Teaching and Working with Robots as a Collaboration

Cynthia Breazeal, Guy Hoffman, Andrea Lockerd
MIT Media Laboratory
77 Massachusetts Ave, NE18-5FL
Cambridge, MA 02139
{cynthiab, guy, alockerd}@media.mit.edu

Abstract

New applications for autonomous robots bring them into the human environment where they are to serve as helpful assistants to untrained users in the home or office, or work as capable members of human-robot teams for security, military, and space efforts. These applications require robots to be able to quickly learn how to perform new tasks from natural human instruction, and to perform tasks collaboratively with human teammates.

Using joint intention theory as our theoretical framework, our approach integrates learning and collaboration through a goal based task structure. Specifically, we use collaborative discourse with accompanying gestures and social cues to teach a humanoid robot a structurally complex task. Having learned the representation for the task, the robot then performs it shoulder-to-shoulder with a human partner, using social communication acts to dynamically mesh its plans with those of its partner, according to the relative capabilities of the human and the robot.

1. Introduction

Robots will inevitably become a part of our daily lives. Our research concerns how people will expect and want to interact with them. For example, efforts are underway in research labs around the world to put robots into homes assisting the elderly (e.g., [19]) and into space working in robot-astronaut teams [2]. As robots move into our natural environment, it is easy to envision situations that afford the need for efficient task learning and collaboration. Consider working together with a robot on a maintaining a flower garden, fixing a car, or cooking a large dinner. In each of these scenarios, one would neither want to wholly relinquish control of the process nor use the robot as a simple-minded tool that needs to be guided each step of the way. The robot should rather act as a partner that can be taught a complex goal-oriented procedure and then effectively collaborate with the human providing appropriate assistance in performing the learned task.

If the robot does not already know how to do a given task, a person must be able to teach the robot in a natural and intuitive manner. The robot, in turn, must be able to quickly learn the new skill from the human from only a few trials (in dramatic contrast to many statistical learning approaches that require hundreds or thousands of trials). Furthermore, once a new skill is learned, the robot should then be competent in its ability to provide assistance; understanding how to perform the task independently as well as how to perform it in partnership with a human.

Our aim is to be able to teach a robot a structurally complex task that can later be performed collaboratively with a human. Ideally, such robots will be as fast and as easy to work with, communicate with, and teach as a person. This paper details our work towards supplying our expressive humanoid robot, Leonardo, with human-centered learning and collaborative abilities.

2. Theoretical Framework

In considering what characteristics a robot must have to effectively work with a human partner, we view both the problem of learning and that of collaborating in terms of dialog and joint intention theory.

2.1. Learning and Joint Intention Theory

Human-style tutelage is a social and a collaborative process [10, 25] and usually takes the form of a dialog, a fundamentally cooperative activity [11]. To be a good instructor, one must maintain an accurate mental model of the learner's state (e.g., what is understood so far, what remains confusing or unknown) in order to appropriately structure the learning task with timely feedback and guidance. The learner (robot or otherwise) helps the instructor by expressing their internal state via communicative acts (e.g., expressions, gestures, or vocalizations that reveal understanding,

confusion, attention, etc.). Through reciprocal and tightly coupled interaction, the learner and instructor cooperate to help both the instructor to maintain a good mental model of the learner, and the learner to leverage from instruction to build the appropriate models, representations, and associations.

Cohen et al. analyzed task dialogs, where an expert instructs a novice assembling a physical device, and found that much of task dialog can be viewed in terms of joint intentions. Their study identified key discourse functions including: organizational markers that synchronize the start of new joint actions ("now," "next," etc.), elaborations and clarifications for when the expert believes the apprentice does not understand, and confirmations establishing the mutual belief that a step was accomplished [6].

We have given our robot a number of social and expressive skills that contribute to the robot's effectiveness in learning through collaborative discussion. For example, joint attention is established both on the object level and on the task structure level. The robot uses subtle expressions to indicate to the human tutor when he is ready to learn something new, and his performance of taught actions provides the tutor with immediate feedback about comprehension of the task. Envelope displays such as gaze aversion, eye contact and subtle nods are used to segment a complex task learning structure in a natural way for the tutor. Additionally, sequencing keywords, like those mentioned above, provide structure to the task.

2.2. Collaboration and Joint Intention Theory

Joint intention theory also motivates our approach to task collaboration. In any collaboration, agents work together as a team to solve a common problem. Team members share a goal and a common plan of execution. This *collaborative plan* does not reduce to the sum of the individual plans [12], but is an interplay of plans inspired and affected by a joint intention.

Several models have been proposed to explain how joint intention relates to individual intention. Searle argues that collective intentions 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 [23]. Bratman's analysis of Shared Cooperative Activity (SCA) introduces the idea of meshing singular sub-plans into a joint activity [4]. In this work, we generalize this concept to a process of dynamically meshing sub-plans.

Bratman also defines certain prerequisites for an activity to be considered shared and cooperative: he stresses the importance of *mutual responsiveness, commitment to the joint activity* and *commitment to mutual support*. Cohen et al. support these guidelines and provide the notion of *joint stepwise execution* [5, 17]. Their theory also predicts that an
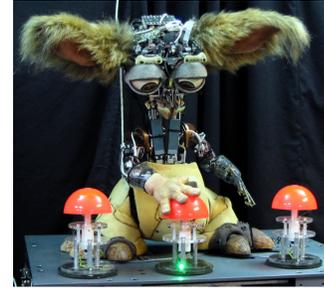

Figure 1. Leonardo performs the steps as he learns them providing the human tutor with valuable error-correcting insight in real time.

efficient and robust collaboration scheme in a changing environment commands an open channel of *communication*. Sharing information through communication acts is critical given that each teammate often has only partial knowledge relevant to solving the problem, different capabilities, and possibly diverging beliefs about the state of the task.

3. Experimental Platform

The physical platform for our research is Leonardo ("Leo"), a humanoid robot with 65 degrees of freedom that has been specifically designed for social interaction using a range of facial and body pose expressions (see Figure 1). Currently, Leo does not speak and therefore relies on gestures and facial expression for social communication. The robot's underlying software architecture consists of the following subsystems: speech recognition and parsing, vision and attention, cognition and behavior, and motor control.

3.1. Perceptual Systems

The robot has both speech and visual inputs. The vision system parses objects from the visual scene such as humans and the robot's toys (e.g., buttons that it can press). These perceptions are sent to the cognitive system along with object attributes (e.g., color, location). The vision system also recognizes pointing gestures and uses spatial reasoning to associate these gestures with their object referent.

The speech understanding system is a Lisp parser based on the NRL Nautilus project [21] with a ViaVoice front end. The system has a limited grammar to facilitate accuracy of the voice recognition. Upon receiving phrases from ViaVoice, the speech understanding system parses these into commands that are sent to the cognitive system.

3.2. Cognitive System

The cognitive system extends the `C5M` architecture, a recent version of the `C4` system described in [3]. It receives a continuous stream of symbols from the vision and speech understanding systems and integrates these into coherent beliefs about objects in the world. The perceptual attributes of a given object are merged together and kept in one structure. For example, information about a button is merged with the corresponding features of location, color and ON/OFF state to form a coherent belief about that button. These belief structures can also be manipulated internally, allowing the cognitive system to add information to its beliefs about the objects in the world (e.g., associating a label with a particular object so that the human can refer to it by name).

On top of these existing processing modules, we have added a set of higher-level cognitive capabilities: goal based decision making, hierarchical task representation, task learning, and task collaboration. These systems are described in the following sections.

4. Task and Goal Representation

Humans are biased to use an intention-based psychology to interpret an agent's actions [7]. Moreover, it has repeatedly been shown that we interpret intentions and actions based on goals rather than specific activities or motion trajectories [26, 9, 1]. A goal-centric view is particularly crucial in a collaborative task setting, in which goals provide a common ground for communication and interaction. All of this suggests that goals and a commitment to their successful completion should be central to task representation, both in learning and collaboration.

4.1. Goal Types

To support this idea, we have extended the notion of the `C5M` *action-tuple* data structure. An action-tuple is a set of preconditions, executables, and until-conditions [3]. Tasks and their constituent actions are variations of this action-tuple structure with the added notion of *goals*.

As the robot learns a new task, it must learn the goals associated with each action, each sub-task, and the overall task. The system currently distinguishes between two types of goals: (a) `state-change` goals that represent a change in the world, and (b) `just-do-it` goals that need to be executed regardless of their impact on the world. These two types of goals differ in both their evaluation as preconditions and in their evaluation as until-conditions. As part of a precondition, a `state-change` goal must be evaluated before doing the activity to determine if it is needed. As an until-condition, the robot shows commitment towards the

`state-change` goal in trying to execute the action, over multiple attempts if necessary, until succeeding to bring about the desired state. This commitment to the successful completion of goals is an important aspect of intentional behavior [4, 5]. Conversely, a `just-do-it` goal will lead to an action regardless of the world state, and will only be performed once.

4.2. Hierarchical Tasks

Tasks are represented in a hierarchical structure of actions and sub-tasks (recursively defined in the same fashion). Since tasks, sub-tasks, and actions are derived from the same action-tuple data structure, they are easily used in a unified way, naturally affording a tree representation for tasks.

A task also encodes the constraints among its actions. Currently we utilize only sequential constraints, but the constraint representation is generic and others could be added in the future.

4.3. Hierarchical Goals

When learning a task, a goal is associated with the overall task in addition to each of the constituent actions. Overall task and sub-task goals are distinct from the mere conjunction of the goals of their actions and sub-tasks, and are learned separately.

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. Overall task goals are evaluated separately from their constituent action goals to determine whether they need to be executed, as well as checking for completion of a task.

One advantage of this top-level evaluation approach is that it is more efficient than having to poll each of the constituent action goals explicitly. Moreover, this goal-oriented implementation supports a more realistic groundwork for intentional understanding—i.e., to perform the task in a way that accomplishes the *overall intent*, rather than just mechanically going through the motions of performing the constituent actions.

5. Task Manager Module

The *task manager* arbitrates between task learning and execution. It listens for a task-related request from the human partner. These can be in the form of: "Leo, do *task x*" or "Let's do *task x*". They can also be questions: "Leo, can you do *task x*?". Upon encountering such a request is, several scenarios can occur: Leo may either know or not know how to perform a task, and the request may be to perform

solo or collaboratively. The task manager distinguishes between these scenarios and starts the proper execution module, answering the person (using head nods and shakes) if the request was a question.

The task manager maintains a collection of known tasks. If Leo is asked to do a task on his own that he already knows, then the task manager executes it by expanding the task's action and sub-tasks onto a *focus stack* (in a similar way to [13]). The task manager proceeds through the actions on the stack popping them as they are done or, for a sub-task, pushing its constituent actions onto the stack.

The major contribution of this work, however, concerns the other two scenarios, learning (Section 6) and collaboration (Section 7). If Leo is asked to do a task that he does not know how to perform, a task learning module is instantiated to learn the task. Alternatively, if collaborative task execution is requested, the task manager starts the collaboration module for that task.

6. Learning in Collaboration with People

There are numerous examples where machine learning is viewed as a search problem, finding an optimal solution to a defined problem. We strongly believe that framing the learning problem in such a way does not take advantage of the wealth of structure and information provided by interaction with a human tutor. We frame the machine learning problem as a collaborative dialog between the human and the robot learner. In our approach, learning takes place through guided experience. This allows the robot to learn a complex task structure quickly from few examples.

As the human teacher leads the robot through a task, the use of sequencing words naturally indicates possible constraints between task steps. Since Leo shows his understanding of a newly learned sub-task or action by actually performing it (Figure 1), failure to comprehend an action or its goal is easily and naturally detected by the tutor. To complement this, we are currently working to incorporate feedback for correcting a task representation or emphasizing a particular segment. In a typical teacher-student interaction, errors are corrected just as they happen in the flow of the interaction; therefore, this type of error correction is likely to be the most natural for the human teacher.

Leo starts the learning process by indicating that he does not know a requested task, shrugging his shoulders and making a confused facial expression. At this stage the human walks the robot through the components of the task, building a new task from his set of known actions and tasks. While in task learning mode, the learning module continually pays attention to what actions are being performed, encoding the inferred goals with these actions. When encoding the goal state of a performed action or task, Leo compares the world state before and after its execution. In the

case that this action or task caused a change of state, this change is taken to be the goal. Otherwise, the goal is assumed to be of the `just-do-it` type. This produces a hierarchical task representation, where a goal is encoded for each individual part of the task as well as for the overall task. When the human indicates that the task is done, it is added to the task manager's collection of known tasks.

This method of goal classification makes some arbitrary, though reasonable, assumptions as to the priority of world state over action. We are currently working on a more flexible method to learn a more general representation for goals (see: Section 9).

Learning is handled recursively, such that a sub-task can be learned within a larger task. If the task manager receives an additional unknown sub-task request, while learning a task, the current learning process is pushed onto a stack and an additional learning thread is started. Once the sub-task learning is complete, it is popped from the stack and its resulting task is added both to the previous learning process and to the task manager's list of known tasks. The original learning process continues, with the newly learned sub-task as part of its task representation.

7. Performing in Collaboration with People

Task collaboration is the joint execution of a common plan. When Leonardo is performing a task alone, he progresses through the task tree until the task's goals are achieved. When collaborating with a human partner, many new considerations come into play. For instance, within a collaborative setting, the task can (and should) be divided between the participants, and the collaborator's actions need to be taken into account when deciding what to do next. Mutual support is provided in the case that one participant is unable to perform a certain action. Finally, a clear channel of communication is 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, progressing towards achieving the joint goal. In order to make the collaboration natural for people, we have implemented a number of the mechanisms that humans use when they collaborate. In particular, we have focused on communication skills to support joint activity (utilizing gestures and facial expressions), dynamic meshing of sub-plans and turn taking.

7.1. Dynamic Meshing of Sub-plans

Leo's intention system is a joint-intention model, which dynamically assigns tasks between the members of the collaboration team. Leo derives his intentions based on a dynamic meshing of sub-plans according to his own actions

and abilities, the actions of the human partner, his understanding of the common goal of the team, and his assessment of the current task state.

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. While usually conforming to this turn-taking approach, our system also supports simultaneous action, in which the human performs an action while Leo is working on another part of the task. If this is the case, Leonardo will re-evaluate the goal state of the current task focus, and might decide to no longer keep this part of the task on his list of things to do.

7.2. Social Communication and Mutual Support

¿From the theoretical work mentioned in Section 2.2, we see that cooperative behavior is an ongoing process of maintaining mutual beliefs, sharing relevant knowledge, coordinating action, and demonstrating commitment to the shared activity. To support this, we have implemented a variety of gestures and other social cues to communicate the robot's internal state during collaboration with the human — such as who the robot thinks is doing an action, or whether the robot believes the goal has been met. For instance, when the human partner changes the state of the world, Leo acknowledges this by glancing briefly towards the area of change before redirecting his gaze to the human. This post-action glance reassures the human 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 sub-task goal, Leo will glance at the change and give a small confirming nod while looking back at the human. Similarly, Leo uses subtle nods while looking at his partner to indicate when he thinks he brought about the completion of a task or sub-task.

This sort of social communication is particularly valuable when the human completes part of the joint plan in parallel to Leo performing a different part of the task, or when the human unexpectedly changes something in the world. The robot's behavior plays a crucial role in establishing mutual beliefs between the teammates on the progress of the shared plan.

Additionally, Leo has the ability to evaluate his own capabilities. If the robot is able to complete the task element, he will offer to do so, but allow the human partner to override this offer (either verbally or by acting on the current goal). Conversely, whenever the robot believes that he cannot successfully perform an action, he will ask the human for help. Since Leonardo does not have speaking capabilities yet, he indicates his willingness to perform an action by pointing to himself, and adopting an alert posture and facial expression (Figure 2(a)). Analogously, when detecting
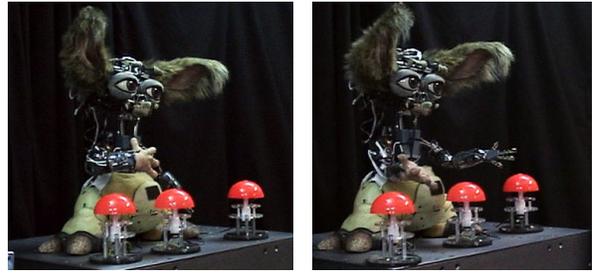
(a) Leo negotiates his turn by gesturing towards himself.

(b) Leo asking for help by gesturing towards the human.

Figure 2. Using communicative gestures.

an inability to perform an action assigned to him, Leo's expression indicates helplessness, as he gestures toward the human in a request for her to perform the intended action (Figure 2(b)). In addition to this gesture, Leo shifts his gaze between the problematic object and the human to direct her attention to what it is that he needs help with.

8. Results and Evaluation

We have tested Leonardo's learning and collaboration abilities on several tasks comprised of simple dialogs and object manipulation skills. In our experimental scenario, there are three buttons in front of Leonardo. The buttons can be pressed ON or OFF which changes their color by switching an LED on or off. Occasionally, a button does not light up when pressed. In our task scenario, this is considered a failed attempt. We designed the tasks to include a number of sequenced steps, such as turning a set of buttons ON and then OFF, and turning a button ON as a single action or as a sub-task of turning all of the buttons ON. The task set represents both simple and complex hierarchies, and has tasks with both `state-change` and `just-do-it` goals.

8.1. Learning the Button-Task with a Human

In our trials, we were able to teach Leonardo all of the above mentioned types of tasks. The robot demonstrated his understanding of nested action by recalling tasks which had been learned as sub-tasks of larger activities. He correctly associated `state-change` goals and `just-do-it` goals while learning new tasks. This was demonstrated in Leo's understanding of when to perform an action and for how long to persist based on its initial success.

In these learning trials, Leo's gestural cues provided much-needed feedback that enabled the tutor to realize

when the robot successfully understood a task and its place in the larger context. Figure 3 diagrams a typical teaching interaction in which Leo was taught to turn two buttons ON and then OFF again.

8.2. Performing the Button-Task with a Human

In the collaboration stage of our trials, the robot displayed successful meshing of sub-plans based on the dynamic state changes of the shared task. These changes were brought about by the robot's own successes and failures as well as the human 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 used subtle nods when he thought he completed a task or sub-task. For instance, in the case of the buttons-ON-then-OFF task, he gave an acknowledgment nod to the human after completing the buttons-ON sub-task and before starting the buttons-OFF sub-task.

Leo's communicative gestures proved particularly valuable when simultaneous action broke the turn-taking protocol. For example, if the human's simultaneous action met a task goal, such as turning the last button ON during the buttons-ON task, Leo glanced at the change and gave a small confirming nod to the human.

Additionally, Leo's need for help displayed his understanding of his own limitations, and his use of gaze and posture served as natural cues for the human to take appropriate action in each case.

See Figure 4 for a transcript of a typical collaborative interaction.

9. Discussion and Future Work

In viewing human-robot interaction as fundamentally a collaborative process and designing robots that communicate using natural human social skills, we believe we will achieve robots that are both intuitive for humans to interact with and that are better equipped to take advantage of our socially structured world. In both collaboration and learning we are utilizing acts that support collaborative dialog, such that the robot is continually communicating its internal state to the human partner, maintaining a mutual belief about the task at hand. We have presented our ability to teach a task to a robot through the course of collaborative dialog, and the ability to coordinate joint intentions to perform the learned task collaboratively. This section details our contributions in relation to prior research followed by plans for future work.

Discourse analysis and collaborative dialog theory have been used in plan recognition [18] and in tutorial systems [22]. Our work, however, takes a different view and uses a collaborative dialog framework for having the robot learn from a human rather than tutor a human.

Not only is dialog a natural form of interaction for the human, it can also provide structure and guidance to the learning process. Other methods that look at robots that learn from people include learning by demonstration or observation [24, 15], and instruction as programming with natural language techniques [16]. [20] is the most similar work to our own and explores a tutelage-inspired paradigm where a robot learns a sequentially structured task from human demonstration. The human uses short verbal commands to frame the interaction into instruction or demonstration episodes, and provides feedback to correct the robot's task model. Their work is interactive and follows a more familiar style of teaching for humans.

Our work goes further to model and represent learning as a collaborative process that leverages human social cues and gestures, tightly coupled turn taking, and dialog. By verbally instructing Leonardo, leading him through the desired task and using gestures to direct attention, this turn taking framework lets the human instructor model the robot's learning progress at each step. This in turn allows the human instructor to provide additional structure that is more appropriate and relevant to the robot's learning state at each step, thereby making the robot's learning process faster and more efficient.

The problem of shoulder-to-shoulder collaboration with a robot is a relatively unexplored field. It is important to distinguish human-robot collaboration from other forms of human-robot interaction. Namely, whereas interaction entails acting *on* someone or something else, collaboration is inherently working *with* others [4, 12].

Much of the current work in human-robot interaction is thus aptly labeled given that the robot (or team of robots) is often viewed as an intelligent tool capable of some autonomy that a human operator commands (perhaps using speech or gesture) to perform a task [14, 21]. This master-slave arrangement does not capture the sense of partnership that we mean by working *jointly with* others as in the case of collaboration.

Human robot collaboration has been studied most notably using autonomous vision-based robotic arms [6], albeit without a social communicative aspect. Other work studies the collaboration between teleoperated humanoids, such as NASA JSCs Robonaut [2] and human teammates. In other teleoperation work, partnership has been considered in the form of *collaborative control* (e.g. [8]), allowing a robot to ask a human for help in resolving perceptual ambiguities. The human is used by the robot as a remote source of information, but not as a peer on a shared task. We propose a different notion of partnership: that of a socially adept autonomous robot working with a human as a
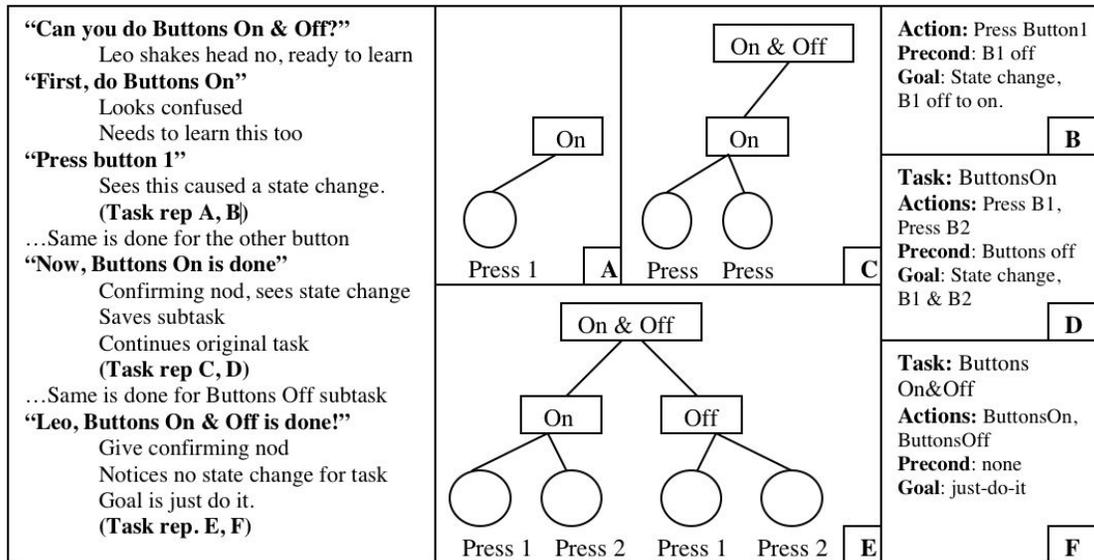
Figure 3. Learning to turn 2 buttons ON and OFF, and the progressive task and goal representation.

Figure 4. Collaborating on a hierarchical task.

member of a collocated team to accomplish a shared task.

In realizing this goal, we believe that robots must be able to cooperate with humans as capable partners and communicate with them intuitively. 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 within the context of the interaction they share. Our goal-driven joint intention based framework is aimed at this promise.

We are currently pursuing a number of extensions to the work presented here. In the learning stage, we are working to give the system a more flexible representation of the possible goals of tasks and actions. In the next iteration,

the robot will make a number of hypothesis about what the goals could be and then become more certain about these assumptions over multiple examples. Additionally, we would like to improve the collaborative interaction and dialog with a richer set of conversational policies. This would be useful for negotiating the meshing of sub-plans during task execution, making this process more flexible. For instance, the current turn taking mechanism works by negotiating task division at each step along the way. Allowing this negotiation to happen in advance as well would speed up the interaction and make it more natural to the human partner.

10. Conclusion

The goal of our work is to make robots more intuitive, efficient, and enjoyable for humans to interact with, to work with, and to teach. To do so, we model these capabilities as fundamentally collaborative processes that rely on open communication using natural human social skills, conventions and understanding. We have presented two important steps toward this goal: the ability to teach a task to a robot through the course of collaborative dialog with gesture and facial expression, and the ability to coordinate joint intentions to perform the learned task collaboratively. Our goal-centric approach at both the task and the action level, based on joint intention theory, proved valuable by establishing a common ground for both learning and collaboration, making them natural for the human, as well as flexible and efficient for the robot.

11. Acknowledgements

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] William Bluethmann, Robert Ambrose, Myron Diftler, Scott Askew, Eric Huber, Michael Goza, Fredrik Rehnmark, Chris Lovchik, and Darby Magruder. Robonaut: A robot designed to work with humans in space. *Auton. Robots*, 14(2-3):179–197, 2003.

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

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

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

[6] P. R. Cohen, H. J. Levesque, J. H. T. Nunes, and S. L. Oviatt. Task-oriented dialogue as a consequence of joint activity. In *Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, Nagoya, Japan, November 1990.

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

[8] Terrence W Fong, Charles 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.

[9] 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.

[10] J. Glidewell, editor. *The social context of learning and development*. Gardner Press, New York, NY, 1977.

[11] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, San Diego, CA, 1975.

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

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

[14] Hank Jones and Pamela Hinds. Extreme work teams: using swat teams as a model for coordinating distributed robots. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 372–381. ACM Press, 2002.

[15] Y. Kuniyoshi, M. Inaba, and H. Inoue. Learning by watching: Extracting reusable task knowledge from visual observation of human performance. *IEEE Transactions on Robotics and Automation*, 10:799–822, 1994.

[16] Stanislao Lauria, Guido Bugmann, Theocharis Kyriacou, and Ewan Klein. Mobile robot programming using natural language. *Robotics and Autonomous Systems*, 38(3-4):171–181, 2002.

[17] 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.

[18] Diane J. Litman and James F. Allen. Discourse processing and commonsense plans. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in communication*, chapter 17, pages 417–444. MIT Press, Cambridge, MA, 1990.

[19] Michael Montemerlo, Joelle Pineau, Nicholas Roy, Sebastian Thrun, and Vandi Verma. Experiences with a mobile robotic elderly guide for the elderly. In *National Conference on Artificial Intelligence*. AAAI, August 2002.

[20] Monica N. Nicolescu and Maja J. Matarić. Natural methods for robot task learning: Instructive demonstrations, generalization and practice. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, Melbourne, Australia, July 2003.

[21] Dennis Perzanowski, Alan C. Schultz, William Adams, Elaine Marsh, and Magda Bugajska. Building a multimodal human-robot interface. *IEEE Intelligent Systems*, 16(1):16–21, 2001.

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

[23] 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.

[24] R. Voyles and P. Khosla. A multi-agent system for programming robotic agents by human demonstration. In *Proceedings of AI and Manufacturing Research Planning Workshop*, 1998.

[25] L. S. Vygotsky. *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge, MA, 1978.

[26] 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.