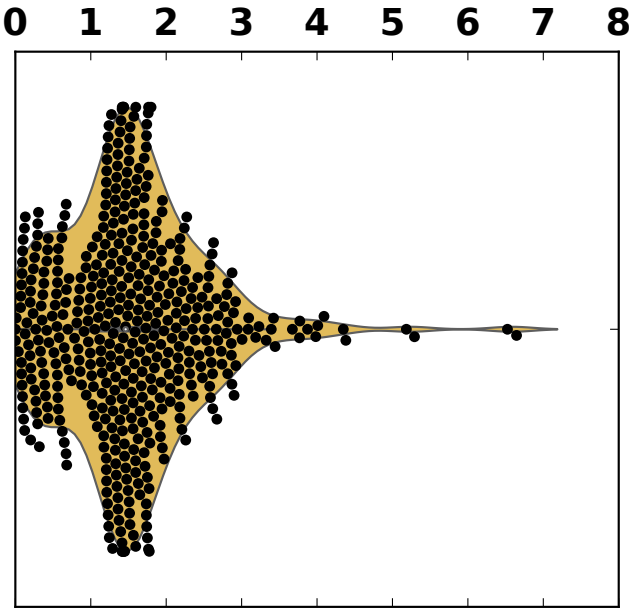


→ the robot personalises its action policies to the child's behaviour.

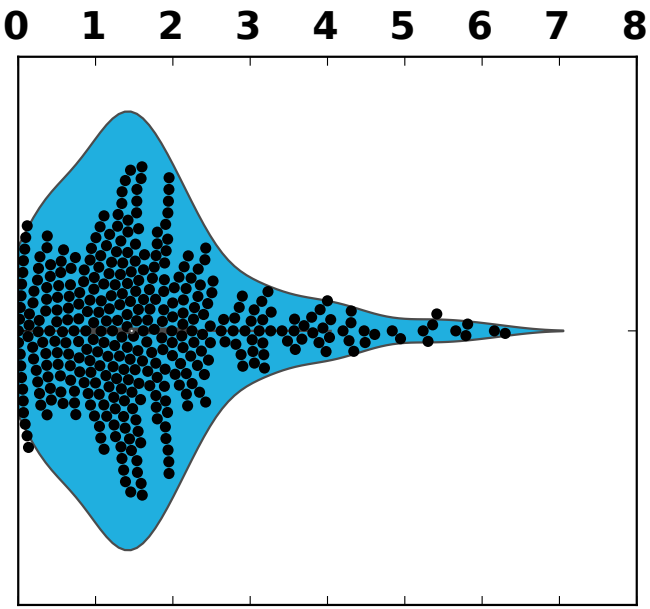
[Senft et al. **Teaching robots social autonomy from in situ human guidance** Science Robotics 2019]

LEARNT ROBOT'S BEHAVIOUR

Time between a child's successful action and a praise:



supervised

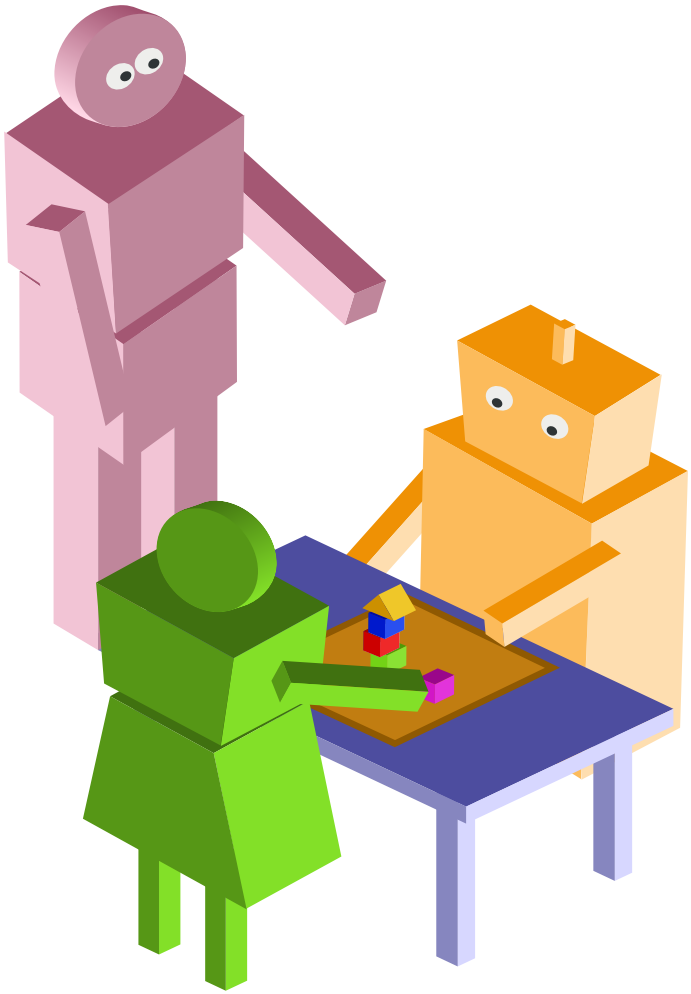


autonomous

→ the robot has also learnt an appropriate social timing.

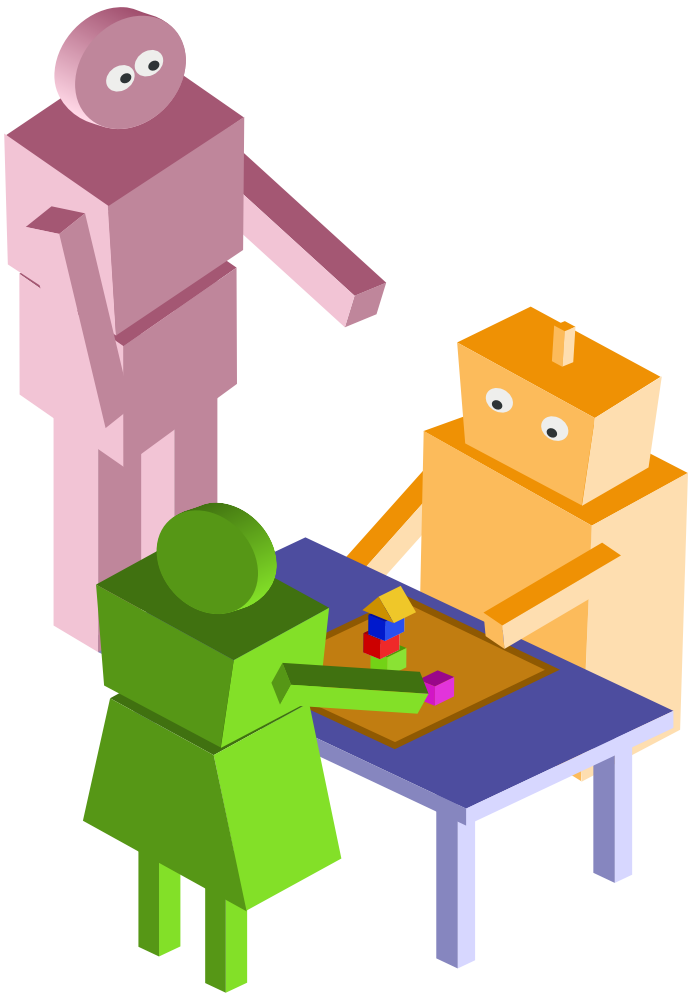
[Senft et al. **Teaching robots social autonomy from in situ human guidance** Science Robotics 2019]

WHAT DOES THAT MEANS FOR THE EXPERT/TEACHER/END-USER?



- **Progressively transferring autonomy** demonstrably works in non-trivial tutoring scenarios
- (it also learns some elements of **social behaviours** and **social timing**)

WHAT DOES THAT MEANS FOR THE EXPERT/TEACHER/END-USER?

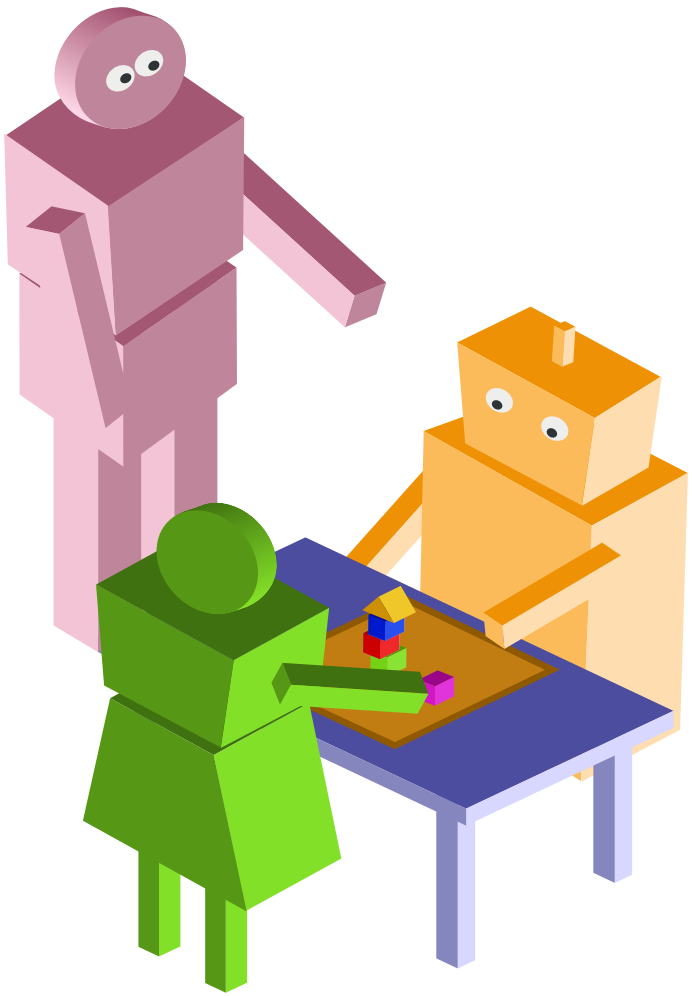


Key properties:

- **progressive autonomy** yet **transparency** of the behaviour;
- **observability** and possibility to **take over**;
- because the training takes place in-situ, the robot behaviours are **situated** and **co-constructed** by the teacher and the child

⇒ good characteristics from a ethics/responsible AI perspective

WHAT DOES THAT MEANS FOR THE EXPERT/TEACHER/END-USER?



Yet:

- Design of the input state tricky and largely task dependent;
- What about more complex social behaviours?
- Would that sustain long-term interactions?

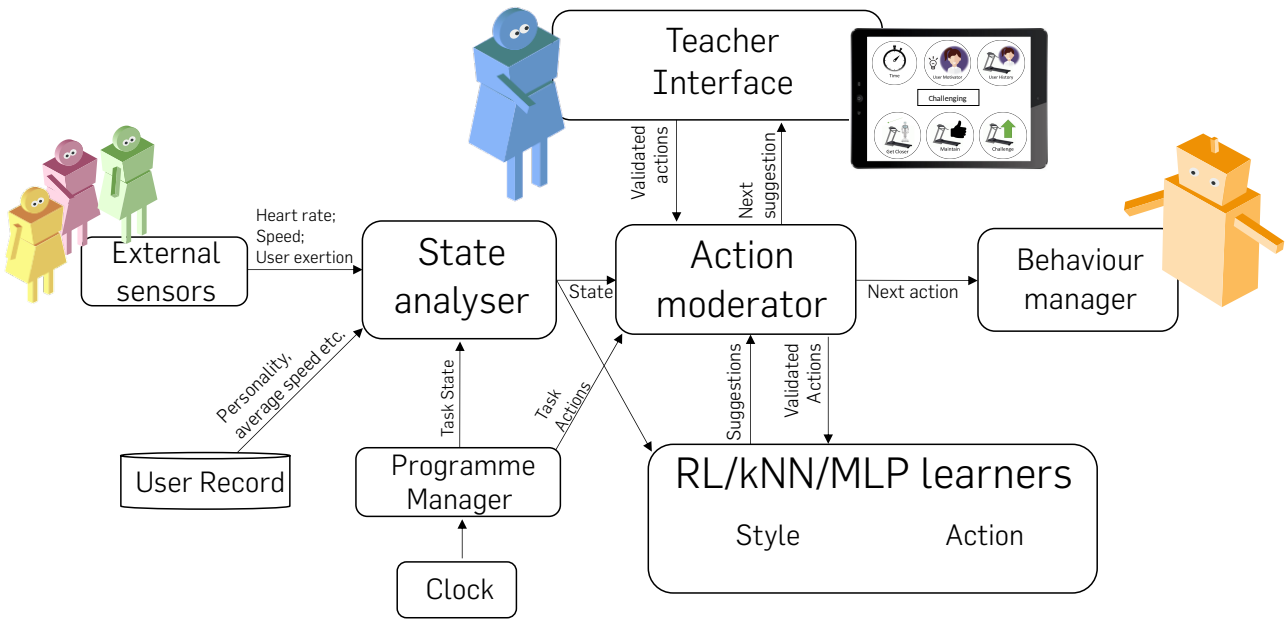
Study 2: gym coach

COUCH-TO-5K STUDY

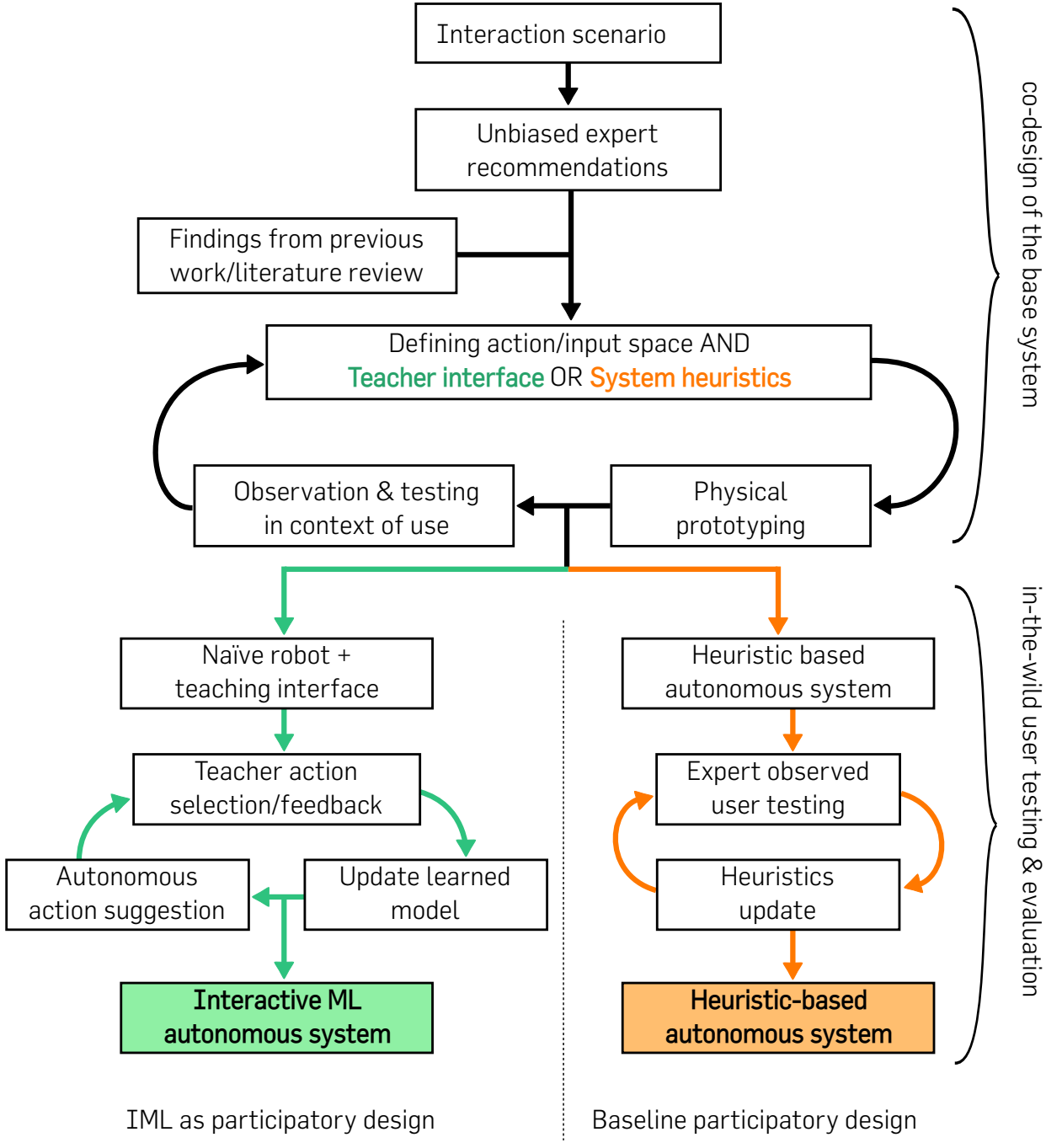
- 9 participants
- 3 months; 27 one-hour sessions per participants
- $|state| = 20$; $|action\_space| = 11$
- Includes participants' personality (Big-5) as input feature

[Winkle et al. **In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment** RSS 2020]

COUCH-TO-5K STUDY



[Winkle et al. **In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment** RSS 2020]


Interaction scenario
Unbiased expert recommendations
Findings from previous work/literature review
Defining action/input space AND
Teacher interface OR System heuristics
Observation & testing in context of use
Physical prototyping
co-design of the base system
Naïve robot + teaching interface
Teacher action selection/feedback
Autonomous action suggestion
Update learned model
Interactive ML autonomous system
Heuristic based autonomous system
Expert observed user testing
Heuristics update
Heuristic-based autonomous system
in-the-wild user testing & evaluation
IML as participatory design
Baseline participatory design

CO-DESIGN FOR REAL-WORLD + LONG-TERM

[Winkle et al. **In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment** RSS 2020]

INPUT & OUTPUT SPACES: MOSTLY TASK-SPECIFIC

Type	Feature	Values	Description
(Dynamic) Task State	Task Action Type	0, 0.5, 1	Whether participant is in warm-up, walk or run
	Session Progress	0-1	Time spent in session/session duration
	Programme Progress	0-1	Time spent on programme/programme duration
	Programme Action Progress	0-1	Time spent on current walk or run action/action duration
	Programme Action Duration	0, 0.5, 1	Current walk/run action length as $\leq$ 3 mins, $\geq$ 20 mins or other
	Time Since Last Action	0-1	Time since last action/60; capped at 1
Dynamic Performance	Relative Speed: Average	0-1	Current speed/(2 x average speed)
	Relative Speed: Best	0-1	Current speed/(2 x personal best speed)
Dynamic Engagement	Heart Rate	0-1	Heart rate/2x resting heart rate capped at 1
	Motivation/Effort	0, 0.5, 1	Self-reported measure in warmup/on check PRE action
	Facial Expression: Lip Pull*	0-1	Normalised action unit returned by OpenFace
	Facial Expression: Mouth Open*	0-1	Normalised action unit returned by OpenFace
Static Engagement	Elaboration level (self)	0-1	Normalised sum of 3 Likert questions
	Elaboration level (expert)	0-1	as above but rated by fitness instructor
	Activity Level	0-1	Likert question response
Static Personality	Extroversion	0-1	Big Five measure normalised with respect to max score
	Agreeableness	0-1	Big Five measure normalised with respect to max score
	Conscientiousness	0-1	Big Five measure normalised with respect to max score
	Emotional Stability	0-1	Big Five measure normalised with respect to max score
	Openness to Experience	0-1	Big Five measure normalised with respect to max score

Features followed by * were ultimately removed due to unreliable detection.

INPUT & OUTPUT SPACES: MOSTLY TASK-SPECIFIC

Type	Feature	Values	Description
(Dynamic) Task State	Task Action Type	0, 0.5, 1	Whether participant is in warm-up, walk or run
	Session Progress	0-1	Time spent in session/session duration
	Programme Progress	0-1	Time spent on programme/programme duration
	Programme Action Progress	0-1	Time spent on current walk or run action/action duration
	Programme Action Duration	0, 0.5, 1	Current walk/run action length as $\leq$ 3 mins, $\geq$ 20 mins or other
	Time Since Last Action	0-1	Time since last action/60; capped at 1
Dynamic Performance	Relative Speed: Average	0-1	Current speed/(2 x average speed)
	Relative Speed: Best	0-1	Current speed/(2 x personal best speed)
Dynamic Engagement	Heart Rate	0-1	Heart rate/2x resting heart rate capped at 1
	Motivation/Effort	0, 0.5, 1	Self-reported measure in warmup/on check PRE action
	Facial Expression: Lip Pull*	0-1	Normalised action unit returned by OpenFace
	Facial Expression: Mouth Open*	0-1	Normalised action unit returned by OpenFace
Static Engagement	**Elaboration level (self)**	0-1	Normalised sum of 3 Likert questions
	Elaboration level (expert)	0-1	as above but rated by fitness instructor
	Activity Level	0-1	Likert question response
Static Personality	**Extroversion**	0-1	Big Five measure normalised with respect to max score
	Agreeableness	0-1	Big Five measure normalised with respect to max score
	Conscientiousness	0-1	Big Five measure normalised with respect to max score
	Emotional Stability	0-1	Big Five measure normalised with respect to max score
	Openness to Experience	0-1	Big Five measure normalised with respect to max score

Features followed by * were ultimately removed due to unreliable detection. **Non-task specific features.**

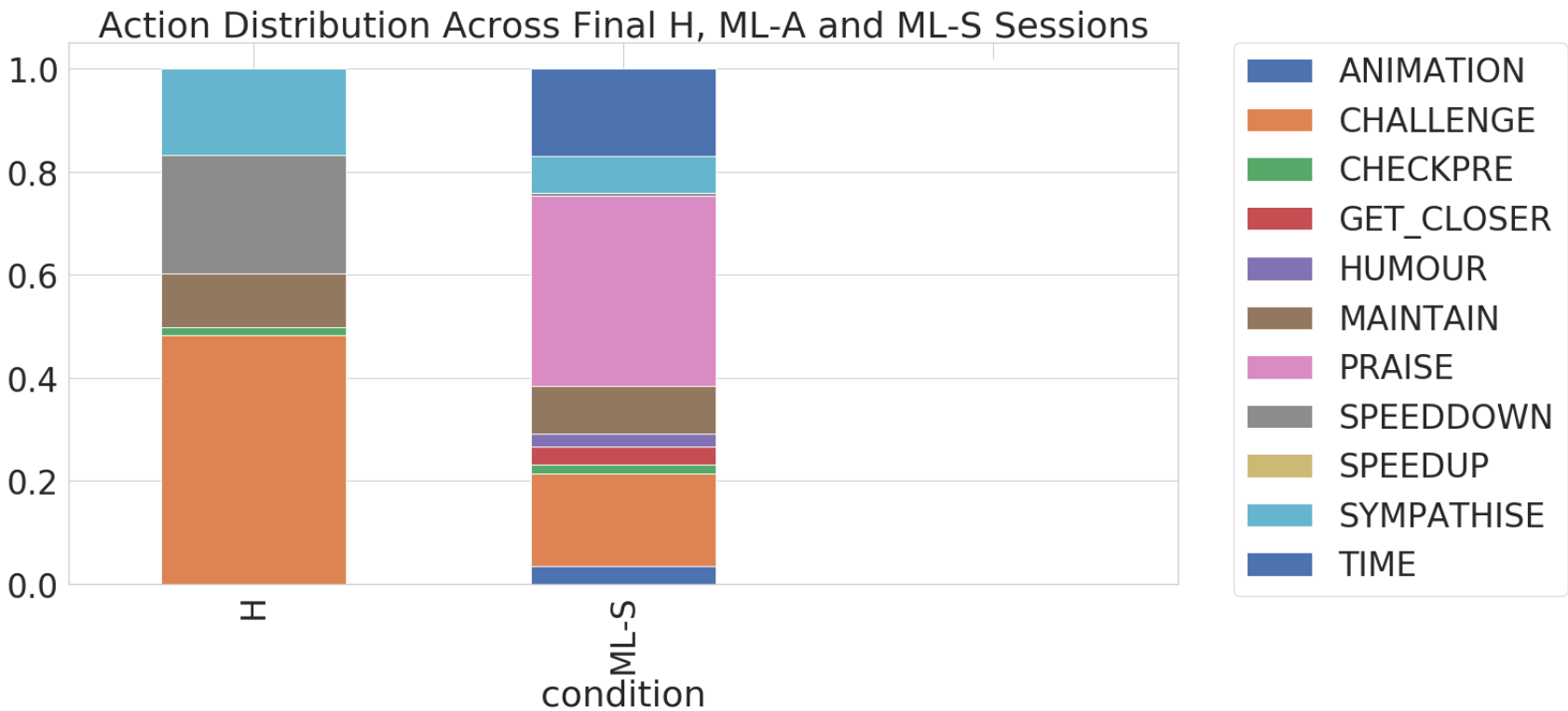
INPUT & OUTPUT SPACES: MOSTLY TASK-SPECIFIC

Full listing of actions as *{action-type, style-modifier}*:

	Social-Supporting Actions								Task Actions		Low Level
	Time	**Social**	**Performance**	**Reward**	**Check User**	**Animation**	**Get Closer**	**Back Off**	**Run**	**Walk**	**Eye Colour**
P	Time	Humour	Maintain	Praise	-	Animation	-	-	Run	Walk	Green
C	Time	Challenge	Speed Up	-	-	-	-	-	Run	Walk	Yellow
S	Time	Challenge Sympathise	Speed Down	Praise	Check PRE	-	-	-	Run	Walk	Blue
N	-	-	-	-	-	-	Get Closer	Back Off	Run	Walk	White

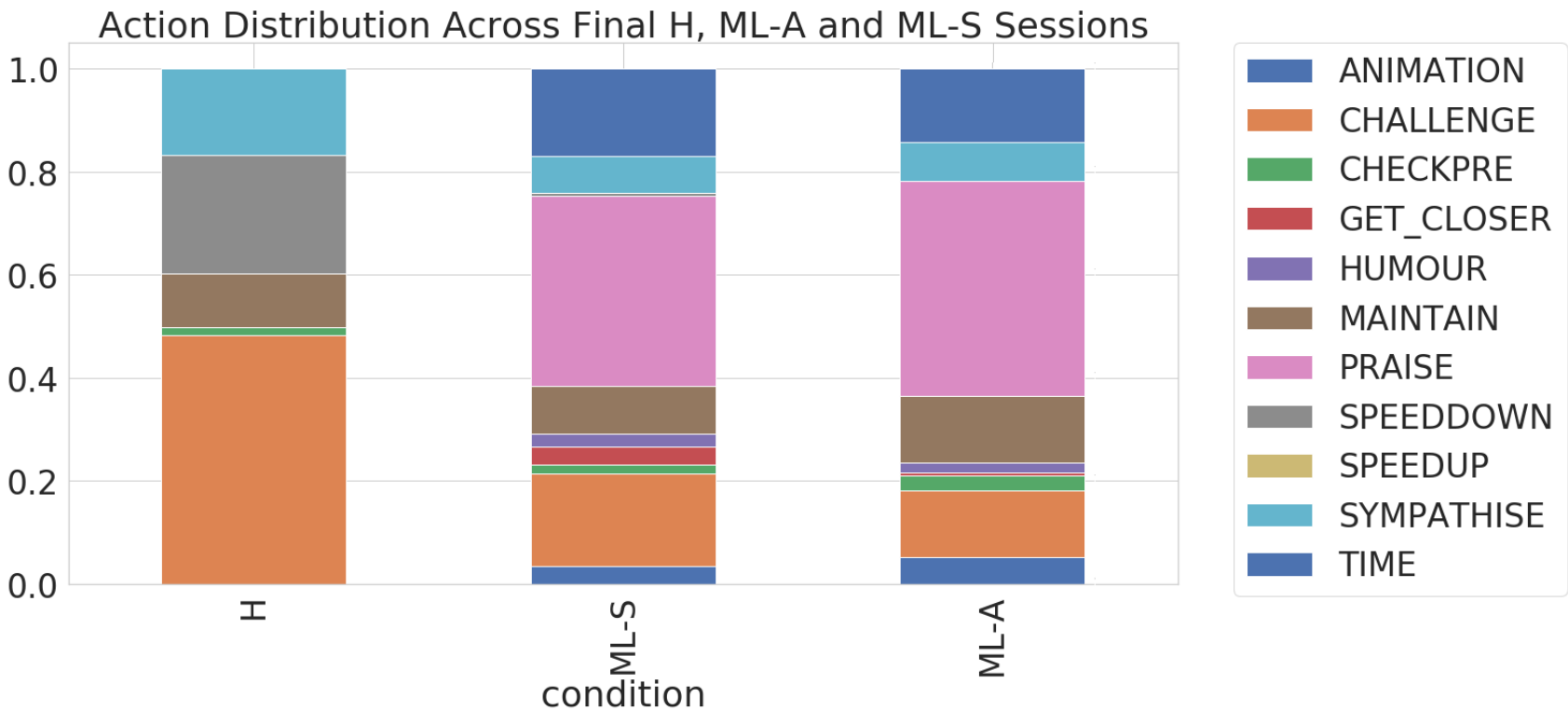
Style modifiers: P = Positive; C = Challenging; S = Sympathetic; N = Neutral.

LEARNT POLICIES



[Winkle et al. **In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment** RSS 2020]

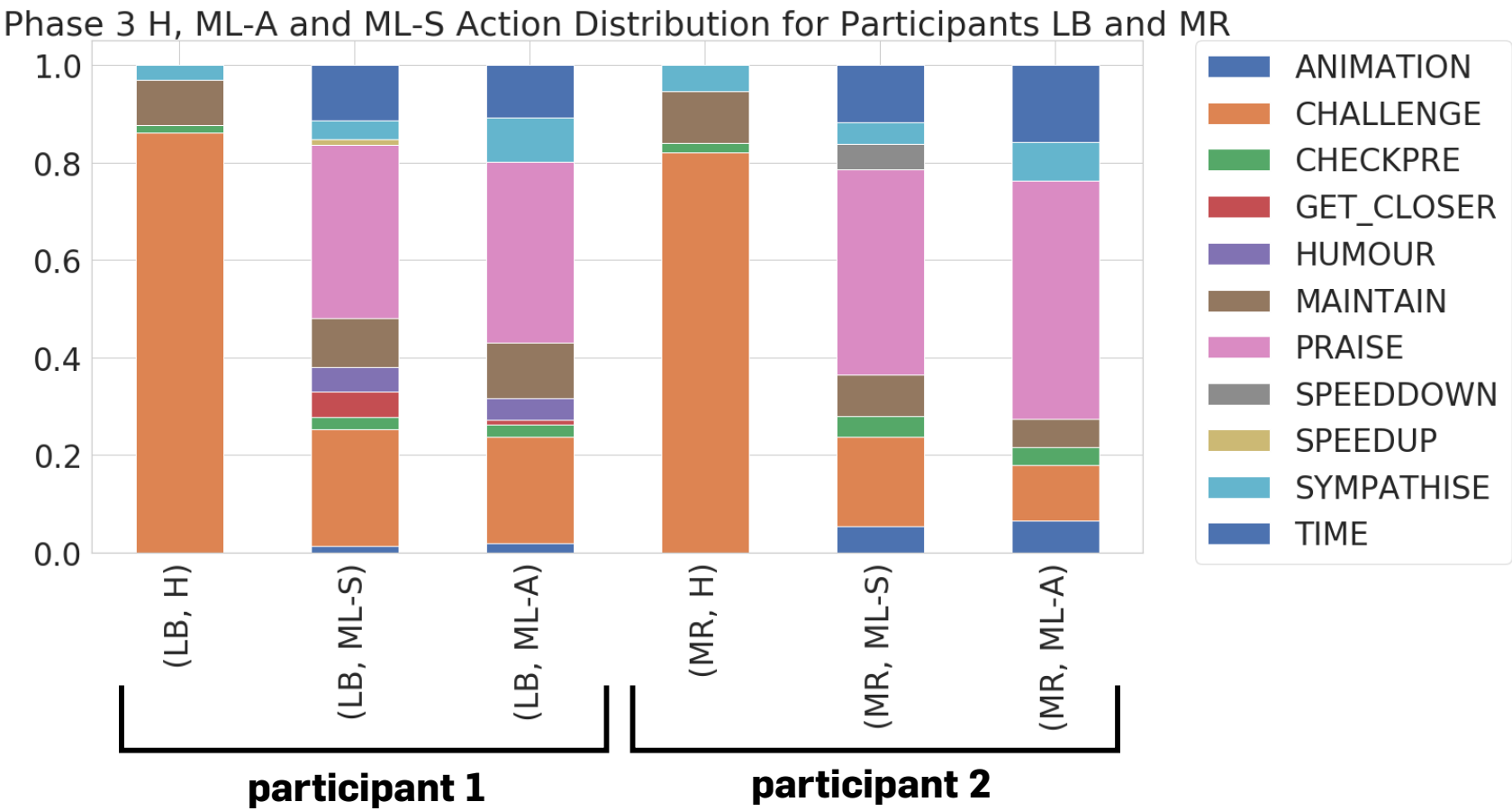
LEARNT POLICIES



(another nail in the expert systems coffin!)

[Winkle et al. **In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment** RSS 2020]

LEARNT POLICIES



[Winkle et al. **In-Situ Learning from a Domain Expert for Real World Socially Assistive Robot Deployment** RSS 2020]

LEADOR: END-TO-END PARTICIPATORY METHODOLOGY

In retrospect:

- a successful technical solution to replace the wizard;
- but equally important, a **end-to-end** participatory design methodology

[Senft at al. **SPARC: Supervised progressively autonomous robot competencies**, ICSR 2015;
Winkle, Senft, Lemaignan **LEADOR: End-to-End Participatory Design of Autonomous Social Robots** FrontiersIn AI & Robotics 2021]

LEADOR: END-TO-END PARTICIPATORY METHODOLOGY

In retrospect:

- a successful technical solution to replace the wizard;
- but equally important, a **end-to-end** participatory design methodology

From a technique:

SPARC: Supervised progressively autonomous robot competencies

to a methodology:

LEADOR: Led-by-Experts Automation and Design Of Robots

[Senft at al. **SPARC: Supervised progressively autonomous robot competencies**, ICSR 2015;
Winkle, Senft, Lemaignan **LEADOR: End-to-End Participatory Design of Autonomous Social Robots** FrontiersIn AI & Robotics 2021]

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

- ? open-ended, underspecified situations

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

- ? open-ended, underspecified situations
- ? rich semantics

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

- ? open-ended, underspecified situations
- ? rich semantics
- ? complex social dynamics (incl. groups)

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

- ? open-ended, underspecified situations
- ? rich semantics
- ? complex social dynamics (incl. groups)
- ? close the "interaction loop"

LET'S REVISIT OUR CHALLENGE

Studies with...

- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

- ? open-ended, underspecified situations
- ? rich semantics
- ? complex social dynamics (incl. groups)
- ? close the "interaction loop"
- ~ sustain long-term autonomous social interactions

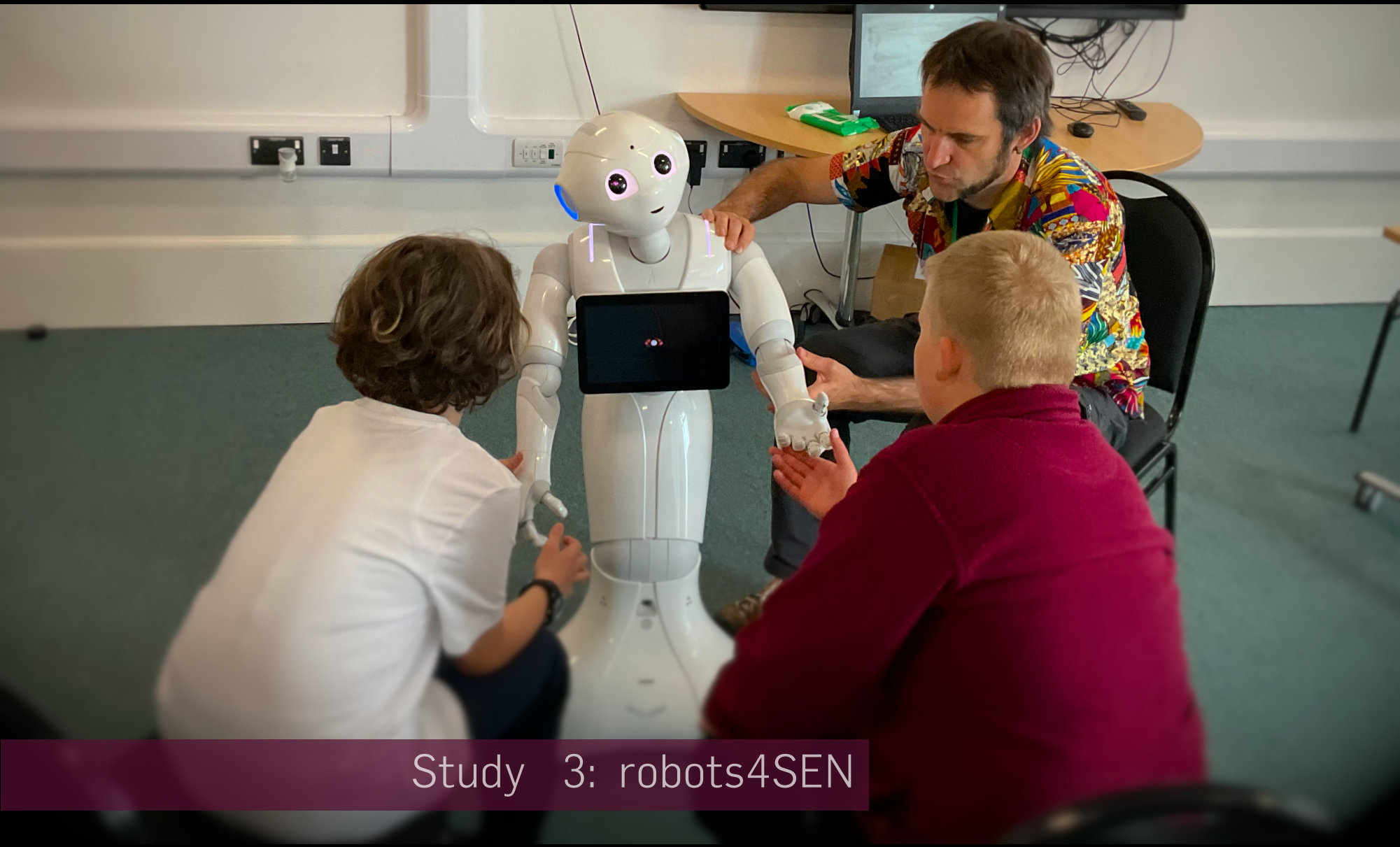
LET'S REVISIT OUR CHALLENGE

Studies with...

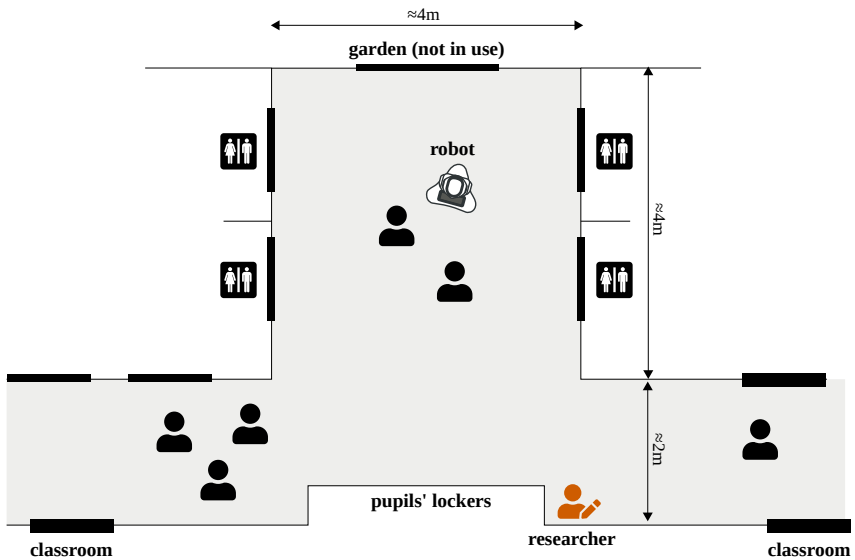
- ✓ real robots
- ✓ real interactions with humans
- ✓ continuous interactions
- ✓ realistic tasks (large state vector & action space)
- ✓ in the wild
- ✓ including social behaviours & social dynamics
- ✓ adaptive autonomous robot

End of story? ...not quite

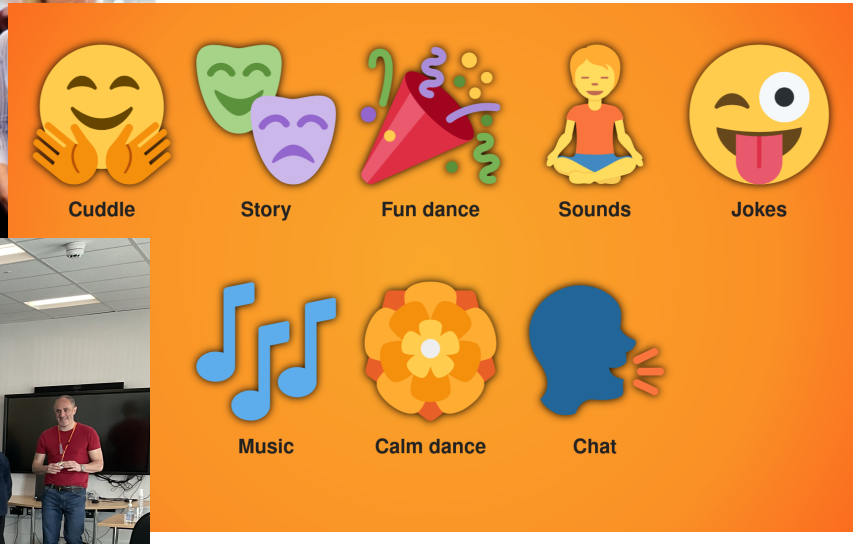
- ⓘ open-ended, underspecified situations
- ⓘ rich semantics
- ⓘ complex social dynamics (incl. groups)
- ⓘ close the "interaction loop"
- ∼ sustain long-term autonomous social interactions
- ∼ real-world robustness


Study 3: robots4SEN

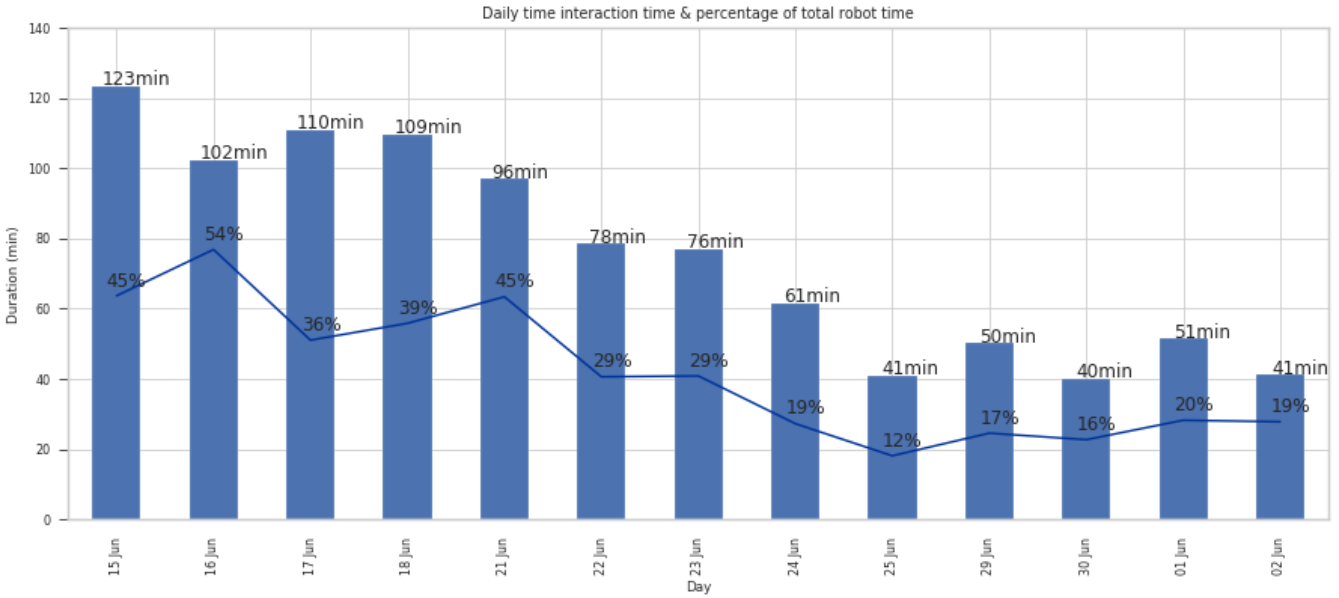
WELL-BEING AND AUTISM

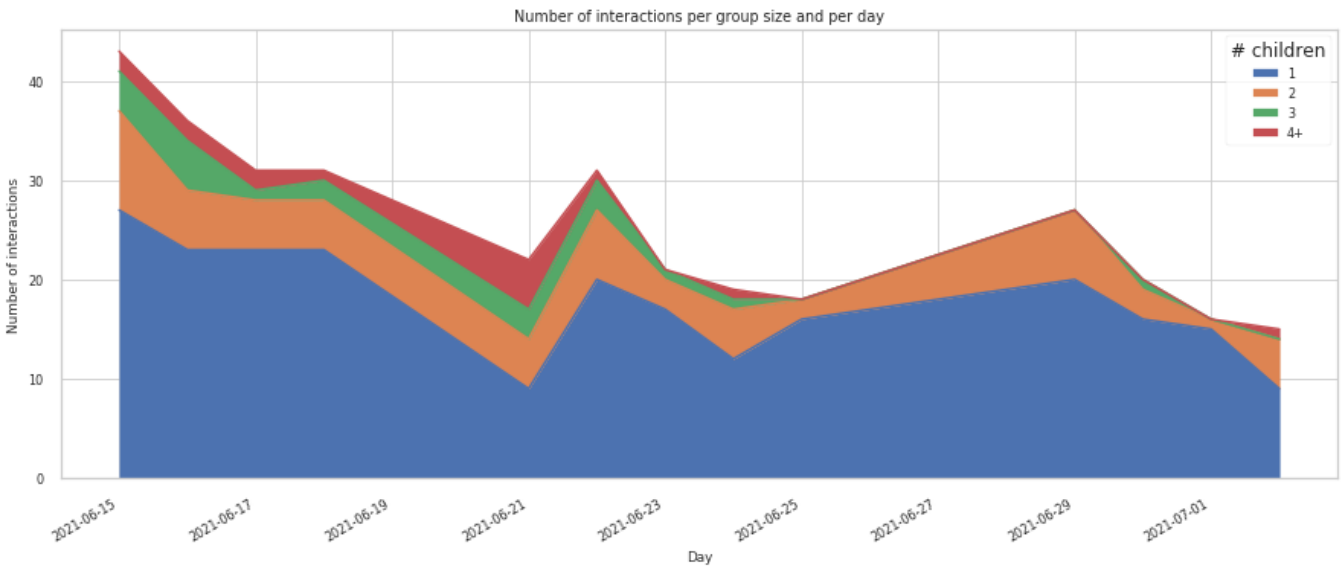


PARTICIPATORY DESIGN WITH AUTISTIC CHILDREN



ENGAGEMENT OVER 3 WEEKS





SOCIAL COMPANIONS TO SUPPORT CHILDREN'S WELL-BEING

- open-ended, underspecified situations
- rich semantics
- complex social dynamics (incl. groups)
- close the "interaction loop"
- sustain long-term autonomous social interactions
- real-world robustness

SOCIAL COMPANIONS TO SUPPORT CHILDREN'S WELL-BEING

- open-ended, underspecified situations
- rich semantics
- complex social dynamics (incl. groups)
- close the "interaction loop"
- sustain long-term autonomous social interactions
- real-world robustness

But limited autonomy! Interactions child-led!

SOCIAL COMPANIONS TO SUPPORT CHILDREN'S WELL-BEING

- open-ended, underspecified situations
- rich semantics
- complex social dynamics (incl. groups)
- close the "interaction loop"
- sustain long-term autonomous social interactions
- real-world robustness

But limited autonomy! Interactions child-led!

$\Rightarrow$ The next frontier

…and now for something more technical

(time permitting…?)

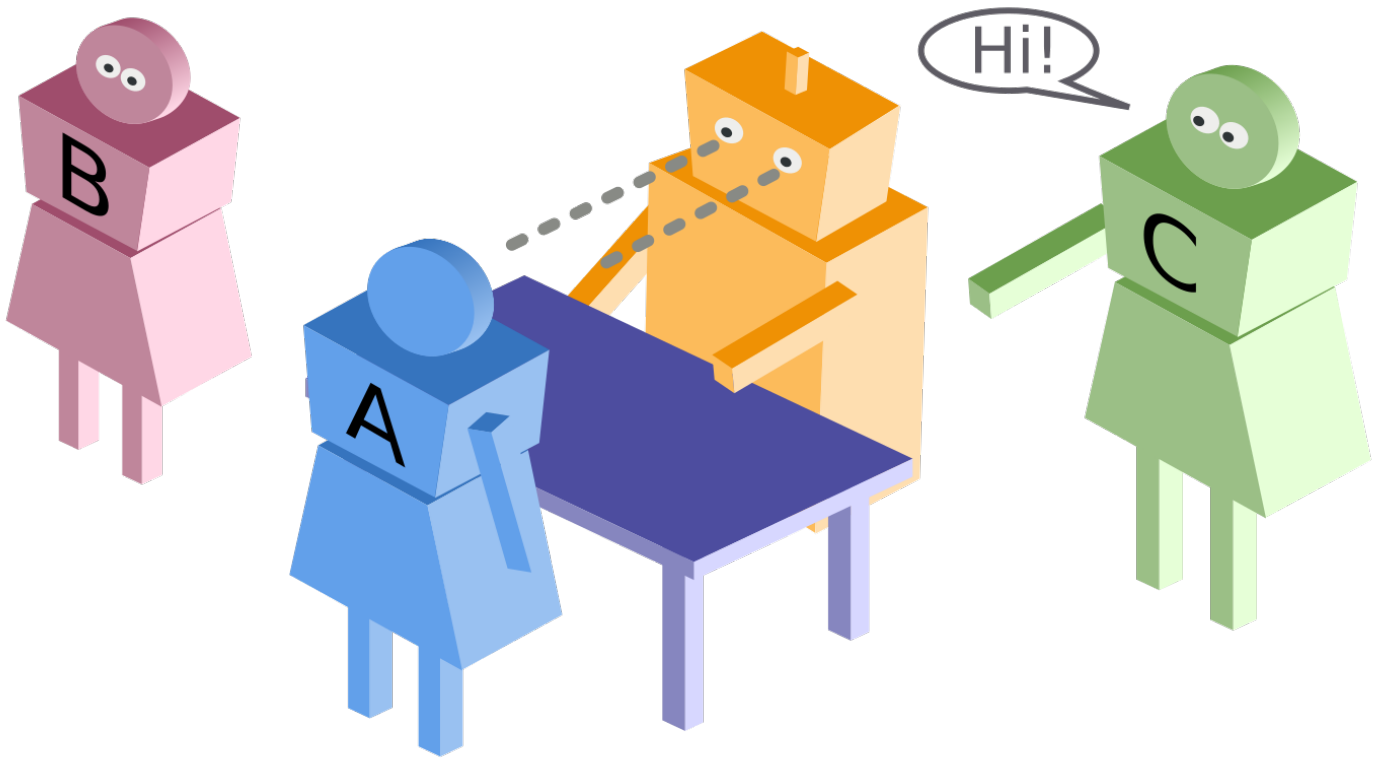
ROS4HRI: BRIEF INTRODUCTION

WHY ROS4HRI?

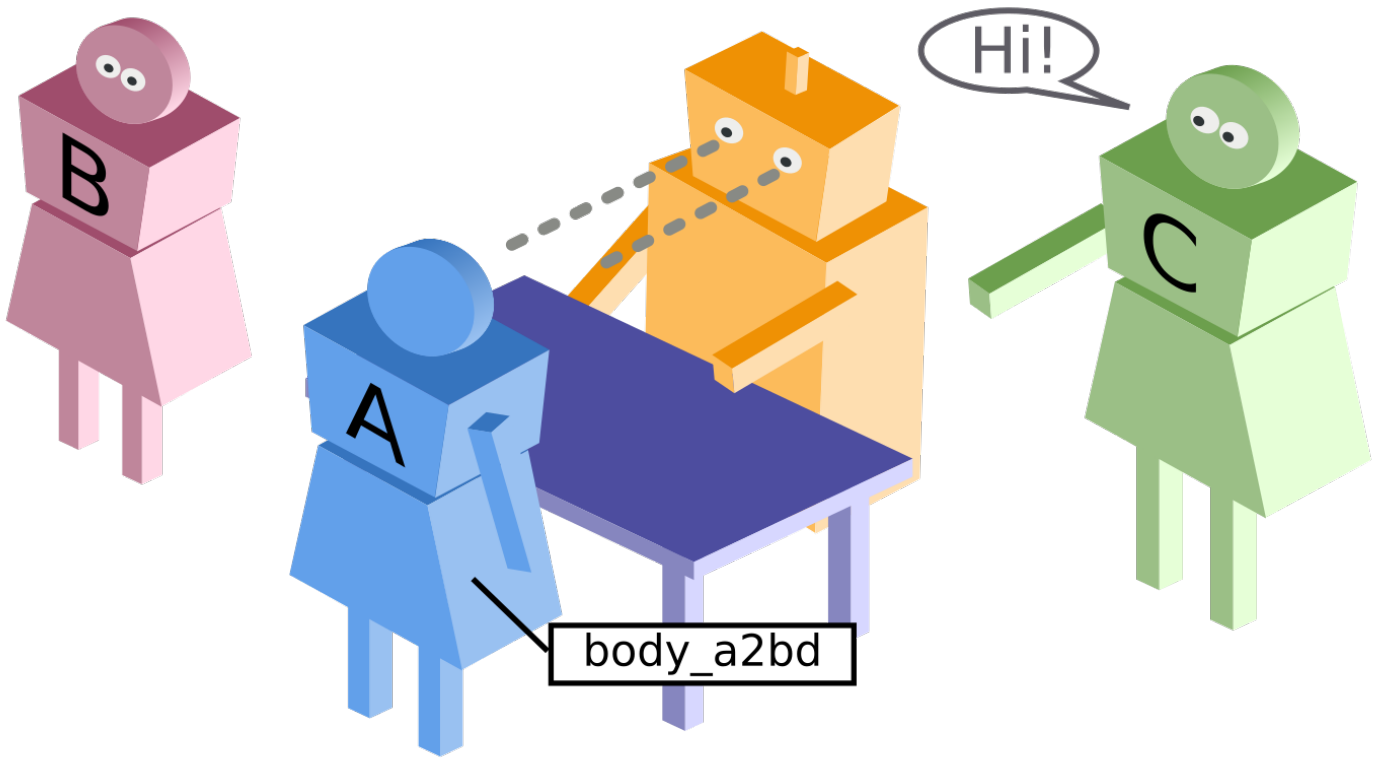
- dealing with humans is actually hard: they keep on disappearing/reappearing; hard to predict where/when; ‘shape’ known at run-time only, etc.
- widely different requirements depending on application: from ‘2D points’ to full online kinematic model.
- no ROS standard for HRI (nothing, nada, rien!)

[Mohamed, Lemaignan, **ROS for Human-Robot Interaction**, IROS 2021]

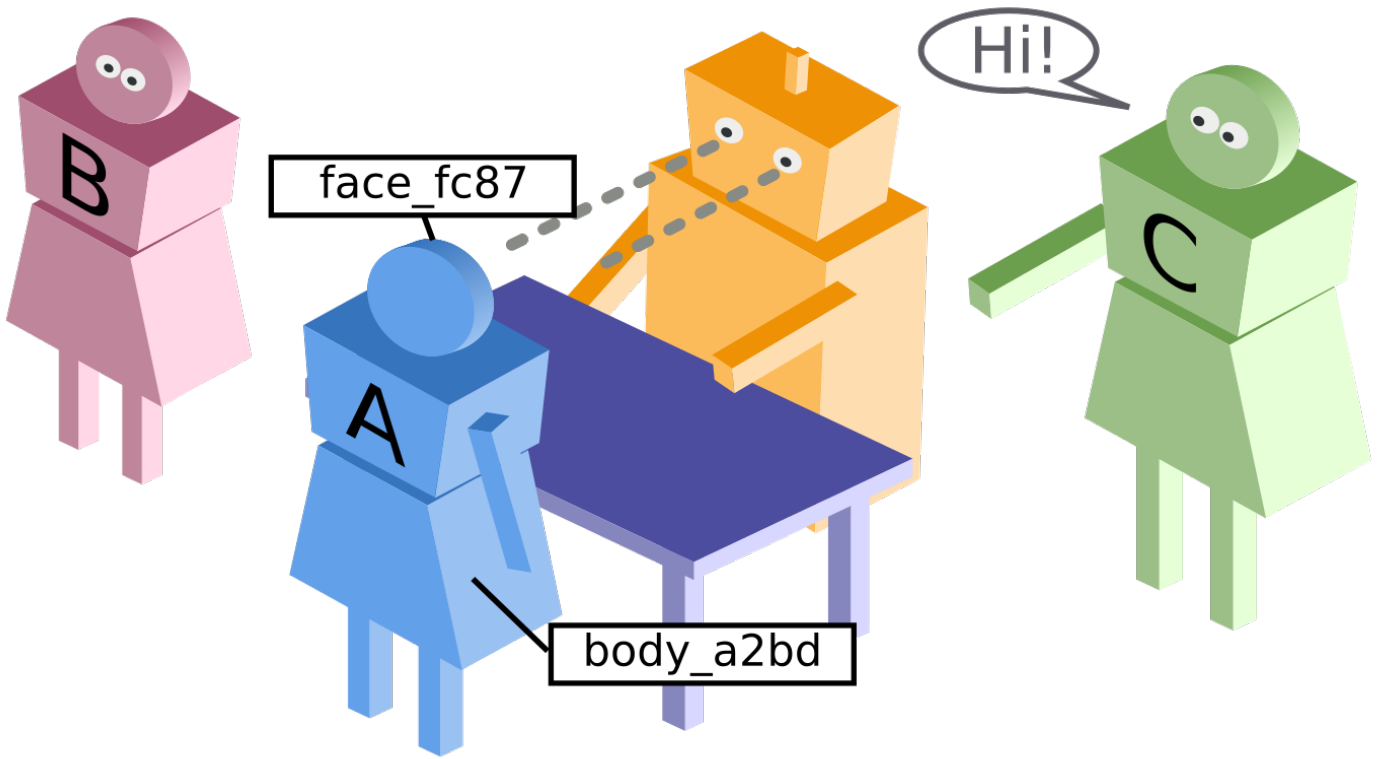
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



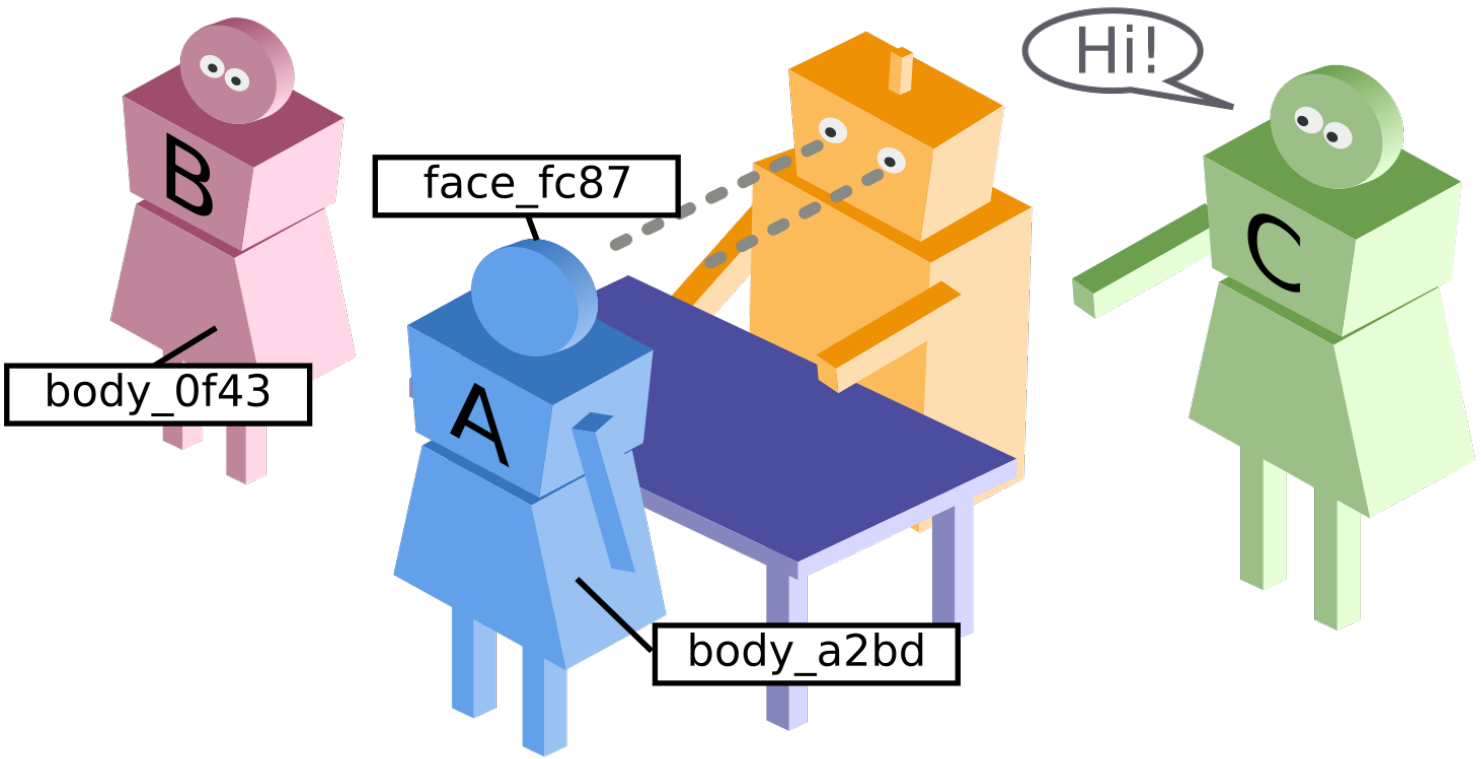
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



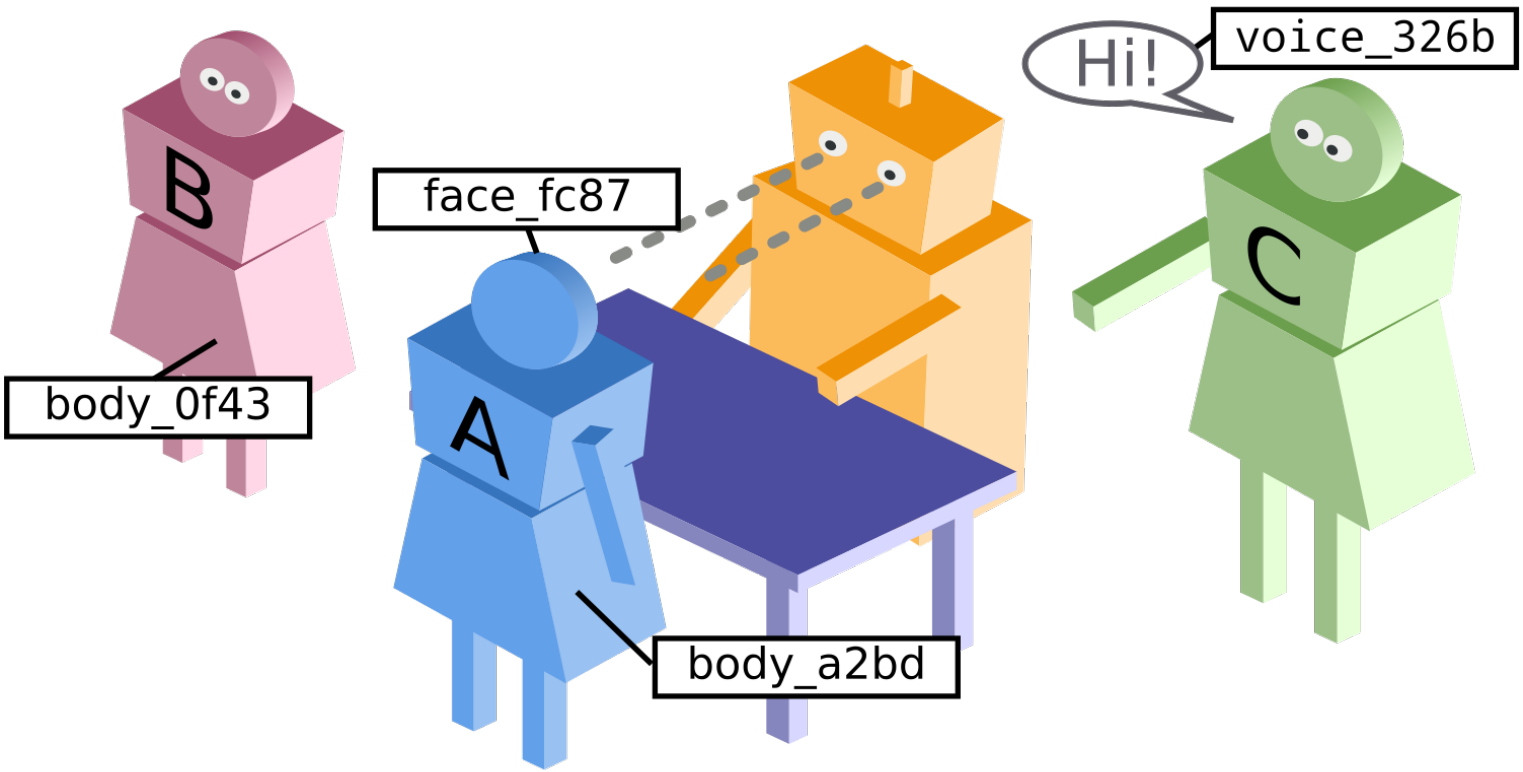
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



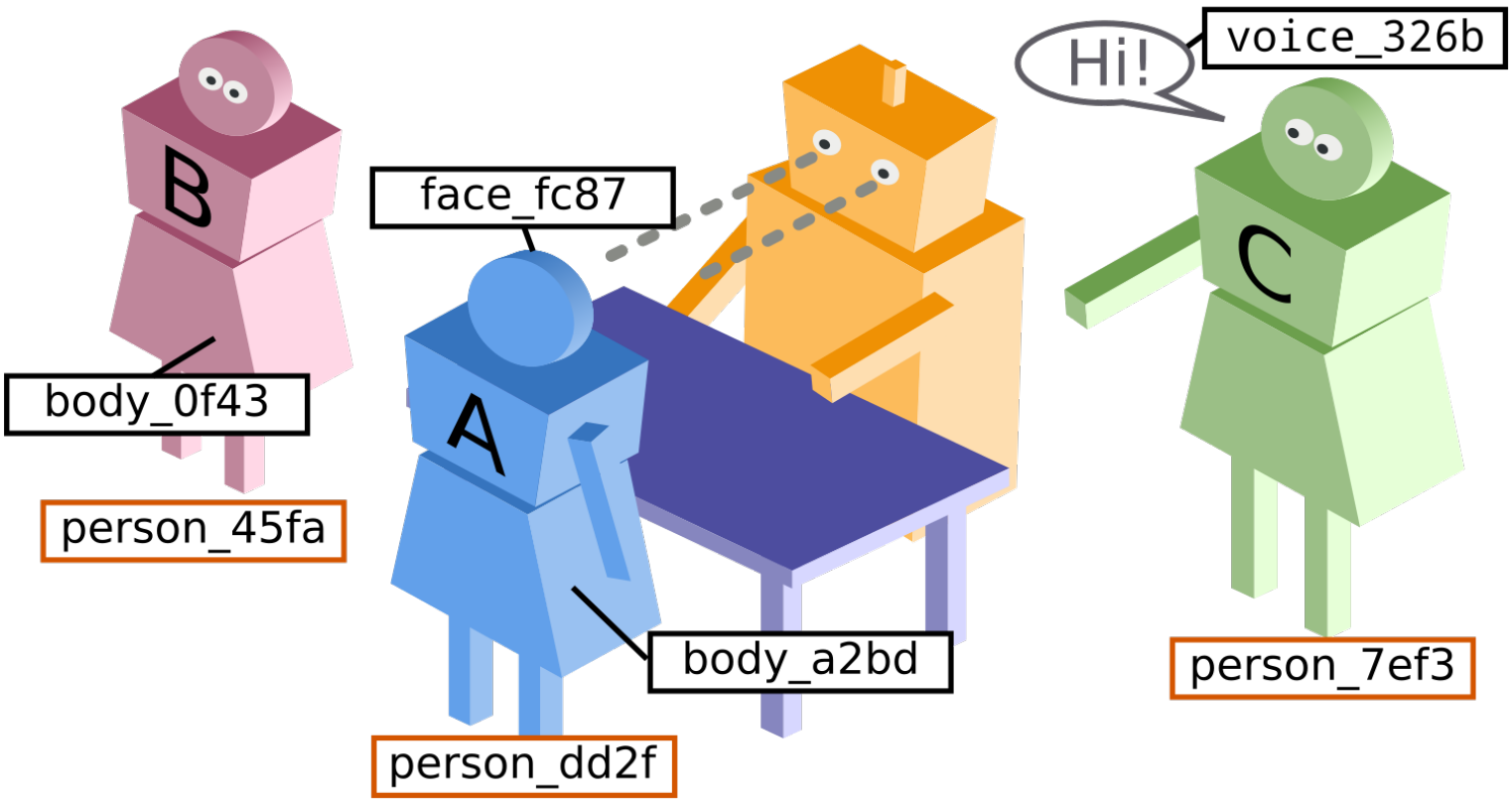
HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



HUMAN REPRESENTATION: PERMANENT VS TRANSIENT IDS



SEMANTIC OF IDS

Face ID	Body ID	Person ID	Interpretation
`#24ac`	∅	∅	Face detected - random id `#24ac` assigned - corresponding TF frame `/face_24ac` is published.
`#d73b`	∅	∅	Face detected (possibly a re-detection of a previous one) - random id `#d73b` assigned + frame published.
∅	`#37ef`	∅	Skeleton detected - id `#37ef` assigned + frame `/body_37ef` published.
`#d73b`	`#37ef`	∅	Face/body matcher merged the face and the skeleton into a single agent. The 2 previous records do not exist any more independently.
∅	∅	`#9d8a`	Human `#9d8a` is known, but not associated to any face or body. Note that TF frame `/person_9d8a` might nevertheless exist (for instance, last known position of the human).
`#d73b`	∅	`#9d8a`	Face `#d73b` is associated to human `#9d8a`. Typical result of successful face recognition.
`#96f1`	∅	`#9d8a`	Human `#9d8a` is now associated to face `#96f1`: this new association might come from the face tracker losing track of a previous face, thus re-assigning a different id to the face. The newly assigned face is however recognized by the face recognition module as being human `#9d8a`.
`#96f1`	`#37ef`	`#9d8a`	The human `#9d8a` is fully tracked: both the head and the body are detected.
∅	`#37ef`	`#9d8a`	Only the body of human `#9d8a` is tracked: this situation typically occurs if the face trackers loses the face after facial recognition successfully identified the human, and while the skeleton tracker still tracks the body.
∅	∅	`#b3da`	`#b3da` is another human, considered by the robot as different to `#9d8a`.

TOPICS STRUCTURE: FACES

Under **`/humans/faces/<faceID>/`** (eg `/humans/faces/bf3d`):

Name	Message type	Description
`/roi`	`hri_msgs/RegionOfInterestStamped`	Region of the face in the source image
`/landmarks`	`hri_msgs/FacialLandmarks`	The 2D facial landmarks extracted from the face
`/facs`	`hri_msgs/FacialActionUnits`	The presence and intensity of facial action units found in the face
`/expression`	`hri_msgs/Expression`	The expression recognised from the face

TOPICS STRUCTURE: BODIES

Under **`/humans/bodies/<bodyID>/`** (eg `/humans/bodies/5e4d`):

Name	Message type	Description
`/roi`	`hri_msgs/RegionOfInterestStamped`	Region of the whole body in the source image
`/skeleton2d`	`hri_msgs/Skeleton2D`	The 2D points of the detected skeleton
`/attitude`	`hri_msgs/BodyAttitude`	Recognised body attitude or gesture

TOPICS STRUCTURE: VOICES

Under **/humans/voices/<voiceID>/**
(eg /humans/voices/dde2):

Name	Message type	Description
/audio	audio_msgs/AudioData	Separated audio stream for this voice
/features	hri_msgs/AudioFeatures	INTERSPEECH'09 Emotion challenge low-level audio features
/is_speaking	std_msgs/Bool	Whether or not speech is recognised from this voice
/speech	std_msgs/String	The live stream of speech recognized via an ASR engine

TOPICS STRUCTURE: PERSONS

Under **/humans/persons/<personID>/**
(eg /humans/persons/45ff):

Name	Message type	Description
/face_id	std_msgs/String (latched)	Face matched to that person (if any)
/body_id	std_msgs/String (latched)	Body matched to that person (if any)
/voice_id	std_msgs/String (latched)	Voice matched to that person (if any)
/location_confidence	std_msgs/Float32	Location confidence; 1 means 'person currently seen', 0 means 'person location unknown'
/softbiometrics	hri_msgs/SoftBiometrics	Soft biometrics like age and gender of the person
/name	std_msgs/String	Name, if known
/native_language	std_msgs/String	IETF language codes like EN_gb, if known

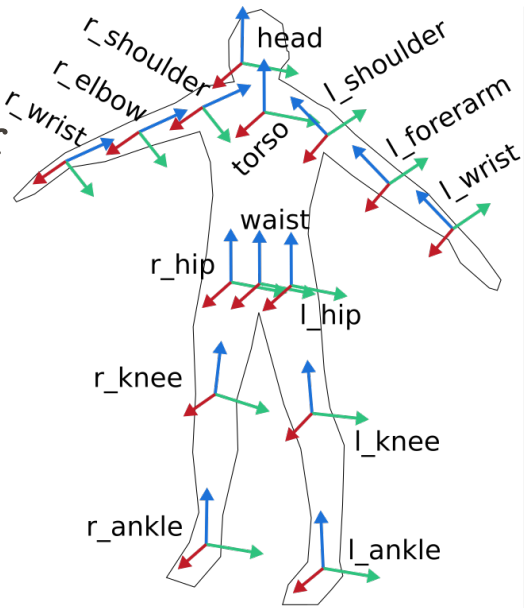
TOPICS STRUCTURE: INTERACTIONS

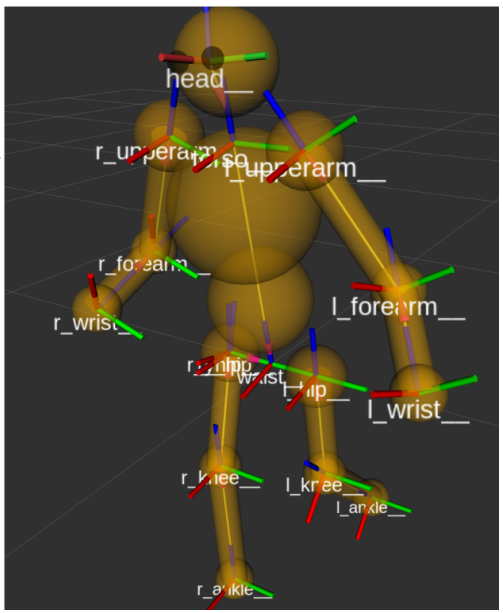
Under **`/humans/interactions/`**:

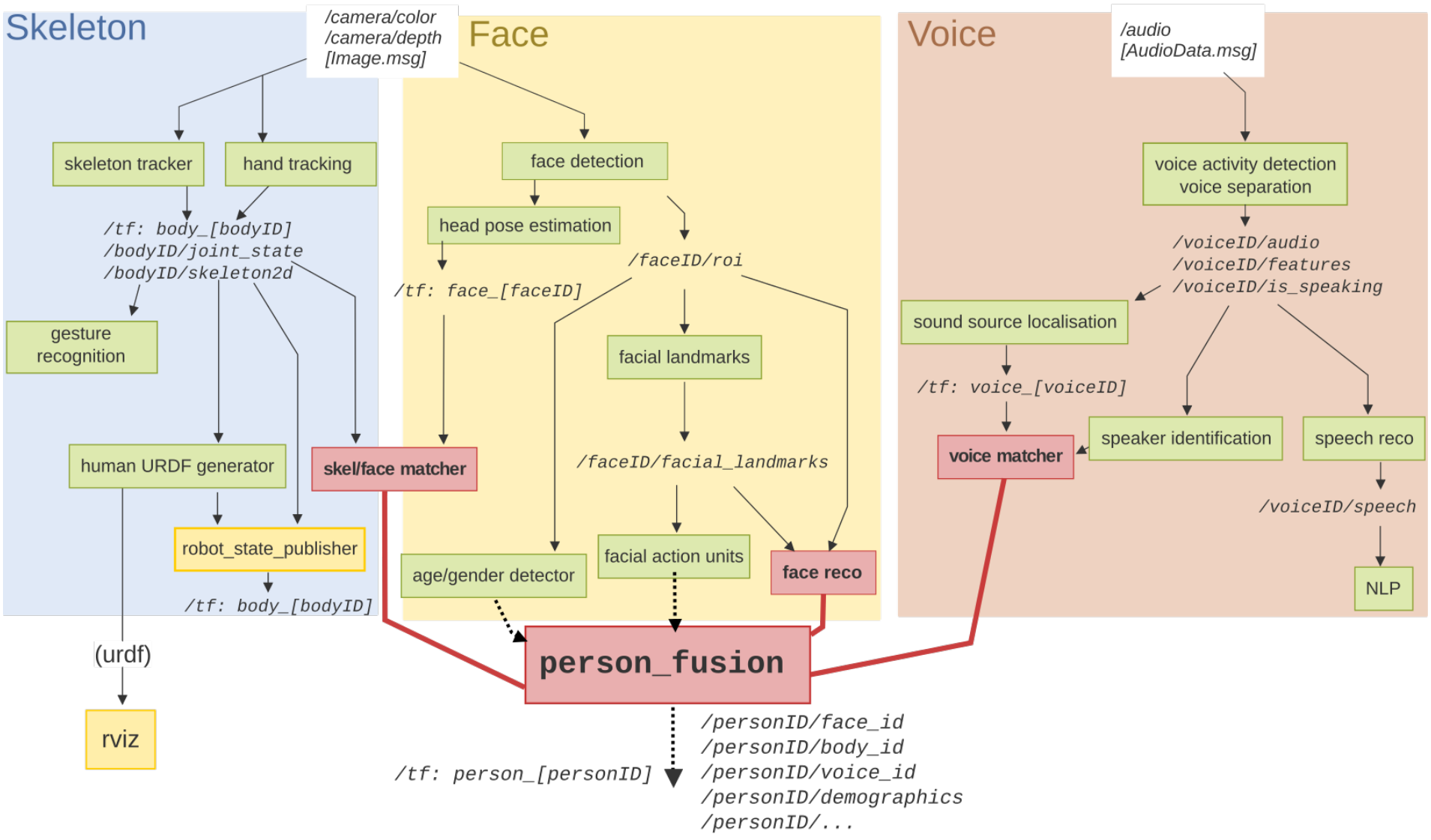
Name	Message type	Description
`/groups`	`hri_msgs/GroupsStamped`	Estimated social groups
`/gaze`	`hri_msgs/GazesStamped`	estimated gazing behaviours

HUMAN PHYSICAL REPRESENTATION

- standard ROS pipeline: joint state (eg OpenPose) -> `robot_state_publisher` + URDF
- URDF generated on the fly, based on person's height (xacro params)
- Follows REP-120 as much as possible.

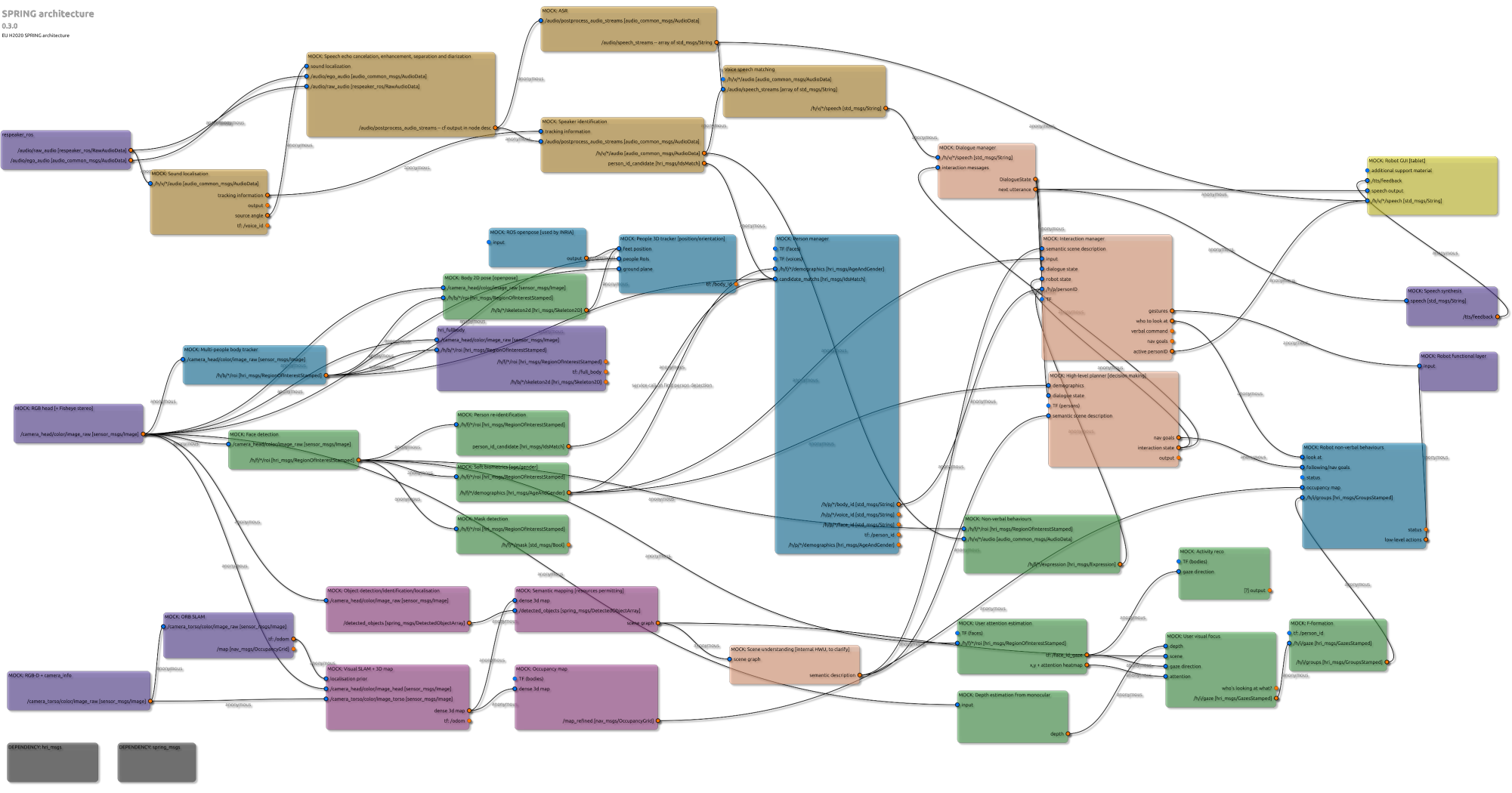


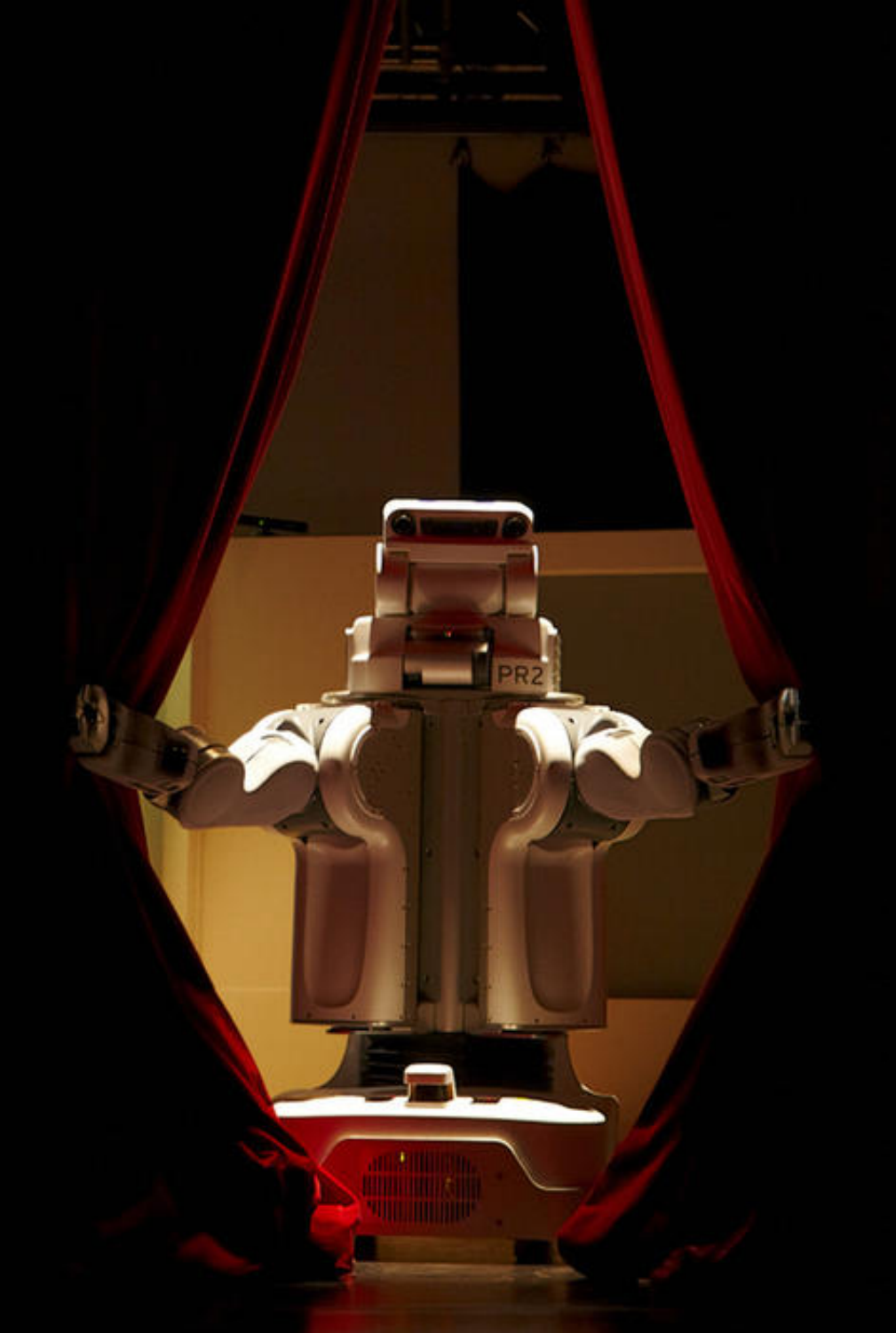



Skeleton
/camera/color
/camera/depth
[Image.msg]
skeleton tracker
hand tracking
/tf: body_[bodyID]
/bodyID/joint_state
/bodyID/skeleton2d
gesture recognition
human URDF generator
skel/face matcher
robot_state_publisher
/tf: body_[bodyID]
(urdf)
rviz
Face
face detection
head pose estimation
/faceID/roi
/tf: face_[faceID]
facial landmarks
/faceID/facial_landmarks
age/gender detector
facial action units
face reco
Voice
/audio
[AudioData.msg]
voice activity detection
voice separation
/voiceID/audio
/voiceID/features
/voiceID/is_speaking
sound source localisation
/tf: voice_[voiceID]
voice matcher
speaker identification
speech reco
/voiceID/speech
NLP
person_fusion
/tf: person_[personID]
/personID/face_id
/personID/body_id
/personID/voice_id
/personID/demographics
/personID/...

ROS4HRI: STATUS

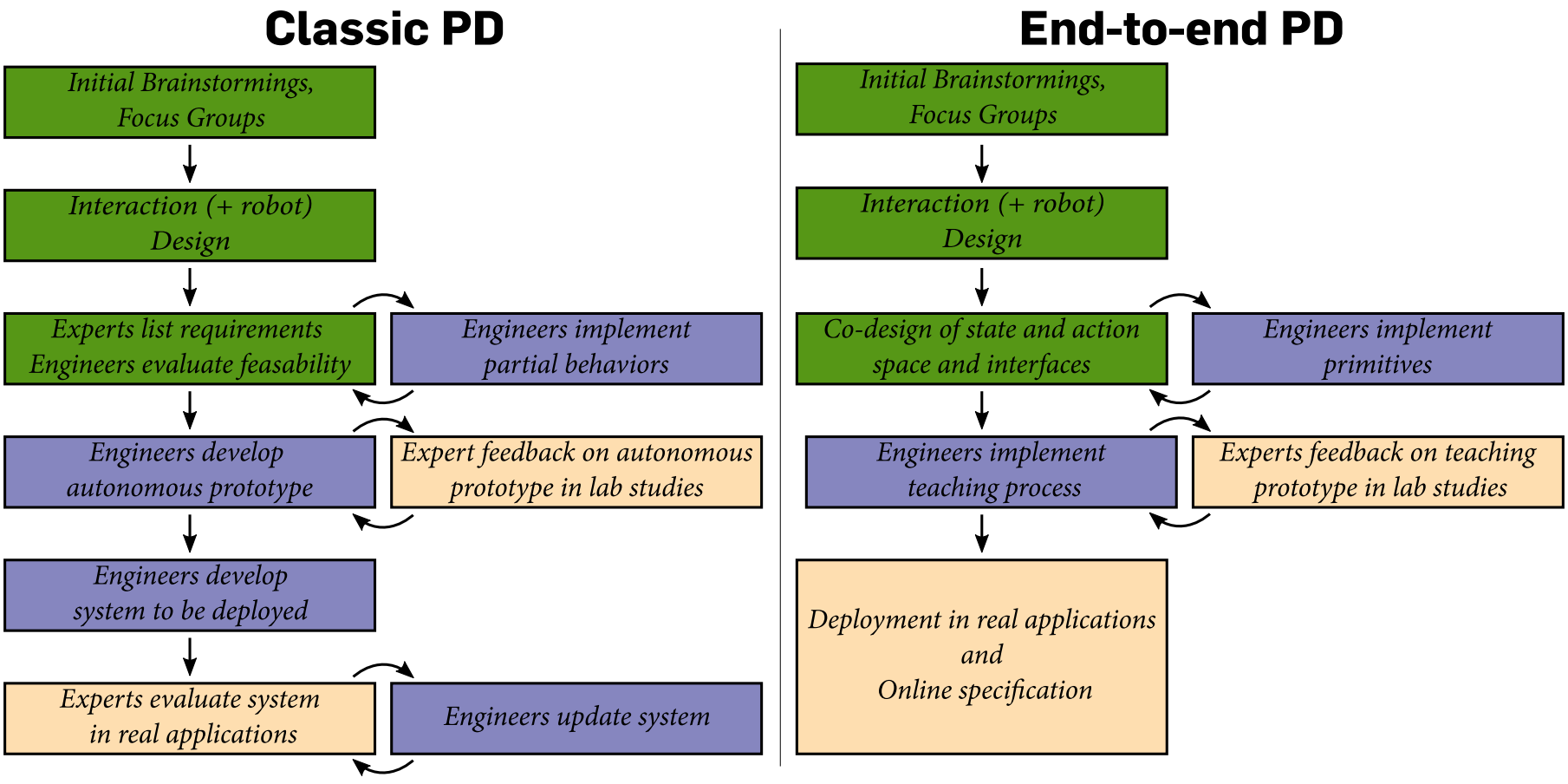
- v0 published, with reference implementation
- on-going work at PAL to use SotA DNNs for eg body posture estimation + better software engineering + real-time performance on ARI. Indicative ETA: 6-8 months
- `libhri` in early spec stage
- modeling gaps; need real-world use-case to better identify them $\rightarrow$ EU SPRING project


SPRING architecture
0.3.0
EU H2020 SPRING architecture
MOCK: ASR
/audio/postprocess_audio_streams [audio_common_msgs/AudioData]
/audio/speech_streams -- array of std_msgs/String
MOCK: Speech echo cancelation, enhancement, separation and diarization
sound localization
/audio/ego_audio [audio_common_msgs/AudioData]
/audio/raw_audio [respeaker_ros/RawAudioData]
/audio/postprocess_audio_streams -- cf output in node desc
MOCK: Speaker identification
tracking information
/audio/postprocess_audio_streams [audio_common_msgs/AudioData]
/h/s/*/audio [audio_common_msgs/AudioData]
person_id_candidate [hri_msgs/IdsMatch]
Voice speech matching
/h/s/*/audio [audio_common_msgs/AudioData]
/audio/speech_streams [array of std_msgs/String]
/h/s/*/speech [std_msgs/String]
respeaker_ros
/audio/raw_audio [respeaker_ros/RawAudioData]
/audio/ego_audio [audio_common_msgs/AudioData]
MOCK: Sound localisation
/h/s/*/audio [audio_common_msgs/AudioData]
tracking information
output
source angle
tf: /voice_id
MOCK: Dialogue manager
/h/s/*/speech [std_msgs/String]
interaction messages
DialogueState
next utterance
MOCK: Robot GUI [tablet]
additional support material
/tts/feedback
speech output
/h/s/*/speech [std_msgs/String]
MOCK: ROS openpose [used by INRIA]
input
output
MOCK: People 3D tracker [position/orientation]
feet position
people RoIs
ground plane
tf: /body_id
MOCK: Person manager
TF (faces)
TF (voices)
/h/f/*/demographics [hri_msgs/AgeAndGender]
candidate_matchs [hri_msgs/IdsMatch]
MOCK: Interaction manager
semantic scene description
input
dialogue state
robot state
/h/p/personID
TF
gestures
who to look at
verbal command
nav goals
active personID
MOCK: Speech synthesis
speech [std_msgs/String]
/tts/feedback
MOCK: Body 2D pose [openpose]
/camera_head/color/image_raw [sensor_msgs/Image]
/h/b/*/roi [hri_msgs/RegionOfInterestStamped]
/h/b/*/skeleton2d [hri_msgs/Skeleton2D]
hri_fullbody
/camera_head/color/image_raw [sensor_msgs/Image]
/h/b/*/roi [hri_msgs/RegionOfInterestStamped]
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
tf: /full_body
/h/b/*/skeleton2d [hri_msgs/Skeleton2D]
service call to full person detection
MOCK: Multi-people body tracker
/camera_head/color/image_raw [sensor_msgs/Image]
/h/b/*/roi [hri_msgs/RegionOfInterestStamped]
MOCK: Robot functional layer
input
MOCK: High-level planner [decision making]
demographics
dialogue state
TF (persons)
semantic scene description
nav goals
interaction state
output
MOCK: RGB head [+ Fisheye stereo]
/camera_head/color/image_raw [sensor_msgs/Image]
MOCK: Face detection
/camera_head/color/image_raw [sensor_msgs/Image]
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
MOCK: Person re-identification
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
person_id_candidate [hri_msgs/IdsMatch]
MOCK: Soft biometrics [age/gender]
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
/h/f/*/demographics [hri_msgs/AgeAndGender]
MOCK: Mask detection
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
/h/f/*/mask [std_msgs/Bool]
/h/p/*/body_id [std_msgs/String]
/h/p/*/voice_id [std_msgs/String]
/h/p/*/face_id [std_msgs/String]
tf: /person_id
/h/p/*/demographics [hri_msgs/AgeAndGender]
MOCK: Robot non-verbal behaviours
look at
following/nav goals
status
occupancy map
/h/l/groups [hri_msgs/GroupsStamped]
status
low level actions
MOCK: Non-verbal behaviours
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
/h/s/*/audio [audio_common_msgs/AudioData]
/h/f/*/expression [hri_msgs/Expression]
MOCK: Activity reco
TF (bodies)
gaze direction
[?] output
MOCK: Object detection/identification/localisation
/camera_head/color/image_raw [sensor_msgs/Image]
/detected_objects [spring_msgs/DetectedObjectArray]
MOCK: Semantic mapping [resources permitting]
dense 3d map
/detected_objects [spring_msgs/DetectedObjectArray]
scene graph
MOCK: ORB-SLAM
/camera_torso/color/image_raw [sensor_msgs/Image]
tf: /odom
/map [nav_msgs/OccupancyGrid]
MOCK: User attention estimation
TF (faces)
/h/f/*/roi [hri_msgs/RegionOfInterestStamped]
tf: /face_id_gaze
x,y + attention heatmap
MOCK: User visual focus
depth
scene
gaze direction
attention
who's looking at what?
/h/l/gaze [hri_msgs/GazesStamped]
MOCK: F-Formation
tf: /person_id
/h/l/gaze [hri_msgs/GazesStamped]
/h/l/groups [hri_msgs/GroupsStamped]
MOCK: Scene understanding [internal HWU, to clarify]
scene graph
semantic description
MOCK: RGB-D + camera_info
/camera_torso/color/image_raw [sensor_msgs/Image]
MOCK: Visual SLAM + 3D map
localisation prior
/camera_head/color/image_head [sensor_msgs/Image]
/camera_torso/color/image_torso [sensor_msgs/Image]
dense 3d map
tf: /odom
MOCK: Occupancy map
TF (bodies)
dense 3d map
/map_refined [nav_msgs/OccupancyGrid]
MOCK: Depth estimation from monocular
input
depth
DEPENDENCY: hri_msgs
DEPENDENCY: spring_msgs



Thank you!

Get the slides:
github.com/severin-lemaignan/presentation-codesign


Classic PD
Initial Brainstormings, Focus Groups
Interaction (+ robot) Design
Experts list requirements Engineers evaluate feasability
Engineers implement partial behaviors
Engineers develop autonomous prototype
Expert feedback on autonomous prototype in lab studies
Engineers develop system to be deployed
Experts evaluate system in real applications
Engineers update system
End-to-end PD
Initial Brainstormings, Focus Groups
Interaction (+ robot) Design
Co-design of state and action space and interfaces
Engineers implement primitives
Engineers implement teaching process
Experts feedback on teaching prototype in lab studies
Deployment in real applications and Online specification