Collaboration in Human-Robot Teams

Article · September 2004

DOI: 10.2514/6.2004-6434

2 authors:

Guy Hoffman
Cornell University

Cynthia Breazeal
Massachusetts Institute of Technology

Collaboration in Human-Robot Teams

Guy Hoffman* and Cynthia Breazeal†

MIT Media Lab, 20 Ames St. E15-468, Cambridge, MA, 02139, USA

Many new applications for robots require them to work alongside people as capable members of human-robot teams. These include—in the long term—robots for homes, hospitals, and offices, but already exist in more advanced settings, such as space exploration. The work reported in this paper is part of an ongoing collaboration with NASA JSC to develop *Robonaut*, a humanoid robot envisioned to work with human astronauts on maintenance operations for space missions. To date, work with Robonaut has mainly investigated performing a joint task with a human in which the robot is being teleoperated. However, perceptive disorientation, sensory noise, and control delays make teleoperation cognitively exhausting even for a highly skilled operator. Control delays in long range teleoperation also make shoulder-to-shoulder teamwork difficult. These issues motivate our work to make robots collaborating with people more autonomous.

Our work focuses on a scenario of a human and an autonomous humanoid robot working together shoulder-to-shoulder, sharing the workspace and the objects required to complete a task. A robotic member of such a team must be able to work towards a shared goal, and be in agreement with the human as to the sequence of actions that will be required to reach that goal, as well as dynamically adjust its plan according to the human's actions. Human-robot collaboration of this nature is an important yet relatively unexplored kind of human-robot interaction.

This paper describes our work towards building a dynamic collaborative framework enabling such an interaction. We discuss our architecture and its implementation for controlling a humanoid robot, working on a task with a human partner. Our approach stems from Joint Intention Theory, which shows that for joint action to emerge, teammates must communicate to maintain a set of shared beliefs and to coordinate their actions towards the shared plan. In addition, they must demonstrate commitment to doing their own part, to the others doing theirs, to providing mutual support, and finally—to a mutual belief as to the state of the task.

We argue that to this end, the concept of task and action goals is central. We therefore present a goal-driven hierarchical task representation, and a resulting collaborative turn-taking system, implementing many of the above-mentioned requirements of a robotic teammate. Additionally, we show the implementation of relevant social skills supporting our collaborative framework.

Finally, we present a demonstration of our system for collaborative execution of a hierarchical object manipulation task by a robot-human team. Our humanoid robot is able to divide the task between the participants while taking into consideration the collaborator's actions when deciding what to do next. It is capable of asking for mutual support in the cases where it is unable to perform a certain action. To facilitate this interaction, the robot actively maintains a clear and intuitive channel of communication to synchronize goals, task states, and actions, resulting in a fluid, efficient collaboration.

I. Introduction

As robots increasingly leave the factory floor and enter human environments, it makes more and more sense to talk about the *human-robot team*, in which people and robots collaborate on tasks, sharing the same workspace and objects. In the future we envision such teams spanning many areas — from kitchens, through offices, to hospitals — but already today we see applications for humanoid robots working in collaboration

*Research Assistant, MIT Media Lab, Cambridge, MA.

†Assistant Professor, MIT Media Lab, Cambridge, MA.

with people. A great example for such an application is NASA JSC's Robonaut.[2] This humanoid robot is envisioned to work shoulder-to-shoulder with astronauts assisting them in space station maintenance operations. The work described in this paper is part of an ongoing collaboration between our research group and the NASA JSC Robonaut project.

Currently, most of the work with Robonaut has been concerned with interactions within a human-robot team, in which the robot is teleoperated. Teleoperation, however, puts a significant cognitive strain on even the most skilled operator. This is due to perceptual disorientation, sensory overload, channel noise, and control delays. In very long range remote operation, these delays can be up to minutes long, making the act of teleoperation even more strenuous. Large control delays also make the the shoulder-to-shoulder interaction between the remotely located human and the robot difficult. The above strongly motivates us to make robots working with humans more autonomous.

A. Robots as Teammates

Although a fully autonomous collaborative robot is far from being an imminent reality, we see an emerging trend of collaborative control for dynamic-autonomous robots.[6] Much of the existing work, though, still views the robot as merely an intelligent tool that a human operator commands, at times relinquishing some level of control. Work that does consider the notion of partnership[13] does so in a teleoperation setting, and views the human mostly as a reliable remote source of information. As a result, collaboration is often viewed as a control or a communications problem. We propose to approach human-robot collaboration from the standpoint of *teamwork*, implying a sense of partnership that occurs when agents work "jointly with" others rather than "acting upon" others.[16]

Such an interpretation of human-robot collaboration requires a social adeptness on the robot's part. It has to reason about our intentions, beliefs, desires, and goals so that it can perform the right actions at the appropriate time. For the human-robot team to succeed, the robot must also communicate its own set of intents and goals to establish and maintain a set of shared beliefs and to coordinate its actions to execute the shared plan.[16] In addition, each teammate must demonstrate commitment to doing their own part, commitment to the other in doing theirs, and commitment to the success of the overall task.[22,10]

B. Outline of the Paper

Section II explores the theoretical foundations that inform our collaborative architecture and related social skills. The next two sections present said architecture: Section III introduces the goal-oriented task representation, which serves as the basis for a collaboration framework described in Section IV. Section V presents an application of our system to a multi-level object manipulation task. Finally, Section VI offers a discussion, comparing the work herein to recent related work in the field of human-robot interaction.

II. Theoretical Foundations

Humans are exceptionally good at working in teams, ranging from the seemingly trivial (i.e. jointly moving a table through a doorway), to the complex (as in sports teams or corporations). What characteristics must a member of a team display to allow for shared activity? What rules govern the creation and maintenance of teamwork? And how does a group of teammates form individual intentions aimed to achieve a joint goal, resulting in a shared activity?

A. Shared Activity

Joint action can best be described as doing something as a team where the participants share the same goal and a common plan of execution. The workings of this inherently social behavior have been of increasing interest to researchers in many fields over the past decade. Grosz—among others—has pointed out that collaborative plans do not reduce to the sum of the individual plans, but consist of an interplay of actions that can only be understood as part of the joint activity.[16]

For example, if we were to move a table jointly through a doorway, your picking up one side of the table and starting to walk through the door does not make sense outside our joint activity. Even the sum of both our picking-up and moving actions would not amount to the shared activity without the existence of a collaborative plan that both of us are sharing (namely to move the table out the door). It seems that we both hold a joint intention, as well as individual intentions related to this joint intention.

The conceptual relationship between individual intentions and joint intentions is not straightforward, and several models have been proposed to explain how joint intentions relate to individual intentions and actions. Searle argues that collective intentions are not reducible to individual intentions of the agents involved, and that the individual acts exist solely in their role as part of the common goal.[28]

In Bratman's detailed analysis of *Shared Cooperative Activity* (SCA), he defines certain prerequisites for an activity to be considered shared and cooperative:[3] he stresses the importance of *mutual responsiveness*, *commitment to the joint activity* and *commitment to mutual support*. His work also introduces the idea of meshing singular sub-plans into a joint activity. In our implementation, we generalize this concept to the idea of dynamically meshing sub-plans.

The Bratman prerequisites guarantee the robustness of the joint activity under changing conditions. In the table-moving example, mutual responsiveness ensures that our movements are synchronized; a commitment to the joint activity reassures each teammate that the other will not at some point drop his side; and a commitment to mutual support deals with possible breakdowns due to one teammate's inability to perform part of the plan. Bratman shows that activities that do not display all of these prerequisites cannot necessarily be viewed as teamwork.

B. Joint Intention

Supporting Bratman's guidelines, Cohen and Levesque propose a formal approach to building artificial collaborative agents.[10] Their notion of *joint intention* is viewed not only as a persistent commitment of the team to a shared goal, but also implies a commitment on part of all its members to a mutual belief about the state of the goal. Teammates are committed to inform the team when they reach the conclusion that a goal is achievable, impossible, or irrelevant. In our table-moving example, if one team member reaches the conclusion that the table will not fit through the doorway, it is an essential part of the implicit collaborative "contract", to have an intention to make this knowledge common. In a collaboration, agents can count on the commitment of other members, first to the goal and then—if necessary—to the mutual belief of the status of the goal. For a more detailed description, see Cohen (1991).[10]

Cohen's Joint Intention Theory predicts that an efficient and robust collaboration scheme in a changing environment requires 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.

C. Common Ground

The above suggests that a central feature of any collaborative interaction is the establishment and maintenance of *common ground*, defined by Clark as

> "the sum of [...] mutual, common, or joint knowledge, beliefs, or suppositions".[9]

Common ground is necessary with respect to the objects of the task, the task state, and the internal states of the team members.

Common ground about a certain proposition p is believed by Clark to rely on a shared base b, that both suggests p and the common knowledge of b. People sharing a common ground must be mutually aware of this shared basis and assume that everyone else in the community are also aware of this basis. A shared basis can be thought of as a signal, both indicating the proposition p, and being mutually accessible to all agents involved. Clark's coins this idea the *principle of justification*, indicating that members of a shared activity need to be specifically aware of what the basis for their common ground is.[9]

This leads to the introduction of so-called *coordination devices* that serve as shared basis, establishing common ground. Coordination devices often include jointly salient events, such as gestural indications, obvious activities of one of the agents, or salient perceptual events (such as a loud scream, or a visibly flashing light).

Finally, an important type of reaching common ground inherent to teamwork is the principle of *joint closure*. Joint closure on a sub-activity is needed to advance a shared activity. Until joint closure is achieved, the sub-activity cannot be regarded as complete by a team. Joint closure is described as "participants in a joint action trying to establish the mutual belief that they have succeeded well enough for current purposes".[9]

D. Goals

A final theoretical note regards goals. Humans are biased to use an intention-based psychology to interpret other agent's actions.[12] Moreover, it has been shown in a variety of experimental settings that from an early age we interpret intentions and actions based on *intended goals* rather than specific activities or motion trajectories.[34,14,1] A person opening a door, once by using her right hand and once by using her left hand, is considered to have performed the same action, regardless of the different motion trajectories. But this person is doing a distinctly different action if she is opening a tap, even though the motion might be very similar to the one used for opening the door. To a human spectator, the goal of the action is more salient than the physical attributes of it.

We argue that a goal-centric view is particularly crucial in the context of teamwork, in which goals often provide a common ground for communication and interaction. Teammates' goals need to be understood and evaluated, and joint goals coordinated. All of this suggests that for a collaborative agent, goals and a commitment to their successful completion should be central to the representation of tasks and actions.

III. Tasks and Goals

The next two sections describe a collaborative interaction architecture aimed to enable an autonomous humanoid robot to execute complexly structured tasks in collaboration with a human teammate. It is inspired by the above theoretical framework, and relies on the concept of goal-oriented atomic actions combined into hierarchical tasks.

Each task is represented as a tree of actions and sub-tasks (recursively defined in the same fashion). Goals are present at every level of a task, so a goal is associated not only with atomic actions, but also with the successful completion of an overall task or subtask. This task goal can be distinct from the goals of each of the task's constituents. (see: Figure A).

A. Goal-Oriented Action

We represent our activities in terms of *action tuples*,[7] with the additional notion of goals. Action tuples are defined as sets of *triggers*, *executables*, *objects*, and *until-conditions* for a certain action (Figure 1). In our implementation, goals play a central role in the *triggers* and the *until-conditions* of the action, as well as in the *objects* of some actions: in most cases, an unachieved goal can trigger the execution of an action tuple, which will then be concerned with its achievement, until the goal has been achieved. Since tasks, sub-tasks, and actions are all action tuples, the above-mentioned tree structure is naturally afforded.

trigger	executable	object	until-condition

Figure 1. An action tuple

Our system currently distinguishes between two types of goals: (a) `state-change` goals that represent a change in the world as a result of the successful application of the activity, 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 a `state-change` goal by executing the action, reattempting if necessary, until the robot succeeds in bringing about the new state. This commitment is an important aspect of intentional behavior.[3,10] Conversely, a `just-do-it` goal will lead to an action regardless of the world state, and will only be performed once.

The executable of atomic actions is one of a set of predefined actions the robot can perform on the tuple's object. The executable of a task consists of its constituent actions with optional restrictions that govern the order and dependencies between the sub-actions or sub-tasks.

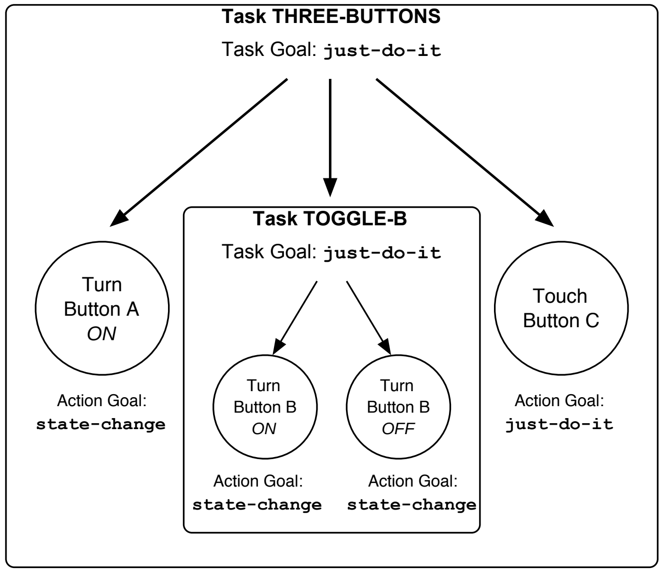

Figure 2. A hierarchical goal-oriented task

A Sample Task

Since our application is currently concerned with the manipulation of buttons, it might make sense at this point to illustrate our data structure using a sample button manipulation task. Our demonstration platform was structured to provide an analog of the Robonaut bolt fastening task, in which objects can be named, pointed to, and operated upon. Tasks can then be learned as sequential, constrained, or hierarchical structures of atomic actions.

In our scenario we consider two types of atomic actions that can be applied on objects of type button: (1) *pressing* a button, toggling its state and (2) *touching* or *pointing to* a button, which has no effect on the button state.

A hierarchical task could be made up of three steps: turning Button A on, then toggling Button B twice, and finally touching Button C. This would be represented by a top-level task including two atomic sub-actions, and one sub-task. The goal of the first action would be a `state-change` goal (resulting in Button A to be on), the goals of both the second sub-task, and the third sub-action would be `just-do-it` goals. The sub-task would—in turn—include two atomic sub-actions, each of which having a `state-change` goal, since they include pressing Button B, changing its state.

Note that the `state-change` goal of the first sub-action implies that, if Button A is already turned on, the executable is not triggered. If, on the other hand, pressing the button fails, the robot will show commitment to the goal by attempting to press Button A again. In contrast, in the case of the last sub-action, the robot is expected to touch Button C once, and only once, regardless of the initial state of the button.

B. The Task Manager

At the top of our architectural hierarchy lies the *Task Manager* module, which keeps a record of all the tasks that the robot knows and their structure. The Task Manager governs the interaction between the robot and the human, relaying control to two submodules, the *Task Learning/Execution* module, and the *Task Collaboration* module.

Initially, the Task Manager waits for a request for the human to perform a task. If this is an unknown task, or if the human requested the robot to perform a task on its own, the Task Learning/Execution module takes control. In the case of an unknown task, the robot will learn the task's structure, its constituent actions and goals. In case of a task that's already known, the robot will attempt to perform the task, while at the

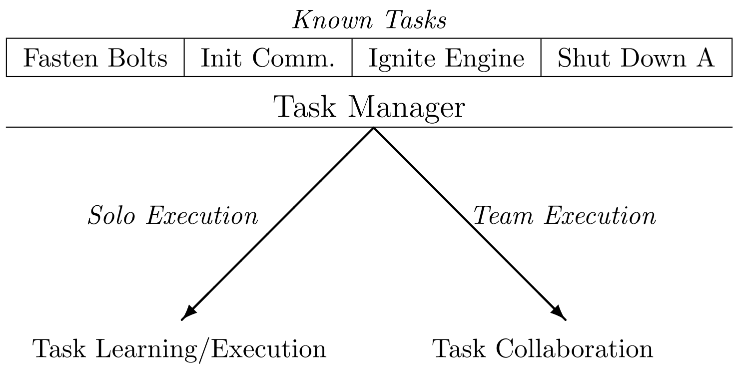

Figure 3. The task manager

same time refining its model of the task by responding to human instruction and feedback. This combined learning/execution approach allows for continuous and efficient refinement of the robot's abilities, employing an intuitive social interface to the human teacher. For a detailed discussion of our system's socially guided learning architecture, please refer to Lockerd et al.[23]

The human collaborator can ask the robot to perform a task as a team effort. In this case, the Task Manager passes the task, as learned by Task Learning/Execution module, on to the Task Collaboration module. During a collaborative interaction, the human can request demonstration of known tasks, to ground the mutual belief as to the structure of these tasks. The rest of this paper focuses on our system's Task Collaboration module.

IV. Collaborative Execution

As described in the previous section, once a task is requested as a team effort, the *Task Collaborator* is engaged to perform the hierarchical task jointly with a human teammate. This collaborative execution draws on the theoretical foundations laid out in Section II, with particular emphasis on the social skills enabling robotic teamwork.

When a robot executes a hierarchical goal-oriented task on its own, the role of the Task Execution module is to unpack the task onto an *focus stack*,[26] and perform each of the actions based on the preconditions of the action. Actions should be performed until the goal conditions of the actions have been achieved. Sub-tasks are similarly performed by recursively unpacking their constituent actions onto the stack.

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 partner'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 an action; and a clear channel of communication has to be used to establish mutual beliefs and maintain common ground (see: Section II). Our implementation of the Task Collaborator module supports these considerations as the robot progresses towards achieving the joint goal. In particular, we have focused on communication acts (utilizing gestures and facial expressions), dynamic meshing of subplans, mutual support, and turn taking.

A. The Task Collaborator

The role of the Task Collaborator module is to translate the joint intention—represented by a task data structure as described in Section III—to the robot's individual intention, which drives its actions. In a collaboration, the robot's plan should always be governed by the team's joint intention, regardless of which teammate is performing a task step at any given time.

At its core, the Task Collaborator subsystem is implemented as a state machine (Figure 4) commanding the interaction flow throughout the collaboration, and triggering the appropriate social behaviors. The collaborator module's states are defined, in rough order of a typical interaction, as follows:

- `COLLAB_NEXT`—The initial state of the system, in which the robot evaluates the currently pertinent goal, and acts upon this evaluation.

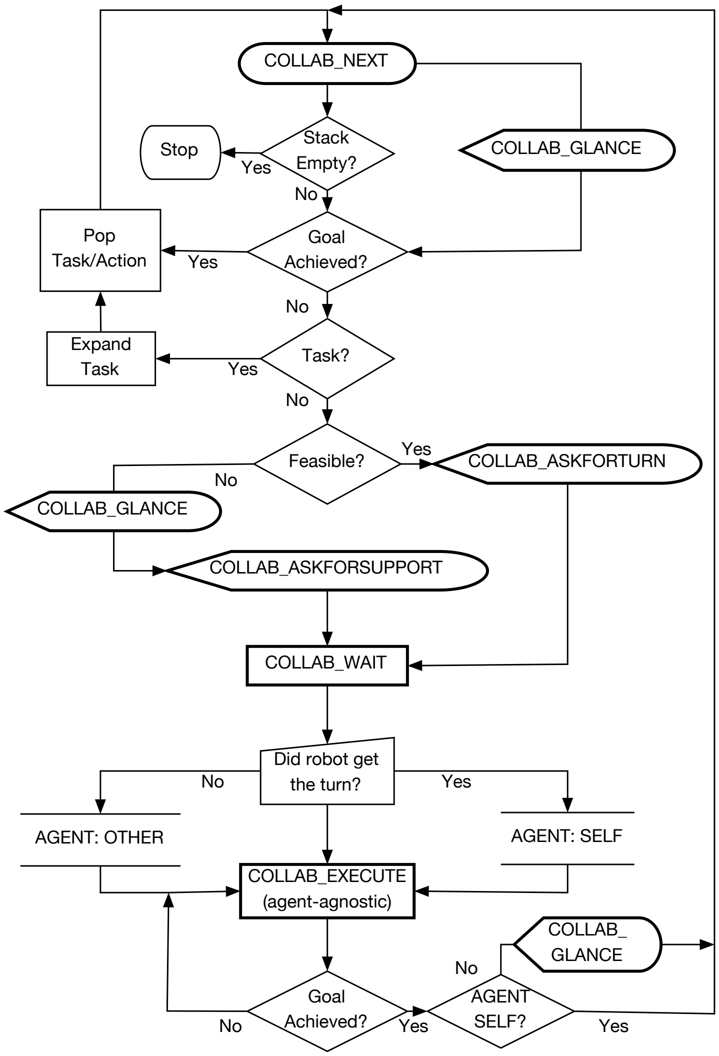

Figure 4. A schematic view of the Task Collaborator module. Note that the `COLLAB_WAIT` state can be terminated by both explicit and implicit turn taking on the human's part; and that in the agent-agnostic `COLLAB_EXECUTING` state the robot may be acting, but may also merely be following the teammate's progress on the action goal.

- `COLLAB_ASKFORTURN`—If the robot is capable of performing the next step of the task, it will offer to take its turn.
- `COLLAB_ASKFORSUPPORT`—If the robot is not capable of performing the next step of the task, it will ask for support from the teammate.
- `COLLAB_WAIT`—Waiting for a response from the other participant in the interaction.
- `COLLAB_EXECUTE`—An agent-agnostic execution step. If the robot is executing the current action, this happens in this state; if the human teammate is executing a step, the robot waits in this state. In both cases the until-condition of the current action is continually evaluated.
- `COLLAB_GLANCE`—Establishing common ground by glancing at an object of attention, both for grounding in-sequence action and for joint closure on out-of-turn action.

Our architecture is goal-based. Therefore a typical operative step begins by evaluating the goal of the currently pertinent task. If it has not been achieved, the robot decomposes the task into its constituent sub-tasks and sub-actions and recursively adds them to the focus stack. If an atomic action reaches the top of the focus stack, it is assigned to an agent in the team (as described in Section B). Currently this can be either AGENT_SELF or AGENT_OTHER, but the framework allows for any number of agents. This so-called derivation of *I*-intentions from *We*-intentions is a fundamental aspect of teamwork.[28] The robot then tracks the performance on the current task step, and dynamically adjust its plan according to the changes in goal satisfaction throughout the collaboration.

B. Turn Taking

In order to assign an agent to a task, the robot negotiates the turn for executing the current action. If the robot can perform the action, it will suggest to do it, otherwise it will ask the human to help (see Section D). Since the robot currently does not speak, it negotiates the assignment of tasks between the agents using gestures.

After a turn-taking proposal is presented to the human teammate, the robot waits for joint closure on this segment of the task dialog (the COLLAB_WAIT state). Expecting joint closure in this case is a natural social behavior, and significantly contributes to the intuitiveness and efficiency of the interaction. This waiting state can be resolved by the human by either explicitly approving the robot's turn, explicitly taking her turn, or implicitly taking a turn by proceeding on the task. After a certain waiting time, no response from the human is understood as implicit agreement to the robot's turn proposal. A possible alternative would be to trigger a repeated offer to take the turn in this case, although in the present framework, we believe it to be a more reasonable behavior to proceed and be prepared to accept intervention if required.

After a turn taking interaction has been completed, an agent is assigned to the current action. The robot follows the overall team intent by tracking the goal of the currently executed action, regardless of the agent of the action (by being in the COLLAB_EXECUTE state, where EXECUTE refers to the team's execution of the current step).

C. Dynamic Meshing of Subplans

While a task step is being performed, the collaboration module keeps track of the state of the current goal. To maintain a mutual belief about the state of the goal, the robot communicates with the human about the state of the overall task goal (through nods and gaze), once the step is completed.

Tracking goals rather than specific actions allows for a dynamic meshing of the teammates' subplans as part of the overall plan[a]. As opposed to following a preset plan based on a specific actions, tracking the goals makes the interaction robust to arbitrary decisions on part of the human teammate, as well as to unexpected changes in the world state. These changes can happen due to external factors or because of a failure on the robot's part. Dynamically adjusting to changes in goal states, the robot can even keep track of simultaneous actions, in which the human performs an action while the robot is working on another part of the task. If this is the case, the robot will take the human's contribution into account and reevaluate the goal state of the current task or task step. The robot then might decide to no longer keep this part of the task on its list of things to do. Particularly in this case, it is crucial for the robot to communicate an acknowledgment of this change of belief to the human in order to maintain mutual belief about the overall task state. In our implementation this is done using gaze and nods.

D. Self Assessment and Mutual Support

In a collaboration, individual team members' capabilities may vary, and teammates must be able to rely on *mutual support*, since they ultimately share a common goal.[3] In our implementation, the robot is aware of its own abilities. This is enforced by the fact that each atomic action must provide an interface to query its feasibility. If the robot is able to complete the task element it will offer to do so. But whenever it believes that it cannot do the action, the robot will ask the human for help. Assuming the human can perform this action, this results in assigning the current action to the human, and putting the robot in the COLLAB_EXECUTING state, evaluating the human's performance by tracking the task's step goal.

[a] The term *dynamic meshing of subplans* is inspired from Bratman's notion, that in a shared cooperative activity the team members' individual plans must not be the same, but they must mesh—i.e. not conflict—to create the joint course of action.[3] A collaborative robot must be able to dynamically adjust its plans to mesh with the human's changing course of action and the changes in the world state.

Since the robot currently does not speak, this, too, is communicated using gestures. In order to complete the request for mutual support, the robot uses gaze to establish a joint object of attention to which this support should be applied.

E. Satisfying the Teamwork Requirements

In this section we will revisit the main requirements of teamwork as laid out in Section II, and evaluate the manner in which they are addressed by the system proposed in this paper. For these purposes, we use the term *persistent goal* as defined in Cohen and Levesque,[10]

> An agent is defined to hold a *persistent goal* to achieve p relative to q when
>
> 1. she believes that p is currently false.
> 2. she wants p to be true eventually.
> 3. 2. will continue to hold until she comes to believe that p is true or will never be true or that q is false.[10]

1. *Individual Plan as part of a Joint Plan*

In a teamwork setting, individual plans should be understood as part of the joint intent,[28] and cannot be modeled as disparate entities, even in coordination with other team members' individual plans. Joint Intention Theory informs us that

> "If a team has a joint persistent goal to achieve p, then each member has p as an individual persistent goal."[10]

Our system corresponds to this approach, as the robot is continually tracking the team's joint plan (in the form of the currently executed team task), and derives its own intentions based on the common plan, the other teammate's actions and changes in the world that apply to the plan's goals.

2. *Commitment to Partner's Action*

In a team, one is not only committed to one's own actions, but also to the appropriate actions of the other team members. Here, *commitment* is understood in the sense of having a persistent goal towards a certain state. Using this definition, according to Cohen and Levesque:

> "[...] an agent x can be committed to another agent y's acting. Just as with committing to her own actions, x would—in that case—not adopt other goals inconsistent with y's acting, would monitor y's success, might request y to do it or help y if need be."[11]

The collaboration system proposed herein shows this kind of persistent intent towards other teammate's actions. After assigning `AGENT_OTHER` to a certain task step (based on the robot's or the human's request), the robot will monitor the teammate's actions with respect to the common plan. This goal evaluation is identical to the robot's evaluation of its own success in achieving a task step's goal, maintaining a parallel between the robot's commitment to its own actions—as in any goal-oriented planner—to the robot's commitment to the teammate's actions, as required in a collaborative goal-oriented planner.

3. *Commitment to Mutual Belief*

One of the defining properties of a joint intention in a team is that team members share not only a commitment to the mutual goal, but—if necessary—a commitment to the mutual belief as to the state of the goal. If one participant reaches the conclusion that the common goal (or a subgoal) is achieved, unachievable or irrelevant, it becomes this participant's goal to inform the other team members of this conclusion.

Our system employs gestures and nods to demarcate task and sub-task boundaries, thus maintaining a continuous mutual belief as to the advancement of the task. This ensures the transparency needed for the human teammate to efficiently collaborate with the robot even on a structurally complex task, and has been found in anecdotal experiments to significantly enhance the collaborative flow. Gaze is used to establish a mutual belief on the object of attention, both in case of the commencement of a new task step, and upon the successful termination of a task step. Nods are also used to indicate that teammates should advance to the next step.

4. *Commitment to Mutual Support*

Since team members share a common goal, and their capabilities may vary, commitment to mutual support is a fundamental requirement of any shared cooperative activity. As stated by Bratman,

> "In [a Shared Cooperative Activity] each agent is committed to supporting the efforts of the other to play her role in the joint activity [...] there must be at least some cooperatively relevant circumstances in which [the participant] would be prepared to provide the necessary help."[3]

In our architecture, the robot is aware of its limitations and will request support from its teammate under the assumption of a shared plan. The robot acts proactively and will assume responsibility for any action it is capable of doing, unless specifically told otherwise. As of now, the robot is not able to detect cases in which the human collaborator is in need of support. This ambitious challenge is left for future work, with preliminary results reported as part of a related research project in our laboratory.[15]

5. *Grounding through Gaze and Demonstration*

In coordinating a joint activity, grounding is key. Not only need there be mutual knowledge on objects and states, but Clark's *principle of justification* also states that

> "[...] people take a proposition to be common ground [...] only when they believe they have a shared basis for that proposition."[9]

Grounding is established in our collaborative framework on various levels. When the human teammate refers to an object, the robot will glance towards the object of attention, grounding this dialog segment through gaze. This is particularly crucial during the preliminary part of the collaboration, in which the human can name the objects around the robot using an intuitive gesture and voice interface ("This is the START button"). In this interaction gaze plays an important role to make the naming process efficient and reliable. Errors in gesture recognition are preemptively and intuitively detected by the human, as the robot's gaze will follow its understanding of the object to which the human points.

On a higher level, grounding on task structure and object understanding can be established through demonstration. During the interaction, known tasks and sub-tasks can be requested of the robot ("Can you fasten the bolts?", "Show me button A"), grounding the team's joint understanding of both processes and objects of operation.

6. *Joint Closure*

As discussed in Section II-C, *joint closure* is defined as

> "participants in a joint action trying to establish the mutual belief that they have succeeded well enough for current purposes".[9]

Joint Closure occurs on two levels, as well. When negotiating turns, the robot expects—as a human would—joint closure on the specific dialog segment, waiting for the human's approval of the agent assignment or, conversely, for the human teammate's adoption of her turn. On the task level, the robot communicates completion of task steps (see Subsection IV-E-3) to provide the human with joint closure regarding the progress of the current task segment.

V. Application

We have applied the framework described in the previous sections to our expressive humanoid robot, Leonardo ("Leo"), shown in Figure 5. The reader will at this point have noticed that some of the social skills presented in this paper make an implicit assumption as to the physical form of the robot. Although not prescribing a single morphology, our approach does assume the robot's ability to direct gaze at its collaborators, as well as at the objects of attention. In addition, the robot must be able to display gestural cues that are human-readable. These gestures can in some cases be replaced by speech, but there is also reason to believe that the use of physical cues, such as subtle nods, body pose, or hand gestures will be more efficiently parsed by human teammates in a rapidly advancing collaborative task setting.

Figure 5. "Leonardo", the expressive humanoid robot used in this study. The left picture shows the robotic structure, the center picture shows the robot when cosmetically finished, the right shows a simulated version of the robot.

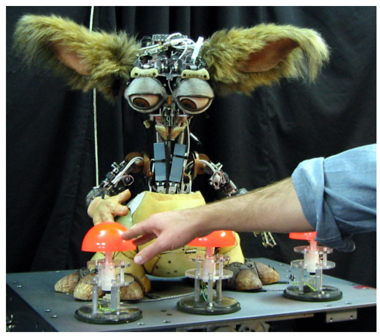

Figure 6. The button pressing task setup.

Leonardo is a 65-degree of freedom (DoF) fully embodied humanoid robot that stands approximately 2.5 feet tall. It is designed in collaboration with Stan Winston Studio to be able to express and gesture to people as well as to physically manipulate objects. The robot is equipped with two 6-DoF arms, two 3-DoF hands, an expressive (24-DoF) face capable of near human-level expression, two actively steerable 3-DoF ears, a 4-DoF neck, with the remainder of the DoFs in the shoulders, waist, and hips.

The robot's perceptual system spans a speech understanding subsystem, based on Sun Microsystem's SPHINX-4 architecture, and two stereo camera systems for detecting humans, gestures and objects in Leo's vicinity. Leo understands a limited grammar tailored for the collaboration task at hand. The robot is additionally equipped with cameras in its eyes, as well as with tactile sensors in its hands, neither of which were used in the experiments described below. For a full description of Leo's perceptual architecture please turn to Breazeal et al.[4]

A. Experimental Setup

In the experimental scenario there are three buttons in front of Leonardo (see: Figure 6). The buttons can be switched ON and OFF (which lights the button up). 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. This rudimentary learning of names is done by pointing to a button, waiting for the robot's confirmation of the button's location (by glancing towards the detected button), and then naming it ("This is the Start button"). Leo can confirm his knowledge of button names by pressing them or pointing to them per request ("Leo, can you show me Button A?").

To test our collaborative task execution implementation, we designed a set of tasks involving a number of sequenced and layered 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 other similar tasks. The task set used represents simple and complex hierarchies and contains tasks with both `state-change` and `just-do-it` goals.

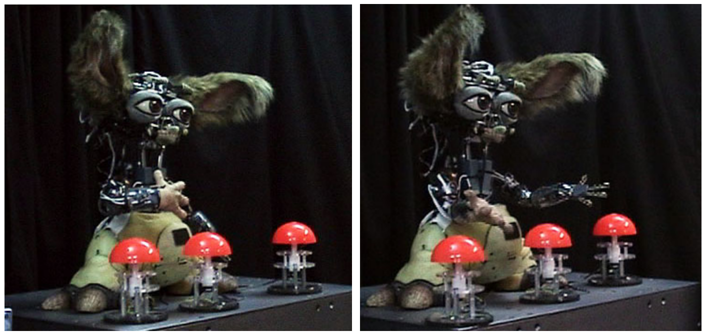

Figure 7. Left: Leo negotiates his turn by gesturing towards himself. Right: Leo asking for help by gesturing towards the human

B. A Sample Interaction

In a typical interaction, Leo displays many of the social skills discussed in the previous sections. To initiate a collaboration, the human proposes a task to the robot: "Let us do the `BUTTONS-ON` Task". If the robot does not know this task he will shrug and cock his head in confusion. If the task is known to the robot, he will glance at the button setup to communicate common ground as to the starting condition of the task.

At every stage of the interaction, either the human should do her part, or the robot his. If Leo can do the next step of the task, he will negotiate this as described in Section IV-B. Since Leonardo does not have speaking capabilities, he indicates his willingness to perform an action by pointing to himself, and adopting an alert posture and facial expression. Similarly, when detecting an inability to perform an action assigned to him, Leo's expression displays helplessness, as he gestures toward the human in a request for her to perform the intended action (Figure 7). Leo also shifts gaze between the problematic button and his partner to direct her attention to what it is he needs help with. 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.

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. The robot needs to then communicate this knowledge to the human in order to maintain mutual belief about the simultaneous action as well as the overall task state. This is once more achieved by using short glances and nods.

Table 1 shows a sample transcript describing a typical interaction between Leonardo and a human teammate. We chose to display the following two-level task: `BUTTONS-ON-AND-OFF`, which is comprised of two sub-tasks, `BUTTONS-ON`, turning Button 1, 2, and 3 *ON* in this order; and `BUTTONS-OFF`, turning Buttons 1, 2, and 3 *OFF* in the same order. This transcript offers a sense of both the goal-driven action and the joint intention related communicative skills fundamental to the collaborative discourse stressed in this paper.

C. Communicative Cues for Task Dialogs

In analysis of human task dialogs, a number of communication acts have been identified,[10] serving to achieve robust team behavior despite adverse conditions, including breaks in communication and other difficulties in achieving the team goals. We argue that similar acts should be used by robots who work in human-robot teams, and have implemented a number of conversational policies enabling the robot to keep a clear channel of communication with the human teammate. Each of these communicative cues conveys an intention, which in turn serves an interaction function, advancing the human-robot collaboration. Naturally, some of these cues are context-dependent. This section presents them in detail.

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

#	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 BUTTONS-ON-AND-OFF"	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 part of sub-task BUTTONS-ON, and suggests doing so.
4	"I can go"	Looks at his partner	
5	Presses Button 1	Looks at Button 1; looks back at partner	Acknowledges partner's action; creates mutual belief on object of attention
6		Points to himself	Leo can do the second part of sub-task BUTTONS-ON, and suggests doing so.
7	"OK, you go"	Presses Button 2, looking at it	Joint closure on the turn taking contribution is achieved; Looking away from the partner while operating establishes turn taking boundaries.
8		Looks back at his partner	Gaze shift signals end of turn
9		Points to himself	He can do the third part of sub-task BUTTONS-ON, and suggests doing so.
10	"I'll go"	Looks at his partner	
11	Presses Button 3	Looks at Button 3	Acknowledges partner's action; creates mutual belief on object of attention
12		Looks back at partner; Nods shortly	Nod creates joint closure in form of a mutual belief as to the sub-task's completion.
13	Moves Button 1 out of Leo's reach		
14		Looks at Button 1, then back at the human partner; extends his arms in "Help me" gesture.	Leo assesses his capabilities and consequently requests support.
15	Presses Button 1	Looks at Button 1; looks back at human	Glance acknowledges partner's action
16		Points to himself	Leo can do the second part of sub-task BUTTONS-OFF, and suggests doing so.
17	"You go"	Presses Button 2, looking at it	Joint closure on the turn taking contribution is achieved; Looking away from the partner while operating establishes turn taking boundaries.
18	At the same time as 17, presses Button 3		
19		Looks back at his partner	Gaze shift signals end of turn
20		Looks at Button 3; looks back at the human	Acknowledges partner's simultaneous action; creates mutual belief on object of attention
21		Nods shortly	Creates joint closure in form of a mutual belief as to the sub-task's completion.

Table 1. Sample task collaboration on a hierarchical goal-driven task.

post-action 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 meets a task goal, such as turning the last button *ON* during the `BUTTONS-ON` task, Leo will glance at the change and give a small confirming nod to the human. Similarly, Leo uses subtle nods when he thinks he completed a task or sub-task. For instance, Leo will give an acknowledgement nod to the human after completing the `BUTTONS-ON` sub-task and before starting the `BUTTONS-OFF` sub-task, in the case of the `BUTTONS-ON-AND-OFF` task. These cues play a crucial role in establishing and maintaining mutual beliefs between the teammates on the progress of the shared plan. Since Leonardo does not currently speak, all of these communicative acts are in the form of gestures. Table 2 provides a summary of Leo's gestural cues used in the application described above.

Conversational Cue	Communicated Intention	Interaction Function
Follows gesture to Object of Attention (OOA)	Establish OOA common ground	OOA set and ready for labeling
Point to object, look to object	Identify a particular object as referential focus (e.g., demonstrate correct association of name with object).	Confirm mutual belief about a particular object referent (e.g., successful identification of the target)
Confirming Nod (short)	Confirmation (e.g., "OK, got it")	Update common ground of task state (e.g., attach label, start learning, etc.)
Affirming Nod (long)	Affirm query (e.g., "Yes, I can")	Affirmation to query
Breaking eye-contact	Executing an action / checking goal	Claiming the turn, pacing human action
Alert expression and gesture	Showing willingness to perform an action	Asking for turn in the joint activity
Leaning forward and raising one ear towards human	Cannot understand (unable to recognize/parse speech)	Cues the human to repeat what was last said
Cocking head and shrugging (expressing confusion)	Cannot perform the request (lack of understanding)	Cues the human to add information or rectify shared beliefs (request clarification or elaboration)
Shake head	Cannot perform the request (lack of ability)	Cues that robot is not able to perform the request

Table 2. Robot's gestures and expressions to support transparent communication of robot's internal state to human.

D. 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.

We are currently underway performing untrained user studies to further investigate the role and effectiveness of the social cues described in this section. Preliminary results indicate significant improvement in perceived transparency of the robot's internal state, some improvement in task efficiency, and anecdotal increase in user politeness as a result of nonverbal social cues.[19]

VI. Related Work

Human-robot collaboration, in the sense that we have tried to address in this paper, is still a relatively unexplored field of study. We distinguish human-robot teamwork from other kinds of human-robot interaction, as those usually view the problem in terms of robot control or human-robot communication with teleoperated or partially-autonomous robots. This does not capture the essence of teamwork as it occurs between human teammates, a social structure which we argue should be the model for collaborative autonomous robots in human-robot teams as well.

Jones and Rock[17] correctly stress the importance of dialog in human-robot communication, but present a two-way communication system that is more concerned with the control of remote robots than with the social aspect of collocated teamwork. Perzanowski et al.,[25] too, model the challenge of human-robot collaboration as a communication protocol. They employ natural language and multi-modal communication, but view the relationship between the human and the robot in a strict master-slave arrangement, that does not conform to the sense of partnership that we mean when we speak of working "jointly with" others.

Collocated human robot collaboration has been studied using autonomous vision-based robotic arms, e.g. Kimura et al.[20] While addressing many of the task representation and labor 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, and does not enable deviations from the plan or dynamic adjustment of individual sub-plans.

Fong et al. consider a working partnership between a human and a robot in terms of *collaborative control*, where human and robot collaborate in vehicle teleoperation.[13] 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 plan issues or perceptual ambiguities. The role of the human in the partnership is not that of a peer working towards a shared goal, but akin to a reliable remote source of information. A similar approach has been taken by Woern and Laengle.[33] More recent work has been evaluating a shoulder-to-shoulder scenario, but still looks at the problem from a purely control or communication oriented standpoint, without regard for the social aspects of teamwork. In contrast, our work explores collaboration where the human and robot work together on a collocated task, in which both the human and the robot can complete steps of the plan, and make use of social communication.

Some work in the field of human and virtual agent teams has the notion of shared plans that must be continually maintained and updated according to changes in the world state. For instance, Traum et al. propose a system in which a human is part of a team of agents that work together in a virtual world.[32] Their architecture addresses plan reassessment and uses dialog models and speech acts to negotiate a plan as a team. Roles are attached to various steps of the plan, and an authority structure helps in negotiating control. Our work differs in two respects from this virtual teamwork scenario. First, in our physically embodied scenario, we explore the issues of face-to-face gestures and socially relevant nonverbal communication acts that facilitate collaboration. Second, we do not utilize an authority structure; instead, the robot and the human negotiate turns in the context of a shared plan.

Employing social cues for dialog and collaboration has been more widely investigated in the field of embodied conversational agents (ECA). Face-to-face multi-modal communication including gesture, gaze, and head pose has been explored in tutorial systems, e.g. by Rickel and Johnson[27] and in embodied dialog systems, e.g. Thórisson.[30] Their work focuses on the agent's display of social cues in a task-oriented virtual character. On the other side of the spectrum, Nakano et al. have studied human means of face-to-face grounding and implemented their findings in the design of an ECA for a collaborative dialog system. Their agent detects and analyzes head pose and gaze of its human conversation partner and offers appropriate elaboration when needed.[24] By the very nature of virtual agents, the tasks in both cases have been primarily informational, and could therefore not capture the physical aspects of shoulder-to-shoulder collaboration between a human and a robot, in particular with regard to object manipulation. Attempts to transfer these important concepts to the realm of robotic agents have so far been rare, rudimentary, and also often information-centric.[21,29]

We see a shifting trend, however, as the need to relinquish control and to allow robots an increased level of autonomy is of growing interest to the robotics community. This has been motivated by both the difficulties of teleoperation over large distances, and a demand for a better robot-to-operator ratio.[18] Recent robot control architectures have begun to combine fully slaved teleoperation with cooperative control, in which the robot and the human divide different aspects of the robot's operation.[13,6] A central goal in this body of work is to be able to place increasing trust in the robots we work with.[5,17] To this end—we argue—the robot must display many of the behaviors one expects from members of a team.

As a whole, one can view the previous work in this field as falling in one of two categories: in one, the robot is viewed as a tool used towards a human's task goal. In the other, the human is merely an accessory in the robot's toolbox. Our perspective is that of a balanced partnership where the human and robot work together on shared task goals. We have thus proposed a different notion of partnership than the one that has 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. Our work also deviates from the main body of publication in that we believe that developing robots with *social skills and understanding* is a critical step towards this goal. 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, as well as communicate their own internal states clearly within the context of a shared interaction.

VII. Conclusion

This paper presents an approach, an architecture, and an implementation aimed at creating robots that can work as capable members in human-robot teams. We are motivated by future applications in which autonomous and semi-autonomous robots work shoulder-to-shoulder with humans, sharing their workspace and tools. The field of robotic maintenance on space missions, currently employing predominantly teleoperated robots, is already in a position to benefit from such robots. Our approach is informed by theories of human teamwork, principally by Joint Intention Theory. Since teamwork is fundamentally a dialog, we also look for insights in Dialog Theory, such as Clark's principles of grounding and joint closure. This body of theoretical work predicts that a capable member of a human-robot team must be able to display an array of social skills.

We therefore present a task encoding that affords social interaction, being goal-oriented on multiple nested levels. We describe a collaborative framework based on this encoding and geared towards robot teamwork, including self-assessment for mutual support, communication to support joint activity, performing dynamic meshing of sub-plans, and negotiating labor division via turn taking. Finally, we present an initial application of this framework in which our humanoid robot is working together on a structured task with a human teammate.

Looking at the road ahead, we believe that thinking about robots in social terms will have profound implications for how we will be able to engage robots in the future—beyond making them appealing, entertaining, or providing an easy interface to their operation. Rather, it is a critical competence that will allow robots to assist us as capable partners.

Obviously, the system described herein merely amounts to initial steps in the direction advocated. We have only begun to explore how the various social aspects of human teamwork can be applied to autonomous robots working with people. As future work, we would like to improve the complexity of the task representation as well as that of the interaction and dialog. Leonardo can understand a few spoken requests of the human, but 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 our task representation so that the robot can represent a larger class of collaborative tasks and more involved constraints between the tasks' action components.

Eventually, a collaborative robot will also have to be able to detect when the human teammate needs assistance, and how to provide appropriate help. While still ambitious, we believe that it is not too early to address this challenge in a limited form. Goal inference and perspective taking will probably prove to be two crucial elements in this endeavor. Therefore, initial work on goal inference for human-robot collaboration is being conducted in our laboratory.[15] We are also collaborating with Alan Schultz and his group at the Naval Research Laboratory on the important task of perspective taking and spatial reasoning, and their applications for collocated collaborative robots.[8,31]

Finally, we are in the process of conducting untrained user studies to evaluate the role and importance of the various social skills proposed in this paper, with promising preliminary results.[19]

Acknowledgments

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. Particular thanks go out to Andrea Lockerd and Jesse Gray for their involvement in this project. Leonardo's speech understanding abilities are developed in collaboration with Sun Microsystems and their `SPHINX-4` system. Stan Winston Studio provided the physical Leonardo robot. Geoff Beatty and Ryan Kavanaugh provided the virtual model and animations. Leonardo's architecture is built on top of the `C5M` code base of the Synthetic Character Group at the MIT Media Lab, directed by Bruce Blumberg. 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]D.A. Baldwin and J.A. Baird. Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5(4):171–178, 2001.

[2]W. Bluethmann, R. Ambrose, M. Diftler, S. Askew, E. Huber, M. Goza, F. Rehnmark, C. Lovchik, and D. Magruder. Robonaut: A robot designed to work with humans in space. *Auton. Robots*, 14(2-3):179–197, 2003.

[3]M. Bratman. Shared cooperative activity. *The Philosophical Review*, 101(2):327–341, 1992.

[4]C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Chilongo. Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics*, 1(2):315–348, 2004.

[5]D. Bruemmer, D. A. Few, M. Goodrich, D. A. Norman, N. Sarkar, J. Scholtz, B. Smart, M. L. Swinson, and H. A. Yanco. How to trust robots further than we can throw them. In *CHI Extended Abstracts*, pages 1576–1577, Vienna, Austria, April 2004.

[6]D. J. Bruemmer, D. D. Dudenhoeffer, and J. Marble. Dynamic autonomy for urban search and rescue. In *2002 AAAI Mobile Robot Workshop*, Edmonton, Canada, August 2002.

[7]R. Burke, D. Isla, M. Downie, Y. Ivanov, and B. Blumberg. CreatureSmarts: The art and architecture of a virtual brain. In *Proceedings of the Game Developers Conference*, pages 147–166, 2001.

[8]N. L. Cassimatis, J. Trafton, A. Schultz, and M. Bugajska. Integrating cognition, perception and action through mental simulation in robots. In *Proceedings of the 2004 AAAI Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems (to appear)*, 2004.

[9]H. H. Clark. *Using Language*. Cambridge University Press, Cambridge, UK, 1996.

[10]P. R. Cohen and H. J. Levesque. Teamwork. *NOÚS*, 35:487–512, 1991.

[11]P. R. Cohen and H. J. Levesque. On team formation. In J. Hintikka and R. Tuomela, editors, *Contemporary Action Theory*. Synthese, 1997.

[12]D. C. Dennett. Three kinds of intentional psychology. In *The Intentional Stance*, chapter 3. MIT Press, Cambridge, MA, 1987.

[13]T. W. Fong, C. Thorpe, and C. Baur. Collaboration, dialogue, and human-robot interaction. In *procofthe10th International Symposium of Robotics Research, Lorne, Victoria, Australia*, London, November 2001. Springer-Verlag.

[14]B. Gleissner, A. N. Meltzoff, and H. Bekkering. Children's coding of human action: cognitive factors influencing imitation in 3-year-olds. *Developmental Science*, 3(4):405–414, 2000.

[15]J. Gray. Goal and action inference for helpful robots using self as simulator. Master's thesis, Massachusetts Institute of Technology, August 2004.

[16]B. J. Grosz. Collaborative systems. *AI Magazine*, 17(2):67–85, 1996.

[17]H. Jones and S. Rock. Dialogue-based human-robot interaction for space construction teams. In *IEEE Aerospace Conference Proceedings*, volume 7, pages 3645–3653, 2002.

[18]D. R. Olsen Jr. and S. B. Wood. Fan-out: measuring human control of multiple robots. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI-04)*, pages 231–238, Vienna, Austria, april 2004.

[19]C. D. Kidd, A. Lockerd, G. Hoffman, M. Berlin, and C. Breazeal. Effects of nonverbal communication in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) in review*. IEEE, April 2005.

[20]H. Kimura, T. Horiuchi, and K. Ikeuchi. Task-model based human robot cooperation using vision. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS'99)*, pages 701–706, 1999.

[21]Hideaki Kuzuoka, Keiichi Yamazaki, Akiko Yamazaki, Jun'ichi Kosaka, Yasuko Suga, and Christian Heath. Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability. In *Proceedings of the 2004 conference on Human factors in computing systems*, pages 183–190. ACM Press, 2004.

[22]H. J. Levesque, P. R. Cohen, and J. H. T. Nunes. On acting together. In *Proceedings of AAAI-90*, pages 94–99, Boston, MA, 1990.

[23]A. Lockerd and C. Breazeal. Tutelage and socially guided robot learning. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS'04)*, Sendai, Japan, September 2004.

[24]Y. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a model of face-to-face grounding. In *Proceedings of the Association for Computational Linguistics*, Sapporo, Japan, July 2003.

[25]D. Perzanowski, A. C. Schultz, W. Adams, E. Marsh, and M. Bugajska. Building a multimodal human-robot interface. *IEEE Intelligent Systems*, 16(1):16–21, 2001.

[26]C. Rich, C. L. Sidner, and N. Lesh. Collagen: Applying collaborative discourse theory to human-computer collaboration. *AI Magazine*, 22(4):15–25, 2001.

[27]J. Rickel and W. L. Johnson. Task-oriented collaboration with embodied agents in virtual worlds. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied conversational agents*, pages 95–122. MIT Press, 2000.

[28]J. R. Searle. Collective intentions and actions. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, chapter 19, pages 401–416. MIT Press, Camridge, MA, 1990.

[29]W. D. Smart, C. M. Grimm, M. Doxin, and Z. Byers. (not) interacting with a robot photographer. In *Proceedings of the AAAI Spring Symposium on Human Interaction with Autonomous Systems in Complex Environments*, Stanford, CA, March 2003.

[30]K. R. Thórisson. Gandalf: an embodied humanoid capable of real-time multimodal dialogue with people. In *Proceedings of the first international conference on Autonomous agents*, pages 536–537. ACM Press, 1997.

[31]J.G. Trafton, A.C. Schultz, N. L. Cassimatis, L. Hiatt, D. Perzanowski, D. P. Brock, M. Bugajska, , and W Adams. Using similar representations to improve human-robot interaction. In *Agents and Architectures (to appear)*, 2004.

[32]D. Traum, J. Rickel, J. Gratch, and S. Marsella. Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *2nd Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems*, pages 441–448, Melbourne, Australia, 2003.

[33]H. Woern and T. Laengle. Cooperation between human beings and robot systems in an industrial environment. In *Proceedings of the Mechatronics and Robotics*, volume 1, pages 156–165, 2000.

[34]A. L. Woodward, J. A. Sommerville, and J. J. Guajardo. How infants make sense of intentional actions. In B. Malle, L. Moses, and D. Baldwin, editors, *Intentions and Intentionality: Foundations of Social Cognition*, chapter 7, pages 149–169. MIT Press, Cambridge, MA, 2001.