Towards Task Understanding through Multi-State Visuo-Spatial Perspective Taking for Human-Robot Interaction

Amit Kumar Pandey and *Rachid Alami*
LAAS-CNRS
Toulouse, France

IJCAI-ALIHT Workshop, July 16-17, 2011, Barcelona

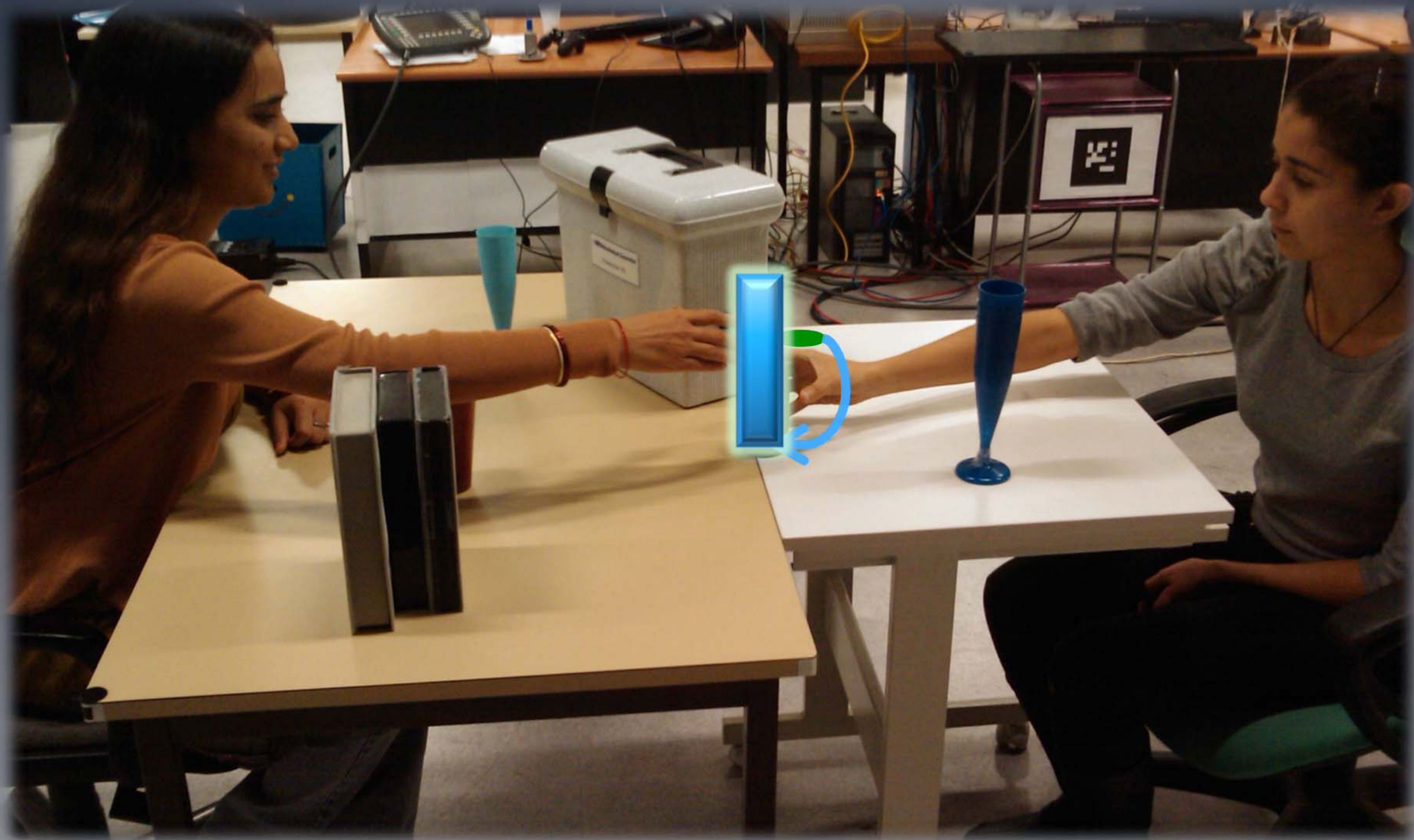
Perspective Taking and Human-Robot Interaction

- Perspective Taking has already been shown important in HRI [Trafton, 2005], in action recognition [Johnson, 2005], in learning [Breazeal, 2006], …
- Then what and why "Multi-State" perspective taking for "task-understanding".

Motivation for Multi-State Perspective Taking

But if she is “good” :

And if she is “better” :



And iff she is the “best” :

But what does this suggest:

We predict various abilities of ourselves and of others

Standing & leaning visibility and reachability of others

Sitting reachability of self and Standing & leaning reachability of others

Sitting but leaning reachability of self and of others

Standing & leaning reachability of self and sitting reachability of others

This suggests that:

- In our day-to-day interactions, we:
  - imagine various achievable states (standing, leaning, etc.)
  - predict various types of abilities (visibility, reachability, etc.)
  - reason on common abilities (visible and reachable by both, etc.)
  - decide mutual comfort levels for a task (commonly reachable without leaning, etc.)

From Human Behavioral Psychology :

- At 12–15 months of age children start showing evidence of understanding :
  - Of <u>occlusion of others line-of-sight</u>,
  - and that an adult is seeing something that they are not when looking at <u>locations behind</u> them or behind barriers.
  - Understanding of

Places/objects Agent Might NOT See | Equal Importance | Places/objects Agent Might See

- e.g. to ***<u>hide</u>*** some object from an agent.

Further from Human Behavioral Psychology :

- From the age of 3 years, children are able to perceive, which places are reachable by them and by others .
- Different types of reach action of the human has been identified and analyzed:

Taxonomy of reach actions:(a) arm-shoulder reach, (b) arm-torso reach, (c) standing reach.

[Gardner, 2001]
[Choi, 2004]

Places/objects Agent Might NOT Reach

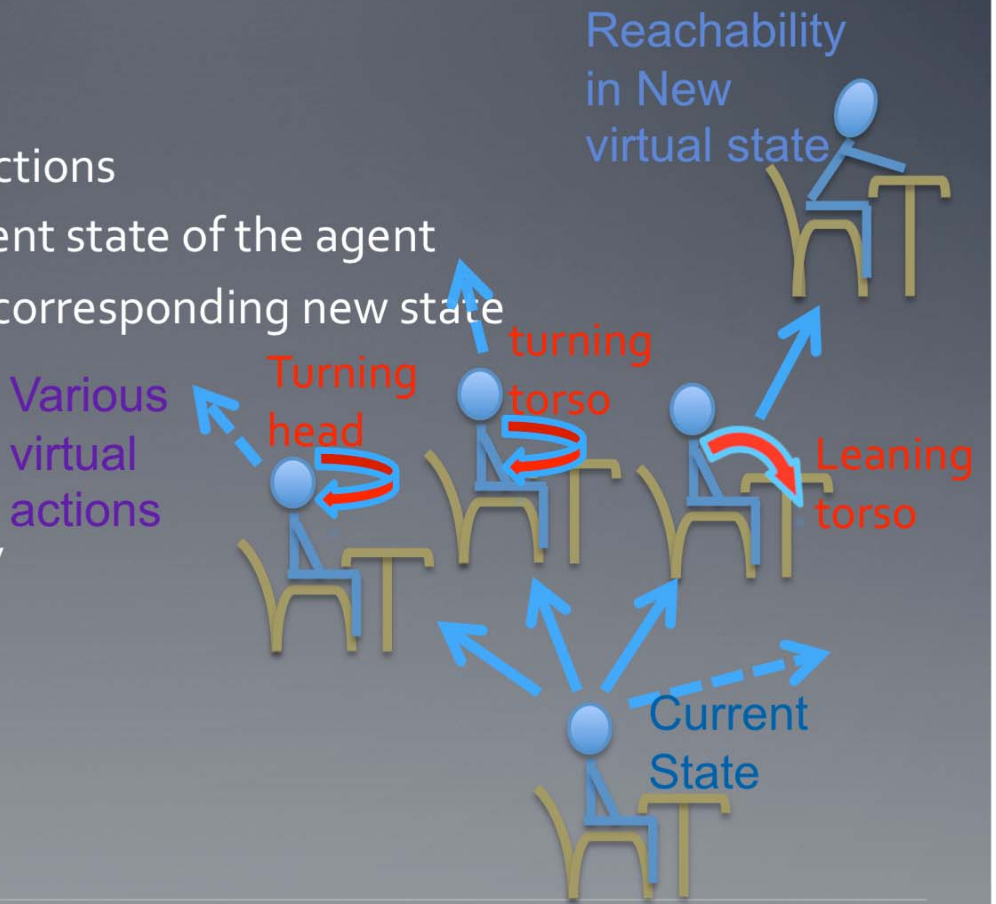
Equal Importance

Places/objects Agent Might Reach

e.g. ***put away*** some object from an agent.

So, we propose...

- to equip our robots to predict various abilities of human as well as of itself from multiple states.
- The idea is:
  - Robot will have a possible set of actions
  - Robot will apply them on the current state of the agent
  - The agent will virtually attain the corresponding new state
- Then predict
  - visibility and reachability
  - non-visibility and non-reachability
- At
  - 3D grid level
  - object level

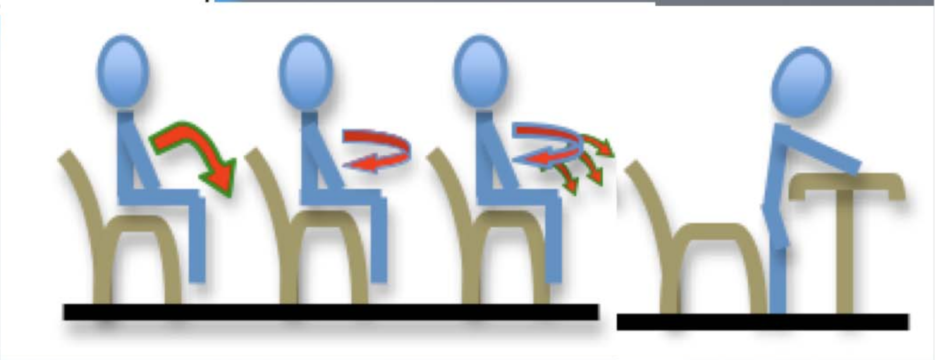


Mightability Analysis: Multi-State perspective taking

- ***Mightability*** : *Might be Able to* …
  - An estimate by applying various virtual actions on the agents.
- Answer to questions about the perceived ability:
  - "if the robot/human will lean forward, it/he/she might be able to reach 'these' places"
  - "if the robot/human will turn around it/he/she might be able to see 'these' objects"
- ***Mightability Map (MM)***: Calculated at 3D grid level
- ***Object Oriented Mightability (OOM)***: Calculated at object level

States for Multi-state Visuo-Spatial Perspective Taking

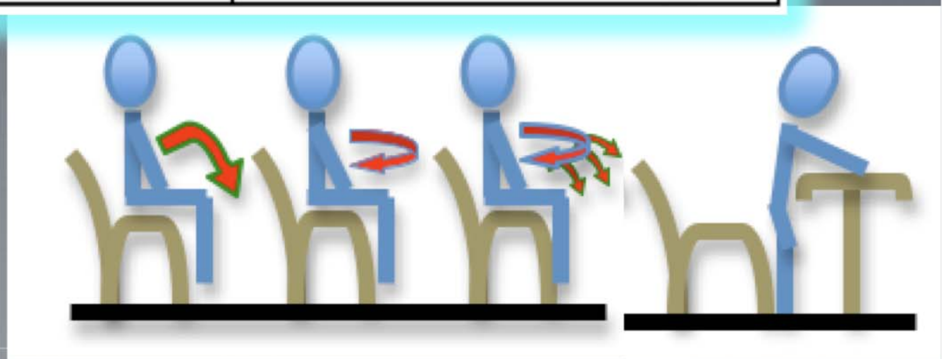
Reachability States	Visibility States
Current	Current
Sitting Straight	Sitting Straight Head
Sitting Turn Around	Sitting Turn Head
Sitting Lean Forward	Sitting Lean Torso and Turn Head
Sitting Turn and Lean	Sitting Turn Torso and Turn Head
Standing Straight	Sitting Turn-Lean Torso and Turn Head
Standing Turn Around	Standing Straight Head
Standing Lean Forward	Standing Turn Head
Standing Turn and Lean	Standing Lean Torso and Turn Head
	Standing Turn Torso and Turn Head
	Standing Turn-Lean Torso and Turn Head



Effort Levels for Multi-state Visuo-Spatial Perspective Taking

Reachability States	Visibility States
Current	Current
Sitting Straight	Sitting Straight Head
Sitting Turn Ar	
Sitting Lean For	
Sitting Turn and	
Standing Stra	
Standing Turn A	
Standing Lean Fo	
Standing Turn an	

Effort to Reach	Effort Level	Effort to See
No_Effort_Required	*Minimum*	*No_Effort_Required*
Arm_Effort		*Head_Effort*
Arm_Torso_Effort		*Head_Torso_Effort*
Whole_Body_Effort		*Whole_Body_Effort*
Displacement_Effort		*Displacement_Effort*
No_Possible_Known_Effort	*Maximum*	*No_Possible_Known_Effort*



Target of Task Understanding:

- Towards separating "meaning" of a task from the "means" to achieve it, in the context of Human-Robot Interaction.
- So that, the understanding:
  - is generalizable to a variety of situations,
  - does not require the learning data for each and every situation,
  - could be shared/transferred among heterogeneous robots.

Existing Works:

- Learning approaches in the context of Human-Robot Interaction:

trajectory based

- Learn trajectory
- with constraint on orientation for "pick-and-place" tasks [Gribovskaya , 2011]
- Adapt to avoid collision for "pour" task [Mhlig, 2009]

- **Hybrid approach**: represent the task in a symbolic sub-task sequence also incorporates trajectory information to perform the task [Ogawara, 2003]

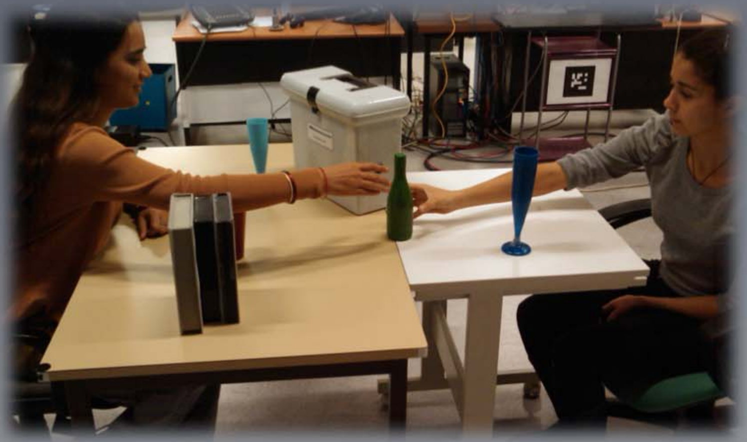
symbolic primitive based

- in terms of **sequence of sub-tasks**
- Place next to: Reach-> grasp-> transfer relative-> release [Chella, 2009]
- Assemble Table: Reach-> Pick-> Place-> Withdraw [Kuniyoshi, 1994]

- in terms of effects
- *'holding object'*, *'hand empty'*, *'object at location'*, etc., for the pick-and-put task [Ekvall, 2008]

But for Make accessible task…

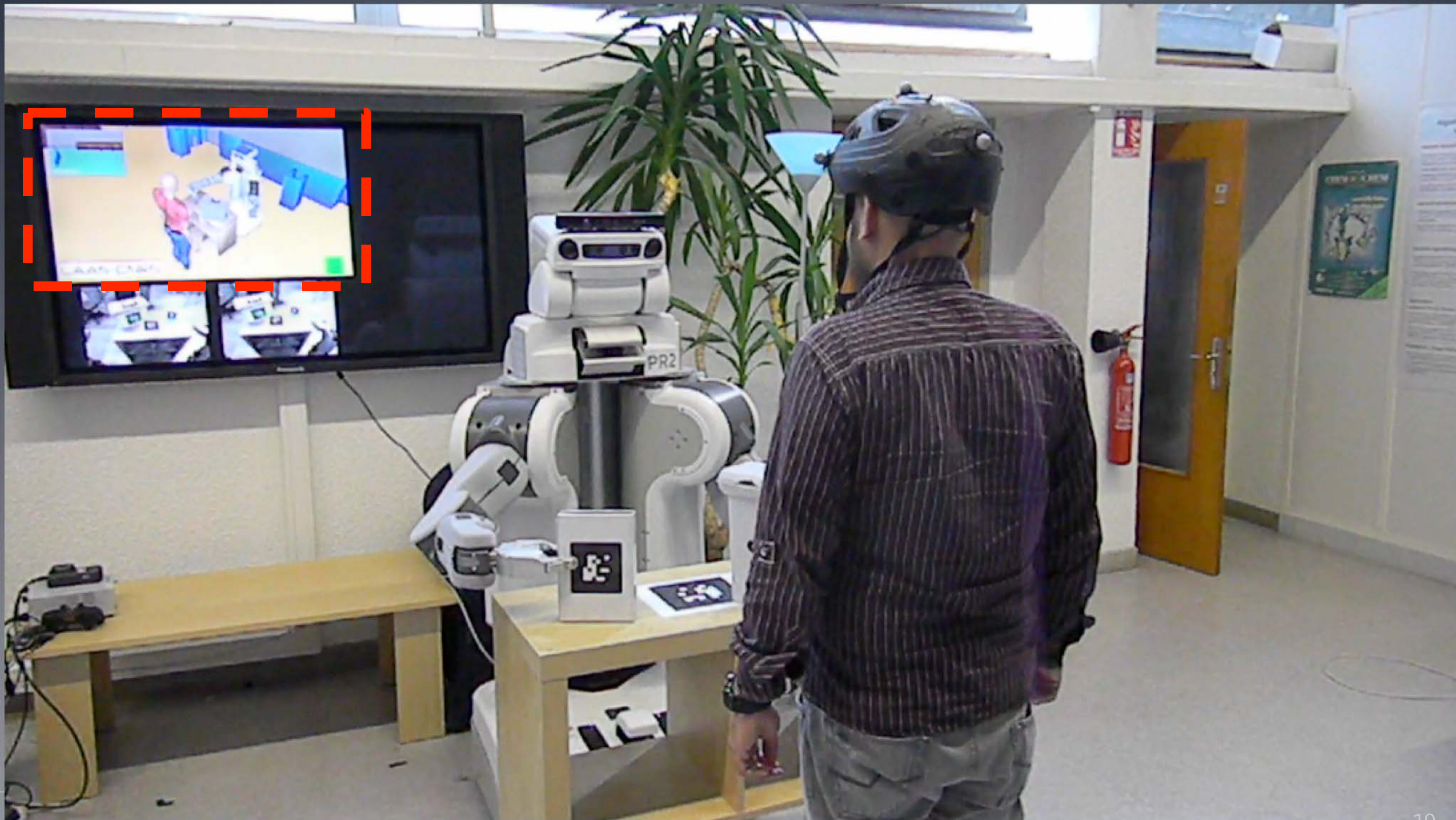
“Reach” -> ”Grasp” -> ”Transfer object” -> “Put Object relative” will not generalize the task in all these cases.



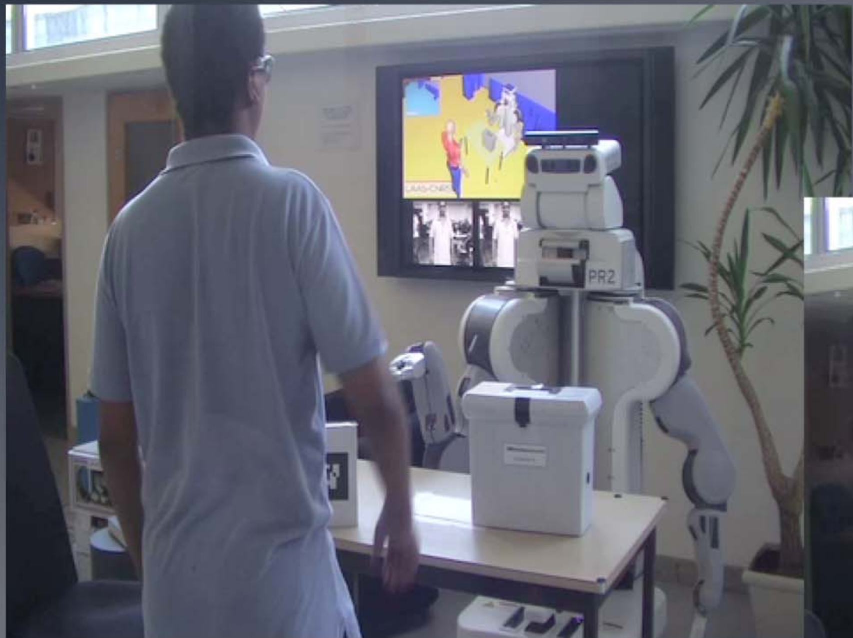
Main Idea:

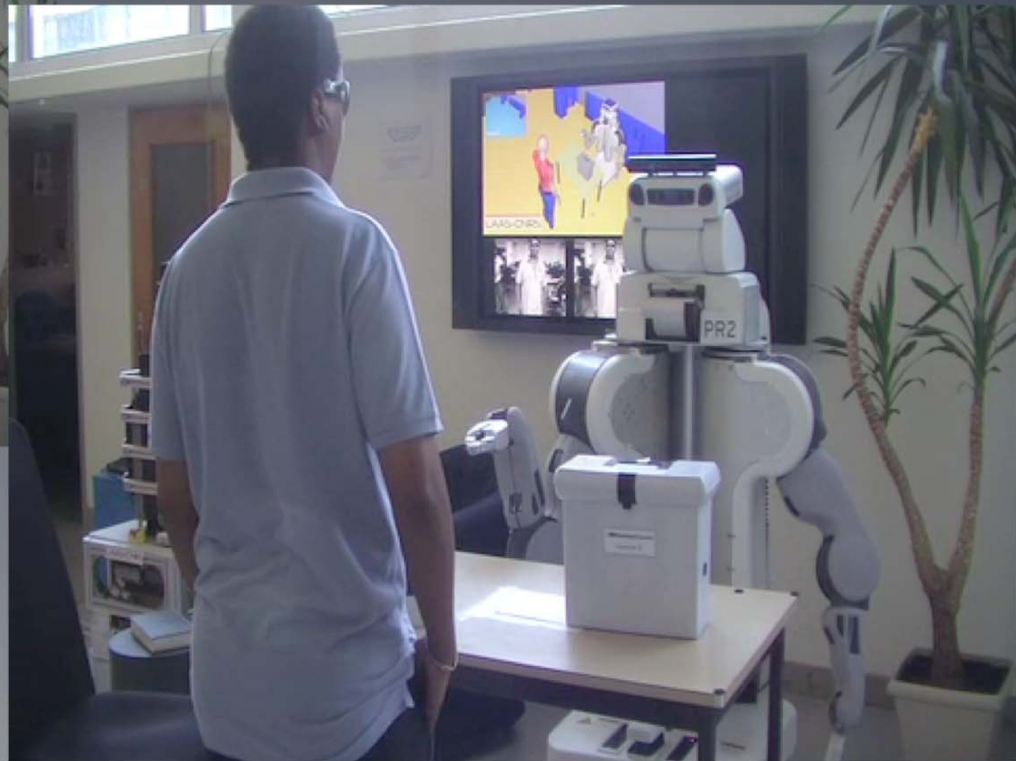
- If robot learns "Making an Object Accessible" means : "Reach" -> "Grasp" -> "Transfer object" -> "Put Object relative", then
  - It is actually learning "How" to perform the task
  - NOT "What" does the task mean, which is:
    - "***the object should be easier to been seen and reached by the person***".

Real Time 3D World Representation, understanding and Agent State Analysis



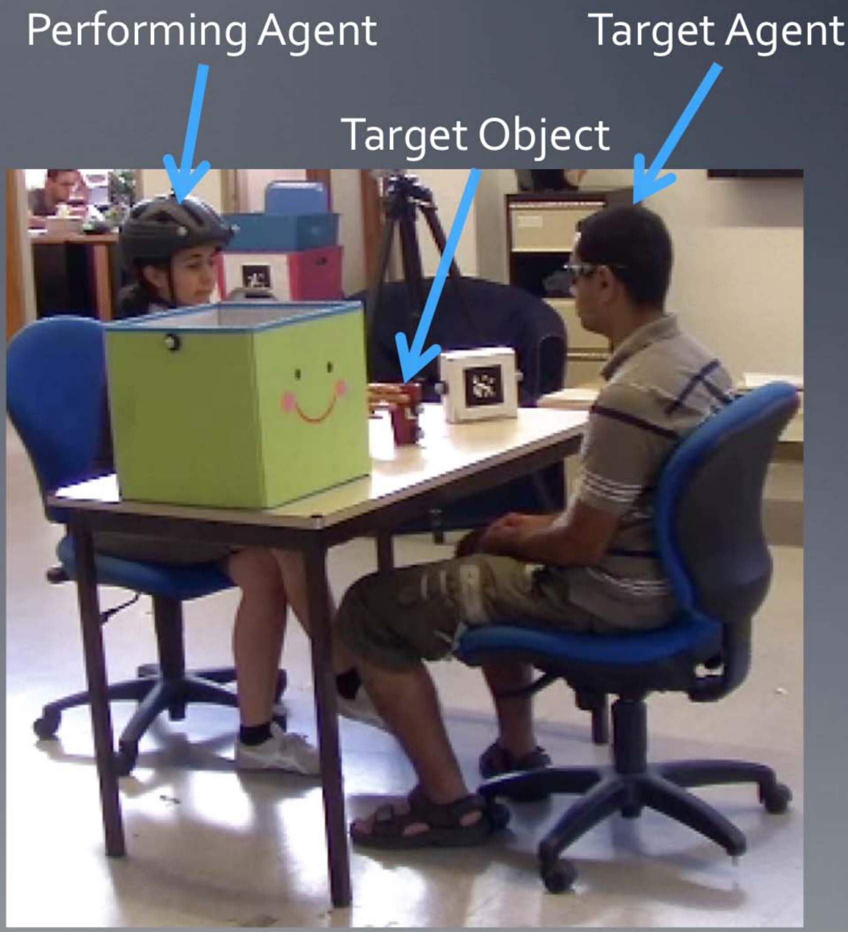
Real Time 3D World Representation, understanding and Agent State Analysis





Few Terms:

- *performing-agent*: the agent who will perform the task
- *target-agent*: for whom the task is being performed,
- *target-object*: on which the task is being performed



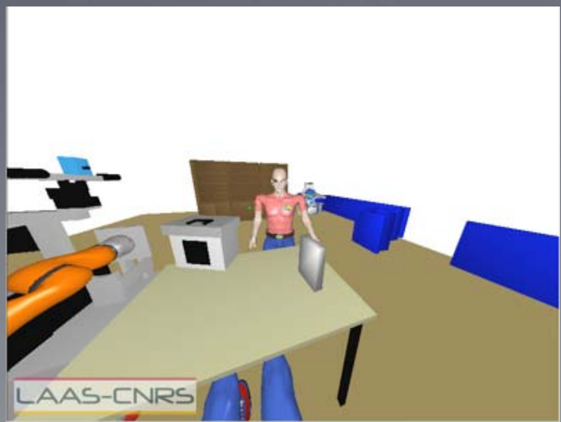
Approach:

- The robot will try to understand the task in terms of the changes on the target-agent's abilities to
  - see the target-object
  - reach the target-object
  - grasp the target-object
  - and the visibility score of the target-object.

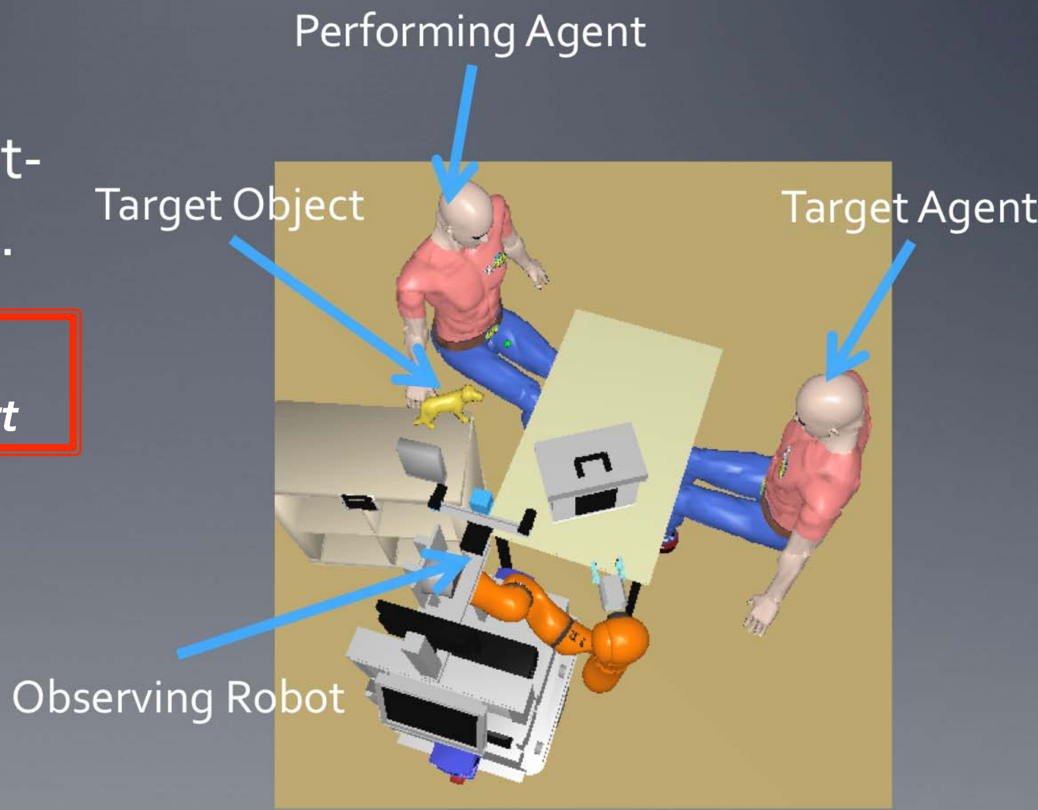
Approach:

1. Find the least effort of target-agent for target-object before the task.

Visibility Effort: ***Whole_Body_Effort***
Rechability Effort: ***Displacement_Effort***



Before Task Scenario from Target-Agents Perspective

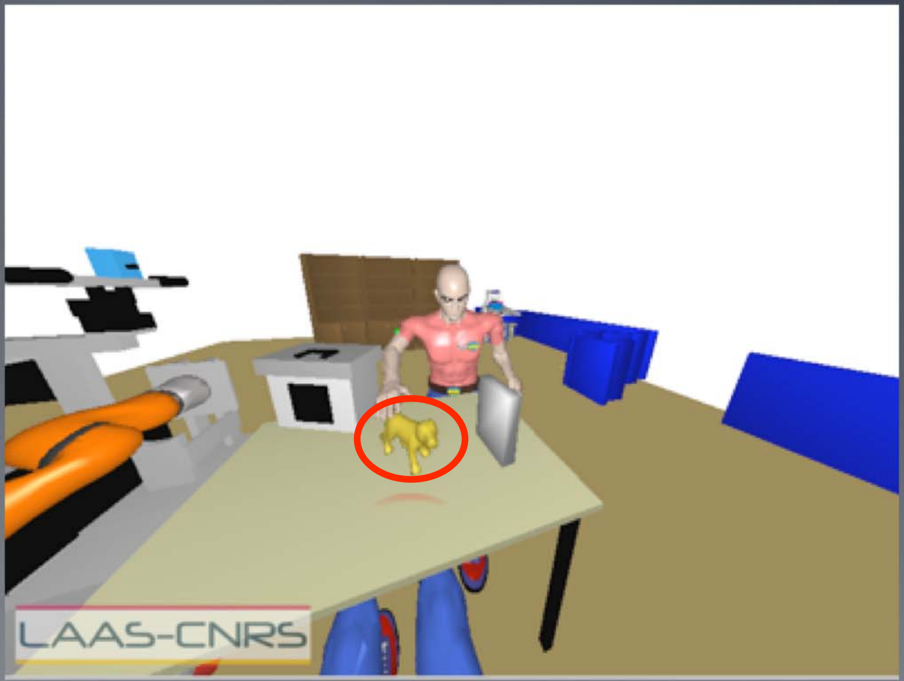


Before Make Accessible Task Scenario

Approach:

2. Find the least effort of target-agent for target-object after the task.

Visibility Effort: ***No_Effort_Required***
Rechability Effort: ***Arm_Torso_Effort***



After Make Accessible Task Scenario from Target-Agents Perspective

Approach:

3. Compare least efforts before and after tasks for the target-agent :

Change in Visibility Effort:
Whole_Body_Effort -> ***No_Effort_Required***

Change in Rechability Effort:
Displacement_Effort -> Arm_Torso_Effort

Approach:

3. Compare least efforts before and after tasks and categorizes the difference as:

Reachability and Visibility	Ability to Grasp	Visibility Score
Easiest_Effort_Maintained (S)	Graspability_Maintained (S)	Almost_Same (S)
Effort_Becomes_Easier (S)	Becomes_Graspable (S)	Increased (S)
Effort_Becomes_Difficult (NS)	Graspability_Lost (NS)	Increased_Significantly (S)
	Still_Not_Graspable (NS)	Decreased (NS)
(S: Supportive, NS: Non Supportive for an agent)		Decreased_Significantly (NS)

For the task "Make Accessible", for the "target-Agent" the "target-object" :

to reach: *Effort_Becomes_Easier,*
to see: *Effort_Becomes_Easier,*
grasp: *Becomes_Graspable,*
visibility score: *Increased*

Continuously Refining the Understanding:

- To avoid
  - over-constrained understanding
  - false association of an ability for a task, e.g. reachability from the target-agent's perspective is not relevant for "hiding an object" task
  - the understanding to become 'rigid' after few observation

- With every new observation of a task:
  - robot compares its past under standing for '*consistency*' or '*contradiction*'.

Observation Occurrence Belief:

Observation Occurrence Belief (OOB) for a particular '*task_type*' for a particular '*ability_type*' as :

the number of times, for the *target-object*, the particular observation (such as *Effort_Becomes_Easier*), has been observed about a particular ability (such as reachability), for a particular task (such as *make accessible*).

$$OOB^{task\_type,\, ability\_type}_{observation\_type} = \frac{N^{task\_type, ability\_type}_{observation\_occurred}}{N^{task\_type}_{demonstrations}}$$

number of times the task has been demonstrated/observed.

Signal of contradiction:

Supportive Observation Occurrence Belief (SOOB) and *Non-Supportive Observation Occurrence Belief (NSOOB)* for a particular ability:

$$SOOB^{task\_type, ability\_type} = \sum_{i=1}^{n_s} OOB_i^{task\_type, ability\_type} \dots$$

$$NSOOB^{task\_type, ability\_type} = \sum_{i=1}^{n_{ns}} OOB_i^{task\_type, ability\_type}$$

If for a particular ability type:

SOOB > 0
and
NSOOB > 0

Reachability and Visibility	Ability to Grasp	Visibility Score
Easiest_Effort_Maintained (S)	*Graspability_Maintained (S)*	*Almost_Same (S)*
Effort_Becomes_Easier (S)	*Becomes_Graspable (S)*	*Increased (S)*
Effort_Becomes_Difficult (NS)	*Graspability_Lost (NS)*	*Increased_Significantly (S)*
	Still_Not_Graspable (NS)	*Decreased (NS)*
*(**S**: Supportive, **NS**: Non Supportive for an agent)*		*Decreased_Significantly (NS)*

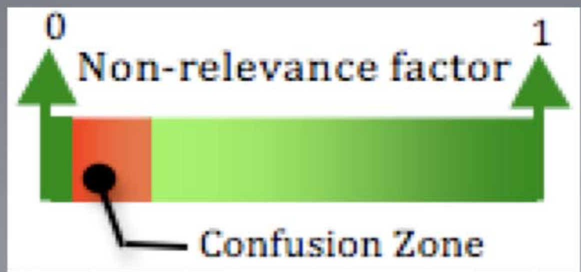
- then there is a contradiction about that particular ability for that particular task.
- the observations for that ability might be just a side effect.

Non-relevance factor:

non-relevance factor' for a particular ability 'a' and for a particular task type, '*t*' :

$$non\_relevance_a^t = 1 - \frac{abs\left(SOOB_a^t - NSOOB_a^t\right)}{\left(SOOB_a^t + NSOOB_a^t\right)}$$

non_relevance factor= 0; no contradiction, ability relevant
non_relevance factor = 1; equal number of contradiction and consistency, ability not-relevant



Robot can communicate confusion to human for resolving.

Example: Show an object task

Observation Occurrence Belief (OOB)						
After Demonstration ->	1		2		3	
Ability of target agent ->	See	Reach	See	Reach	See	Reach
Easiest_Effort_Maintained	0	0	0	0.5	0	0.33
Effort_Becomes_Easier	1	1	1	0.5	1	0.33
Effort_Becomes_Difficult	0	0	0	0	0	0.33
SOOB	1	1	1	1	1	0.66
NSOOB	0	0	0	0	0	0.33
Non Relevance Factor	**0**	**0**	**0**	**0**	**0**	**0.67**

Reachability of target agent becomes irrelevant for "show" task

Result: Various tasks Understood:

Task	*Visibility*	*Reach*	*Vis. Score*	*Grasp*	*N*	*T (s)*
Show	Supp	Not Relv	Supp	Not Relv	4	0.48
Hide	Non-Supp	Not Relv	Non-Supp	Not Relv	3	0.67
Make Accessible	Supp	Supp	Supp	Supp	3	0.4
Give	Supp	Supp	Supp	Supp	2	0.42
Put Away	Supp	Non-Supp	Supp	Non-Supp	3	0.51
Hide Away	Non-Supp	Non-Supp	Non-Supp	Non-Supp	2	0.83

Supp: Supportive, **Non-Supp**: Non-Supportive, **Not-Relv**: Not-Relevant

N: Number of times the task has been demonstrated
T: Average processing time in s, per demonstration

Discussion on Potential Applications:

• Robot understands the task in terms of the *desirable changes* in the *target-agent*'s visuo-spatial abilities on the *target-object*, and is not bound to learn relative distances or trajectory, such "Understandings" could be

• Generalized to novel scenarios: On different object, different spatial arrangements of agents, on different support plane, etc…

• **Transfer among heterogeneous agents:**
From JIDO to PR2 to HRp2

• Generalization for multiple target-agents : e.g. Hide from two humans, etc.

• Greater flexibility to the symbolic planner : If Robot knows "Show" means Object should be visible, it could even plan to displace the occluding object instead of directly manipulating the object to show.

• Used to predict action and show proactive behavior

Understanding Hide Task:

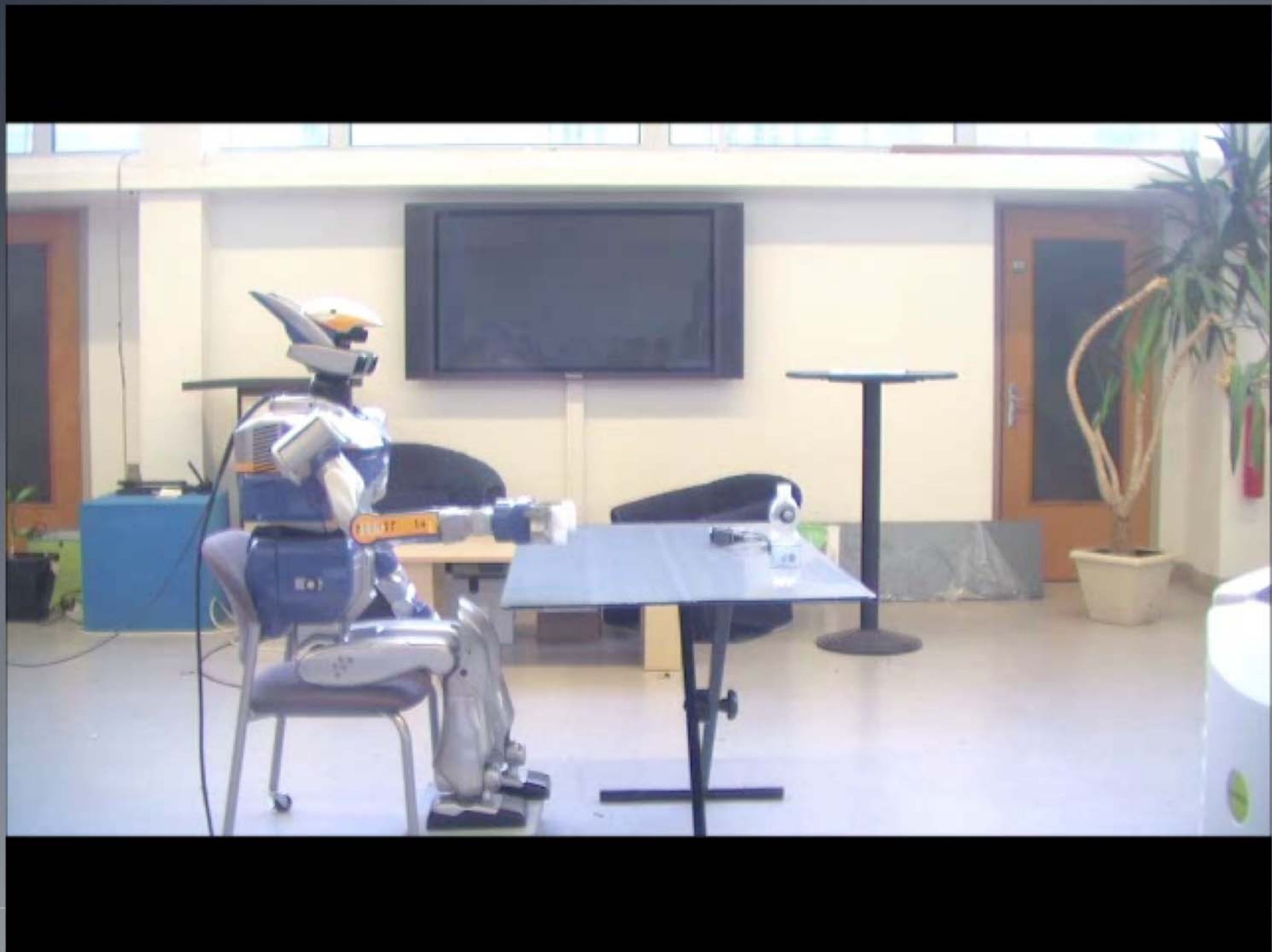
Summary:

- Understanding tasks in terms of effect from change in agent's ability point of view.
- Towards separating task understanding from its execution.
- Shown multi-state perspective taking as an aspect of such understanding.
- Equipped the robot to understand a set of Human-Human interaction tasks and execute them for Human-Robot Interaction.

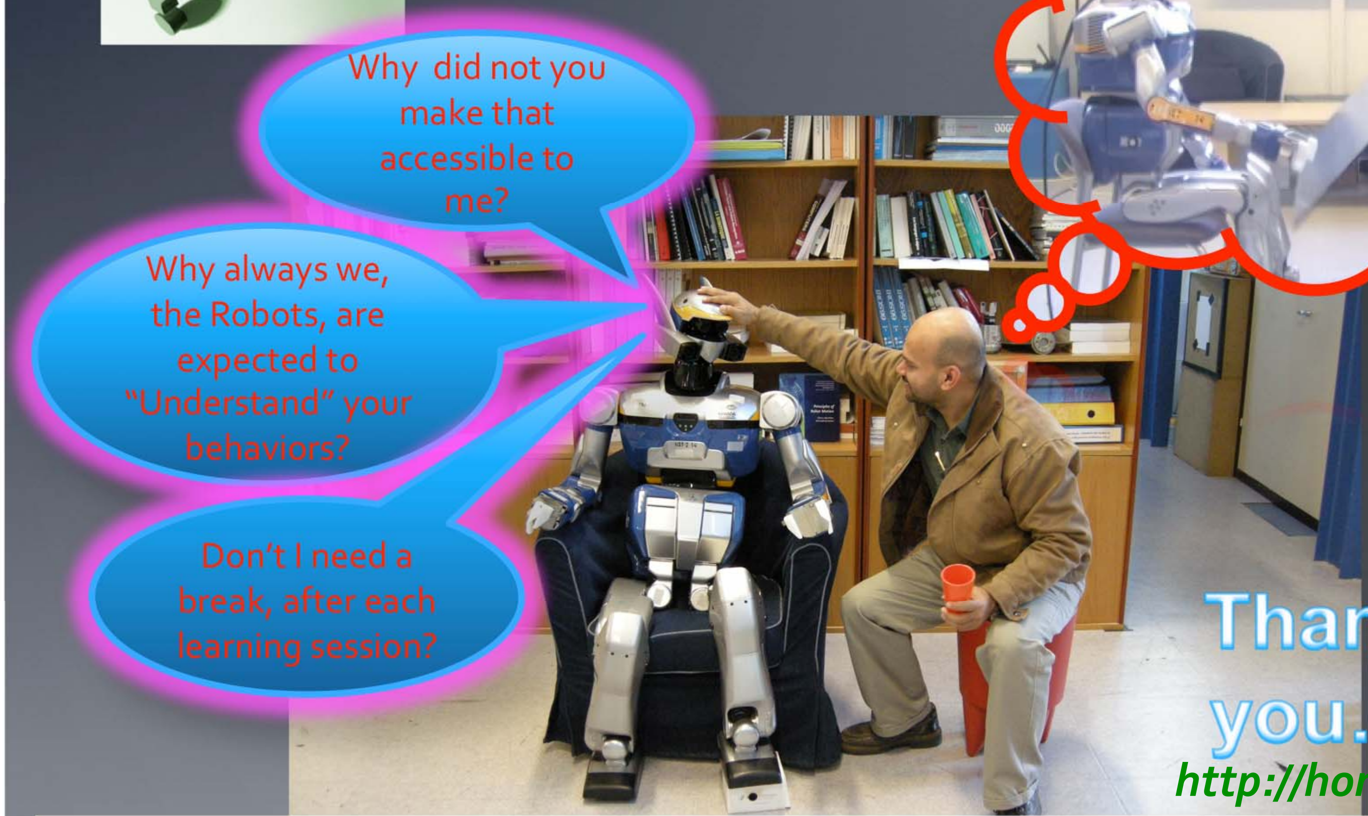
Future work:

- Perform such analysis from performing agent's perspective to **learn the performer's preferences** .
- ***Incorporate different states*** (***object in hand***, ***object on support***, etc.) and ***actions*** understanding (***grasping***, ***lifting***, ***moving hand***, etc.) during the task performance to ***learn execution preferences***.
- To incorporate additional primitives to understand more complex task: "take an object", "Dump into trash bin an object, etc."
- To autonomously finding inter-task relations, such as '*give*' could be '*show*' with some additional constraints

Expected Result :



Expected Result :



Thank you…

http://homepages.laas.fr/akpandey/

CHRIS
Cooperative Human Robot Interaction Systems

COOPERATION