Coordination Mechanisms in Human-Robot Collaboration

Bilge Mutlu, Allison Terrell, Chien-Ming Huang
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706 USA
{bilge, aterrell, cmhuang}@cs.wisc.edu

***Abstract*—Robots are envisioned to collaborate with people in tasks that require physical manipulation such as a robot instructing a human in assembling household furniture, a human teaching a robot how to repair machinery, or a robot and a human collaboratively completing construction work. These scenarios characterize *joint actions* in which the robot and the human must effectively communicate and coordinate their actions with each other in order to successfully achieve task goals. Drawing on recent research in cognitive sciences on joint action, this paper discusses key mechanisms for effective coordination—joint attention, action observation, task-sharing, action coordination, and perception of agency—toward informing the design of communication and coordination mechanisms for robots. It presents two illustrative studies that explore how robot behavior might be designed to employ these mechanisms, particularly joint attention and action observation, to improve measures of task performance and perceptions of the robot in human-robot collaboration.**

***Index Terms*—Human-robot collaboration, coordination, joint action, joint attention, action observation, gaze, dialogue, repair**

I. Introduction

Collaboration involves multiple parties working together to successfully achieve a shared goal, coordinating their actions in order to most effectively use their individual capabilities or skills. These *joint actions* differ from individual actions in that they involve shared intentions and goals and require contributions from all collaborators, which in turn requires collaborators to coordinate their individual actions and communicate their knowledge, intentions, and goals with each other to ensure a shared understanding of the task. Human-robot collaborations pose a similar challenge; the robot must appropriately communicate its knowledge, intentions, and goals with its human collaborators, while correctly understanding and interpreting those of its collaborators, and coordinate its individual actions with those of humans.

Joint actions may involve collaborators performing the same action or carrying out complementary actions toward a shared goal [1]. For instance, in the context of assembling a transmission, an expert mechanic and an apprentice might both be engaged in the action of cleaning the parts of the transmission before assembly, carrying out the same action on different parts. Alternatively, the mechanic and the apprentice might carry out complementary actions, the mechanic installing gears on the shaft while the apprentice prepares the next part or passes tools to the mechanic. The actions of the mechanic and the apprentice are also complementary in a training scenario in which the expert instructs the apprentice in carrying out the assembly task, monitors task progress, and provides the apprentice with feedback as necessary toward successfully completing the assembly task. The joint actions described in these situations all require the mechanic and the apprentice to coordinate their actions and to communicate to achieve this coordination [2].

This scenario illustrates the many domains in which humans and robots might work collaboratively to carry out physical tasks such as production assembly, training, occupational therapy, and routine household work. Humans and robots will similarly need to coordinate their actions to successfully complete joint tasks and to effectively communicate the to achieve this coordination. This paper seeks to answer the following questions: What are the mechanisms people use to achieve this form of coordination? How might robots employ these mechanisms? It presents ongoing work on designing coordination mechanisms for human-robot collaboration, covering two studies that explore (1) how robots might improve *joint attention* by disambiguating spatial references in a collaborative sorting task and (2) how robots might use *action observation* to monitor task progress and provide appropriate feedback toward improving outcomes in an assembly task.

The remainder of this paper outlines some of the key mechanisms humans use to communicate and coordinate actions and the mechanisms research in robotics has explored to date (Section II), describes the two studies on designing such mechanisms for robots (Sections III), and discusses open questions and remaining challenges on designing mechanisms for effective human-robot collaboration (Section IV).

II. Coordination Mechanisms in Joint Action

Many collaborative scenarios envisioned for human-robot interaction require *joint action* in which multiple agents engage in social interaction to coordinate their actions toward changing their environment [1]. Joint action is enabled by a number of cognitive and communicative mechanisms that facilitate shared attention, representations, and goals between the agents toward successfully and effectively complete the task at hand. Recent research in cognitive science has proposed that humans draw generally on five key mechanisms to perform joint action: (1) joint attention, (2) action observation, (3) task-sharing, (4) action coordination, and (5) perception of agency [1]. Research in robotics, particularly in cognitive human-robot interaction,

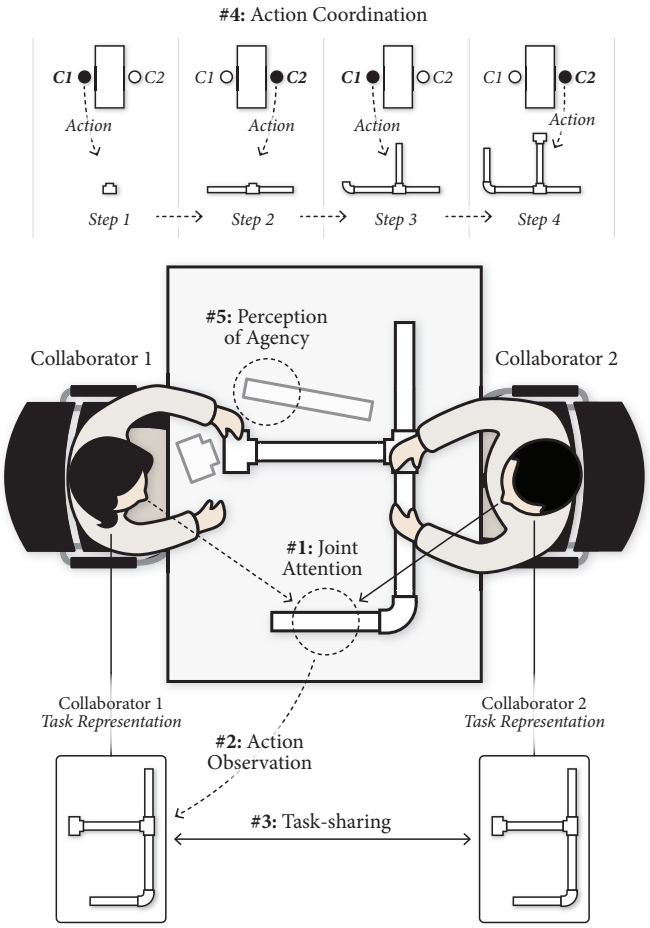

Fig. 1. The five key coordination mechanisms suggested by Sebanz et al. [1]: joint attention, action observation, task-sharing, action coordination, and perception of agency.

has also explored how robots might be designed to support some of these mechanisms. The paragraphs below describe these five mechanisms, drawing on the categorization proposed by Sebanz et al. [1], and review relevant work to date on human-robot joint action.

a) Joint attention: Individuals engaging in joint action draw on several aspects of language use, including spatial reference frames [3], gaze cues [4], and gestures and pointing [5], [6], and demonstrative task actions [2], to establish a "perceptual common ground" [1]. Research has shown that lacking the ability to attend to the same environment impairs joint action performance in assembly tasks, leading to increased task time and errors [7], suggesting that joint attention is a crucial mechanism for successful joint action. To achieve joint attention, participants seek to direct their partners' attention to a particular object or area of interest or to themselves [2] using deictic or conceptual references such as "on the left" or "behind the chair" [3], by looking, pointing, or gesturing toward these targets in the environment, [4]–[6], or by demonstrating "material signals," i.e., demonstrative actions toward establishing joint attention [2]. When a participant engages in such prompts, other participants seek to "align" their perspectives to that of the participant toward establishing a joint understanding of the topic at hand [8].

Among the mechanisms discussed here, joint attention has arguably received the most attention by robotics researchers. This research includes the development of models and implementations of gaze- and pointing-based skills for establishing joint attention [9], models of spatial referencing that enabled a robot to interpret directions such as "in front of" [10], and models that enabled the robot to take its user's perspective toward aligning its perspective with that of its user [11], [12].

b) Action observation: While the joint attention mechanism helps participants align their perspectives on task-relevant information in the environment toward establishing a "perceptual common ground," a common understanding of what actions the participants are taking is facilitated by the "action observation" mechanism [1]. Neuroscience research on "mirror neurons" (e.g., [13]) suggest that observations of material signals such as demonstrating intended actions [2] or performing actions activate mental representations of the actions in the observer, establishing a "procedural common ground" [14].

The research on "mirror neurons" and "motor resonance" has inspired the development of a number of "simulation-theoretic" models of human-robot interaction that involved a robot imitating or simulating the behaviors of its user in order to learn from or make inferences about its partner's actions [15]–[17]. For instance, Gray et al. [16] proposed a model of action observation that enabled a robot to observe the actions of its human counterpart and match these actions to actions in its repertoire toward making inferences about its users task progress and goals in order to anticipate its user's needs and offer relevant help.

c) Task-sharing: Sebanz and Frith [18] suggest that participants engaged in joint action also draw on a "task-sharing" mechanism to form shared representations of the task, particularly how individuals might respond to specific task or environmental conditions with specific actions, and predict the actions of their partners. While studies on action observation and coordination in robotics also involve matching perceived actions with shared representations (e.g., [17]), the models described in these studies do not help robots form generalized task representations toward predicting action outcomes using contextual cues.

d) Action coordination: A unique characteristic of joint actions is that participants engaged in a task have shared intentions and goals and engage in "planned coordination" to realize these intentions and goals [19]. This coordination involves developing a plan that specifies the joint action outcome and the part that each agent plays in achieving this outcome. Knoblich et al. [19] suggest that a shared task representation and joint perceptions are necessary elements in action coordination. Coordinated joint actions frequently involve participants performing complementary actions in order to reach a common goal [1].

Research in robotics has also explored how humans and robots might coordinate their actions toward accomplishing a shared goal. This research has drawn on Joint Intention Theory [20] and Activity Theory [21] to develop appropriate shared

task representations and strategies for task decomposition [22]–[24]. For instance, Breazeal et al. [22] proposed a model of human-robot joint action based on Joint Intention Theory that dynamically assigned sub-tasks of a joint goal to the robot or a human collaborator based on each actor's abilities and the current task state. Work on robotics on action observation, particularly recent work by Bicho et al. [17], has also considered how goals inferred from action observation might be combined with contextual cues and shared task knowledge to infer what actions the robot might take to complement the actions of its user.

e) Perception of Agency: A mechanism that functions hand-in-hand with action coordination is perception of agency in joint action, which the ability to distinguish among the actions of different actors and their effects [1]. Work in robotics on this mechanism is limited to an exploration of how robots might distinguish themselves from "animate others" in the mirror [25].

III. Coordination Mechanisms in Human-Robot Joint Action

The mechanisms described above make up a rich design space for designing similar cognitive and communicative mechanisms for robots to support human-robot joint actions. This section briefly describes two studies that explore this design space, particularly how robots might draw on these mechanisms to improve human-robot collaboration in physically situated joint action scenarios. The first study explores how robots might improve *joint attention* by disambiguating spatial references, enabling its users to better align their perspective with that of the robot, in a collaborative sorting task. The second study explores how robots might use *action observation* and conversational repair to improve joint task performance in a collaborative assembly task.

A. Study 1: Disambiguating Spatial References in Joint Action

One of the challenges in successfully establishing joint attention is the ambiguity in deictic communication that can arise from the speaker and the hearer taking different spatial perspectives [26]. Collaborators seek to resolve these ambiguities using linguistic references such as "on your left" [3], gaze shifts directed at referenced objects [4], pointing or gesturing toward objects [5], [6], or physically manipulating them [2]. Studies on joint attention have highlighted the importance of gaze cues in aligning perspectives between communicators [27] and improving the hearer's comprehension of spoken information [4], [28].

How might robots similarly use gaze cues to disambiguate spatial references to objects in joint action? Prior work in human-robot interaction has shown that robots' use of gaze cues in joint attention improves the perceived effectiveness of the interaction and perceived competence, sociability, and naturalness of the robot [29]. When gaze cues are used congruently with speech, participants also judge the accuracy of speech references faster than they do when no gaze cues are present or when they are not congruent with speech [30].

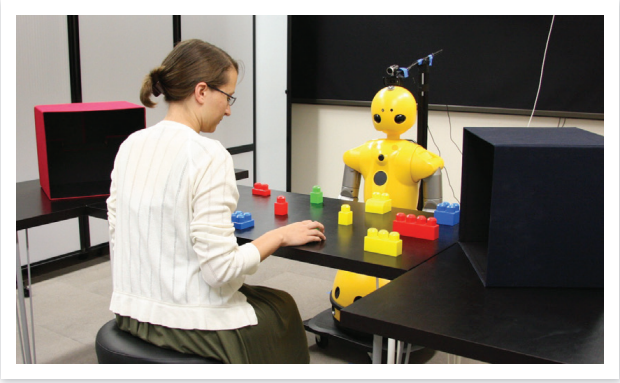

Fig. 2. In Study 1, a robot instructed participants to sort objects in different colors and shapes that were laid out on a table to different boxes.

These results provide evidence that the appropriate use of gaze cues by a robot improves perceived aspects of human-robot interaction and facilitated joint attention. This study built on these results to explore how these cues facilitate joint attention in physically situated human-robot collaboration.

To gain a better understanding of how gaze cues might disambiguate spatial references, thereby facilitating joint attention in human-robot joint action, the study involved a collaborative physical sorting task that a humanlike robot and a participant worked together to complete [24]. In the task, the robot instructed participants to move objects with different colors and shapes to boxes with different colors (Figure 2). Four variations of deictic gaze cues accompanied the robot's verbal instructions: (1) *congruent*, the robot's gaze toward the referenced object preceded the corresponding linguistic reference by approximately 800 to 1000 milliseconds, as suggested by research on human joint attention [31], [32], (2) *temporally incongruent*, the robot's gaze toward the referenced object followed the corresponding linguistic reference by approximately 500 to 1,000 milliseconds, aligning with the listener's gaze instead of the speaker's gaze [33], (3) *spatially incongruent*, the robot followed the timing in the *congruent* condition but directed its gaze toward an object that was different from the object that it referred to in its speech, and (4) *no gaze*, the robot did not direct its gaze toward any objects but instead tracked the participant's face to signal that it was attending to the participant. The outcome measures included the time it took participants to locate the object that the robot instructed them to move and their perceptions of the robot.

The results showed that participants located referenced objects in less time when the robot displayed congruent gaze cues to signal its referential attention than they did in other conditions. More specifically, participants in the temporally incongruent, spatially incongruent, and no-gaze conditions located the referenced object approximately 300 to 800 milliseconds after the robot provided the corresponding linguistic reference. This finding matches those in the joint attention literature; in the absence of speaker gaze cues, hearers look toward the referent approximately 500 to 1000 milliseconds after the onset of the spoken reference [33]. This finding suggests that, in the baseline conditions (i.e., temporally

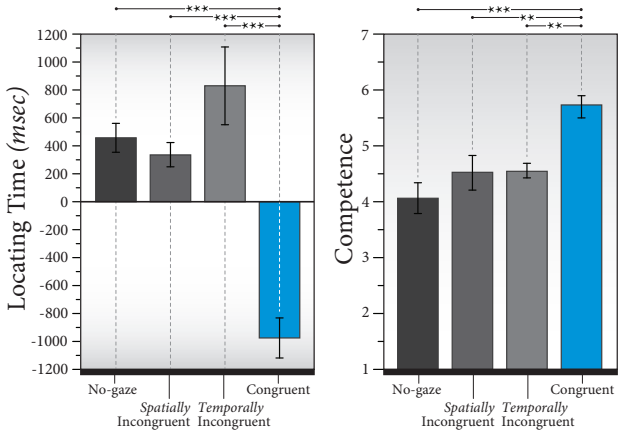

Fig. 3. Results showed that congruent gaze cues improved task performance and perceptions of the robot's competence. (**) and (***) denote significant differences with $p < .01$ and $p < .001$, respectively.

incongruent, spatially incongruent, and no-gaze conditions), participants did not utilize the robot's gaze cue to locate the referenced object. Participants in the congruent condition, on the other hand, drew on the information from the robot's gaze in performing the sorting task, leading to more effective human-robot collaboration. Moreover, participants perceived the robot that exhibited congruent gaze behavior to be more competent than the robot that exhibited other baseline behaviors. Figure 3illustrates these results. The reader can find other results and details of the data analysis in Huang and Mutlu [24].

The results of this study suggest that robot gaze cues, when used congruently with speech references, enable more effective joint attention by disambiguating spatial references in speech, improving task performance in joint action and the perceptions of the robot.

B. Study 2: Repairing Task Breakdowns in Joint Action

In joint action, participants must establish and maintain common ground through observing each others' communicative acts [2] and task actions [1]. However, breakdowns in communication [34] and common ground [35] frequently occur and can impede progress in the interaction until they are resolved or cause future breakdowns if left untreated [36]. Communicators seek to address these breakdowns through a process of conversational *repair* that results in all participants having a similar understanding of the information being discussed [8], [14], [36].

To better understand how breakdowns and repairs occur over the course of an interaction, this study included an observation of eight human dyads performing an assembly task in which one participant instructed the second to configure a set of plastic pipes. The analysis of this data involved modeling which participant initiated the breakdown, what information was missing or incorrect to cause the breakdown, and what information the instructor needed to detect and resolve the breakdown (e.g., visual or verbal information). The resulting model represented three types of triggers for repair—requests, hesitancy, and mistake detection—and how repair might be carried out for each type of trigger. *Requests* are initiated by participants to seek clarification or further information. *Hesitancy* occurs when a participant is expected to carry out a task action but delays the action due to uncertainty. Finally, *mistake detection* involves a participant identifying mistakes in the actions of the other participant such as attempting to install an incorrect part in a particular step of an assembly sequence.

The study next involved building an autonomous robot system with the ability to monitor these triggers by observing its user's task actions and task-related behaviors and interpreting speech. In the implementation of this system, a humanlike robot instructed participants to complete the same assembly task used in the observation study described above. To monitor and interpret the speech and visual information necessary to detect breakdowns and offer appropriate repair, the system included four modules: vision, listening, speech, and control. The vision module evaluated the workspace to detect mistakes during the assembly task, such as the user choosing an incorrect part for a particular step or installing the part in the wrong location. The listening module detected and categorized utterances from the participant, such as the participant asking for clarification on which part to use or requesting repetition of the instruction. The control module determined whether repair was needed by evaluating information from the vision and listening modules. When the control module determined that repair was needed, it passed the type of breakdown detected to the speech module, which then provided the participant with appropriate feedback on what action to take.

A controlled laboratory experiment with human participants tested whether the repair provided by the robot improved human-robot collaboration in measures of task performance and perceptions of the robot in the instructional assembly task described above (Figure 4). The design of the experiment followed a between-participants design, each participant interacting with an autonomous robot operating under one of three repair models: no repair, simple repair, and humanlike repair. In the *no-repair* condition, the robot did not provide participants with repair when it observed breakdowns or respond to requests for repair, waiting until the current step of the assembly to be completed successfully in order to proceed with instructions on the next step. In the *simple-repair* condition, the robot offered "yes/no" responses to any questions that could be answered with a "yes" or a "no." Other questions such as "Which pipe do I need?" resulted in the robot repeating the instruction.

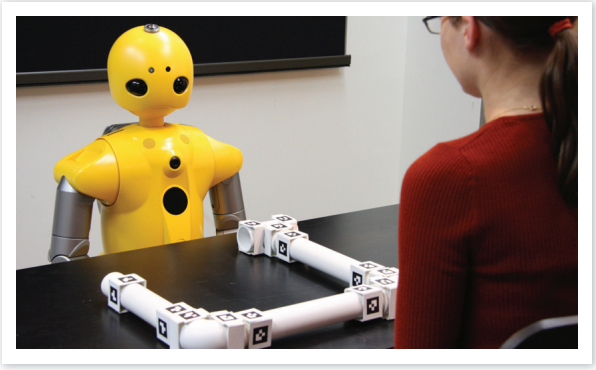

Fig. 4. In Study 2, the robot instructed participants in completing an assembly task while it monitored task progress for breakdowns and rectified them.

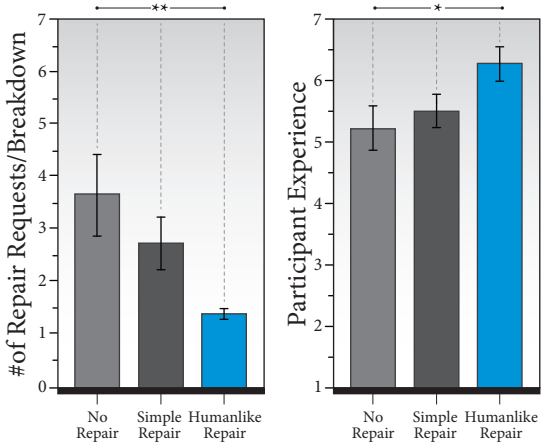

Fig. 5. Results showed that humanlike repair reduced requests for repair by participants and improved participant experience while simple repair did not. (*) and (**) denote significant differences with $p < .05$ and $p < .01$, respectively.

Finally, in the *humanlike-repair* condition, the robot followed the model outlined above, providing appropriate repair for all breakdowns it detected.

The results showed that participants requested significantly fewer repairs per breakdown in the humanlike-repair condition compared with the no-repair condition, while there were not significant differences in the number of requests per breakdown between the simple- and no-repair conditions. Similarly, participants reported a significantly more positive experience in the human-repair condition compared with the no-repair condition, while these differences were not observed between the simple-repair and no-repair conditions. These results are summarized in Figure 5. The reader should refer to future publications for other findings and further detail on data analysis.

These results suggest that robots might improve task performance and user experience in human-robot collaboration by monitoring for breakdowns through action observation and dialogue and by rectifying them through appropriate feedback. The findings of the study highlight action observation and conversational repair as effective mechanisms for supporting human-robot joint action.

IV. Discussion

The studies described above illustrate the growing body of work in robotics on the development of coordination mechanisms to support human-robot collaboration. The findings from these studies indicate that coordination mechanisms, such as joint attention and action observation, and communicative mechanisms, such as conversational repair, have the potential to improve task efficiency in human-robot collaboration and perceptions of robots by human collaborators. The broader set of mechanisms described in Section II offer a rich and promising space for design and research that future work in robotics should explore. The studies described above illustrate this promise while highlighting a number of key challenges that future work should also consider. The paragraphs below discuss some of these challenges and potential directions for addressing them.

A. Integrating Real-time Action and Behavior Recognition

Many of the coordination mechanisms in joint action require that collaborators observe and make sense of task-related actions and communicative acts, such as gaze shifts toward a referent, spatial references in speech, or manipulating an object to demonstrate possible actions. To effectively utilize these mechanisms in human-robot collaboration settings, robots must detect, recognize, and interpret human actions and behaviors in real time. For instance, the system described in Study 2 recognized task actions by tracking augmented-reality (AR) tags installed on the parts that the participants were asked to assemble. However, such instrumentation of the environment and objects is not feasible or realistic for more complex tasks. Future work in this area must inform the development of robust action and behavior recognition algorithms and refine them for use in joint action situations.

B. Beyond Repairing Breakdowns

Results from Study 2 underline the importance of repairing communication breakdowns that are essential to collaboration. However, the conversational repair approach that this study adopted sought to detect and repair breakdowns only after they occurred, while conversational repair in human communication involves mechanisms for anticipating breakdowns and warning communication partners of unavoidable breakdowns or preventing avoidable breakdowns *before* breakdowns occur [34]. To anticipate breakdowns before they occur, robots will need to employ richer models of communication and task breakdowns and draw on real-time action and behavior recognition. The development of these models should also include the design of appropriate strategies for preventing or warning about anticipated breakdowns.

C. Pushing the Envelope in Collaboration Scenarios

Finally, most human-robot collaboration scenarios considered by research to date including the two studies described above use "toy" tasks or tasks that are simplified for proof-of-concept implementations or evaluations. However, such joint action situations do not warrant the use of many of the mechanisms described in Section II. For example, tasks that require only a few steps to complete require very little action coordination. Similarly, agency problems—failure to distinguish between the effects of one's own actions and those of others—do not arise in instructional scenarios employed in the two studies described above. Future work should consider real-world tasks that introduce more realistic coordination challenges and require robots to employ the broader set of coordination mechanisms. Using such tasks might also serve as gateways for real-world applications.

Future work that explores the broader set of mechanisms described in Section II and the challenges discussed above will enable robot systems that support real-world human-robot collaborations.

V. Conclusion

Robots are envisioned to collaborate with people in tasks that require *joint action*—multiple agents engaging in social interaction to coordinate their actions toward changing their environment [1]. Research in cognitive sciences suggests that, to achieve this coordination, humans draw on a number of communicative and cognitive mechanisms including joint attention, action observation, task-sharing, action coordination, and perception of agency [1], [19]. Research in robotics has also explored how these mechanisms might support human-robot collaboration, contributing to a growing body of work on the development of coordination mechanisms for robots.

This paper outlined key coordination mechanisms that humans employ in joint action and presented two studies that explored how robots might employ such mechanisms toward improving human-robot collaboration. The first study investigated how robots might use gaze cues to improve joint attention by disambiguating spatial references in a collaborative sorting task and showed that appropriate use of gaze cues to support joint attention result in significant improvements in task performance and perceptions of the robot compared. The second study involved the development of an autonomous robot system that monitored collaborator actions for task breakdowns through action observation and used conversational repair strategies to rectify these breakdowns. An evaluation of the system showed that the use of repair strategies significantly reduced requests for help by the participants and improved participant experience with the robot.

These studies illustrate the promise that coordination mechanisms hold for improving human-robot collaboration and underline the rich space these mechanisms offer for design and research that future work should explore. The studies also highlight a number of challenges, including the need for robust action and behavior tracking, richer models of repair, and real-world human-robot collaboration tasks, that future work should consider to enable robots that effectively collaborate with people.

VI. Acknowledgments

This work was supported by National Science Foundation awards 1017952 and 1149970.

References

[1] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends Cogn Scis*, vol. 10, no. 2, pp. 70–76, 2006.

[2] H. H. Clark, "Coordinating with each other in a material world," *Discourse studies*, vol. 7, no. 4-5, pp. 507–525, 2005.

[3] M. F. Schober, "Spatial perspective-taking in conversation," *Cognition*, vol. 47, no. 1, pp. 1–24, 1993.

[4] D. C. Richardson and R. Dale, "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension," *Cognitive Science*, vol. 29, no. 6, pp. 1045–1060, 2005.

[5] H. H. Clark and S. E. Brennan, *Grounding in communication*. American Psychological Association, 1991, pp. 127–149.

[6] A. Bangerter, "Using pointing and describing to achieve joint focus of attention in dialogue," *Psych Sci*, vol. 15, no. 6, pp. 415–419, 2004.

[7] H. H. Clark and M. A. Krych, "Speaking while monitoring addressees for understanding," *Journal of Memory and Language*, vol. 50, no. 1, pp. 62–81, 2004.

[8] S. Garrod and M. J. Pickering, "Why is conversation so easy?" *Trends Cogn Scis*, vol. 8, no. 1, pp. 8–11, 2004.

[9] B. Scassellati, "Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot," *Computation for metaphors, analogy, and agents*, pp. 176–195, 1999.

[10] R. Moratz, K. Fischer, and T. Tenbrink, "Cognitive modeling of spatial reference for human-robot interaction," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 04, pp. 589–611, 2001.

[11] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz, "Enabling effective human-robot interaction using perspective-taking in robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 35, no. 4, pp. 460–470, 2005.

[12] R. Ros, S. Lemaignan, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken, "Which one? grounding the referent based on efficient human-robot interaction," in *Proc ROMAN'10*, 2010, pp. 570–575.

[13] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annu. Rev. Neurosci.*, vol. 27, pp. 169–192, 2004.

[14] H. H. Clark, *Using language*. Cambridge Univ. Press, 1996.

[15] M. N. Nicolescu and M. J. Mataric, "Linking perception and action in a control architecture for human-robot domains," in *Proc HICSS'03*, 2003.

[16] J. Gray, C. Breazeal, M. Berlin, A. Brooks, and J. Lieberman, "Action parsing and goal inference using self as simulator," in *Proc ROMAN'05*, 2005, pp. 202–209.

[17] E. Bicho, W. Erlhagen, L. Louro, and E. Costa e Silva, "Neuro-cognitive mechanisms of decision making in joint action: A human–robot interaction study," *Hum Movement Sci*, vol. 30, no. 5, pp. 846–868, 2011.

[18] N. Sebanz and C. Frith, "Beyond simulation? neural mechanisms for predicting the actions of others," *Nature neuroscience*, vol. 7, no. 1, pp. 5–6, 2004.

[19] G. Knoblich, S. Butterfill, and N. Sebanz, "Psychological research on joint action: Theory and data," *Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 54, p. 59, 2011.

[20] P. R. Cohen and H. J. Levesque, "Teamwork," *Nous*, pp. 487–512, 1991.

[21] A. N. Leont'ev, "The problem of activity in psychology," *Journal of Russian and East European Psychology*, vol. 13, no. 2, pp. 4–33, 1974.

[22] C. Breazeal, G. Hoffman, and A. Lockerd, "Teaching and working with robots as a collaboration," in *Proc AAMAS'04*, 2004, pp. 1030–1037.

[23] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, and R. Chatila, "Task planning for human-robot interaction," in *Proc sOc-EUSAI'05*, 2005, pp. 81–85.

[24] C.-M. Huang and B. Mutlu, "Robot behavior toolkit: Generating effective social behaviors for robots," in *Proc HRI'12*, 2012.

[25] P. Michel, K. Gold, and B. Scassellati, "Motion-based robotic self-recognition," in *Proc IROS'04*, vol. 3, 2004, pp. 2763–2768.

[26] G. Retz-Schmidt, "Various views on spatial prepositions," *AI magazine*, vol. 9, no. 2, p. 95, 1988.

[27] D. C. Richardson, R. Dale, and N. Z. Kirkham, "The art of conversation is coordination common ground and the coupling of eye movements during dialogue," *Psych sci*, vol. 18, no. 5, pp. 407–413, 2007.

[28] M. K. Tanenhaus, M. J. Spivey-Knowlton, K. M. Eberhard, J. C. Sedivy *et al.*, "Integration of visual and linguistic information in spoken language comprehension," *Science*, vol. 268, no. 5217, pp. 1632–1634, 1995.

[29] C.-M. Huang and A. L. Thomaz, "Effects of responding to, initiating and ensuring joint attention in human-robot interaction," in *Proc ROMAN'11*, 2011.

[30] M. Staudte and M. W. Crocker, "Visual attention in spoken human-robot interaction," in *Proc HRI'09*, 2009, pp. 77–84.

[31] Z. M. Griffin, "Gaze durations during speech reflect word selection and phonological encoding," *Cognition*, vol. 82, no. 1, pp. B1–B14, 2001.

[32] A. S. Meyer, A. M. Sleiderink, and W. J. M. Levelt, "Viewing and naming objects: eye movements during noun phrase production," *Cognition*, vol. 66, no. 2, pp. B25–B33, 1998.

[33] B. Fischer, "Attention in saccades," in *Visual attention*. Oxford University Press, 1998, pp. 289–305.

[34] H. H. Clark, "Managing problems in speaking," *Speech communication*, vol. 15, no. 3, pp. 243–250, 1994.

[35] G. Klein, P. J. Feltovich, J. M. Bradshaw, and D. D. Woods, "Common ground and coordination in joint activity," *Organizational simulation*, pp. 139–184, 2005.

[36] C. Zahn, "A reexamination of conversational repair," *Communications Monographs*, vol. 51, no. 1, pp. 56–66, 1984.