

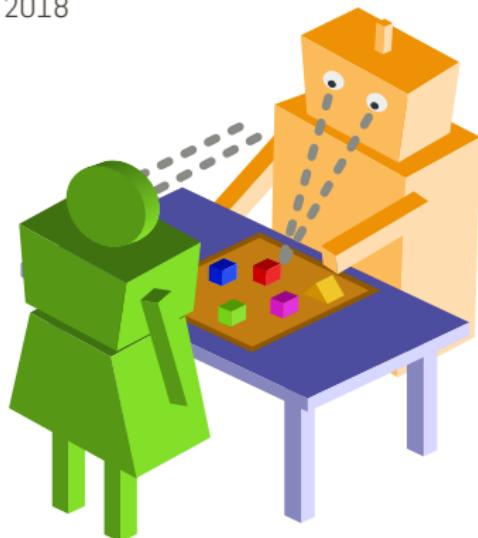
learning & child-robot interaction

a glimpse into the next challenges

Symposium on Robots for Language Learning – 13 Dec. 2018

Séverin Lemaignan

Bristol Robotics Lab University of the West of England





This presentation is released under the terms of the
Creative Commons Attribution-Share Alike license.

You are free to reuse it and modify it as much as you want as long as
(1) you mention me as being the original author,
(2) you re-share your presentation under the same terms.

You can download the sources of this presentation here:
github.com/severin-lemaignan/l2tor-symposium2018-technical-challenges-cri

WHAT CHALLENGES ARE WE TALKING
ABOUT, EXACTLY?

What about some math now?

Yeah, yeah...

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

Speech recognition

Individual webcams

Imitative
learning

Embedded sensors
(Arduino)

Subjective
data

Microphones, Kinect II

SURFACE FUNCTIONS FOR SOCIAL COGNITION

WITH
**PLYMOUTH
UNIVERSITY**



FINE, BUT MAYBE WE DO NOT REALLY
NEED SOCIAL ROBOTS?

What challenges are we talking about, exactly?

Social interactions?

A horizontal sequence of 20 circles, starting with a black circle followed by 19 white circles.

Learning social behav

○○○○○○○○○○

SOCIAL OR NOT SOCIAL?



Non-social

Interaction spectrum

Social

What challenges are we talking about, exactly?

Social interactions?

Learning social behav

○○○○○○○○○○

SOCIAL OR NOT SOCIAL?



Cellulo



CoWriter

What challenges are we talking about, exactly?

Social interactions?

Learning social behav

NON-SOCIAL INTERACTION

What is the most effective learning tool in a classroom?

What challenges are we talking about, exactly?

ooo

Social interactions?

ooo●oooooooooooooooooooo

The challenge of "doing together"

ooooooooooo

Learning social beha

ooooooooooo

NON-SOCIAL INTERACTION

What is the most effective learning tool in a classroom?





Pens and paper are pervasive...



WORDMANIA



...so should robots

CELLULO: DESIGN PRINCIPLES



- **ubiquitous:** a pervasive yet unremarkable tool that blend into the daily learning routine; has to be trustworthy (i.e. reliable), readily replaceable (i.e. cheap, no affective bonding), intuitive (i.e. few simple affordances)

CELLULO: DESIGN PRINCIPLES



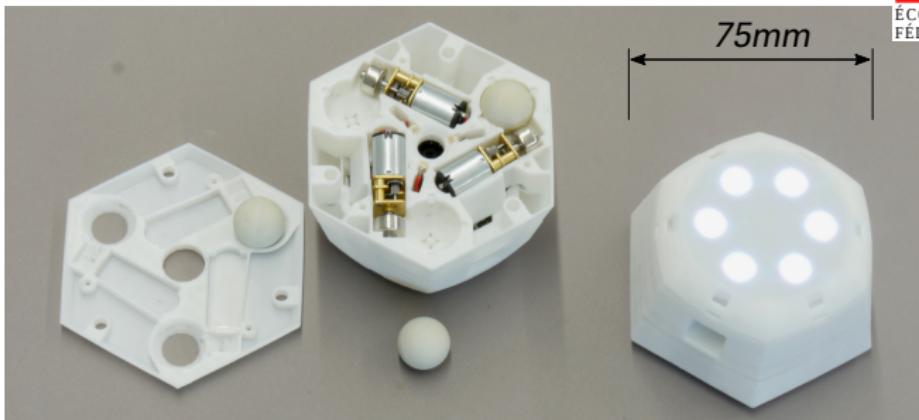
- **ubiquitous:** a pervasive yet unremarkable tool that blend into the daily learning routine; has to be trustworthy (i.e. reliable), readily replaceable (i.e. cheap, no affective bonding), intuitive (i.e. few simple affordances)
- **versatile:** applicable to a broad range of learning scenarios; the robots' hardware, appearance and interaction modalities must not imply or be constrained to specific use cases

CELLULO: DESIGN PRINCIPLES



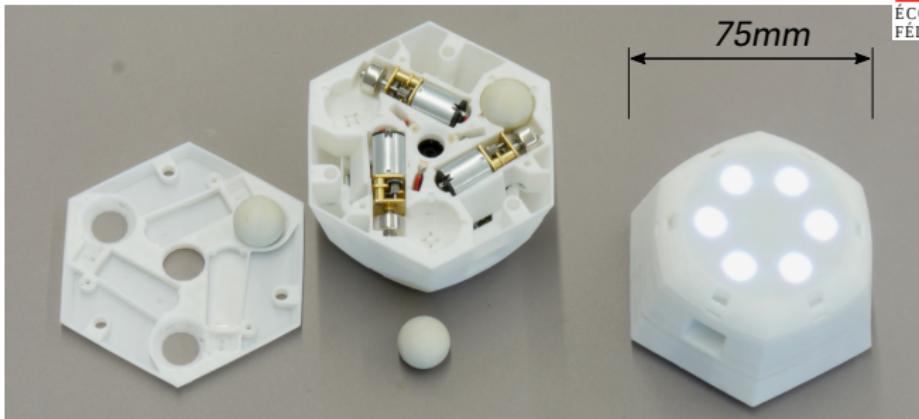
- **ubiquitous:** a pervasive yet unremarkable tool that blend into the daily learning routine; has to be trustworthy (i.e. reliable), readily replaceable (i.e. cheap, no affective bonding), intuitive (i.e. few simple affordances)
- **versatile:** applicable to a broad range of learning scenarios; the robots' hardware, appearance and interaction modalities must not imply or be constrained to specific use cases
- **practical:** to gain field acceptance in the classrooms, educative robots must critically represent a net educative gain and must not incur higher workload for the teachers

CELLULO: HARDWARE



- o Holonomic motion
- o Sub-mm absolute localisation (no external hardware)
- o Haptic feedback + tactile RGB LED buttons
- o Bluetooth

CELLULO: HARDWARE



- o Holonomic motion
- o Sub-mm absolute localisation (no external hardware)
- o Haptic feedback + tactile RGB LED buttons
- o Bluetooth
- o Affordable (prototype: €125)



INTERACTION WITH THE PAPER



Critically, Cellulo is meant as an **interaction between (classroom-friendly) paper and the robots**.

INTERACTION WITH THE PAPER



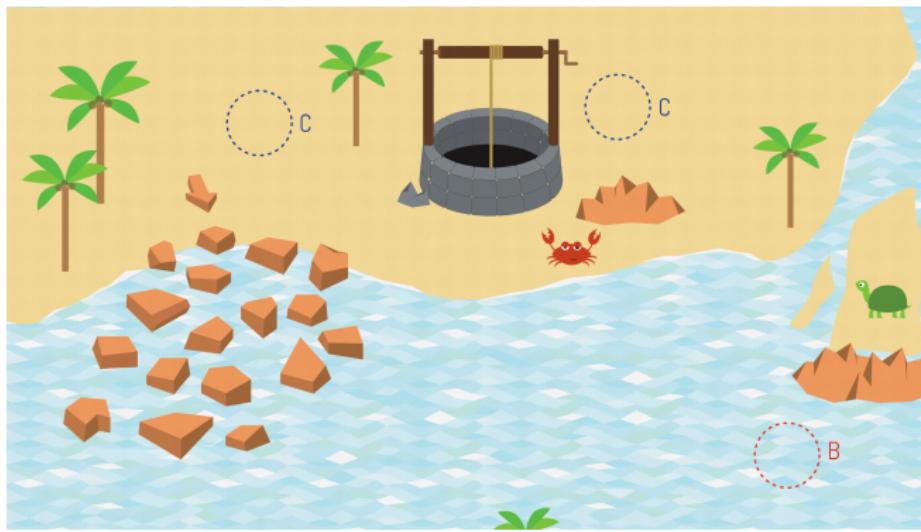
Critically, Cellulo is meant as an **interaction between (classroom-friendly) paper and the robots.**

Achieved through a **paper-based absolute localisation system**

INTERACTION WITH THE PAPER

Critically, Cellulo is meant as an **interaction between (classroom-friendly) paper and the robots**.

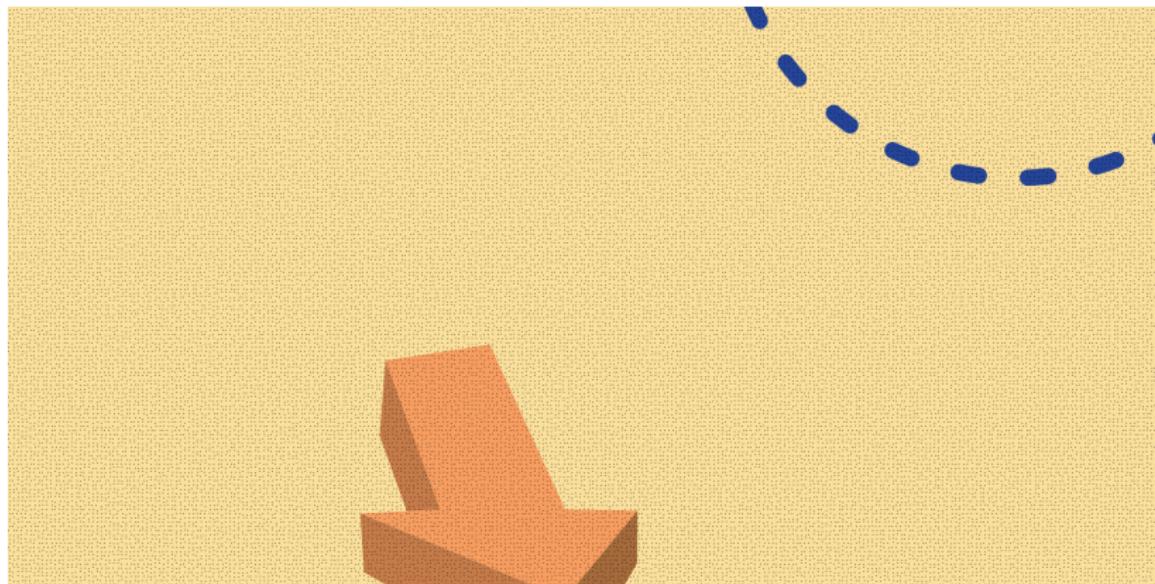
Achieved through a **paper-based absolute localisation system**



INTERACTION WITH THE PAPER

Critically, Cellulo is meant as an **interaction between (classroom-friendly) paper and the robots**.

Achieved through a **paper-based absolute localisation system**



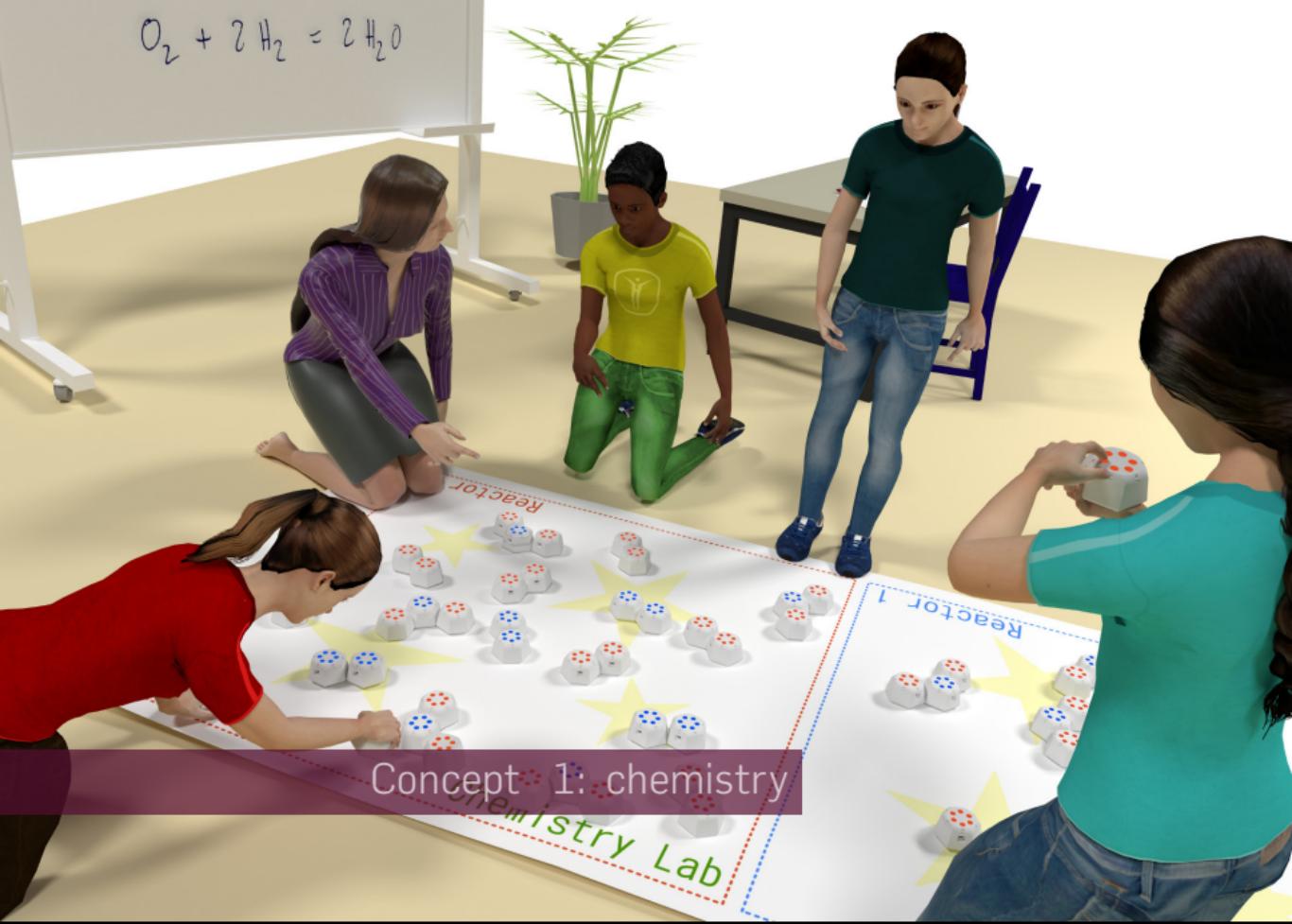
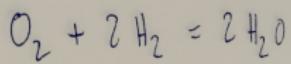
INTERACTION WITH THE PAPER



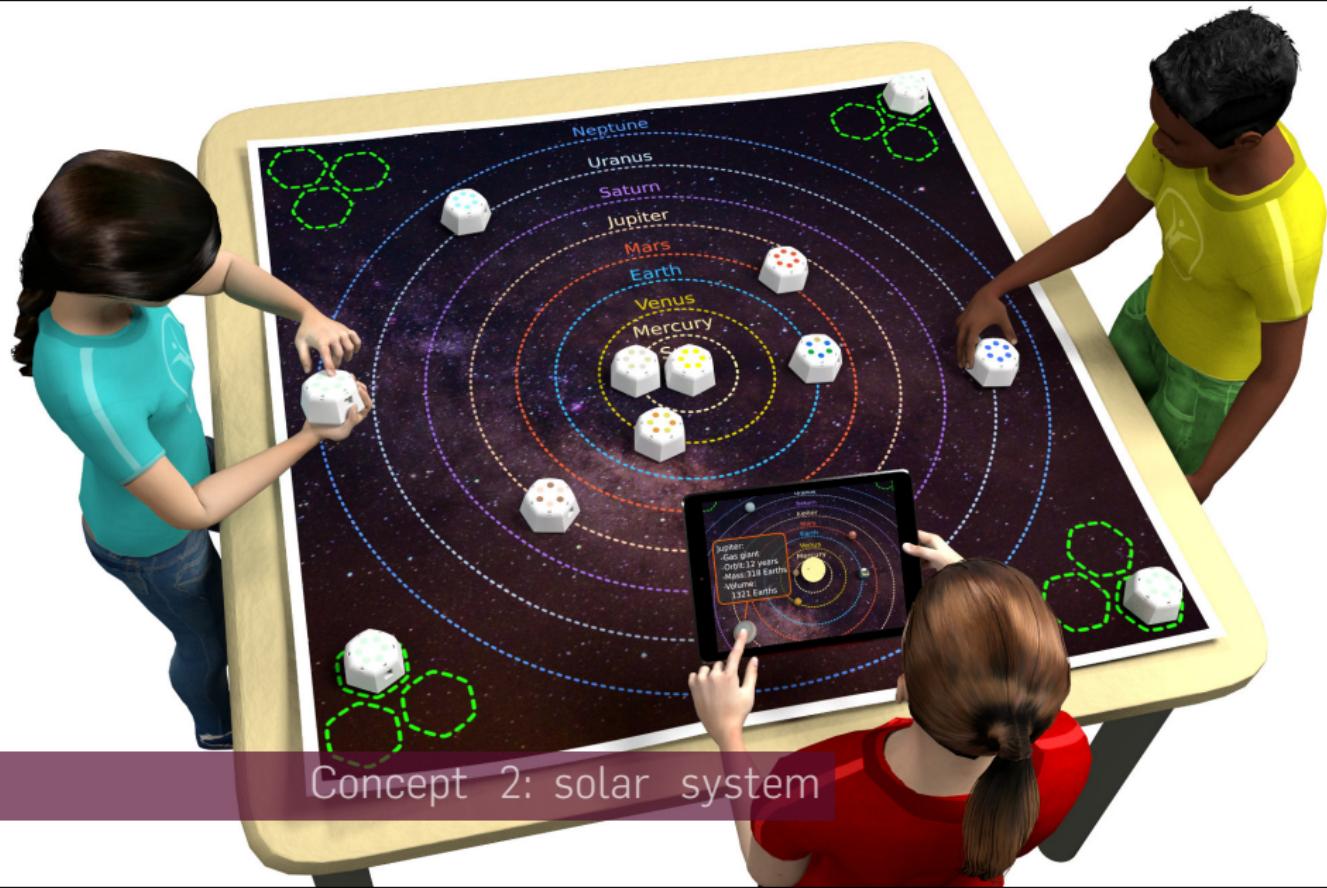
Critically, Cellulo is meant as an **interaction between (classroom-friendly) paper and the robots**.

Achieved through a **paper-based absolute localisation system**

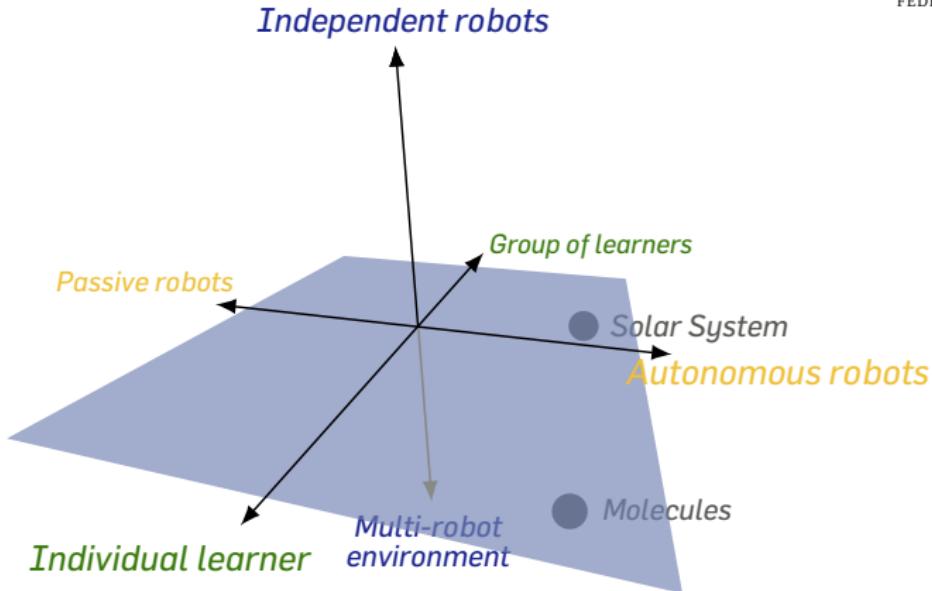
- even more than 'classroom-friendly', paper is 'teacher-friendly'
- easy to manipulate, copy, print, cutout, dispose...
- unique activity IDs: drop the robots onto the sheet, it recognizes the activity



Concept 2: solar system



INTERACTION DESIGN SPACE



...at the other end of the spectrum...

TECHNICAL CHALLENGES



- Get a child-proof robot to write...

TECHNICAL CHALLENGES



- o Get a child-proof robot to write...
- o ...badly...

TECHNICAL CHALLENGES



- Get a child-proof robot to write...
- ...badly...
- Make it able to learn...

TECHNICAL CHALLENGES



- Get a child-proof robot to write...
- ...badly...
- Make it able to learn...
- ...with the help of children

What challenges are we talking about, exactly?

ooo

Social interactions?

oooooooooooo●oooooo

The challenge of "doing together"

oooooooooooo

Learning social beha

oooooooooooo

THE COWRITER PROJECT



Can we address children' hand-writing impairments with robots?

THE COWRITER PROJECT



Can we address children' hand-writing impairments with robots?

- Robots do not know how to write!

THE COWRITER PROJECT



Can we address children' hand-writing impairments with robots?

- Robots do not know how to write!
- Learning by Teaching

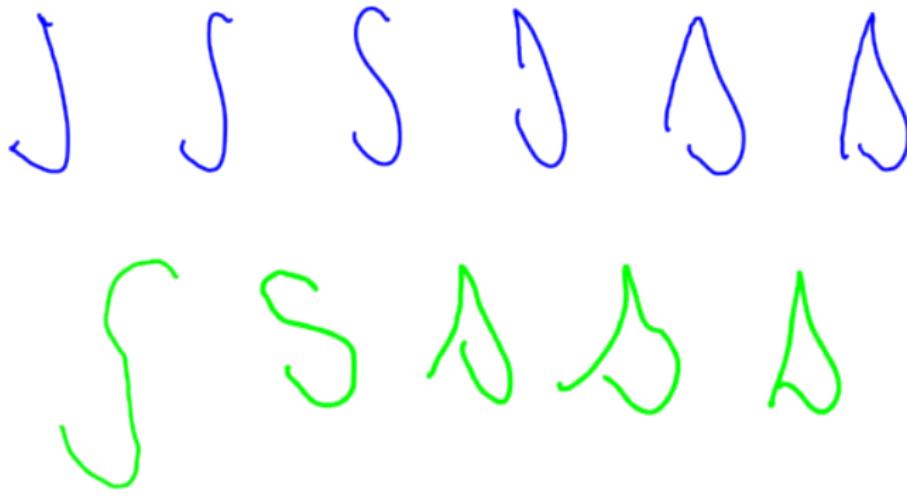
THE COWRITER PROJECT



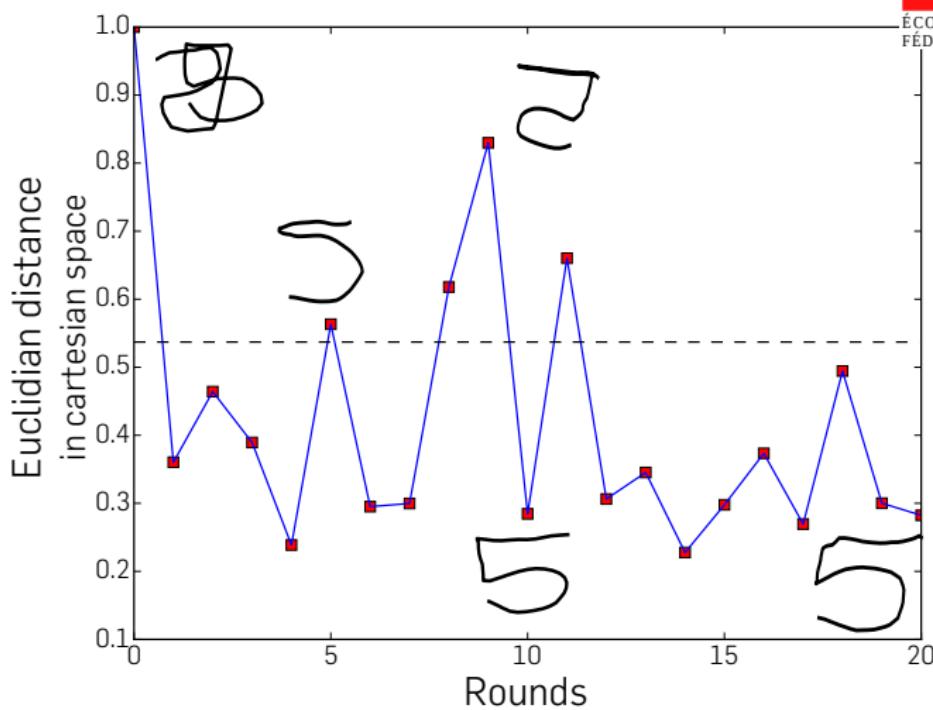
Can we address children' hand-writing impairments with robots?

- Robots do not know how to write!
- Learning by Teaching
- (nice side-effect: we can adapt to each child and each disabilities)

LEARNING FROM DEMONSTRATION



LEARNING TO DRAW A 5





BEFORE – AFTER



salut mimi
nous persons
que c'est un
corps
est ce que tu penses
croire en des
photos de
la lune

BEFORE – AFTER

salut mimi
nous pensons
que c'est un
corps
est ce que tu peux
croire des
photos de
la base

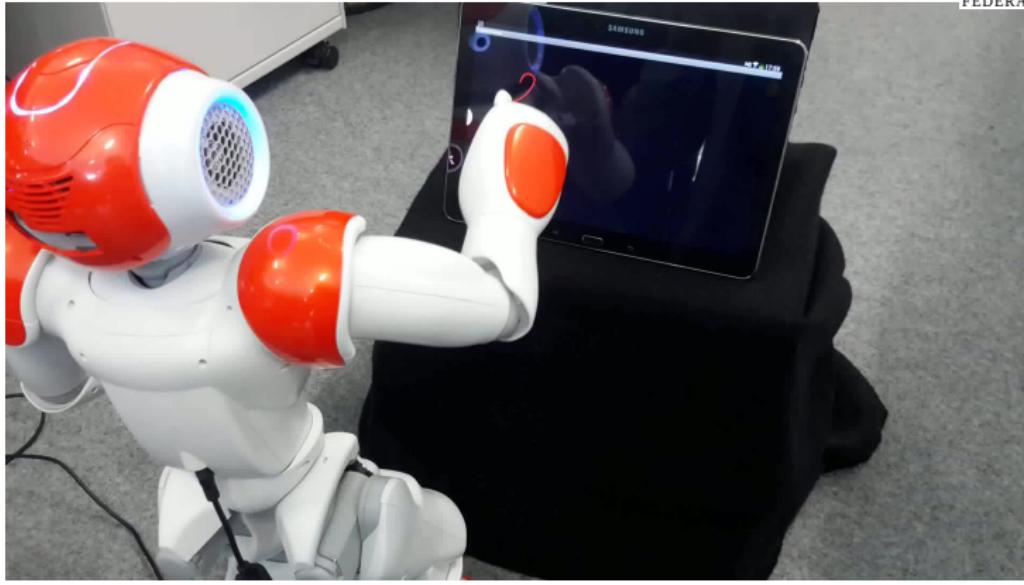
salut mimi
nous pensons
que c'est un
corps
est ce que tu peux
envoyer des
photos de
la base

BEFORE – AFTER

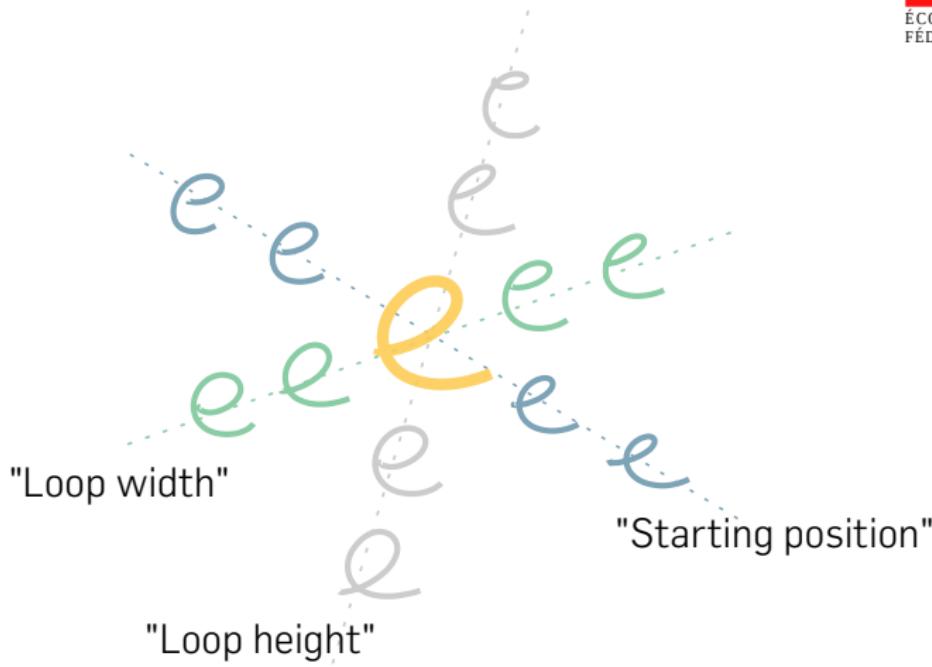
salut mimi
nous pensons
que c'est un
corps
et que tu peux
croire des
photos de
la base

salut mimi
nous pensons
que c'est un
corps
est ce que tu peux
envoyer des
photos de
la base

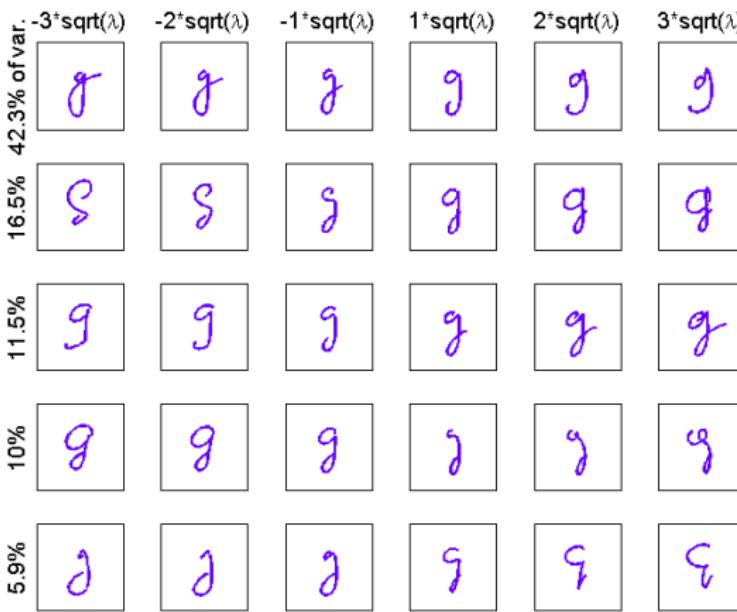
COWRITER IMPLEMENTATION



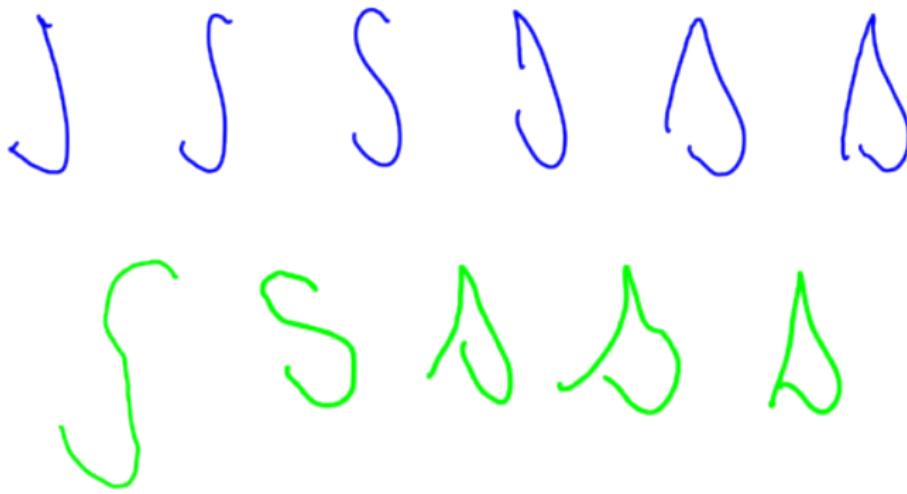
COWRITER IMPLEMENTATION

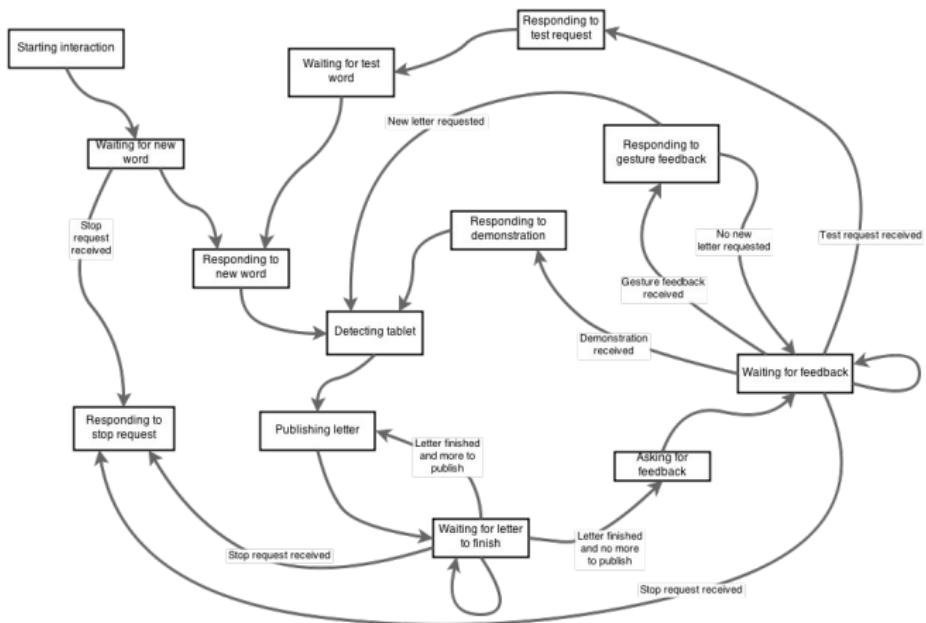


COWRITER IMPLEMENTATION



COWRITER IMPLEMENTATION





THE CHALLENGE OF “DOING TOGETHER”



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

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

⇒ **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

⇒ **explicit cognitive steps**

...hard ones, though:

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

What challenges are we talking about, exactly?

ooo

Social interactions?

oooooooooooooooooooo

The challenge of "doing together"

ooo●oooo

Learning social behav

oooooooooooo

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?

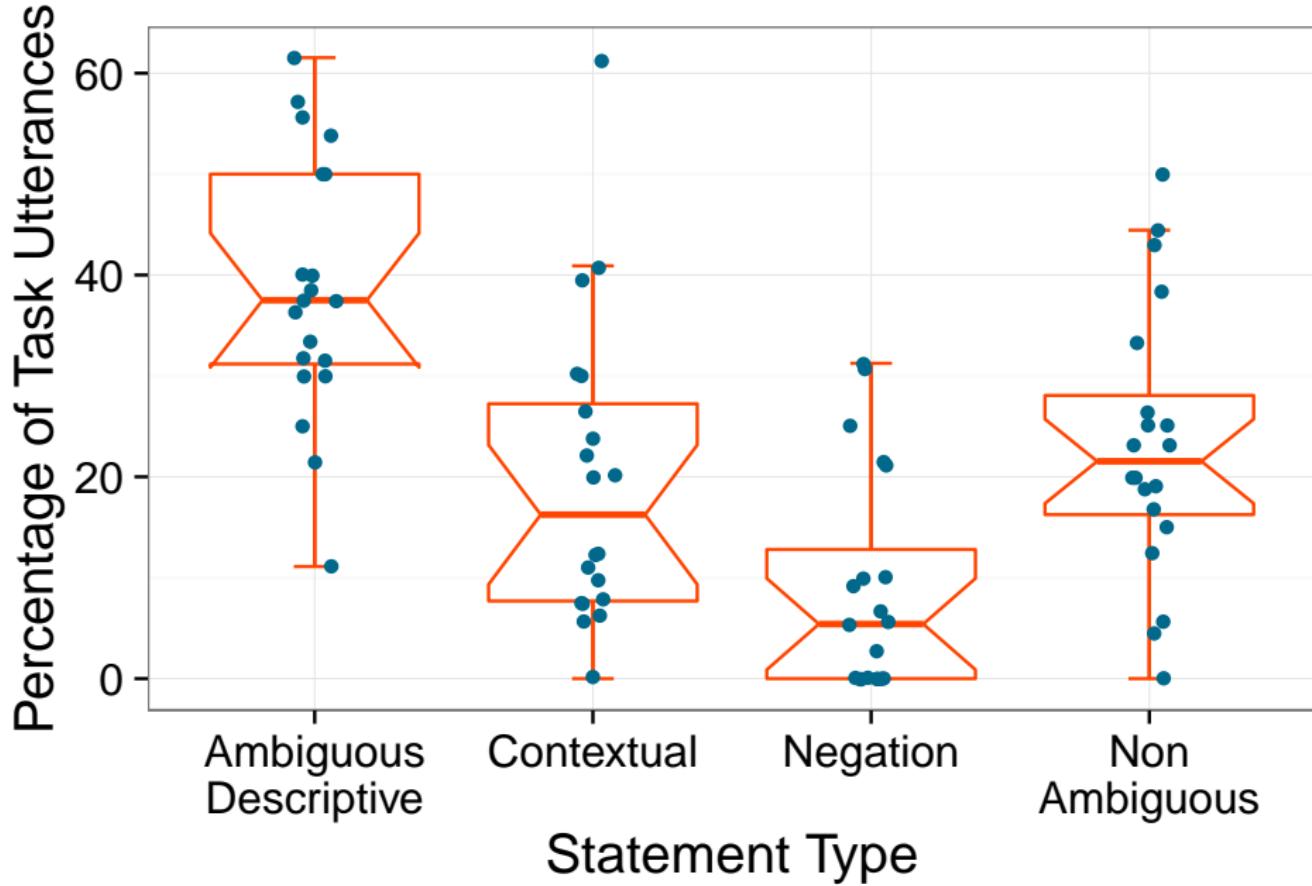
ONE EXAMPLE: GROUNDING OF SPATIAL LANGUAGE



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 tries to signal missing/misunderstood informations

SURFACE ALIGNMENT; GROUNDING CRITERION

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

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

SURFACE ALIGNMENT; GROUNDING CRITERION

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

- *Communication is a dynamic social process*: the partner often tries 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

SURFACE ALIGNEMENT; GROUNDING CRITERION

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

- *Communication is a dynamic social process*: the partner often tries 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
- ⇒ we typically only reach *partial (or surface) alignment* – full alignment is usually not required

SURFACE ALIGNEMENT; GROUNDING CRITERION

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

- *Communication is a dynamic social process*: the partner often tries 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
- ⇒ we typically only reach *partial (or surface) alignment* – full alignment is usually not required
- **do we need a high-level Theory of Mind at all?**

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

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

Can we model & generate them?

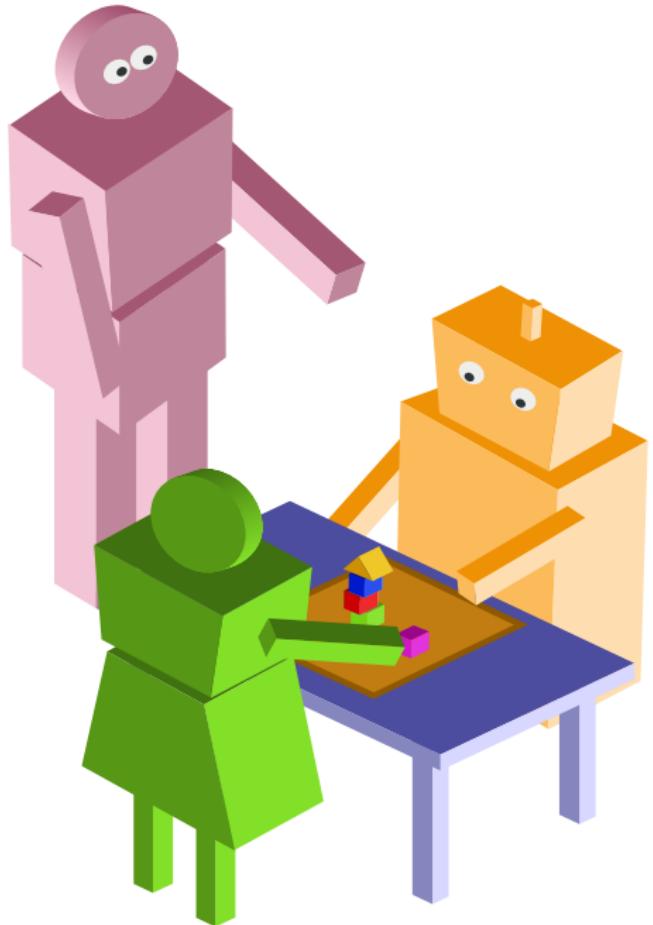
THE PROBLEM

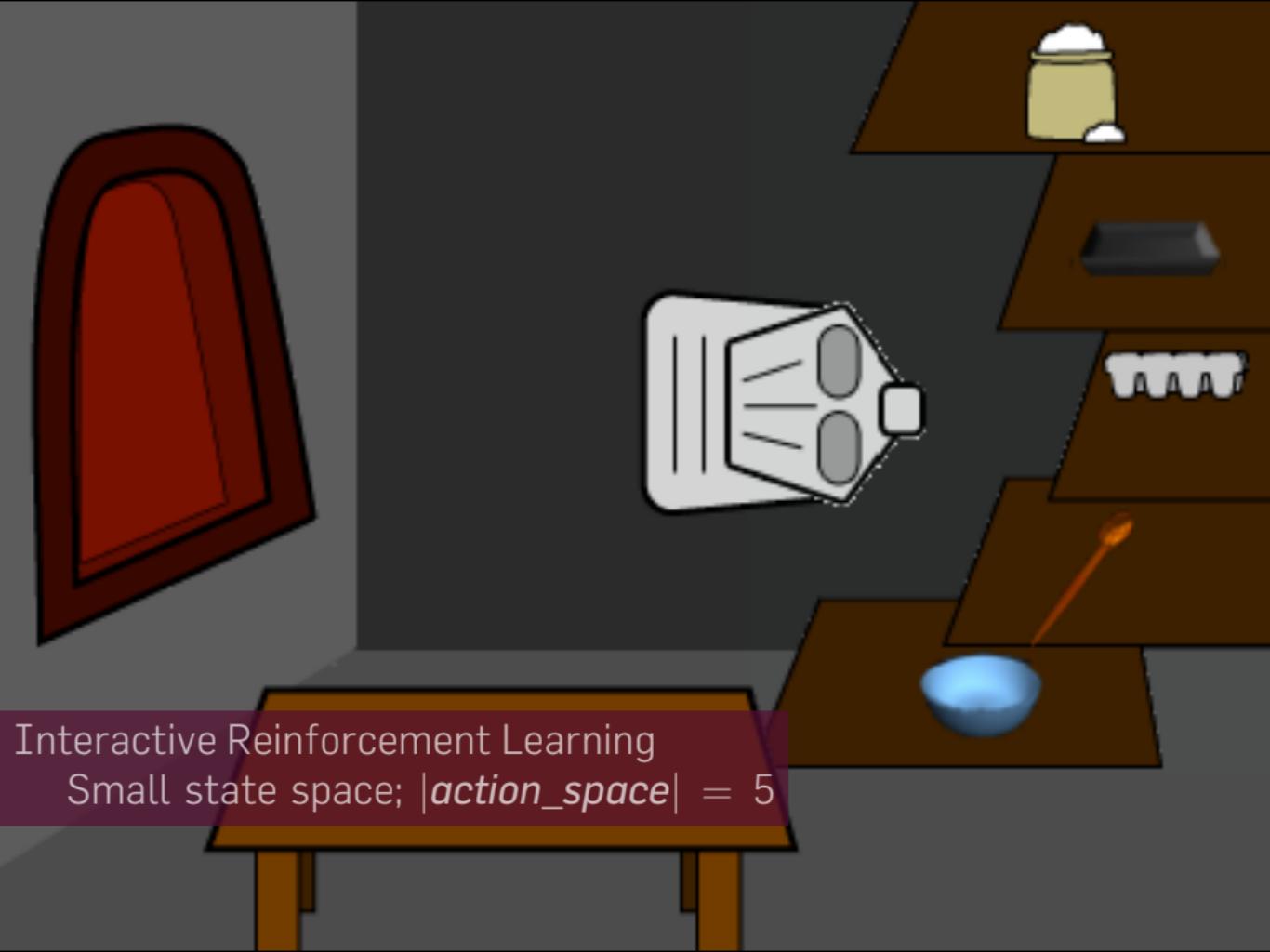
- These mechanisms are unfortunately often ill-defined, and particularly difficult to turn into equations (or controllers, in our case)
- not close-form equation of social interactions \Rightarrow data-driven approaches?

2 INSTANCES

- SPARC: transferring social skills from a human expert to an autonomous robot
- PInSoRo: learning to recognise complex social situations from child-child interactions

LEARNING SOCIAL AUTONOMY FROM HUMANS





Interactive Reinforcement Learning
Small state space; $|action_space| = 5$

...WELL, WELL...



Can we tackle much more complicated cases?

- real robot?

...WELL, WELL...



Can we tackle much more complicated cases?

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

...WELL, WELL...



Can we tackle much more complicated cases?

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

...WELL, WELL...



Can we tackle much more complicated cases?

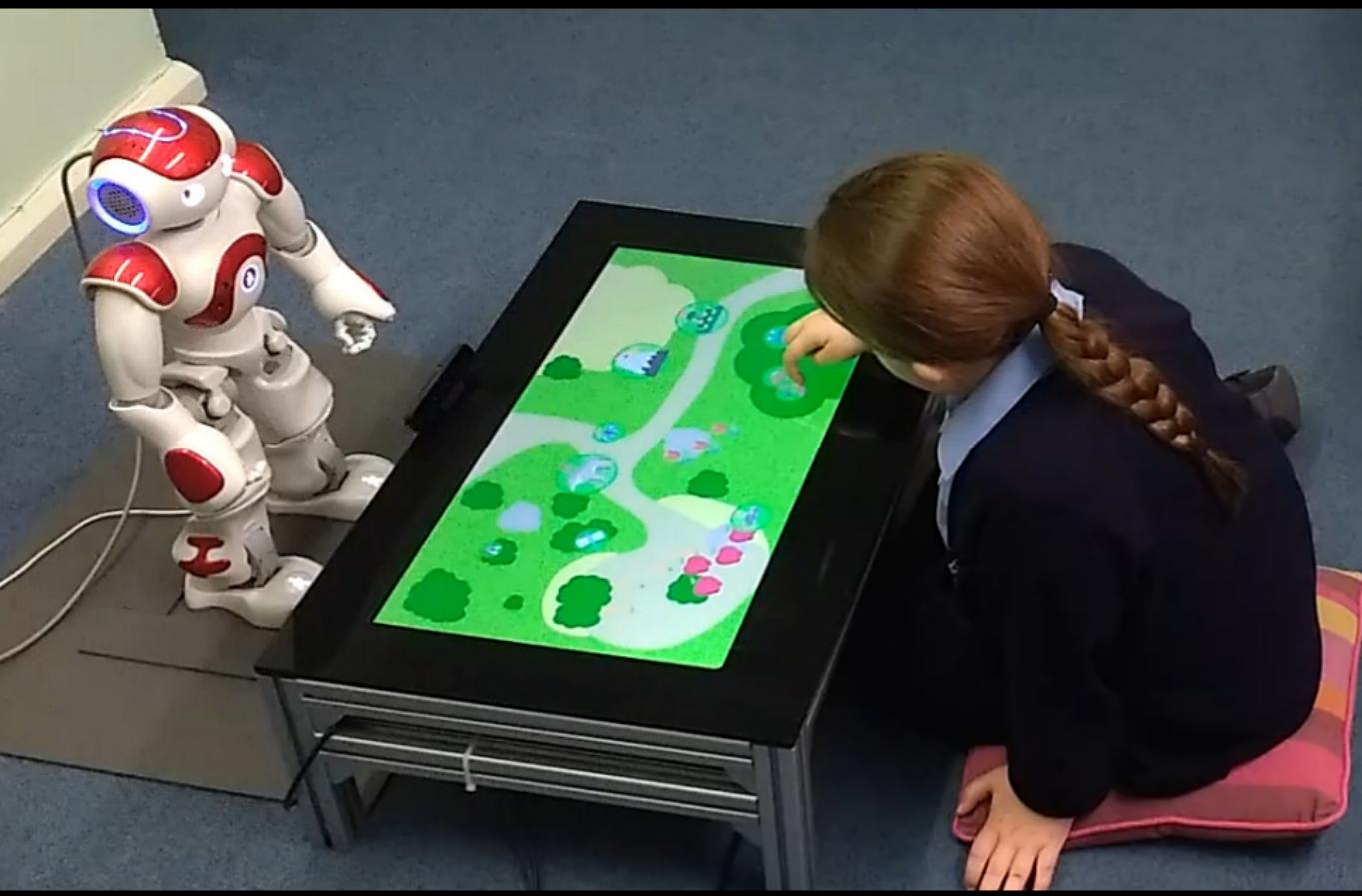
- real robot?
- real interaction (...with a human!)?
- continuous interaction?
- more realistic task (state vector & action space)?

...WELL, WELL...



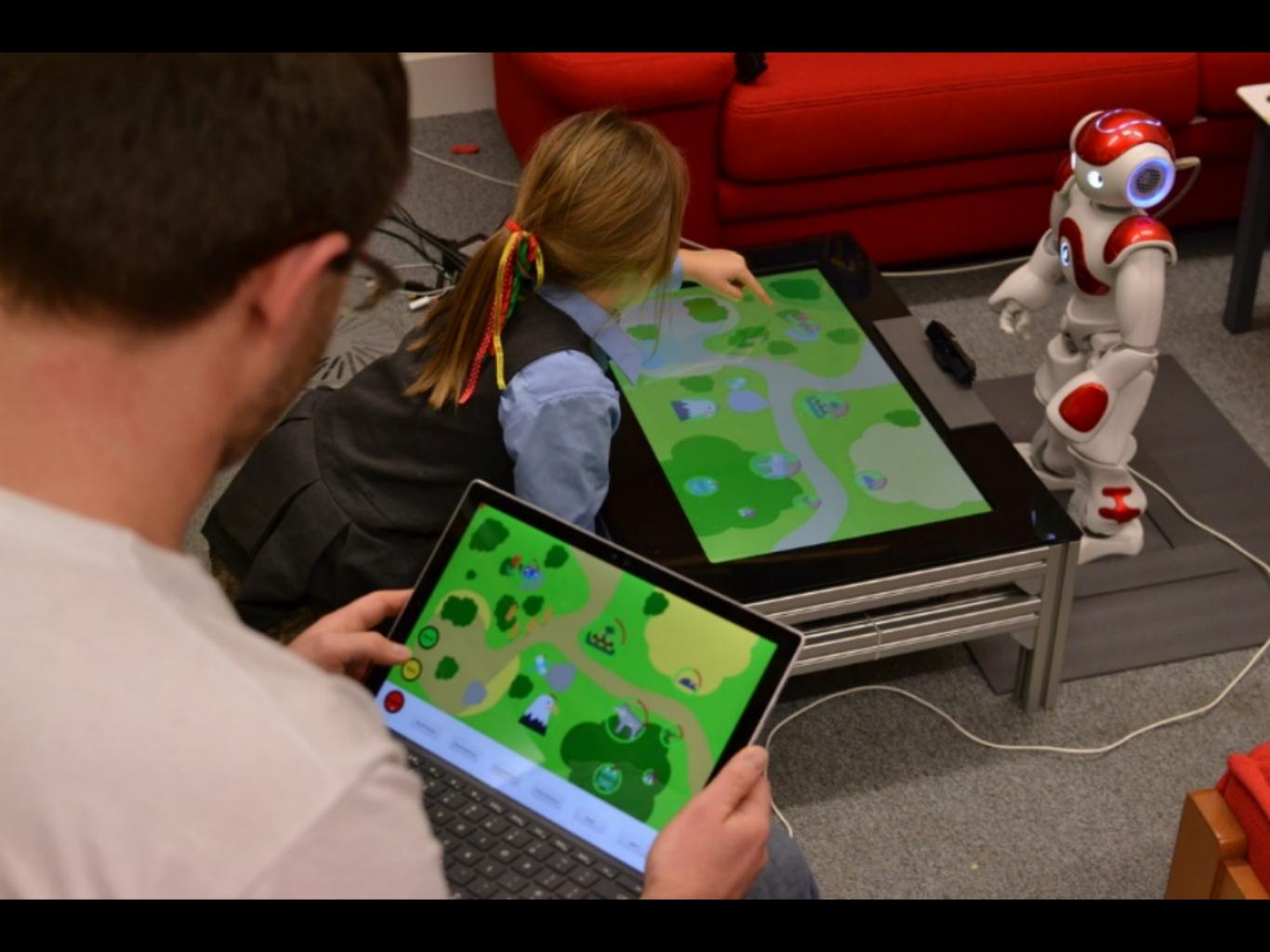
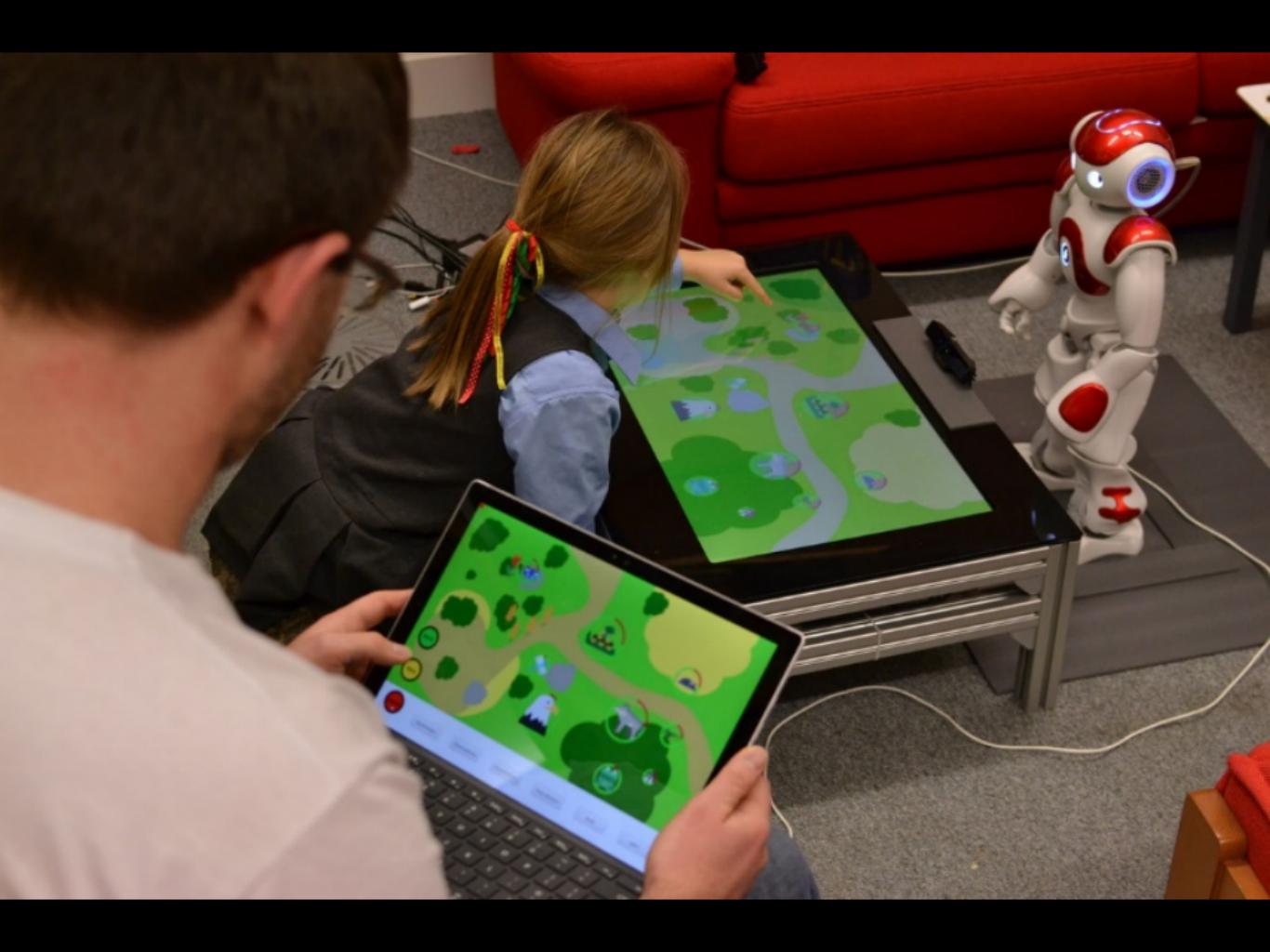
Can we tackle much more complicated cases?

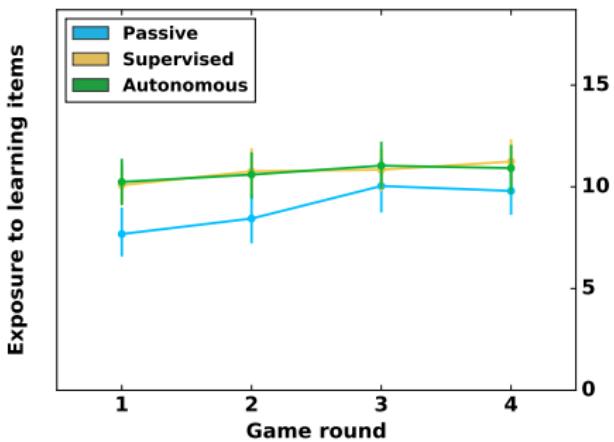
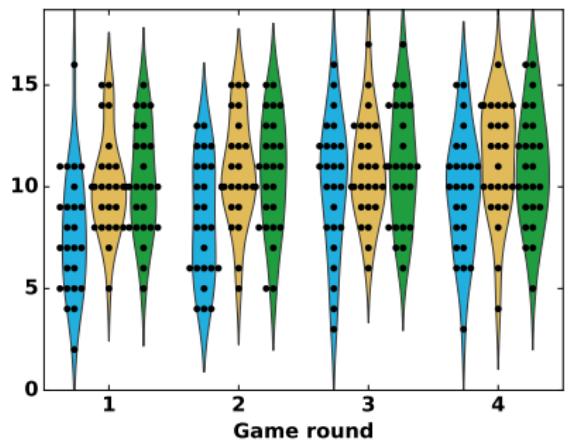
- real robot?
- real interaction (...with a human!)?
- continuous interaction?
- more realistic task (state vector & action space)?
- also including social behaviours?





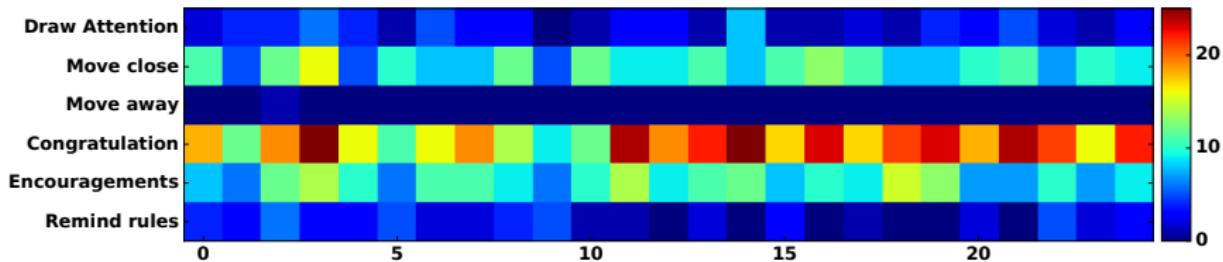
$state \in [0.; 1.]^{210}$ | $action_space| = 655$



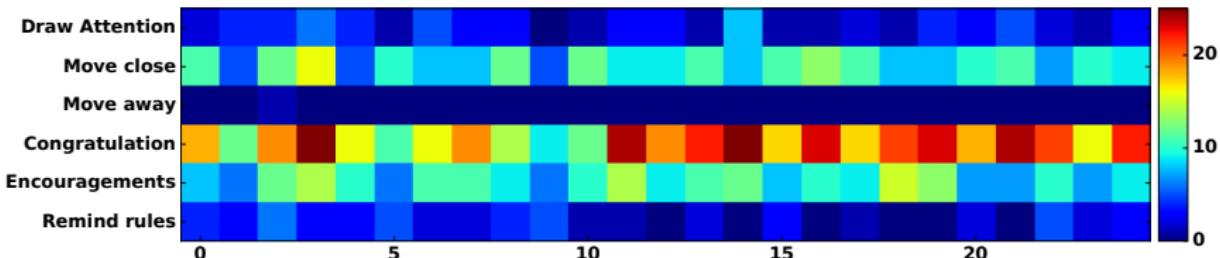


Learning-related game actions

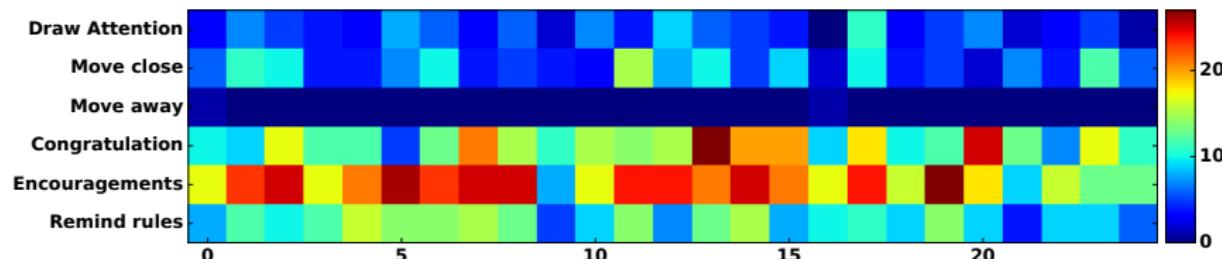
Supervised



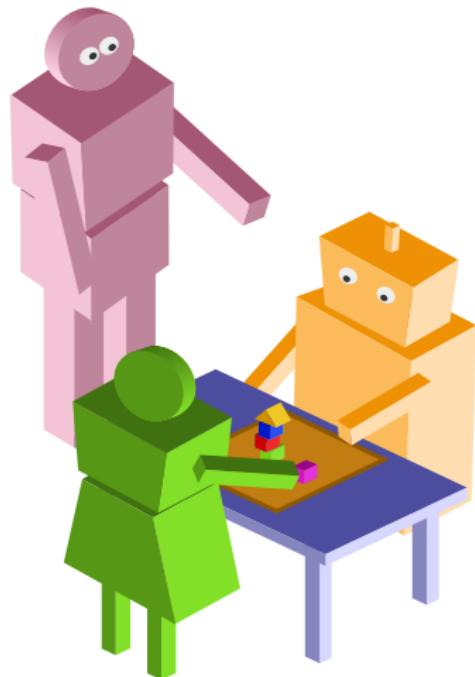
Supervised



Autonomous



TAKE-HOME MESSAGE FOR JOINT ACTION



- this example is about tutoring, however **progressively transferring autonomy** is a general principle
- it works well for relatively high-dimensional problems
- it also works for **social behaviours**

WHAT ABOUT MORE SUBTLE SOCIAL
DYNAMICS?

What challenges are we talking about, exactly?

ooo

Social interactions?

oooooooooooooooooooo

The challenge of "doing together"

oooooooooooo

Learning social behav

oooooooooooo

TO STUDY SOCIAL DYNAMICS, WE NEED...

What challenges are we talking about, exactly?

ooo

Social interactions?

oooooooooooooooooooo

The challenge of "doing together"

oooooooooooo

Learning social behav

oooooooooooo

TO STUDY SOCIAL DYNAMICS, WE NEED...

A **task!**

TO STUDY SOCIAL DYNAMICS, WE NEED...

A task!

...that exhibits:

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

TO STUDY SOCIAL DYNAMICS, WE NEED...

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

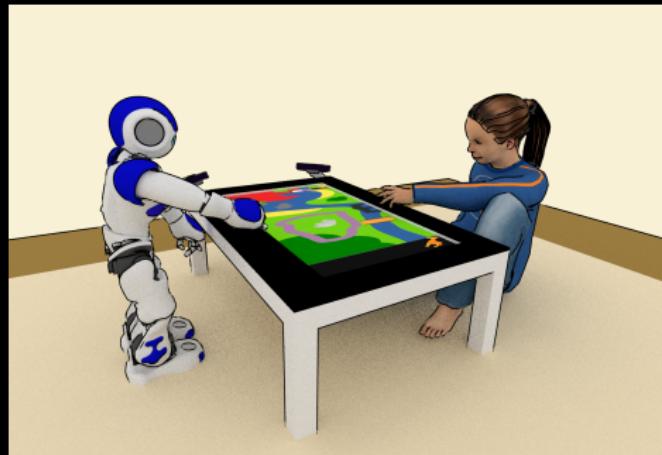
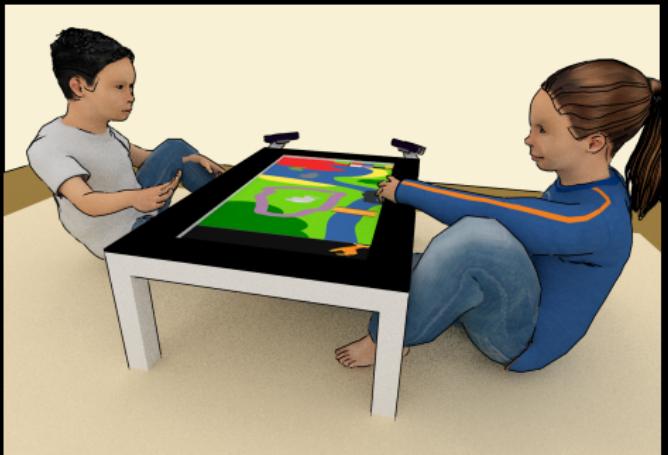
“Just play! Enjoy yourselves!”

- **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



THE PINSORO DATASET

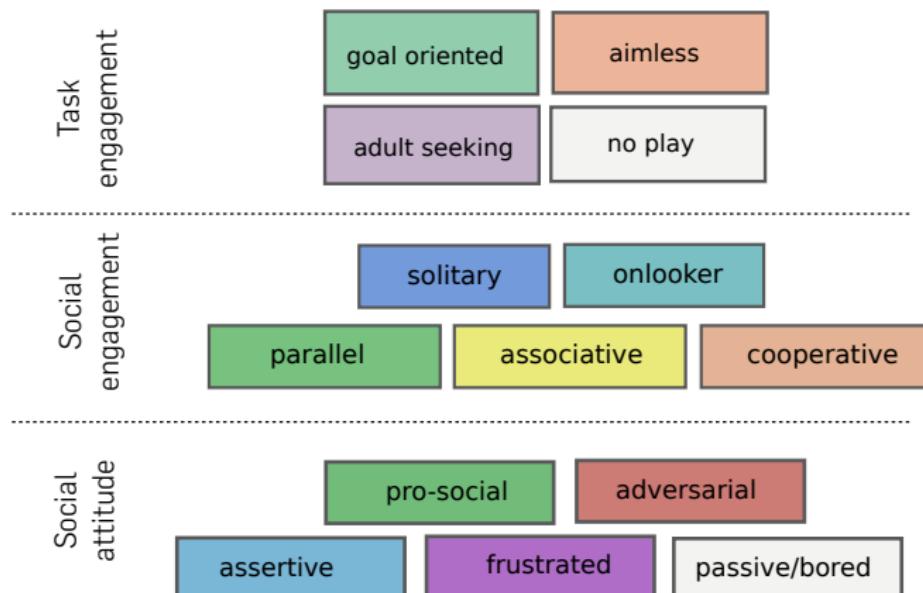
- 120 children, 4 to 8 years old
- 75 interactions
 - 90 children playing with another child,
 - 30 playing with a robot
- About 45h+ of recordings; 2M+ frames; \approx 2TB
- average duration of freeplay interactions: 24min in child-child condition; 19min in child-robot condition

Large open dataset: **freeplay-sandbox.github.io**

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	
+ post-process	optical flow, audio features facial action units...	

13000+ ANNOTATIONS



Attitude: passive

Social engag.: onlooker

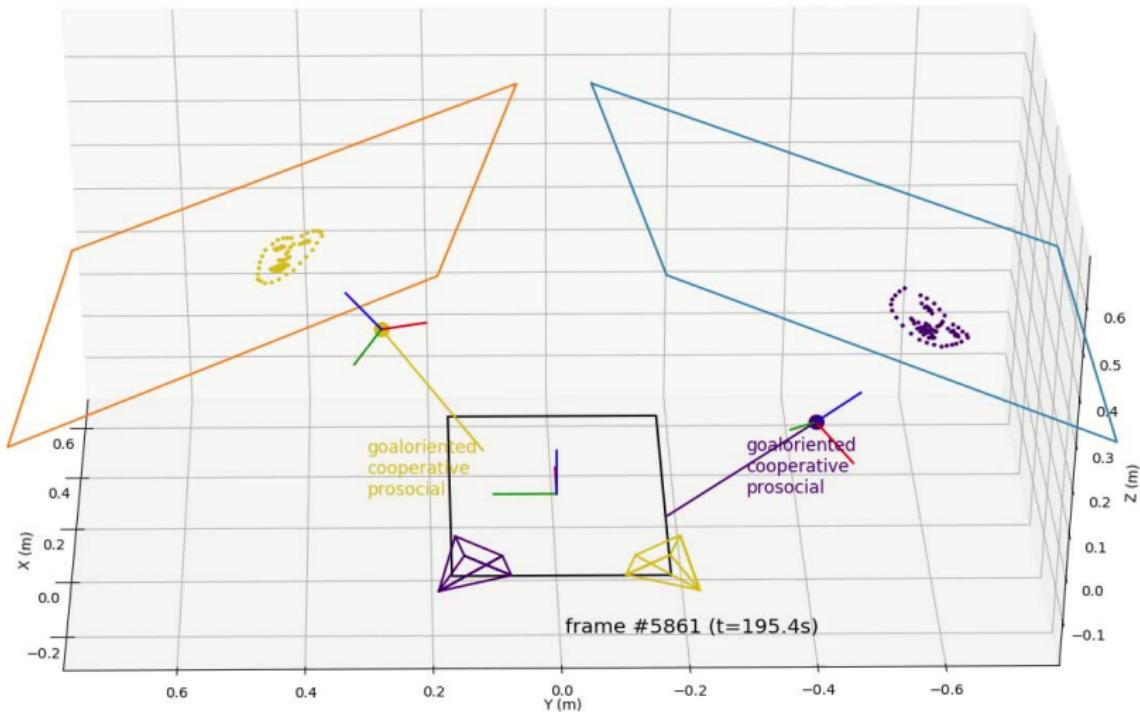
Task engag.: no play

Attitude: passive

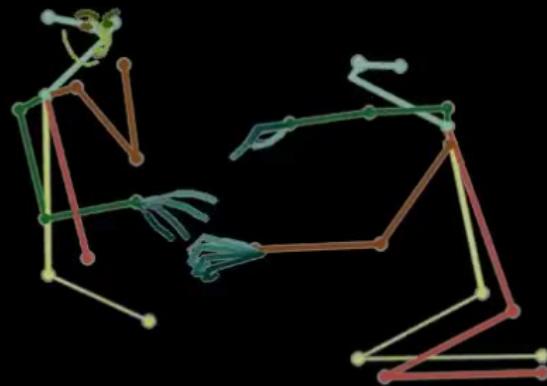
Social engag.: solitary

Task engag.: goal oriented

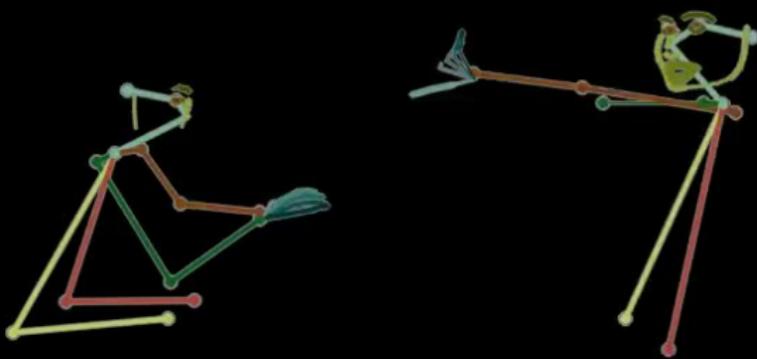




TOWARDS DATA-DRIVEN SOCIAL DYNAMICS?







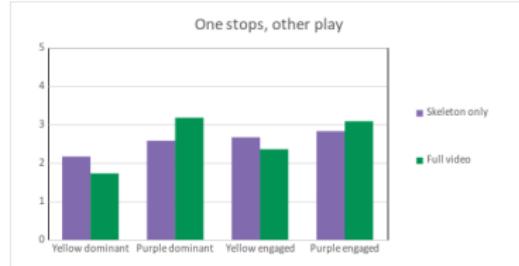
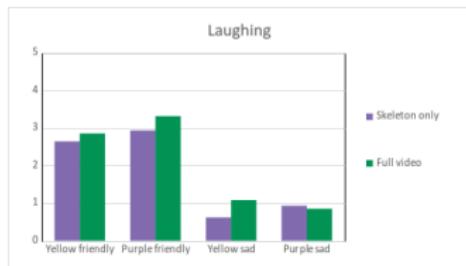


WHAT DO YOU SEE?



20 30-secs clips with a range of social situations; 200 participants on m-turk.

t-test between skeleton only and full video-streams show no difference in perception for the vast majority of the 11 tested constructs (cooperative, competitive, friendly, sad, engaged,...).



WHAT DO YOU SEE?



20 30-secs clips with a range of social situations; 200 participants on m-turk.

t-test between skeleton only and full video-streams show no difference in perception for the vast majority of the 11 tested constructs (cooperative, competitive, friendly, sad, engaged,...).

⇒ **30-secs long sequences of body postures and facial landmarks of dyads should be sufficient to recognise a social situation**

What challenges are we talking about, exactly?

ooo

Social interactions?

oooooooooooooooooooo

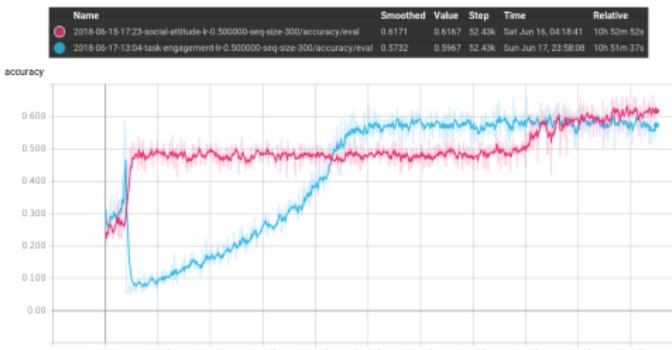
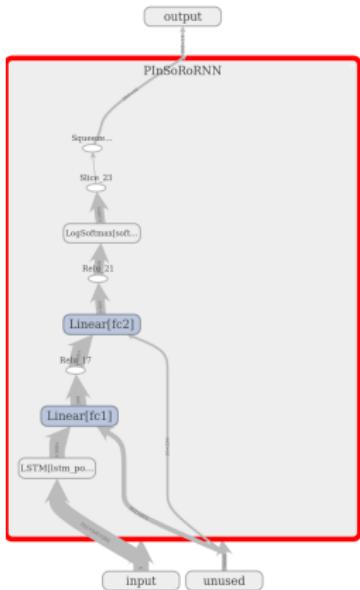
The challenge of "doing together"

oooooooooooo

Learning social beha

oooooooooooo

DATA CRUNCHING GOING ON!



pytorch; trained on 10 epochs x 2M datapoints; **WIP!!**

ULTIMATELY...

Real-time identification by the robot of...

- o 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!**

NOW FOR THE DISCUSSION

- o to reduce the socio-cognitive cost of collaboration, rely as much as possible on **implicit (sub-conscious) social mechanisms**



NOW FOR THE DISCUSSION



- o to reduce the socio-cognitive cost of collaboration, rely as much as possible on **implicit (sub-conscious) social mechanisms**
- o (do not be scared of ambiguous/partially defined instructions!)

NOW FOR THE DISCUSSION



- to reduce the socio-cognitive cost of collaboration, rely as much as possible on **implicit (sub-conscious) social mechanisms**
- (do not be scared of ambiguous/partially defined instructions!)
- however, **communication dynamics** & the **recognition of grounding errors** should be research priorities

NOW FOR THE DISCUSSION

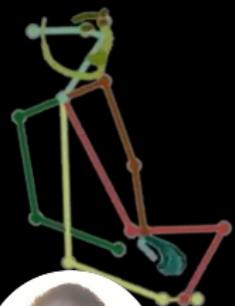


- to reduce the socio-cognitive cost of collaboration, rely as much as possible on **implicit (sub-conscious) social mechanisms**
- (do not be scared of ambiguous/partially defined instructions!)
- however, **communication dynamics** & the **recognition of grounding errors** should be research priorities
- **how does that relate to the higher-level concept of 'theory of mind'?**

Attitude: pro-social

Social engag.: parallel play

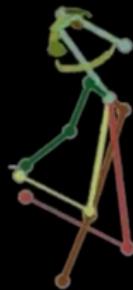
Task engag.: goal oriented



Attitude: pro-social

Social engag.: parallel play

Task engag.: goal oriented



Thank you!

SOME MORE STUFF

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

What challenges are we talking about, exactly?

Social interactions?

ooo

oooooooooooooooooooo

The challenge of "doing together"

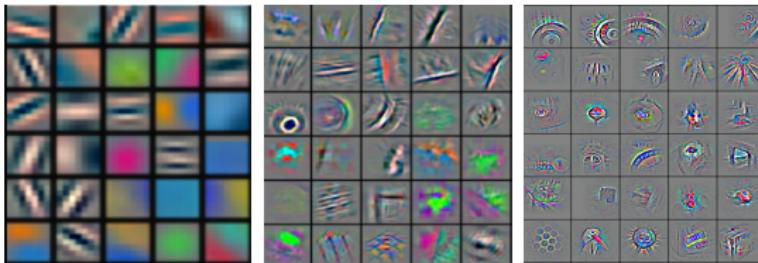
Learning social behav

oooooooooooo

DEEP NETWORKS \equiv BLACK BOXES?



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



What challenges are we talking about, exactly?
ooo

Social interactions?
oooooooooooooooooooo

The challenge of "doing together"
oooooooooooo

Learning social behavior
oooooooooooo

DEEP NETWORKS ≡ BLACK BOXES?



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