Robots that Work in Collaboration with People

Guy Hoffman and Cynthia Breazeal

MIT Media Lab
20 Ames St, E15-468
Cambridge, MA 02139
{guy, cynthiab}@media.mit.edu

Abstract

Developing robots with social skills and understanding is a critical step towards enabling them to cooperate with people as capable partners, to communicate with people intuitively, and to learn quickly and effectively from natural human instruction. These abilities would enable many new and exciting applications for robots that require them to play a long-term, supportive, and helpful role in people's daily lives. This paper describes our work towards building sociable autonomous robots that can work in collaboration with people. Our approach puts an emphasis on task dialog and social communication under the theoretical framework of joint intention theory.

Introduction

Many of the most useful and new applications for autonomous robots require them to work alongside people as capable, cooperative, and socially savvy partners. For instance, robots are being developed to provide the elderly with assistance in their homes. Such a robot should be persuasive in ways that are sensitive to the person, for example helping to remind them when to take medication, without being annoying or upsetting. In other applications, robots are being developed to serve as members of human-robot teams. NASA JSC's *Robonaut* is a great example (Bluethmann et al. 2003). This humanoid robot is envisioned to work shoulder-to-shoulder with astronauts assisting them in space station maintenance operations.

To provide a human teammate with the right assistance at the right time, a robot partner must not only recognize what the person is doing (i.e., his observable actions) but also understand the intentions or goals being enacted. This style of human-robot cooperation strongly motivates the development of robots that can infer and reason about the mental states of others within the context of the interaction they share.

For applications where robots interact with people as partners, it is important to distinguish **human-robot collaboration** from other forms of human-robot interaction (HRI). Namely, whereas interaction entails action *on* someone or something else, collaboration is inherently working *with* others (Bratman 1992, Grosz 1996)

Much of the current work in human-robot interaction is thus aptly labeled given that the robot (or team of robots) is often viewed as an intelligent tool capable of some autonomy that a human operator commands, perhaps using speech or gesture as a natural interface (Jones and Rock 2002, Perzanowski et al. 2001). This sort of master-slave arrangement does not capture the sense of peer-to-peer partnership that we are pursuing in our own work.

Work with *Robonaut* at NASA JSC has investigated performing a joint task between a human and a teleoperated humanoid robot, whereas our robot is completely autonomous. In other teleoperation work, the notion of partnership has been considered in the form of *collaborative control* (Fong, Thorpe, and Baur 2001) allowing the robot to ask a human for help in resolving perceptual ambiguities. In this approach, the human is used as a remote source of information by the robot. In contrast, our work explores the scenario where a collocated human and robot work on a shared physical task.

This kind of human-robot collaboration is thus an important yet relatively unexplored kind of human-robot interaction. Our paper describes how we apply our theoretical framework (based on joint intention theory) to enable an expressive humanoid robot, *Leonardo*, to work shoulder-to-shoulder with a human teammate on a joint task. To this end, *"Leo"* uses collaborative discourse, gesture, and accompanying social cues. Leo is shown in Figure 1(a).

Joint intention theory

What characteristics must a robot have to work effectively with its human collaborator? To answer this, we look to insights provided by *Joint Intention Theory* (Cohen and Levesque 1991). According to this framework, joint action is conceptualized as doing something together as a team where the teammates share the same goal and a common plan of execution. This *collaborative plan* does not reduce to the sum of the individual plans (Grosz 1996), but

American Association for Artificial Intelligence (www.aaai.org).

consists of an interplay of actions inspired and affected by a joint intention.

Several models have been proposed to explain how joint intention relates to individual intention. Searle argues that collective intentions ("*We*-intentions") are not reducible to individual intentions of the agents involved ("*I*-intentions"), and that the individual acts exist solely in their role as part of the common goal (Searle 1990). Bratman's analysis of Shared Cooperative Activity (SCA) introduces the idea of meshing singular sub-plans into a joint activity (Bratman 1992). We generalize this concept to the idea of dynamically meshing sub-plans.

Bratman also defines certain prerequisites for an activity to be considered shared and cooperative; he stresses the importance of *mutual responsiveness, commitment to the joint activity* and *commitment to mutual support*. Cohen and his collaborators support these guidelines and provide the notion of *joint stepwise execution* (Cohen and Levesque 1991). Their theory also predicts that an efficient and robust collaboration scheme in a changing environment with partial knowledge commands an open channel of communication. Sharing information through communication acts is critical given that each teammate often has only partial knowledge relevant to solving the problem, different capabilities, and possibly diverging beliefs about the state of the task.

Our work integrates these ideas to model and perform collaborative tasks.

Modeling collaborative tasks

Humans are biased to use an intention-based psychology to interpret an agent's actions (Dennett 1987). Moreover, it has repeatedly been shown that we interpret intentions and actions based on goals, not specific activities or motion trajectories (e.g. Baldwin and Baird 2001). A goal-centric view is particularly crucial in a collaborative task setting, in which goals provide a common ground for communication and interaction.

All of this argues that goals and a commitment to their successful completion must be central to our intentional representation of tasks, especially if those should be performed in collaboration with others.

Intention and Task Representation

We represent tasks and their constituent actions in terms of *action tuples* (Burke et al. 2001) with the additional notion of goals. These goals play a central role both in the *precondition* that triggers the execution of a given action tuple, and in the *until-condition* that signals when the action tuple has successfully completed.

Our task representation currently distinguishes between two types of goals: (a) *state-change* goals that represent a change in the world, and (b) *just-do-it* goals that need to be executed regardless of their impact on the world. These two types of goals differ in both their evaluation as preconditions and in their evaluation as until-conditions. As part of a precondition, a state-change goal must be evaluated before doing the action to determine if the action is needed. As an until-condition, the robot shows commitment towards the state-change goal by executing the action, over multiple attempts if necessary, until the robot succeeds in bringing about the desired new state. This commitment is an important aspect of intentional behavior (Bratman 1992, Cohen and Levesque 1991). Conversely, a just-do-it goal will lead to an action regardless of the world state, and will only be performed once.

Tasks are represented in a hierarchical structure of actions and sub-tasks (recursively defined in the same fashion). Since tasks, sub-tasks, and actions are derived from the same *action tuple* data structure, a tree structure is naturally afforded. It should be noted that goals are also associated with the successful completion of an overall task or sub-task, separate from the goals of each of the task's constituents.

Intention and Decision-Making

When executing a task, goals as preconditions and until-conditions of actions or sub-tasks manage the flow of decision-making throughout the task execution process. Additionally, overall task goals are evaluated separately from their constituent action goals. This top-level evaluation approach is not only more efficient than having to poll each of the constituent action goals, but is also conceptually in line with a goal-oriented hierarchical architecture. For example, consider a task with two actions. The first action makes some change in the world (and has a state-change goal), and the second action reverses that change (also a state-change goal). The overall task goal has no net state change and becomes a just-do-it goal even though its constituent actions both have state-change goals.

Task manager

The *task manager module* maintains a collection of known task models and their associated names. Given this set of tasks, the robot listens for speech input that indicates a task-related request from the human partner. These can be in the form of: "Leo, do *task x*" or "Leo, let's do *task x*." These requests can also be made in the form of a question: "Leo, can you do *task x*?" In the case of a question, given Leonardo has no speech generating capabilities yet, the robot will answer by either nodding "yes" or shaking its head "no." If the robot does not recognize the name of the requested task, or if the robot does not know how to perform it, he looks puzzled or shrugs his shoulders "I don't know."

The task manager distinguishes between requests for autonomous task completion and invitations to task collaboration, and starts the appropriate execution module. If Leo is asked to do a known task on his own, then the task manager executes it autonomously by expanding the task's actions and sub-tasks onto a focus stack (in a similar way to Grosz and Sidner 1990). The task manager proceeds to work through the actions on the stack popping them as they are done and, upon encountering a sub-task, pushing its constituent actions onto the stack. The robot thus progresses through the task tree until the task's goals are achieved.

The major contribution of this work, however, concerns the collaborative scenario: If a collaborative task execution is requested, the task manager starts the collaboration module to jointly execute a common plan.

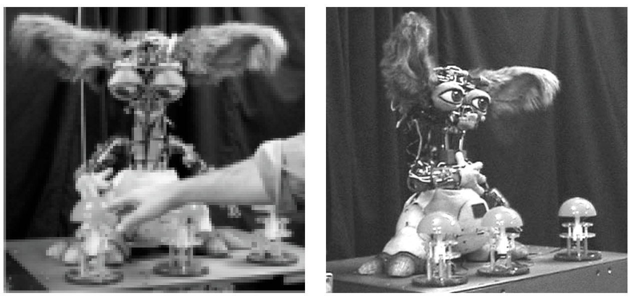

Figure 1: (a) Leonardo participating in a collaborative button-pressing task. (b) Leonardo negotiating his turn for an action he is able to perform.

Performing Tasks with Humans

When collaborating with a human partner, many new considerations come into play. For instance, within a collaborative setting the task can (and should) be divided between the participants, the collaborator's actions need to be taken into account when deciding what to do next, mutual support must be provided in cases of one participant's inability to perform a certain action, and a clear channel of communication must be used to synchronize mutual beliefs and maintain common ground for intentions and actions.

Our implementation supports these considerations as Leonardo participates in a collaborative discourse while progressing towards achieving the joint goal. To do so, and to make the collaboration a natural human interaction, we have implemented a number of mechanisms that people use when they collaborate. In particular, we have focused on communication acts to support joint activity (utilizing gestures and facial expressions), dynamic meshing of sub-plans, turn taking, and an intuitive derivation of *I*-intentions from *We*-intentions.

Experimental Setup

In our experimental scenario there are three buttons in front of Leonardo. The buttons can be switched ON and OFF (which changes their color). Occasionally, a button that is pressed does not light up, and in our tasks this is considered a failed attempt. We use tasks comprised of vision and speech recognition and simple manipulation skills. For instance, Leonardo can learn the names of each of the buttons and is able to point to and press the buttons.

To test our collaborative task execution implementation, we designed a set of tasks involving a number of sequenced steps, such as turning a set of buttons ON and then OFF, turning a button ON as a sub-task of turning all the buttons ON, turning single buttons ON and others. This task set represents simple and complex hierarchies and contains tasks with both state-change and just-do-it goals.

Dynamic Meshing of Sub-plans

Leo's intention system is a joint-intention model that dynamically assigns tasks between the members of the collaboration team. Leo derives his *I*-intentions based on a dynamic meshing of sub-plans according to his own actions and abilities, the actions of the human partner, Leo's understanding of the common goal of the team, and his assessment of the current task state.

Leonardo is able to communicate with the human teammate about the commencement and completion of task steps within a turn-taking interaction. Specifically, the robot is able to recognize changes in the task environment, as well as successes and failures on both Leo's and his teammate's side. Most importantly, Leonardo is able to communicate to the human teammate the successful completion or inability to accomplish a crucial task step to the complete joint action.

Self-Assessment and Mutual Support

At every stage of the interaction, either the human should do her part in the task or Leo should do his. Before attempting an element of the task, Leo negotiates who should complete it. For instance, Leo has the ability to evaluate his own capabilities. In the context of the button task, Leonardo can assess whether he can reach each button or not. If he is able to complete the task element (e.g., press a particular button) then he will offer to do so. Conversely, whenever he believes that he cannot do the action (e.g., because he cannot reach the button) he will ask the human for help.

Since Leonardo does not have speaking capabilities yet, he indicates his willingness to perform an action by pointing to himself, and adopting an alert posture and facial expression (see: Figure 1(b)). Analogously, when detecting an inability to perform an action assigned to him, Leo's

#	Human	Leonardo	Notes
1	"Leo, let's do task BUTTONS"	Shrugs "I don't know"	Leo does not know this task.
2	"Let's do task BUTTON-ONE"	Looks at the buttons	Leo acknowledges that he understands the task, and visibly establishes mutual belief on the task's initial conditions.
3		Points to himself	He can do the first (and only) part of the task, and suggests doing so.
4	"OK, you go"	Presses button one, looking at it	Looking away from the partner while operating establishes turn taking boundaries.
5		Looks back at his partner	Gaze shift is used to signal end of turn
6		Nods shortly	Communicates the robot's perceived end of task
7	"Leo, let's do task BUTTON-ONE"	Looks at the buttons; points to himself	As in steps 2-3
8	"I'll go "	Looks at his partner	
9	Presses button one	Looks at button one	Acknowledges partner's action, creates mutual belief
10		Nods shortly	Communicates perceived end of task.
11	Moves button one out of Leo's reach		
12	"Let us do task BUTTON-ONE"	Looks at buttons	Leo acknowledges that he understands the task, and visibly establishes mutual belief on the task's initial conditions.
13		Looks at button one, then back at the human partner; extends his arms in "Help me" gesture.	Leo assesses his capabilities and consequently requests support.
14	Presses button one	Looks at button one; looks back at human; nods shortly.	Glance acknowledges partner's action and creates mutual belief as to the task's completion.
15	"Let us do task BUTTON-ONE-AND-TWO"	Looks at buttons	Leo acknowledges that he understands the task, and visibly establishes mutual belief on the task's initial conditions
16		Points to himself	He can do the first part of the task, and suggests doing so.
17	"OK, you go"	Presses button one, looking at it	
18	At the same time as 17, presses button two		
19		Looks at button two; looks back at the human; nods shortly	Acknowledges partner's simultaneous action, creates mutual belief as to the task's completion.

Table 1: Task collaboration transcript on a single-level task. This table shows a sample transcript describing a characteristic interaction between Leonardo and a human teammate. We chose to display the following simple, non-hierarchical tasks for reasons of transcript brevity: *BUTTON-ONE* – Toggle button one, *BUTTON-ONE-AND-TWO* – Turn buttons one and two ON. While these do not illustrate the Leonardo's full range of goal-oriented task representation, they offer a sense of the joint intention and communicative skills fundamental to the collaborative discourse stressed in this paper.

expression indicates helplessness, as he gestures toward the human in a request for her to perform the intended action. Additionally, Leo shifts his gaze between the problematic button and his partner to direct her attention to what it is that the robot needs help with.

Communication to Support Joint Activity

While usually conforming to this turn-taking approach, the robot can also keep track of simultaneous actions, in which the human performs an action while Leo is working on another part of the task. If this is the case, Leonardo will take the human's contribution into account and reevaluate the goal state of the current task focus. He then might decide to no longer keep this part of the task on his list of things to do. However, the robot needs to communicate this knowledge to the human to maintain mutual belief about the overall task state.

We have implemented a variety of gestures and other social cues to allow the robot communicate his internal state during collaboration – such as who the robot thinks is doing an action, or whether the robot believes the goal has been met. For instance, when the human partner unexpectedly changes the state of the world, Leo acknowledges this change by glancing briefly towards the area of change before redirecting his gaze to the human. This post-factum glance lets the human know that the robot is aware of what she has done, even if it does not advance the task.

If the human's simultaneous action contributes in a positive way to the task, such as turning ON a button during the buttons-ON sub-task, then Leo will glance at the change and give a small confirming nod to the human. Similarly, Leo uses subtle nods while looking at his partner to indicate when the robot thinks a task or sub-task is completed. For instance, Leo will give an acknowledgement nod to the human when the buttons-ON sub-task is completed before starting the buttons-OFF sub-task (in case of the buttons-ON-then-OFF task).

All of these play an important role in establishing and maintaining mutual beliefs between human and robot on the progress of the shared plan. See Table 1 for a sample interaction transcript of our system.

Results and Future work

In summary, during the trials for the collaborative button task, Leonardo displayed successful meshing of sub-plans based on the dynamic state changes as a result of his successes, failures, and the partner's actions. Leo's gestures and facial expressions provided a natural collaborative environment, informing the human partner of Leo's understanding of the task state and his attempts to take or relinquish his turn. Leo's requests for help displayed his understanding of his own limitations, and his use of gaze and posture served as natural cues for the human to take appropriate action in each case. See Table 1 for a transcript of a typical collaborative interaction.

As future work, we would like to improve the complexity of the task representation as well as the interaction and dialog. Leonardo can understand a few spoken requests of the human, but he does not speak himself. Although his gestures and facial expressions are designed to communicate his internal state, combining this with an ability to speak would give the robot more precision in the information that he can convey. We would also like to implement a richer set of conversational policies to support collaboration. This would be useful for negotiating the meshing of sub-plans during task execution to make this process more flexible and efficient. We continue to make improvements to Leonardo's task representation so that he can represent a larger class of collaborative tasks and more involved constraints between the tasks' action components.

Conclusion

Building sociable robots has profound implications for how we will be able to engage robots in the future – far beyond making them appealing, entertaining, or providing an easy interface to their operation. It is a critical competence that will allow robots to assist us as capable partners.

This paper presents an overview of our work to build sociable robots that work cooperatively with people using natural dialog, gesture, and social cues. We have presented how our ideas, informed by joint intention theory, can be applied to building and demonstrating robots that engage in self-assessment and provide mutual support, communicate to support joint activity, perform dynamic meshing of sub-plans, and negotiate task division via turn taking.

Acknowledgements

The work presented in this paper is a result of the ongoing efforts of the graduate and undergraduate students of the MIT Media Lab Robotic Life Group and our collaborators. Leonardo's speech understanding abilities are developed in collaboration with Alan Schultz and his group at the Naval Research Lab. Stan Winston Studio provided the physical Leonardo robot. Leonardo's architecture is built on top of the *C5M* code base of the Synthetic Character Group at the MIT Media Lab. Geoff Beatty and Ryan Kavanaugh provided the virtual model and animations. This work is funded in part by a DARPA MARS grant and in part by the *Digital Life* and *Things that Think* consortia.

References

Bluethmann, W., Ambrose, R., Diftler, M., Huber, E., Goza, M., Lovchik, C., and Magruder, D. 2003. Robonaut: A Robot Designed to Work with Humans in Space. *Autonomous Robots* 14: 179-207.

Bratman, M. 1992. Shared Cooperative Activity. *The Philosophical Review* 101,2: 327-341.

Grosz, B. 1996. Collaborative Systems: 1994 AAAI Presidential Address. *AI Magazine* 2,17: 67-85.

Jones, H., and Rock, S. 2002. Dialog-based Human-robot Interaction for Space Construction Teams. *Proc. IEEE Aerospace Conference*.

Perzanowski, D., Schultz, A., Adams, W., Marsh, E., and Bugajska, M. 2001. Building a Multimodal Human-Robot Interface, *IEEE Intelligent Systems*: 16-20.

Fong, T., Thorpe, C., and Baur, C. 2001. Collaboration, Dialogue, and Human-Robot Interaction. *Proc. of the International Symposium of Robotics Research*.

Cohen, P., and Levesque, H. 1991. Teamwork. *NOÚS* 25: 487-512.

Searle, J. 1990. Collective Intentions and Actions. In Cohen, Morgan and Pollack eds.: *Intentions in communication*, MIT Press: Chapter 19.

Dennett, D. 1987. Three kinds of intentional psychology. In *The Intentional Stance*, MIT Press: Chapter 3.

Baldwin, D., and Baird, M. 2001. Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5, 4: 171-178.

Burke, R., Isla, D., Downie, M., Ivanov, Y., and Blumberg, B. 2001. Creature Smarts: The art and architecture of a virtual brain. *Proc. of the Computer Game Developers Conference*.

Grosz, B. J. and Sidner, C. L. 1990. Plans for discourse. In Cohen, P. R., Morgan, J., and Pollack, M. E., eds. *Intentions in communication*: chapter 20, 417–444. MIT Press, Cambridge, MA.