Social Robots: Beyond Tools to Partners

C. Breazeal, J. Gray, G. Hoffman, M. Berlin
MIT MEDIA LAB
20 Ames St, E15-468
Cambridge, MA, 02139, USA
E-mail cynthiab@media.mit.edu

Abstract

This paper presents our efforts towards building sociable autonomous robots that can work in collaboration with people. In teamwork, it is critical that a robot partner be able to infer and understand the human's goals and intentions in order to anticipate the person's needs and offer appropriate assistance in a timely manner. We present our framework for human-robot collaboration based on joint intention theory, and our initial efforts to develop a simulation-theoretic approach for anticipating the human partner's goals to proactively offer help. 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.

1 Introduction

Many new applications for personal or professional service 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. Many of these interactions require the robot to work jointly with the person, such as helping them to get out of bed, to walk up stairs, to prepare a meal together, etc. In other applications, robots are being developed to serve as members of human-robot teams for urban search and rescue, space exploration, construction tasks, and more.

1.1 Related work

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 [1].

Much of the current work in human-robot interaction is thus aptly labelled 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) [2]. For instance, work on *Robonaut* at NASA JSC has investigated performing a joint task between a human and a teleoperated humanoid robot [3]. This sort of master-slave arrangement does not capture the sense of partnership that we mean when we speak of working "jointly with" others as in the case of collaboration.

In other teleoperation work, the notion of partnership has been considered in the form of *collaborative control* [4]. The robot maintains a model of the user, can take specific commands from the operator, and also has the ability to ask the human questions to resolve issues in the plan or perceptual ambiguities. The role of the human in the partnership is to serve as a reliable remote source of information for the robot. A similar approach has been explored by Woern and Laengle [5].

Kimura *et al.* explore human-robot collaboration with vision-based robotic arms [6]. While addressing many of the task representation and labour division aspects necessary for teamwork, it views the collaborative act as a planning problem, devoid of any social aspect. As such, it does not take advantage of the inherent human expertise in generating and understanding social acts. As a result, the interaction requires the human teammate to learn gestures and vocal utterances akin to programming commands.

In sum, on one hand previous works have dealt with the scenario of a robot being the tool towards a human's task goal, and on the other, the human being the tool in a robot's task goal. In contrast, our perspective is that of a balanced partnership where the human and robot maintain and work together on shared task goals. We propose a different notion of partnership that has not been addressed in prior works: that of an autonomous robot working with a human as a member of a collocated team to accomplish a shared task.

1.2 Robots as partners

In realizing this goal, we believe that robots must be able to cooperate with humans as capable partners and communicate with them intuitively in human terms. For instance, consider the following collaborative task where a human and a humanoid robot work together shoulder-to-shoulder. The shared goal of the human and the robot is to assemble a physical structure. Both have different capabilities---the human being more dexterous. The task requires different tools and different equipment. Given these constraints, the human's responsibility is to operate the tools necessary to assemble the structure. The robot's responsibility is to be a helpful assistant, providing the human with the appropriate tools at the right time, sharing relevant knowledge, and helping to manoeuvre the awkward pieces of the assembly into place so that they may be fastened together by the human.

To be an effective assistant, the robot must be able to infer the human's goals and intentions from her observable behaviour and task context in order to anticipate her needs and offer relevant and timely assistance. For instance, if the human is fumbling with an awkward piece of equipment, her surface behaviour will not match typical successful instances of enacting the goal. Ideally, the robot could infer what the human is trying to do and consequently hold the awkward piece steady to help the human accomplish her goal, or possibly complete the goal for her.

This collaborative scenario motivates our current efforts in two significant directions. The first part of this paper presents an overview of our implementation of collaborative processes, communication policies, goal-oriented task representation and evaluation procedures to support joint action. This work enables our expressive humanoid robot, *Leonardo* (see Figure 1), to work shoulder-to-shoulder with a human teammate towards accomplishing a joint task --- an excerpt is given in Appendix A (see [7] for a detailed technical presentation).

Next, for a robot partner to provide a human teammate with the right assistance at the right time, it must not only recognize what the person is doing (i.e., his observable actions) but also understand the intentions or goals being enacted. Hence, the second part of this paper presents our initial efforts to enable Leonardo to infer the intended goal of its human collaborator and to proactively help her to achieve it.

2. Human-robot collaboration

What characteristics must a robot have to work effectively with its human collaborator? Bratman's analysis of Shared Cooperative Activity (SCA) [8] 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 [9] support these guidelines and provide the notion of *joint stepwise execution*. They also stress the importance of *communication* between teammates to achieve efficient and robust collaboration within a changing environment 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. These core collaborative processes enable individuals (with their respective goals and plans) to perform as a team to pursue a common goal with a shared plan of execution. Our work integrates these ideas to model and perform collaborative tasks by a human-robot team. We summarize our approach below and refer the reader to Breazeal *et. al*. [7] for a detailed presentation of our collaborative processes and communication policies.

2.1 Task representation and evaluation

Humans are biased to use an intention-based psychology to interpret an agent's actions [10]. A goal-centric view is particularly crucial in a collaborative task setting, in which goals provide a common ground for communication and interaction. This argues that goals and a commitment to their successful completion must be central to an intentional representation of tasks, especially if those should be performed in collaboration with others.

To satisfy this requirement, tasks are represented in a hierarchical structure of goal-directed actions and sub-tasks (recursively defined in the same fashion). The *task manager module* maintains a collection of known task models (i.e., recipes [1, 11]) and their associated names. Tasks, sub-tasks, and actions are derived from the same *action tuple* data structure consisting of a *precondition*, an *executable* (e.g., action), and *until-condition* (e.g., expected result) [12]. 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. The *until-condition* ensures that the robot demonstrates commitment to its goals --- causing the robot to reattempt failed actions to achieve the intended result. This commitment is an important aspect of intentional behaviour [1, 8, 9].

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. The task manager is responsible for expanding the task's actions and sub-tasks onto a focus stack (in a similar way to [11]) and 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. Importantly, 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 important to support *dynamic evaluation* of the overall task state given that the human teammate can make contributions (that might advance or hinder) the overall task goal at any time. This allows the robot to dynamically reconsider its own contributions and to coordinate them with those of the human.

2.2 Collaborative policies

When collaborating with a human partner, however, 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 (see Appendix A for an example interaction). 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 [9, 11].

Leo's intention system is a *joint-intention model* that dynamically assigns tasks between the members of the collaboration team. Leo derives his own 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.

For instance, 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. If he is able to complete the task element then he will offer to do so by pointing to himself (to communicate "I can do it."). Conversely, whenever he believes that he cannot do the action, he will ask the human for help by looking at the problematic task element and then looking to the human while gesturing toward her (to communicate "Can you do it?").

Leo 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 re-evaluate 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. The robot communicates this knowledge to the human to maintain mutual belief about the overall task state using a variety of gestures and other social cues. 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. Conversely, if the human's simultaneous action contributes in a positive way to the task, then Leo will glance at the change and give a small confirming nod to the human.

All of these communicative acts (and more), framed within a turn-taking interaction as a collaborative dialog, play an important role in establishing and maintaining mutual beliefs between human and robot on the progress of the shared plan. See Table 1 in appendix A for a sample interaction transcript of our system during a button-pressing task.

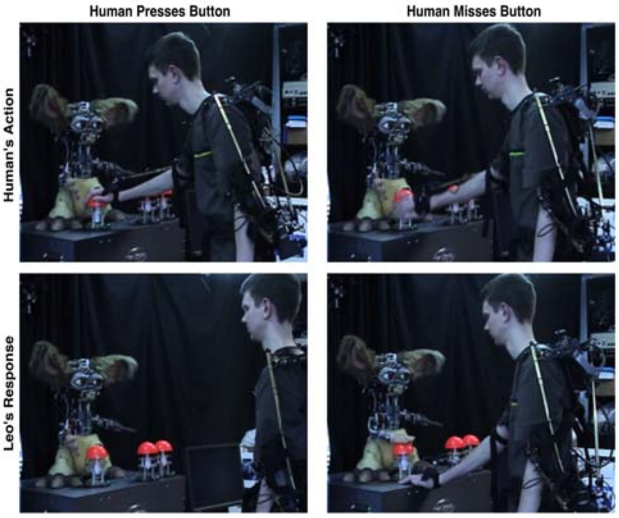

Figure 1: Goal inference for a button-pressing task.

3 Toward Understanding Human Intent

In general, our collaborative processes and communication policies for coordinating joint action work well when the human-robot teammates adhere to the shared plan and each teammate can achieve their negotiated goal. It becomes more interesting, however, when unexpected problems arise --- for instance, when the human teammate is having difficulty achieving her agreed upon goal. In this case, the robot must be able to first determine what she is trying to accomplish order to figure out how to best help her.

Our approach for inferring, relating to, and reasoning about the intentions of others is based on the theoretical framework of Simulation Theory [13]. Simulation theory argues that we exploit our own psychological responses in order to simulate others' minds to infer their mental states. With respect to action recognition and goal inference, this implies that the ability for a robot to perform an action should also help it to recognize that same action when performed by others. It further implies that a robot could leverage its own action/goal representation to infer goals of others.

To implement a working model of goal inference based on a simulation-theoretic approach, our implementation decomposes into three main sub-problems. First, movement data about the human must be perceived, and then converted into a similar representation as the internal movement data of the robot. Next, the robot can then attempt to determine if any motor skills within its own repertoire could produce the observed movement. Finally, if it finds such a motor skill, the robot can discover which of its actions could produce such a skill, and, based on the ending conditions of its own action, try to guess the goal of the human by assuming that the human is sufficiently “like-me” (a foundational assumption according to developmental psychology accounts for the acquisition of a theory of other minds [18]). In principle, this approach could be extended to not only infer the goals of others, but other mental states as well (motivations, emotions, desires, beliefs, etc).

3.1 Perceptual Mapping

The idea of sharing a representation between motor actions or behaviours produced by others and oneself has been discussed in the fields of philosophy [13], developmental psychology [14, 15], autism [16] and neuroscience [17]. Mirror neurons are seen as evidence for this shared representation, as they have been shown to fire similarly in both cases [17]. The discovery of mirror neurons is also evidence for Meltzoff and Moore’s (1997) Active Intermodal Mapping Hypothesis (AIM), which proposes a modality independent representation common to perception and production. This shared representation is thought to be a starting point for a simulation theoretic process of inferring goals and intentions in humans [14, 15].

A number of motor learning efforts for robotic systems have looked to mirror neurons for their biological inspiration (see [19] for a review); some even used simulated mirror neurons to recognize movements both of the robot and of the human interacting. Similar to these works, we have chosen to use sets of joint angles, broken down by time into individual poses, as the intermodal space. These body configurations (poses) string together to form movement trajectories over time. This is the same representation used by Leonardo to control motor production. The first step, then, is getting perceptual data into this space.

Melzoff and Moore hypothesize that human infants learn the intermodal mapping during early imitative exchanges with adults. Inspired by this, Leonardo learns this mapping via imitation as well. The mapping from observed human movement data (in our experiments we used a motion capture suit for the right arm of a human demonstrator) into poses in Leo's own joint space is achieved by training a Radial Basis Function (RBF) model to correlate poses of the human with poses in Leo's joint space. An RBF needs a set of known good mappings from one vector space to another, from which it can extrapolate the mappings for new unseen input points. To train the RBF model, an imitative interaction is used to acquire a number of matching pose pairs. Each pair must contain two representations of the same actual pose, one in Leo's joint space and the other represented as a frame of motion capture data. Building on the work described in [20], Leonardo progresses through 18 training poses (selected to provide a good distribution over his movement space) where the robot leads the interaction by adopting a pose that the human then imitates. After the human has mimicked all 18 poses, the robot has acquired enough data and can train the model.

3.2 Movement Matching

Once Leonardo is able to represent observed movements in terms of its own joint angles, the next challenge is to recognize observed trajectories in terms of its own repertoire of actions. This is equivalent to matching these observed sequences to existing sequences in Leo's motor production repertoire as represented in its posegraph [21]. This is accomplished by first “parsing” observed movement into discrete units. The *segmentor process* divides up the observed movement of the participant by detecting pauses in the motion of the end effector (the position of the end effector is calculated for a given joint configuration using our skeletal model of the robot). These chunks are then handed to the *matcher process* to be matched against movements in Leo's repertoire.

The *matcher* uses a representation of end effector movement called the *Movement Axis Model.* For any given sequence of observed end effector positions, the matcher computes a Movement Axis Model that consists of the average position of the end effector, as well as the direction and length of the “major axis” of movement of the end effector (computed by finding the two end effector positions in the movement that are furthest away from each other; the vector between these two positions is the “axis of the movement”). In addition, the Movement Axis Model for each movement in Leo's repertoire is computed once and then stored for later use.

The Movement Axis Model representation allows observed movement to be compared to existing movements within Leonardo’s repertoire. The distance between the model for the observed movement and the model for each of Leo's own movements is calculated, and the most similar model in Leo’s repertoire is declared to be the match. The distance between any two Movement Axis Models is computed as the sum of three quantities: 1) The distance between the average end effector positions; 2) The difference between the lengths of the major axes; and 3) the angular distance between the two major axes, multiplied by the average length of the two major axes.

3.3 Goal Inference

Finally, these low-level movement trajectories are mapped onto goals in Leo's internal behaviour representation. These goals are then be hypothesized to be the goals of the human participant. As described in section 2.1, high-level tasks are represented as action tuples. Each action tuple encodes a goal-directed behavior --- the desired result is specified by the *do-until* condition. Thus the task of goal inference is to map low-level movements onto the *do-until* contexts of specific action tuples.

When an observed movement is matched to a movement in Leo's motor repertoire, a search is performed to find the action tuple that would most likely generate the observed movement. Our initial implementation is quite simple and makes a strong assumption that there is a one-to-one mapping between movements and action tuples. We are currently developing a more sophisticated goal-inference mechanism, where the search for the most relevant action tuple incorporates not only movement information, but also information about the object being targeted, as well as clues about the task context and social context of the behaviour. Extending this search is as an interesting and important area of ongoing work, along with developing a richer representation of goals.

3.4 Helpful Acts

The *goal inference module* operates in conjunction with a module that monitors the participant's progress towards the inferred goal. If the participant finishes her movement but the goal condition did not occur, the *progress module* detects this failure and motivates the robot to assist the participant by achieving the goal for her. In the simplest case, the robot's corresponding action tuple is activated, causing Leo to perform a similar movement to attempt to satisfy the desired goal condition. For instance, in Leo's button pushing tasks, the robot can help the human to activate or deactivate a particular button if her previous attempt had failed (see Figure 1).

4 Conclusion

This paper presents an overview of our work to build 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. We have outlined our initial efforts in using a simulation-theoretic approach to give Leonardo the ability to model, infer, relate to the intended goal of its human collaborator.

Ongoing work includes extending the perceptual mapping task to the rest of the robot's body and to incorporate visual observation of human movement, developing more sophisticated representations for movement and similarity metrics for movement matching, and bolstering our goal inference mechanisms with additional sources of information (task information, perceptual context, etc.) to enable the robot to perform more advanced joint tasks with humans.

Acknowledgements

This work is funded in part by a DARPA MARS grant and in part by the *Digital Life* and *Things that Think* consortia.

References

[1] Grosz, B. "Collaborative Systems: 1994 AAAI Presidential Address", *AI Magazine*, Vol. 2, No. 17, pp. 67-85, 1996.

[2] Jones, H., and Rock, S. "Dialog-based Human-robot Interaction for Space Construction Team",*Proc. IEEE Aerospace Conference*, 2002.

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

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

[5] H. Woern & T. Laengle, A. "Cooperation between Human Beings and Robot Systems in an Industrial Environment", *M&R*, Vol. 1, pp. 156-16, 2000.

[6] H. Kimura, T. Horiuchi and K. Ikeuchi, "Task-Model Based Human Robot Cooperation Using Vision", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'99)*, pp. 701-706, 1999.

[7] Breazeal, C., Hoffman, G. and Lockerd, A. "Teaching and Working with Robots as a Collaboration", *Proceedings of AAMAS-04*, 2004.

[8] Bratman, M. "Shared Cooperative Activity", *The Philosophical Review*, Vol. 101, No. 2. pp. 327-341, 2002.

[9] Cohen, P., and Levesque, H. "Teamwork", *NOÚS*, Vol. 25, pp. 487-512, 1991.

[10] Baldwin, D., and Baird, M. "Discerning intentions in dynamic human action", *Trends in Cognitive Sciences*, Vol. 5, No. 4, pp. 171-178, 2001.

[11] B. J. Grosz and C. L. Sidner, "Plans for discourse" in P. R. Cohen and J. Morgan and M. E. Pollack (eds.), *Intentions in communication* (MIT Press, Cambridge, MA), pp. 417—444. 1990.

[12] Burke, R., Isla, D., Downie, M., Ivanov, Y., and Blumberg, B. Creature Smarts: The art and

architecture of a virtual brain. *Proc. of the Computer Game Developers Conference*, 2001.

[13] Gordon, R. “Folk Psychology as Simulation,” *Mind and Language*, Vol. 3, No. 2, pp. 158-171, 1986.

[14] A. Meltzoff & M. K. Moore, “Explaining facial imitation: A theoretical model”, *Early Development and Parenting*, Vol. 6, pp. 179-192, 1997.

[15] Meltzoff, A. N. 1995 Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* Vol. 31, pp. 838-850.

[16] Williams, J.H.G., Whiten, A., Suddendorf, T., & Perrett, D.I. “Imitation, mirror neurons and autism”,. *Neuroscience and Biobehavioral Reviews*.

[17] Gallese, V. and Goldman, A.” Mirror neurons and the simulation theory of mind-reading”,.*Trends in Cognitive Sciences*. Vol. 2, No. 12, pp. 493-501, 1998.

[18] Woodward, A. L., Sommerville, J. A., and Guajardo, J. J. “How infants make sense of intentional action”, *Intentions and Intentionality: Foundations of Social Cognition*, pp. 149-169. 2001.

[19] S. Schaal. “ Is imitation learning the route to humanoid robots?”, *Trends in Cognitive Sciences,*Vol. 3, pp. 233-242. 1999.

[21] M. Downie. “Behavior, animation, and music: the music and movement of synthetic characters”. MIT MAS Master’s thesis, Cambridge, MA. 2000.

#	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: Sample task collaboration on single-level tasks

Appendix A: Task Collaboration Transcript

Table 1 shows a sample transcript describing a characteristic task collaboration 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 presented in this paper.