Analysis of empirical results on argumentation-based dialogue to support shared decision making in a human-robot team

Mohammad Q. Azhar[1] and Elizabeth I. Sklar[2]

Abstract— This paper reports on a study conducted with human-robot teams in which *computational argumentation-based dialogue* was implemented as part of a framework designed to support collaborative decision making. The focus here is on detailed analysis of how argumentation-based dialogue was employed during the study, illustrating ways in which this type of dialogue impacted how users engaged with the robot. The analysis demonstrates that users took advantage of the capability and reached shared decisions as a result of exchanging information with the robot and coming to agreement about robot actions and the team's task environment.

I. INTRODUCTION

We are interested in situations where the initiative for collaboration in a human-robot team can emanate from either the human or the robot, where discussion about actions can ensue, *reasons* for and against taking particular actions can be passed back and forth, and the responsibility for actions can flow to either the human or the robot, until a mission is completed. For example, suppose a human-robot team is asked to fetch an orange ball from an unfamiliar environment. If the human asks a robot to help her look for such a ball which she can pick up, the robot may wander around their environment taking pictures and sending them to the human, without knowing whether it has captured an image of a ball or a round fruit that is orange-colored. Feedback from the human about the content of the image would improve the likelihood that the human-robot team obtains a picture of an orange ball, because humans have better abilities to distinguish between items that closely resemble each other. If the robot does not have a manipulator or the image shows that the orange ball has rolled under a piece of furniture such that the robot cannot reach it, then the robot will need help from the human to pick up the ball. Feedback from the robot about the location where the image was taken and directions for how to get there would improve the likelihood that the human-robot team retrieves the ball, because robots have better abilities to map their environment and perform path planning using the map.

If the robot and human disagree about an aspect of their task, the ability to communicate the reasons for their individual opinions—the *evidence* that led each to reach their own *conclusion*—can be invaluable in resolving the conflict. Thus, rich discussion enables *shared decision making* about

the content of candidate images, *agreement* about an image that indeed contains an orange ball and *guidance* for retrieving the ball. In particular, reaching agreement—evolving to a state of *shared belief* regarding the task at hand—is a long-studied topic in artificial intelligence (AI) and includes work on *joint intentions* [1], *shared plans* [2] and *teamwork* [3].

For *human-robot interaction (HRI)* systems to be truly collaborative, participants must be able to engage in opportunistic exchange that can adjust dynamically as the situation unfolds. However, dialogue that facilitates opportunistic exchange of ideas is not well supported in today's human-robot systems. Upon experiencing (or expecting to experience) failure or discovering new opportunities—at moments unforeseen by the human collaborator—the robot, as well as the human, needs to be able to take the *initiative* [4] in an ongoing or new conversation.

This paper reports on the use of *computational argumentation-based dialogue* [5], [6] to support shared decision making in human-robot teams, in particular focusing on analysis of interaction logs taken from a study in which such a dialogue framework was implemented and evaluated. The argumentation-based dialogue theory underlying the framework is described elsewhere [7]. Experimental results from the study that focus on task performance statistics (such as how long it took for human-robot teams to make decisions, how long it took for robots to perform actions and how far robots traveled) are also reported elsewhere [8]. In this paper, we focus on detailed examination of *how* decisions were made, by analyzing the situations in which *computational argumentation-based dialogue* was invoked, by both humans and the robot, and how the use of this dialogue capability influenced actions in the experimental domain.

II. APPROACH

Our technical approach is centered on the application of *computational argumentation*, a formal logic-based model of reasoning in which *claims* are examined with respect to evidence that either *supports* the claim or *conflicts* with it [9], [10]. We have developed a software framework called *ArgHRI* [7] which employs *computational argumentation-based dialogue* [5], [6] to enable a robot to interact with a human and exchange ideas. The formal theory allows the robot to manage its beliefs in a structured way. Beliefs are represented using first-order logic and maintained using fundamental elements of computational argumentation [6], [11], which include a store of the robot's committed beliefs about itself and its environment, its beliefs about the human and a record of what has been said in a dialogue.

[1]Mohammad Q. Azhar is with the Borough of Manhattan Community College, City University of New York, USA <mazhar@bmcc.cuny.edu>

[2]Elizabeth I. Sklar is with the Department of Informatics, King's College London, UK <elizabeth.sklar@kcl.ac.uk>

	$(b \in H) \in R$	$(\neg b \in H) \in R$	$(?b \in H) \in R$
$b \in R$	agreement *(no dialogue)*	disagreement *(persuasion)*	lack of info *(info-seek)*
$\neg b \in R$	disagreement *(persuasion)*	agreement *(no dialogue)*	lack of info *(info-seek)*
$?b \in R$	lack of info *(info-seek)*	lack of info *(info-seek)*	shared lack *(inquiry)*

Fig. 1. Cases for different types of dialogues (from [7]), based on the beliefs of the robot (R) and its beliefs about the human (H). $b \in R$ means that the robot believes b. $(b \in H) \in R$ means that the robot believes that the human believes b. The notation $?b$ means that the robot (or human) has no knowledge of or information about b (or not b, written $\neg b$).

The argumentation-based dialogue system developed for ArgHRI is an implementation of a *dialogue game*, adapted from [12], [6]. This formal model prescribes a structure for defining *protocols* for utterances associated with each type of dialogue, essentially a list of legal moves that each participant in the dialogue game is allowed to select from. In related work, we have defined the protocols for each of the types of dialogue implemented in ArgHRI [13], [7], including state machines that control the possible sequences of locutions, from start state to termination state. One of the advantages of employing argumentation-based dialogue games is that it has been proven formally that the rules for each type of dialogue guarantee termination [6].

Three types of argumentation-based dialogue have been implemented in ArgHRI: **information-seeking** [5]—where one participant seeks answers to questions from another participant, who is believed by the initiating participant to know the answers; **inquiry dialogue** [14]—where the participants collaborate to answer a question or questions whose answers are not known to any participant; and **persuasion dialogue** [15]—where one participant seeks to persuade another party with a different opinion to adopt a belief or point-of-view. Examples of human-robot scenarios where each type of dialogue may be applied include the following. The robot could ask the human for information that the robot does not have and believes that the human has, in order to prevent errors; this is an example of an *information-seeking* dialogue. The robot and human may agree to seek an answer to an unknown query because neither of them has enough information to make an informed decision; this is an example of a *inquiry* dialogue. The robot might discover information that the human does not possess or that contradicts something the human knows, in order to correct the human's misconception(s) and pre-empt possible failure; this is an example of a *persuasion* dialogue.

Figure 1 lists the conditions under which the robot can initiate each of these three types of dialogue. A belief, b, is considered, and a decision is made about which type of dialogue is appropriate based on the robot's belief in b (or its negation, $\neg b$), taken in combination with the robot's beliefs about the human's beliefs. Note that there is also the case where the robot or human "do not know b", represented as $?b$ (i.e., neither has information about b or $\neg b$).

An argumentation-based dialogue game is defined by the following [16][17]: a set of rules that defines the pre-conditions or circumstances under which the dialogue game can begin; the complete set of possible moves consisting of statements issued by one participant and directed toward the other participant; a set of *protocols* that governs which moves a participant can make in each dialogical context and when each participant can move; a set of rules that defines the circumstances under which each participant may commit to a conclusion (*acceptance*); and a set of rules that enables a dialogue to reach its termination condition.

Following the rules of dialogue games [12], participants can only utter beliefs that they can *support* in the formal sense of argumentation. This means that a participant is allowed to utter b, where b is represented as an atomic *fact* in the participant's internal set of beliefs, or c, where c is a conclusion that can be drawn from a set of arguments (S, c). Here, S is considered the *support* for the conclusion c, such that each element of S is either an atomic fact (like b) or can be derived from rules that are part of the participant's set of beliefs. For example, the rule: $\mathsf{GoTo}(t, loc) \rightarrow \mathsf{At}(t+1, loc)$ says that if the predicate on the left is true, then the predicate on the right can be concluded as the result of applying the rule. An atomic fact, b, can represent fixed knowledge about the world or about the domain or the robot's environment, all of which could be established *a priori* and remain unchanged during a robot's mission. In addition, b can represent information about the robot or its environment that does change during a mission, e.g., the output of a $\mathsf{Sense}(t, loc)$ action.

ArgHRI employs the *ArgTrust* argumentation engine [18] to calculate the support for a specific conclusion based on the robot's beliefs. It is used by the robot to reason about how a goal might be achieved or to resolve conflicts found within its internal set of beliefs or between its beliefs and its beliefs about the human's beliefs. Conflicts can be formally computed in two ways: *undermining*, where the human's conclusion of an argument conflicts with the conclusion of the robot's arguments or vice versa; and *rebuttal*, where the conclusion of the human's argument conflicts with some element in the support of the robot's argument or vice versa. The output from ArgTrust is an *accept* or *defeat* (*reject*) predicate, containing evidence that supports or attacks a conclusion, respectively; the output can also be *undecided*, if the evidence is inconclusive.

The robot is actualized for ArgHRI through the *HRTeam* robot operating environment [19]. The HRTeam framework is structured around a client-server architecture that comprises an *agent* layer (for intelligence), a *robot* layer (for actuating robot behaviors) and a centralized server for passing messages between nodes on each layer. The robot layer is built on *Player/Stage* [20], which supports easy switching between physical or simulated robots. Figure 2 shows the HRTeam arena (a) and robot (b) that were employed in ArgHRI and used for the experiments described here.

III. EXPERIMENTS

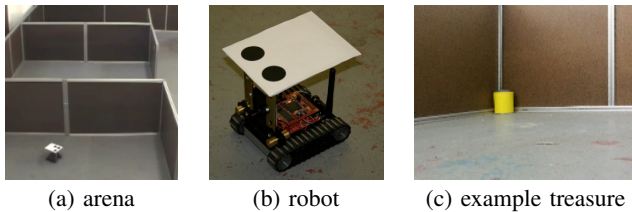The experiments analyzed here are based on two user studies conducted using the ArgHRI framework to play an

(a) arena   (b) robot   (c) example treasure

Fig. 2. The ArgHRI+HRTeam experimental setup.

Fig. 3. The ArgHRI user interface

adaptation of the *Treasure Hunt Game (THG)* [21] [7]. Our version of the THG involves a human and a robot player who work together as a team. The robot operates inside the arena and has the ability to move around, use sensors to gather data and communicate with the human. The human operates outside the arena and has the ability to receive data from the robot about the arena and communicate with the robot. Their task is framed as a real-time strategy game in which they must locate objects, or "treasures," in the arena. The robot has an energy level that decreases when it moves, when it gathers sensor data, and when it transmits sensor data to the human. The robot does not have enough energy to perform an exhaustive search of the arena to find all the treasures. Thus, the shared mission of the THG is for the human-robot team to find and correctly identify as many treasures in the arena as possible before the robot loses all of its energy. The human-robot team's score in the game is the number of points earned by correctly identifying treasures. They lose points by incorrectly identifying treasures (see Figure 2c).

In order to conduct controlled experiments, we designed a series of three decision points to test the shared decision-making capabilities of the human-robot teams. First, the human and robot decide *where to go*—which rooms should be visited in order to look for treasure (since an exhaustive search is not possible). Second, the human and robot decide *how to get there*—the order in which the agreed-upon rooms should be visited. The human plans a travel sequence to visit the rooms, and so does the robot, using the A* path planner [22]. Then, the human and robot engage in dialogue to identify any conflicts in their plans and reach agreement about which path plan the robot will follow. Third, the human and robot decide *what is found there*—whether the sensor data collected by the robot (e.g., images) contains treasures.

Here we analyze dialogue transcripts from two user studies, each conducted under a different operating condition. The same user interface was employed for both operating conditions, shown in Figure 3. Under one condition, human subjects interacted with *physical* robots operating in the physical arena shown in Figure 2a. Under the other, humans interacted with *simulated* robots operating in a virtual version of this arena. The data for each condition is analyzed independently here. In both studies, each human subject played two treasure hunt games: one with a robot capable of engaging in dialogue, which produced the data analyzed here; and one with a robot unable to engage in dialogue[1].

[1]This provided a baseline control for comparing task performance metrics, but the data is not analyzed here, because there was no dialogue.
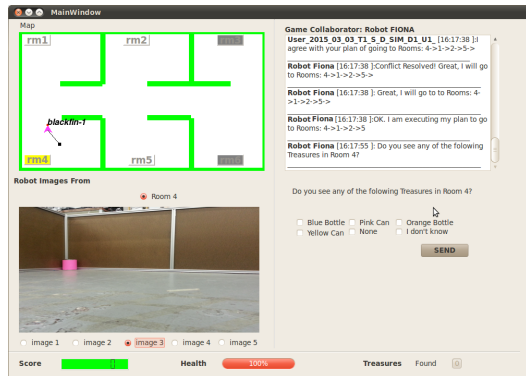
Statistical analysis comparing objective performance metrics, which included distance travelled by robots and game score obtained by each team, showed significantly better results for the teams with the dialogue-enabled robot for both operating conditions. Statistical analysis comparing subjective results, in which users indicated their level of satisfaction and trust working with the robot, were also significantly better for the teams with the dialogue-enabled robot. See [8] for details.

In the physical robot experiment, there were 27 participants. In the simulated robot experiment, there were 33 participants. In the physical experiments, 20 participants were male and 7 were female. In the simulation experiment, 22 participants were male and 11 were female. In both experiments, most of the participants were undergraduates (97%) and ranged in age from 18 to 24 (74%), while the remaining 26% of the participants were 25 to 39 years old. About half of the participants had no prior experience with robots; the rest had less than one year of experience.

IV. RESULTS

For each treasure-hunt game, the human-robot team has three *decision points* in which to discuss where to search, how to get there and the identity of any treasure found. With 60 participants, there were at least 180 unique opportunities for the robot or human to initiate dialogue. If the robot visited more than one room, there were opportunities for additional dialogues, because the third decision point was revisited each time the robot collected sensor data in a room. At each decision point, depending on the pre-conditions met (see Figure 1), one of three types of argumentation dialogue (information-seeking, inquiry, or persuasion) was triggered. We analyze the dialogue transcripts in several ways.

First, we examine the number of opportunities where dialogue could have occurred in each game. The first two decision points occurred once per game, whereas the third decision point occurred each time the robot captured images. There were 3-4 treasures hidden per game; we compute summary statistics for the first five times that the third decision point was encountered. Figure 4 presents summary statistics for the average number of times, under each experimental operating condition, the opportunity to engage in dialogue

	no dialogue (agreement)		dialogue occurred	
(a) first decision point, *where to go*				
physical	9	(30%)	18	(70%)
simulation	6	(18%)	27	(82%)
(b) second decision point, *how to get there*				
physical	12	(40%)	18	(60%)
simulation	11	(33%)	22	(67%)
(c) third decision point, *what is found there*, for the first five candidate treasures				
physical	31	(23%)	104	(77%)
simulation	49	(30%)	116	(70%)

Fig. 4. Number of arg.-based dialogues triggered per decision point

(a) physical

(b) simulation

Fig. 6. Breakdown of dialogue usage, by dialogue type.

was taken. When a decision point is reached and the human and robot agree, there is no dialogue. The high percentages in the figure for cases where dialogue did take place indicate how frequently the human and robot disagreed. On average, under the physical operating condition, 69% of opportunities triggered dialogue; under the simulated operating condition, 73% of opportunities triggered dialogue. *This result shows the value of providing the opportunity for dialogue, even in the simple, controlled experimental domain employed here.*

The number of times that each human-robot team reached the third decision point varied, depending on how quickly the team was playing, how many rooms the robot was able to search, how many images the robot was able to capture and send to the human, and how much energy the robot had left. Figure 5 shows the percentage of dialogue opportunities triggered at the third decision point, as games progressed, averaged over human subjects, grouped by operating condition. The plot clearly illustrates that the proportion of players who reached beyond a fifth opportunity declined dramatically.
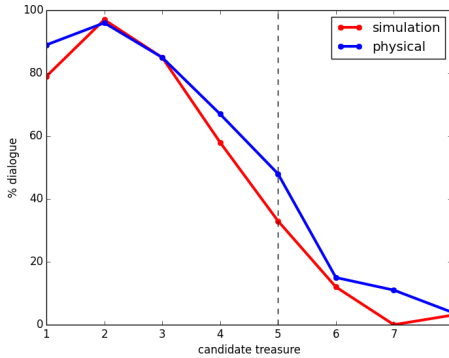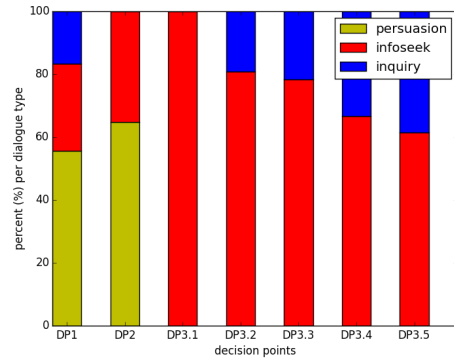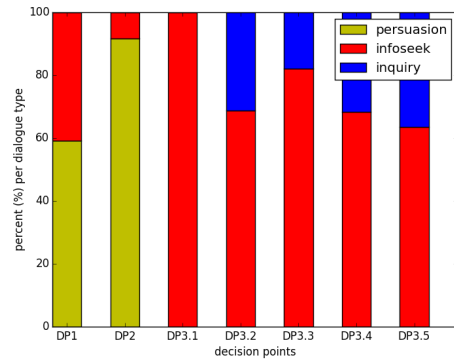

Fig. 5. Percentage of opportunities in which arg.-based dialogue was triggered at 3rd decision point, per successive candidate treasure (1st-8th).

Second, we look at the breakdown for each type of dialogue. Figure 6 illustrates the percentage for each type of dialogue that occurred in the physical (a) and simulation (b) operating conditions. For the first two decision points, the least frequently used dialogue was *inquiry*, which is triggered when neither the human nor the robot have knowledge about the belief under consideration. The *persuasion* dialogue was not employed for the third decision point, primarily because

the human users were able to draw conclusions about images, while the robots were unable to interpret the image data. Consequently, the first instance of decision point 3 was always an *information-seeking* dialogue, initiated by the robot, who made the initial assumption that the human could provide information about the image. If this were not the case, then the robot (or human) would initiate an *inquiry* in the second (and later) instance(s) of the third decision point.

Third, we examine the dialogue outcomes. Figure 7 contains the number of times a dialogue initiator was *challenged*, and the number of times a dialogue terminated in *acceptance* or *rejection*. A *challenge* move is invoked when one participant does not agree with a belief put forth by the other participant. The response is for the challenged participant to provide their *support* (*evidence*) for their belief, in a series of moves in the dialogue. It is a notable result that more than half the decisions for the first two decision points were challenged, yet more than 85% (on average) were subsequently accepted. *These data clearly illustrate the effectiveness of the dialogues.* Many fewer challenges occurred at the third decision point. We surmise that this is because the robot had limited image processing capabilities and so was unable to contradict the human in most cases.

This analysis demonstrates the rich opportunities for information exchange and shared decision making afforded by argumentation-based dialogue as implemented in our ArgHRI framework. The information-seeking dialogues dur-

	challenge		accept		reject	
(a) first decision point, *where to go*						
phys	10	(56%)	16	(89%)	2	(11%)
sim	16	(59%)	25	(93%)	2	(7%)
(b) second decision point, *how to get there*						
phys	9	(53%)	15	(88%)	2	(12%)
sim	13	(59%)	17	(77%)	5	(23%)
(c) third decision point, *what is found there* for the first five candidate treasures.						
phys 1	1	(4%)	23	(96%)	1	(4%)
phys 2	7	(27%)	20	(77%)	6	(23%)
phys 3	3	(13%)	20	(87%)	3	(13%)
phys 4	5	(28%)	13	(72%)	5	(28%)
phys 5	3	(23%)	10	(77%)	3	(23%)
sim 1	5	(19%)	23	(88%)	3	(12%)
sim 2	9	(28%)	27	(84%)	5	(16%)
sim 3	3	(11%)	25	(89%)	3	(11%)
sim 4	4	(21%)	16	(84%)	3	(16%)
sim 5	3	(27%)	11	(100%)	0	(0%)

Fig. 7. Analysis of argumentation-based dialogues triggered during each decision point.

ing the first two decision points showed that the human sought the robot's help. Our analysis suggests that the humans most frequently asked for the robot's help when there was a lack of information at the first decision point. The robot challenging the human during inquiry or information-seeking dialogue showed that the robot attempted to correct the human. For example, the robot had knowledge of the color of a treasure, but did not know its shape; the robot believed that the human could identify both color and shape by looking at images captured in the arena; thus, the robot initiated information-seeking dialogues. If the human chose the wrong color for the treasure, then the robot would challenge the human. If the human responded to the challenge stating she was unsure, then the information-seeking dialogue would terminate and trigger an inquiry dialogue, where the robot would propose a treasure to the human based on the color the robot believed it was. If the human agreed with the robot, then the robot successfully provided information to the human and they collaboratively identified the treasure. Our analysis suggests that humans ultimately agreed with the robot even though they frequently challenged the robot. We believe the ability to argue or challenge is crucial for successful collaboration. When the robot initiated a persuasion dialogue, this was an indication that the robot believed it had formulated a better decision than the human. For example, if the robot successfully persuaded the human to agree with its proposed path to visit an agreed-upon room, then the robot would not travel as far.

V. RELATED WORK

The robot that is capable of collaborating with a human needs to be able to make high-level decisions by communicating with human peers about joint actions [23], [24], [25]. Current research in human-robot dialogue explores a wide range of opportunities and challenges and represents diverse research areas, including robotics, multi-modal interfaces, natural language processing, spoken dia-

logue systems, human-computer interaction and human-robot interaction [26]. Thus as a broad community, we are still in the early stages of research on enabling dialogue for fluent human-robot interaction [26], [27].

We divide current work on human-robot dialogue into three categories, as follows. The "how to say it" problem concerns the *mode of delivery* (e.g., speech, gestures) and *language* in which content is expressed (e.g., natural language generation) [28], [29]. The TeamTalk project [30] explores multi-modal interaction, including spoken dialogue as well as mouse clicks and pen gestures to support human-robot collaboration for search tasks. In this architecture, a multi-agent dialogue manager helps prioritize tasks and process input from the human; but the robot does not take initiative. The "when to say it" problem concerns the *timing of delivery* and *turn-taking* issues, i.e., figuring out which partner has the floor to speak or act during a conversation [31]. The CoBOT project investigates situations where a robot asks a human for help [32], exploring when it is appropriate to ask a human for assistance. The "what to say" problem concerns the selection of *content* and can be considered from an abstract, conceptual level, aligning with the types of logic-based belief representation systems frequently employed in many AI and HRI systems. For example, it is common for a system to represent states and actions using predicates like put(A,B), which is then translated using a script template into an English sentence like Put box A on top of box B.

Fischer [33] studied how human-robot dialogue can be designed to reduce uncertainty about joint tasks and robot capabilities. The results of an experiment involving 22 human participants confirmed that the content of feedback positively affects human-robot interaction by lowering user uncertainty during interaction. The author concluded that user expectations of robot capabilities and appearance affect not only human-robot relationships but also human-robot dialogues. Lee & Makatchev [34] analyzed the dialogue content from interactions with a robot receptionist to investigate the types of dialogues used by human participants. The results indicated that 41.54% of the dialogues were task-specific questions (e.g., location of offices or where to get a taxi) and related to seeking information. About 30% of the dialogues were related to chatting about the robot, 20% of the dialogues were related to greetings (saying hello), and 10% of the dialogues related to impolite behaviors (insulting the robot).

Very few researchers have implemented and tested interactive systems that employ computational argumentation-based dialogue [35], [36]. To the best of our knowledge, we are the first to demonstrate this capability with physical robots.

VI. DISCUSSION

Humans seek help when they are uncertain about something. Results from our user studies suggest that most human participants did not know where to search, were without clues, and as a result, sought help during the where-to-search dialogue. This theory was validated by our analysis of the dialogue data showing who engaged in information-seeking dialogues. On average, 91% of the dialogues that

occurred during the first decision point ended in agreement, even though the robot was challenged 57% of the time.

In contrast, humans are reluctant to get help when they feel certain about something. For instance, results from our user studies reported that robot peers challenged humans, on average, more than half the time at the first two decision points, and 20% of the time at the third decision point. At the third decision point, the challenges were intended to prevent the team from mistakenly identifying a treasure incorrectly. Note that this includes repeated challenges that continued until a dialogue terminated. Less than 15% resulted in rejection. The robot in our set-up could only recognize colors and did not know how to detect treasures, so it appeared to be less confident about the image analysis evidence it provided to humans. The results suggest that robots may have failed to prevent human errors in these cases because the human participants were more confident about their selections. It may also be that humans unwarrantedly trust their judgment over the robot. However, if humans think they know something but are not absolutely sure, or there is evidence presented that suggests otherwise, they will change their minds. For instance, results from our user studies demonstrated that the robot persuaded 85% of their human teammates to change course during the where-to-search and how-to-get-there discussions.

VII. CONCLUSION

The detailed analysis presented here contributes to the complete story behind our empirical exploration of the influence of computational argumentation-based dialogue to facilitate shared decision making in human-robot teams. While the task performance metrics described in [8] demonstrate that our methodology works, the analysis here illustrates *how* it works. The positive impact of our computational argumentation-based dialogue model as a structured means to exchange the reasons for making decisions has been clearly demonstrated. Future work will explore extending the methodology to multi-robot and multi-human environments.

REFERENCES

[1] H. J. Levesque, P. R. Cohen, and J. H. T. Nunes, "On Acting Together," in *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI)*, 1990.
[2] B. J. Grosz and L. Hunsberger and S. Kraus, "Planning and Acting Together," *AI Magazine*, vol. 20, no. 4, 1999.
[3] M. Tambe, "Towards Flexible Teamwork," *Journal of Artificial Intelligence Research*, vol. 7, 1997.
[4] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proc of CHI*, 1999.
[5] D. Walton and E. C. W. Krabbe, *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, 1995.
[6] S. Parsons, M. Wooldridge, and L. Amgoud, "Properties and complexity of formal inter-agent dialogues," *Journal of Logic and Computation*, vol. 13, no. 3, 2003.
[7] E. I. Sklar and M. Q. Azhar, "Argumentation-based dialogue games for shared control in human-robot systems," *Journal of Human-Robot Interaction*, vol. 4, no. 3, 2015.
[8] M. Q. Azhar, "Toward an argumentation-based dialogue framework for human-robot collaboration," Ph.D. dissertation, City University of New York, 2015.
[9] T. J. M. Bench-Capon and P. E. Dunne, "Argumentation in artificial intelligence," *Artificial Intelligence*, vol. 171, no. 10-15, 2007.
[10] I. Rahwan and G. R. Simari, Eds., *Argumentation in Artificial Intelligence*. Springer, 2009.
[11] E. I. Sklar and S. Parsons, "Towards the Application of Argumentation-based Dialogues for Education," in *Proc of AAMAS*, 2004.
[12] P. McBurney and S. Parsons, "Games that agents play: A formal framework for dialogues between autonomous agents," *Journal of Logic, Language, and Information*, vol. 11, no. 3, 2002.
[13] E. I. Sklar, M. Q. Azhar, S. Parsons, and T. Flyr, "A Case for Argumentation to Enable Human-Robot Collaboration (Extended Abstract)," in *Proc of AAMAS*, 2013.
[14] P. McBurney and S. Parsons, "Representing epistemic uncertainty by means of dialectical argumentation," *Annals of Mathematics and Artificial Intelligence*, vol. 32, no. 1–4, 2001.
[15] H. Prakken, "Formal systems for persuasion dialogue," *Knowledge Engineering Review*, vol. 21, no. 2, 2006.
[16] P. McBurney and S. Parsons, "Dialogue Game Protocols," in *Communication in Multiagent Systems*. Springer, 2003, vol. LNCS 2650.
[17] ——, "Dialogue games for agent argumentation," in *Argumentation in Artificial Intelligence*. Springer, 2009.
[18] Y. Tang, E. I. Sklar, and S. Parsons, "An argumentation engine: ArgTrust," in *ArgMAs Workshop at AAMAS*, 2012.
[19] E. I. Sklar, A. T. Özgelen, J. P. Muñoz, J. Gonzalez, M. Manashirov, S. L. Epstein, and S. Parsons, "Designing the HRTeam Framework: Lessons Learned from a Rough-and-Ready Human/Multi-Robot Team," in *ARMS Workshop at AAMAS*, 2011.
[20] B. Gerkey, R. T. Vaughan, and A. Howard, "The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems," in *Proc of the Intl Conf on Advanced Robotics (ICAR)*, 2003.
[21] E. G. Jones, B. Browning, M. B. Dias, B. Argall, M. Veloso, and A. Stentz, "Dynamically formed heterogeneous robot teams performing tightly-coordinated tasks," in *Proc of ICRA*, 2006.
[22] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimal cost paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, 1968.
[23] J. Scholtz, "Theory and evaluation of human robot interactions," in *Proc of the Annual Hawaii Intl Conf on System Sciences*, 2003.
[24] G. Hoffman and C. Breazeal, "Collaboration in human-robot teams," in *Proc of the AIAA Intelligent Systems Technical Conf*, 2004.
[25] C. Breazeal, G. Hoffman, and A. Lockerd, "Teaching and working with robots as a collaboration," in *Proc of AAMAS*, 2004.
[26] B. Mutlu, S. Andris, and S. Allison, "Enabling human-robot dialogue," in *Robots that Talk and Listen: Technology and Social Impact*, J. Markowitz, Ed. Walter de Gruyter GmbH & Co KG, 2015.
[27] J. Peltason and B. Wrede, "The curious robot as a case-study for comparing dialog systems," *AI Magazine*, vol. 32, no. 4, 2011.
[28] B. Mutlu, "Designing embodied cues for dialog with robots," *AI Magazine*, vol. 32, no. 4, 2011.
[29] R. Simmons, M. Makatchev, R. Kirby, M. K. Lee, I. Fanaswala, B. Browning, J. Forlizzi, and M. Sakr, "Believable robot characters," *AI Magazine*, vol. 32, no. 4, 2011.
[30] M. R. Marge, A. K. Pappu, B. Frisch, T. K. Harris, and A. I. Rudnicky, "Exploring spoken dialog interaction in human-robot teams," in *Proc of Robots, Games, and Research: Success stories in USARSim IROS Workshop*. ACM, 2009.
[31] A. Thomaz and C. Chao, "Turn Taking Based on Information Flow for Fluent Human-Robot Interaction," *AI Magazine*, vol. 32, no. 4, 2011.
[32] S. Rosenthal, J. Biswas, and M. Veloso, "An effective personal mobile robot agent through symbiotic human-robot interaction," in *Proc of AAMAS*, 2010.
[33] K. Fischer, "How people talk with robots: Designing dialogue to reduce user uncertainty," *AI Magazine*, vol. 32, no. 4, 2011.
[34] M. K. Lee and M. Makatchev, "How do people talk with a robot?: an analysis of human-robot dialogues in the real world," in *Proc of CHI*. ACM, 2009.
[35] A. Toniolo, H. Li, T. J. Norman, N. Oren, W. R. Ouyang, M. B. Srivastava, T. Dropps, J. A. Allen, and P. Sullivan, "Enabling intelligence analysis through agent-support: The CISpaces toolkit," in *Proc of AAMAS*, 2015.
[36] E. I. Sklar, S. Parsons, Z. Li, J. Salvit, S. Perumal, H. Wall, and J. Mangels, "Evaluation of a trust-modulated argumentation-based interactive decision-making tool," *Journal of Autonomous Agents and Multi-Agent Systems*, vol. 30, no. 1, 2016.