



HAL
open science

A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management

Grégoire Milliez, Matthieu Warnier, Aurélie Clodic, Rachid Alami

► To cite this version:

Grégoire Milliez, Matthieu Warnier, Aurélie Clodic, Rachid Alami. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. IEEE RO-MAN 2014, Aug 2014, Edinburgh, United Kingdom. pp.FrAT2.4, <10.1109/ROMAN.2014.6926399>. <hal-01064546>

HAL Id: hal-01064546

<https://hal.science/hal-01064546v1>

Submitted on 16 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management

Gregoire Milliez^{1,2}, Matthieu Warnier^{1,2}, Aurelie Clodic^{1,2}, Rachid Alami^{1,2}

Abstract—In daily human interactions, spatial reasoning occupies an important place. In this paper we present a situation assessment reasoner that generates relevant symbolic information from the geometry of the environment with respect to relations between objects and human capabilities. The role of SPARK (*SPAtial Reasoning and Knowledge*) component is to permanently maintain a state of the world in order to provide a basis for the robot to plan, to act, to react and to interact. More precisely, we describe here the way the system manages the hypotheses to be able to handle such knowledge in a flexible manner. Equipped with such capabilities, a robot that will interact with humans should be able to extract, compute or infer these relations and capabilities in order to communicate and interact efficiently in a natural way. To illustrate our work, we will explain how the robot is able to manage and update agents beliefs and pass Sally-Anne test. This work is part of a broader effort to develop a complete decisional framework for human-robot interactive task achievement.

I. INTRODUCTION

Since the beginning of robotics, one key quality needed by a robot has been to reason about its environment. Equipped with perception abilities, the robot must be able to extract information that would help it to achieve its task. It is the well-known perception-action loop.

With the recent advances in robotics, robots have begun to appear in our daily lives. The world of a robot is now populated with humans with whom it needs to interact. An important challenge for researchers is to adapt the robot's reasoning capabilities to this new world, which is by default shaped for humans. Human-robot interaction requires to equip the robot with explicit reasoning on the human and on its own capacities to achieve its tasks in a collaborative way with a human partner.

In our concrete context, a robot and one or several human(s) share a physical environment, typically a workshop or a domestic environment. The environment is composed of static walls and furnitures. What is dynamic is the fact that humans and robot move and manipulate objects.

The role of SPARK (*SPAtial Reasoning and Knowledge*) component described here in the robot control architecture is to permanently maintain a state of the world in order to provide a basis for the robot to plan, to act, to react and to interact. In addition, SPARK has been designed in order to take into account the following considerations.

First, the robot must be able to exhibit knowledge in a human-understandable manner. The robot must be able to handle spatial reasoning and anchoring where "anchoring is the process to establish and maintain in time the correspondence between symbolic knowledge and sensory data"[1].

Secondly, it must be able to gain explicit reasoning on the human it interacts with. It means that not only the knowledge must be grounded between the robot and its interactor but also that the robot must be able to maintain an explicit representation of the knowledge of its interactor apart of its own knowledge. That will allow the robot to compare its own beliefs with the one of the human and to infer similarities as well as differences and ambiguities. Thus, the robot must be able to handle a kind of "theory of mind"¹.

Equipped with such capabilities, a robot who will interact with humans should be able to extract, compute or infer these relations and capabilities in order to communicate and interact efficiently in a natural way.

To achieve this we identified 3 main ingredients. First, the consideration of perspective-taking, i.e. the ability of the robot not only to build a model of the world for itself but also to estimate what its human partners perceive. Secondly, the ability to compute efficiently affordances for itself and to estimate the affordances of its human partners in a given situation Finally, the ability to maintain a history of beliefs based on presence and focus of attention of humans which will enable reasoning on divergent beliefs. We will present how these features are implemented in SPARK as a permanent activity based on inter-related processes.

This work is part of a broader effort to develop a decisional framework for human-robot interactive task achievement, embedded in a cognitive architecture, involving a knowledge base, a task planner, a motion planner and a supervisor. In addition, the approach introduced in this work not only facilitates the interaction between robots and humans, but also bridges the gap between high level tasks and low level movement actions [2].

Section II reviews related works and give the context of our contribution. Section III explains how our system manage perception raw data to exhibit meaningful information. Section IV exposes how our system is able to update its knowledge and to use it in interaction context.

¹ CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France ² Univ de Toulouse, LAAS, F-31400 Toulouse, France
Firstname.Name at laas.fr

¹http://en.wikipedia.org/wiki/Theory_of_mind

II. RELATED WORK

Building and updating an intermediate 3D geometric model is common in robotic architecture, however our system is used as a hub for both sensor fusion and geometric reasoning and uses a rich 3D model extracted from the Move3D environment [3]. SPARK can be compared to the Grounded Situation Model (GSM) introduced by Mavridis and Roy [4] in the sense that they both provide an amodal physical representation of the world used as a mediator between the sensor space and symbolic models. [5] presents a survey of such symbol grounding systems in robotics. [6], [7] present examples of the use of such system.

Applications of spatial reasoning [8] are multiple. It has been used for instance for natural language processing for applications such as direction recognition [9], [10] or language grounding [11]. [12] presented a spatial reasoner integrated in a robot which computes symbolic positions of objects

We use perspective taking and some elements of theory of mind techniques to efficiently compute perspective-aware models of the world.

Perspective Taking is a human ability which allows one to see things from other's point of view. Studied in psychology literature [13], [14], this ability is crucial when interacting with people by allowing one to reason on others' understanding of the world in terms of visual perception, spatial descriptions, affordances and beliefs, etc. Therefore, in the last years it has been gradually employed in Human-Robot Interaction. [15] presents a learning algorithm that takes into account information about a teacher's visual perspective in order to learn a task. [16] apply visual perspective taking for action recognition between two robots. [17] use both visual and spatial perspective taking for finding out the referent indicated by a human partner.

In psychology, Theory of mind (ToM) is defined as an understanding of other peoples mental states (their thoughts, feelings, desires, motivations, intentions). It includes perspective taking ability. Visual perspective taking is one of the most significant ToM precursor. ToM encompasses a wide range of skills from instantaneous visual perspective skill to complex interpretation of other agent intents, plans, feelings occurring on a long time period. Increased ToM skills directly lead to increased performance when interacting with other agent in a collaborative as well as a competitive context. Being able to attribute false belief (to recognize that someone else has different beliefs about the physical world) has been considered as a milestone in ToM development. In psychology literature the false belief task was formulated in [18]. Breazal in [19] proposed one of the first human robot implementation and proposed some more advanced goal recognition skills relying on this false belief detection.

This paper will present an evolution of the previous work [20] [21] [22] with a refined management of divergent beliefs, temporal reasoning on data and improved inferring capabilities. These improvements allow the robot to pass Sally and Anne test [23] (see Section IV B), to make inferences not only on position properties but also on dialogue,

and other human actions.

III. SPARK

SPARK is a component responsible for 3D environment management and spatial reasoning. The goal of SPARK is to get the geometric position of each entity and to generate symbolic facts from this geometric data. With the symbolic facts and geometric data we can then create collaborative plans, generate efficient dialog and interact with the world.

A. Data Gathering

To understand its environment a robot needs first to perceive it. Figure 1 presents an example of a scene as it is represented in SPARK.



Fig. 1. Rendering of an environment as it is build and maintain by the system using SPARK data for furnitures, objects, humans and robot.

To obtain this rendering, it needs to collect data from its environment by using sensors. In SPARK, we need to collect data concerning 3 entities: human, objects and robot itself. This data will be used together to reason on the environment and build symbolic facts.

Concerning humans, we need to get not only the position of each human but also the position of their hands, head and shoulders. We also need an identifier to recognize each human. To put humans at the right place in the model, we make a projection in SPARK using the capture device position. To recognize a human that has already been seen, we compare its input identifier with the ones that are already in the current database.

The robot posture and position are collected directly from the robot internal sensors.

At last, we need to get the objects position and identifier. This is usually done using object recognition algorithm based on vision. Once we get the relative position of the objects, we send it to SPARK. To put objects at the right place in the model, we make a projection in SPARK using the robot's vision position device.

B. Object Position Hypotheses

When humans lead activities at home, they usually involve object manipulation (cooking, cleaning, etc). As robots are made to help humans, they should be able to manipulate objects as well as track them to collaborate with them. Our

goal is not to explain perception acquisition but how reasoning on perception data can improve object detection and tracking.

Tracking an object is not a simple task. Indeed, it can be quite small so they frequently get occluded or may go out of vision field. Humans are able to know an object's location even without direct perception. Humans constantly make position hypotheses on objects they can't see and perform reasoning on these hypotheses. To work with humans robot should be able to perform this kind of reasoning as well.

To implement the ability of reasoning on the robot we added several position hypotheses that the robot makes when an object is not visible.

As a basic hypothesis, when the robot can't see an object (object is out of the vision field or occluded), it assumes that the object is at the same place and orientation where it was perceived the last time. That is why in fig. 1, even if the robot can't see the blue book anymore (occluded by pink box) SPARK maintain its position. This hypothesis is used as default hypothesis, i.e. if we have another hypothesis of an object's position we will choose this last hypothesis over the default one.

The other position hypotheses are generated according to robot and human actions. If the robot grabs an object, the object will most probably be hidden by its own hand, but we know where it is (in the robot's hand). We have the same kind of hypotheses for human, using a monitoring sphere to know when human picks up an object. This is basically just a sphere positioned around the objects. If a human puts his hand inside the sphere, we assume that he picked up the object, unless we still see it on table. Finally, we also implemented a hypothesis on containing the objects. If an agent drops an object in another one, even if we will not be able to perceive the object anymore, we know roughly its position (center of the containing object).

To sum up, the robot uses first the perception, then if it's not available it uses position hypothesis. If none of these are available for the current object, it uses the last perceived position as shown in fig. 2. The position hypotheses have to be managed and updated in real time to keep a consistent model of the world. If an object has an hypothesis and is perceived then we compare the position perceived from the one given by the hypothesis. If they are too far from each other, the hypothesis is deleted as it is considered as impossible. If an object should be seen and is not seen, then we delete the default hypothesis that the object is at the last perceived position and set it as unknown.

These hypotheses concerning objects' position allow the robot to have a more efficient tracking of objects, specially when they are occluded or not perceived.

C. Perspective Taking and Belief Management

As explained in the previous section, humans also make hypotheses concerning objects. What comes with every hypothesis is the possibility of making mistakes. In the field of HRI, the robot has to understand the human. That is why we think, knowing what humans believe and when they have wrong information is key for interacting with them.

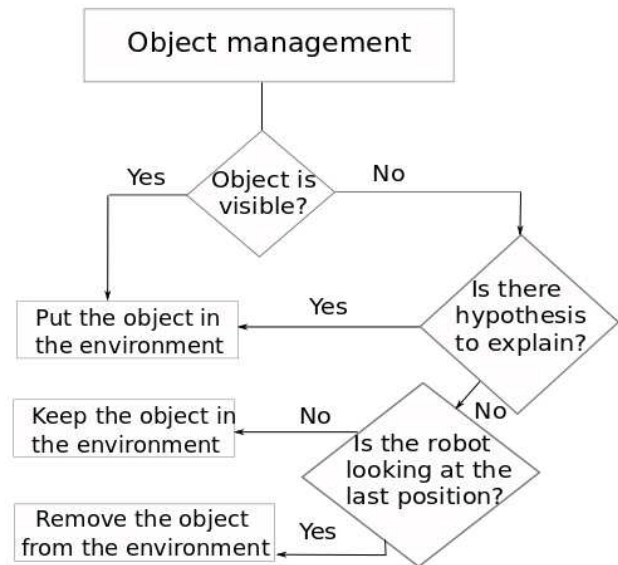


Fig. 2. Schema of SPARK reasoning to manage object position. The hypothesis on object position can come from occlusion due to hiding the object, object in container or object grasped by agent.

The first step in knowing what a human thinks is to be aware of the information he gets. There are two ways a human can get information about his environment: either he gets it directly (or can infer) from his perception or from dialog. As humans we constantly use what psychologist call "perspective taking", which means being able to put yourself into someone else's shoes. This helps to adapt our dialog according to the mental state of the other participant. To implement this behavior on a robot, we focused on localization, perception and human robot dialog aspects to infer what information the human can get.

In SPARK we have the position of humans (see III 1.). We use it to calculate affordances of each human toward elements of the scene he can interact with. To estimate what is visible for a human, it computes which objects are present in a cone, emerging from human's head. If the object can be directly linked to the human's head with no obstacle and if it is in the field of the view cone, then we assume that the human sees the object and hence has knowledge of its position. If an obstacle is occluding the object, then it won't be visible for the human. Concerning the reachability, a threshold of 1 meter is used to determine if the human can reach an object or not. Then, we have different models of the world for each agent. One for the world as perceived by the robot and one for each human perception (as the robot thinks human sees it). Each of these models is independent and logically consistent.

In some cases human and robot models can hold different values. It can come from different visions of the scene (some objects properties are accessible only for the robot so the human does not have knowledge on them). It can come from divergent believes between the human and the robot (human believes the object as property P when robot

believes this property is false for the object). In this second case, perspective taking is not enough to understand human's wrong believes. A divergence belief management is needed.

This Management relies on data from the environment as well as from affordances and supervisor. This way, the robot can generate beliefs according to the task stated in correlation with affordances as shown in fig. 3.

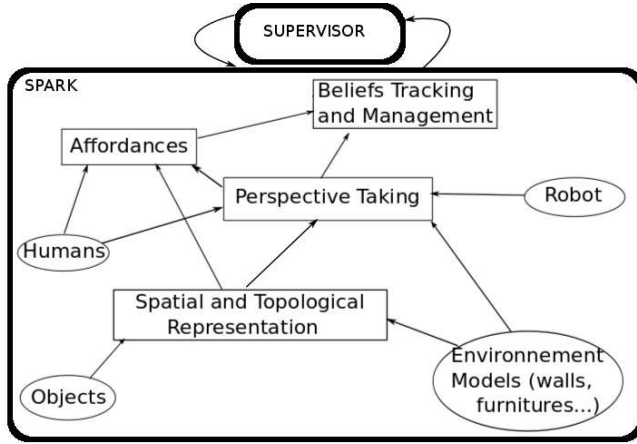


Fig. 3. Schema of SPARK reasoning to generate beliefs

As an example, we consider a human who leaves the interacting area with the belief that an object A has property P in state s1. When he will come back, even if object A has changed it's property to state s2, he will probably think the property is still valid as it was before until he is able to update his belief about the property by using dialog or perception. To manage this kind of situation, when a human leaves the area, we keep his beliefs outdated until he comes back.

Here are the beliefs of agents about the object before human comes back:

```

ROBOT                                GREG
obj_A property_P state2  obj_A property_P state1

```

Once he comes back, there are 3 different situations:

- First situation, human notices the object new property state and automatically corrects his belief. We then also update the belief vector with the last perception data.

```

ROBOT                                GREG
obj_A property_P state2  obj_A property_P state2

```

- 2nd situation, human notices that the object has changed but can't infer what is the new property state. We then update the belief vector to remove his belief, i.e., robot knows human knows he doesn't know the object new property.

```

ROBOT                                GREG
obj_A property_P state2  obj_A property_P unknown

```

- 3rd situation, human can't notice that the object property has changed. The human will then keep his wrong belief. We assume here that if the human is able to perceive an object property, he is then aware of the object property state. As an example for position property, if there is no obstacle between his eyes and the object and the object is in the human field of view, he is then aware of its position.

```

ROBOT                                GREG
obj_A property_P state2  obj_A property_P state1

```

Knowing these beliefs is a great help to the robot to understand human and interact with him. The robot takes human's

perception into account to have appropriate interpretation of human requests, to proactively inform or warn the human about a missing or wrong belief and also to generate a collaborative plan. Thus, the robot has to reason on what a human can see, reach and knows to get these skills.

IV. IMPLEMENTATION AND RESULTS

A. Implementation

SPARK can use different sensors as input. Once the data is given to SPARK the processing is identical. We present briefly the sensors and modules we use in the experiments.

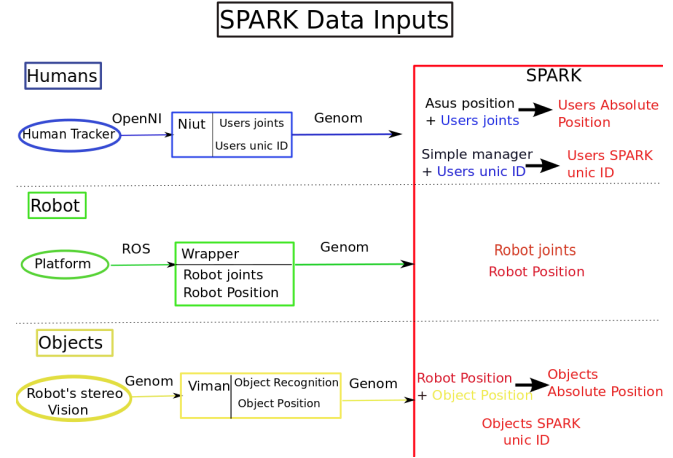


Fig. 4. Schema of SPARK input implementation

To track the human posture, we use an Asus xtion. As moving head is an important part of communication for the robot to show its attention, and to avoid tracking loss, we put the Asus xtion on a stable platform behind the robot. The module niut manages human tracking using the OpenNI API. It also detects the average hue of humans' top in order to identify them. To get the human average hue, we extract the position of the torso and then project it in RGB image coordinate. If available, we also use the coordinates of shoulders and hips to define the rectangle in which the hue will be computed, otherwise we create a rectangle around the torso position that we scale according to the human depth. By giving defined hues to the module we can then recognize the humans and send their posture data to SPARK with a unique id. This also avoid false detection issues and unregistered human to disturb experiments. To put humans at the right place in the model, we make a projection in SPARK using the Asus xtion position.

We get the robot's posture and position directly from its internal sensor using ROS.

To get the objects position we created the Viman module. This module uses the stereo vision of the robot to recognize and localize objects. To do so, viman scans RTags on objects. Once we get the relative position of the objects, we send it to SPARK. To put objects at the right place in the model, we make a projection in SPARK using the robot's head position. Fig 4 illustrates this implementation.

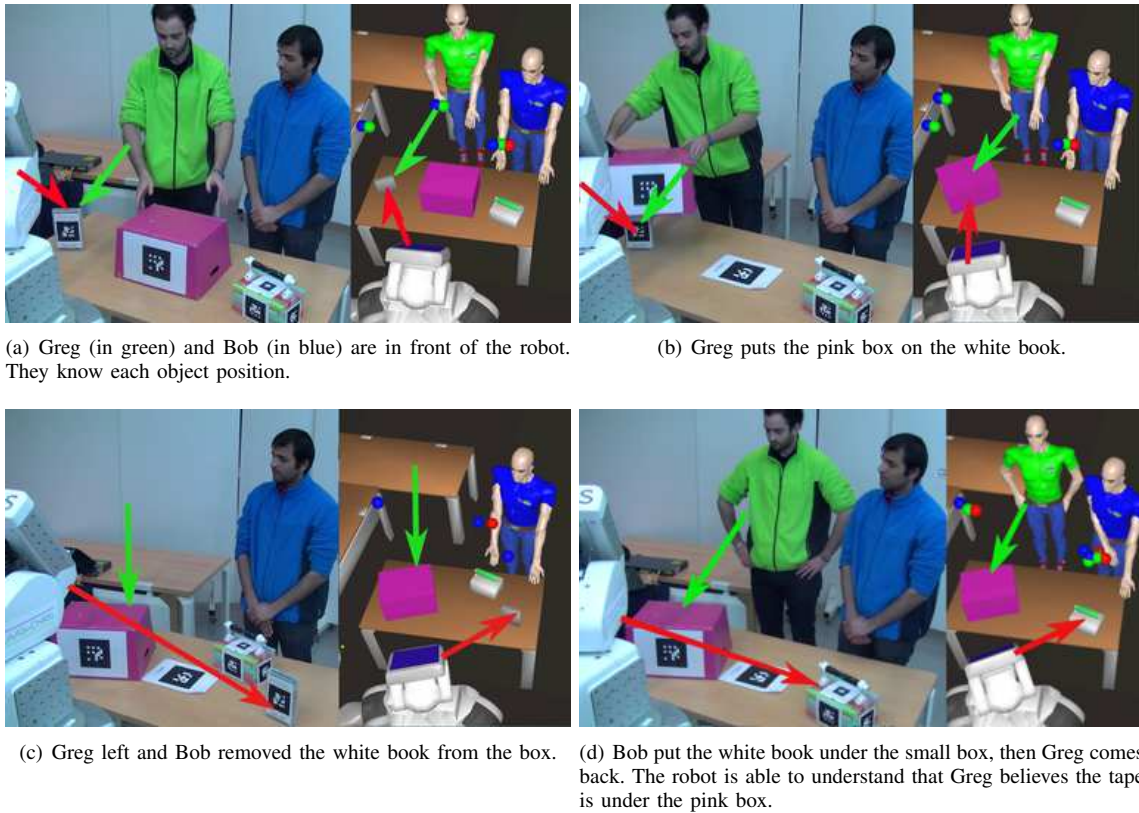


Fig. 5. Sally and Anne test in real (left part) and as perceived by the robot (right part). To help the understanding we explicitly show beliefs on the white book position (with green arrow for Greg and red one for the robot). The small colored spheres in the rendered scene are used as indications about the reachability of an object for an agent.

B. Results

We introduced above how we keep the track of distinct beliefs for each agent. We believe this feature is helpful for understanding human speech, action and focus of attention, i.e. to interact with humans. As the robot knows the human beliefs it can decide what information human needs and then whether to speak or not according to the current situation or plan realization. The belief management features presented above has been used on a robot in order to test them ².

1) *Sally and Anne test*: For better understanding of the scenarios, fig. 5 and fig. 6 are images from the real experiment with screen-shoots from the 3D model display of the spatial reasoning module.

In the 1st scenario, in order to test our system and illustrate the robot ability to detect *False/Distinct belief on object position* we decided to run Sally and Anne test. In our experiment, two users Greg in green and Bob in blue are in front of the robot. the scene is composed by a white book and 2 boxes standing on a table (fig. 5(a)). Greg puts the white book under the pink box (fig. 5(b)). Then Greg leaves. While Greg is away, Bob takes the white book and puts it under the small box (fig. 5(c) 5(d)). Then when Greg comes back, we ask the robot where Greg will look for the box.

We will show below some interesting symbolic facts for Greg and for the robot corresponding to 5(d).

ROBOT	GREG
PINK_BOX isVisibleBy GREG	PINK_BOX isVisibleBy GREG
SMALL_BOX isVisibleBy GREG	SMALL_BOX isVisibleBy GREG
W_BOOK notVisibleBy GREG	W_BOOK notVisibleBy GREG
W_BOOK isIn SMALL_BOX	W_BOOK isIn PINK_BOX
W_BOOK hasKnownLocation true	W_BOOK hasKnownLocation true

As the robot knows that the white book is not visible by Greg, it didn't update his belief concerning this object location. With the help of our system, the robot is able to observe that, as Greg can't see what has changed, he has a divergent belief about the object position. If the human and robot has to fulfill a task involving this object, the robot will know it has to inform the human about the object location.

2) *Divergent belief management*: To go further and illustrate the 3 different situations exposed in III C, we run another scenario. In this second experiment, two users Greg and Bob are in front of the robot. They have a white book, and a white box that will be use for vision occlusion (fig. 6(a)). Once Bob left (fig. 6(b)) Greg takes the white book and put it behind the white box so that the object is not visible in the human perspective (fig. 6(b)). Then Greg leaves and Bob comes back. The robot is able to understand that, because the white book is occluded, Bob doesn't know where it is. However, Bob can see that the white book

²Videos of the experiment can be found in <http://homepages.laas.fr/gmilliez/roman2014/>

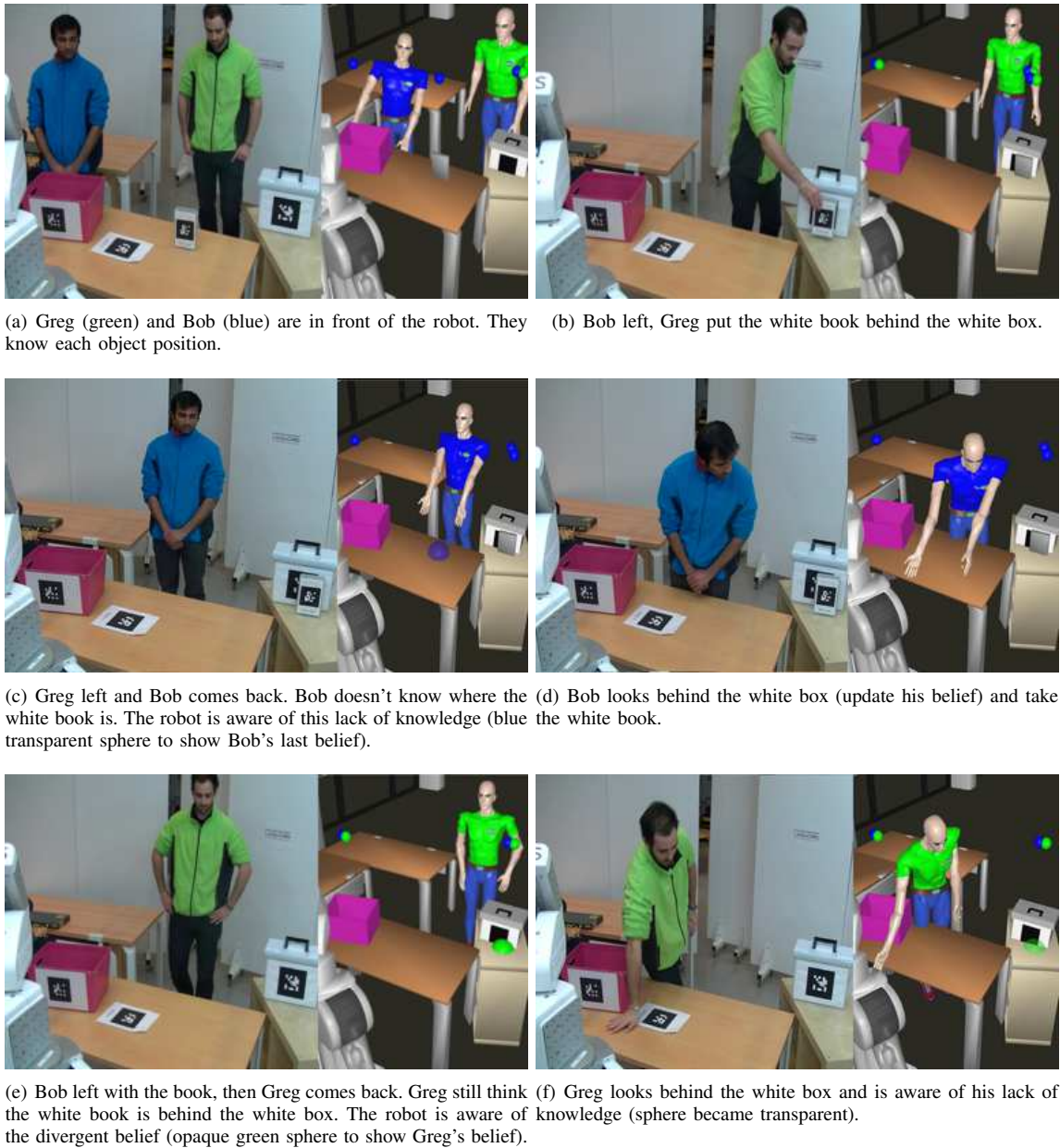


Fig. 6. Divergent belief scenario. in real (left part) and as perceived by the robot (right part).

is no longer where he thought it would be. This lack of knowledge is represented by a blue transparent sphere where Bob expected the object to be (fig. 6(c)).

Below are some symbolic facts for Bob and for the robot.

ROBOT	BOB
WHITE_BOX isVisibleBy ROBOT	WHITE_BOX isVisibleBy ROBOT
WHITE_BOX isVisibleBy BOB	WHITE_BOX isVisibleBy BOB
W_BOOK notVisibleBy BOB	W_BOOK notVisibleBy BOB
W_BOOK hasKnownLocation true	W_BOOK hasKnownLocation false

In the next step, Bob looks behind the white box and update his belief (fig. 6(d)). Then he leaves with the white book. When Greg comes back, as the object was occluded before and is still not visible, the robot is able to know that Greg has a wrong belief on the white book position. This

wrong belief is represented by a green opaque sphere where Greg thinks the white book is (fig. 6(e)). Then Greg looks behind the white box. He has no longer a divergent belief but he still ignores the white book position. As the robot saw this, the green sphere become transparent (fig. 6(f)).

Without our algorithm, at each steps the robot would assume that the human holds the same beliefs as itself (i.e. the human knows the new object position).

3) *Dialog disambiguation:* Now we will show how dialog could benefits from our system. At the end of the scenario of fig 6, Bob left with the white book. The robot was able to see this action by using the monitoring spheres. Now, let's assume that in addition to this setup, a black book stands on the table but is hidden by the pink box on Bob's side. So the

black book is not visible by Bob and is visible by the robot. Consequently, if Bob asks the robot "where is the book?", as the robot knows Bob took the white one, even if both books are currently not visible by Bob it understands that Bob speaks about the black book. The robot will answer: "It is on the table behind the pink box". Such dialog ability is only possible if the robot holds correct assumption concerning human's knowledge as done by our system. Without the temporal reasoning on human actions, robot would have to ask "which book are you talking about?".

Now, come back to the end of the scenario of fig 6 where Greg has a wrong belief about the white book's position (symbolized by an opaque green sphere). Seeing Greg trying to have a look behind the white box, the robot can infer that he's looking for the white book. Consequently, it can say proactively : "The object you are looking for was taken by Bob". Such proactive dialog ability is possible with the help of our system because it allows to infer human's intention from human's (wrong or lack of) belief and to talk proactively to the human to correct it.

This level of human understanding allows the robot to interact in a more natural way with humans.

V. CONCLUSION

In this article we have presented our spatio-temporal reasoning system SPARK. We exposed how reasoning on objects while interacting with humans helped us to develop the ability of making hypotheses to improve object tracking. To make the robot understand humans while interacting with them we have also developed perspective taking to understand the situation of the human. And at last, to manage human's potentially false/distinct or lacking belief on object position we developed a new feature based on distinct beliefs state model management for each agent. With this feature robot can be aware of humans' belief states.

Experiments were carried out, revealing our robot's capacity to pass Sally and Anne test and assess human beliefs. We show the use of this feature in dialog and task planning. It has yielded promising first results. This is a useful improvement of its ToM that leads to more natural interaction with the human and more efficient dialog. In part IV, we explain how we are able to apply this for knowledge on position. However the algorithm could easily be extended to manage non position-related beliefs. The robot could reason on agent beliefs about indirectly perceivable and inferable properties such as temperature, content, weight, ownership, etc. The robot should be able to sense, infer or be informed about these attributes and assess human perception or beliefs about them. It would also be interesting to have a real time line managing the events history with time stamp. As an example if an object has been painted by a user A and 5 minutes after a user B come close to it, without knowledge of the fresh painting, the robot should be able to warn him and tell him "be careful, this object was painted 5 minutes ago". This would be useful to proactively give information but also to learn habits of each human, his action after an event or the order in which events happen. Robot could also give

information of what happened in a room while the human was away.

Acknowledgments: This work has been supported by l'Agence Nationale pour la Recherche under projet reference ANR-12-CORD-0021 (MaRD).

REFERENCES

- [1] S. Coradeschi and A. Safotti, "An introduction to the anchoring problem." *Robotics and Autonomous Systems*, 2003.
- [2] R. Alami, M. Warnier, J. Guitton, S. Lemaignan, and E. A. Sisbot, "When the robot considers the human..." in *Proceedings of the 15th International Symposium on Robotics Research*, 2011.
- [3] T. Simon, J.-P. Laumond, and F. Lamiroux, "Move3d: a generic platform for path planning," in *4th International Symposium on Assembly and Task Planning*, 2001.
- [4] N. Mavridis and D. Roy, "Grounded situation models for robots: Bridging language, perception, and action," in *In AAAI-05 Workshop on Modular Construction of Human-Like Intelligence*, 2005.
- [5] S. Coradeschi, A. Loutfi, and B. Wrede, "A short review of symbol grounding in robotic and intelligent systems," *KI - Kunstliche Intelligenz*, vol. 27, no. 2, 2013.
- [6] M. Daoutis, S. Coradeschi, and A. Loutfi, "Grounding commonsense knowledge in intelligent systems." *Journal of Ambient Intelligence and Smart Environments*, 2009.
- [7] S. Lemaignan, A. Sisbot, and R. Alami, "Anchoring interaction through symbolic knowledge," in *Proceedings of the 2011 Human-Robot Interaction Pioneers workshop*, 2011.
- [8] J. O'Keefe, *The Spatial Prepositions*. MIT Press, 1999.
- [9] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward understanding natural language directions," in *HRI*, 2010.
- [10] C. Matuszek, D. Fox, and K. Koscher, "Following directions using statistical machine translation," in *Int. Conf. on Human-Robot Interaction*. ACM Press, 2010.
- [11] S. Tellex, "Natural language and spatial reasoning," Ph.D. dissertation, MIT, 2010.
- [12] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock, "Spatial language for human-robot dialogs," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, no. 2, pp. 154-167, 2004.
- [13] J. Flavell, *Perspectives on Perspective Taking*. L. Erlbaum Associates, 1992, pp. 107-139.
- [14] B. Tversky, P. Lee, and S. Mainwaring, "Why do speakers mix perspectives?" *Spatial Cognition and Computation*, vol. 1, no. 4, pp. 399-412, 1999.
- [15] C. Breazeal, M. Berlin, A. Brooks, J. Gray, and A. Thomaz, "Using perspective taking to learn from ambiguous demonstrations," *Robotics and Autonomous Systems*, 2006.
- [16] M. Johnson and Y. Demiris, "Perceptual perspective taking and action recognition," *Advanced Robotic Systems*, 2005.
- [17] J. Trafton, N. Cassimatis, M. Bugajska, D. Brock, F. Mintz, and A. Schultz, "Enabling effective human-robot interaction using perspective-taking in robots," *IEEE Transactions on Systems, Man, and Cybernetics*, 2005.
- [18] H. Wimmer and J. Perner, "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception," *Cognition*, vol. 13, no. 1, pp. 103 - 128, 1983.
- [19] C. Breazeal, J. Gray, and M. Berlin, "An embodied cognition approach to mindreading skills for socially intelligent robots," *I. J. Robotic Res.*, 2009.
- [20] M. Warnier, J. Guitton, S. Lemaignan, and R. Alami, "When the robot puts itself in your shoes. Explicit geometric management of position beliefs," in *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012.
- [21] S. Lemaignan, R. Ros, E. A. Sisbot, R. Alami, M. Beetz, Broqure, D. Sidobre, and R. Alami, "Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction." *International Journal of Social Robotics*, 2012.
- [22] E. Sisbot, R. Ros, and R. Alami, "Situation assessment for human-robot interaction," in *IEEE ROMAN*, 2011.
- [23] F. U. Baron-Cohen S, Leslie AM, "Does the autistic child have a 'theory of mind'?" *Cognition*, vol. 21, no. 1, pp. 37 - 46, 1985.