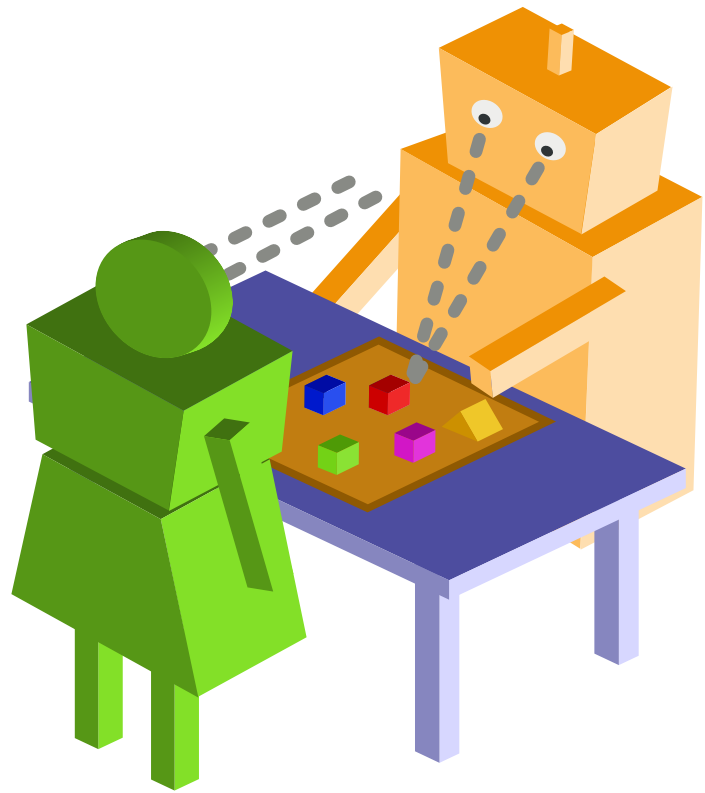
cognition, HRI & data

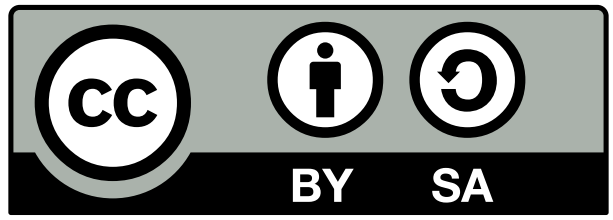
how to *do* together?

UKRAS19 – 24th Jan. 2019



Séverin Lemaignan

Bristol Robotics Lab University of the West of England



You can download the sources of this presentation here:

github.com/severin-lemaignan/2019-ukras-talk

JOINT ACTION WITH A ROBOT?

MODEL-BASED JOINT ACTION

1. establish a joint goal
2. plan for the robot
3. plan for the human in order to build a set of priors
4. execute the robot plan
5. monitor progress of the partner towards the goal
6. (loop back)

[Lemaignan et al. **Artificial Cognition for Social Human-Robot Interaction: An Implementation** Artificial Intelligence 2017]

MODEL-BASED JOINT ACTION

1. establish a joint goal
2. plan for the robot
3. plan for the human in order to build a set of priors
4. execute the robot plan
5. monitor progress of the partner towards the goal
6. (loop back)

⇒ **explicit cognitive steps**

MODEL-BASED JOINT ACTION

1. establish a joint goal
2. plan for the robot
3. plan for the human in order to build a set of priors
4. execute the robot plan
5. monitor progress of the partner towards the goal
6. (loop back)

⇒ **explicit cognitive steps**

...hard (and brittle) ones, though:

- *how to communicate/agree on goals & plans?*
- *what about the human's own plans?*
- *what to do when we're going off track?*
- *monitoring/recognising error situations*
- *...*

[Lemaignan et al. **Artificial Cognition for Social Human-Robot Interaction: An Implementation** Artificial Intelligence 2017]

WE CAN DO A BIT ALREADY

[Lemaignan et al. **Artificial Cognition for Social Human-Robot Interaction: An Implementation** – Artificial Intelligence 2017]

HOW DO HUMANS PERFORM TASKS TOGETHER?

Collaborating is a costly socio-cognitive activity.

HOW DO HUMANS PERFORM TASKS TOGETHER?

Collaborating is a costly socio-cognitive activity.

Humans are good at it.

They are also really good at minimizing the cost involved.

HOW DO HUMANS PERFORM TASKS TOGETHER?

Collaborating is a costly socio-cognitive activity.

Humans are good at it.

They are also really good at minimizing the cost involved.

The key mechanism: **prefer implicit to explicit**

HOW DO HUMANS PERFORM TASKS TOGETHER?

Collaborating is a costly socio-cognitive activity.

Humans are good at it.

They are also really good at minimizing the cost involved.

The key mechanism: **prefer implicit to explicit**

...which is closely related to: **be lazy**

First, go for the simple – if possibly ambiguous – actions; and, if really needed, repair

HOW DO HUMANS PERFORM TASKS TOGETHER?

Collaborating is a costly socio-cognitive activity.

Humans are good at it.

They are also really good at minimizing the cost involved.

The key mechanism: **prefer implicit to explicit**

...which is closely related to: **be lazy**

First, go for the simple – if possibly ambiguous – actions; and, if really needed, repair

What does "be lazy" mean for robots?

ONE EXAMPLE: GROUNDING OF SPATIAL LANGUAGE

Ambiguities arise easily when describing spatial scenes.

How do we solve them?

SURFACE ALIGNMENT; GROUNDING CRITERION

Psycholinguistics provides a lot of the foundational work on these questions.

- ***Communication is a dynamic social process***: the partner often to signal missing/misunderstood informations

[Clark and Wilkes-Gibbs, ***Referring as a collaborative process***, Cognition 1986
Pickering and Garrod, ***Alignment as the basis for successful communication***, Research on Language and Computation 2006]

SURFACE ALIGNMENT; GROUNDING CRITERION

Psycholinguistics provides a lot of the foundational work on these questions.

- ***Communication is a dynamic social process***: the partner often to signal missing/misunderstood informations
- Repairing is generally less costly than avoiding ambiguities in the first place

[Clark and Wilkes-Gibbs, ***Referring as a collaborative process***, Cognition 1986
Pickering and Garrod, ***Alignment as the basis for successful communication***, Research on Language and Computation 2006]

SURFACE ALIGNMENT; GROUNDING CRITERION

Psycholinguistics provides a lot of the foundational work on these questions.

- ***Communication is a dynamic social process***: the partner often to signal missing/misunderstood informations
- Repairing is generally less costly than avoiding ambiguities in the first place
- You only ever need to reach the ***grounding criterion***, ie ***enough*** mutual understanding for the task

[Clark and Wilkes-Gibbs, ***Referring as a collaborative process***, Cognition 1986
Pickering and Garrod, ***Alignment as the basis for successful communication***, Research on Language and Computation 2006]

SURFACE ALIGNMENT; GROUNDING CRITERION

Psycholinguistics provides a lot of the foundational work on these questions.

- ***Communication is a dynamic social process***: the partner often to signal missing/misunderstood informations
- Repairing is generally less costly than avoiding ambiguities in the first place
- You only ever need to reach the ***grounding criterion***, ie ***enough*** mutual understanding for the task
- $\Rightarrow$ we typically only reach ***partial (or surface) alignment*** – full alignment is usually not required

[Clark and Wilkes-Gibbs, ***Referring as a collaborative process***, Cognition 1986
Pickering and Garrod, ***Alignment as the basis for successful communication***, Research on Language and Computation 2006]

IN SOCIAL HUMAN-ROBOT INTERACTION

Well studied in communication (cf back-channeling)

Can we expand this line of thought to sHRI in general?

Most of our social and behavioural alignment comes from sub-conscious social mechanisms:

- entrainment (coupling),
- mimicry,
- implicit turn-taking,
- joint attention
- ...and others

IN SOCIAL HUMAN-ROBOT INTERACTION

- These mechanisms are unfortunately often ill-defined, and particularly difficult to turn into equations (or controllers, in our case)
- no close-form equation of social interactions $\Rightarrow$ data-driven approaches?

TOWARDS THE DATA-DRIVEN STUDY OF SOCIAL DYNAMICS

DATA!

If we want to use machine learning, we need data.

DATA!

If we want to use machine learning, we need data.

...a task that exhibits:

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

DATA!

If we want to use machine learning, we need data.

...a task that exhibits:

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

while being...

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

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**

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**

- focus on children
- with a little bit of scaffolding & framing

SOCIAL DYNAMICS TO BE OBSERVED

- the **task engagement**

  are the participants 'on task' or not?

SOCIAL DYNAMICS TO BE OBSERVED

- 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?

SOCIAL DYNAMICS TO BE OBSERVED

- 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...

SOCIAL DYNAMICS TO BE OBSERVED

- 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

SOCIAL DYNAMICS TO BE OBSERVED

- 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

SOCIAL DYNAMICS TO BE OBSERVED

- 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**

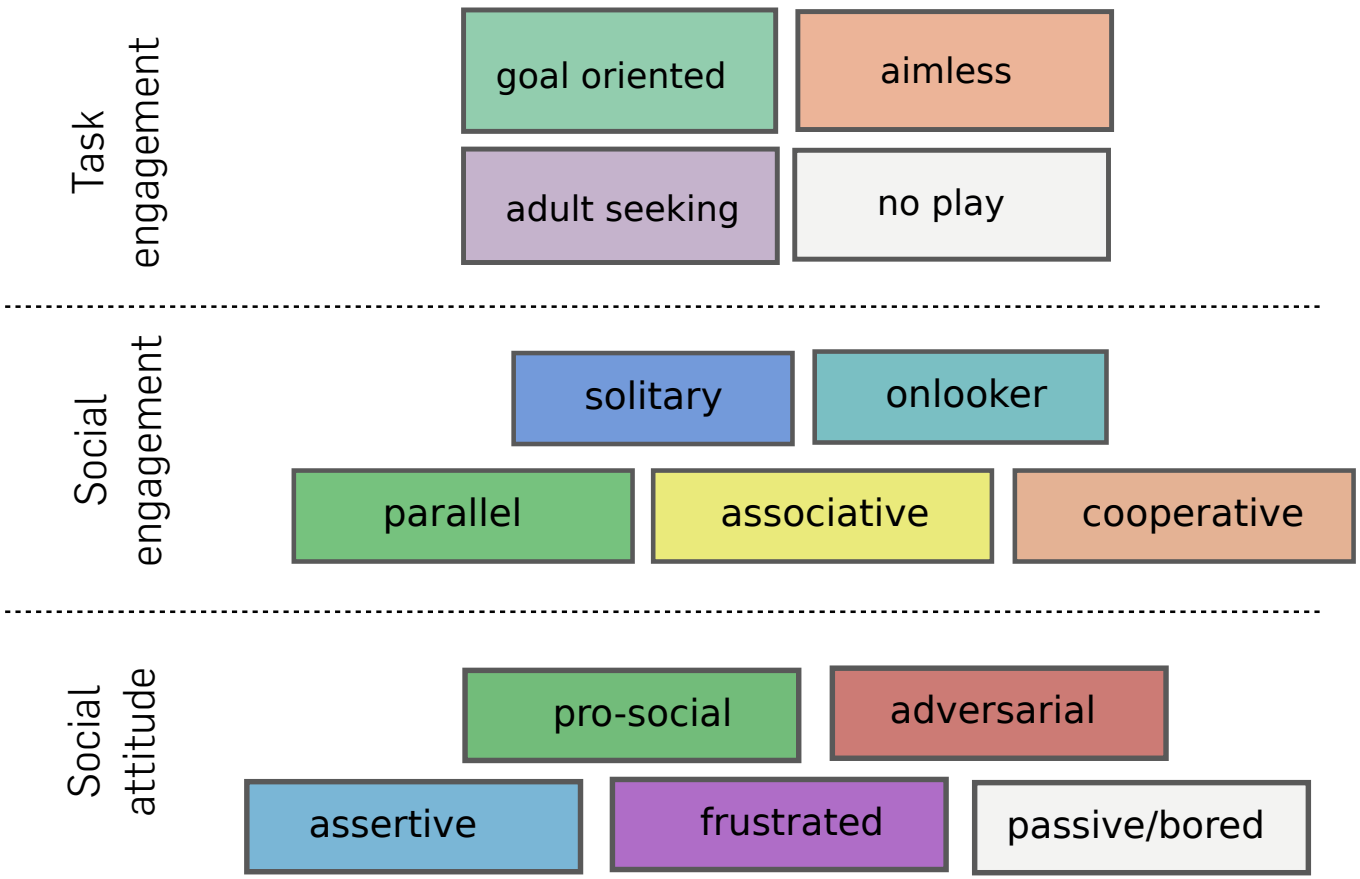
THE PINSORO DATASET

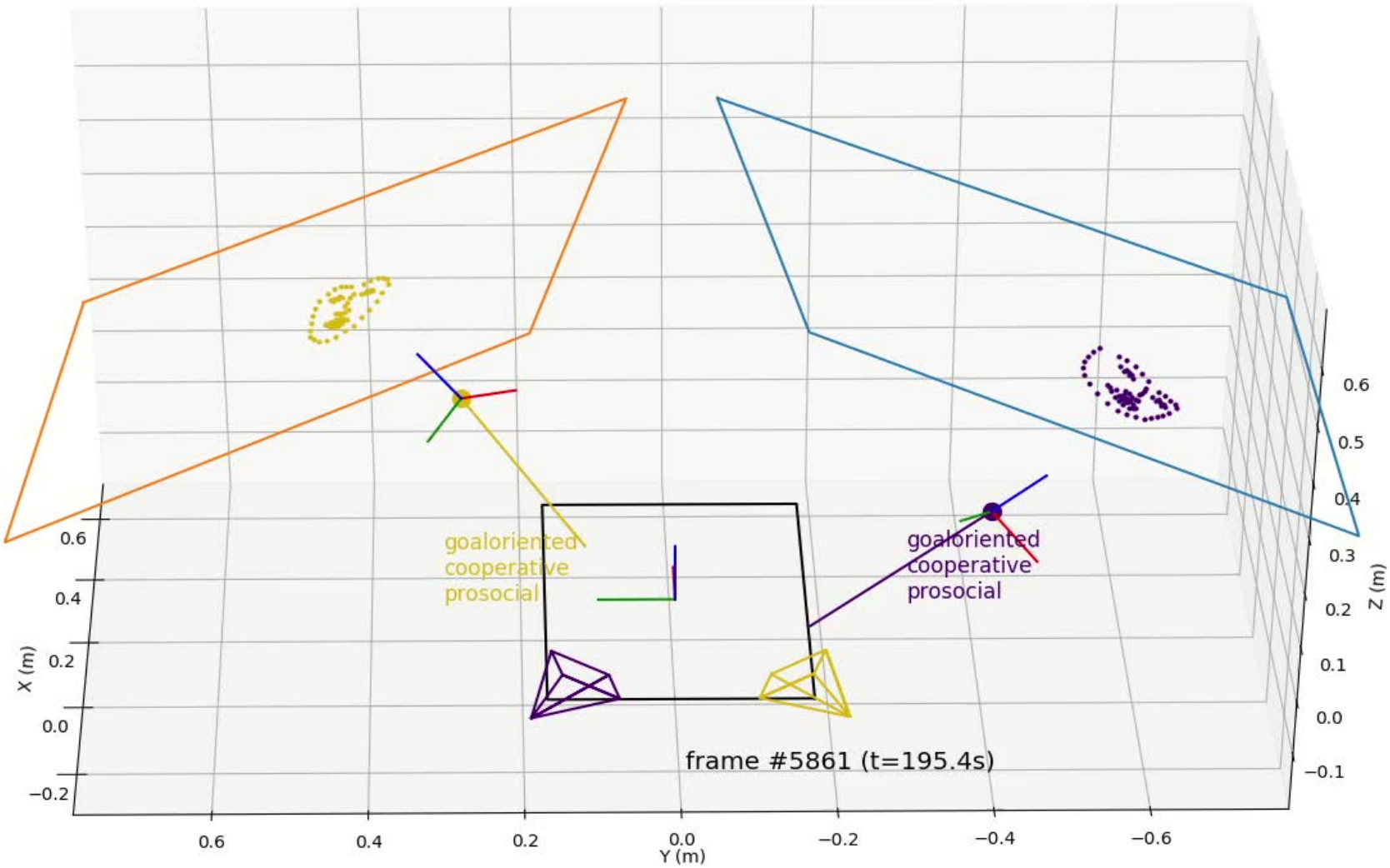
- 120 children, 4 to 8 years old
- 75 interactions

  90 children playing with another child,
  30 playing with a robot
- About 45h+ of recordings; 2M+ frames; ≈ 2TB
- average duration of freeplay interactions: 24min in child-child condition; 19min in child-robot condition

[Lemaignan et al. **The PInSoRo dataset: Supporting the data-driven study of [...]** PLOS One 2018]

13000+ ANNOTATIONS




goaloriented
cooperative
prosocial
goaloriented
cooperative
prosocial
frame #5861 (t=195.4s)
X (m)
Y (m)
Z (m)
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.5
0.3
0.1
−0.1

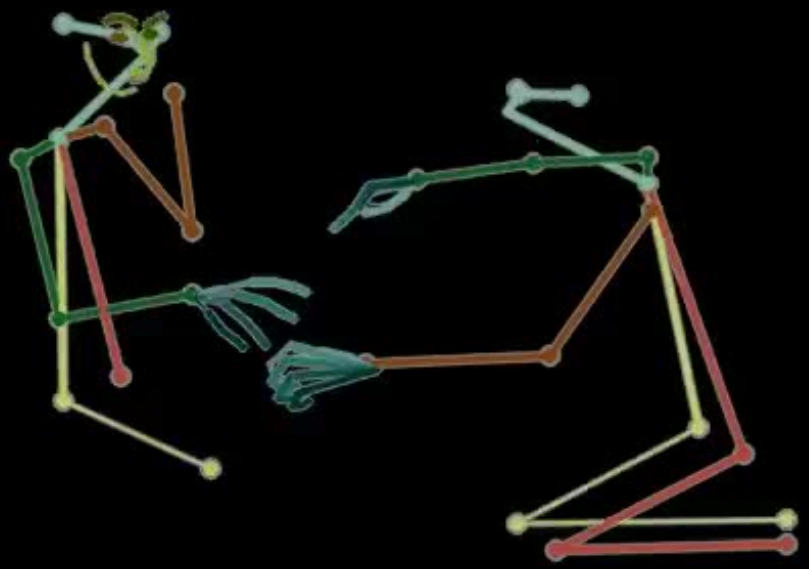
LARGEST OPEN DATASET OF NATURAL SOCIAL INTERACTIONS

Anonymised version (7.2GB) available on-line.
Grab it now!

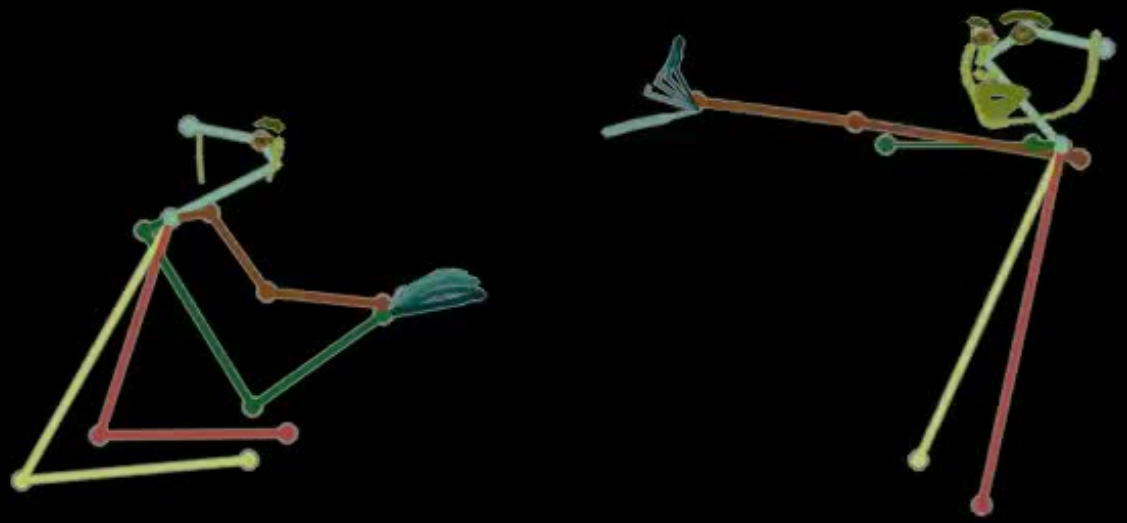
freeplay-sandbox.github.io

Open data! Hosted on EU's zenodo

CASE STUDY: UNDERSTANDING SOCIAL KINEMATICS




Focused!
ILLUMINATED POLITICAL
& PHYSICAL




open me

200 participants, 4 clips each, on MTurk

For each construct, calculate Δ and Σ

EFA: EXPLORATORY FACTOR ANALYSIS

	Factor 1 *full-scene*	**Factor 2** *full-scene*	**Factor 3** *full-scene*
Δ Sad	0.41		
Σ Sad		0.72	
Δ Happy	0.49		
Σ Happy			-0.55
Δ Angry	0.40		
Σ Angry		0.81	
Δ Excited	0.53		
Σ Excited			-0.71
Δ Calm	0.45		
Σ Calm			
Δ Friendly	0.69		
Σ Friendly			-0.43
Δ Aggressive	0.78		
Σ Aggressive		0.80	-0.36
Δ Engaged			0.65
Σ Engaged			-0.64
Δ Distracted			0.65
Σ Distracted		0.63	
Δ Bored			0.61
Σ Bored		0.58	0.48
Δ Frustrated	0.53		
Σ Frustrated		0.70	
Δ Dominant	0.75		
Σ Dominant		0.53	
Δ Submissive	0.68		
Σ Submissive		0.54	

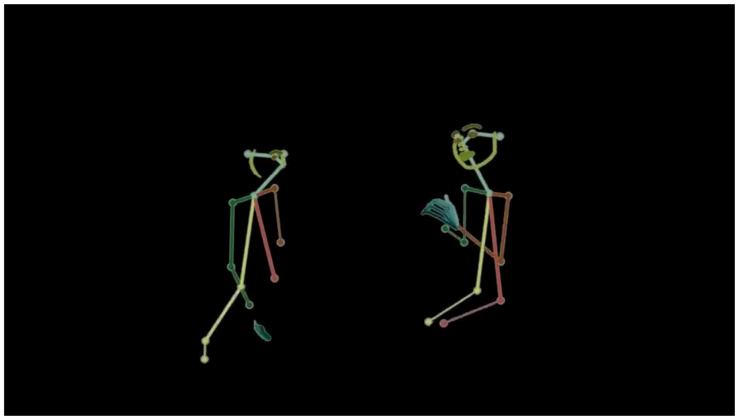
EFA: EXPLORATORY FACTOR ANALYSIS

	Factor 1: imbalance *full-scene*	**Factor 2: (negative) valence** *full-scene*	**Factor 3: engagement** *full-scene*
Δ Sad	0.41		
Σ Sad		0.72	
Δ Happy	0.49		
Σ Happy			-0.55
Δ Angry	0.40		
Σ Angry		0.81	
Δ Excited	0.53		
Σ Excited			-0.71
Δ Calm	0.45		
Σ Calm			
Δ Friendly	0.69		
Σ Friendly			-0.43
Δ Aggressive	0.78		
Σ Aggressive		0.80	-0.36
Δ Engaged			0.65
Σ Engaged			-0.64
Δ Distracted			0.65
Σ Distracted		0.63	
Δ Bored			0.61
Σ Bored		0.58	0.48
Δ Frustrated	0.53		
Σ Frustrated		0.70	
Δ Dominant	0.75		
Σ Dominant		0.53	
Δ Submissive	0.68		
Σ Submissive		0.54	

EFA: EXPLORATORY FACTOR ANALYSIS

	Factor 1: imbalance		Factor 2: (negative) valence		Factor 3: engagement	
	full-scene	*mov.-alone*	*full-scene*	*mov.-alone*	*full-scene*	*mov.-alone*
Δ Sad	0.41	0.52				
Σ Sad			0.72	0.53		0.49
Δ Happy	0.49	0.53				
Σ Happy				-0.51	-0.55	
Δ Angry	0.40	0.62				
Σ Angry			0.81	0.85		
Δ Excited	0.53	0.63				
Σ Excited					-0.71	
Δ Calm	0.45	0.63				
Σ Calm				-0.45		
Δ Friendly	0.69	0.56				
Σ Friendly				-0.60	-0.43	
Δ Aggressive	0.78	0.79				
Σ Aggressive			0.80	0.72	-0.36	
Δ Engaged		0.39			0.65	0.52
Σ Engaged					-0.64	-0.64
Δ Distracted					0.65	0.63
Σ Distracted			0.63			0.82
Δ Bored		0.44			0.61	0.54
Σ Bored			0.58		0.48	0.83
Δ Frustrated	0.53	0.61				
Σ Frustrated			0.70	0.69		
Δ Dominant	0.75	0.81				
Σ Dominant			0.53	0.52		
Δ Submissive	0.68	0.72				
Σ Submissive			0.54			

THREE CONSTRUCTS TO RULE THEM ALL



Interaction imbalance

Interaction valence

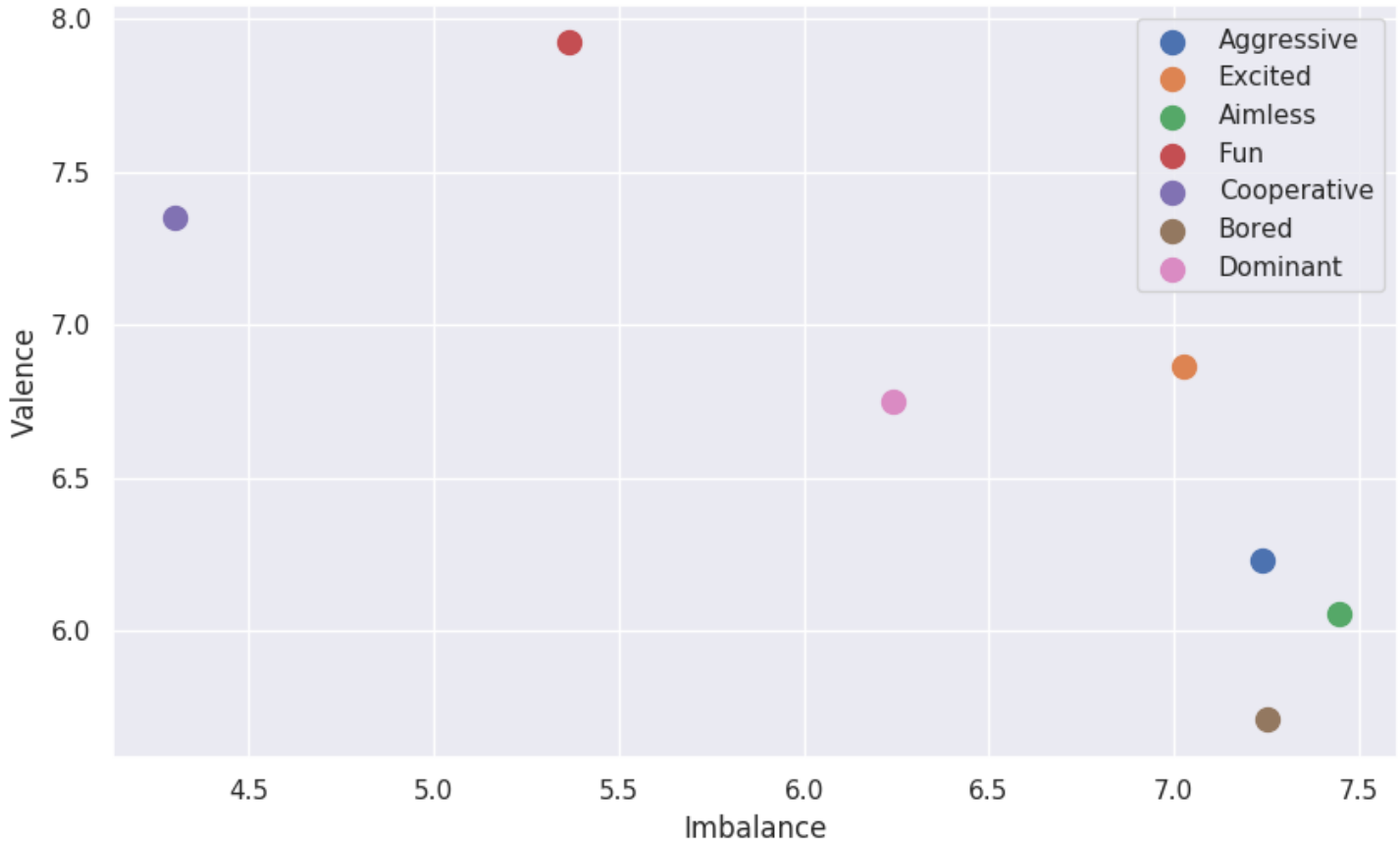
Engagement

MEAN EFA PROJECTION OF CLIPS PER SOCIAL SITUATION

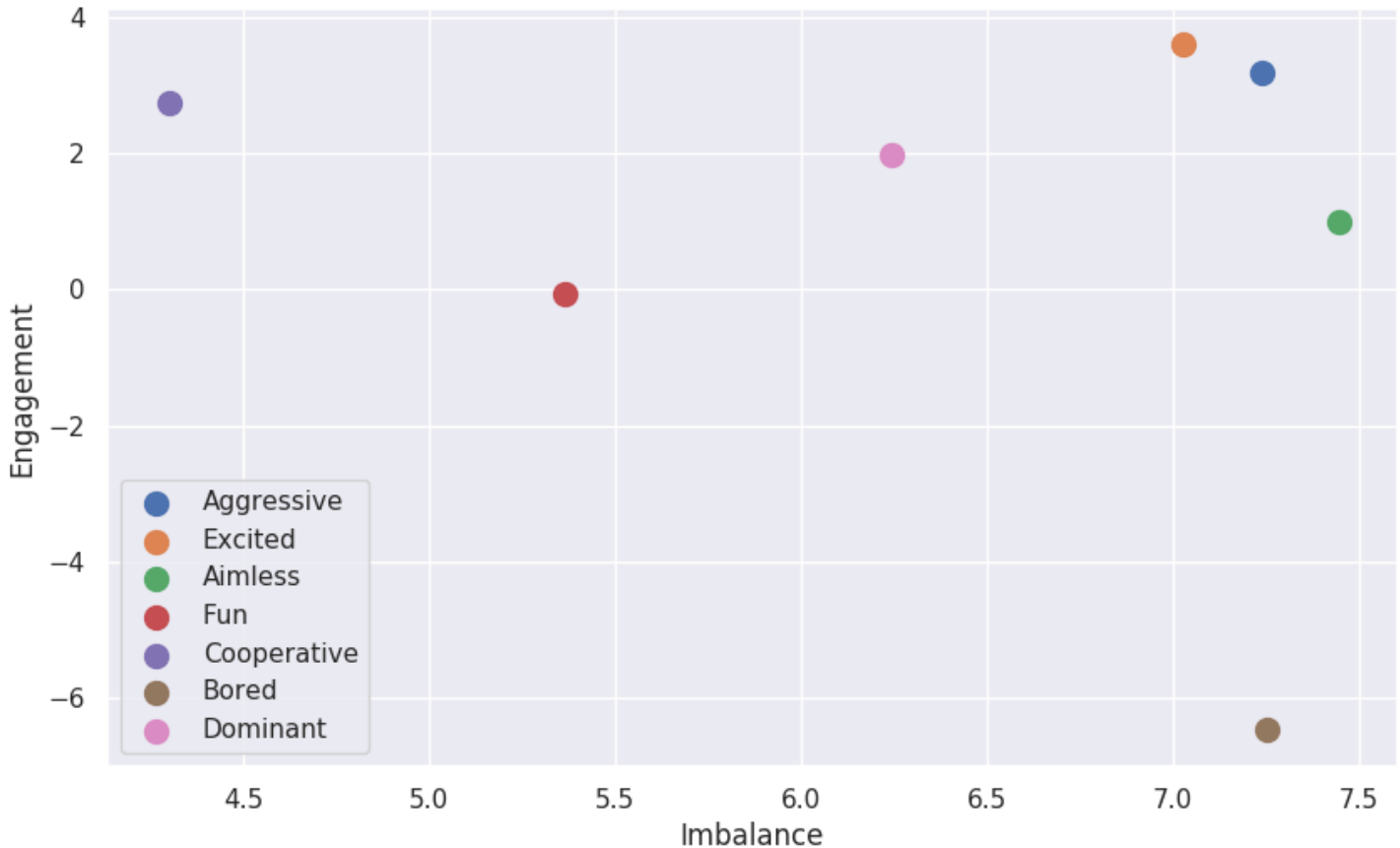
The 20 clips were labelled after their salient social features (*aggressive*, *excited*, *aimless*, *fun*, *cooperative*, *bored*, *dominant*).

What happen if we project the ratings for 'aggressive' clips, 'excited' clips, etc. onto the 3 EFA factors?

MEAN EFA PROJECTION OF CLIPS PER SOCIAL SITUATION



MEAN EFA PROJECTION OF CLIPS PER SOCIAL SITUATION



WHAT NEXT?


Attitude: passive
Social engag.: onlooker
Task engag.: no play
Attitude: passive
Social engag.: solitary
Task engag.: goal oriented

ULTIMATELY...

Real-time identification by the robot of...

- the **task engagement**

  is my partner 'on task' or not?

ULTIMATELY...

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?

ULTIMATELY...

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...

ULTIMATELY...

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

ULTIMATELY...

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

ULTIMATELY...

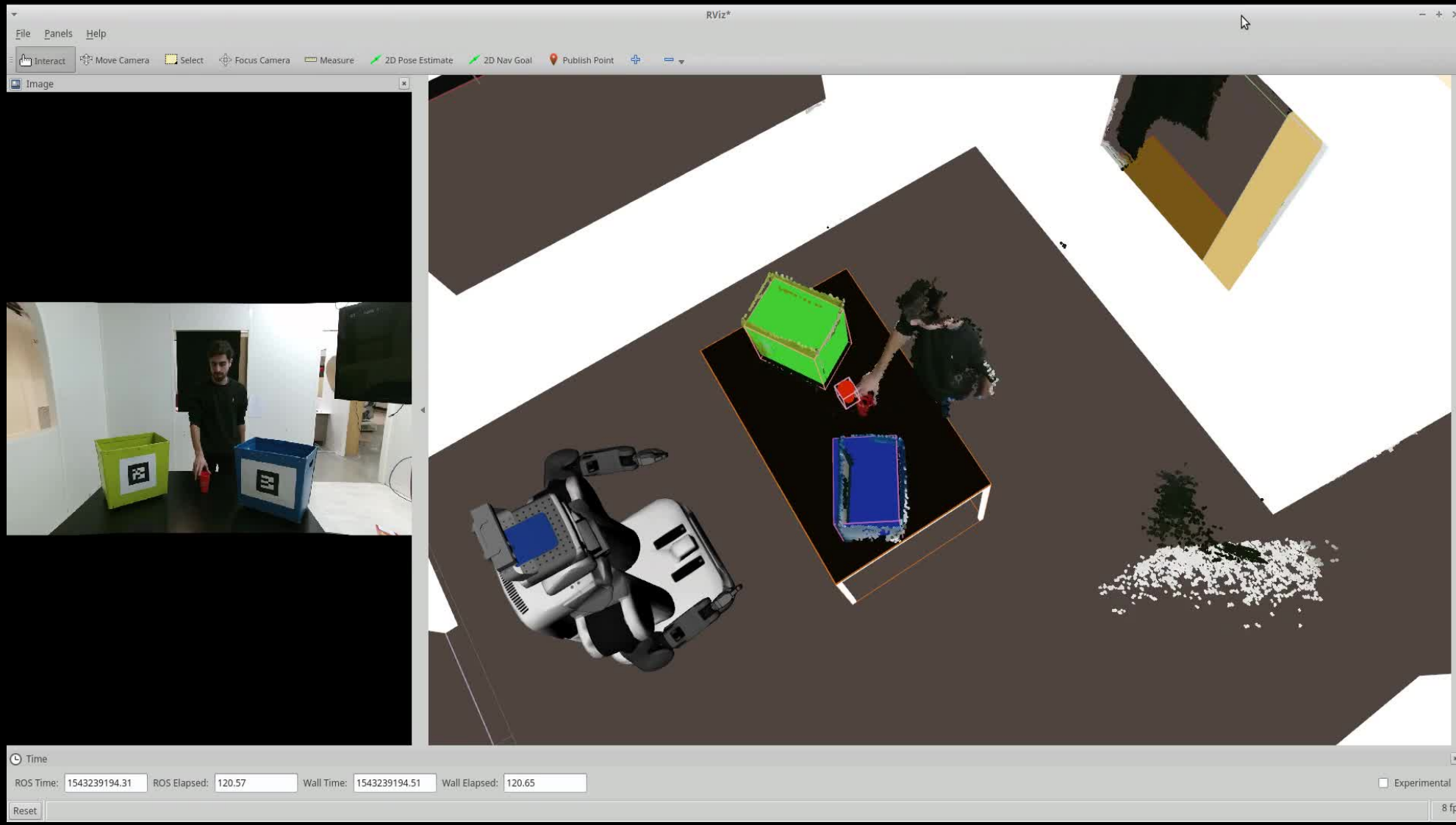
Real-time identification by the robot of...

- 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

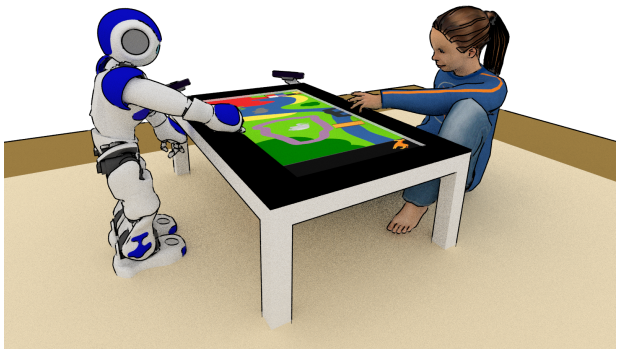
Social behaviours; Social dynamics: **generation as well!**

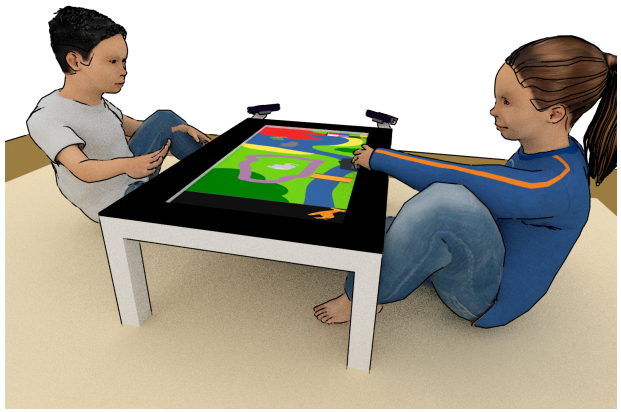
Thank you!

SOME MORE STUFF


RViz*
File
Panels
Help
Interact
Move Camera
Select
Focus Camera
Measure
2D Pose Estimate
2D Nav Goal
Publish Point
Image
Time
ROS Time: 1543239194.31
ROS Elapsed: 120.57
Wall Time: 1543239194.51
Wall Elapsed: 120.65
Experimental
Reset

TWO BASELINES



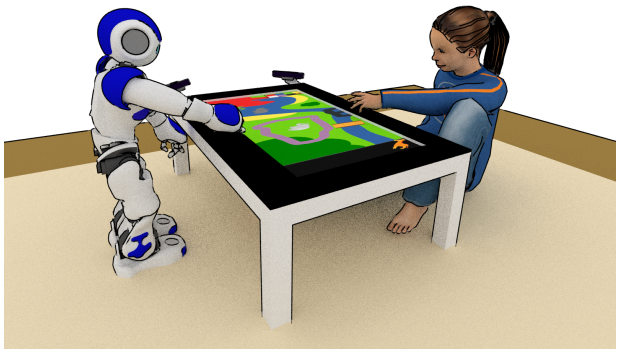


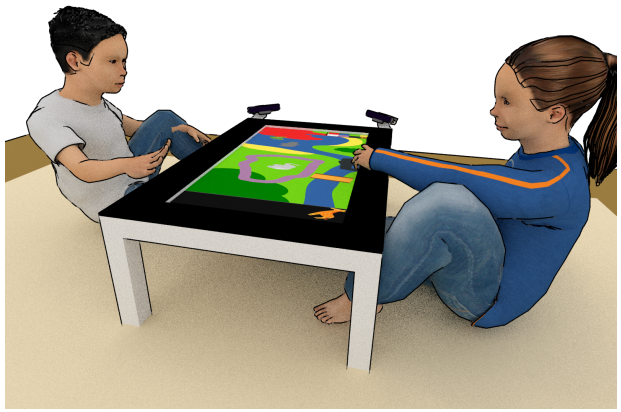
child - non-social robot

child - child

richness of social interactions

TWO BASELINES





child - non-social robot

your next socio-cognitive model

child - child

richness of social interactions

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

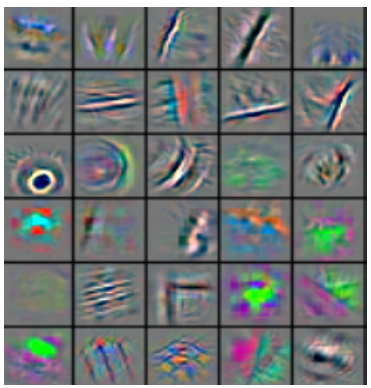
DBSoC: Deep Behavioural Social Cloning – LfD + CNNs + LSTM

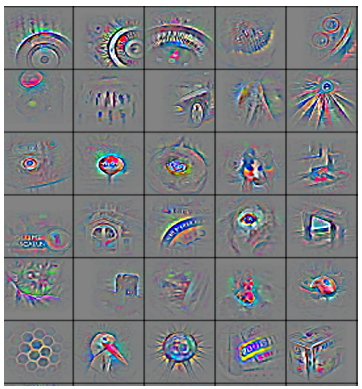
Two tasks for a telepresence robot:

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

DEEP NETWORKS ≡ BLACK BOXES?

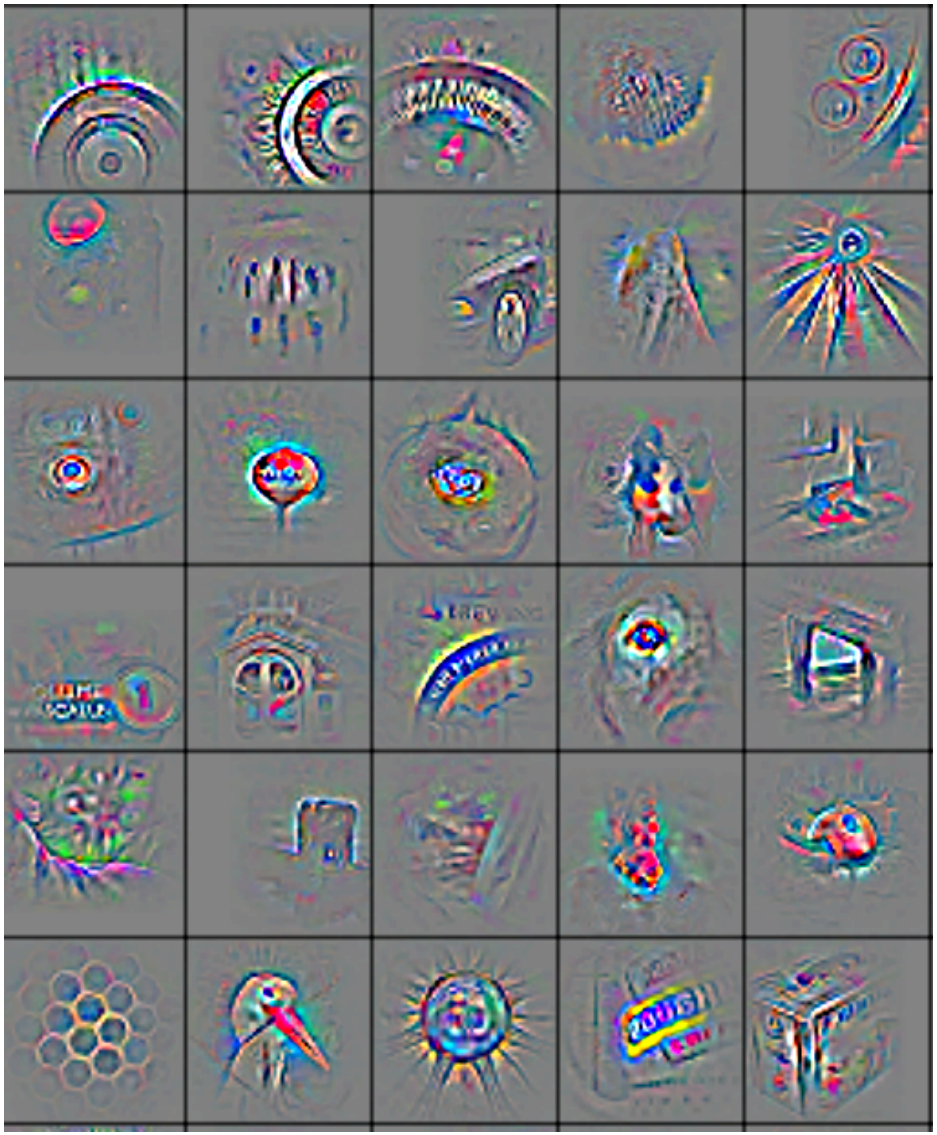
→ Low-Level features → Mid-Level features → High-Level features → Trainable classifier





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

DEEP NETWORKS ≡ BLACK BOXES?



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

WHAT DID WE RECORD?

Domain	Type	Details
child × 2	audio	16kHz, mono, semi-direction
	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	
+ post-process	optical flow, audio features	
	facial action units...	