Towards Human-Aware Cognitive Robots

Rachid Alami, Raja Chatila, Aurélie Clodic, Sara Fleury
Matthieu Herrb, Vincent Montreuil, Emrah Akin Sisbot
LAAS-CNRS
7, Avenue du Colonel Roche 31077 Toulouse, France
Firstname.Name@laas.fr

Abstract

Human-robot interaction requires the robot to explicitly reason on human environments and on its own capacities to achieve tasks in a collaborative way with a human partner. We have devised a decisional framework for human-robot interactive task achievement, embedded in a cognitive architecture, and that is aimed to allow the robot not only to accomplish its tasks but also to produce behaviours that support its commitment vis-a-vis its human partner and to interpret human behaviours and intentions. Together and in coherence with this framework, we develop and experiment various task planners and interaction schemes that allow the robot to select and perform its tasks while taking into account explicitly the human abilities as well as the constraints imposed by the presence of humans, their needs and preferences. We present the first results obtained by our "human-aware" task and motion planners and discuss how they can be extended.

Introduction

The introduction of robots in our daily life raises an additional key issue to the "standard challenge" of autonomous robots: the presence of humans in their environment and the necessity to interact with them. Clearly, the ability to interact with humans should be taken into account in all steps of the robot design. We are conducting research on robot decisional abilities taking into account explicit reasoning on the human environment and on the robot capacities to achieve its tasks in such a context. This research is conducted in the framework of the Cogniron project (http://www.cogniron.org/) that aims at studying and developing the cognitive capacities of a cognitive robot companion.

A cognitive robot should exhibit, besides the more "classical" robotic functions of navigation, motion planning, etc. more semantic capabilities in world and context interpretation, decision-making and human-robot interaction and communication. Learning is a fourth cognitive capacity that has to be transversal to all the robot functions. One of the central issues is the cognitive architecture that organises the robot capabilities. This paper first describes the architecture we propose, then focuses on the framework for decisional human-robot interactive task planning, refinement and execution that allows the robot to select and perform its tasks while taking into account explicitly human abilities as well as constraints imposed by the presence of humans, their needs and preferences.

In the the next section we present a cognitive robot architecture. Then, the human-robot interactive framework is described. The following sections present specific HRI issues in symbolic action planning as well as in motion planning. The last section discusses various extensions.

A Robot Cognitive Architecture

The term "cognitive architectures" is commonly used in the AI and cognitive sciences communities to designate the organisation of systems designed to model the human mind. Two of them, SOAR and ACT-R, are large long term projects and claim objectives of generality.

SOAR (State, Operator And Result) (Rosenbloom, Laird, & A. Newell 1993; Lehman, Laird, & Rosenbloom 2006) is a system aiming at modelling human cognition and at implementing instances of an artificial cognitive system in several domains, ranging from robotics to air combat. It is based on Newell seminal work on theories of cognition (Newell 1990). Operational knowledge in SOAR is represented by production rules. To achieve a goal, the rules' conditions are matched to a "working memory" which contents is coded as sets of {attribute-values}. The working memory includes different kinds of knowledge. Current goals, states and operators are in a context stack in this memory. Hence SOAR implements the concept of "problem spaces" introduced by Newell. Learning in SOAR is mainly based on a mechanism called "chunking" (other mechanisms such as reinforcement learning are being added). This process is similar to identifying macro-operators, i.e. new rules that abstract the succession of rules selected to achieve a goal.

ACT-R (Adaptive Control of Thought-Rational) is a cognitive architecture proposed by John Anderson and colleagues, in development over several years, that represent for its authors a model of the human mind (Anderson *et al.* 2004). The general concept in ACT-R is a classical rule based system. Knowledge is organised in a declarative memory about facts and events and their relationships (the data structures in this memory are called ”chunks” . The rules are selected according to their matching the chunks in the memory, and to their cost and probability of success. Their actions change the declarative memory or trigger actions in the world. However, there is in addition a learning process. The declarative chunks have indeed an associated ”base level” which increases according to their past selection.

There are many similarities between these two main cognitive architecture. They are in agreement on symbolic representations at high levels of abstractions and they both represent operational knowledge by production rules. They both put emphasis on learning mechanisms based on remembering and abstracting previous successful courses of actions. They also both are not very much concerned with real time operation and building representations from sensory data. More precisely, they both say this is important, but at the same time they don't provide a clear approach for achieving it. The question of linking symbolic and subsymbolic representations is actually not really addressed. From a robotics standpoint, this however is a central question.

We propose an architecture (figure 1) that integrates concurrent processes and interacting subsystems performing according to different temporal properties and implementing the cognitive capabilities of interpretation, decision-making, learning and communication. It is in some aspects similar to Aura (Arkin & Balch 1997), when considered at an abstract level (there are functionalities of learning and interaction), but the nature of the exchanges, the global economy of the system is different. There are interaction and learning capabilities in SOAR and ACT-R as well, that we also consider differently here.

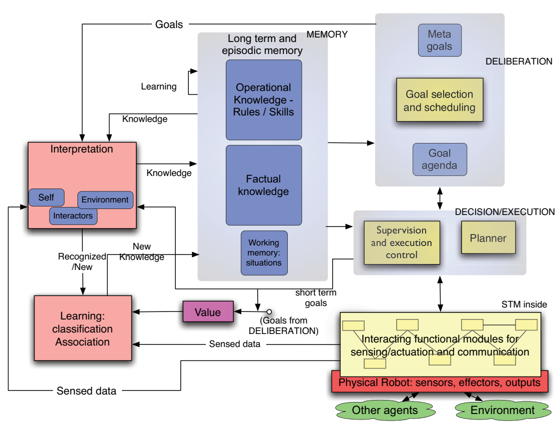

Figure 1: A global cognitive architecture for a cognitive robot. A ”box” in this figure is not necessarily a uniquely geographically located process; it rather represents several processes participating in the same function concurrently. Memory and representations are distributed across the processes even if there is a central repository depicted.

We overview next the learning, perceptual and decisional features of this architecture before focusing on the Human-robot interaction issues.

The Central and Pervasive Role of Learning. We contend that a cognitive robot must have, at the core of its operation, a permanent learning activity. This is indeed the only manner to achieve an open-ended system, i.e., a system able to operate without strong a priori limitations due to predefined representation primitives or operational capabilities. There is a permanent flow of data incoming from the robot's sensors and low level processing functions. The perceived environment includes known and unknown objects and situations. Therefore a permanent process should be able to interpret the flow, discriminating what is already known, and classify data to produce new representations of space, objects and situations. The architecture has to include such an interpretation/learning activity. Interpretation and learning concern the environment, and the robot itself. In our context of a robot companion, we put also a special focus on the robots' interlocutors and interactors who are not just part of the environment. To cope with the information flow, not all incoming data should produce new categories. A saliency measure (or information measure) can filter out part of the data.

The robot acquires knowledge from perception (including through interaction), but it also should be able to produce new knowledge through internal structuring, associations, deduction, abduction and induction mechanisms. The memory that contains factual knowledge is not just a repository of sensed data but an active process which transforms this data into linked structures through those mechanisms. Robot operational knowledge in terms of skills, rules and scripts (i.e. organised procedures) have to be permanently modified and improved, through their evaluation with respect to the achieved tasks in a supervised or a non-supervised manner. We seek a learning process that produces new skills from more basic ones, by associating perception and actuation to produce sensori-motor representations. This association must take into account the context and the pursued goals. We propose to do this by the use of a value function that expresses how much the associations enable to accomplish the goal, i.e., their utility, taking also into account their cost. This value function itself should not be predefined once and for all, but learned so that new associations can be made according to new goals of the robot. This is a reinforcement learning process.

Perception and Interpretation. It has already been largely discussed in the design of control architectures that the system's knowledge cannot be stored in a central unique data base. Representations are rather distributed and are adapted to the processes that handle them (or even confused with them in the case of neural architectures). The architecture we propose has this distributed memory feature. Each processing function has its own representations (e.g., the local environment representations for obstacle avoidance are points or edges for the avoidance servoing), whereas there is a more central representation of a situation as a result of the interpretation process. The coherence of all these representation is an important issue, addressed by Horswill in CEREBUS (Horswill 2001) where he proposes an approach based on a tagging scheme so that a given object can have different representations which remain related to each other. The "interpretation" in our architecture is a data abstraction process that combines sensed data and extracts more global and semantic representations from them. The issue of coherence is to be addressed in the different instances of the architecture. The "Memory" in the architecture is distributed and comprises factual knowledge, which are the representations the robot knows and their temporal relations (episodic memory), and operational knowledge which is the set of rules and procedures representing its capabilities and skills. A situation resulting form interpreting the perceived world is in a working memory is included here.

Decision-Making and Reflection. We consider that the robot has permanent goals on the one hand (such as keeping its energy level sufficiently high, or keeping human satisfaction high, or keeping its integrity), called "metagoals", and goals that arouse from the actual situation as perceived by the interpretation process. In general the robot will be facing multiple conflicting goals, and even if some of the robot's goals are strong, to be implemented by sensori-motor reflexes (such as non collision), the robot might find itself in situations where other priorities have to be taken into account as well. The architecture includes two main decisional processes, named "deliberation" and "decision/execution". The role of the first process is to solve those multiple goal situations and produce a "goal agenda" that will be in turn solved by the decision/execution level whose role is to plan the course of actions that will accomplish these goals. In deciding on its goal agenda or in deciding the course of actions, the system uses the knowledge in the memory, including the operational knowledge and robot state.

Note that the decision/execution system is the decisional component of a hybrid architecture (Alami *et al.* 1998) (the "execution control level" is included in the supervisory module), and the set of "interacting modules" is its functional level. The planning system here is not in charge, but is used as a resource when necessary. The set of sensori-motor functional modules operate and interact based on a Finite State Machine model.

Human-Robot Interaction.

Human-robot interaction (see (Fong, Nourbakhsh, & Dautenhahn 2003) for a general survey on HRI) and communication is not explicitly depicted in the architecture schema. The processes responsible for interaction are part of the perceptual, decisional and action processes. One can consider that there is an interaction process "floating" upon the architectural components that are part of an interaction (data processing, language understanding, dialogue generation, planning for physical interaction, etc.). Interaction is therefore an intricate activity of the system itself. The inputs from robot sensors include the communication inputs and the actions include those that are produced for communication. Interpretation of the exchanges with humans is done within the general interpretation process, and shown in the "interactors" box. Dialogue and other interactions are part of the decisional processes.

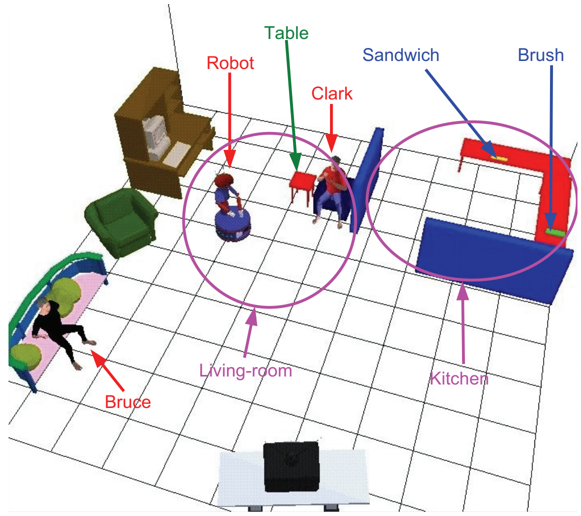

Figure 2: Fetch-an-carry tasks in a human environment.

In our context the human is physically present in the vicinity of the robot, is sensed by the robot and may even participate to the task performance. In relation with this, a number of recent contributions about close interaction deal with the notion of physical and mental safety (Nonaka *et al.* 2004) or the introduction of emotions and/or cognitive models in robotic structures (Breazeal 1998; Nakajima *et al.* 2004). Very often, HRI is merged into the task performance. This tends to reduce HRI to a (sometimes very sophisticated) human interface.

Our aim is to endow the robot with an explicit consideration of humans and with the ability to manage its interactions with them. This must be considered at the task/motion planning and execution level, as well as in the system architecture design (Tambe 1997; Kawamura, Nilas, & Mugumura 2003; Scerri *et al.* 2003; Fong *et al.* 2005).

One key source of inspiration of our work is the Joint Intention theory (Cohen & Levesque 1991; 1990; Kumar *et al.* 2002). It is based on the notion of commitment for team members and defines for a team the concept of Joint Persistent Goal. These definitions constitute a basis for the elaboration of cooperation schemes between heterogeneous agents. We follow a stream similar to (Feil-Seifer & Mataric 2005; Buchsbaum *et al.* 2005; Trafton *et al.* 2005). Indeed, we believe that an effective implementation of this theory can be achieved, when limited to a clearly defined context in which the robot will deal explicitly with the actions, beliefs or intentions of the human partner.

An illustrative scenario

Let us consider the situation illustrated by Figure 2. There are two persons named Bruce and Clark, and a robot named Robot.

Clark wants to eat something. Robot knows that there is a sandwich in the kitchen. It also has to clean the table. The brush is also in the kitchen. Consequently, there are two goals to achieve : (1) clean the table near Clark with the brush and (2) make Clark have the sandwich.

Let us examine some relevant HRI issues in this context. Robot needs specific decisional capabilities in order to elaborate plans that are "legible" (i.e. "understandable") and "socially acceptable" by the humans that are involved in the task or simply present in its vicinity. This has consequences on the tasks that the robot will perform but also on its motions. Not only the robot has to elaborate human-friendly task and motion plans but it has also to continuously observe human activity. Indeed, it has to ensure, when necessary, that the persons involved in the task are doing their part and that its presence and behaviour are accepted.

A Decisional framework

In the decision level of the architecture (see figure 1, bottom-right)(Alami *et al.* 1998), we have introduced a decisional layer for interaction called "InterAction Agents" (IAAs). They are similar to proxies but are directly implemented on the robot side as a representative of a human agent. To make the interaction more explicit we have defined a complete process of establishing a common goal, achieving it and verifying commitment of all agents involved. Besides, relevant IAA models should be devised and used in the robot planning activities. Such models will range from high-level specifications of the human abilities and preferences to geometric attributes such as position, posture or visibility regions.

We envision HRI in a context where two agents (a human and a robot) share a common space and exchange information through various modalities(Clodic *et al.* 2005; Alami *et al.* 2005).

Interaction happens as a consequence of an explicit request of the human to satisfy a goal or because the robot finds itself in a situation where it is useful if not mandatory.

In both cases, the robot has a goal to satisfy. An important issue is the notion of engagement, a process in which the robot will have to establish, maintain and terminate a connection with a human partner. Besides conversation, such a process will provide a framework for robots performing tasks in a human context.

This covers goal establishment, selection of an incremental refinement of the task that is intended to be achieved, and execution monitoring. This context will be used by the robot in order to follow human task performance, to monitor his/her commitment to the common goal, and even to influence it.

The proposed decisional framework (Clodic *et al.* 2005) consists of several entities, having each a specific role as illustrated by Figure 3.

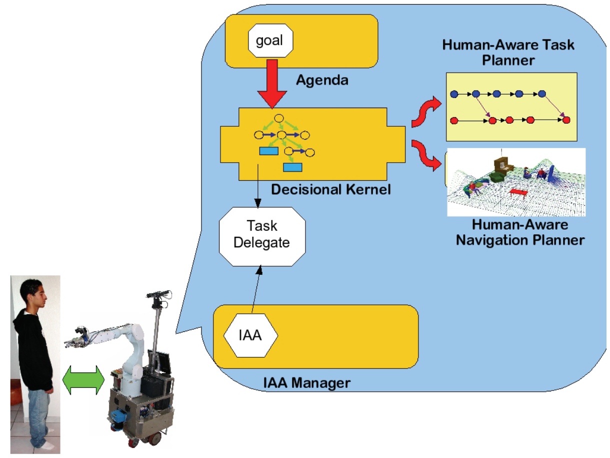

Figure 3: Decisional framework for a HRI-enabled robot: the IAA (InterAction Agent) represents the human state, abilities and preferences. Such information is used by the Human-Aware planners and by the Decisional Kernel.

The HRI we consider in this context is the common achievement of tasks by two agents - a robot and a human - in order to satisfy a joint goal. The human involvement may range from a direct participation to the task achievement, to a simple "acceptance" of robot activity in his close vicinity.

The Agenda Several goals may be sought at a given time, involving possibly several persons. At any moment, there may be several active, inactive and suspended goals. The Agenda manages the current set of robot goals. It ensures the consistency between active goals, and determines their priorities, and their causal links. Based on data provided by the Supervision Kernel, the Agenda determines the relevance of goals and decides to create, suspend, resume or abandon a goal. When a goal is created, it may be associated to the robot alone or to a "team" of agents.

The IAA Manager The humans encountered by the robot are represented by entities called "InterAction Agents" (IAAs). An IAA is created dynamically and maintained by the "IAA Manager". IAAs are containers for various information associated to a human: not only information provided by perception but also its abilities and preferences. This information will be typically used by the planners described in the next sections.

The Task Delegates The set of active goals entails the incremental execution of a set of tasks, some of them involving interaction with humans. Each task corresponding to an active or a suspended goal is represented by an entity called "Task Delegate" that is in charge of monitoring the progress towards the goals of both the robot and the IAA and to assess the level of commitment of the associated person.

The Robot Supervision Kernel The Robot Supervision Kernel is responsible for all tasks selection, refinement and execution. It maintains an integrated view of all robot activities and ensures a global coherence of robot behaviour. It is the only entity that can send execution requests to the functional level.

For each new active goal the Robot Supervision Kernel creates a Task Delegate, selects or elaborates a plan and allocates the roles of each team member. For all the other active goals, the Robot Supervision Kernel has already a plan and is in charge of the execution of the robot part. Whenever an elementary action is performed, the Robot Supervision Kernel forwards this information to all active Tasks Delegates. Depending on the context, the planning process can be more or less elaborated. The planning activity associated to a task is a "continuous process"; it provides, incrementally, the next sub-tasks to achieve. It has also to state, depending on the context, on the feasibility or relevance of the task.

The next sections discuss related issues at task level - HATP, a "Human-Aware Task Planner" - and at motion level - HAMP, a "Human-Aware Motion Planner".

Human-Aware Task Planning

Context The main point here is how high level robot task planning skills should be developed in order to allow it to act as an assistant.

In such a scheme, the robot plans for itself and anticipates the human behaviour in order not only to assess the feasibility of the task (at a certain level) before performing it, but also to share the load between itself and the human (negotiation), and to explain/illustrate a possible course of actions.

One major point is that the robot must not only perform its tasks but also act in a way judged as "acceptable" and "legible" by humans. Other desired features, that fall in the same category, are "predictability" and "directability" (Klein *et al.* 2004).

Representing social constraints We have elaborated a formalisation where both the robot and the human are represented in terms of actions they can perform.

It is based on a formalisation where both the robot and the human are represented in terms of actions they can perform. A "team" composed of two "agents" (the robot and a human) can be represented as: $(A_{human}, C^{ctxt}_{human})$ and $(A_{robot}, C^{ctxt}_{robot})$ where A_i are sets of actions and C^{ctxt}_i are their context-dependent associated costs.

The introduction of costs allows to select preferred behaviours. Indeed, at this level, it is possible to deal with social constraints that can be represented as:

- costs/utilities that denote the difficulty and the pleasure an agent has in an action realisation
- undesirable states (from the human side)
- desirable or undesirable sequences of actions that may induce a robot behaviour that is not understandable (legible) by its human partner
- synchronisations and protocols that may represent social or cultural conventions

Relevant action models and planning algorithms have still to be devised. In a first tentative, we have used an existent planner, in order to assess the pertinence of the approach. A HTN (Hierarchical Task Network) planner SHOP2(Nau *et al.* 2003) has been used mainly because it permits to specify costs for actions and encode procedural knowledge. Examples involved domestic like situations where the robot essentially various actions in interaction and/or in presence of humans.

An example We illustrate here below a use of the current version of HATP for the scenario described above. Two agents are directly involved: Clark and Robot. We assume that they can perform the same set of actions: $A_{Clark} = A_{robot}$.

Typical actions are:

- `(GOTO ?dest)`: moving to from current place to a specified destination `?dest`.
- `(TAKE ?obj)`: picking a object that is placed near the agent
- `(PUT ?obj)`: releasing a grasped object
- `(GIVE ?obj ?a)`: handing the grasped object directly to another agent `?a`.
- `(USE_BRUSH ?furniture)`: cleaning a piece of furniture `?furniture`

In this very simple example, we provide a set of human preferences to the planner. We specify an "undesirable state" corresponding to the situation where Robot holds simultaneously food in one hand and a cleaning object on the other hand. We also specify a (socially) undesirable sequences of actions; for instance, the sequence in which Robot puts an object near a human agent A_i and immediately after A_i takes the same object.

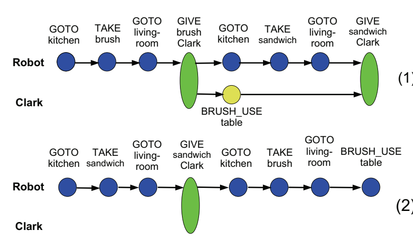

Figure 4: Plans synthesised by HATP.

Figure 4 illustrates two plans elaborated by HATP under different conditions related to the "level tiredness" of Clark. In Plan (1) is a plan, the robot takes into account that Clark is tired and prefers not to move. However, Clark prefers to clean himself the table for efficient reasons. Plan (2) corresponds to the case where Clark is really tired so he does not want to do anything except getting his sandwich.

Note that in both cases, the planner has avoided to produce a plan where Robot gets the brush and the sandwich at the same time even though it is far more efficient in terms of number of actions or energy.

Human-aware motion planning

The presence of humans in the environment raises also new issues to the classic motion/manipulation task planning (Chatila *et al.* 2002; Pacchierotti, Christensen, & Jensfelt 2005; Sisbot *et al.* 2005). Classic motion planners that consider only obstacles and free space are clearly insufficient.

For instance, Figure 5 illustrates two paths generated by a standard motion planner. Both paths are uncomfortable: (1) the robot "springs out" close and then move too close to the seated person, (2) the robot moves in the back of the person.

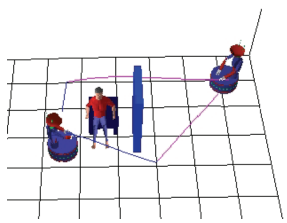

Figure 5: A path produced by a conventional planner: an efficient trajectory, that does not into account the "human parameter", makes the robot move too close to people and sometimes behind them.

We claim that a human-aware motion planner must not only elaborate safe robot paths(Kulic & Croft 2004), but also plan "good", socially acceptable and "legible" paths. Our aim is to build a motion planner that takes explicitly into account the human partner by reasoning about his accessibility, his vision field and potential shared motions.

While several contributions take into account the robot's and humans safety, very few papers, in our knowledge, deal with comfort and legibility issues and often in an ad hoc manner. We believe that our approach can be more generic. We introduce two criteria to the motion planning stage to ensure safety and comfort. The first criterion, called Safety Criterion, mainly focuses on ensuring the humans' safety by controlling the distance between robot and humans present in the environment. The robot, unless necessary, must avoid to approach too much humans. In some cases a given perimeter around humans must not be allowed to pass through.

The second criterion, called Visibility Criterion, takes into account the humans field of view and robot's relative position to it. Humans tend to feel safer and more comfortable when the robot is in their sight. It is preferable that the robot chooses a path as visible as possible to ensure this property. The visible and invisible zones (to the humans' field of view) in the environment can be ranked proportionally to the minimum angular deviation from the humans gaze. Indeed, one can consider this Visibility Criterion as a proportion to the "humans effort to keep the robot in his sight by turning the head or the body". Another aspect concerns zones that are hidden (in the human perspective) by a walls or obstacles of a given height. The sudden appearance of the robot can cause fear and surprise especially if the obstacle is close to the human.

Note that other aspects should be taken into account like speed (time to contact) and acceleration of the robot (or of a part of its structure) particularly when it is in the close vicinity of humans.

We are investigating various minimisation criteria based on a weighted combination of distance, visibility and comfort for computing a satisfactory path and velocity profile. The two criteria mentioned above are represented by numerical potentials stored in 2D grids combining various costs. These costs are highly related to the humans' state, capabilities and preferences. Figure 6 shows safety criterion costs?

A first navigation planner (Sisbot *et al.* 2005) has been built in order to study motion in the vicinity of humans as well approach motions to human. The chosen criteria are based on user trials that have been conducted by (Walters *et al.* 2005).

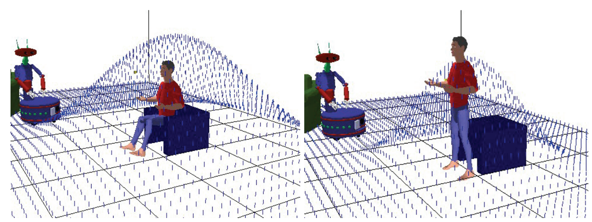

Figure 6: Human relative "Safety grids" when human is sitting and standing are different.

Back to the example: To illustrate the results obtained by our motion planner, we show how the actions selected by HATP are refined and executed at geometric level.

As an input, the motion planner receives the action `(GOTO KITCHEN)` together with a set of complementary information: the next possible action `(TAKE BRUSH)`, the current state of the world S^v which contains the positions and states of the robot, the humans and the objects.

These information are used to adapt HAMP's criteria. For example, there is no human in the kitchen. When planning motion for `(GOTO KITCHEN)`, visibility looses its importance because the robot is already seen and there is nobody at the destination point. This will not be the case when the robot will plan a trajectory from the kitchen to the room where Clark and Bruce are present.

In Figure 7-a, one can see the path generated by HAMP for `(GOTO KITCHEN)`. Although the choice of the final point of the path is not made automatically in the current implementation, the path produced by HAMP takes into account human safety and comfort by staying in the visibility of both persons.

When performing `(GOTO LIVING_ROOM)`, we can see in Figure 7-b that HAMP finds a path that avoids springing out from the kitchen wall too close to the seated person. The robot chooses a path that keeps a certain distance to this wall.

In Figure 7-c, we can see that Bruce came to talk to Clark; so the robot calculates a different trajectory which stays in

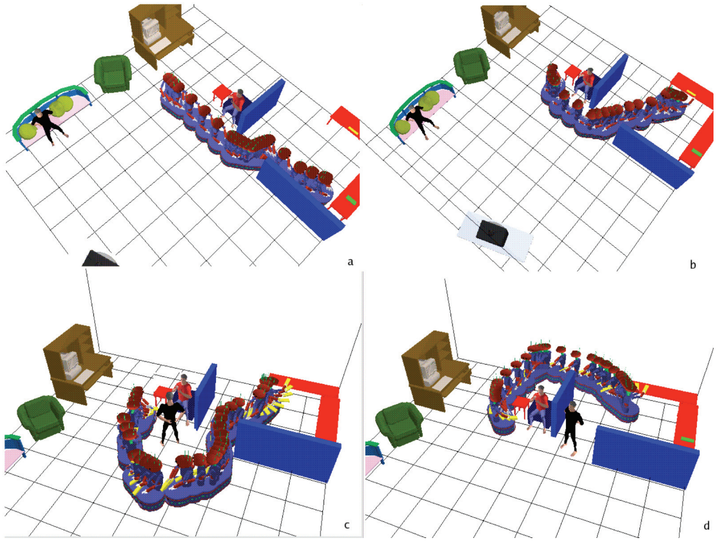

Figure 7: Paths generated by the Human-Aware Motion Planner. Note that (1) the robot avoids to "burst" in the close vicinity of humans, (2) that it tries to stay in their field of sight and (3) that it avoids to approach humans from behind.

Clark's visibility and avoids to pass close to Bruce back.

In the last Figure, the original path is blocked and the robot computes an alternative trajectory (Figure 7-d).

Discussion and future work

The design choices and the results presented here are still preliminary. While the general scheme we propose might be difficult to implement in a general sense, we believe that it is a reasonable challenge to implement it in the case of a personal robot assistant essentially devoted to fetch-and-carry, as well interactive manipulation tasks and associated activities. The robot would operate in a known in-door environment (acquired in a preliminary phase).

Fetch-and-carry and object manipulation task need 3D geometric planning. One challenging problem would be to extend the approach discussed above to the situation where a robot has to hand an object to human. Indeed, there is a need to take into account visibility and reach, in terms of kinematic constraints, of the human partner.

Besides, the robot should produce motion that is acceptable and easily "legible". The human partner should easily understand by observing the robot motion that it is intending to hand an object (Figure 8).

One additional difficulty when considering such issues is the construction of a coherent formalisation that allows to take into account various constraints of different nature. For instance, some of them can be best expressed geometrically while others may be expressed in terms of temporal or causal links between actions.

Acknowledgements

The work described in this paper was conducted within the EU Integrated Project COGNIRON FP6-IST-FET-002020 and the French National Program ROBEA.

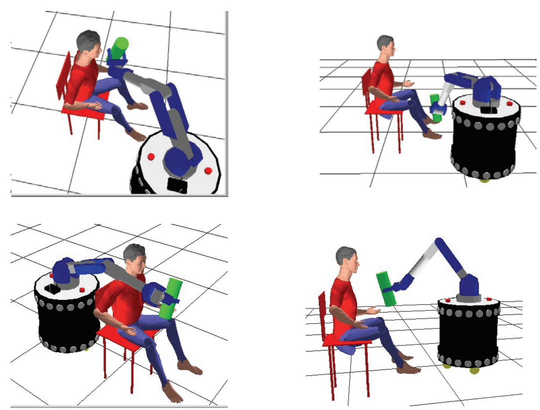

Figure 8: Four ways of handing an object to a person. Only one seems acceptable.

References

Alami, R.; Chatila, R.; Fleury, S.; Ghallab, M.; and Ingrand, F. 1998. An architecture for autonomy. *Internatonal Journal of robotics Research, Special Issue on Integrated Architectures for Robot Control and Programming* 17(4).

Alami, R.; Clodic, A.; Montreuil, V.; Sisbot, E.; and Chatila, R. 2005. Task planning for human-robot interaction. *Proceedings sOc-EUSAI'2005 conference*. Grenoble, France,.

Anderson, J. R.; Bothell, D.; Byrne, M. D.; Douglas, S.; Lebiere, C.; and Qin, Y. 2004. An integrated theory of the mind. *Psychological Review* 111(4):1036–1060.

Arkin, R. A., and Balch, T. 1997. Aura: Principles and practice in review. *Journal of Experimental and Theoretical AI* 9(2).

Breazeal, C. 1998. A motivational system for regulating human-robot interaction. *National Conference of American Association for Artificial Intelligence (AAAI)*.

Buchsbaum, D.; Blumberg, B.; Breazeal, C.; and Meltzoff, A. N. 2005. A simulation-theory inspired social learning system for interactive characters. *IEEE International Workshop on Robots and Human Interactive Communication*.

Chatila, R.; Alami, R.; Simeon, T.; Pettre, J.; and Jaillet, L. 2002. Safe, reliable and friendly interaction between humans and humanoids. *Rapport LAAS No02530, 3rd IARP International Workshop on Humanoid and Human Friendly Robotics* 83–87.

Clodic, A.; Montreuil, V.; Alami, R.; and Chatila, R. 2005. A decisional framework for autonomous robots interacting with humans. *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*.

Cohen, P. R., and Levesque, H. J. 1990. Intention is choice with commitment. *Artificial Intelligence* 42(2-3):213–361.

Cohen, P. R., and Levesque, H. J. 1991. Teamwork. *Nous* 25(4):487–512.

Feil-Seifer, D., and Mataric, M. J. 2005. A multi-modal approach to selective interaction in assistive domains. *IEEE International Workshop on Robots and Human Interactive Communication* 416–421.

Fong, T. W.; Nourbaksh, I.; Kunz, C.; Flckiger, L.; Schreiner, J.; Ambrose, R.; Burridge, R.; Simmons, R.; Hiatt, L. M.; Schultz, A.; Trafton, J. G.; Bugajska, M.; and Scholtz, J. 2005. The peer-to-peer human-robot interaction project. *AIAA Space 2005*.

Fong, T.; Nourbakhsh, I.; and Dautenhahn, K. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems, Special Issue on Socially Interactive Robots* (42 (3-4)):146–163.

Horswill, I. 2001. Tagged behavior-based systems: Integrating cognition with embodied activity. *EEE Intelligent Systems, September/October 2001, pp 30-38*.

Kawamura, K.; Nilas, P.; and Mugumura, K. 2003. An agent-based architecture for an adaptive human-robot interface. *Proceedings of the 36th Hawaii International Conference on System Sciences (HCISS)*.

Klein, G.; Woods, D. D.; Bradshaw, J. M.; Hoffman, R. R.; and Feltovich., P. J. 2004. Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems* 19(6):91–95.

Kulic, D., and Croft, E. 2004. Safe planning for human-robot interaction. *Proc. in IEEE Int. Conf. on Robotics & Automation, New Orleans, USA*.

Kumar, S.; Huber, M. J.; Cohen, P. R.; and McGee, D. R. 2002. Toward a formalism for conversation protocols using joint intention theory. *Computational Intelligence* 18(2).

Lehman, J. F.; Laird, J.; and Rosenbloom, P. 2006. A gentle introduction to soar : An architecture for human cognition: 2006 update. *http://ai.eecs.umich.edu/soar/docs/GentleIntroduction-2006.pdf*.

Nakajima, H.; Morishima, Y.; Brave, S.; Yamada, R.; Maldonado, H.; Nass, C.; Arao, M.; and Kawaji, S. 2004. Toward an actualization of social intelligence in human and robot collaborative systems. *International Conference on Intelligent Robots and Systems (IROS)*.

Nau, D.; Au, T.-C.; Ilghami, O.; Kuter, U.; Murdock, J. W.; Wu, D.; and Yaman, F. 2003. Shop2: An htn planning system. *Journal of Artificial Intelligence Research* (20):379–404.

Newell, A. 1990. *Unified Theories of Cognition*. Harvard University Press, Cambridge, Massachusetts.

Nonaka, S.; Inoue, K.; Arai, T.; and Mae, Y. 2004. Evaluation of human sense of security for coexisting robots using virtual reality. *IEEE International Conference on Robotics and Automation (ICRA)*.

Pacchierotti, E.; Christensen, H. I.; and Jensfelt, P. 2005. Human-robot embodied interaction in hallway settings: a pilot user study. *14th IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man), Nashville, USA*.

Rosenbloom, P.; Laird, J.; and A. Newell, e. 1993. *The Soar Papers: Research on Integrated Intelligence*. MIT Press, Cambridge, Massachusetts.

Scerri, P.; Pynadath, D. V.; Johnson, L.; Paul, R.; Schurr, N.; Si, M.; and Tambe, M. 2003. A prototype infrastructure for distributed robot-agent-person teams. *The Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Sisbot, E. A.; Alami, R.; Simeon, T.; Dautenhahn, K.; Walters, M.; Woods, S.; Koay, K. L.; and Nehaniv, C. 2005. Navigation in presence of humans. *IEEE-RAS International Conference on Humanoid Robots, Humanoids2005, Tsukuba, Japan*.

Tambe, M. 1997. Agent architectures for flexible, practical teamwork. *Proceedings of the National Conference on Artificial Intelligence (AAAI)*.

Trafton, J. G.; Schultz, A. C.; Bugajska, M.; and Mintz, F. 2005. Perspective-taking with robots: Experiements and models. *IEEE International Workshop on Robots and Human Interactive Communication (RO-MAN)*.

Walters, M.; Dautenhahn, K.; Boekhorst, R.; Koay, K. L.; Kaouri, C.; Woods, S.; Nehaniv, C.; Lee, D.; and Werry, I. 2005. The influence of subjects personality traits on personal spatial zones in a human-robot interaction experiment. *14th IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man), Nashville, USA*.