

WITH PLYMOUTH UNIVERSITY

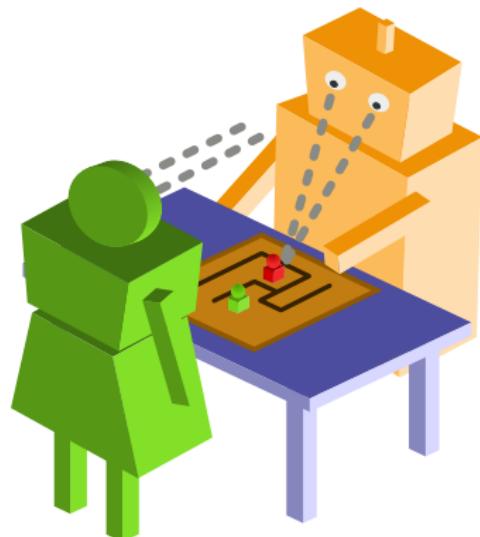


From Children Free Play to Robot AI on the way to artificial social cognition in HRI

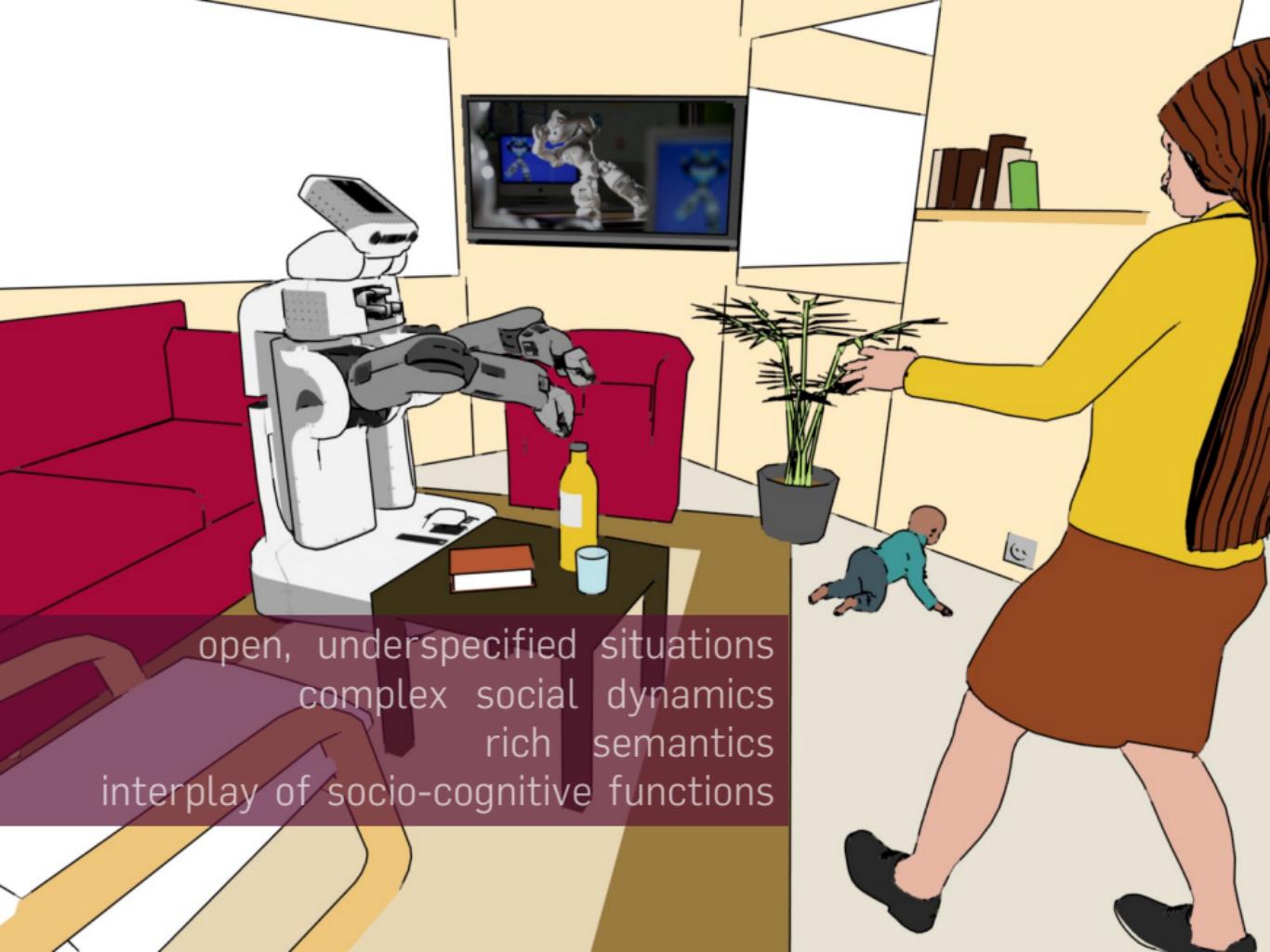
@CNRS-LAAS Toulouse – 11 jan. 2018

Séverin Lemaignan

Centre for Robotics and Neural Systems
Plymouth University



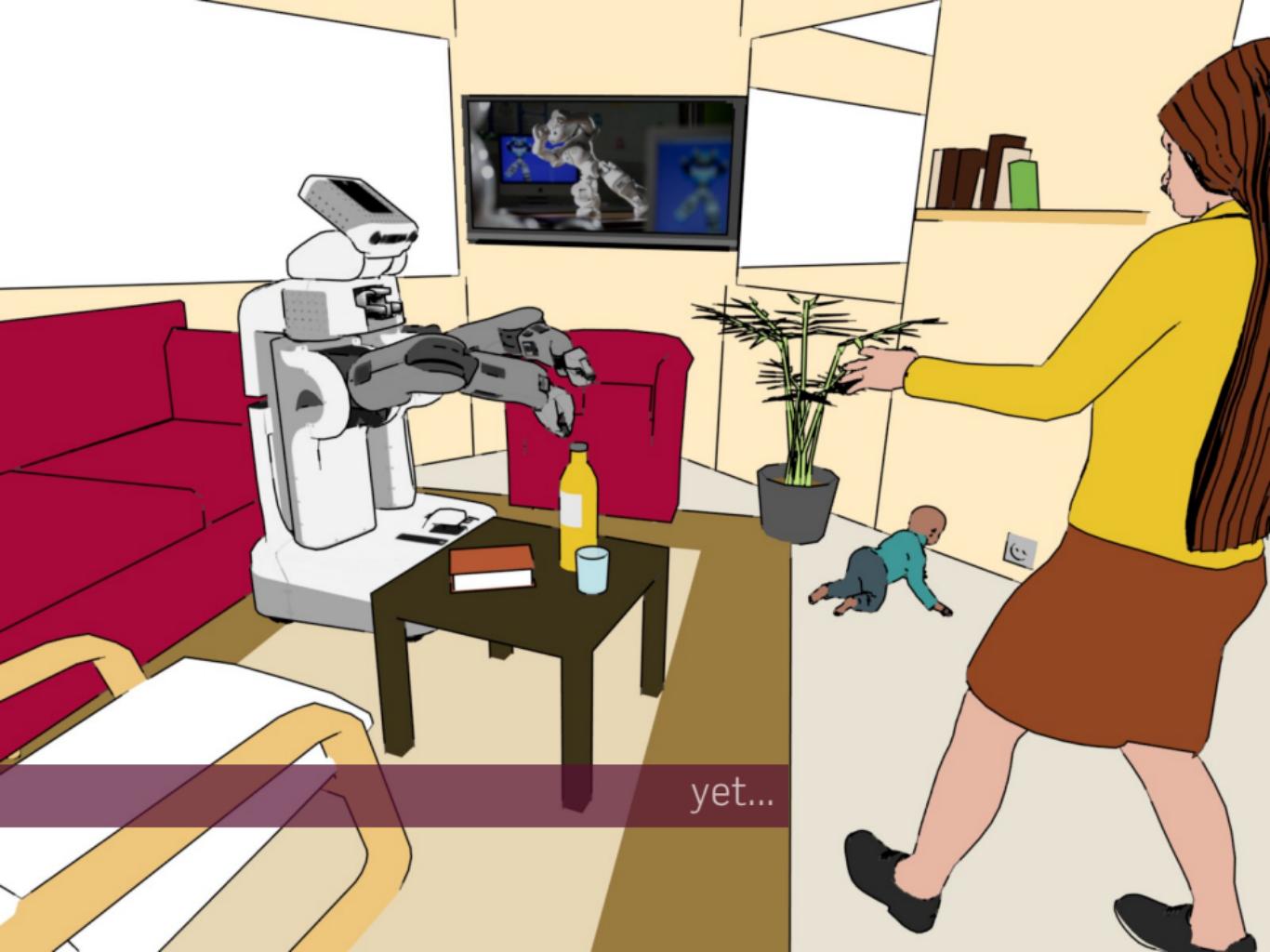
**How to push back the boundaries of social
robotics?**



open, underspecified situations
complex social dynamics
rich semantics
interplay of socio-cognitive functions

SURFACE FUNCTIONS FOR SOCIAL COGNITION





yet...

WHAT METHODOLOGY FOR SOCIAL HRI?

- typical socio-cognitive tasks **too simple and constrained**
- yet, research *in the wild* is **difficult to conduct rigorously and to replicate**

WHAT METHODOLOGY FOR SOCIAL HRI?

- typical socio-cognitive tasks **too simple and constrained**
- yet, research *in the wild* is **difficult to conduct rigorously and to replicate**

Finding the right task is difficult

- natural interactions ⇐ meaningful task
- realistic with today's technologies
- practical, reproducible and measurable
- focus on social cognition

THE FREE PLAY PARADIGM

FREE PLAY

“Just play! Enjoy yourselves!”

- **rich set of cognitive and social dynamics**; importance of motivation/drive; **uncertain and unexpected situations**
- what is the right action policy? Focus instead on the **social policy**



STAGES OF PLAY

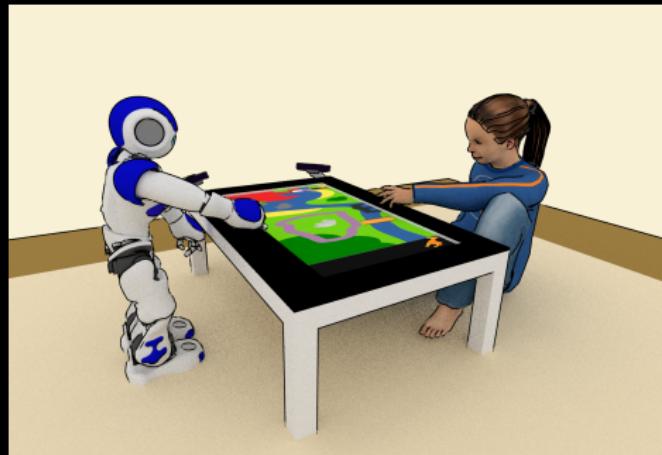
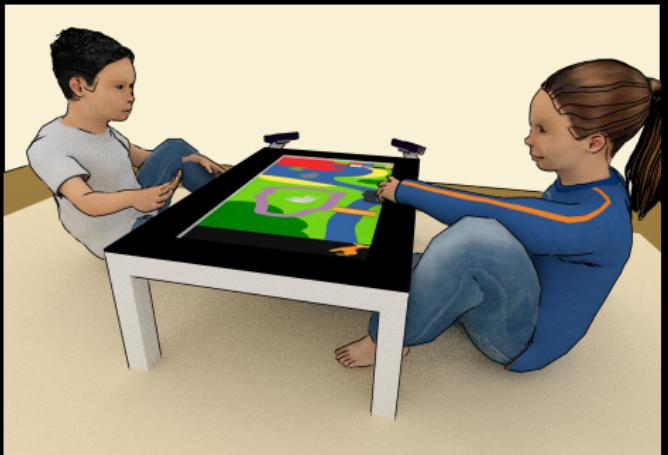
In developmental psychology, Parten's **stages of play**:

1.  **Solitary (independent) play**
2.  **Onlooker play**
3.  **Parallel play**
4.  **Associative play**
5.  **Cooperative play**



Can we make it work for HRI?





freeplay_sandbox_analysis_Free-play sandbox analysis tool - rqt

File Help

Bag

audio

env_camera

l_camera

r_camera

0m0s | 0m05s | 0m10s | 0m15s | 0m20s | 0m25s | 0m30s | 0m35s | 0m40s | 0m45s | 0m50s

1489767416.1165 Mar 17 2017 16:16:56.115 4.956s > 365.16 MB

D C O - O X Sandtray D C O - O X

D C O - O X l_camera D C O - O X r_camera D C O - O X

The interface displays four video timelines at the top left: 'audio' (empty), 'env_camera' (showing two children playing with a robot on a table), 'l_camera' (showing a child from the side), and 'r_camera' (showing a child from the side). A timeline bar at the top shows time points from 0m0s to 0m50s. Below the timelines are playback controls (rewind, play, fast forward). The center-right features a 3D map of a room with various objects labeled: 'odom', 'rhino', 'cube_29', 'cube_20', 'cube_37', 'cube_24', 'cube_22', 'cube_32', 'toychild4', 'cube_38', 'cube_25', 'cube_11', 'cube_5', 'child', 'lion', 'cube_34', 'cube_Stephan', 'cube_14', and 'cube_6'. The bottom section contains three large image frames showing the children playing with the robot on the table.

'SANDBOXED FREE PLAY' EXPERIMENTAL PARADIGM

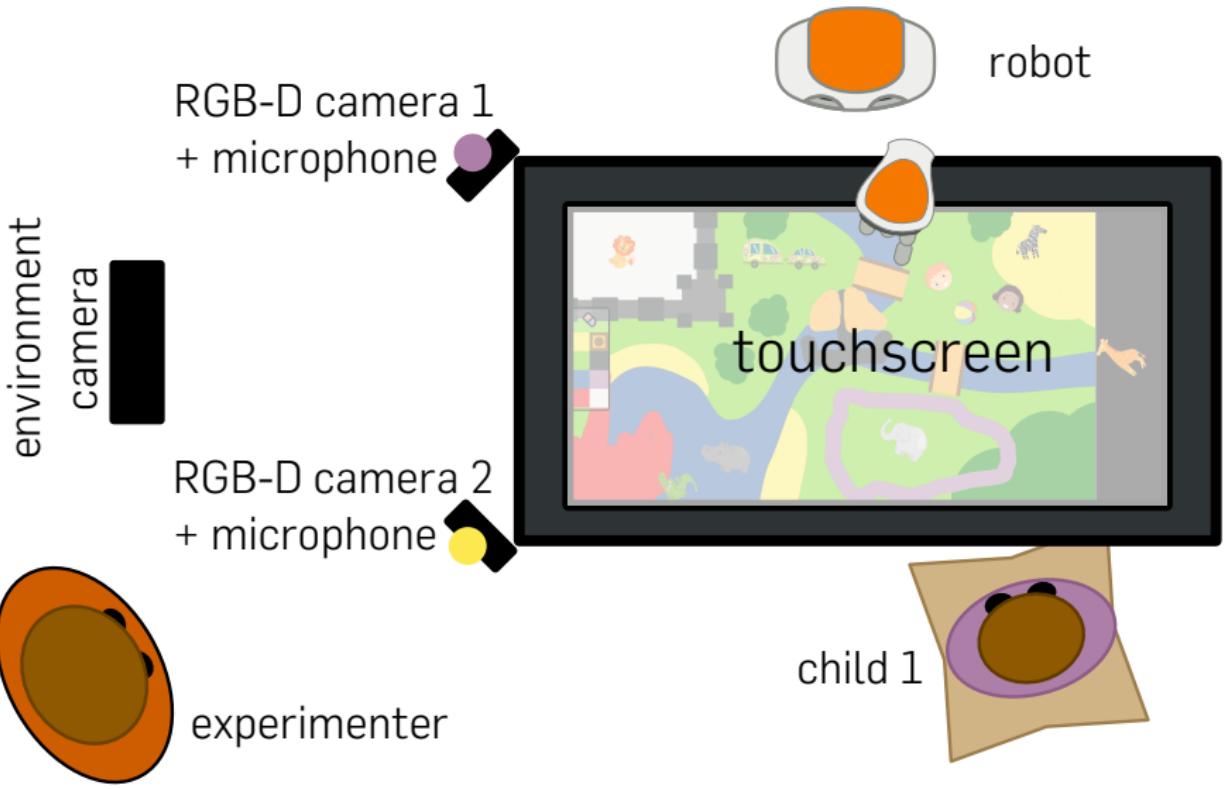
- **Structured methodology** (sandtray) yet **loosely structured task** (free play)

'SANDBOXED FREE PLAY' EXPERIMENTAL PARADIGM

- **Structured methodology** (sandtray) yet **loosely structured task** (free play)
- physical playground → **replaced by large touchscreen**: escape perception and manipulation in dense & cluttered scene (but *only* that)

'SANDBOXED FREE PLAY' EXPERIMENTAL PARADIGM

- **Structured methodology** (sandtray) yet **loosely structured task** (free play)
- physical playground → **replaced by large touchscreen**: escape perception and manipulation in dense & cluttered scene (but *only* that)
- importantly, **perception and interaction with the partner is unimpaired**



CHECK OUR 'SHOPPING LIST'

We were looking for a task that exhibits...

- open, underspecified situations
- complex social dynamics
- natural interactions
- rich semantics
- interplay of many socio-cognitive functions

CHECK OUR 'SHOPPING LIST'

We were looking for a task that exhibits...

- open, underspecified situations
- complex social dynamics
- natural interactions
- rich semantics
- interplay of many socio-cognitive functions

And as well:

- reproducible/replicable experimental procedure
- clear quantitative metrics
- practical

WHAT DO WE WANT TO OBSERVE?

- o the ~~semantic of the interaction~~

WHAT DO WE WANT TO OBSERVE?

- the **semantic of the interaction**
- the **task engagement**
are the participants 'on task' or not?

WHAT DO WE WANT TO OBSERVE?

- the **semantic of the interaction**
- the **task engagement**
 - are the participants 'on task' or not?
- the **interaction flow & situation awareness**
 - what is happening *now*? what is the next expected action?

WHAT DO WE WANT TO OBSERVE?

- the **semantic of the interaction**
- the **task engagement**
 - are the participants 'on task' or not?
- the **interaction flow & situation awareness**
 - what is happening *now*? what is the next expected action?
- the **social attitude**
 - pro-social, hostile, assertive ('bossy'), passive...

WHAT DO WE WANT TO OBSERVE?

- the **semantic of the interaction**
- the **task engagement**
 - are the participants 'on task' or not?
- the **interaction flow & situation awareness**
 - what is happening *now*? what is the next expected action?
- the **social attitude**
 - pro-social, hostile, assertive ('bossy'), passive...
- the **social dynamics**
 - entrainment (coupling), mimicry, implicit turn-taking, joint attention

WHAT DO WE WANT TO OBSERVE?

- the **semantic of the interaction**
- the **task engagement**
 - are the participants 'on task' or not?
- the **interaction flow & situation awareness**
 - what is happening *now*? what is the next expected action?
- the **social attitude**
 - pro-social, hostile, assertive ('bossy'), passive...
- the **social dynamics**
 - entrainment (coupling), mimicry, implicit turn-taking, joint attention

WHAT DO WE WANT TO OBSERVE?

- the **semantic of the interaction**
- the **task engagement**
 - are the participants 'on task' or not?
- the **interaction flow & situation awareness**
 - what is happening *now*? what is the next expected action?
- the **social attitude**
 - pro-social, hostile, assertive ('bossy'), passive...
- the **social dynamics**
 - entrainment (coupling), mimicry, implicit turn-taking, joint attention

→ **paradigm for socio-cognitive investigation**

SANDTRAY PARADIGM: MORE APPLICATIONS

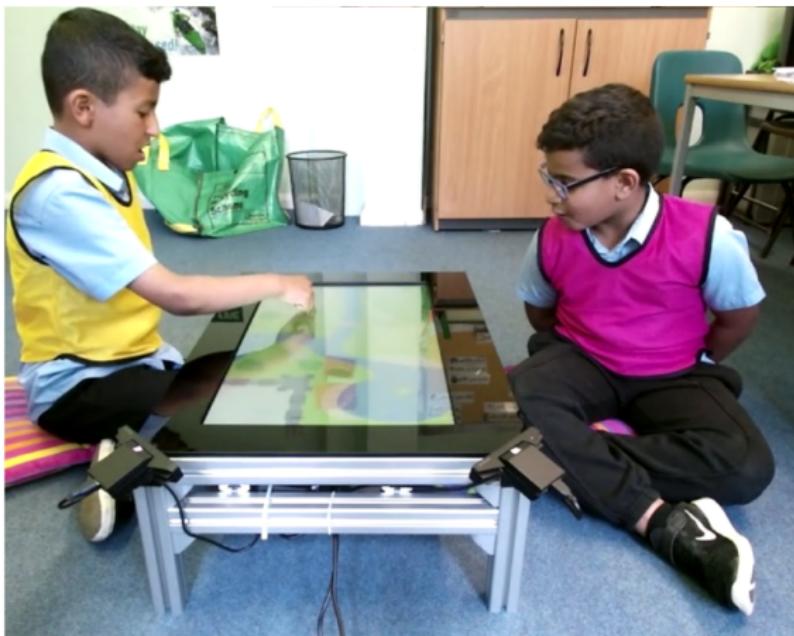
SPATIAL REASONING AND PERSPECTIVE TAKING

→ on-going PhD work by Christopher Wallbridge



SPATIAL REASONING AND PERSPECTIVE TAKING

→ on-going PhD work by Christopher Wallbridge



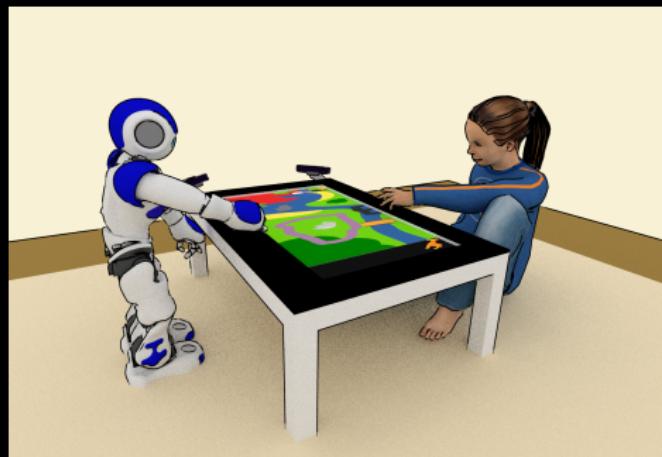
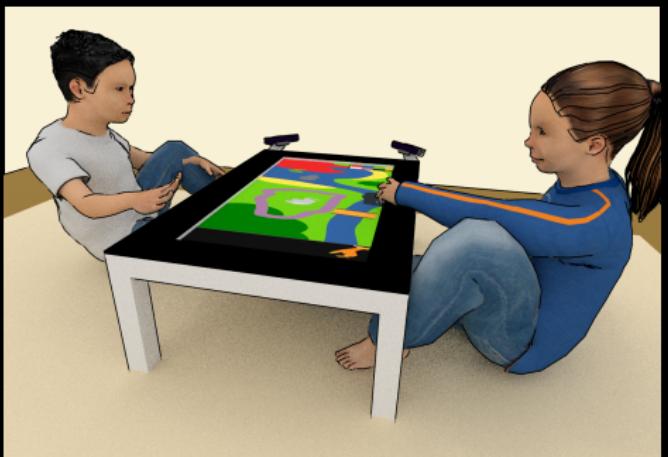
(take home message: ambiguity is good for you!)

SUPERVISED AUTONOMY

→ on-going PhD work by Emmanuel Senft



THE PINSORO DATASET



THE PINSORO DATASET

- 120 children, 4 to 8 years old
- 75 interactions
 - 90 children playing with another child,
 - 30 playing with a robot
- About 45h+ of recordings; 2M+ frames; \approx 2TB



TWO BASELINES



child - non-social robot



child - child

richness of social interactions

TWO BASELINES



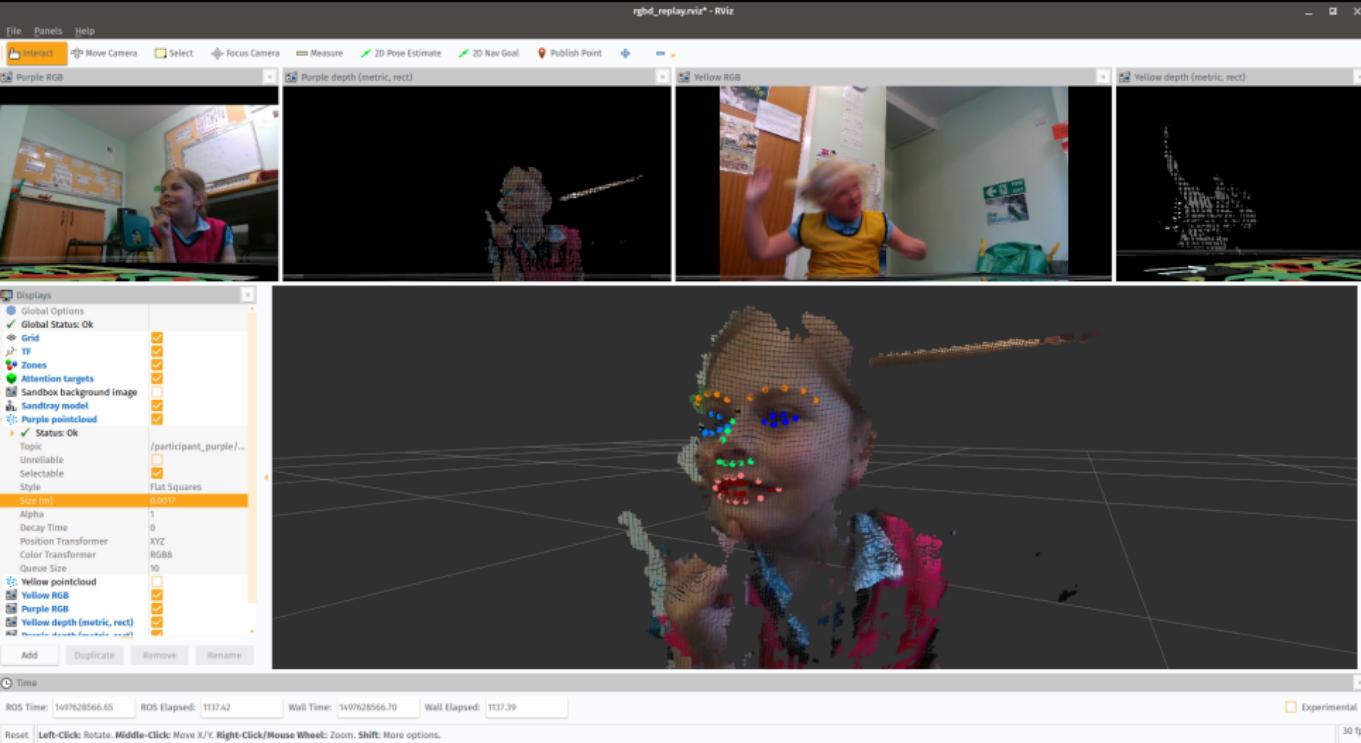
child - non-social robot



your next
socio-cognitive
model

child - child

richness of social interactions



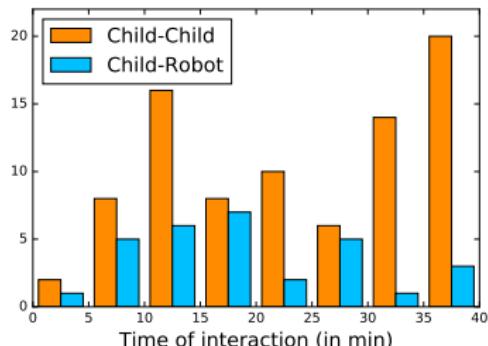
WHAT DID WE RECORD?

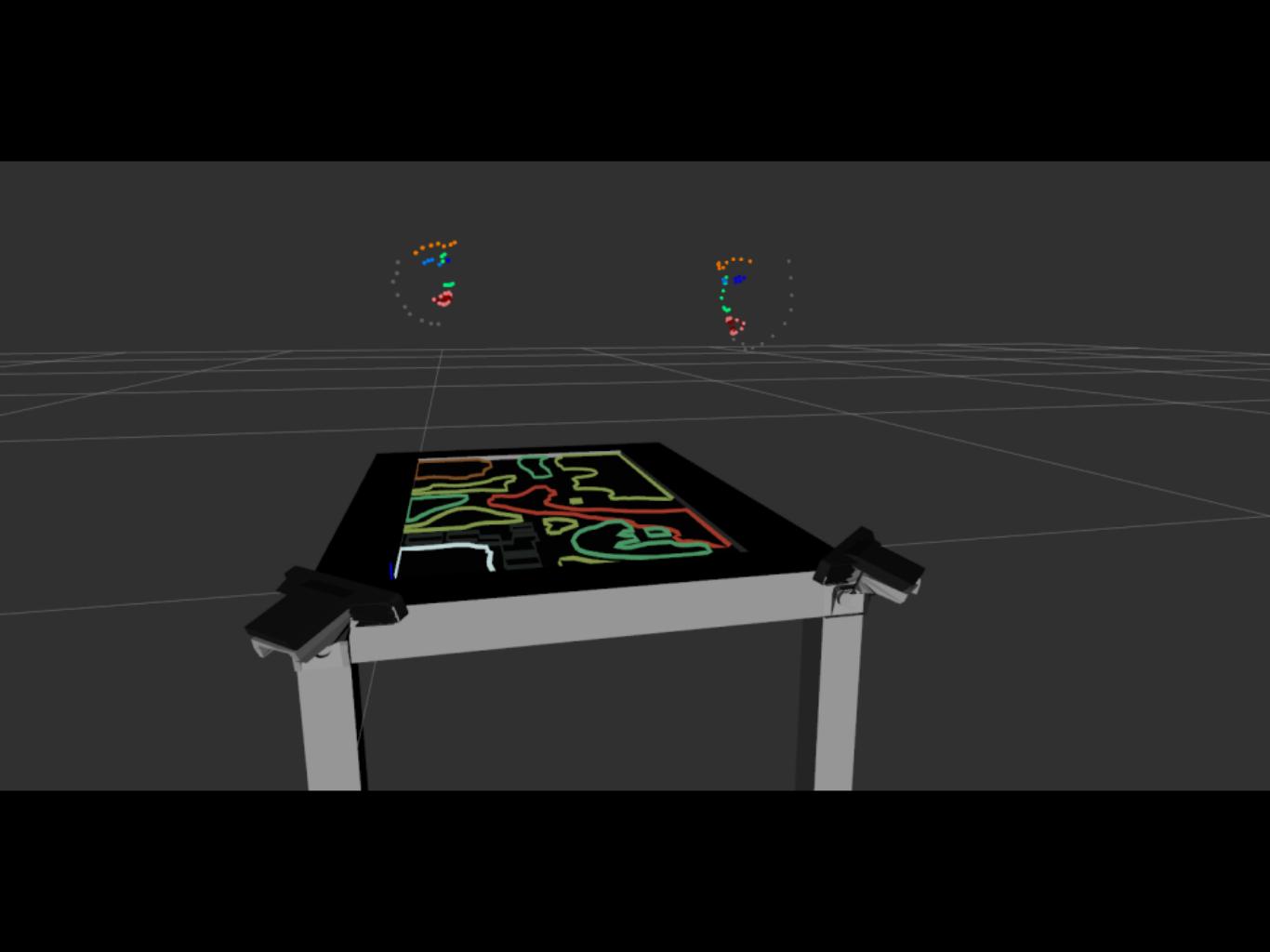
Domain	Type	Details
child × 2	audio	16kHz, mono, semi-directional
	face (RGB)	qHD (960x540), 30Hz
	face (depth)	VGA (640x480), 30Hz
	facial features	70 2D points, 30Hz
	skeleton	15 2D points, 30Hz
	hands	20 x 2 2D points, 30Hz
environment	RGB	qHD (960x540), 29.7Hz
touchscreen	background drawing (RGB)	4Hz
	touches	6 points multi-touch, 10Hz
	items position and orientation	(x,y,theta), 10Hz
annotations	timestamped annotations of social behaviours	
other	static transforms between touchscreen and facial cameras	
	cameras calibration informations	



DATASET

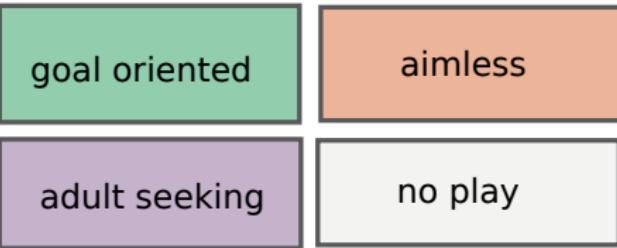
- 120 children, 4 to 8 years old; 45h+ of recordings;
- average duration of freeplay interactions: 24min in child-child condition; 19min in child-robot condition
- facial features extracted in 98% of frames



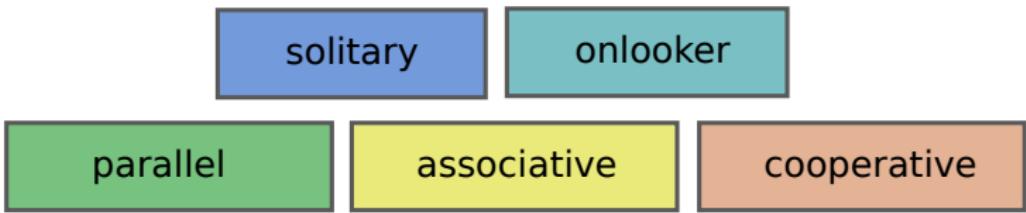




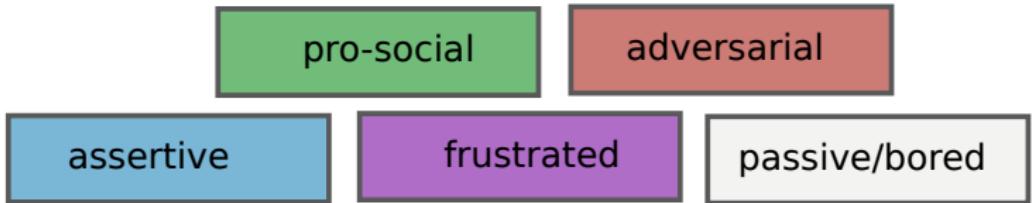
Task engagement

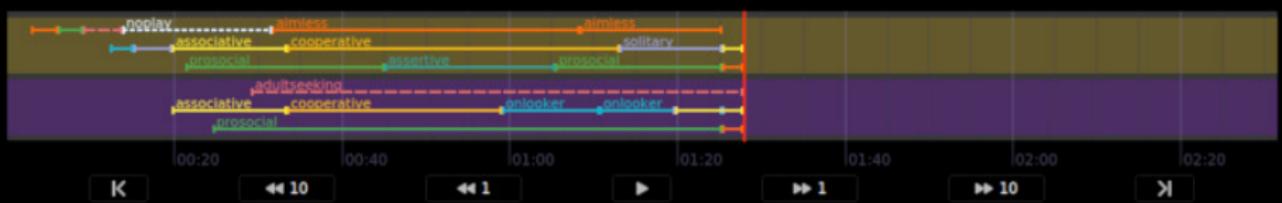


Social engagement



Social attitude





01:28 {88.0326s}

Auto-saving to /home/slemaignan/freeplay_sandbox/data/2017-06-12-143746652201/freeplay.annotations.yaml

SOCIAL ANNOTATIONS

So far,

- 85% of the dataset annotated
- 11500+ annotations
- average duration of coded episodes: 46 seconds
- 15% double-coded

Attitude: passive

Social engag.: onlooker

Task engag.: no play

Attitude: passive

Social engag.: solitary

Task engag.: goal oriented

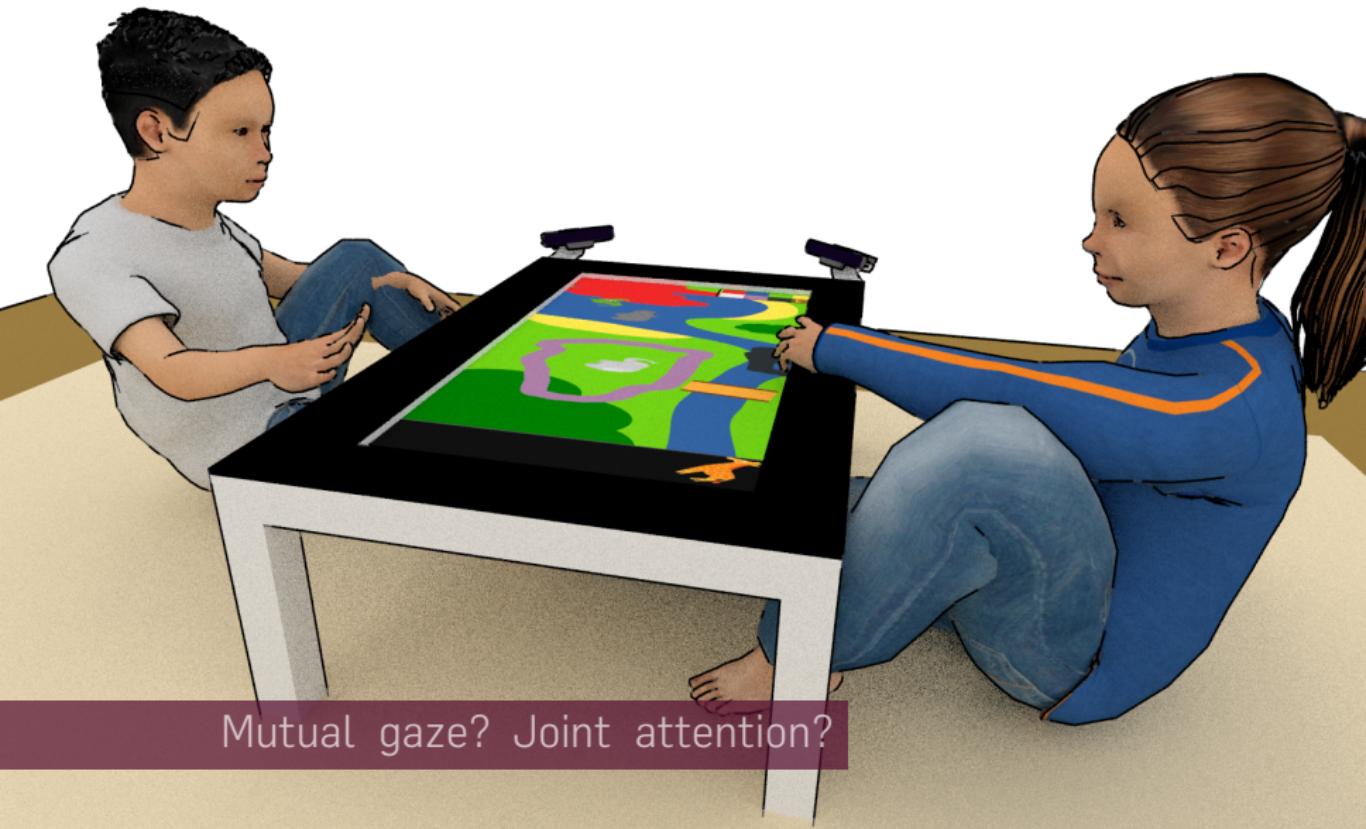


Anonymised version (only 5.7GB...) available
on-line. Grab it now!

freeplay-sandbox.github.io

Open data! Hosted on EU's **zenodo**

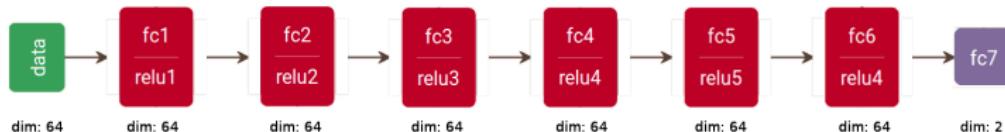
WHAT DOES THIS DATASET MEAN TO
HRI?



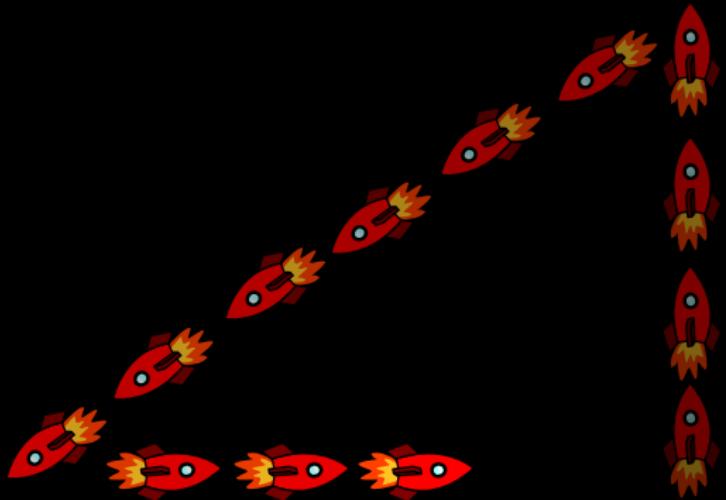
Mutual gaze? Joint attention?

TRAINING FOR GAZE ESTIMATION

End-to-end training to map 2D facial features to gaze location

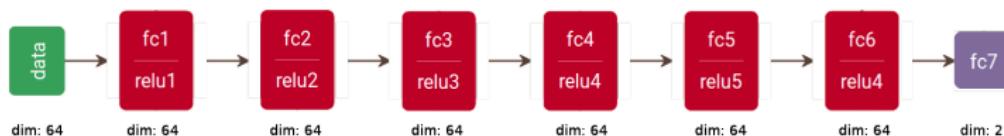


- **Input:** 32 2D points (eyes, eyebrows, nose, ears, shoulders)
- **Output:** 2D gaze location on the interactive table

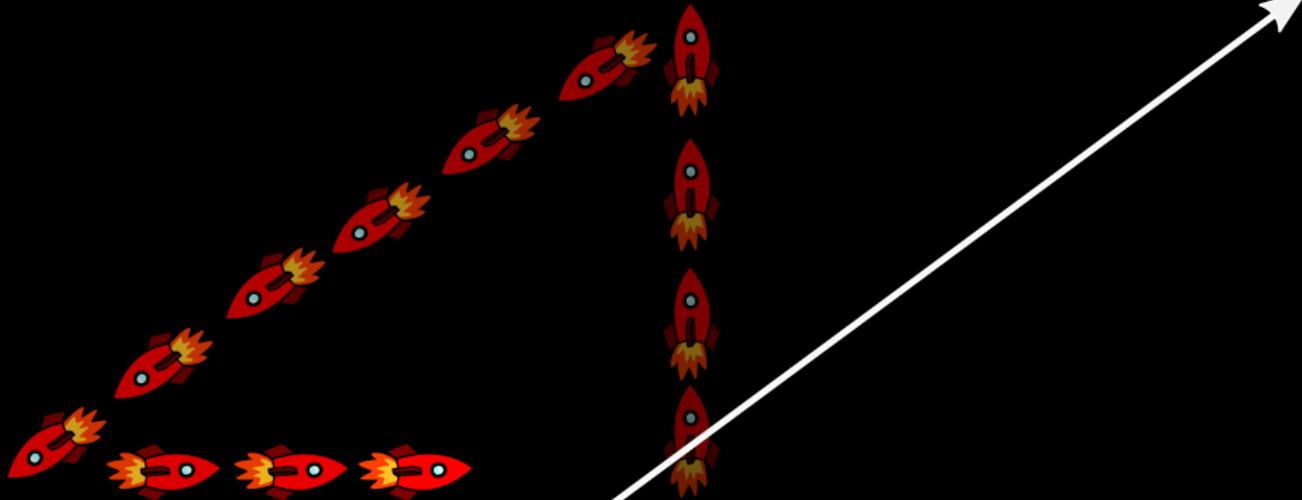


TRAINING FOR GAZE ESTIMATION

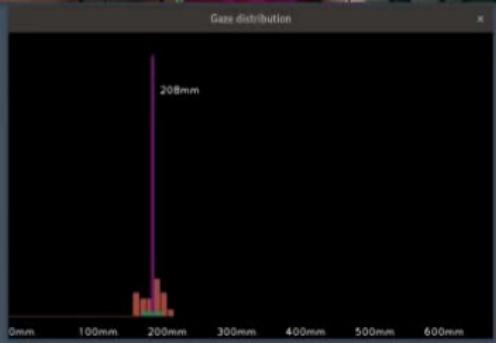
End-to-end training to map 2D facial features to gaze location

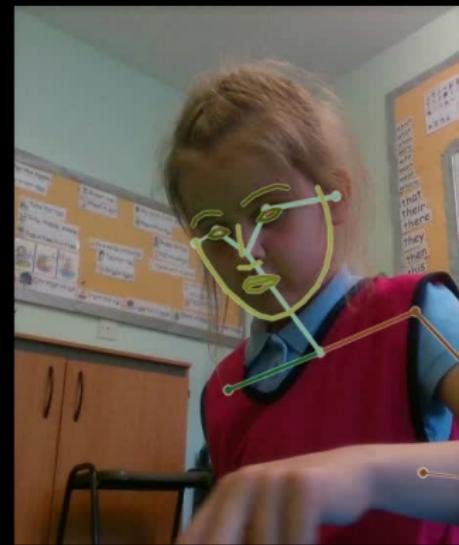
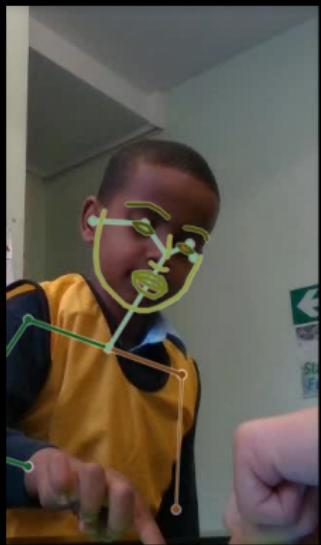


- **Pros:** no calibration; no eye tracking device; 2D images
- **Cons:** require a initial ground truth; accurate?



~70cm

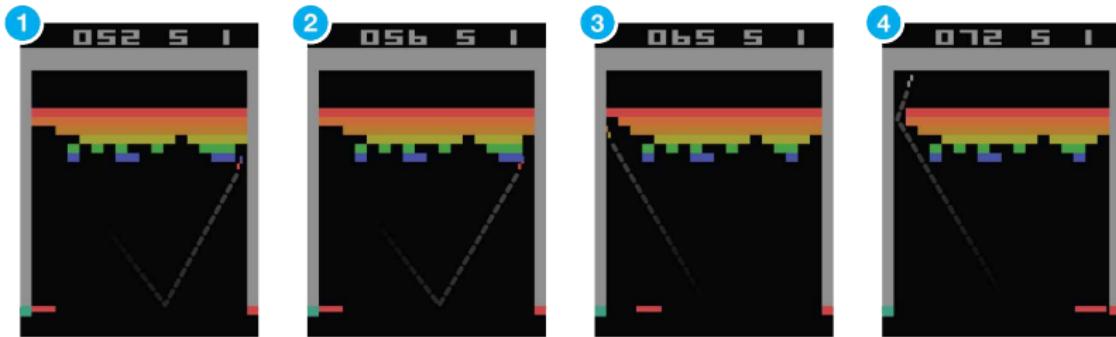




What about social HRI?

DEEP LEARNING AND ROBOTICS

LEARNING COMPLEX BEHAVIOURS



- Inputs: raw screen image + score
- from the outside, looks like planning

LEARNING COMPLEX BEHAVIOURS



- Inputs: raw screen image + score
- from the outside, looks like planning
- ~~1.000.000~~ **500** games to play a good human-level



LEARNING SEQUENCES

Ogata's demonstration (published this year in IEEE RAL):

- 2 arms, 12 DoF
- Inputs: on-board 112×112 px camera, joint state

LEARNING SEQUENCES

Ogata's demonstration (published this year in IEEE RAL):

- 2 arms, 12 DoF
- Inputs: on-board 112×112 px camera, joint state
- training from 40 teleoperated demonstrations of ≈ 70 s

LEARNING SEQUENCES

Ogata's demonstration (published this year in IEEE RAL):

- 2 arms, 12 DoF
- Inputs: on-board 112×112 px camera, joint state
- training from 40 teleoperated demonstrations of ≈ 70 s

Not only learning poses, but **sequences as well.**

Time-Delay Neural Network (TDNN) to learn to predict the next step (no RNNs!).

DEEP LEARNING OF SOCIAL INTERACTIONS?

Attitude: passive

Social engag.: onlooker

Task engag.: no play

Attitude: passive

Social engag.: solitary

Task engag.: goal oriented



AIM

Real-time identification by the robot of...

- the **task engagement**
is my partner 'on task' or not?

AIM

Real-time identification by the robot of...

- the **task engagement**
is my partner 'on task' or not?
- the **interaction flow & situation awareness**
what is happening right now? should I do something?

AIM

Real-time identification by the robot of...

- the **task engagement**
is my partner 'on task' or not?
- the **interaction flow & situation awareness**
what is happening right now? should I do something?
- the **social attitude**
Pro-social, hostile, assertive ('bossy'), passive...

AIM

Real-time identification by the robot of...

- the **task engagement**
is my partner 'on task' or not?
- the **interaction flow & situation awareness**
what is happening right now? should I do something?
- the **social attitude**
Pro-social, hostile, assertive ('bossy'), passive...
- the **social dynamics**
entrainment (coupling), mimicry, turn-taking, joint attention

AIM

Real-time identification by the robot of...

- the **task engagement**
is my partner 'on task' or not?
- the **interaction flow & situation awareness**
what is happening right now? should I do something?
- the **social attitude**
Pro-social, hostile, assertive ('bossy'), passive...
- the **social dynamics**
entrainment (coupling), mimicry, turn-taking, joint attention

AIM

Real-time identification by the robot of...

- the **task engagement**
is my partner 'on task' or not?
- the **interaction flow & situation awareness**
what is happening right now? should I do something?
- the **social attitude**
Pro-social, hostile, assertive ('bossy'), passive...
- the **social dynamics**
entrainment (coupling), mimicry, turn-taking, joint attention

Social behaviours; Social dynamics: **generation as well!**

SOME BUILDING BLOCKS EXISTS

- **Multi-modal fusion**
e.g. Noda et al. **Multimodal integration learning of robot behavior using DNN**, Robotics and Autonomous Systems 2014
- **Behavioural sequences recognition**
How et al. **Behavior recognition for humanoid robots using long short-term memory**, IJARS 2016 → *LSTM to recognise Nao behaviours*
Shiarlis et al. **Acquiring Social Interaction Behaviours for Telepresence Robots via Deep Learning from Demonstration**, IROS 2017

SOME BUILDING BLOCKS EXISTS

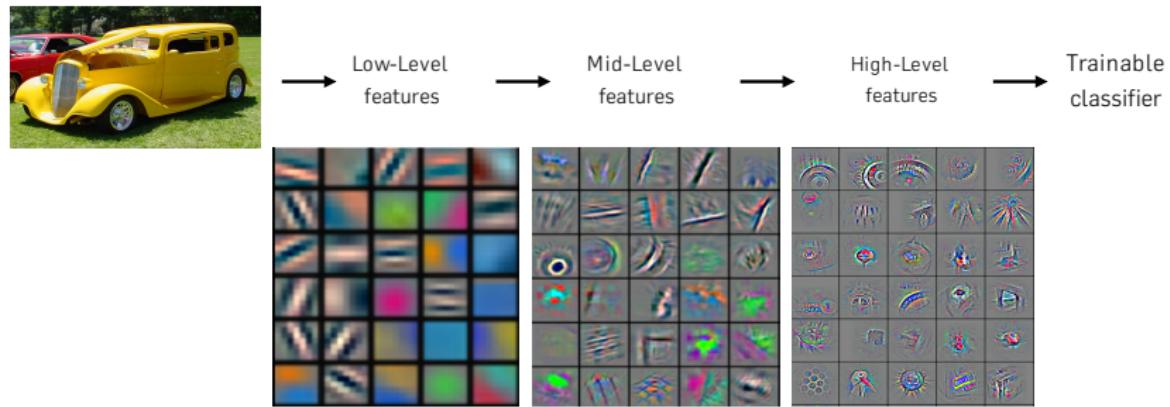
- **Multi-modal fusion**
e.g. Noda et al. **Multimodal integration learning of robot behavior using DNN**, Robotics and Autonomous Systems 2014
- **Behavioural sequences recognition**
How et al. **Behavior recognition for humanoid robots using long short-term memory**, IJARS 2016 → *LSTM to recognise Nao behaviours*
Shiarlis et al. **Acquiring Social Interaction Behaviours for Telepresence Robots via Deep Learning from Demonstration**, IROS 2017

DBSoC: Deep Behavioural Social Cloning – LfD + CNNs + LSTM

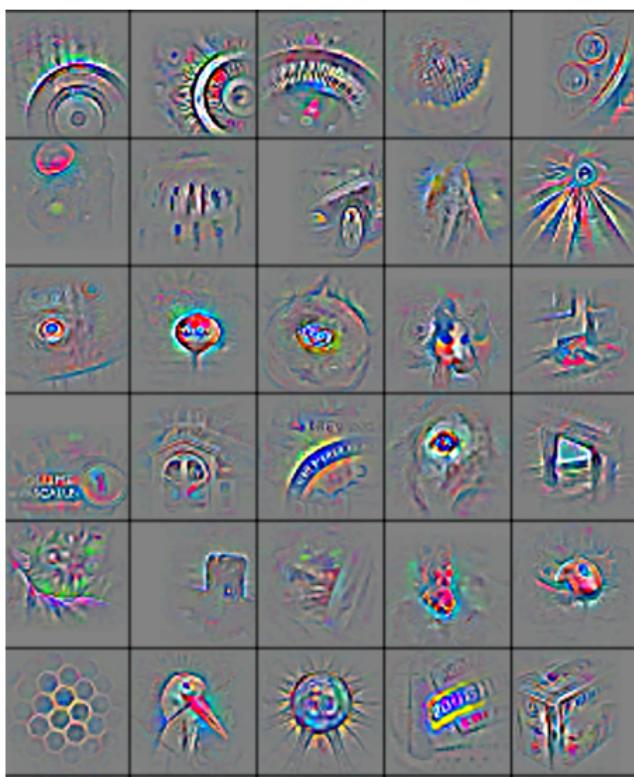
Two tasks for a telepresence robot:

1. position itself in a (dynamic) group of persons
2. follow 2 persons

DEEP NETWORKS ≡ BLACK BOXES?



DEEP NETWORKS ≡ BLACK BOXES?

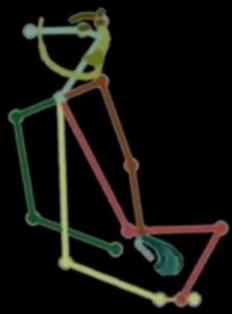


[taken from a NIPS2015 tutorial by Geoff Hinton, Yoshua Bengio & Yann LeCun]

Attitude: pro-social

Social engag.: parallel play

Task engag.: goal oriented



Attitude: pro-social

Social engag.: parallel play

Task engag.: goal oriented



The problem is framed, data is available, next step: **mining it!**

BACK TO THE BIGGER PICTURE

ONE QUESTION

Can sociality emerge from interaction?

ONE QUESTION

Can sociality emerge from interaction?

Both “emerge” as *arise from* and “emerge” as in *emergent paradigm of cognition*

ONE QUESTION

Can sociality emerge from interaction?

Both “emerge” as *arise from* and “emerge” as in *emergent paradigm of cognition*

“Social cognition arising in interaction”? → a situated & embodied view on cognition

DEEP LEARNING OF SOCIAL INTERACTIONS

Working hypothesis: **Sociality emerges from interaction**

DEEP LEARNING OF SOCIAL INTERACTIONS

Working hypothesis: **Sociality emerges from interaction**
⇒ **learning how to interact might lead to an emergent socio-cognitive behaviour**

DEEP LEARNING OF SOCIAL INTERACTIONS

Working hypothesis: **Sociality emerges from interaction**

⇒ **learning how to interact might lead to an emergent socio-cognitive behaviour**

⇒ **(Deep) learning of socio-cognitive human-robot interactions**

(Supervised (or unsupervised!) recurrent neural networks to model others' minds → a connectionist theory of mind!)

DEEP LEARNING OF SOCIAL INTERACTIONS

Working hypothesis: **Sociality emerges from interaction**

⇒ **learning how to interact might lead to an emergent socio-cognitive behaviour**

⇒ **(Deep) learning of socio-cognitive human-robot interactions**

(Supervised (or unsupervised!) recurrent neural networks to model others' minds → a connectionist theory of mind!)

...towards a principled model of social cognition?

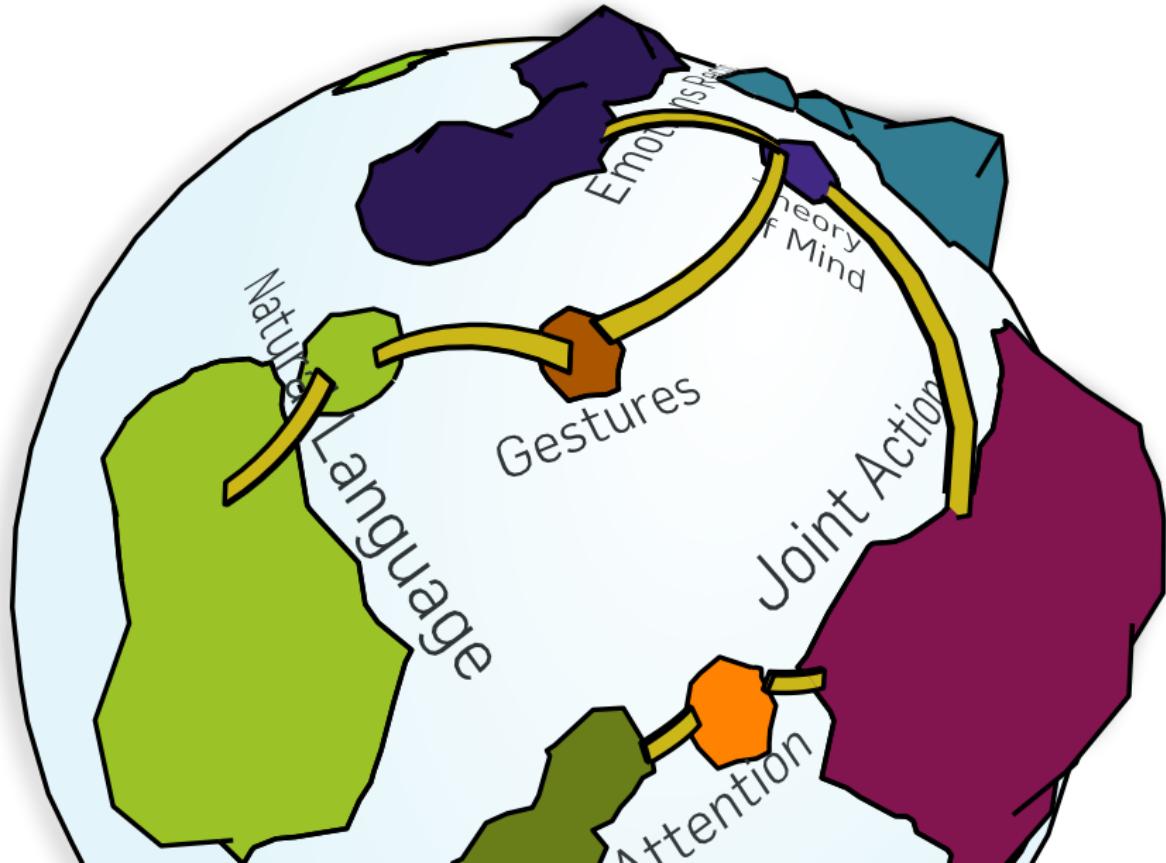
A MODEL?

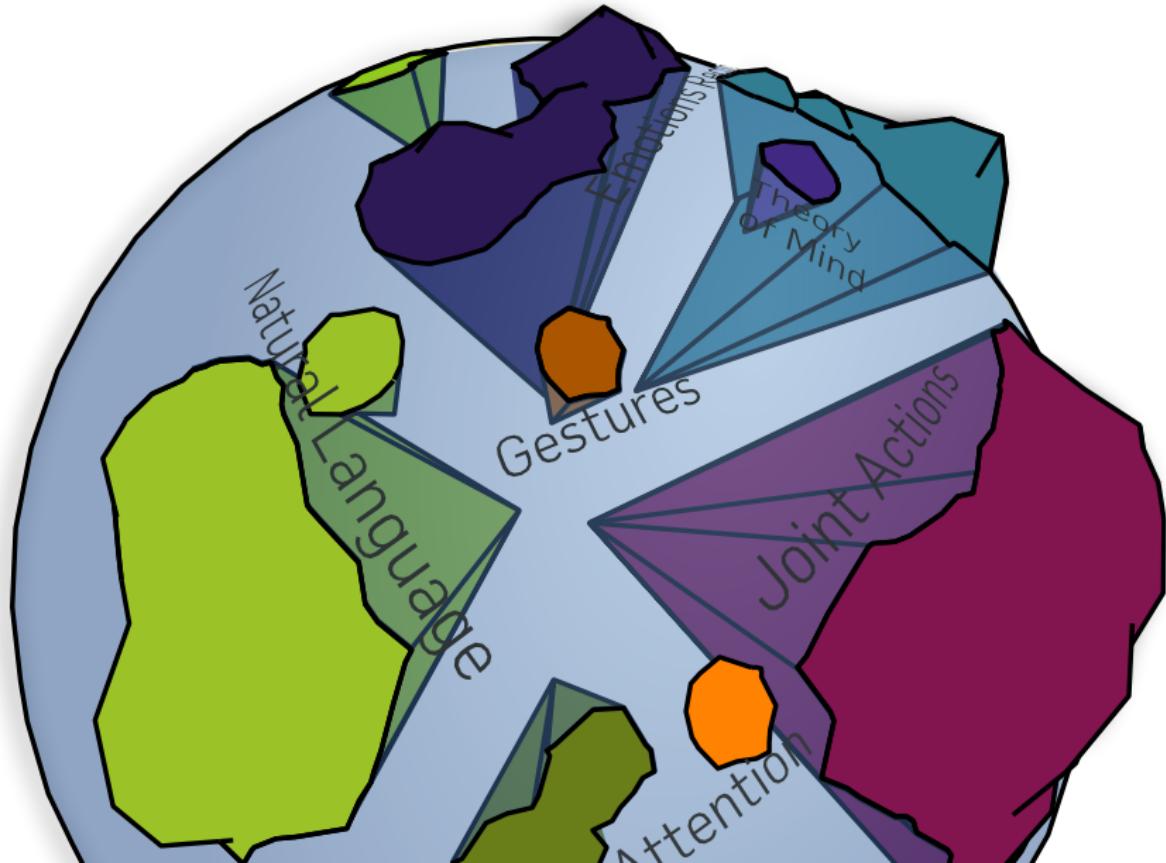
Models attempt to *explain*:

“identifying the causes for an event or phenomenon of interest”

“unifying disparate phenomena”









A MODEL?

Models attempt to *explain*:

“identifying the causes for an event or phenomenon of interest”

“unifying disparate phenomena”

A model's value is gained from

“predicting facts that, absent the theory, would be antecedently improbable”

TOWARDS DEVELOPMENTAL SOCIO-ROBOTICS?

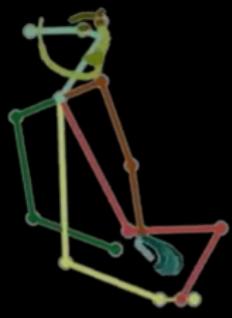
Emergence of Parten's stages of play?

1.  Solitary (independent) play
2.  Onlooker play
3.  Parallel play
4.  Associative play
5.  Cooperative play

Attitude: pro-social

Social engag.: parallel play

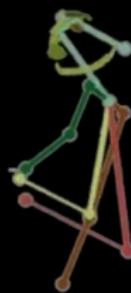
Task engag.: goal oriented



Attitude: pro-social

Social engag.: parallel play

Task engag.: goal oriented



Thank you!

freeplay-sandbox.github.io