A Case for Argumentation to Enable Human-Robot Collaboration

Elizabeth Sklar[1,2], Mohammad Q. Azhar[2], Todd Flyr[2], and Simon Parsons[1,2]

[1]Brooklyn College and [2]The Graduate Center
The City University of New York, New York, USA
sklar@sci.brooklyn.cuny.edu,mqazhar@gmail.com,
twf@fastmail.fm,parsons@sci.brooklyn.cuny.edu

Abstract. A case is made for *logical argumentation* as a means for enabling true collaboration between human and robot partners. The majority of human-robot systems involve interactions in which a human requests a robot to perform a task, and the robot reports its findings. The relationship between human and robot is one in which the robot is subordinate, and all high-level decision making is performed by the human. In contrast, when humans collaborate with each other, they interact in various relationships, some of which are subordinate, while others are truly collaborative. Successful instances of such relationships involve *dialogue* in which each party presents ideas, these are discussed, and a shared conclusion is agreed upon. This type of dialogue, which promotes dynamic exchange of ideas, does not exist in today's human-robot systems. Indeed the primary focus in human-robot dialogue is on the method of delivery, while the content is typically chosen from scripted sequences. However, in order to enable human-robot partnerships, both parties must be able to participate in constructive dialogue where the content and sequence of utterances can adjust dynamically as the discussion ensues. Argumentation is a method that can support such needs, as is demonstrated here.

1 Introduction

Humans interact with each other in a range of relationships, some of which are subordinate, such as boss-employee or parent-child, and others are collaborative, such as two lumberjacks each holding one end of a two-man cross-cut saw or two software engineers engaged in pair-programming. In many productive human-human relationships, the skills of each human complement each other, for example, a graphic designer and a web programmer collaborating on building a web site, or a composer and lyricist collaborating on writing a hit song. Each of these relationships, to be successful and productive, relies on some amount of communication—*dialogue*—in which each party presents their ideas, which are discussed together, and a shared conclusion is agreed upon by both parties.

In contrast, the vast majority of human-robot relationships are ones in which the human is the master and tells the robot what to do. For example, a workplace

robot might be asked to deliver letters or packages in an office building, a nurse robot might be asked to dispense medication to a patient, or a search-and-rescue robot might be asked to explore a region inside a damaged building. In each of these cases, the need for communication is limited to the human commanding or requesting a robot to perform a task, and the robot reporting its findings to the human. There is no exchange of ideas, hence no dialogue in the formal sense described above.

This type of limited exchange puts many restrictions on the human-robot interaction. For example, if a robot fails at its assigned task, it can only report to the human that it has failed; it cannot discuss the reasons for failure or possible follow-on courses of action—as two humans would when collaborating on a task. Or, if a robot disagrees with its assigned task, perhaps because it knows of a reason why the task may fail, *a priori*, before even attempting the task, the robot has no way of explaining this to the human. Additionally, the human-robot team is constrained by the human's scope of information and ideas: the robot cannot recognize new or unexpected opportunities and interrupt its task to suggest an alternate activity.

Dialogue that is founded on unscripted and opportunistic exchange of ideas does not exist in today's human-robot interaction (HRI) systems. The current focus in most human-robot dialogue work is on natural language architectures [1] or delivery methods [2–5], rather than dynamic content selection. For human-robot systems to be truly collaborative, participants need to be able to engage in constructive dialogue that can adjust dynamically as the dialogue and situation unfolds. *Argumentation* is a well-founded theoretical method that can support such needs. *Argumentation-based dialogue* can be used to handle the kinds of example situations listed above: recovering from failure, pre-empting failure, and revising plans dynamically. In this paper, we make a case for argumentation in order to enable true human-robot collaboration.

2 Architecture

This work contributes to multiple areas: argumentation, human-robot interaction and human-agent interaction, by filling in the details for combining and implementing theoretical models of logical argumentation and argumentation-based dialogue in a dynamic, real-time setting. While there is a large literature discussing the theoretical underpinnings of argumentation [6, 7] and dialogue [8], the only implemented systems are off-line decision-making tools [9, 10]. In contrast, the system described here outlines an end-to-end solution, which is necessary for an actual implementation and addresses questions such as how and when to update the beliefs of an agent engaged in a real-time dialogue, and how and when to initiate a dialogue.

Figure 1 illustrates a classic three-step agent controller architecture [11]: first, an agent senses its environment; second, the agent formulates a plan about what to do; and third, the agent acts out its plan; then the process loops back to the first step. Although modern architectures frequently employ a less sequential

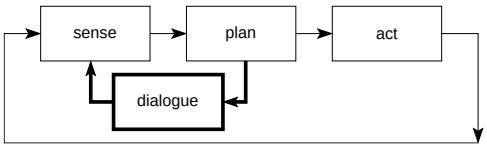

Fig. 1. Agent control architecture, with dialogue step added.

strategy, these three fundamental components are widely used. In our work, we are concerned with situations in which an agent incorporates *dialogue* in its decision-making process about what to do, in order to collaborate with a human partner[1]. Thus, we extend the classic architecture by adding a dialogue step. This dialogue step could be considered part of, or separate from, the planning step. For now, we take the easier course of considering it separately, but in the future will look at ways to build plans that combine robot actions and speech acts [12] that support argumentation-based dialogues.

As shown in Figure 1, an inner loop is added to the classic architecture, for the agent to sense the environment again after dialogue: since the agent's environment is dynamic, conditions may change during a possibly lengthy dialogue. If no (significant or relevant) changes occur, then the return loop through sense and plan after dialogue will be redundant: no changes will be deemed necessary and the agent will execute the original plan. However, if changes have occurred, then re-planning will be required. Overall, it is less costly to re-sense and re-assess the original plan than to attempt a plan that is no longer valid.

Next, we explain each step and introduce some notation.

0. Agent Ag starts with an initial belief state, $Ag.\Sigma_0$, where Σ_t represents the agent's beliefs at time t.
   (Beliefs are discussed in more detail in Section 3.)
1. Agent Ag observes its environment, at time t, using its sensors:

$$obs_t \leftarrow Ag.sense(Env_t)$$

   and then updates its beliefs, based on its observations:

$$Ag.\Sigma_t \leftarrow update(Ag.\Sigma_{t-1}, obs_t)$$

2. Agent Ag plans which action to perform:

$$Ac_t \leftarrow action()$$

3. Agent Ag performs the selected action, Ac_t.
4. Go to step 1.

[1] The process outlined here could also be applied to agent-agent dialogues, but to focus the scope of this paper, we only discuss human-agent interactions here.

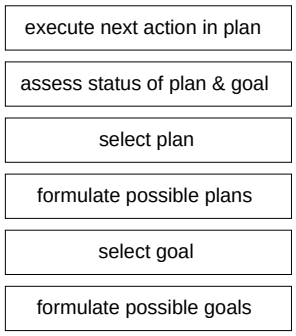

Fig. 2. A cognitive architecture to support collaboration, which implements a stack of hierarchically ordered shared decisions.

Step 2 is extended in our context as follows. The choice of action is based on executing that action in order to achieve a plan, which is selected in order to achieve a goal. In a collaborative human-robot setting, the robot and human should reach agreement about the goal(s) that they are trying to achieve, and then about the plan that they will attempt in order to achieve their joint goal. We model this joint process using a cognitive architecture based loosely on FORR [13] and the BDI model [14]. Essentially, the process is broken down into hierarchically ordered decisions, each of which must be accepted by both participants (human and robot) before considering the next decision. This process is implemented in the robot as a stack, as illustrated in Figure 2. Each time the robot's control loop iterates, the items on the stack are considered, starting at the top. If any item is triggered, then it executes and the control loop iterates again; otherwise that item is popped off the stack and the next item (below it on the stack) is considered.

The first time the control loop runs, there are no goals or plans. This means that the top five items on the stack pop off, and the formulate possible goals item is triggered. The human and robot collaborate to agree on a set of goals, and $\mathcal{G}oals$ is instantiated.

The second time the control loop runs, the top four items on the stack pop off because there is no goal selected and thus no plans; so the select goal item is triggered. Here, the human and robot collaborate to select $g \in \mathcal{G}oals$.

The third time the control loop runs, the top three items on the stack pop off because—although there is now a goal (g)—there are no plans yet defined to achieve that goal. So, the formulate possible plans is triggered, and the human and robot define a set of possible $\mathcal{P}lans$ to achieve g.

In the fourth iteration of the control loop, the top two items on the stack pop off because a particular plan, p, has not yet been selected. The select plan item is triggered, and the human and robot interact to agree on which $p \in \mathcal{P}lans$ will be attempted.

The next iteration of the control loop calls for the human and robot to assess the status of g and p, with respect to the current state of the environment. If g is no longer valid, then it is discarded and so is p; and the next time through the control loop, the top five items will pop off the stack and the process will start over with selecting a new goal. If p is no longer valid (but g is), then the process will start over with selecting a new plan. If g and p are both valid, then the top of the stack will trigger, and the robot will execute next action in plan, matching the conclusion of Step 2:

$$Ac_t \leftarrow action()$$

In order for the human and robot to reach agreement about goals and plans, they engage in a *dialog game* [8] in which they exchange locutions according to a protocol. The set of dialogues we propose for human-robot interaction are covered in the next section.

3 Dialogue

We model HRI dialogues between two agents: R, the robot, and H, the human. First, we define some notation, and then we explain how different types of dialogues from the literature can be used to facilitate exchange of ideas between human and robot, with the aim of reaching agreement (though this is not always the outcome).

The following information is represented by the robot:

- Δ_R is the set of beliefs that the robot has about its environment and about the world in general
- $\Gamma_R(H)$ is the set of beliefs that the robot has about the human, that is what the robot believes that the human believes
- CS_R is the robot's "commitment store" [15], the set of propositions that have been put forth in the dialogue by the robot
- CS_H is the human's commitment store, the set of propositions that have been put forth in the dialogue by the human
- $\mathcal{G}oals_R$ is the set of robot's goals
- $\mathcal{P}lans_R$ is the set of robot's plans

We define:

$$\Sigma_R = \Delta_R \cup \Gamma_R(H) \cup CS_R \cup CS_H \cup \mathcal{G}oals_R \cup \mathcal{P}lans_R$$

as the complete set of information that the robot can use in the dialogue. We note that Σ_R may be inconsistent. Following generally-accepted rules of dialogue games [8], the robot is only allowed to utter locutions that make use of information from Σ_R. We make the assumption that the human participant follows the same rules.

Note that we only represent the beliefs of the robot—we do not pretend to know what the human is thinking. However, $\Gamma_R(H)$ can be a proxy for the

human's beliefs, since it represents what the robot believes that the human believes. These beliefs are acquired through dialogue, based on what the human says to the robot. Conceptually, we can say that:

$$\mathcal{G}oals \subseteq \mathcal{G}oals_R \cup \mathcal{G}oals_H$$

with no loss of credibility, since $\mathcal{G}oals$ is the set of goals that the human and robot have agreed upon, and $\mathcal{G}oals_H \subseteq CS_H$ can be the subset of the human's goals that the human has discussed with the robot (which are necessarily in CS_H, the human's commitment store). We say "subset" because we assume that the human typically does not mention to the robot *every* goal in her head, though she could (which this framework also allows). Similarly, we can define:

$$\mathcal{P}lans \subseteq \mathcal{P}lans_R \cup \mathcal{P}lans_H$$

where $\mathcal{P}lans$ is the set of plans that the human and robot have agreed upon, and $\mathcal{P}lans_H \subseteq CS_H$ can be the subset of the human's plans that the human has discussed with the robot.

Each type of dialogue affects the beliefs or the actions of the robot. While the robot's full set of beliefs is represented as Σ_R, a single belief is denoted b, where $b \in \Delta_R \cup \Gamma_R(H)$. The full set of possible actions that the robot or the human could perform is represented as $\mathcal{A}ctions$, while a single action is denoted a, where $a \in \mathcal{A}ctions$, $\mathcal{A}ctions_R \subseteq \mathcal{A}ctions$ is the set of actions the robot is able to perform, $\mathcal{A}ctions_H \subseteq \mathcal{A}ctions$ is the set of actions the human is able to perform, and a sequence of actions represents a plan, $\{a_0, a_1, ..., a_{n-1}\} = p$ where $p \in \mathcal{P}lans$ (as in the previous section). Similarly (as previously), the set of possible goals is represented as $\mathcal{G}oals$, while a single goal is denoted g, where $g \in \mathcal{G}oals$.

The remainder of this section describes the pre-conditions, outcomes and protocols for each type of dialogue, mentioned above, that is relevant in an HRI scenario. The pre-conditions and outcomes express how the relevant components of the robot's belief set are considered and updated. Note that the updating of beliefs is treated abstractly here. We make the assumption that when the dialogues described below terminate with *accept*, then beliefs are updated; but we leave more complete discussion of this aspect for later. Indeed, a detailed discussion of belief revision is beyond the scope of this paper; but this topic is being investigated in our work (and many others') and will be discussed in future work.

These definitions are helpful, as will be seen in Section 5, in deciding which dialogue to apply in a given situation. As stated earlier, we describe a system that is used *by a robot* to decide which dialogue to use and what to say during the dialogue; the focus, below, is on the robot's mental state.

3.1 Dialogues for discussing Beliefs

Three types of dialogues allow the human and robot to discuss their beliefs. These are:

- *Information-seeking* [16]: where one agent asks another agent a question that the first agent believes the other agent can answer;
- *Inquiry* [17]: where two agents collaboratively answer a question that neither knew before the conversation; and
- *Persuasion* [18]: where one agent tries to alter the beliefs of another agent (which could result in either adding new information or revising existing information in the other agent's set of beliefs).

Each is discussed below.

An Information-seeking dialogue is used when the robot asks a question of the human, or the human asks a question of the robot. The agent posing the question should have reason to believe that the respondent will know the answer. Hence the second pre-condition specifies that the belief being inquired about, b, should be contained in the robot's beliefs about the human's beliefs ($\Gamma_R(H)$) (because the robot is seeking for the human to say something about b). Since the robot begins an information-seeking dialogue about b believing that the human believes b, the robot is checking its belief rather than starting from ignorance.

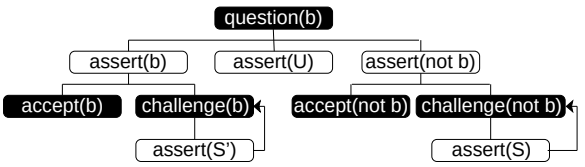

Fig. 3. Information-seeking dialogue protocol

The protocol for Information-seeking is shown in Figure 3. The diagram is read top-down, starting with the pre-conditions, then a tree illustrating the allowable sequence of locutions, then the outcomes. In the tree, the first locution (at the top) is uttered by the participant who initiates the dialogue. The next layer down contains the possible responses by the other participant, and so on. The (possible) locutions uttered by the initiator are outlined in green, whereas the possible responses uttered by the other participant in the dialogue are outlined in black, to make it easier to keep track of which speaker can say what. Some locutions cause the dialogue to terminate, such as accept(b). Other locutions cause the dialogue to loop back, such as assert(S), where S is the support for an argument. For example, if a participant's assertion is challenged, then that participant presents the support for its assertion, $A = (S, b)$, where S is the support for the argument that has b as its conclusion. If the challenger accepts all the elements of the support, then the conclusion is accepted. Otherwise, the

dialogue reaches an impasse where the only possible move is for one participant to repeat itself. This is taken to indicate that the dialogue terminates [19].

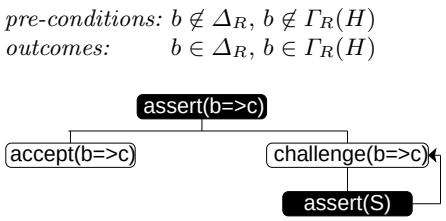

Fig. 4. Inquiry dialogue protocol

An Inquiry dialogue is used when the robot and human work together to answer a question. This will result in updating the robot's beliefs with the answer, b. The protocol for an Inquiry dialogue is shown in Figure 4. The diagram is read in the same way as Figure 3.

Finally, a Persuasion dialogue is used when the robot wants to alter the human's beliefs, or vice versa. Such a dialogue is appropriate when the robot believes b and believes the human does not believe b. This type of dialogue is helpful for error prevention: for example, when the human asks the robot to execute a plan which the robot thinks will fail. The robot will need to convince the human to abandon her belief in the plan. The protocol for a Persuasion dialogue is shown in Figure 5. Again, the diagram is read in the same way as Figure 3.

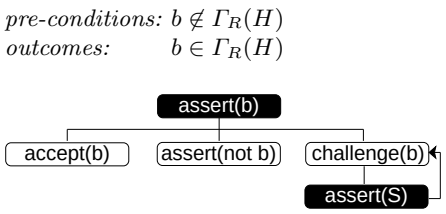

Fig. 5. Persuasion dialogue protocol

3.2 Dialogues for discussing Plans and Goals

Two types of dialogues allow the human and robot to discuss their plans and goals. These are:

- *Negotiation* [20]: where two agents attempt to reach agreement about allocation of a scarce resource; and
- *Deliberation*: [21] where agents collaboratively decide what action to take.

Each is discussed below.

A Negotiation dialogue is typically used when two agents need to reach an agreement about allocating resources. Here, we adapt this type of dialogue for allocating *tasks* between the human and robot, which could be for reaching agreement about joint plans or shared goals. Figure 6 shows the protocol for a Negotiation dialogue. Instead of discussing a belief, b, this dialogue is used to discuss a task, k. This type of dialogue also makes use of a special type of connective: $k \rightsquigarrow j$, which can be read as: *if* k *then* j, in other words, the robot might say to the human: *if you do* k*, then I will do* j [22]. The Negotiation dialogue starts with one participant requesting that the other perform a particular task, k.

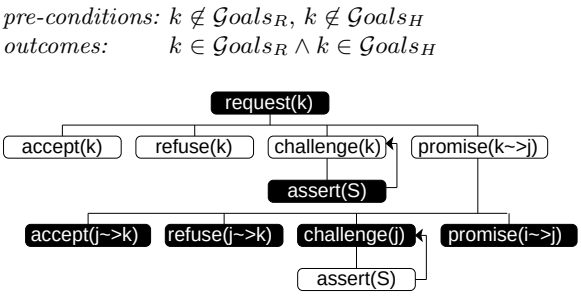

Fig. 6. Negotiation dialogue protocol

A Deliberation dialogue can be used when the human and robot need to decide on a plan. They share their intentions, i.e., set of possible plans, and together select which plan to follow. The Deliberation dialogue opens with one agent proposing that an action, a, be undertaken. The second agent can either accept(a) the proposal, agreeing to execute the specified action. Alternatively, the second agent can express a preference for a different action z, in place of that originally proposed, a: propose(z>a).

pre-conditions: $a \notin \mathcal{P}lans_R$
outcomes: $a \in \Gamma_R(H)$, $a \in \mathcal{P}lans_R$

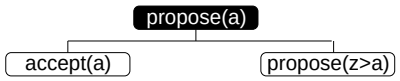

Fig. 7. Deliberation dialogue protocol

4 Formal Argumentation

The dialogues from the previous section are constructed on top of a formal system of argumentation. This allows the semantics of the locutions to be defined in terms of formal arguments, and hence related back to the contents of the robot's knowledge base. The formal argumentation system we use [23] starts with the idea that we deal with a robot which has access to a knowledge base, Σ (exactly the Σ from the previous section), containing formulae in some language $\mathcal{L}$. An *argument* is then:

Definition 1 (Argument). *An* argument *A from a knowledge base $\Delta_i \subseteq \mathcal{L}$ is a pair (S, p) where p is a formula of $\mathcal{L}$ and $S \subseteq \Delta_i$ such that:*

1. *S is consistent;*
2. *$S \vdash p$; and*
3. *S is minimal, so there is no proper subset of S that satisfies the previous conditions.*

S is called the *support*, or *grounds*, of A, and p is the *conclusion* of A. Any $s \in S$ is called a *premise* of A. The key aspect of argumentation is the association of the grounds with the conclusion. The conclusion and support are exactly those elements being exchanged in the dialogues described in the previous section.

The language, $\mathcal{L}$, that we use here is constructed from:

- $\mathcal{L}^{prop}$, a set of atomic propositions;
- $\mathcal{L}^{pref}$, a set of formulae of the form: $p_1 > p_2$, where $p_1, p_2 \in \mathcal{L}^{prop}$; and
- $\mathcal{L}^{def}$, a set of defeasible Horn clauses of the form:

$$p_1 \wedge \ldots \wedge p_n \Rightarrow c$$

where $\Rightarrow$ is defeasible (rather than material) implication and

$$p_1, \ldots, p_n, c \in \mathcal{L}^{prop} \cup \mathcal{L}^{pref}$$

Thus, $\mathcal{L} = \mathcal{L}^{prop} \cup \mathcal{L}^{pref} \cup \mathcal{L}^{def}$. Inference in this system is by a defeasible form of generalized modus ponens (DGMP):

$$\frac{p_1, \ldots, p_n \qquad p_i \wedge \ldots \wedge p_n \Rightarrow c}{c} \tag{1}$$

and if p follows from a set of formulae S using this inference rule alone, we denote this by $S \vdash^{DHC} p$.

It is typical that from the data a given individual Ag_i has about a situation, we can construct a set of arguments that conflict with each other. We might have an argument (S, p) in favor of some decision option, and another argument $(S', \neg p)$ against it (in this case we say the arguments *rebut* each other). We might also have a third argument $(S'', \neg s)$ where $s \in S$ is one of the grounds of the first argument (in this case we say that $(S'', \neg s)$ *undermines* (S, p)). Finally, we might have a fourth argument $(S''', \neg i)$ where i is one of the conclusions to one of the steps in (S, p). (This is another form of rebut, rebuttal of a sub-argument.) Argumentation provides a principled way—or rather a number of alternative ways—for Ag_i to establish which of a conflicting set of arguments are *acceptable* [24]. In other words, when an agent should use the locution accept.

As mentioned above, we use argumentation because it allows us to link what the robot says and does to what it has in its knowledge base. For example [25, 19], we can link the assert locution to the robot's knowledge base by only allowing the robot to assert propositions b that are the conclusions of acceptable arguments. Similarly, we can restrict the robot to only accept propositions for which it has an acceptable argument. (Doing this leads to some desirable properties of the dialogue, as discussed in [19].)

5 An Example

In related work [26], we have implemented a human/multi-robot framework in which a human interacts with multiple robots to achieve a collaborative task. For the purposes of collecting data from human subjects, the task is a *treasure hunt game* in which the robots explore a region inaccessible to the human, send sensor data (e.g., camera images) to the human to interpret, and the human identifies "treasure" items in the images. The robots' power is limited (i.e., battery life; treated like "health points" in a video game), and is depleted by moving and by sending images to the human. The robots (for purposes of the game) possess only minimal image processing capabilities and must rely on the human to identify anything of interest in an image. If the human believes that she has found a treasure in an image, then she can send the image to the "game master" for verification. If she is correct, then the human-robot team is rewarded with points. If she is incorrect, then she and her teammates lose points. The goal of the game is to locate as many treasures as possible before the robots run out of power. A view of the treasure hunt arena, with three robots, is shown in Figure 8. This is the view that the human player has during the game.

Following the logic described in Figure 2, the first task is for the human and robots to agree on where the robots should search. For the purposes of using

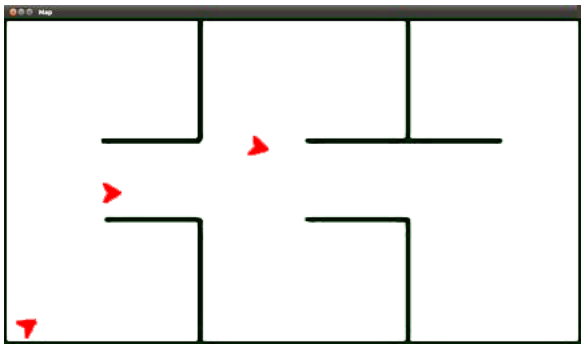

Fig. 8. Experimental arena

this game scenario to illustrate the dialogue implementation, we only consider a dialogue between the human and one of the robots[2].

In our example, the human and robot would like to achieve the goal:

$$found(treasure)$$

The robot holds the following information about the world:

$$\begin{aligned} at(treasure, region1) &\Rightarrow found(treasure) \\ at(treasure, region2) &\Rightarrow found(treasure) \\ &\dots \\ at(treasure, regionN) &\Rightarrow found(treasure) \end{aligned}$$

In order to achieve $at(treasure, region_i)$, the following must be true:

$$\begin{aligned} in(picture, treasure) \wedge \\ have(picture, region_i) \quad &\Rightarrow at(treasure, region_i) \end{aligned}$$

In order to achieve $have(picture, region_i)$, the following must be true:

$$\begin{aligned} at(robot, region_i) \wedge \\ has(robot, camera) \wedge \\ sense(camera) \quad &\Rightarrow have(picture, region_i) \end{aligned}$$

[2] Though the human/single-robot method can be applied in parallel to multiple robots, by having the human engage in multiple conversations at once—like conducting multiple "chat" sessions with different, separate people in an instant messenger client. We note that this obviously will not scale for very many robots (unless the human is a teenager, in which case her capacity to chat with multiple friends at once appears limitless to most adults), but testing the scalability is beyond the scope of this paper. We focus on the human/one-robot mode.

In order to achieve $in(picture, treasure)$, the following must be true:

$$\begin{aligned} has(human, picture) \wedge \\ analyse(picture) \wedge \\ match(picture, treasure) \quad \Rightarrow in(picture, treasure) \end{aligned}$$

The robot knows that $sense()$ is an action that it can perform, $sense() \in \mathcal{A}ctions_R$, and that $analyse()$ is an action that the human can perform, $analyse() \in \mathcal{A}ctions_H$. In solving our example, the robot determines that it has to negotiate with the human to perform the analysis, so it opens a Negotiation dialogue with the human:

$$R : request(analyse(picture))$$

The human could respond with any of the locutions in the second level of the tree in Figure 6. For the sake of the example, suppose that the human does not have a picture to analyse, so she responds with:

$$H : promise(analyse(picture)) \Rightarrow have(picture, region_i)$$

In other words: the human is bargaining by making a promise to analyse a picture if she has one. We can assume that she knows that the robot has to supply the picture, so the robot agrees:

$$R : accept(analyse(picture)) \Rightarrow have(picture, region_i)$$

At the end of this negotiation, the human and robot have agreed to achieve two tasks[3]:

$$\begin{aligned} analyse(picture) \\ have(picture, region_i) \end{aligned}$$

The next step is to formulate a plan to accomplish these tasks. The robot can take the initiative and open a Deliberation dialogue:

$$R : propose(at(robot, region1))$$

which the human will agree with:

$$H : accept(at(robot, region1))$$

unless the human doesn't want the robot to go to $region1$, so she might express a preference instead of accepting the robot's proposal:

$$H : propose(at(robot, region4) > at(robot, region1))$$

Thus the human expresses a preference for the robot to go to $region4$ instead of $region1$.

[3] We could also think of these as *subgoals*.

Now, the robot might wonder why the human has a preference for $region4$. Perhaps the human knows something about $region4$ that makes it more likely to find the treasure there. So the robot can suspend the Deliberation dialogue and open an Information-seeking dialogue in order to get more information from the human:

$$R : question(region4 > region1)$$

to which the human would assert her preference:

$$H : assert(region4 > region1)$$

which the robot would challenge:

$$R : challenge(region4 > region1)$$

forcing the human to defend her stance by presenting all her evidence from which she draws the conclusion of preferring $region4$ over $region1$ (as explained in Section 3.1).

The great advantage for using argumentation dialogues is that the rules for these interchanges have been well laid out in the literature, for example [19]. The rules force a dialogue either to terminate with agreement or end in an impasse. There is a guarantee that a dialogue will never continue infinitely, because neither participant is allowed to repeat a locution within any of the dialogue protocols. In addition, there is a formal framework that defines how to combine dialogues as we have described here [8]. Work in this area is ongoing.

6 Related Work

Scholtz [27] defines 5 different roles that humans may undertake when functioning in the same physical space as a robot: (i) bystander, (ii) supervisor, (iii) operator, (iv) mechanic or programmer, and (v) teammate. In the first case (i), the human is an observer who has no physical interaction or direct communication with the robot. In the next three cases (ii–iv), the human has a dominating role over the robot in which the human either tells the robot what to do (ii and iii) or actually programs the robot to perform a task (iv). The fifth case (v) is the instance we are interested in: the case where the robot and human interact as peers. Here, they must collaborate, which means discussing ideas about which task(s) to perform and how to perform the task(s). Just like in any effective human-human collaboration, they should reach agreement about what to do and how to do it before either partner takes any action, and, as we have shown, argumentation-based dialogue is a way to achieve this. There is little other work that has moved in the direction of cooperative human-robot interaction, and here we briefly survey what there is.

In [28] the robot is intended to assist an astronaut working on the space station. The robot handles tools, greatly improving the efficiency of the task, but the need is primarily one of convenience and there is little need for assistance on the human's part. Other robots, such as the one developed by [29], use human

teachers in a learning phase, then use that information to be more autonomous at a later stage. [30] demonstrates a robot with affect as it wanders terrain with a human but warns when its battery power is low, while [31] makes humans part of an "operating system" where a computer determines, distributes and assigns tasks to human and robot alike. In this application, while humans and robots are team members, communication between humans and robots is not a priority.

Much of the work on human-robot cooperation involves less explicit communication than we have explored here. [32] had a person take a mobile robot by its arm and lead it (by pushing and pulling) around an obstacle course. In a later phase, the human is led by the robot. In the phases where the human is leading, the robot is learning the course. The robot's sensors are of limited capability, and thus needs human assistance in learning. In other versions of the experiment, the human is blindfolded and the robot leads.

[33] takes a similar approach with a robot that leads people through an office building to attend meetings. The authors describe this as a symbiotic relationship. The robot in question is a mobile platform but it cannot open doors or serve coffee. It can also have trouble localizing. Thus the robot is programmed to ask for human help.

In work where communication is important, the focus is less on the content of the communication than the delivery. For example the robot in [30] detects negative valence in the human's voice as well as some degree of facial recognition of stress. It associates affect with its task list which helps determine priority and it can express worry in its synthesized voice. Similarly, the robot in [34] uses aspects such as eye contact, proximity, and vocal cues to more effectively persuade a human subject to do things.

7 Summary

We have presented a model for human-robot interaction that supports flexible and dynamic argumentation-based dialogue. Although this work is preliminary, the ideas contribute not only to human-robot interaction, but also to argumentation, by outlining an end-to-end framework that combines theoretical models of logical argumentation and argumentation-based dialogue applied to a real-time setting. Our next steps include investigation of belief revision, as indicated in Section 3, and specification of termination conditions for each type of dialogue. Parallel work includes an implementation of the theory outlined here and a user study, to test the efficacy of our method with human subjects in playing multiple scenarios of the treasure hunt game [35]. In addition, the framework is being applied to a human-agent interaction environment in which human users reason about uncertain information received in real-time from multiple sources.

Acknowledgments

Funding for this work was provided in part by the National Science Foundation under #11-16843 and by the US Army Research Office under the Science of

Security Lablet grant. We would like to thank reviewers of earlier versions of this paper for helpful comments.

References

1. Lemon, O., Gruenstein, A., Peters, S.: Collaborative activities and multi-tasking in dialogue systems. TAL: Special Issue on Dialogue **43**(2) (2002)
2. Kirby, R., Broz, F., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A., Wang, J.: Designing robots for long-term social interaction. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (2005) 2199–2204
3. Krestin, F.: How people talk with robots: Designing dialogue to reduce user uncertainty. AI Magazine **32**(4) (2011)
4. Peltason, J., Wrede, B.: The curious robot as a case-study for comparing dialog systems. AI Magazine **32**(4) (2011)
5. Thomaz, A., Chao, C.: Turn taking based on information flow for fluent human-robot interaction. AI Magazine **32**(4) (2011)
6. Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation **1** (2010)
7. Rahwan, I., Simari, G.R., eds.: Argumentation in Artificial Intelligence. Springer Verlag (2009)
8. McBurney, P., Parsons, S.: Games that agents play: A formal framework for dialogues between autonomous agents. Journal of Logic, Language, and Information **11**(3) (2002)
9. Fabregues, A., Sierra, C.: Dipgame: a challenging negotiation testbed. Engineering Applications of Artificial Intelligence **24**(7) (2011) 1137–1146
10. Judson, P.N., Fox, J., Krause, P.J.: Using new reasoning technology in chemical information systems. Journal of Chemical Information and Computer Sciences **36** (1996) 621–624
11. Nilsson, N.J.: Technical note no. 323. Technical report, SRI International (1984)
12. Austin, J.L.: How to do things with words. Volume 88. Harvard University Press (1975)
13. Epstein, S.L.: For the right reasons: The forr architecture for learning in a skill domain. Cognitive Science **18**(3) (1994) 479–511
14. Bratman, M.E., Israel, D.J., Pollack, M.E.: Plans and resource-bounded practical reasoning. Computational Intelligence **4**(4) (1988)
15. Hamblin, C.L.: Mathematical models of dialogue. Theoria **37** (1971) 130–155
16. Walton, D.N., Krabbe, E.C.W.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. State University of New York Press, Albany, NY, USA (1995)
17. Hulstijn, J.: Dialogue Models for Inquiry and Transaction. PhD thesis, Universiteit Twente (2000)
18. Prakken, H.: Formal systems for persuasion dialoge. Knowledge Engineering Review **21**(2) (2006)
19. Parsons, S., Wooldridge, M., Amgoud, L.: Properties and complexity of formal inter-agent dialogues. Journal of Logic and Computation **13**(3) (2003) 347–376
20. Rahwan, I., Ramchurn, S.D., Jennings, N.R., McBurney, P., Parsons, S., Sonenberg, L.: Argumentation-based negotiation. Knowledge Engineering Review **18**(4) (2003)

21. McBurney, P., Parsons, S.: A denotational semantics for deliberation dialogues. In: Proc of AAMAS. (2004)
22. Amgoud, L., Parsons, S., Maudet, N.: Arguments, dialogue and negotiation. In: Proc of ECAI. (2000)
23. Tang, Y., Cai, K., McBurney, P., Sklar, E., Parsons, S.: Using argumentation to reason about trust and belief. Journal of Logic and Computation **22)**(5) (2012) 979–1018
24. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. The Knowledge Engineering Review (2011)
25. Amgoud, L., Maudet, N., Parsons, S.: An argumentation-based semantics for agent communication languages. In: Proceedings of the Fifteenth European Conference on Artificial Intelligence. (2002)
26. Sklar, E., Parsons, S., Epstein, S.L., Özgelen, A.T., Muñoz, J.P., Schneider, E., Costantino, M., Abbasi, F., Aragon, K., Green, A., Hernandez, J., Ibraheem, I., Namalu, A., Yalkabov, S., Wan, J.: Demonstration: Investigating Human/Multi-Robot Team Interaction. In: AAAI Robotics and Multimedia Fair, Toronto, Canada (July 2012)
27. Scholtz, J.: Theory and Evaluation of Human Robot Interactions. In: Hawaii International Conference on System Science (HICSS). Volume 36. (January 2003)
28. Trafton, J.G., Cassimatis, N.L., Bugajska, M.D., Brock, D.P., Mintz, F.E., Schultz., A.C.: Enabling effective human-robot interaction using perspective-taking in robots. IEEE Transactions on Systems, Man and Cybernetics, Part A **35**(4) (2005) 460–470
29. Chrysanthakopoulos, G., Shani, G.: Augmenting appearance-based localization and navigation using belief update. In: Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems, Toronto (2010) 559–566
30. Scheutz, M., Schermerhorn, P., Kramer, J.: The utility of affect expression in natural language interactions in joint human-robot tasks. In: Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot interaction. (2006) 226–233
31. Fong, T.W., Nourbakhsh, I., Ambrose, R., R. Simmons, a.S., Scholtz, J.: The peer-to-peer human-robot interaction project. In: Proceedings of the AIAA SPACE Conference. (2005)
32. Ogata, T., Sugano, S., Tani, J.: Open-end human robot interaction from the dynamical systems perspective: Mutual adaptation and incremental learning. In: Proceedings of the International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems. (2004) 435–444
33. Rosenthal, S., Biswas, J., Veloso, M.: An effective personal mobile robot agent through symbiotic human-robot interaction. In: Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems. (2010) 915–922
34. Chidambaram, V., Chiang, Y.H., Mutlu, B.: Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In: Proceedings of the 7th ACM/IEEE Conference on Human-Robot Interaction. (2012)
35. Azhar, M.Q., Schneider, E., Salvit, J., Wall, H., Sklar, E.I.: Evaluation of an argumentation-based dialogue system for human-robot collaboration. In: Workshop on Autonomous Robots and Multirobot Systems (ARMS) at Autonomous Agents and MultiAgent Systems (AAMAS). (2013)