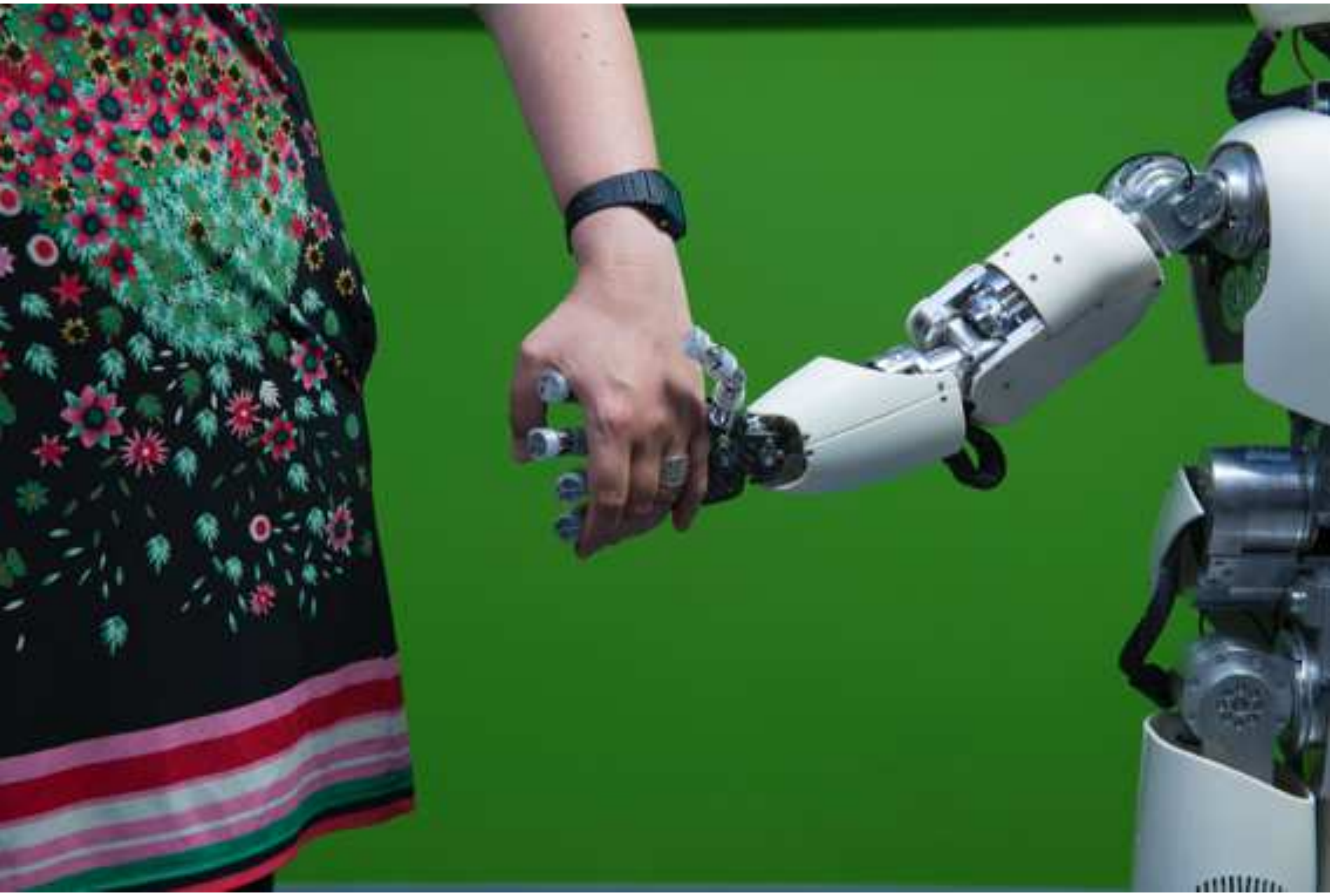
Grasping, vision and interaction for object manipulation with iCub



Serena Ivaldi

Team LARSEN, INRIA
IAS Lab, TU Darmstadt

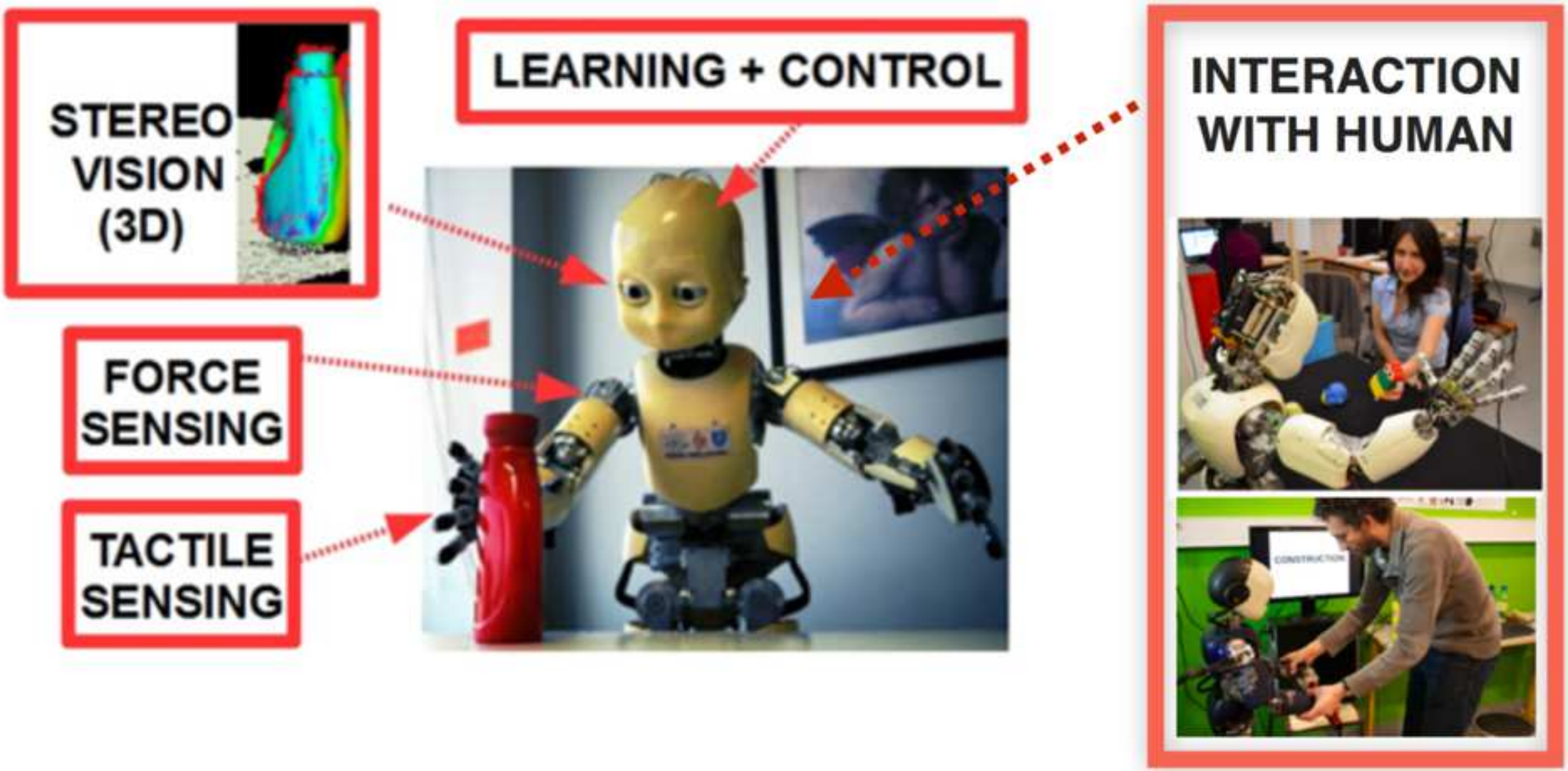
serena.ivaldi@inria.fr

larsen

informatiques mathématiques

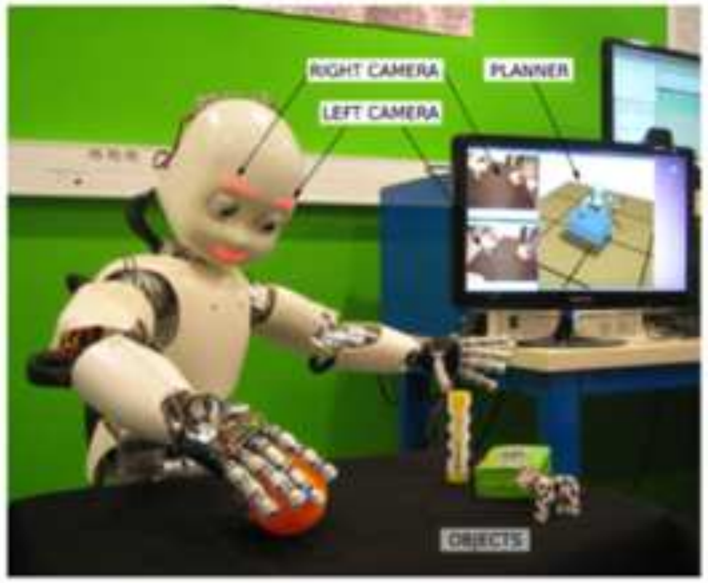
TECHNISCHE UNIVERSITÄT DARMSTADT

Outline of the talk



Outline of the talk

Multimodal learning of the visual appearance of objects (w/ Kinect)

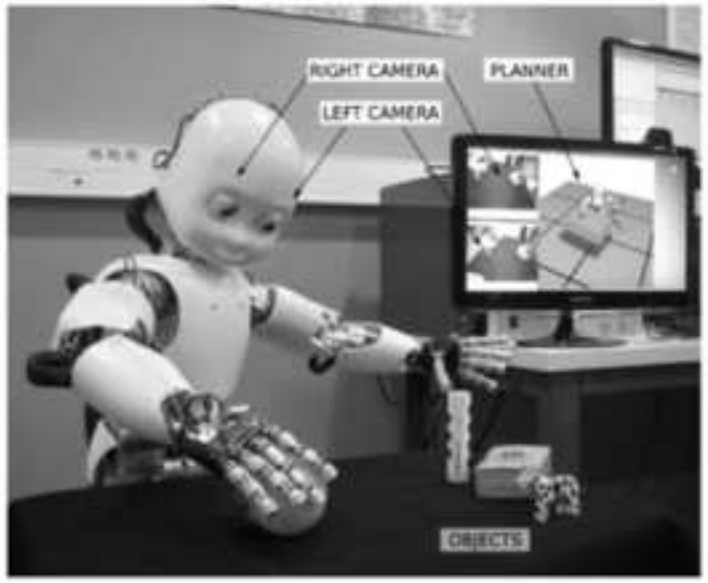


Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Physical interaction: even non-experts can teach iCub how to assemble objects

Outline of the talk

Multimodal learning of the visual appearance of objects (w/ Kinect)



Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Physical interaction: even non-experts can teach iCub how to assemble objects

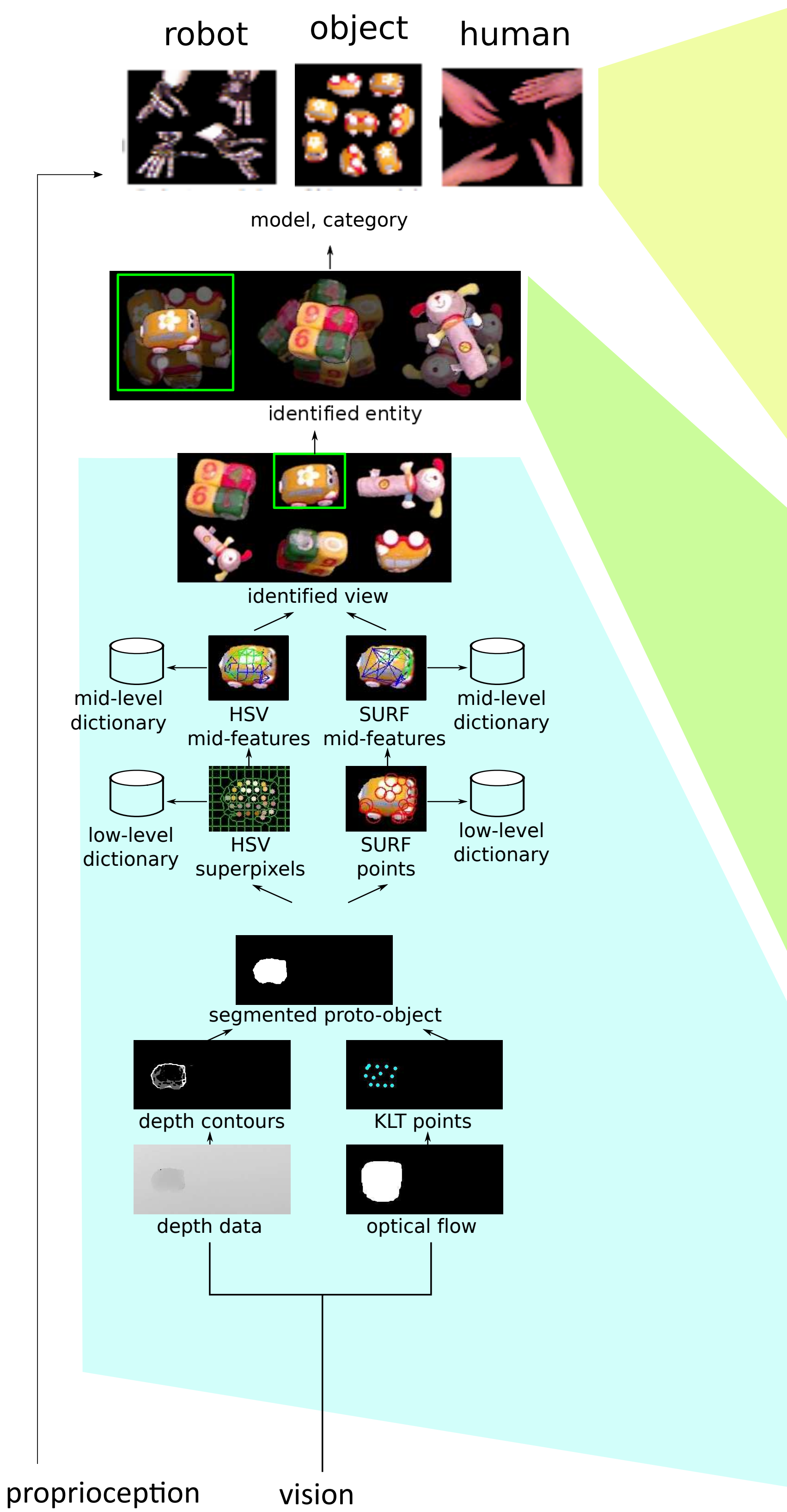
Learning to identify objects

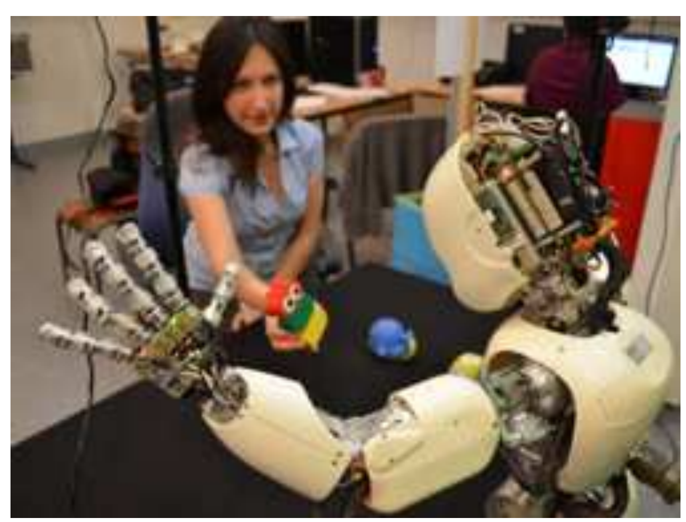
What should the robot do to learn the objects appearance?

- intuitively, focus on the most "complex objects"
- manipulate the object to update its model
- choose the manipulations that provokes a new object appearance
- get help from the human (teacher)

Ivaldi et al, Object learning through active exploration, IEEE Trans Aut Ment Develop, 2014

Multimodality for object learning





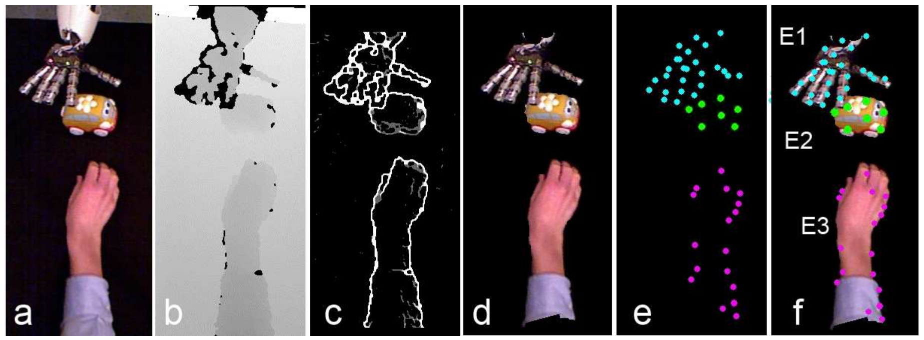
exploration and interaction
(better models with categories)

active exploration
(better models)

observation
(pure vision: models and entities)

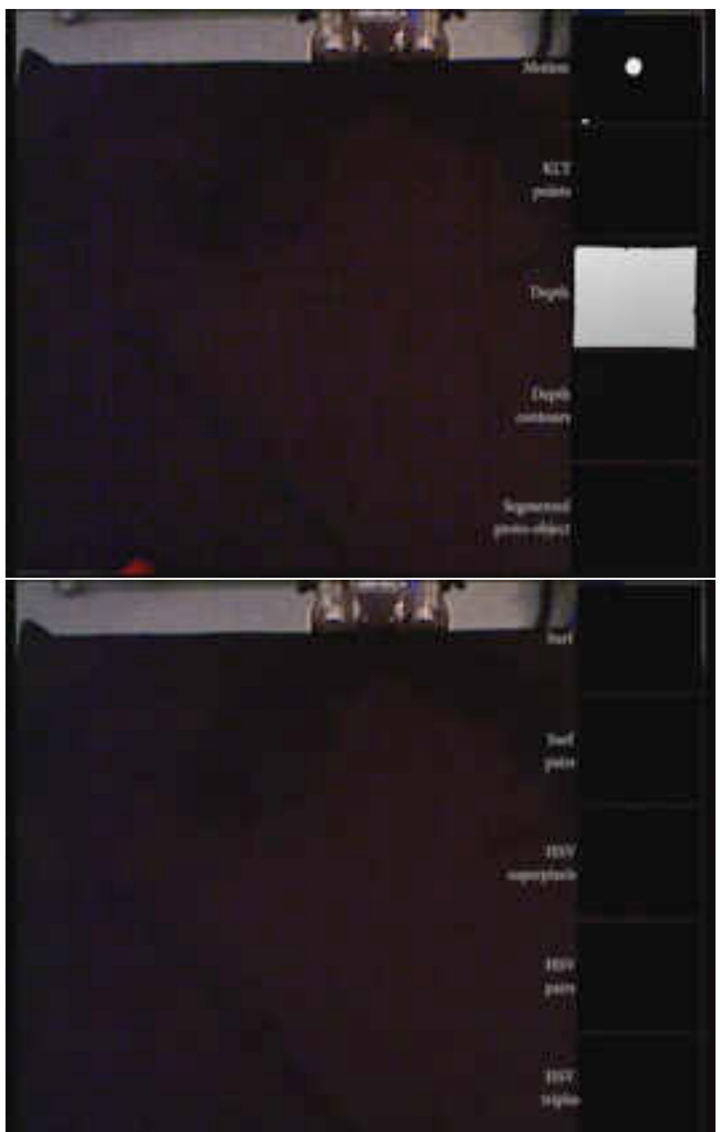
Observation alone is not enough

vision →



entities + collected views →

model →



The robot learns the objects demonstrated by the human.

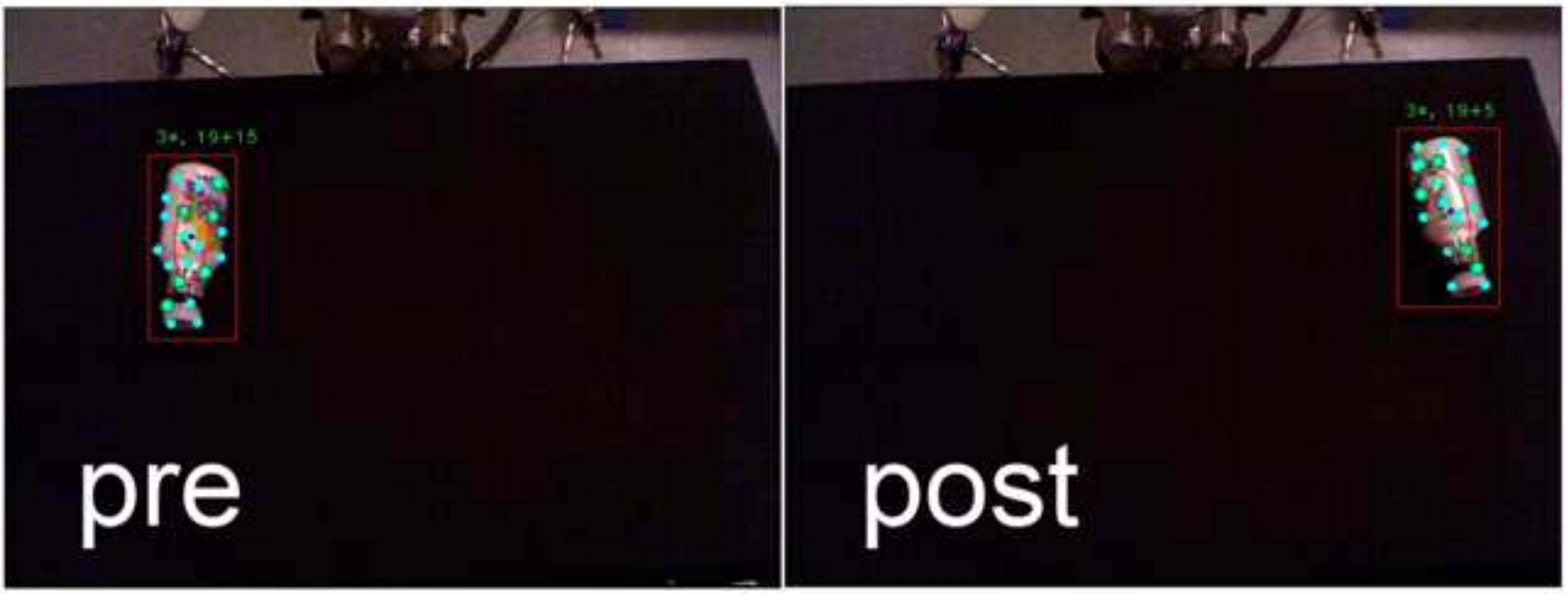
The robot has not yet learnt to identify its body, hence all entities are labeled by an "unknown" category.

grasp
lift
throw

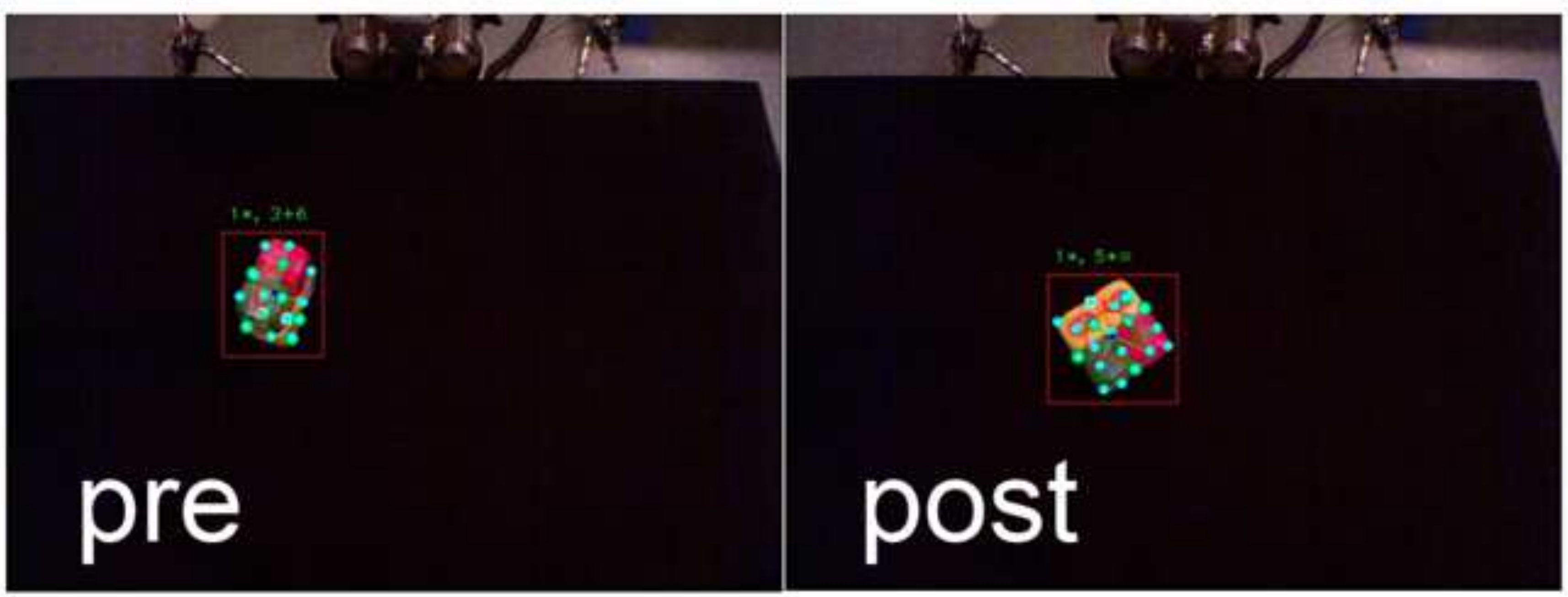
grasp
lift
rotate
put

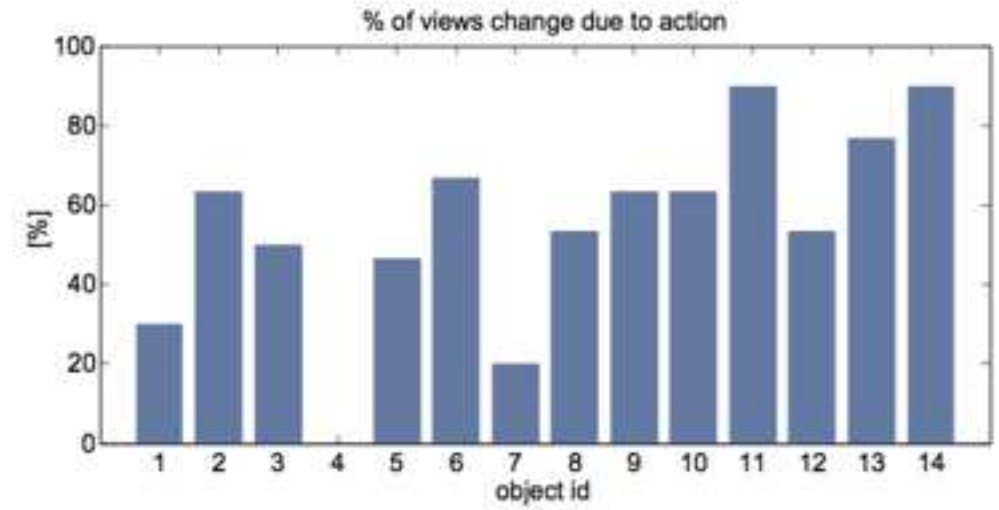
Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

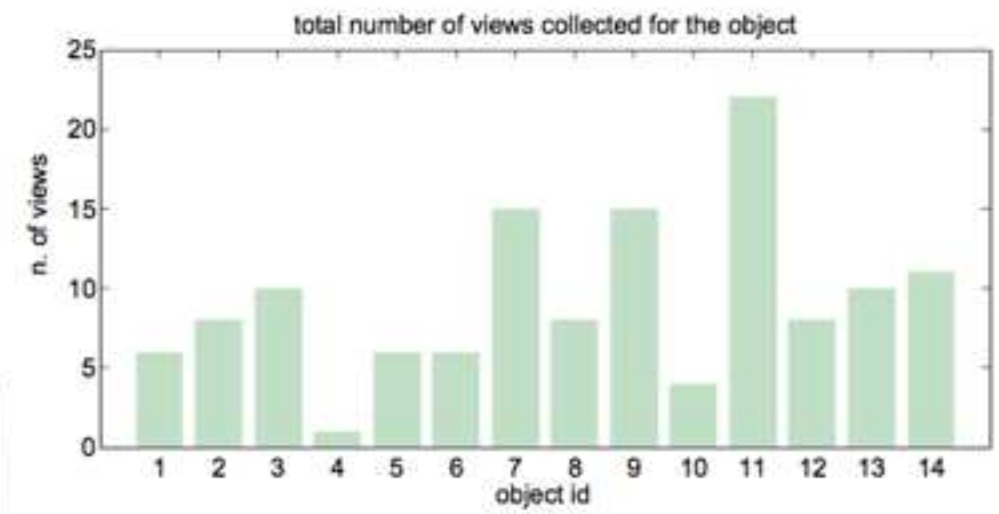
action does not change the view



action provokes a new view

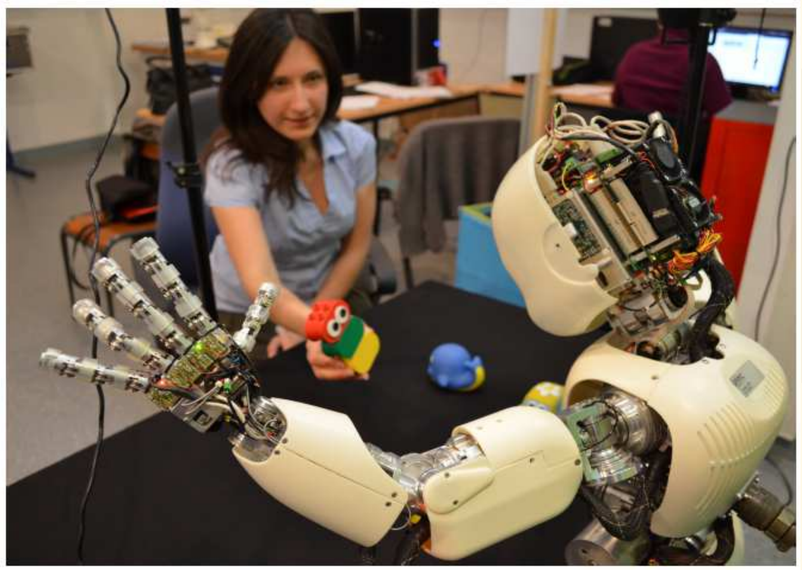






Active exploration & social guidance

social exploration

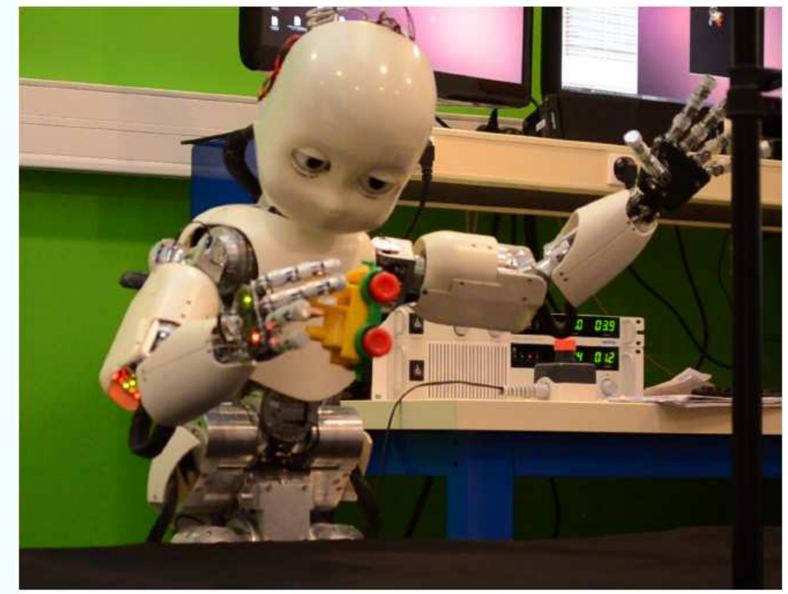


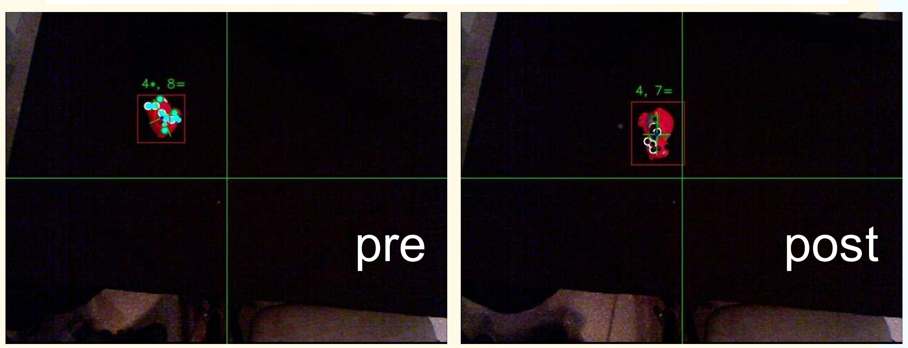
intrinsic motivation
SGIM-ACTS

exploration strategy

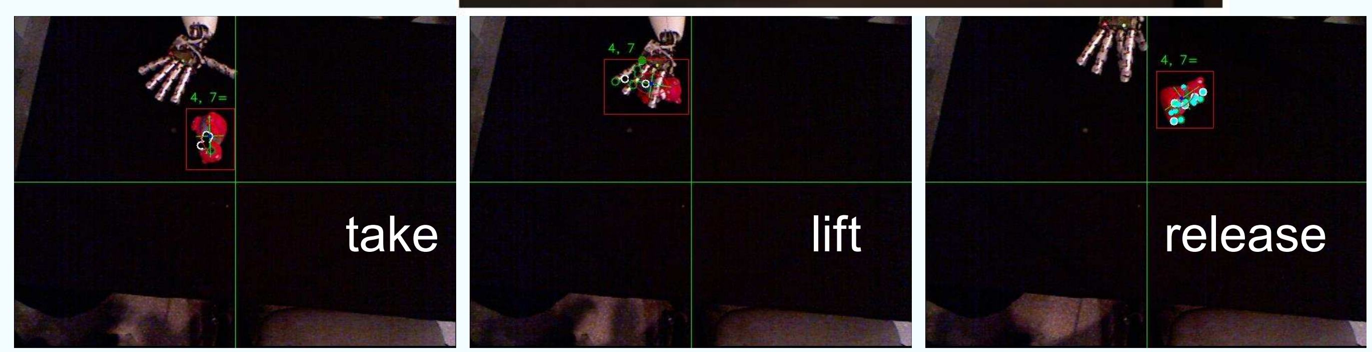
actor action object

autonomous exploration

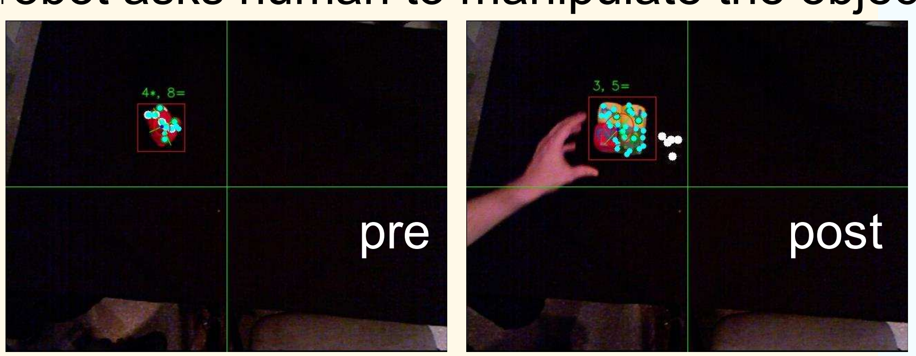




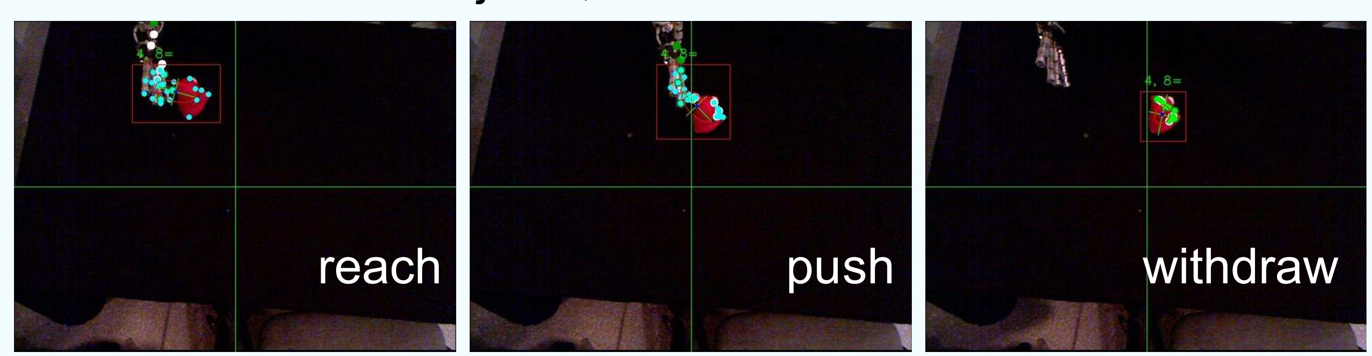
robot asks human to manipulate the object



robot lifts the objects, then makes it fall on the table



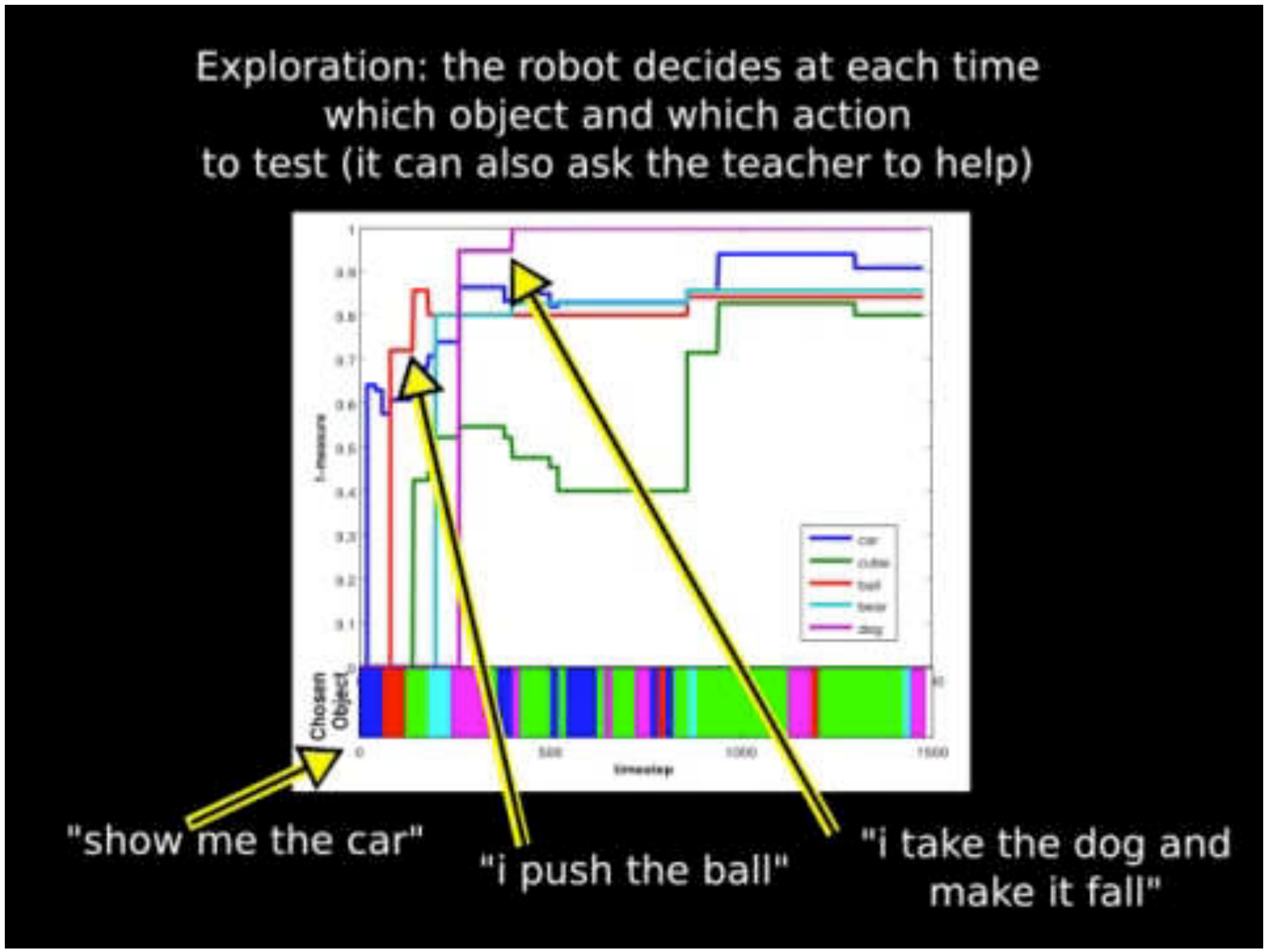
robot asks human to show a new object



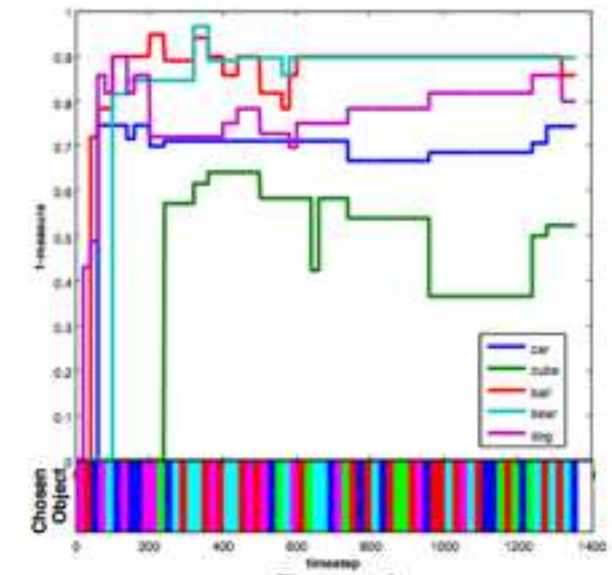
robot pushes the object

Curiosity-driven exploration of objects

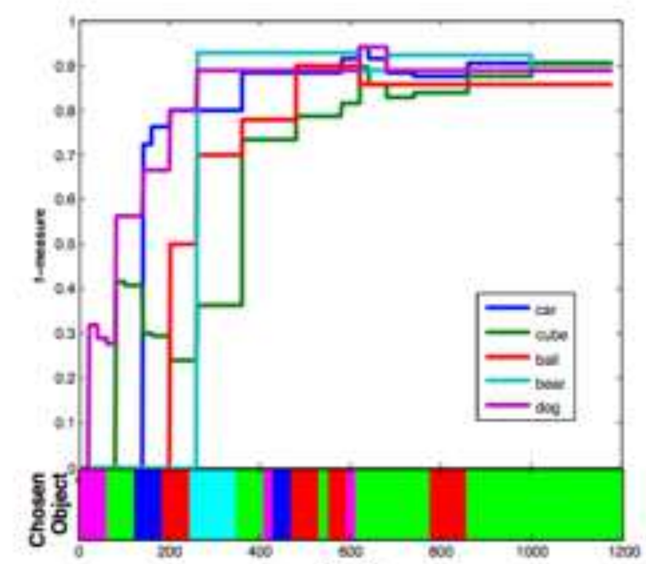
- Focusing on the objects that are not yet learned
- choosing the appropriate action for each object



ball → yellow car → red bear → …



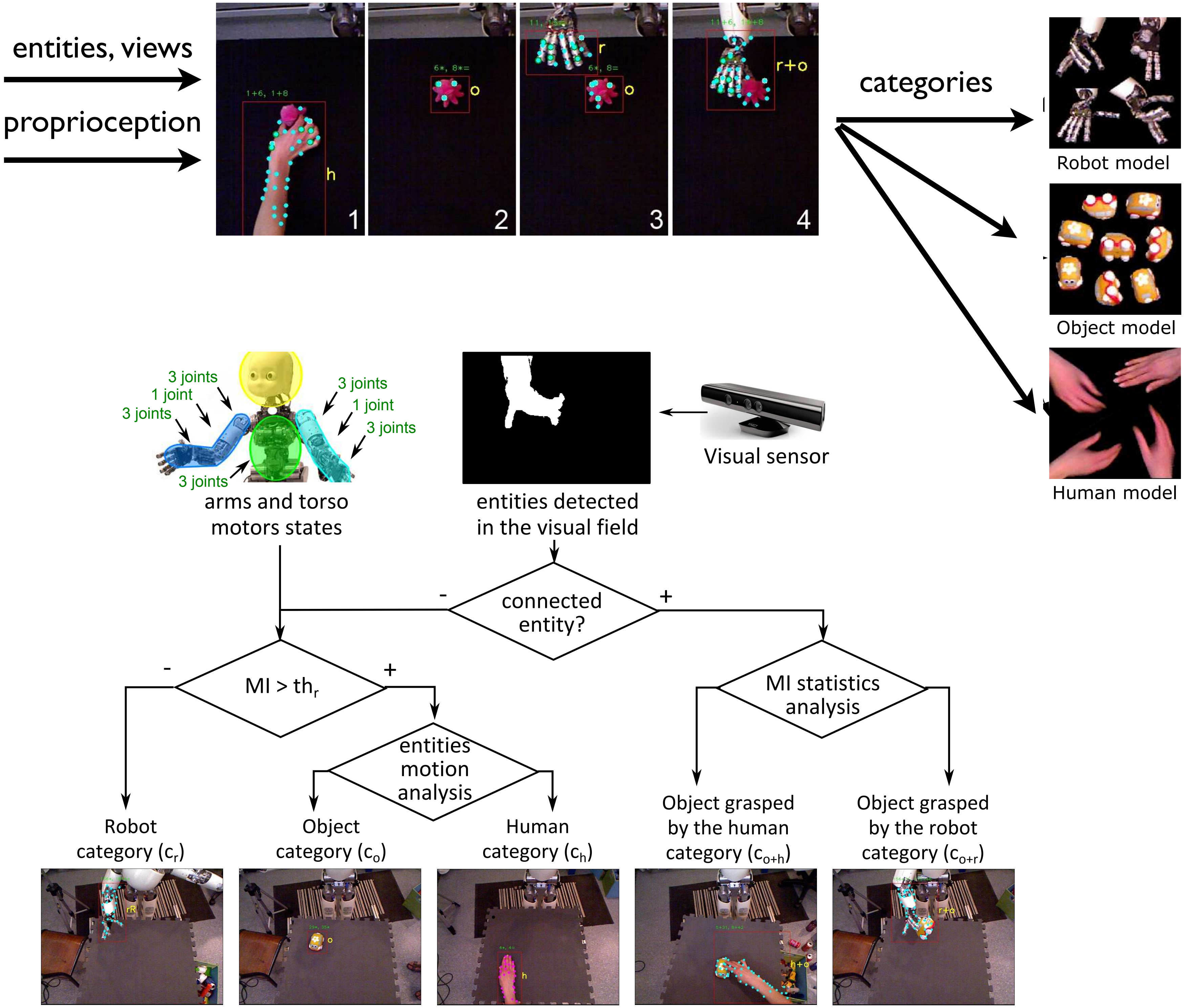
Random



Intrinsic motivation

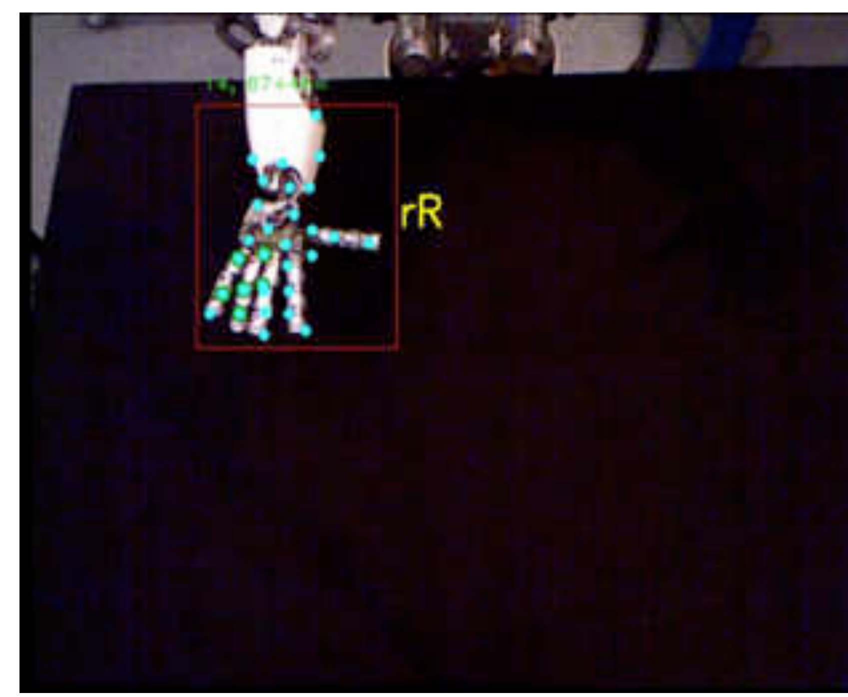
Ivaldi, Nguyen, Lyubova, Droniou, Padois, Filliat, Oudeyer, Sigaud (2014) Object learning through active exploration. IEEE Transactions on Autonomous Mental Development.

Better learning with action and interaction



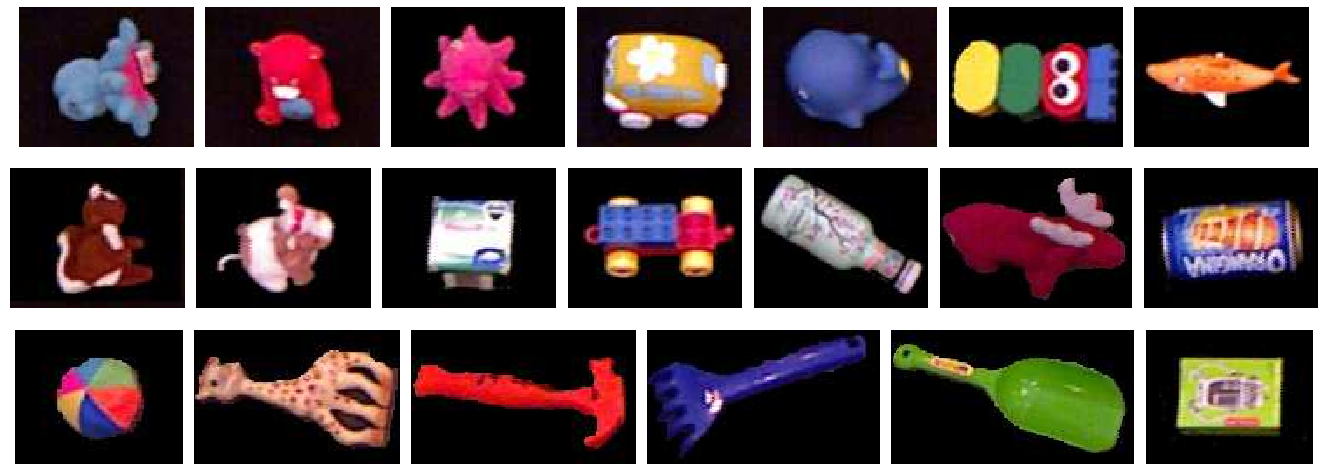
The robot learns the objects through manipulation.

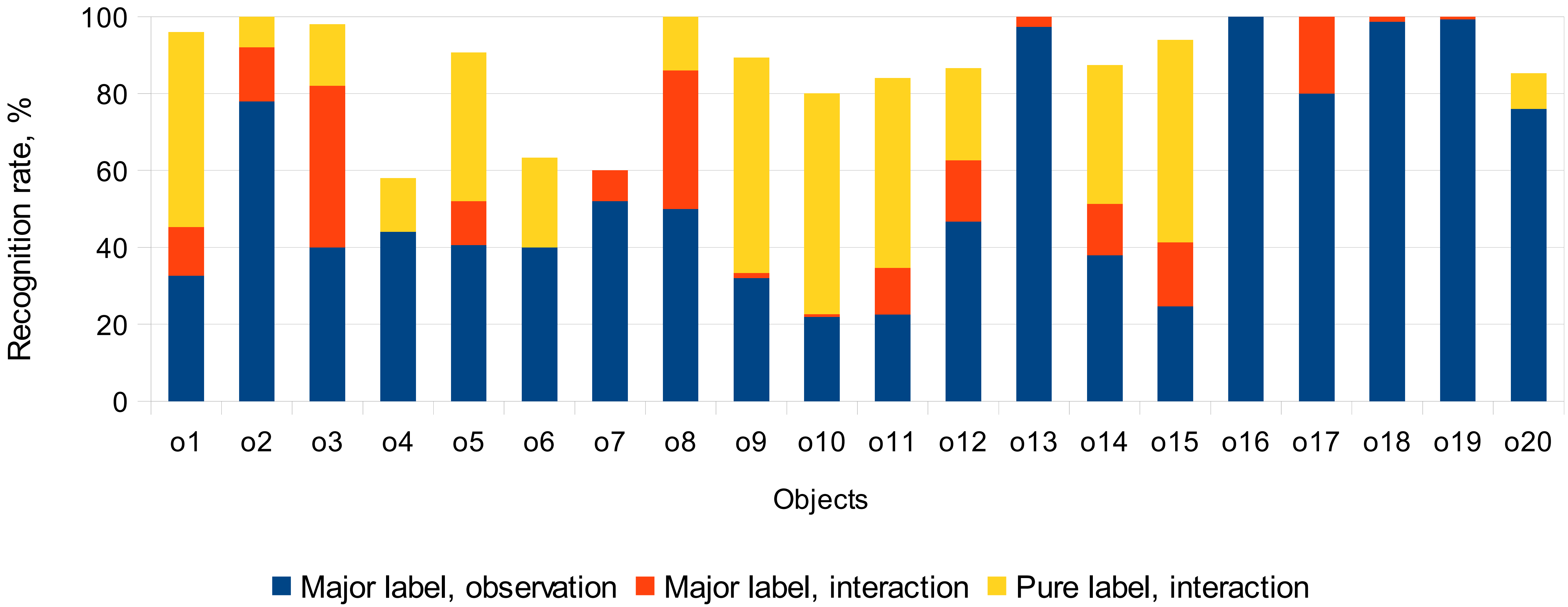
The robot learns to identify its body, hence entities can be categorized as "robot hand", "human hand" and "object".



Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

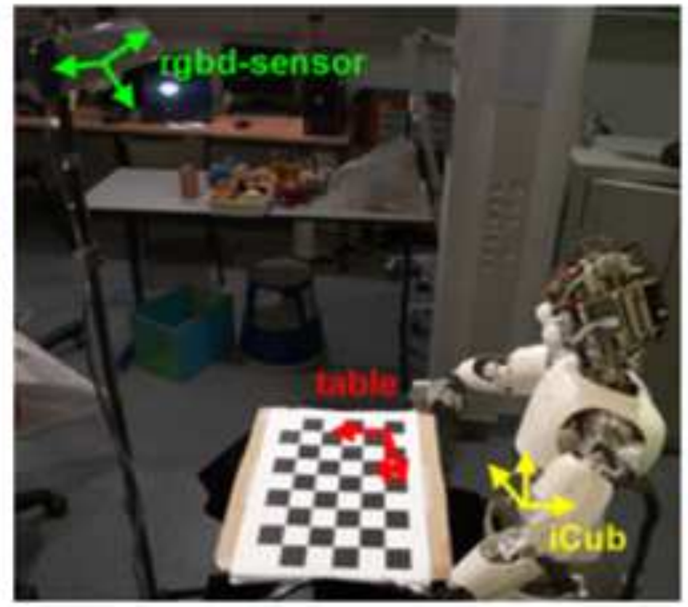
Better object recognition



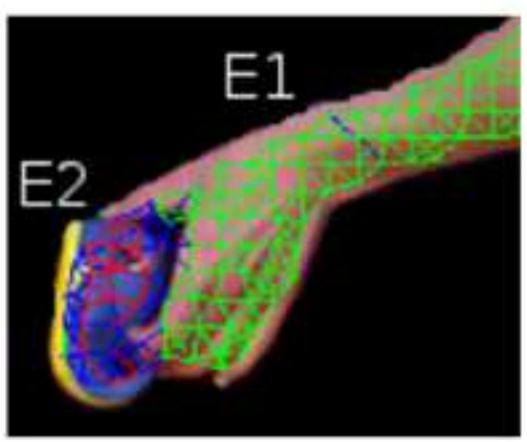


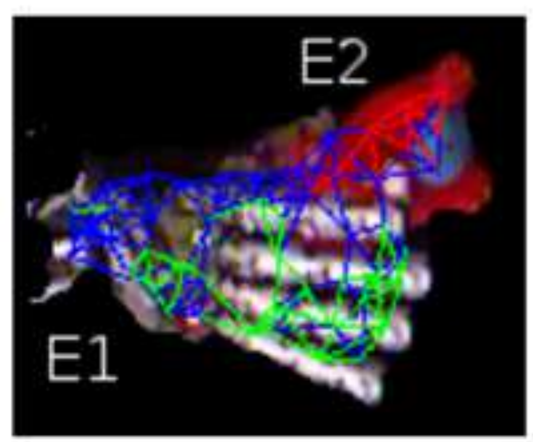
Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

Visual learning using the Kinect



- Calibration only once (if Kinect is fixed)
- High-resolution images with depth image
  - Useful to retrieve the top (max_z) of each object and adapt the grasp
  - Many feature points: better models

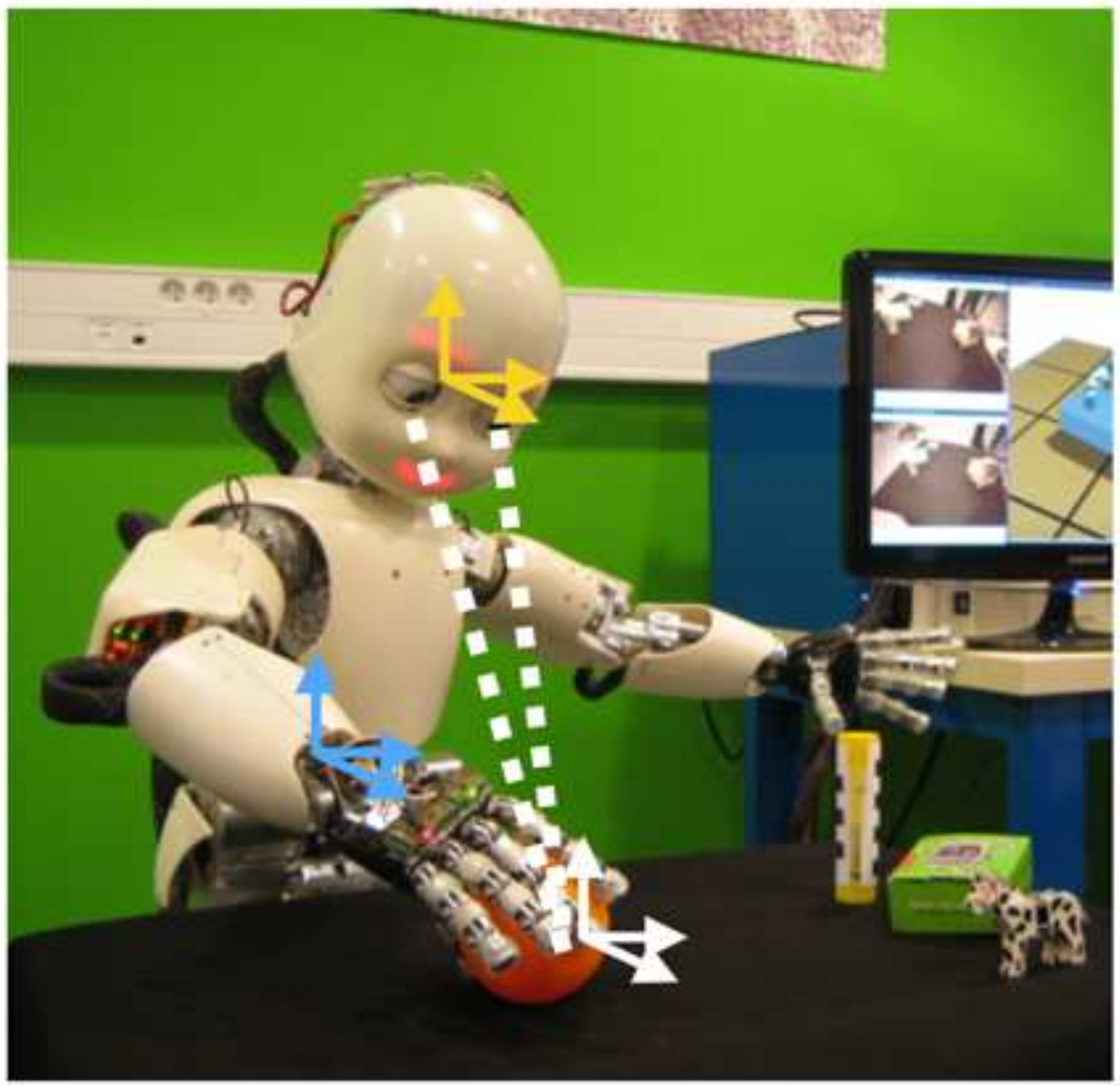




Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

Can we do the same with the eyes' cameras?

- Ideal grasping of perfectly localised objects using the eyes' cameras: eye-hand calibration + object pose (vision) + object size/shape (vision) + correct grasp = success!

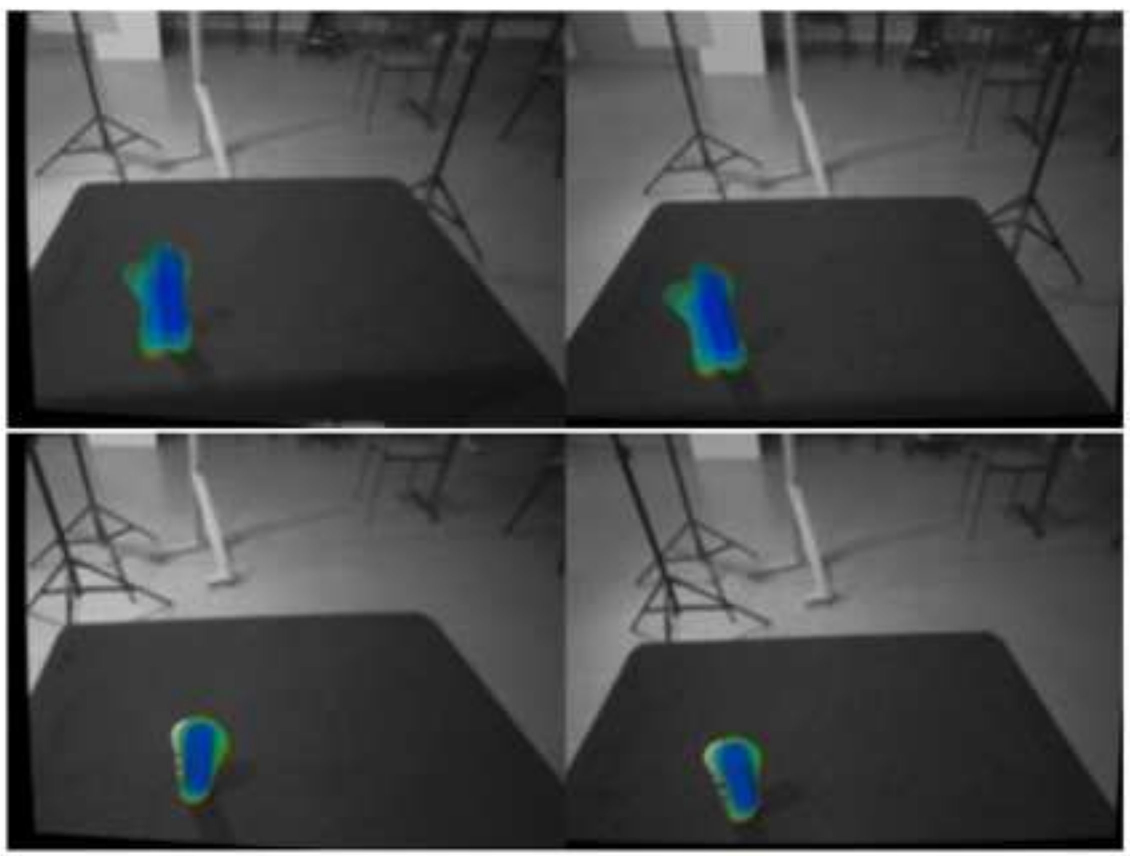


success!

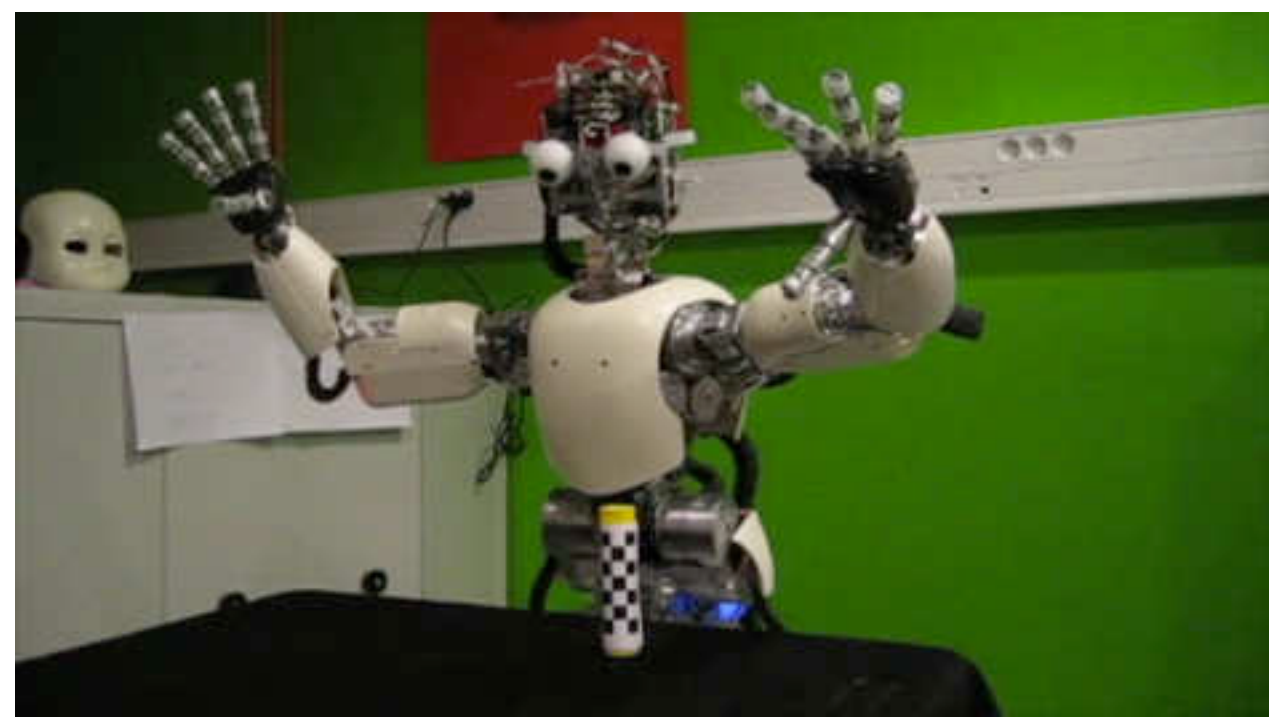
Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. Robotics and Autonomous Systems, 62(12): 1742-1754.

Unfortunately, the cameras bring limitations…

- Error in object pose estimation is inevitable, particularly when the object pose is estimated through low-resolution cameras
- Grasping is very sensitive to the accuracy of the object pose estimation
  ➡ failure!



inaccurate object pose estimation

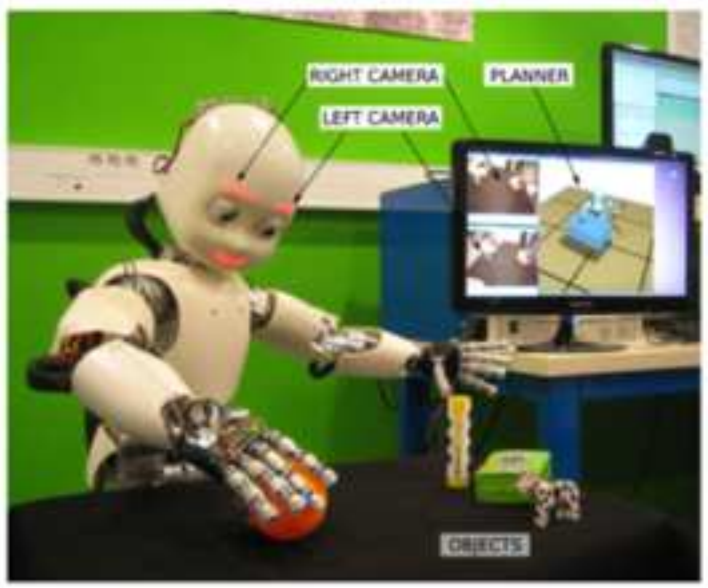


failure!

Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. Robotics and Autonomous Systems, 62(12): 1742-1754.

Outline of the talk

Multimodal learning of the visual appearance of objects (w/ Kinect)



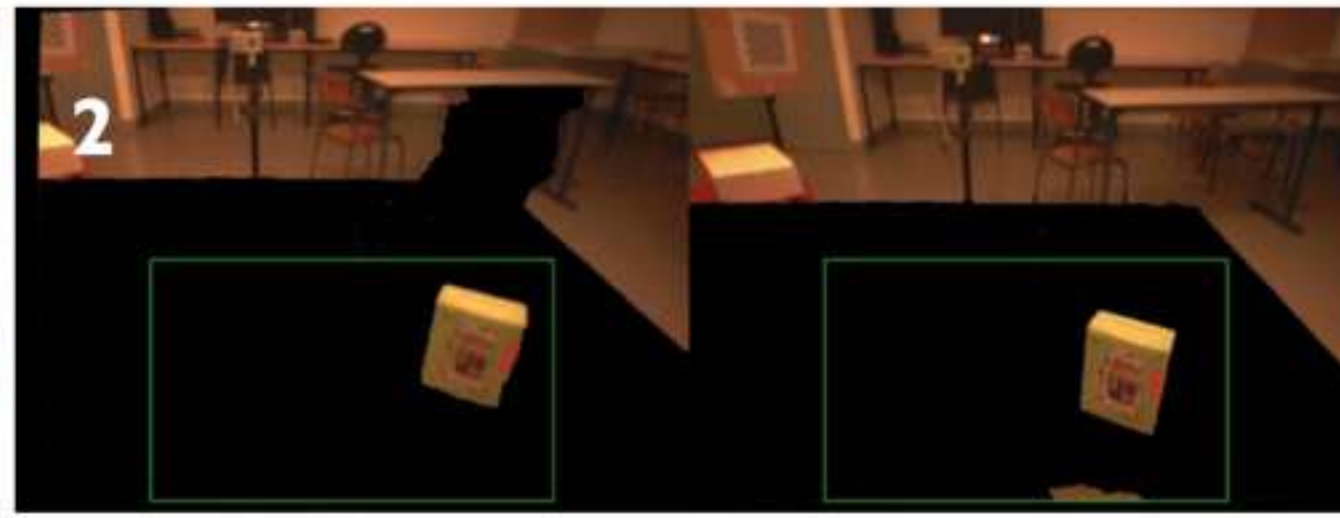
Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Physical interaction: even non-experts can teach iCub how to assemble objects

Unfortunately, the cameras bring limitations…

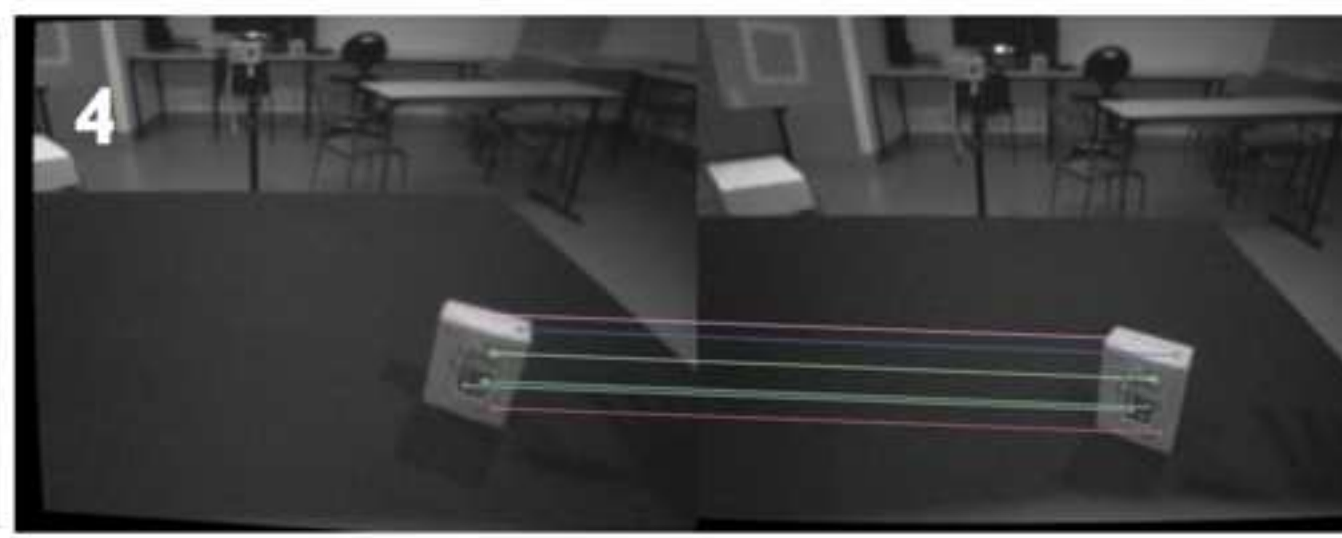
- Extracting the point cloud from the stereo cameras of iCub

camera images after distortion correction



extracted objects (w/ Grabcut algorithm)

2D features (SURF)



matched features

Grasping objects localised from noisy point clouds

- **Problem:** the point cloud from the stereo cameras has too few points to run classical algorithms, such as the Iterative Closest Point (ICP, in PCL)

20 points　　8 points　　9 points

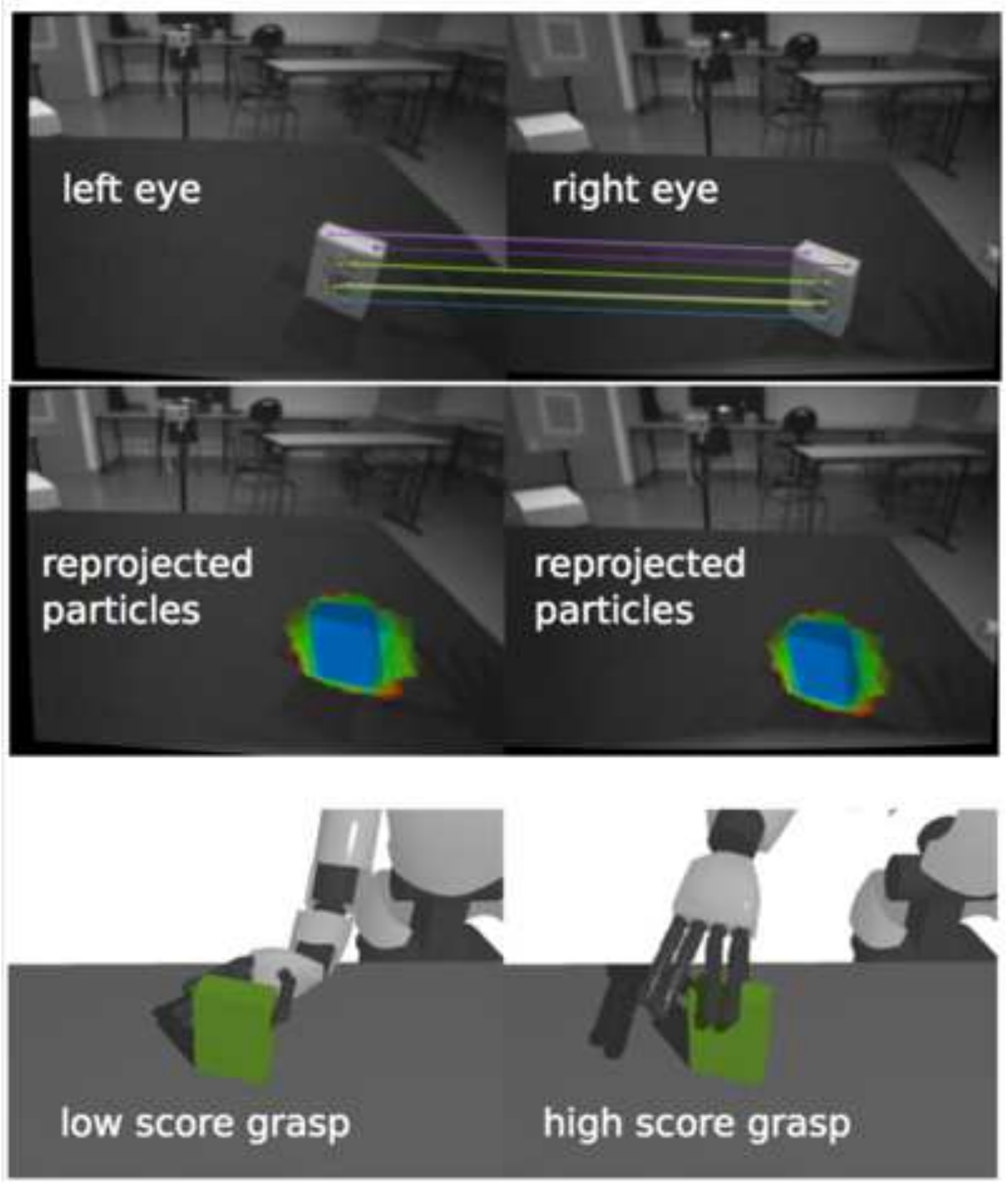
- Small errors in the estimated pose may cause the planned grasp to fail
- Difficult to validate a grasp when tactile or force sensing is missing

➡ **find grasps that are less sensitive to the pose uncertainty**

- **Intuition**: exploit the uncertainty in the object pose estimation
- **Not a pose, rather a distribution**

Grasping objects localised from noisy point clouds

➡ **Proposed method:**
grasp planning method that explicitly considers the pose uncertainty to compute the best grasp configuration

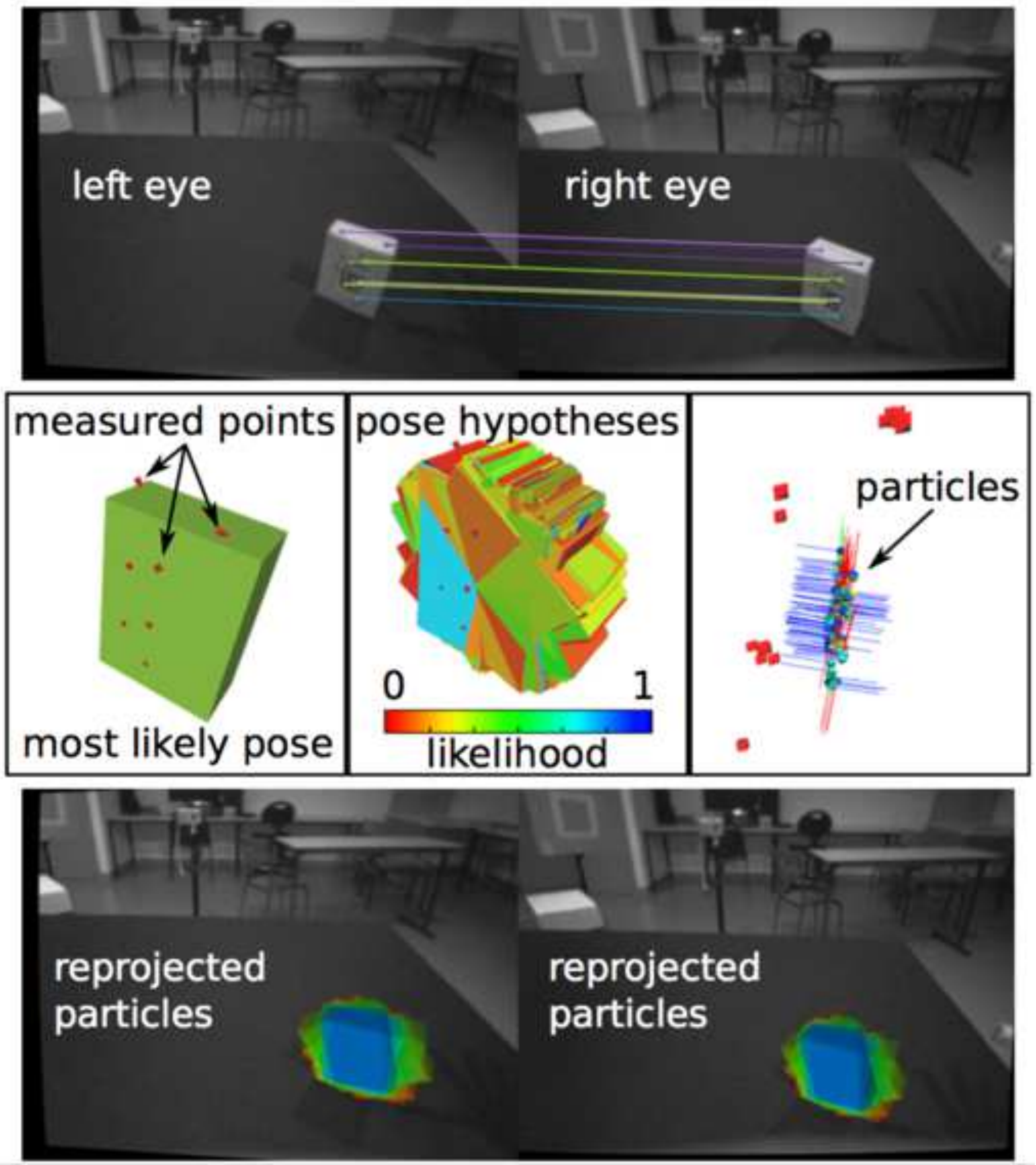


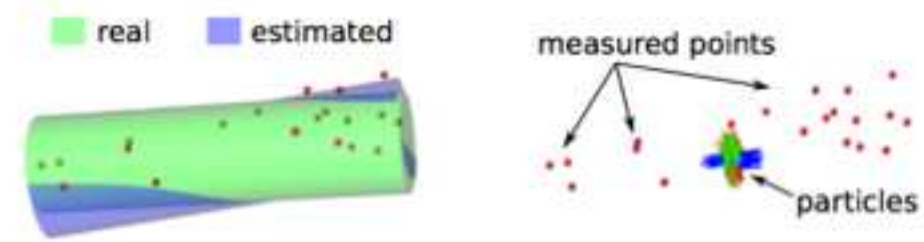
- Inputs: point cloud, object model (primitive or 3D mesh)
- Step 1: Estimate the probability distribution of the object pose with a set of particles/hypotheses
- Step 2: Build a set of stable grasps & compute scores

Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. RAS

Grasping objects localised from noisy point clouds

- Step 1: Estimate the probability distribution of the object pose and a set of particles/hypotheses

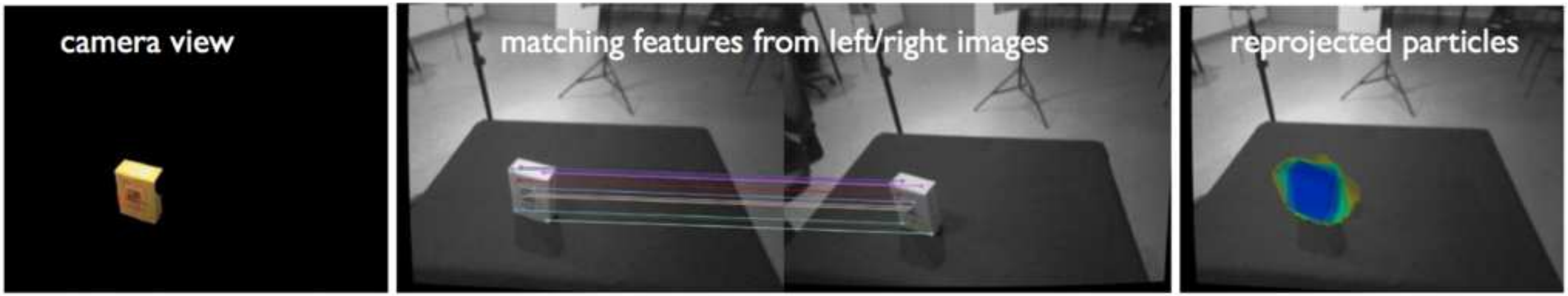




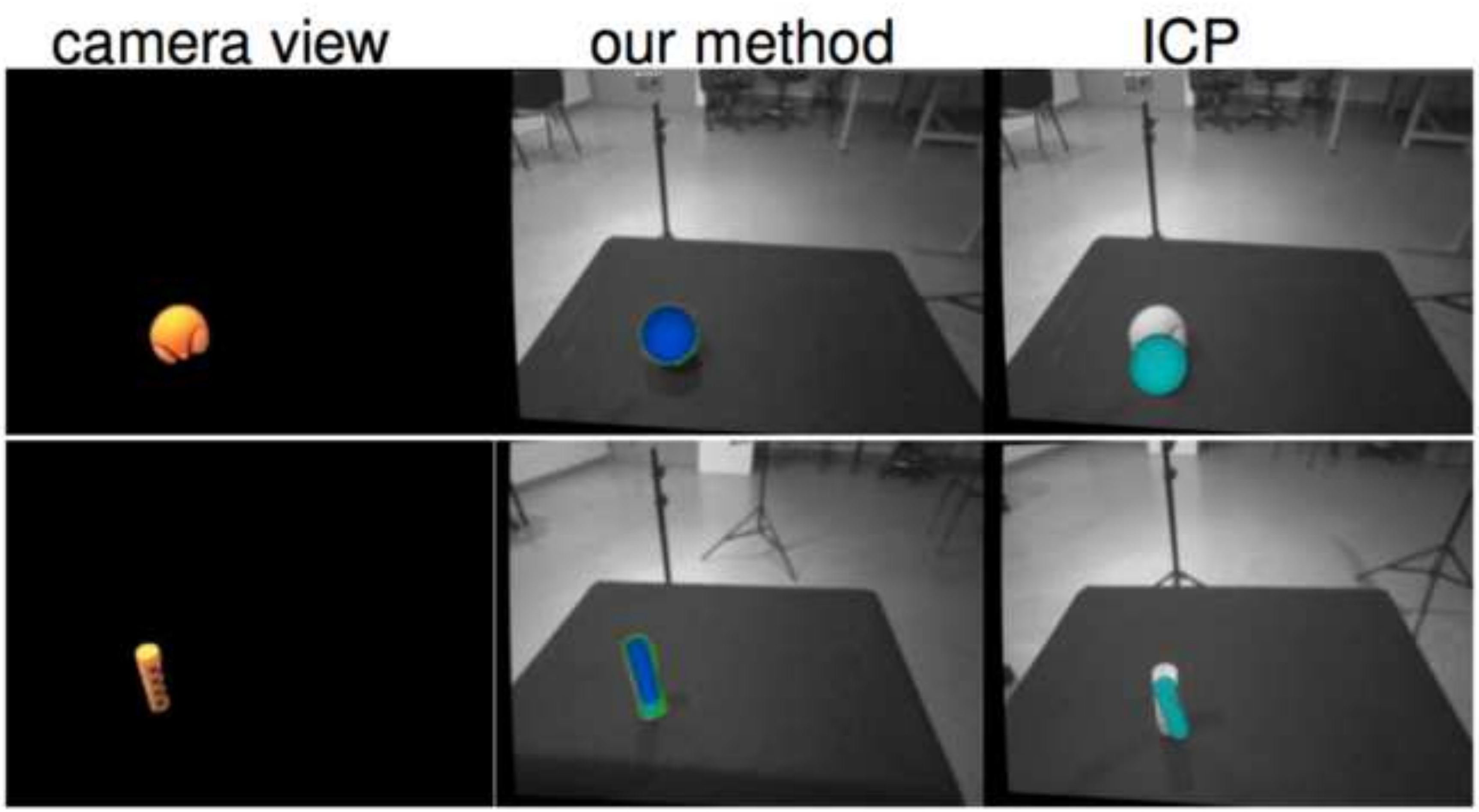
- a particle X is an hypothesis on the object pose
- we have m measured points (from the point cloud)
- we can compute the probability of a candidate object pose X given m measured points d

$$p(X|d_1, \ldots, d_m)$$

Grasping objects localised from noisy point clouds



- note: the most likely pose (blue) does not fit perfectly to the real object pose
  ➡ interest for reasoning with a distribution and not with a single best estimate



Grasping objects localised from noisy point clouds

- Step 2: Build a set of stable grasps & compute scores

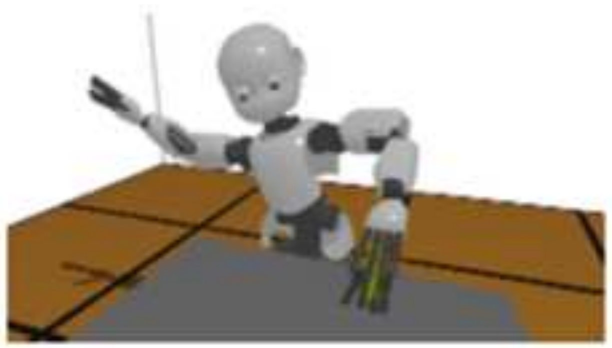
n=144 evaluations of grasps

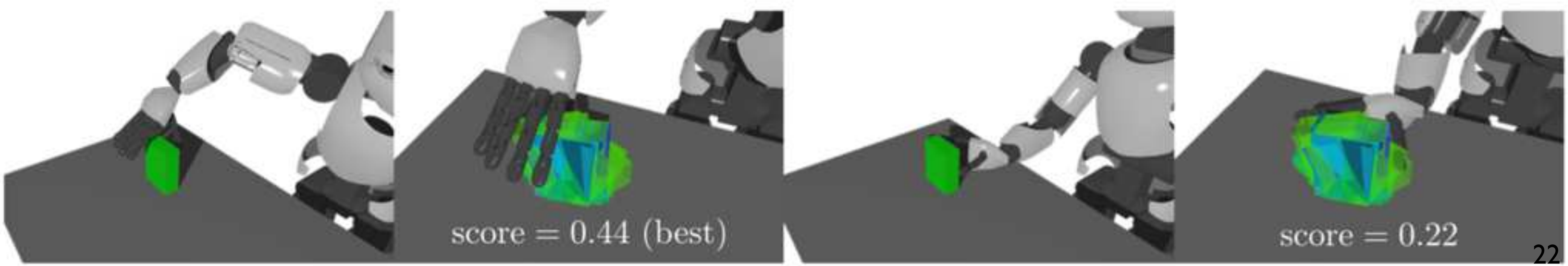
compute probability of success S of the candidate grasp T:

$$p(S\,|T_{grasp}) = \frac{1}{n}\sum_{i=1}^{n} p(X_i|d_1,\ldots,d_m)\,p(S\,|T_{grasp},X_i)$$

probability of the object pose X given the observations d, computed at step 1

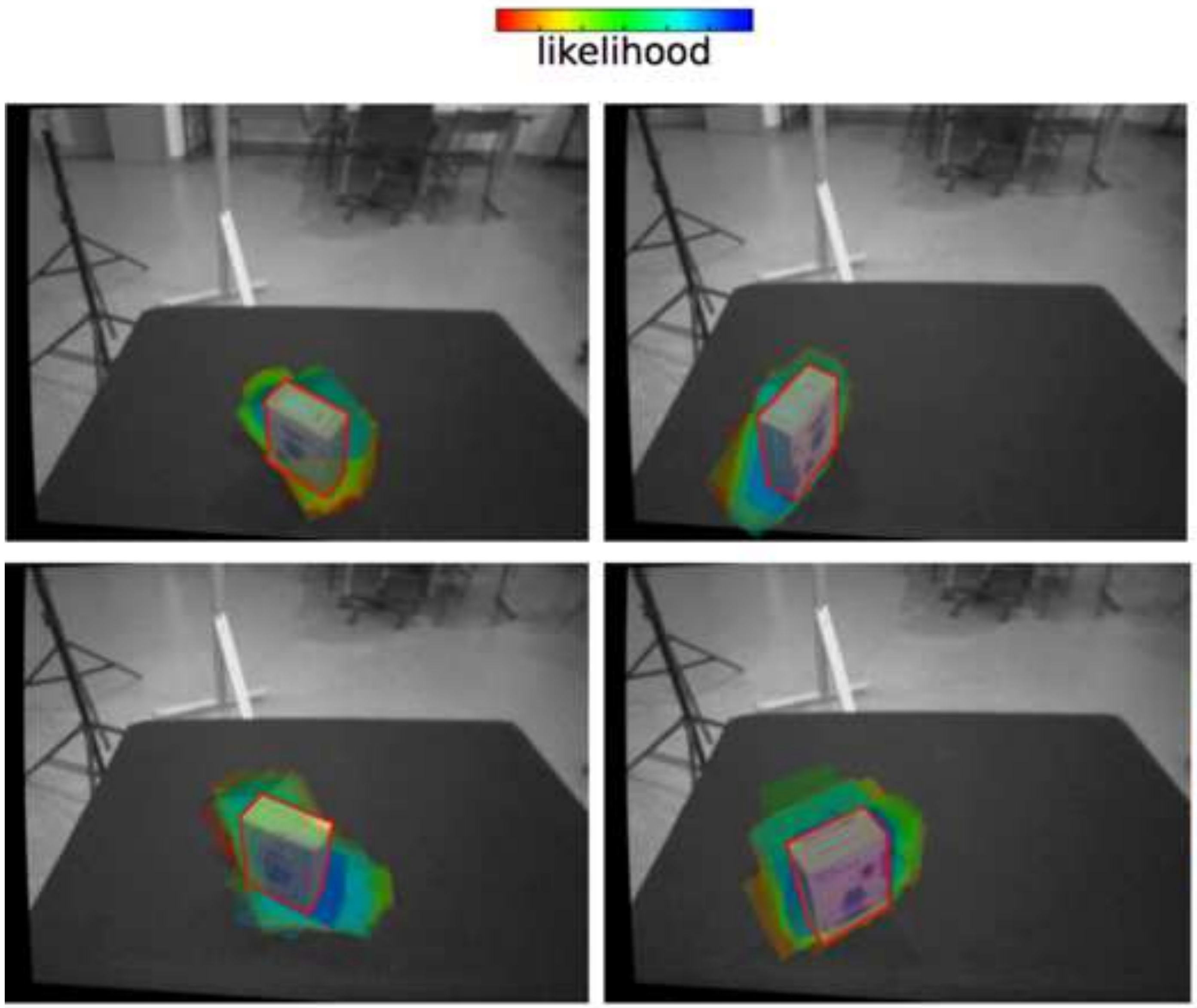
= {0=invalid, 1=valid}, evaluated in simulation



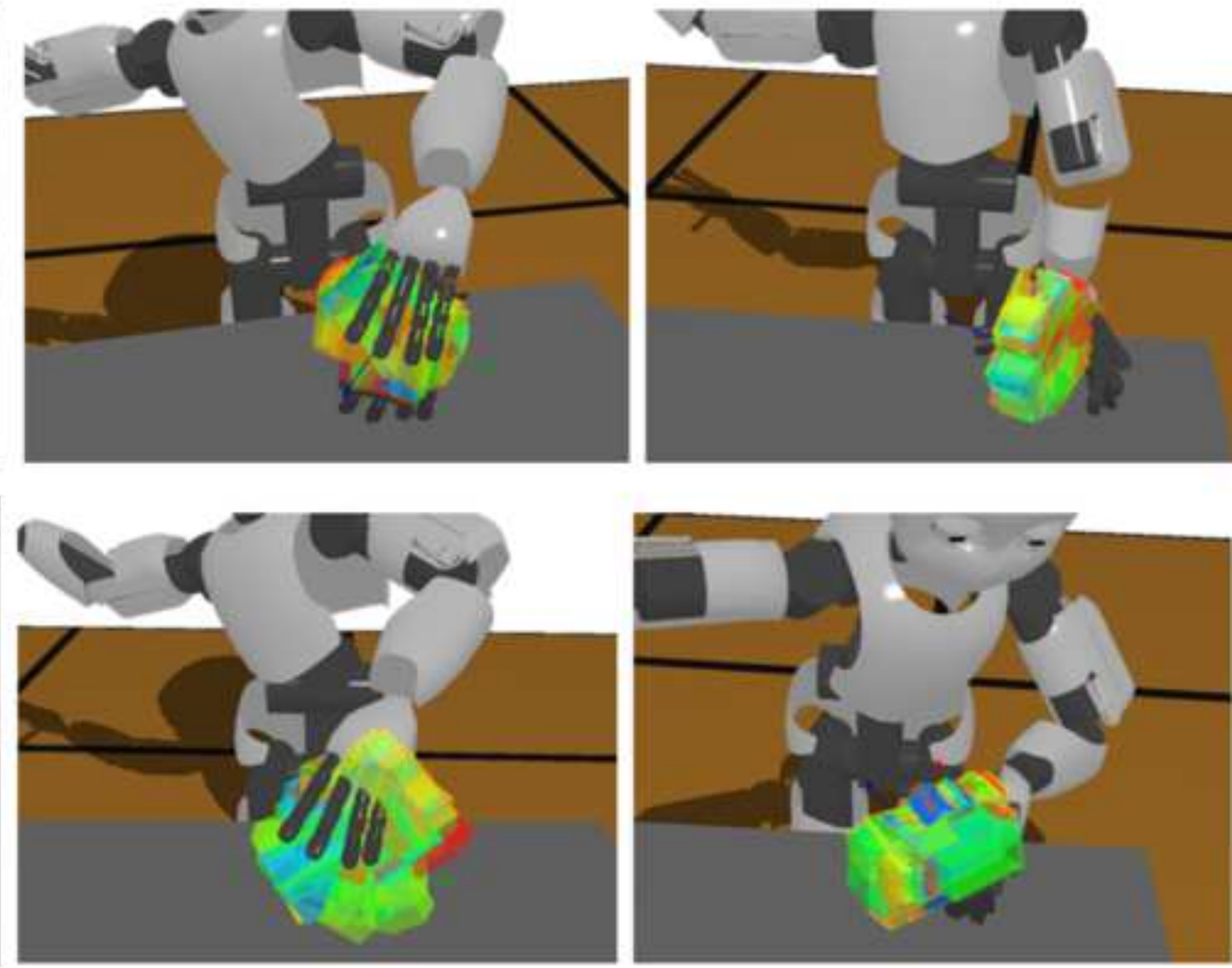


Grasping objects localised from noisy point clouds

- Each different pose & orientation of the object yield different grasps



Reprojection of the particle set on the left image



Grasp that was ranked first in the likelihood to succeed

Grasping objects localised from noisy point clouds

- **We use the probability distribution of the object pose to help selecting the grasp that is more likely to succeed considering the possible poses**

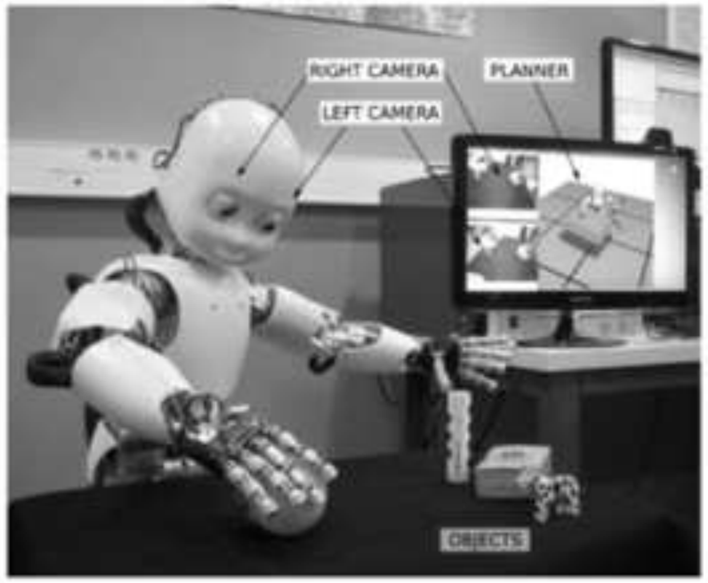
- **Pro**:
  - It is possible to plan a successful grasp direction from a sparse noisy point cloud acquired by (noisy) stereo cameras
  - It can help compensating the absence of tactile sensing in the fingers

- **Cons**:
  - Need a prior object model
  - Computational time required to find the most suitable grasp (~seconds, less than ICP in any case)
    => learning?

Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. Robotics and Autonomous Systems, 62(12): 1742-1754.

Outline of the talk

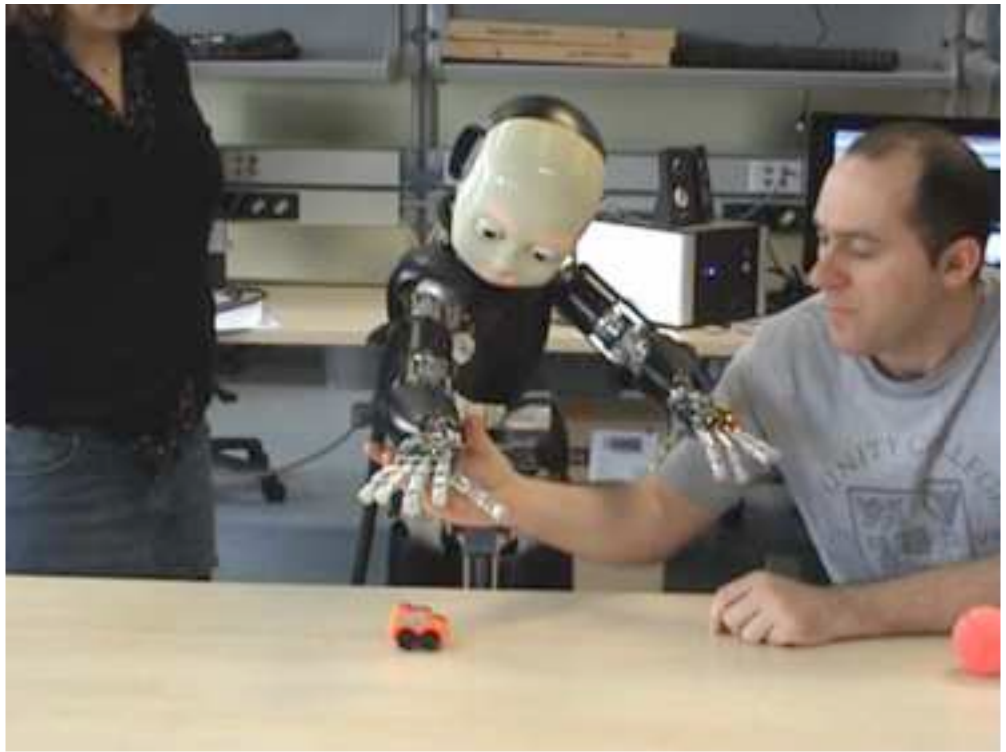
Multimodal learning of the visual appearance of objects (w/ Kinect)

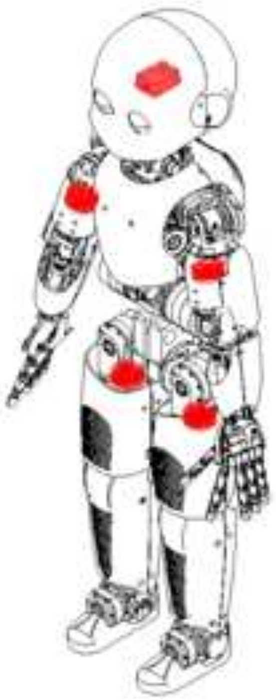


Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Physical interaction: even non-experts can teach iCub how to assemble objects

Teaching object manipulation via physical HRI

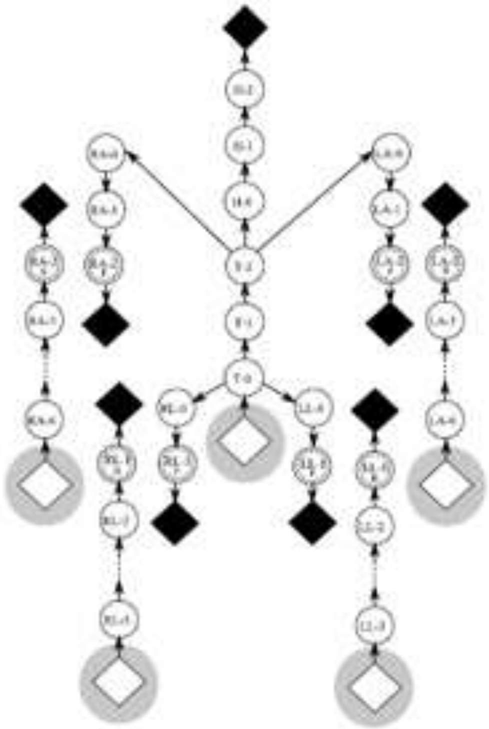




Inertial sensor

F/T sensor

contacts
by skin

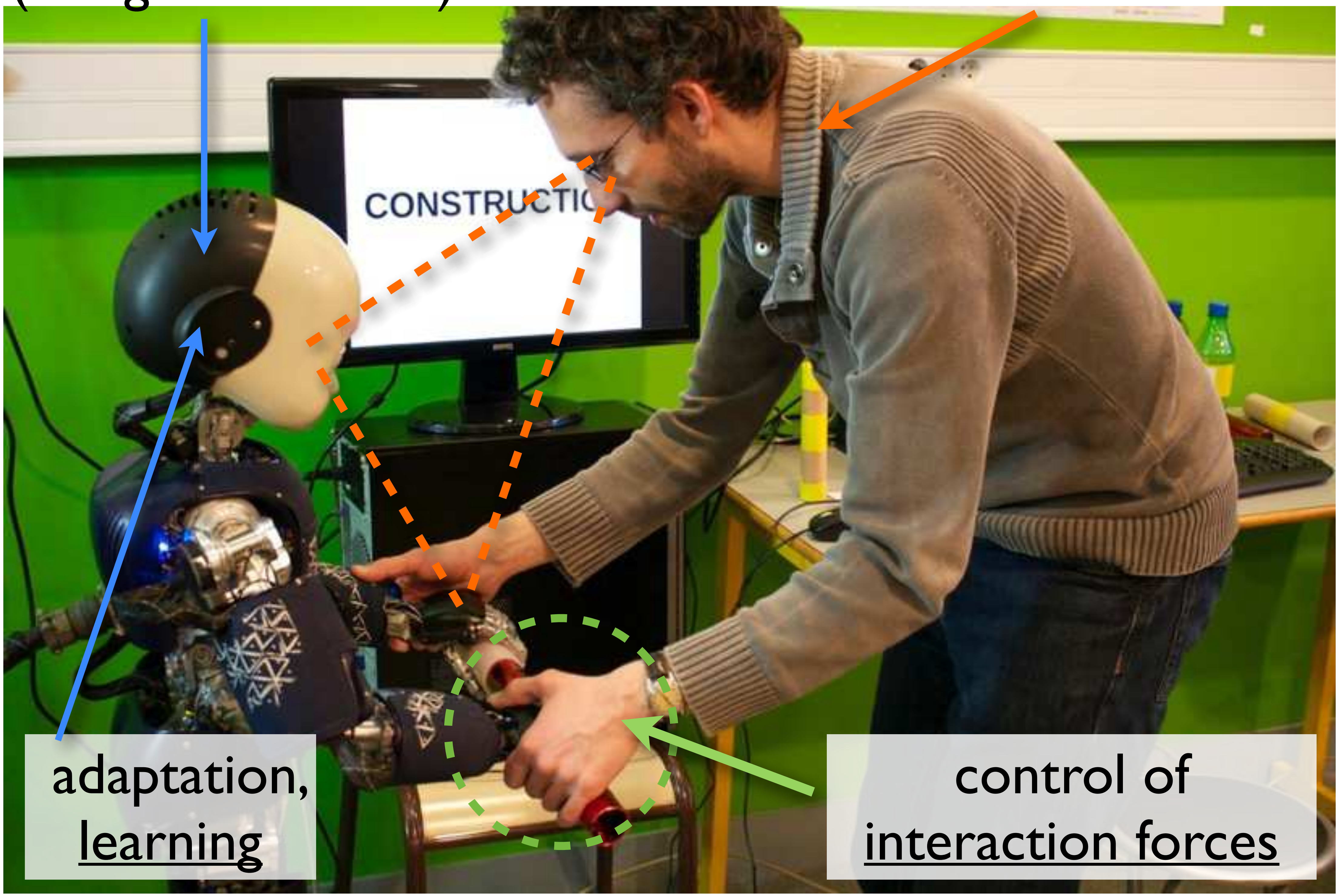


Ivaldi, Fumagalli, Randazzo, Nori, Metta, Sandini. Computing robot internal/external wrenches by inertial, tactile and FT sensors: theory and implementation on the iCub. HUMANOIDS 2011, Autonomous Robots 2012

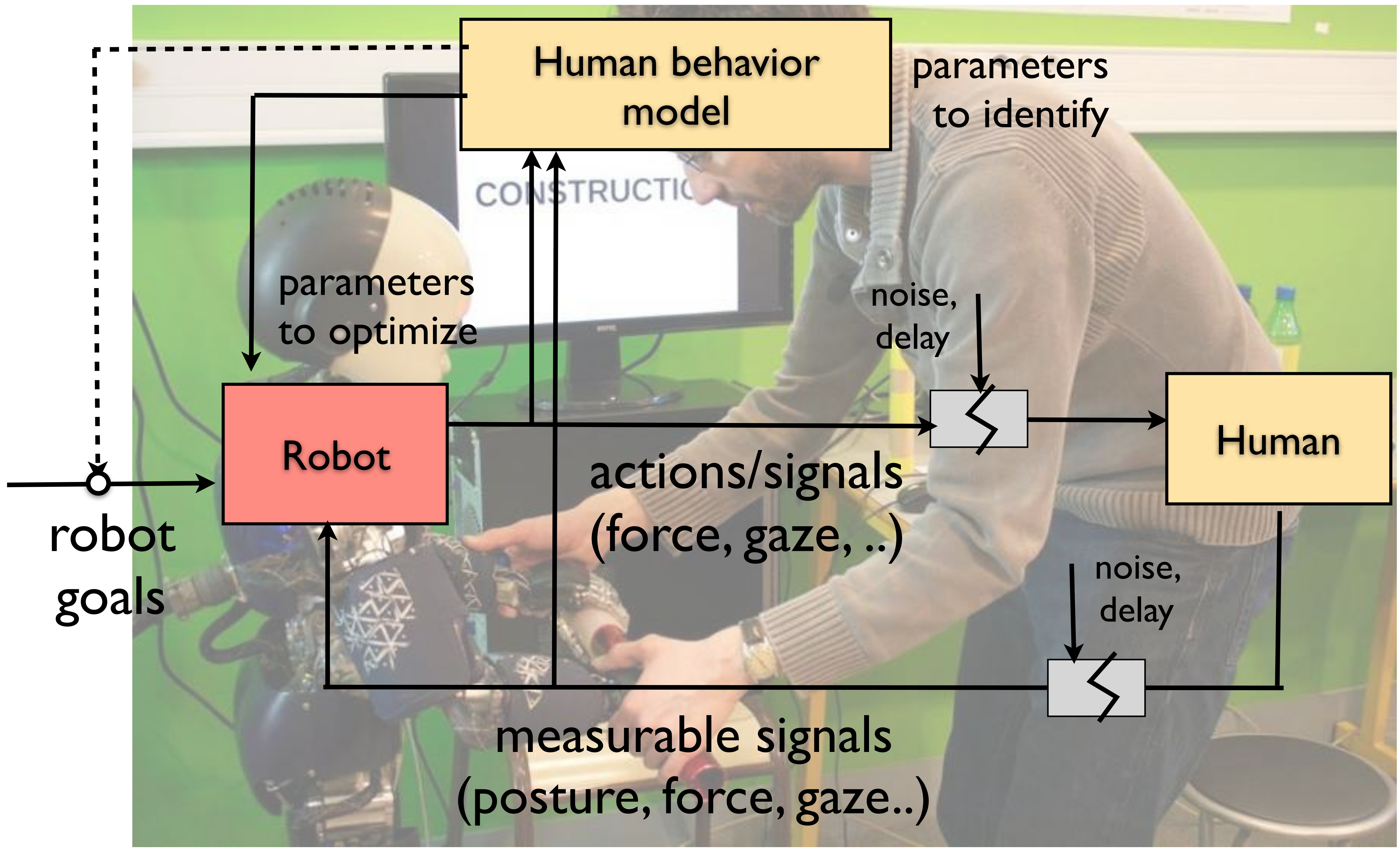
Physical & social interaction

multimodal “behavior” control
(use/give feedback)

verbal/non-verbal signals



Physical & social interaction



Ordinary people teach iCub how to assembly an object

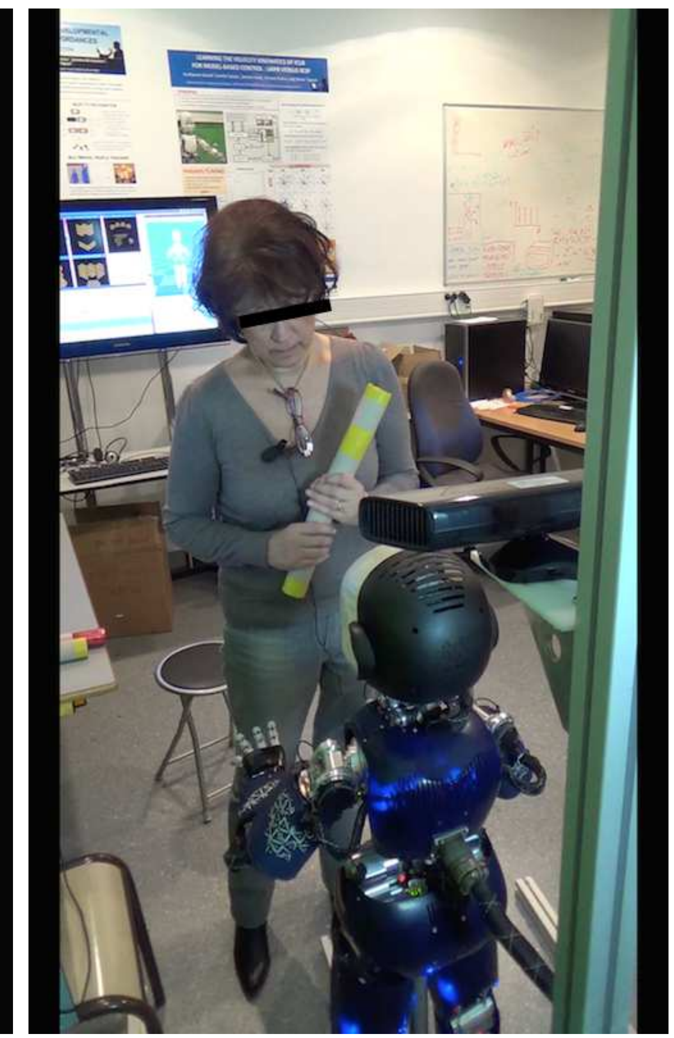
1. How do people behave (gaze, touch, posture, ...) during physical interaction?
2. How much force do they apply on the robot?
3. Do these measures change depending on their expertise with robots, their personality and attitudes?

Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Ordinary people teach iCub how to assembly an object

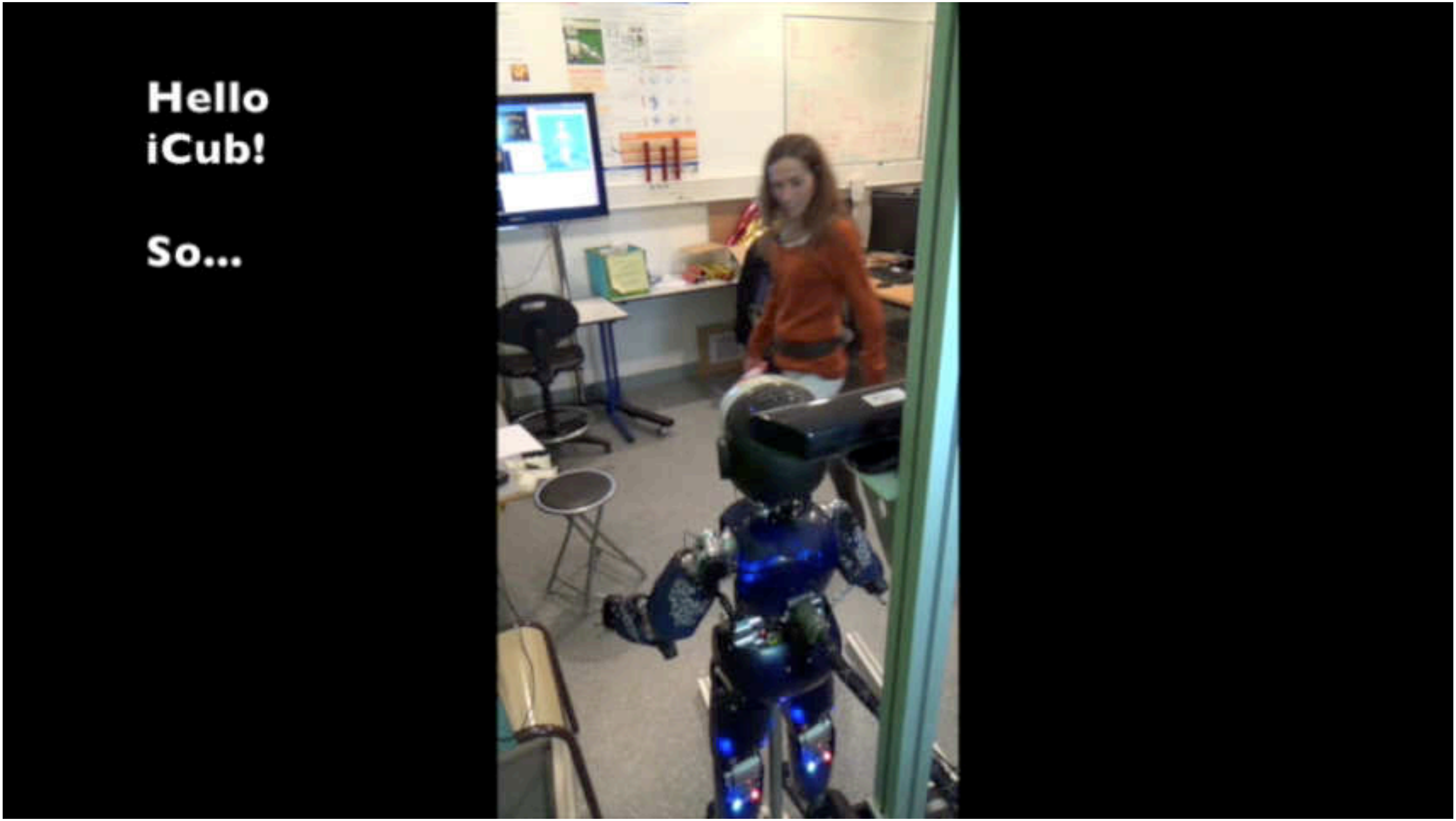
- 56 subjects
- age : 36,95±14,32 (min 19, max 65)
- sex : 19 male, 37 females



Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Individual factors appear in the interaction

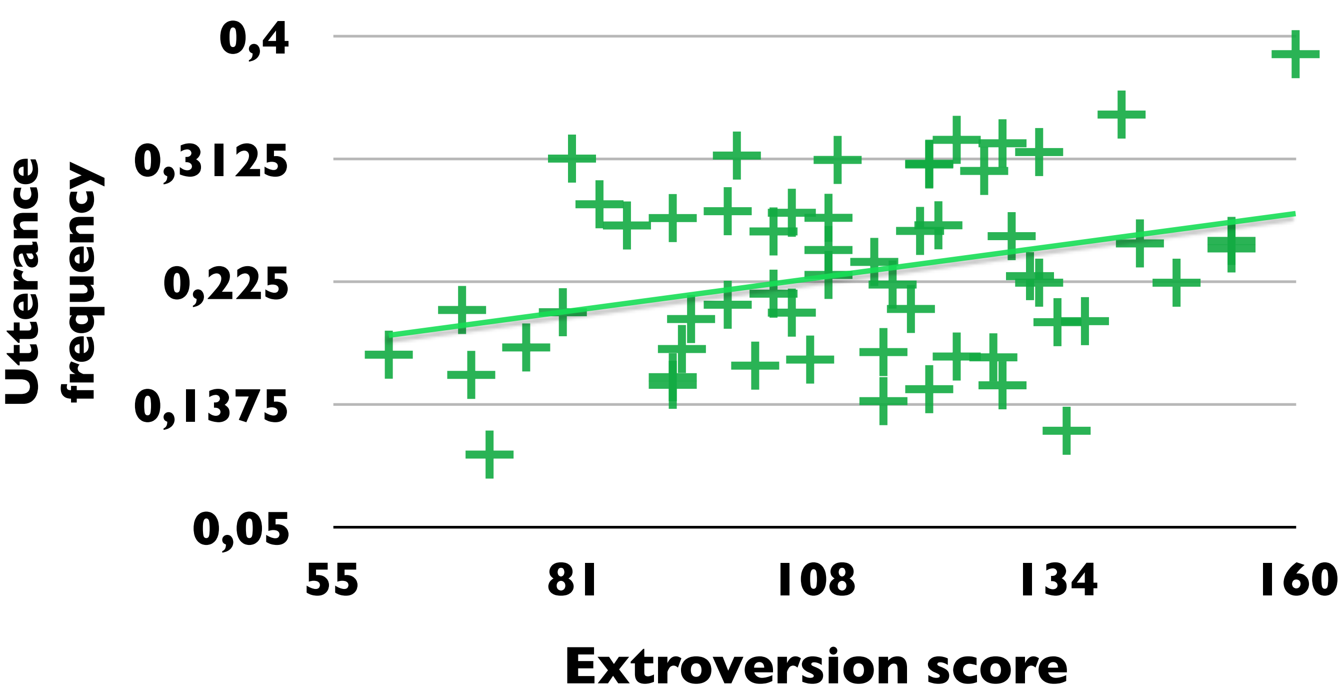


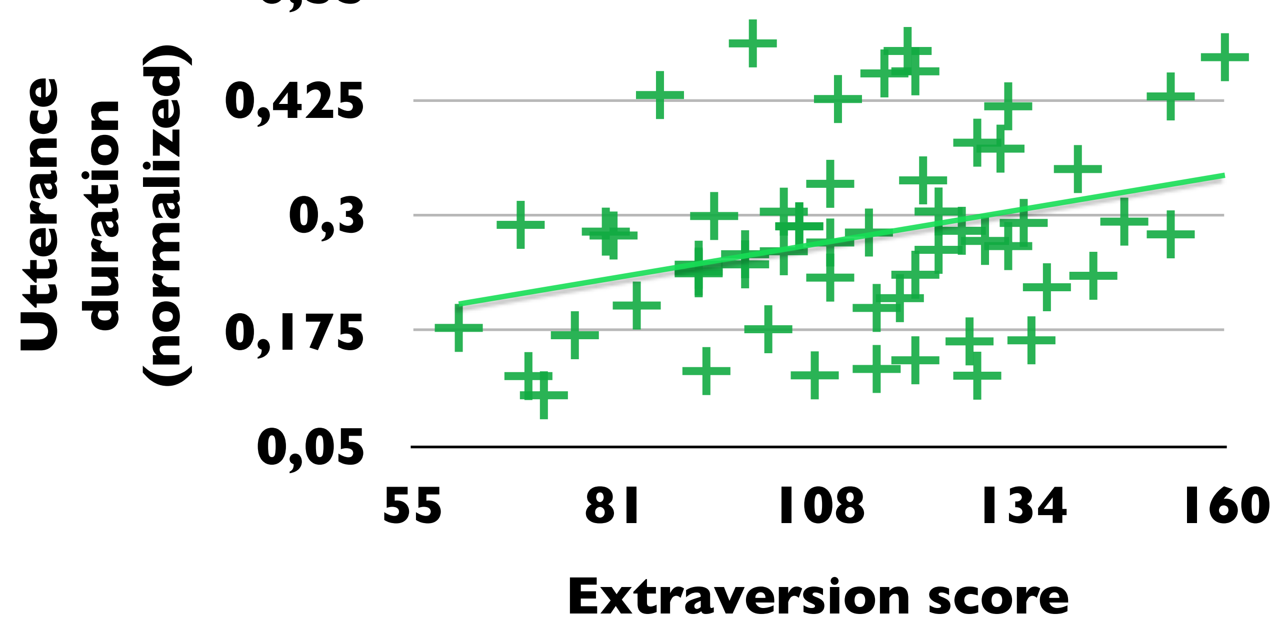
Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Assembly: personality effects on speech

Extroverts talk more to the robot

Variable	Extroversion score
Utterance frequency	r^2= 0,318 ; **p=0.017 <0.05**
Utterance duration	r^2= 0,321 ; **p=0.016<0.05**



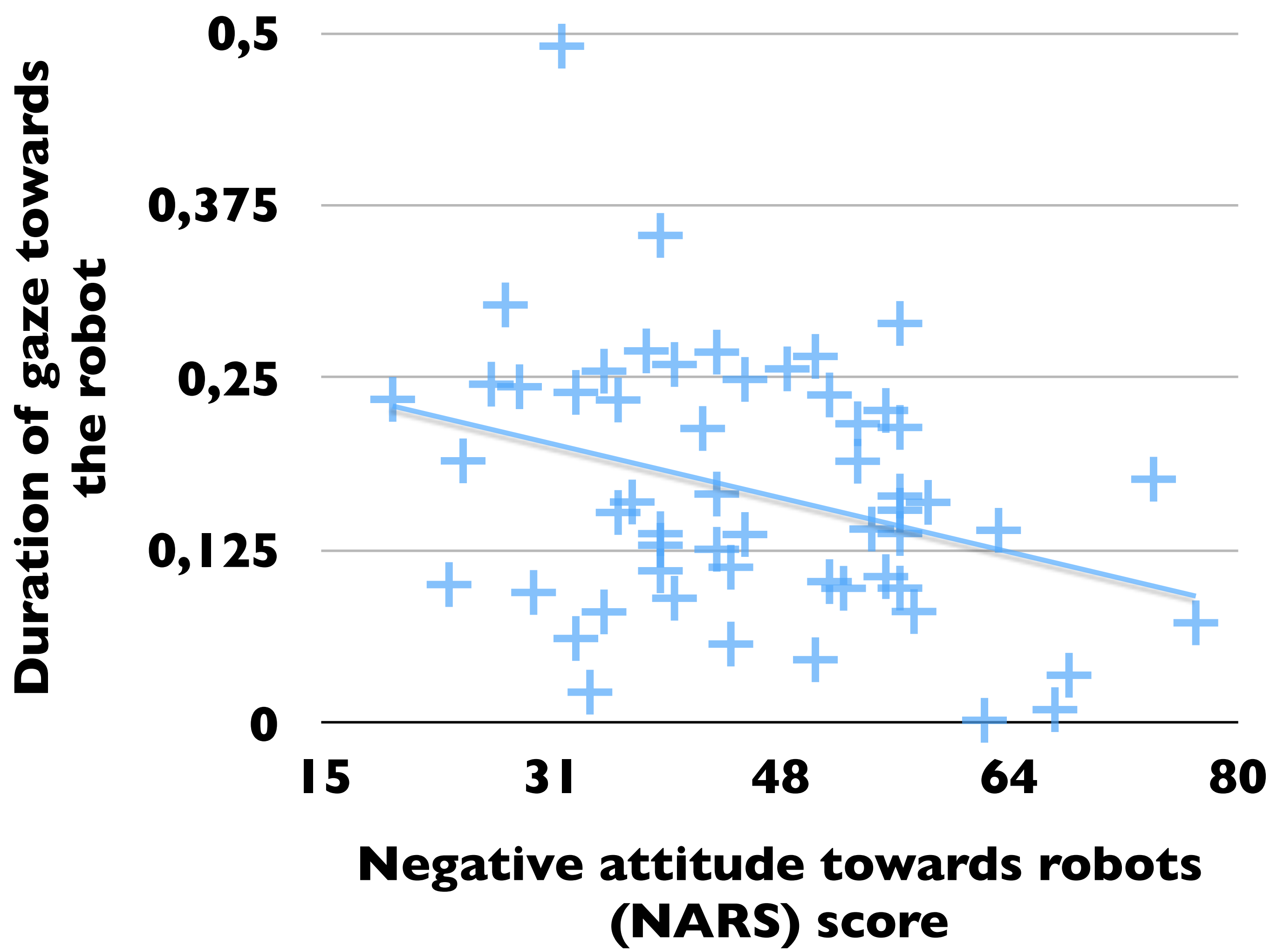


Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Assembly: personality effects on gaze

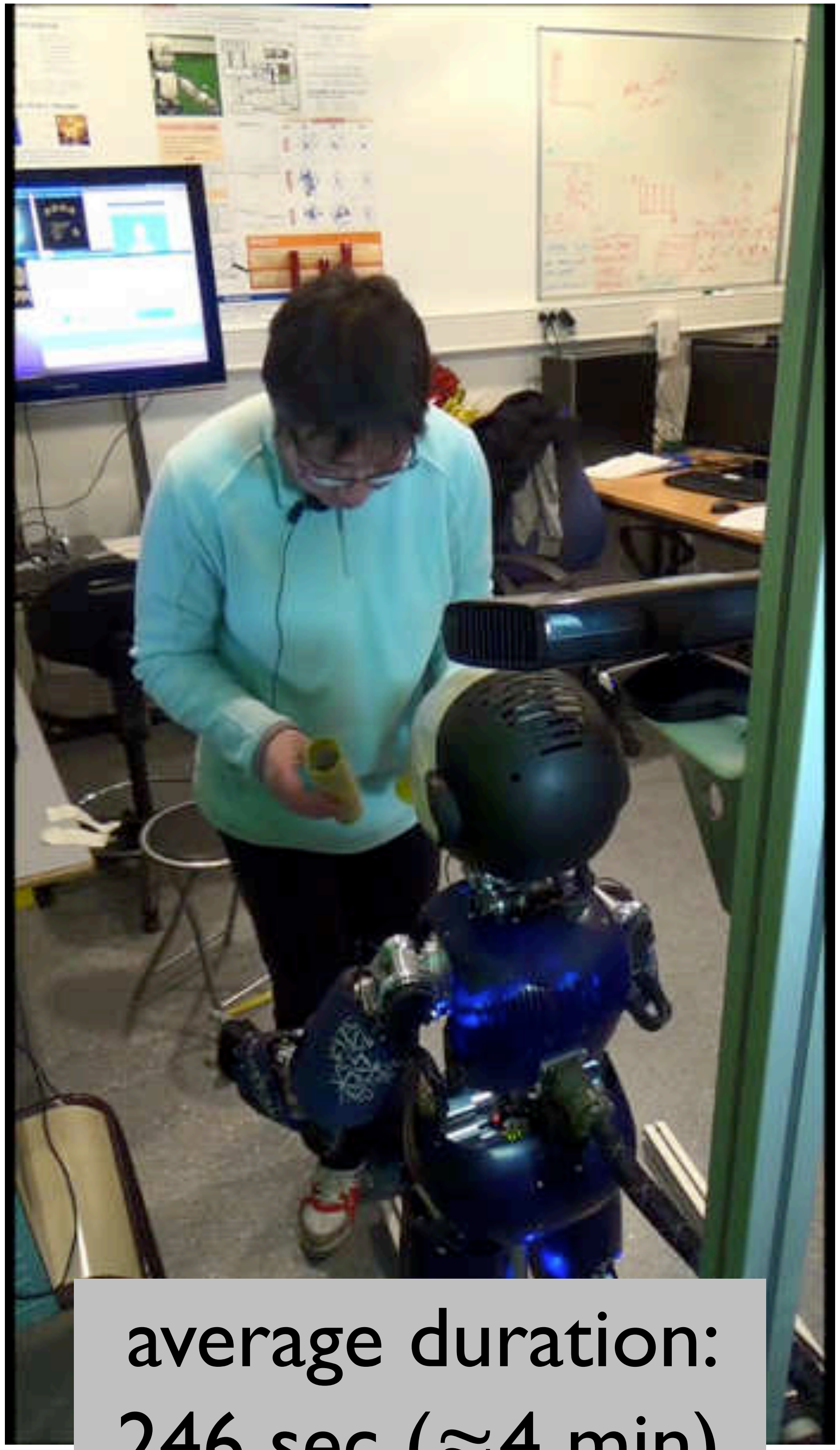
People with negative attitude towards robots look at the robot face for shorter time, and more at the hands where the physical interaction occurs.

Variable	Score "negative attitude towards robots"
Gaze towards face duration	r^2= -0,331 ; p=0.013<0,05
Gaze towards hands duration	r^2= 0.355 ; p=0.007 < 0.05



Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

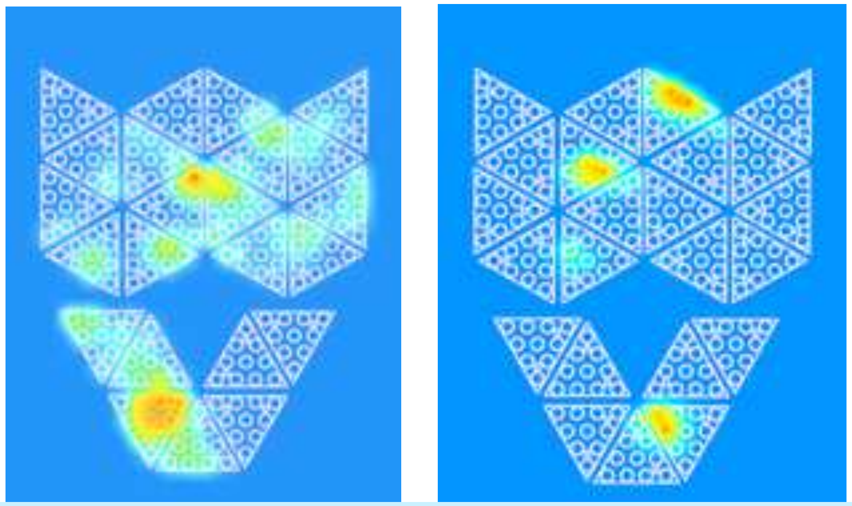
Tactile signatures during teaching



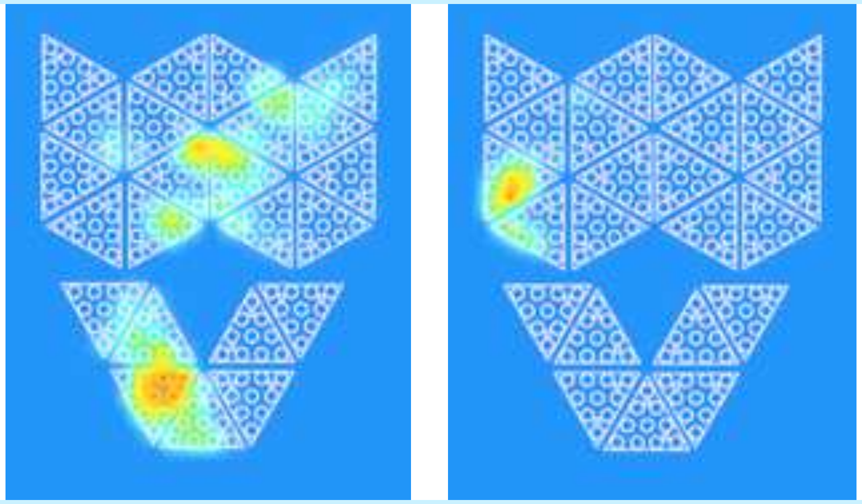
average duration:
246 sec (≈4 min)

faster
less force

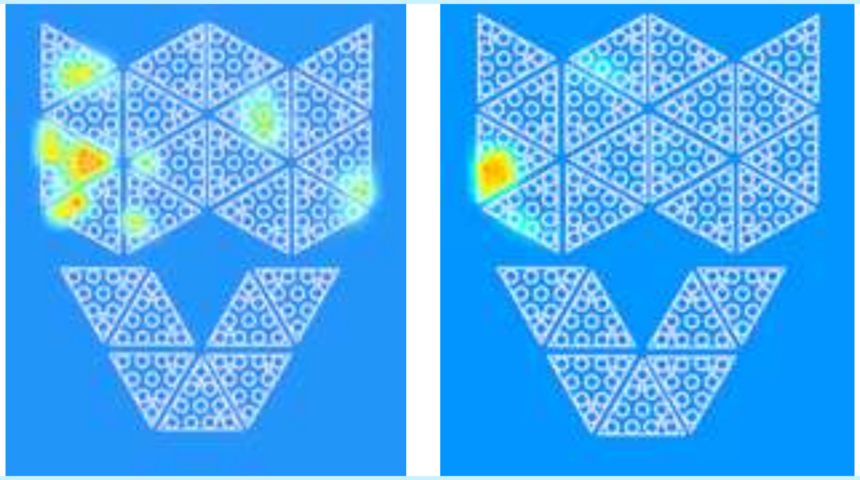
Learning effect



1st trial

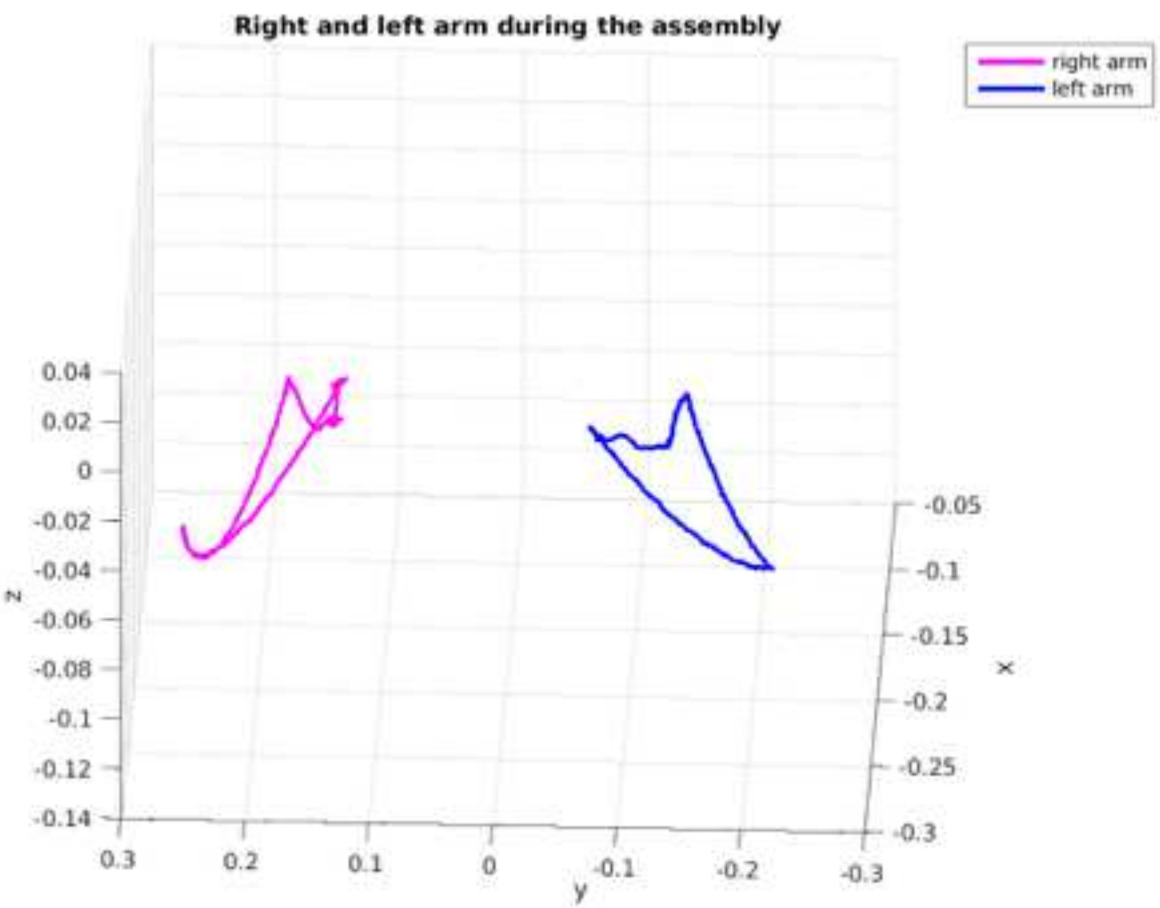


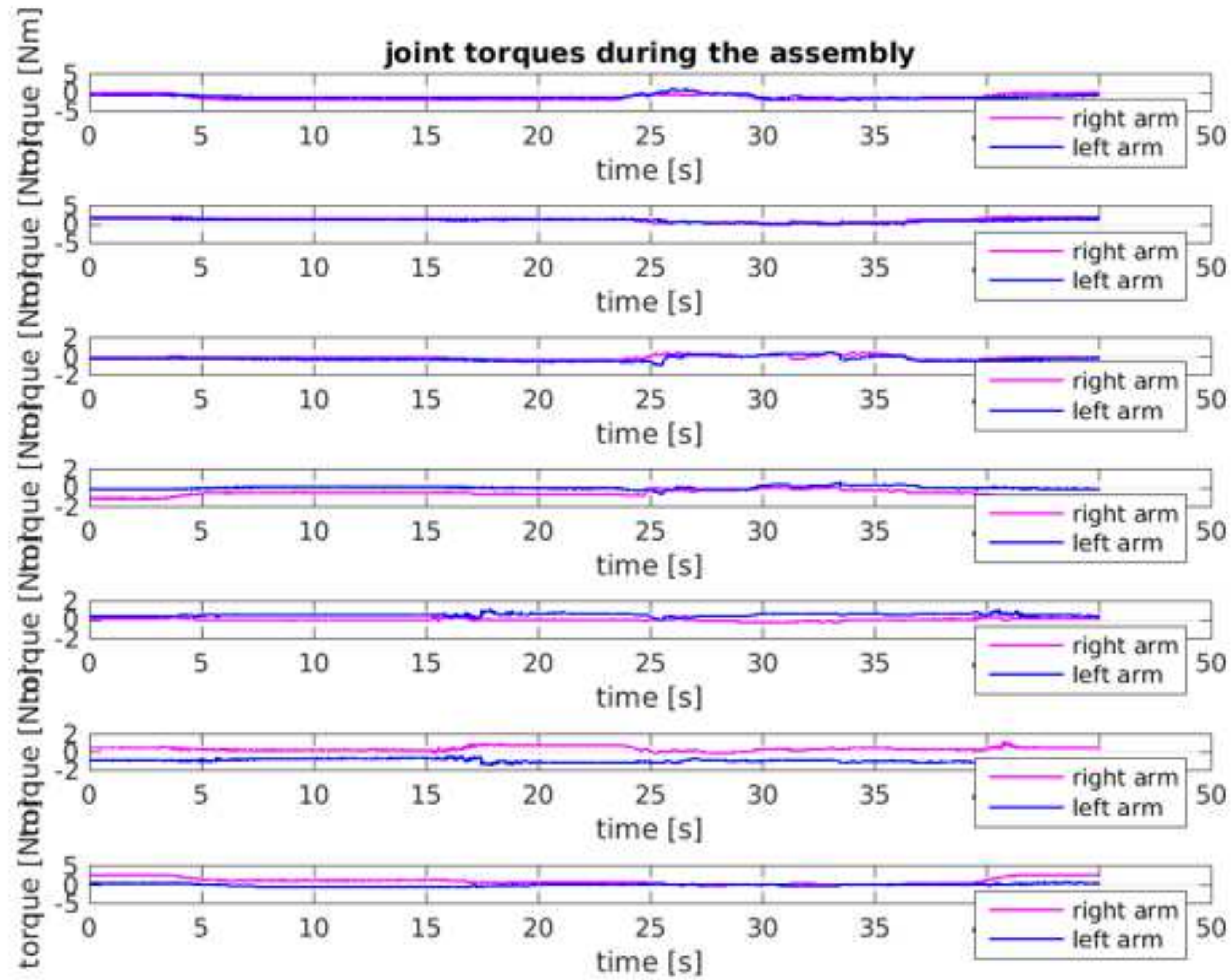
2nd trial

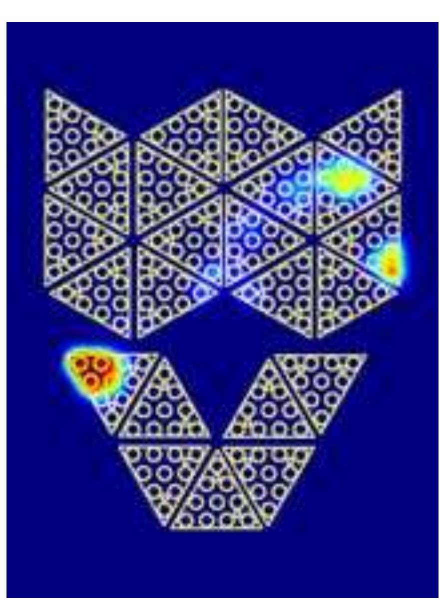


3rd trial

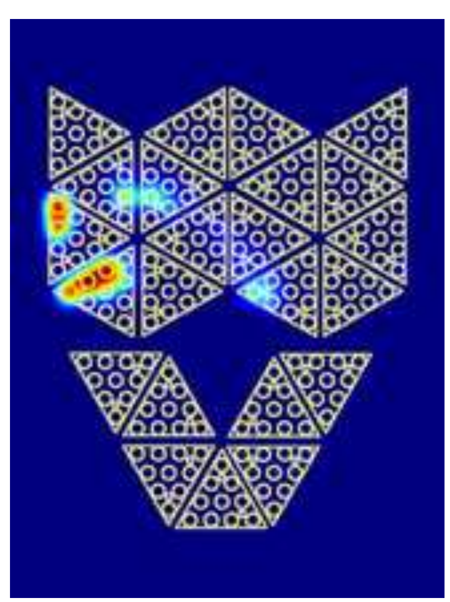
Demonstration from the expert







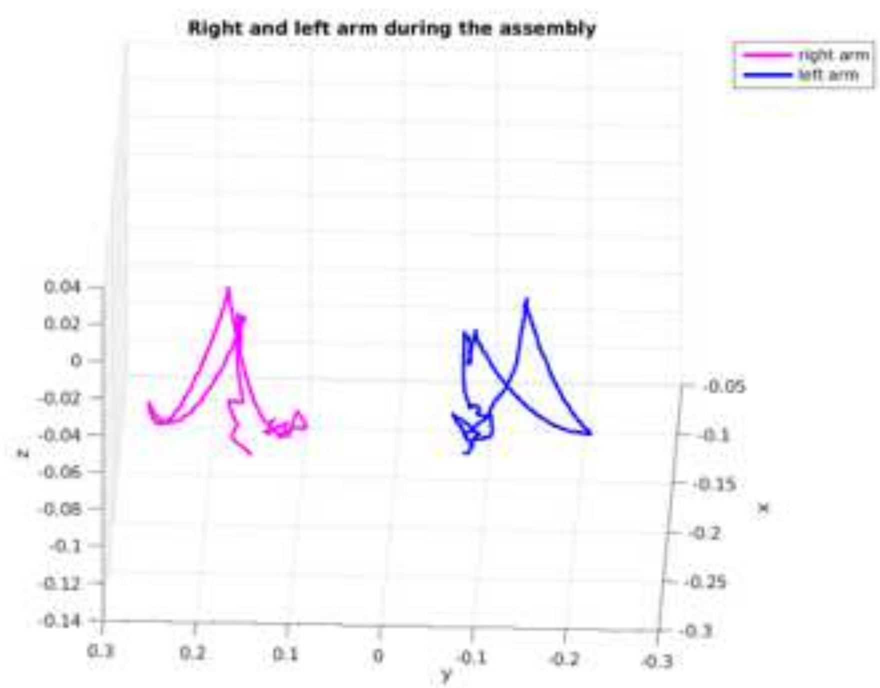
Right Forearm

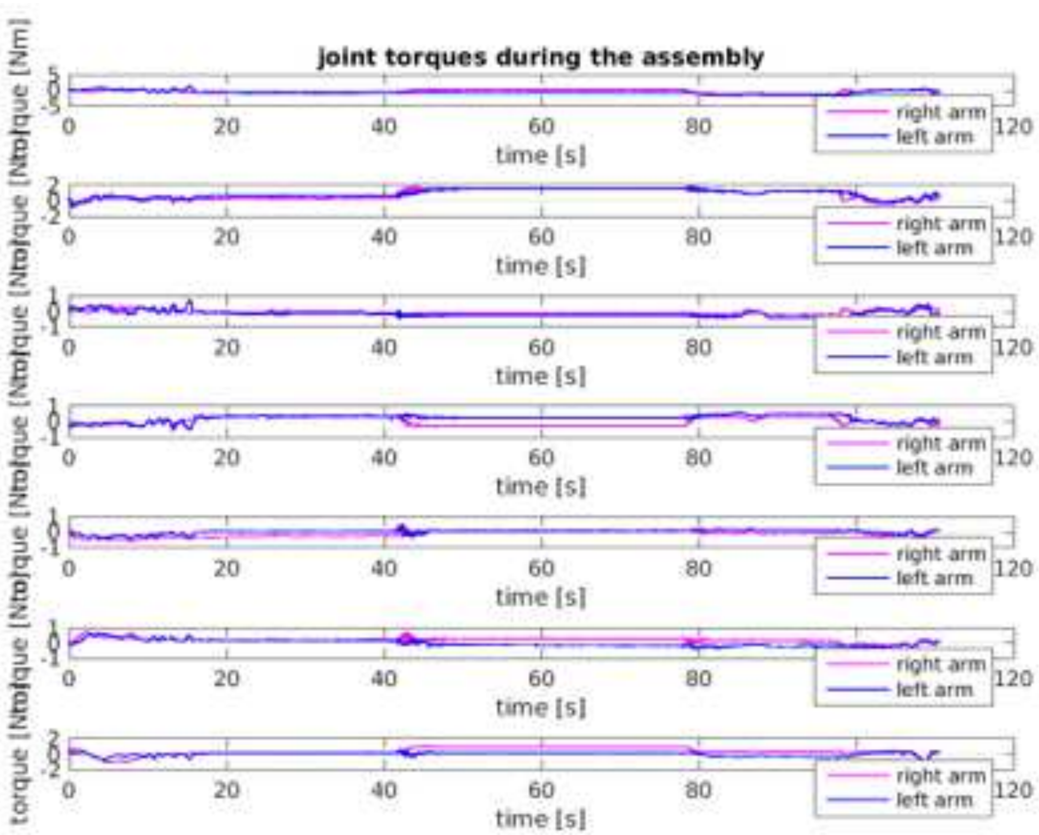


Left Forearm

Trials of the non-expert #62

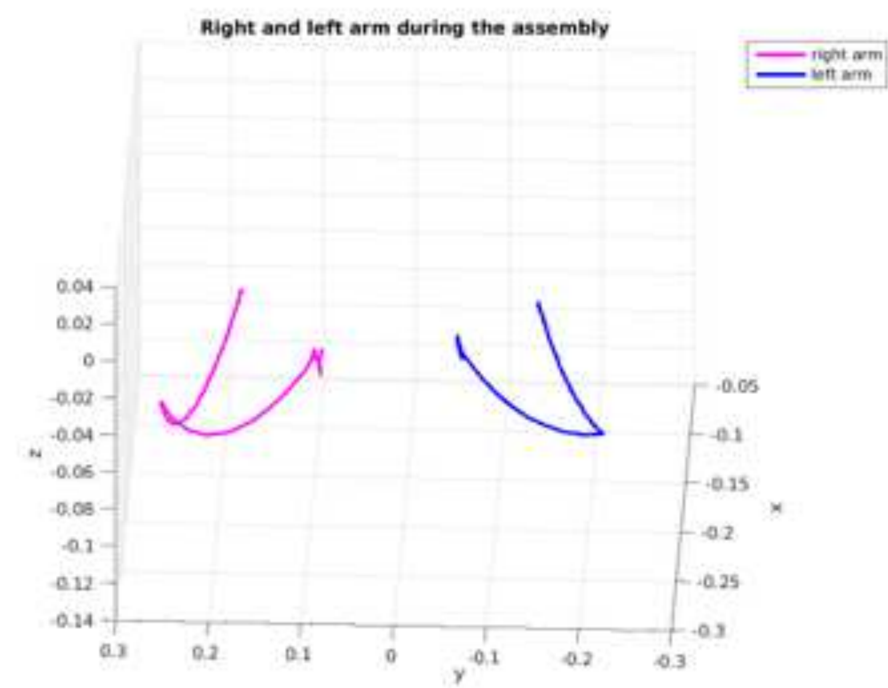
Trial #2

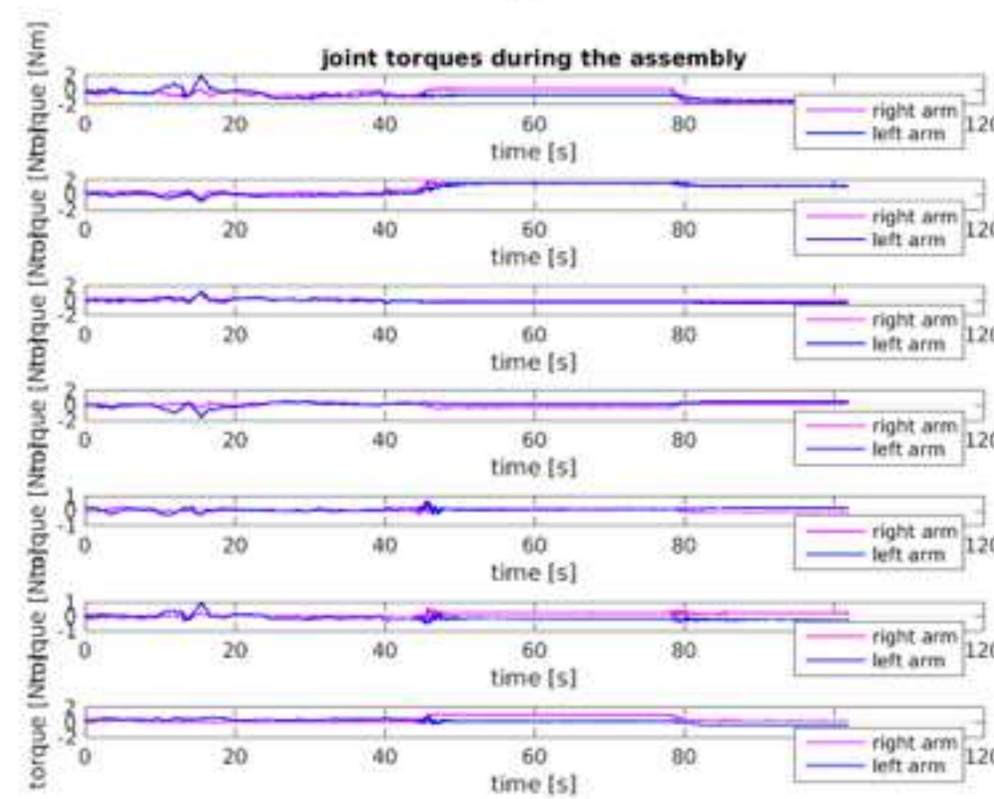




- smoother
- more precise trajectory

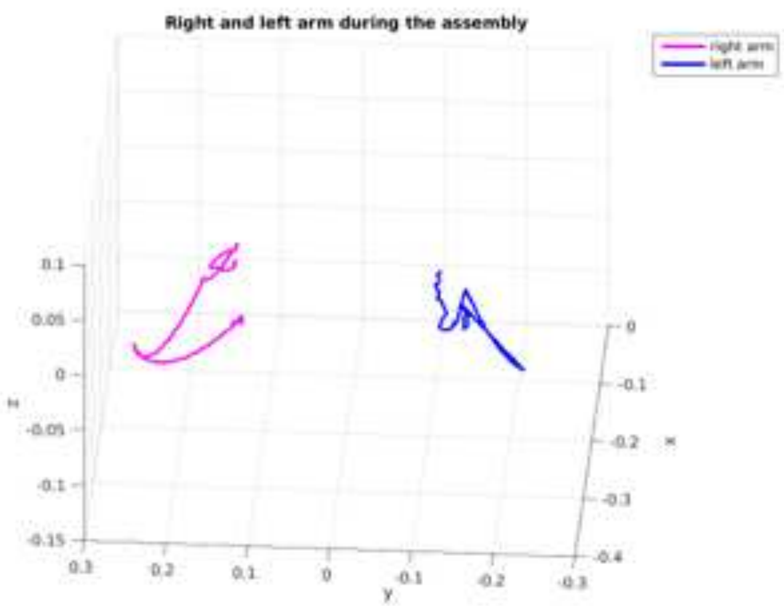
Trial #3

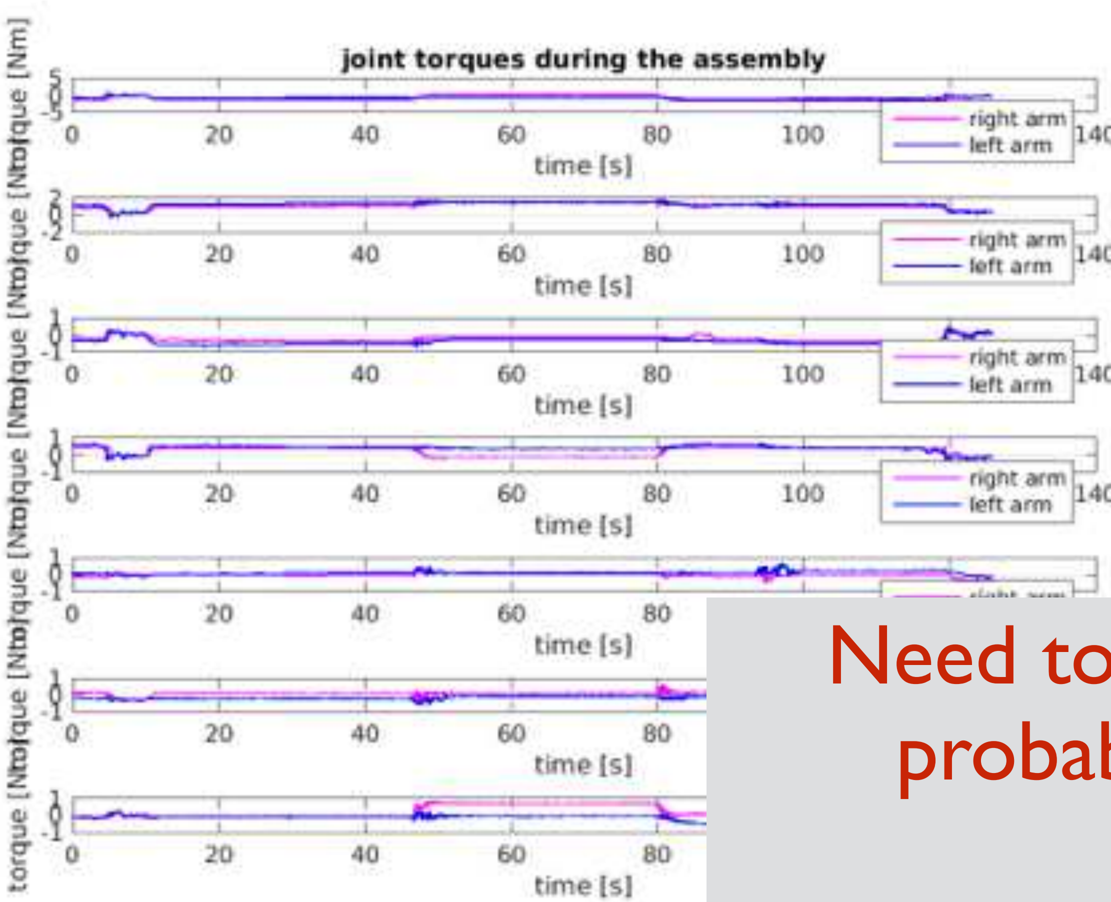




Trials of the non-expert #58

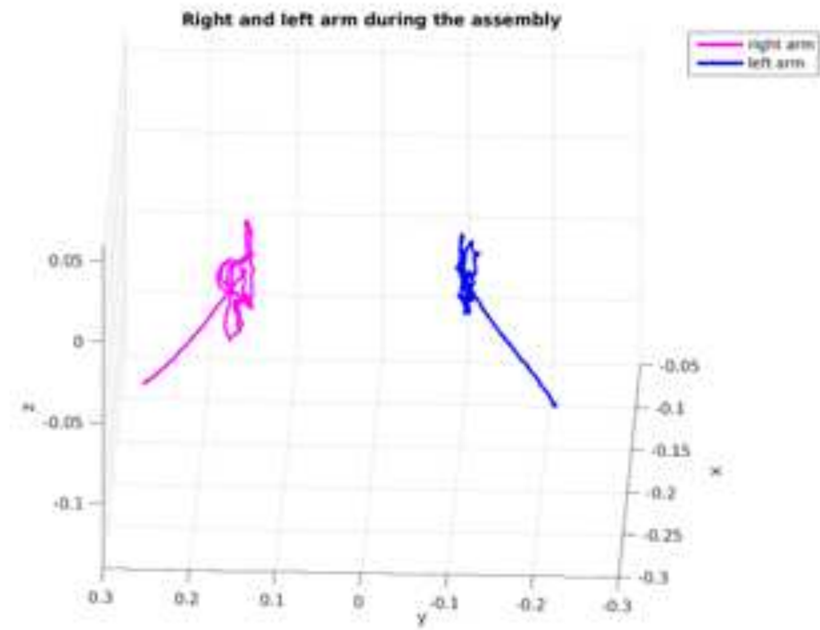
Trial #2

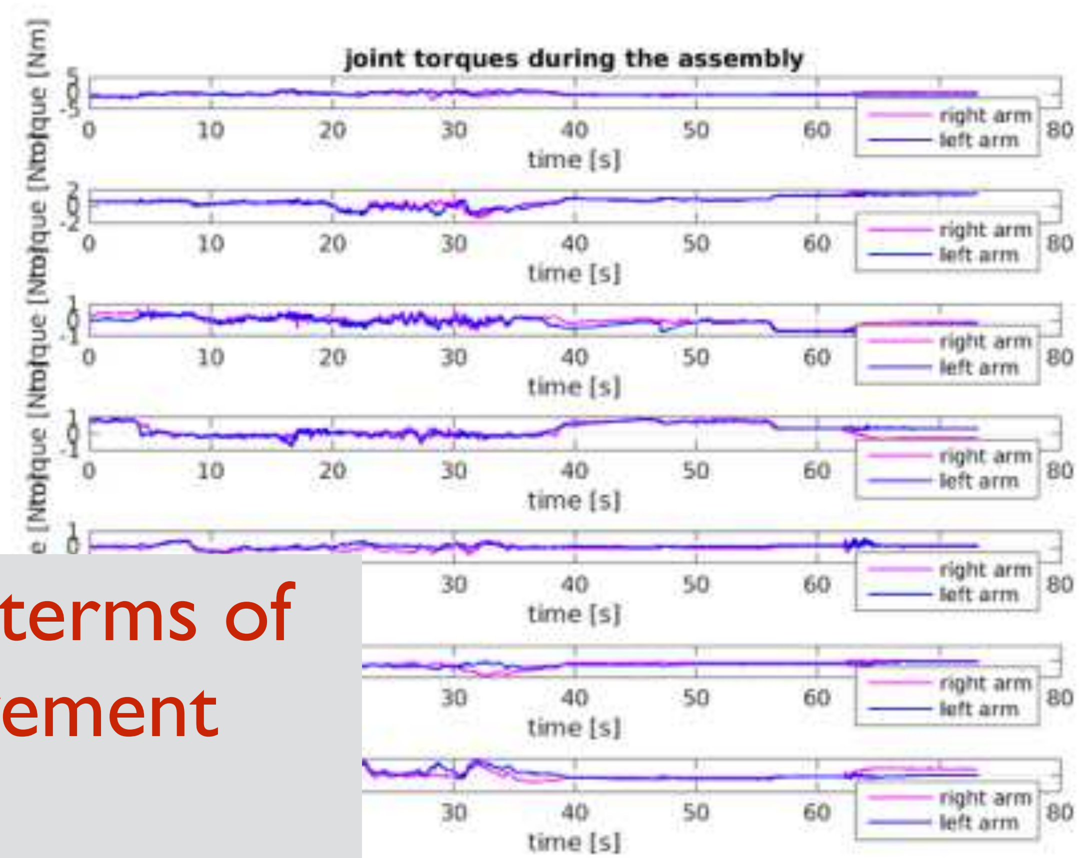




- faster
- precise alignment of the cylinders

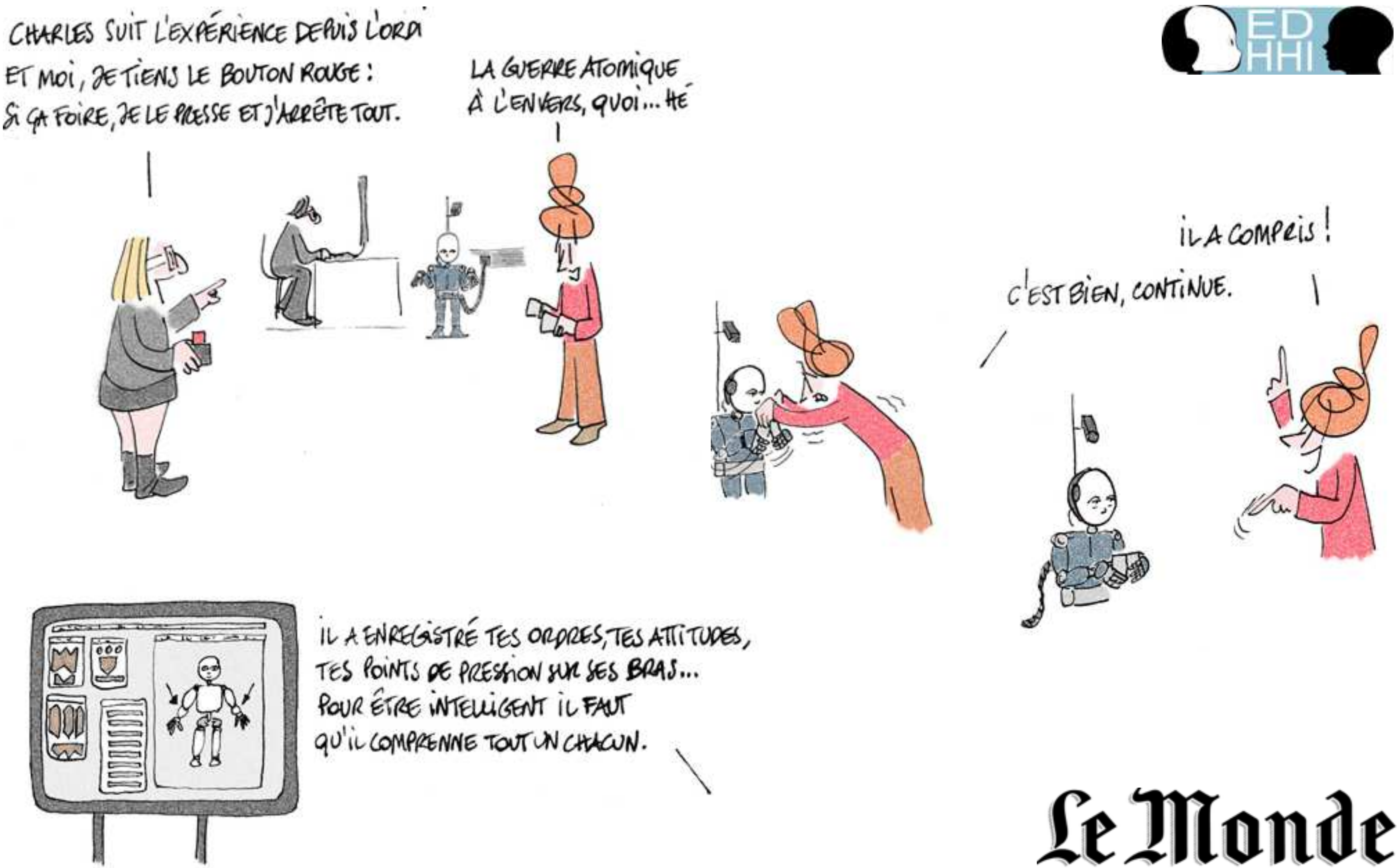
Trial #3





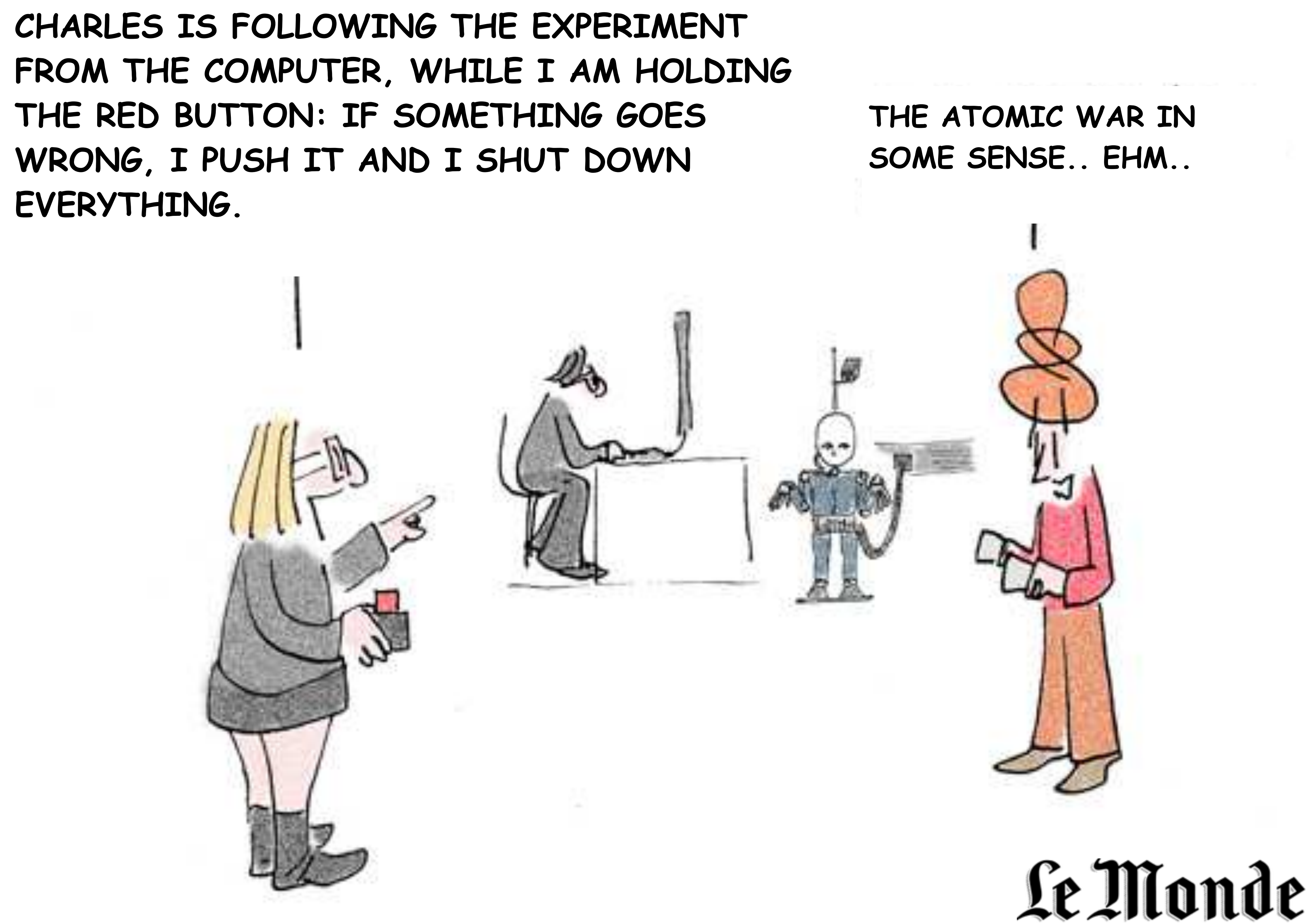
Need to reason in terms of probabilistic movement primitives.

The experiment seen by an artist :)



Comics by Fiamma Luzzati - Le Monde - April 2014

Thank you!
Questions ?



Comics by Fiamma Luzzati - Le Monde - April 2014

Postdocs wanted!

Open postdoc position for 2016 for the project "Learning to walk with iCub" within the ERC Resibots

contacts:
serena.ivaldi@inria.fr, jean-baptiste.mouret@inria.fr

Open postdoc position for 2017 for the project H2020 AnDy - "Ergonomics models for human-robot collaboration"

contacts:
serena.ivaldi@inria.fr