Argumentation-Based Dialogue Games for Shared Control in Human-Robot Systems

Elizabeth I. Sklar
Department of Informatics, King's College London
and
M. Q. Azhar
Borough of Manhattan Community College, City University of New York

Dialogue can support exchange of ideas and discussion of options as a means to enable shared decision making for human-robot collaboration. However, dialogue that supports dynamic, evidence-backed exchange of ideas is a major challenge for today's human-robot systems. The work presented here investigates the application of *argumentation-based dialogue games* as the means to facilitate flexible interaction, including unscripted changes in initiative. Two main contributions are provided in this paper. First, a methodology for implementing multiple types of argumentation-based dialogues for human-robot interaction is detailed. This includes explanation about which types of dialogues are appropriate given the beliefs of the participants and how multiple dialogues can occur simultaneously while maintaining a consistent set of beliefs for the participants. Second, a formal definition is presented for the *Treasure Hunt Game (THG)*, a test environment that provides rich opportunities for experimentation in shared human-robot control, as well as motivating and engaging experiences for human subjects.

Keywords: human-robot interaction, argumentation, argumentation-based dialogue

1. Introduction

Humans interact with each other in many types of relationships, ranging from *subordinate*, where one person instructs or commands another, to *collaborative*, where the skills of one person complement those of another. In a subordinate relationship, the leader takes responsibility for making decisions about joint actions and actions that affect others. In contrast, partners in collaborative relationships share decision making. They exchange ideas and discuss options, and they jointly arrive at decisions about dependent and related actions. Such shared decision making is enabled using conversation—*dialogue*—that allows each partner to communicate ideas and adjust their beliefs according to new and/or contrasting ideas presented by others.

Most human-robot relationships today are subordinate, where a human leader maintains the locus of control and effectively tells the robot what to do. The human leader sets overall goals and assigns to the robot tasks to achieve those goals; and the robot then defines its own series of subgoals in order to accomplish its assigned tasks. For example, a human leader may tell a robot to go to a

Journal of Human-Robot Interaction, Vol. 4, No. 3, 2015, Pages 120-148. DOI 10.5898/JHRI.4.3.Sklar

Domain	*Human user*	*Robot tasks*
Search-and-rescue (Murphy, Casper, & Micire, 2001) (Yanco et al., 2006)	First responder	Search for victims; communicate with victims; find safe path to victim for first responders
Humanitarian de-mining (Santana, Barata, & Correia, 2007) (Habib, 2007)	NGO worker	Find mines; find safe path to mine for demining specialist
Manufacturing (Alers, et al., 2014)	Factory worker	Assemble products
Health aid (Matthews, 2002)	Patient	Administer medication; assist with physical therapy
Geriatric companion (Wada, Shibata, Saito, & Tanie, 2002)	Elderly person	Administer medication; observe behaviour; engage in exercise; read out loud; answer telephone/door
Tutor (Castellano et al., 2013)	Student	Play educational games; encourage learning activities

Table 1: Example domains, users and tasks found in HRI literature.

particular location, and the robot will execute its own path-planning behaviour to select waypoints and its own motion behaviour to travel to each waypoint. However, this mode of interaction limits the robustness of the human-robot partnership, because it does not take full advantage of the robot's sensory and/or processing potential. If a robot fails at its assigned task, it will typically only report that failure has occurred and not (be able to) elaborate on the reason(s) for failure. In the example above, if the robot cannot go to the location assigned by the human because there is a large obstacle blocking access, the robot cannot engage the human in discussion about alternative goals.

Dialogue that facilitates opportunistic exchange of ideas is not well supported in today's human-robot systems. Current work on dialogue in the *human-robot interaction (HRI)* community is focused on challenges in *natural language* dialogue systems, such as architectures (Bohus, Raux, Harris, Eskenazi, & Rudnicky, 2007; Lemon, Gruenstein, & Peters, 2002) and multimodal delivery methods (Bohus, Horvitz, Kanda, Mutlu, & Raux, 2011; Modayil, 2010; Torrey, Powers, Marge, Fussell, & Kiesler, 2006). However, for HRI systems to be truly collaborative, participants must be able to engage in opportunistic dialogue that can adjust dynamically as the situation unfolds. 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* (Carbonell, 1970; Horvitz, 1999) in an ongoing or new conversation.

Within the domains and situations typically explored in the HRI literature, we identify three specific cases where the ability to exchange of ideas opportunistically would broaden the scope of human-robot capabilities and improve success rates: (1) responding to discovery, (2) pre-empting failure, and (3) recovering from failure. Illustrative examples of domains, tasks, and users commonly found in the HRI literature are listed in Table 1.

In response, we investigate the application of *argumentation-based dialogue games* as the means to facilitate opportunistic exchange of ideas. *Argumentation* (Rahwan & Simari, 2009) is a well-

founded theoretical method, based in logic, in which agents put forth claims and produce evidence that support (or attack) the claims. Argumentation is extensively explored within the multi-agent systems community. *Argumentation-based dialogue* (Prakken, 2006; McBurney & Parsons, 2002; Hulstijn, 2000; Walton & Krabbe, 1995) is a formal system in which agents exchange arguments with specific goals in mind respecting what the dialogue should achieve. A *persuasion dialogue* (Prakken, 2006) is where one agent tries to alter the beliefs of another agent. An *information-seeking dialogue* (Walton & Krabbe, 1995) is where one agent asks a question for which it believes the other agent knows the answer. An *inquiry dialogue* (McBurney & Parsons, 2001) is where two agents collaboratively seek the answer to a question for which neither knows the answer. In this paper, we demonstrate how these three types of dialogue can be used individually or in combination to address the needs cited above (responding to discovery, pre-empting failure and recovering from failure). While the argumentation-based dialogue literature provides formal definitions for these types of dialogue and proposes rules for how each might implemented in isolation, there is no comprehensive, implemented system that supports all types of dialogue and allows agents to interleave partial dialogues. In addition, aside from our preliminary work (Sklar, Azhar, Parsons, & Flyr, 2013), logical argumentation has not been applied to human-robot interaction. Our contribution here is three-fold: (1) We provide a methodology for implementing multiple types of dialogues; (2) we detail how multiple dialogues can occur simultaneously, while maintaining a consistent set of beliefs for the agents engaged in the dialogue(s); and (3) we demonstrate how our method can be applied to extend the current capabilities of HRI systems.

2. Background: Argumentation Theory

In this section, we provide the essential technical background on *argumentation theory* that we will need to demonstrate how argumentation-based dialogue can be used to extend current HRI capabilities. We use the formal system from Parsons, Wooldridge, and Amgoud (2003a) and Parsons, McBurney, Sklar, and Wooldridge (2007).

2.1 Argumentation

An agent Ag maintains a set of beliefs, Σ, containing formulae from a *propositional language*, $\mathcal{L}$. $\mathcal{L}$ contains *atomic propositions*, p_i, which are individually either `true` or `false`. An *inference mechanism* $\vdash_{\mathcal{L}}$ is associated with $\mathcal{L}$, such that

$$S \vdash_{\mathcal{L}} c$$

means that c can be proven from S using rules and propositions contained in the language $\mathcal{L}$. A rule

$$p_1 \wedge p_2 \wedge \ldots \wedge p_n \rightarrow c$$

derives, or proves, an agent's conclusion c when every p_i listed in the rule is either a member of Σ, or can be derived as the conclusion of another member of Σ. The agent's set of beliefs, Σ, may be inconsistent; in other words, Σ may contain both p and $\neg p$ (not p; i.e., if p is true, then $\neg p$ is false).

Definition 1 (Argument) *An **argument** A is a pair (S, c) where c and $S = \{s_1, s_2, \ldots, s_n\}$ are formulae of some language $\mathcal{L}$ and S is a subset of Σ, such that:*

1. *S is consistent;*
2. *$S \vdash_{\mathcal{L}} c$; and*
3. *S is **minimal**, meaning that no proper subset of S satisfying both (1) and (2) exists.*

*S is called the **support** of A; and c is the **conclusion** of A.*

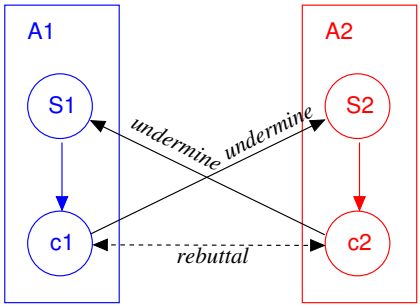


Figure 1. Forms of **attack** between arguments: c1 rebuts c2, and, symmetrically, c2 rebuts c1; c1 undermines S2; and c2 undermines S1.

We can also define an argument in terms of *evidence* by stating that S is the set of evidence in Σ that supports the conclusion c. Thus, an *argument* is a logical entity that consists of both the conclusion and the evidence supporting that conclusion. Formally, the support is a consistent minimal set of formulae from which the conclusion can be derived using some inference mechanism. We write $\mathcal{A}(\Sigma)$ to denote the set of all possible arguments that could be made from Σ.

Since Σ may be inconsistent (as mentioned earlier), arguments in $\mathcal{A}(\Sigma)$ may conflict. We identify two ways in which arguments may conflict: (1) *undermining*—where the conclusion of one argument conflicts with some element in the support of another argument; and (2) *rebuttal*—where the conclusion of one argument conflicts with the conclusion of another argument. These are generally called *attack* relations between arguments and are illustrated in Fig. 1.

Arguments can also *support* each other. We identify two ways in which arguments may offer support (Cohen, Parsons, Sklar, & McBurney, 2014): (1) *premise-support (p-support)*—where one argument is part of the support for another argument; and (2) *conclusion-support (c-support)*—where two non-intersecting sets of propositions support the same conclusion. These are illustrated in Fig. 2.

Formal definitions for these four concepts are listed below. In all cases, let $A_1 = (S_1, c_1)$ and $A_2 = (S_2, c_2)$ be arguments in $\mathcal{A}(\Sigma)$.

Definition 2 (Rebuttal) *A_1 **rebuts** A_2 iff $\neg c_2 \equiv c_1$. Symmetrically, A_2 rebuts A_1, since $\neg c_1 \equiv c_2$.*

Definition 3 (Undermine) *A_1 **undermines** A_2 iff there is some $p \in S_2$ such that $\neg p \equiv c_1$. (Prakken, 2010)*

Definition 4 (C-support) *A'_1 **c-supports** argument $A_1 = (S_1, c_1)$ iff there is some argument $A'_1 = (S'_1, c_1) \in \mathcal{A}(\Sigma)$ such that $S'_1 \cap S_1 = \emptyset$. (Cohen et al., 2014)*

Definition 5 (P-support) *A_1 **p-supports** A_2 iff there is some $p \in S_2$ such that $p \equiv c_1$. (Cohen et al., 2014)*

Next, we apply these definitions, particularly the two forms of attack, to the notion of *acceptability*. That is, if an argument is attacked, can it still be accepted as a valid argument? There are quite a number of different methods in the argumentation literature for computing acceptability (Prakken, 2010), some of which are based on the notion of preferences (Modgil & Prakken, 2013) between attacks. For example, when one piece of evidence comes from a more trusted source than another piece of evidence, an agent may be more inclined to believe the evidence from the more trusted

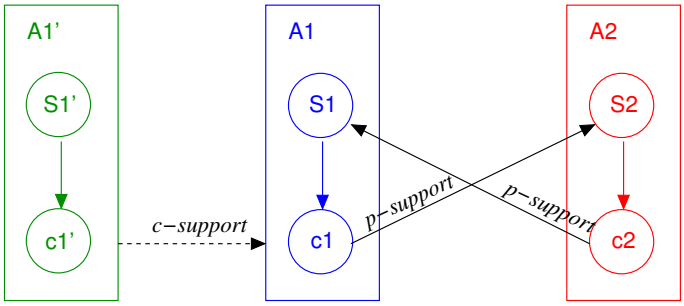


Figure 2. Forms of **support** between arguments: c1 $\in$ S2 and thus p-supports c2; c2 $\in$ S1 and thus p-supports c1; S1' c-supports c1, where S1 $\cap$ S1' $= \emptyset$.

source and hence prefer arguments supported by that evidence over other arguments supported by weaker evidence (Sklar, Parsons, & Singh, 2013). Detailed discussion of acceptability is beyond the scope of this paper, but the general concept is necessary for what follows.

2.2 Argumentation-Based Dialogue

In an **argumentation-based dialogue**, two (or more) agents participate in a structured interaction following a set of rules. The basic rules of a two-agent[1] argumentation-based dialogue state that:

1. Agents take turns putting forth *utterances*, alternating between them.
2. The agent that presents the first utterance is the agent with the *initiative* in the dialogue.
3. All utterances are constructed from the agents' beliefs.
4. The *axiomatic semantics* (McBurney & Parsons, 2009) of each *type* of dialogue dictate which utterance(s) can be invoked by each participant at distinct points during the interchange.
5. No two utterances can be repeated (i.e., agents cannot "say" the same thing twice).

This last rule is important because it guarantees that all dialogues must terminate in either *agreement*, *disagreement*, or *stalemate*.

In the context of human-robot dialogue, we model the robot's set of beliefs as $R.\Sigma$. According to the argumentation-based dialogue rules, we make the assumption that the robot operates within the constraints of its beliefs—in other words, it does not perform an action that it does not know how to perform, and it cannot "say" anything about concepts it does not know about. For example, we assume that a ground-based robot cannot fly because it does not possess motors that will lift it off the ground, or that a robot equipped with only sonar sensors (and no camera) will be unable to detect colors. However, we also assume that a robot can *learn* new things, such as complex actions comprised of atomic actions that it already knows how to perform, or color properties of objects that a trusted collaborator can detect and provide reliable information about. We very loosely express the notion of learning by saying that any change in the robot's beliefs, $R.\Sigma$, represents learning—with the caveat that real discussion of the myriad methods of machine learning, knowledge acquisition and belief revision are beyond the scope of this article—we only use the term "learning" here for convenience and leave extended discussion to colleagues and future work.

In addition to Σ, each agent in an argumentation-based dialogue stores the set of past utterances in the dialogue. This is referred to as its *commitment store*, CS. We think of this as an agent's "public knowledge," since it contains information that is shared with other agents. In contrast, the contents

[1] Note that these rules can be extended to dialogues involving more than two agents, but for simplicity and with respect to the human-robot context presented here, we only discuss two-agent dialogues in this article.

of Σ are "private." In the description that follows, we use Δ to denote **all** the information available to an agent, which includes Σ and CS, as well as other partitions of the agent's knowledge base (some of which are discussed below, while others are beyond the scope of this article). Thus, in an interaction between two agents, Ag_i and Ag_j, the beliefs available to the first agent are represented as $Ag_i.\Delta = Ag_i.\Sigma \cup Ag_i.CS \cup Ag_j.CS$, and the beliefs available to the second agent are represented as $Ag_j.\Delta = Ag_j.\Sigma \cup Ag_j.CS \cup Ag_i.CS$.

Further, we distinguish a subset of Σ, namely Γ (Sklar & Azhar, 2011; Sklar & Parsons, 2004), which represents an agent's beliefs about another agent (or human—i.e., any participant in a dialogue). For agent Ag_i, its beliefs about other agents, $Ag_i.\Gamma$ can be described as n additional subsets (one for each "other" agent):

$$Ag_i.\Gamma = Ag_i.\Gamma(Ag_1) \cup Ag_i.\Gamma(Ag_2) \cup \ldots \cup Ag_i.\Gamma(Ag_n)$$

where each $Ag_i.\Gamma(Ag_j)$ represents agent Ag_i's beliefs about what agent Ag_j believes. In the HRI setting, we use $R.\Gamma(H)$ to represent the robot's beliefs about what the human believes.

This is an important concept in our work, because we do not claim to know the human's beliefs. We only infer the human's beliefs from her interactions with the robot in our HRI system; thus we only represent $R.\Gamma(H)$ and do not explicitly represent $H.\Sigma$. Note that we can represent the human's commitment store, $H.CS$, since this contains the human's public knowledge—an aggregate of all the beliefs the human has put forth in the dialogue.

3. Approach: Argumentation-Based Dialogue Games

We begin our discussion of *argumentation-based dialogue games* for HRI by explaining the notation we use for describing a game between a robot, R, and a human, H:

- $R.\Sigma$ represents the robot's set of beliefs, as described in the previous section.
- $R.\Gamma(H)$ represents the robot beliefs about the human's beliefs. (As mentioned in the previous section, we do not pretend to be able to know what the human actually believes, so instead of representing the human's beliefs as $H.\Sigma$, we represent the robot's beliefs about what the human believes—i.e., beliefs for which the robot has evidence due to something the human has said or done in their interaction.)
- b represents a *belief*. For example, if the robot believes it is in location (x, y), then we could have:

$$b = \ulcorner \mathsf{at}(R, (x, y)) \urcorner$$

We use the corner quotation marks to delineate an atomic belief. Depending on context, a belief b may be atomic or may be compound. For example if a robot believes that it sees a red ball ahead, then we could have:

$$b = \ulcorner \mathsf{at}(R, (x, y)) \urcorner \wedge \ulcorner \mathsf{at}(object, (x \pm \epsilon, y \pm \epsilon)) \urcorner \wedge \ulcorner \mathsf{isa}(object, ball) \urcorner \wedge \ulcorner \mathsf{has}(object, red) \urcorner$$

- $\neg b$ represents *disbelief* in b. For example, if the robot believes it sees a red ball ahead but the human tells the robot that she believes that the object the robot sees is a red box, then b could represent the robot's belief that the object is a red ball and $\neg b$ the human's belief that the object is not a red ball:

$$\begin{aligned} b &= \ulcorner \mathsf{isa}(object, ball) \urcorner \\ \neg b &= \ulcorner \neg \mathsf{isa}(object, ball) \urcorner \end{aligned}$$

- $?b$ represents the situation where the robot or human has no information about b; so neither believes nor disbelieves b.

	$b \in R.\Gamma(H)$	$\neg b \in R.\Gamma(H)$	$?b \in R.\Gamma(H)$
$b \in R.\Sigma$	*case 1* agreement *(no dialogue)*	*case 4* disagreement *persuasion dialogue*	*case 7* lack of knowledge *information-seeking dialogue*
$\neg b \in R.\Sigma$	*case 2* disagreement *persuasion dialogue*	*case 5* agreement *(no dialogue)*	*case 8* lack of knowledge *information-seeking dialogue*
$?b \in R.\Sigma$	*case 3* lack of knowledge *information-seeking dialogue*	*case 6* lack of knowledge *information-seeking dialogue*	*case 9* shared lack of knowledge *inquiry dialogue*

Table 2: Cases for different types of dialogues

Table 2 lists the possible cases for justifying different types of dialogue between the robot and the human. The rows signify the robot's beliefs, as contained in $R.\Sigma$. The columns signify the robot's beliefs about the human's beliefs, $R.\Gamma(H)$ (per earlier discussion). The combinations condense into the following four situations:

- **agreement** (because beliefs do not conflict);
- **disagreement** (because beliefs conflict);
- **lack of knowledge** (because one of the parties in the dialogue has no knowledge about a belief, thus agreement or disagreement is not yet possible); and
- **shared lack of knowledge** (because neither party has knowledge about a belief).

Each situation is discussed below.

Agreement *(cases 1 and 5)*. Either the robot believes b, and the human believes b; or the robot believes $\neg b$, and the human believes $\neg b$. These cases are represented formally as:

$$\langle b \in R.\Sigma \rangle \wedge \langle b \in R.\Gamma(H) \rangle$$

or:

$$\langle \neg b \in R.\Sigma \rangle \wedge \langle \neg b \in R.\Gamma(H) \rangle$$

respectively. In these cases, the robot and the human *agree* about b or $\neg b$; so no dialogue is necessary.

Disagreement *(cases 2 and 4)*. Either the robot believes $\neg b$, and the human believes b; or the robot believes b, and the human believes $\neg b$. These cases are represented formally as:

$$\langle \neg b \in R.\Sigma \rangle \wedge \langle b \in R.\Gamma(H) \rangle$$

or:

$$\langle b \in R.\Sigma \rangle \wedge \neg \langle b \in R.\Gamma(H) \rangle$$

respectively. These are cases of *disagreement*, which warrants a *persuasion* (Prakken, 2006) dialogue where either the robot initiates a dialogue to convince the human to change her belief to b or $\neg b$, or the human initiates a dialogue to convince the robot to change its belief to b or $\neg b$. For example, the robot believes it sees a red ball and the human believes the robot sees a red box. The

human can initiate a persuasion dialogue to convince the robot that the object it sees is a box, by presenting evidence that the object it sees is shaped like a cube.

Lack of Knowledge *(cases 3, 6, 7 and 8).* Either the robot has no knowledge about b, and the human believes b or $\neg b$; or the human has no knowledge about b, and the robot believes b or $\neg b$. These cases are represented formally as:

$$\langle ?b \in R.\Sigma \rangle \wedge \Big\langle \langle b \in R.\Gamma(H) \rangle \vee \langle \neg b \in R.\Gamma(H) \rangle \Big\rangle$$

or:

$$\langle ?b \in R.\Gamma(H) \rangle \wedge \Big\langle \langle b \in R.\Sigma \rangle \vee \langle \neg b \in R.\Sigma \rangle \Big\rangle$$

These are cases of *lack of knowledge* on the part of either the robot or the human, which warrants an *information-seeking* (Walton & Krabbe, 1995) dialogue to be initiated by the party who is lacking knowledge. For example, the robot captures an image but cannot detect anything in the image, and the human believes there is a red box in the image. The robot can initiate an information-seeking dialogue to learn what the human sees in the image.

Shared Lack of Knowledge *(case 9).* Neither the robot nor the human has any knowledge about b. This case is represented formally as:

$$\langle ?b \in R.\Sigma \rangle \wedge \langle ?b \in R.\Gamma(H) \rangle$$

This is a case of *shared lack of knowledge*, which warrants an *inquiry* (McBurney & Parsons, 2001) dialogue to be initiated by either the robot or the human. For example, the robot captures an image but cannot figure out what is in the image, and the human also cannot figure out what is in the image. The robot might be able to detect color, but not shape; and the human might be able to discern shape but not color. So the robot can initiate an inquiry dialogue in which it proposes that it sees a red object in the image; the human might counter that she sees a box in the image; and together they can learn that there is a red box in the image.

Now that we have identified the reasons for which three different types of dialogue may be required, we next detail the inner workings of each dialogue. This involves first describing the **protocol** for each type of dialogue, and then describing the **axiomatic semantics** for each type of utterance mentioned in the dialogue protocols.

3.1 Dialogue Protocols

A *dialogue protocol* specifies the utterance that is employed at the start of a dialogue by the participant who initiates the dialogue, followed by the set of possible utterances that can be invoked in response, and so forth. These are illustrated graphically in Fig. 3. In the discussion below, the participant who initiates the dialogue is Ag_i, and the respondent is Ag_j. This general notation allows the discussion to hold no matter whether the robot or the human is the initiator.

Persuasion dialogue protocol. The protocol for a *persuasion* dialogue is illustrated in Fig. 3a. The reason to invoke a persuasion dialogue is when the initiator, Ag_i, believes something that she wants to convince another agent, Ag_j, to believe. Thus, before the dialogue begins, we have $b \notin Ag_i.\Gamma(j)$; and, if successful, after the dialogue ends, we will have $b \in Ag_i\Gamma(j)$. The opening utterance in a persuasion dialogue is assert(b). According to the rules of dialogue games, the belief b must be available to Ag_i, i.e., $b \in Ag_i.\Sigma \cup Ag_i.CS \cup Ag_j.CS$.

The simplest response to an assert is simply to accept, which agent Ag_j can present if Ag_j holds the same belief (i.e., $b \in Ag_j.\Sigma$) or if Ag_j contains an argument that either *p-support*s or *c-support*s (S, b) (as illustrated in Fig. 2). However, if $Ag_j.\Sigma$ contains arguments that *undermine* or

rebut (S, b) (as illustrated in Fig. 1), then Ag_j can *attack* the assertion by presenting a challenge. When an assertion is attacked, the agent that uttered the assertion (Ag_i) is required to provide the *support* for the assertion. The support is a set containing all the arguments in $Ag_i.\Sigma$ that *p-support* or *c-support* the argument (S, b). Every element in S must be accepted by Ag_j in order for (S, b) to be accepted, and hence for b to be accepted. So the process is an iterative one in which Ag_i cycles through each $s \in S$, eliciting a response to each s in turn. If every $s \in S$ is accepted, then the argument (S, b) is accepted and hence the conclusion of the argument, b, is accepted, which terminates the dialogue. Conversely, if any s is rejected (by Ag_j), then the argument (S, b) may be rejected and the dialogue will terminate. Alternatively, the rejection can be questioned (by Ag_i), by pausing the dialogue and initiating a second-level, *embedded* dialogue (illustrated in Fig. 5). For example, if Ag_j rejects s, then Ag_i could initiate an information-seeking dialogue by opening with question$(\neg s)$. The notion of embedded dialogues is discussed ahead in Section 3.3.

In the case that $\neg b \in Ag_j.\Sigma$ or $(S, \neg b) \in Ag_j.\Sigma$, then the response from Ag_j can be assert$(\neg b)$. Here, Ag_i will issue a challenge with respect to $\neg b$, since there is clearly a conflict because Ag_i had asserted b to begin with. The iterative challenge process (as above) will then take place with Ag_i in the role of challenger and Ag_j in the role of defender. The same termination conditions apply as above: either all the support $s \in (S, \neg b)$ is accepted, in which case $\neg b$ is accepted; or any s is rejected, in which case, $\neg b$ is rejected; and the dialogue terminates.

Information-seeking dialogue protocol. The protocol for an *information-seeking* dialogue is illustrated in Fig. 3b. The reason to invoke an information-seeking dialogue is when the initiator, Ag_i, wants to acquire information that she believes another agent, Ag_j, possesses. Thus, before the dialogue begins, we have $?b \in Ag_i.\Sigma$ and $b \in Ag_i.\Gamma(j)$. If successful, after the dialogue ends, Ag_i will have acquired information about the belief, which could be either b or $\neg b$. The opening utterance in an information-seeking dialogue is question(b). The respondent can reply by asserting either b or $\neg b$, which is why the dialogue may terminate satisfactorily with the initiator believing either b or $\neg b$, as well as confirming the other agent's belief or disbelief in b.

The processes for handling assertions and challenges in an information-seeking dialogue are the same as detailed above for persuasion dialogue. The only difference is that an additional possible response exists to the opening utterance: assert$(\mathcal{U})$. This is invoked if $?b \in Ag_j.\Sigma$, and so the dialogue terminates and Ag_i's beliefs are updated to: $?b \in Ag_i.\Gamma(j)$. The updates to Ag_i's beliefs upon acceptance are shown in the figure.

Inquiry dialogue protocol. The protocol for an *inquiry* dialogue is illustrated in Fig. 3c. The reason to invoke an inquiry dialogue is when the initiator, Ag_i, wants to acquire information that she believes another agent, Ag_j, does not possess either—so the goal is for the two agents to learn this information together. Thus, before the dialogue begins, we have $?b \in Ag_i.\Sigma$ and $?b \in Ag_i.\Gamma(j)$. If successful, after the dialogue ends, both agents will have acquired information about the belief, which could either be to believe b or $\neg b$. The opening utterance in an inquiry dialogue is propose$(a \rightarrow b)$. The explanation, elaborated in (Parsons, Wooldridge, & Amgoud, 2003b), is as follows. Note that (Parsons et al., 2003b) use the assert proposition in an inquiry dialogue, whereas we introduce propose in order to distinguish from the use of assert for persuasion. We make the assumption that the agents are already aware of the existence of b[2], so the purpose of the inquiry dialogue is to establish the veracity of b and the evidence which implies b (i.e., a) being either `true` or `false`. Hence, the opening gambit in the inquiry dialogue is a proposal by the initiator that b is implied by the proposition a. The respondent can either agree with the proposal, by issuing the utterance accept$(a \rightarrow b)$, or the respondent can challenge the proposal. In the latter case, the reply to the challenge utterance consists of providing support, S, for the proposition that was challenged

[2] We could engage in a philosophical debate about this question—whether the agents know about the existence of b—but such discussion is beyond the scope and purpose of this paper.

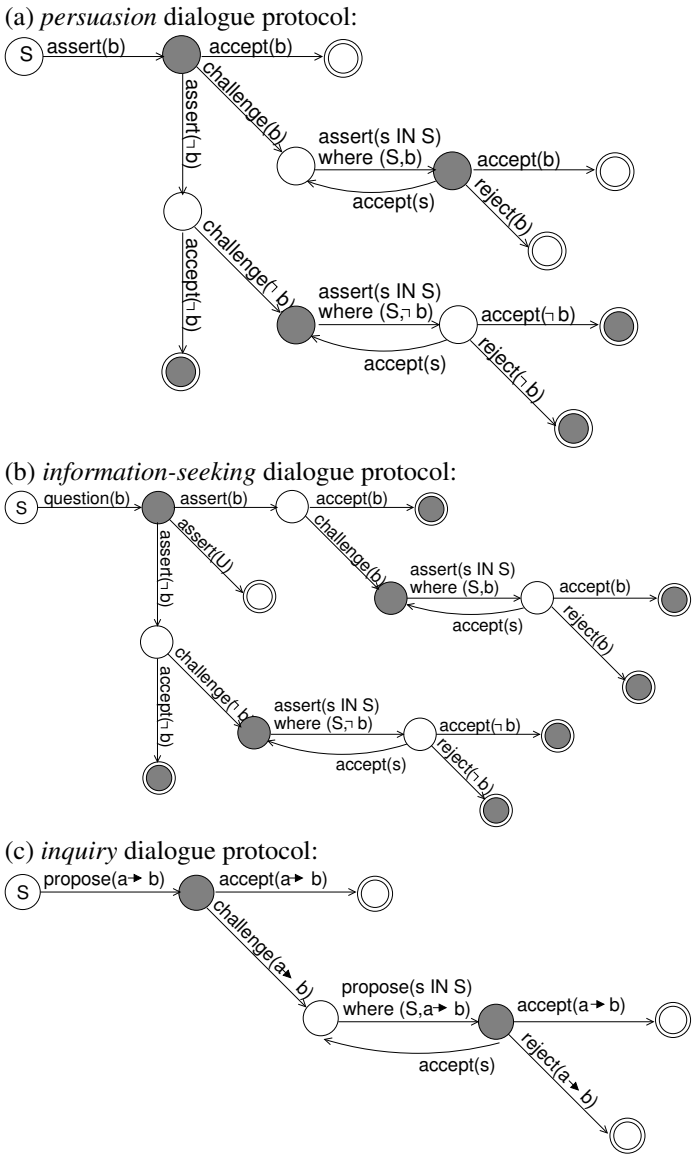


Figure 3. Dialogue protocols, drawn as state machines. The start state is indicated with an S. Termination states are indicated with double circles. States shown without fill are states in which the initiating agent is expected to make a move in the dialogue game; states filled in grey are states in which responding agent is expected to make a move.

(i.e., $a \rightarrow b$). This can continue iteratively, by proposing each element in the set of support $s \in S$, until either all the support is accepted or any element s is rejected.

LOCUTION:	PRE-CONDITIONS:	POST-CONDITIONS:
assert(b)	1. $b \in Ag_i.\Sigma$ 2. $(S, b) \in \mathcal{A}(Ag_i.\Sigma)$ 3. $b \notin Ag_i.\Gamma(j)$	1. $Ag_i.CS \leftarrow$ assert(b)
assert(S, b)	1. $b \in Ag_i.\Sigma$ 2. $(S, b) \in \mathcal{A}(Ag_i.\Sigma)$ 3. $b \notin Ag_i.\Gamma(j)$ 4. $(S, b) \notin Ag_i.\Gamma(j)$	1. $Ag_i.CS \leftarrow$ assert(S, b)
assert$(\mathcal{U})$ *(terminates dialogue)*	1. $?b \in Ag_i.\Sigma$	1. $Ag_i.CS \leftarrow$ assert$(\mathcal{U})$ 2. $Ag_i.\Sigma$: no change 3. $Ag_i.\Gamma(j) \leftarrow ?b$
challenge(b)	1. $b \in Ag_j.CS$ 2. $b \notin Ag_i.\Sigma$ 3. $(S, b) \in Ag_i.\Gamma(j)$	1. $Ag_i.CS \leftarrow$ challenge(b)
propose$(a \rightarrow b)$	1. $a \in Ag_i.\Sigma$ 2. $b \notin Ag_i.\Sigma$ 3. $b \notin Ag_i.\Gamma(j)$	1. $Ag_i.CS \leftarrow$ propose$(a \rightarrow b)$
question(b)	1. $?b \in Ag_i.\Sigma$ 2. $b \in Ag_i.\Gamma(j)$	1. $Ag_i.CS \leftarrow$ question(b)
accept(b)* *(terminates dialogue)*	1. $b \notin Ag_i.\Sigma$ 2. $b \in Ag_j.CS$ 3. $b \in Ag_i.\Gamma(j)$ 4. $(S, b) \in \mathcal{A}(Ag_i.\Gamma(j))$	1. $Ag_i.CS \leftarrow$ accept(b) 2. $Ag_i.\Sigma \leftarrow \{b\}$ 3. $\mathcal{A}(Ag_i.\Sigma) \leftarrow \{(S, b)\}$ 4. $Ag_i.\Gamma(j)$: no change
reject(b)* *(terminates dialogue)*	1. $b \notin Ag_i.\Sigma$ 2. $(S, b) \notin \mathcal{A}(Ag_i.\Sigma)$ 3. $b \in Ag_j.CS$	1. $Ag_i.CS \leftarrow$ reject(b) 2. $Ag_i.\Sigma$: no change 3. $Ag_i.\Gamma(j)$: no change

Figure 4. Axiomatic Semantics.

3.2 Axiomatic Semantics

The previous section described the protocols for three types of dialogue: persuasion, information-seeking and inquiry. In all, six different utterances, or *locutions*, are specified in the protocols. These are: accept, assert, challenge, propose, question, and reject. The *axiomatic semantics* for each type of locution are detailed in Fig. 4. These are described from the perspective of the speaking agent, Ag_i, uttered to a listening agent, Ag_j. A set of *pre-conditions* is listed for each locution (middle column in the figure), indicating what conditions must be true in order for the locution to be uttered. When multiple pre-conditions are listed, then all of them must be true. A set of *post-conditions* is also listed for each locution (rightmost column in the figure). Four of the locutions can be presented at the beginning or middle of a dialogue: assert, challenge, propose and question. After the intermediate locutions are presented, only the commitment store ($Ag_i.CS$) of the speaking agent is updated—with the locution that was uttered. In this way, the commitment store functions as a kind of "chat log."

A dialogue typically terminates when one of two locutions is presented: accept or reject. The post-conditions for these locutions include updating the speaking agent's commitment store ($Ag_i.CS$), as above. Because these locutions indicate the termination of the dialogue, the speaking

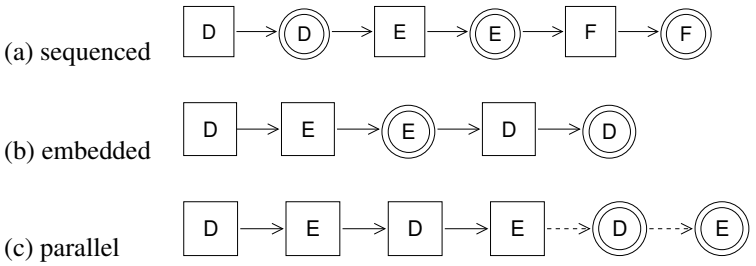


Figure 5. Different combinations of dialogues. The boxes indicate the commencement of a new dialogue, and the double-circles indicate the termination of a dialogue. In the *sequenced* combination (a), dialogue D starts and ends before dialogue E begins; and E begins and ends before dialogue F commences; and so forth. In the *embedded* combination (b), dialogue D starts; then dialogue E begins and ends, before D has terminated—so that E is entirely nested within the middle of D. In the *parallel* combination (c), dialogue D starts; then dialogue E begins; then D continues; then E continues—locutions from the two dialogues are interleaved, and either dialogue may terminate before the other.

agent's belief set ($Ag_i.\Sigma$) and beliefs about the listening agent ($Ag_i.\Gamma(j)$) may also be updated. For accept($\neg b$) and reject($\neg b$), values of b in the pre- and post-conditions are replaced with $\neg b$. Note that a special form of the assert locution, assert($\mathcal{U}$), also terminates a dialogue. This is uttered when the speaker has no knowledge about the question just asked.

3.3 Control Layer

In order to implement the dialogue games described above, particularly in a human-robot environment designed to support fluid and spontaneous exchange of ideas, we incorporate the notion of a *control layer* (McBurney & Parsons, 2002). A control layer consists of rules that determine when to start and end a dialogue (*commencement* and *termination* rules, respectively) and help keep track of which dialogue(s) are active at any given time. This construct also allows multiple dialogues to occur simultaneously.

When two agents (e.g., a human and a robot, in our HRI context) share decisions and perform a mission together, they will need to interact and likely engage in multiple dialogues. The dialogues may be interleaved with actions, for example, they may first engage in a dialogue in which they agree for the robot to collect some sensor data from a particular location. Then the robot goes to the location, gathers data and engages the human in another dialogue in order to discuss the data. The dialogues may also be interleaved with each other, for example, the robot may begin an inquiry dialogue to propose that it go to a location and take sensor data. The human may agree with the idea of collecting sensor data, but disagree about the location; in which case, a persuasion dialogue will be initiated by the human before the robot's inquiry dialogue has terminated. Fig. 5 illustrates the ways in which multiple dialogues may occur. A *sequenced* dialogue combination is where multiple dialogues occur one after the other, so that one dialogue terminates before another dialogue commences. An *embedded* dialogue combination is where one dialogue commences, and before it terminates, a new dialogue commences and terminates. A *parallel* dialogue combination is where one dialogue commences, and before it terminates, a new dialogue commences; then the

first dialogue continues, before the second has terminated; and so on, so that the dialogues are interleaved.

4. Application: ArgHRI

The previous sections of this paper have described logical argumentation theory and dialogue games, which were developed for multi-agent interaction. As mentioned in the introduction, our work involves applying this theory to the human-robot domain. In this section, and for the remainder of the paper, we shift our focus to HRI and detail how we have applied the theory to obtain a flexible system for shared human-robot decision making which can handle unexpected input from the human and the robot and the physical world. Our system is called *ArgHRI*.

There are a number of key differences between multi-agent and HRI forms. First, in multi-agent interaction, all participants in a dialogue are agents; thus their beliefs are all modelled computationally and their actions are controlled. In a traditional multi-agent environment—as opposed to a multi-robot environment—the agents are instantiated in software and act in a virtual world, whereas robots are embodied and act in the physical world. While robots' beliefs can also be modelled, their actions are non-deterministic because they function in a noisy world; whereas most virtual agent worlds are deterministic, especially agent-only worlds (i.e., without human interaction). *Human-Computer Interaction (HCI)* is a broad and extremely challenging field, incorporating disparate topics ranging from interface design to human factors to natural language understanding and generation. So the shift from agent-agent dialogues to human-robot dialogues entails two significant steps: (1) from the virtual to the physical world; and (2) from agent-only interactions to interaction with humans. Our approach involves two primary components: (1) a robot control architecture that incorporates the argumentation and dialogue game theory described above; and (2) a human interface that facilitates communication with the robot and enforces the rules of the dialogue games. These are each discussed below.

4.1 Robot Control Architecture

Fig. 6 illustrates Nilsson's classic three-step robot control architecture (Nilsson, 1984): first, the robot *senses* its environment; second, the robot formulates a *plan* about what to do; third, the robot *acts* out its plan; and then the process loops back to the first step. Although modern architectures frequently employ a less sequential strategy, these three fundamental components are widely used. We are concerned with situations in which the robot interacts with a human in a shared decision-making step where the human and robot discuss and reach agreement about what the robot should do. Thus, we extend the classic architecture by adding a *dialogue* step, as shown in the figure (step 2*). 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, and leave for future work investigation of ways to build plans that combine robot actions and speech acts (Austin, 1975).

As shown in Fig. 6, we add an inner loop to the classic architecture, for the robot to sense its environment again after dialogue. Since the robot'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 not introduce any changes to the robot's 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. The details of the processing steps are as follows:

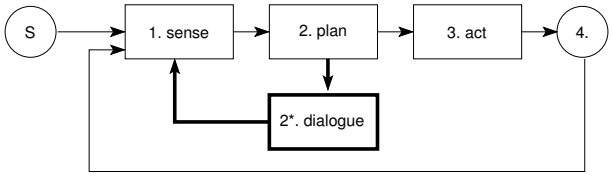


Figure 6. Robot control architecture, with dialogue step added.

S. The robot R starts with an initial belief state:
$R.\Sigma_0$ (at time $t = 0$)

1. The robot R senses its environment, at time t:
$R.obs_t \leftarrow R.sense(Env_t)$
and then updates its prior beliefs, based on its observations:
$R.\Sigma_t \leftarrow update(R.\Sigma_{t-1}, obs_t)$

2. The robot R plans which action to perform:
$R.Ac_t \leftarrow action()$

2*. The robot R discusses its plan with human H to reach agreement:
$R.Ac_t \leftarrow R.dialogue(H)$
The plan may change or stay the same.
Re-sense (step 1) and re-plan (steps 2 and 2*), if necessary
(i.e., if the environment has changed).

3. The robot R performs the selected action, $R.Ac_t$.

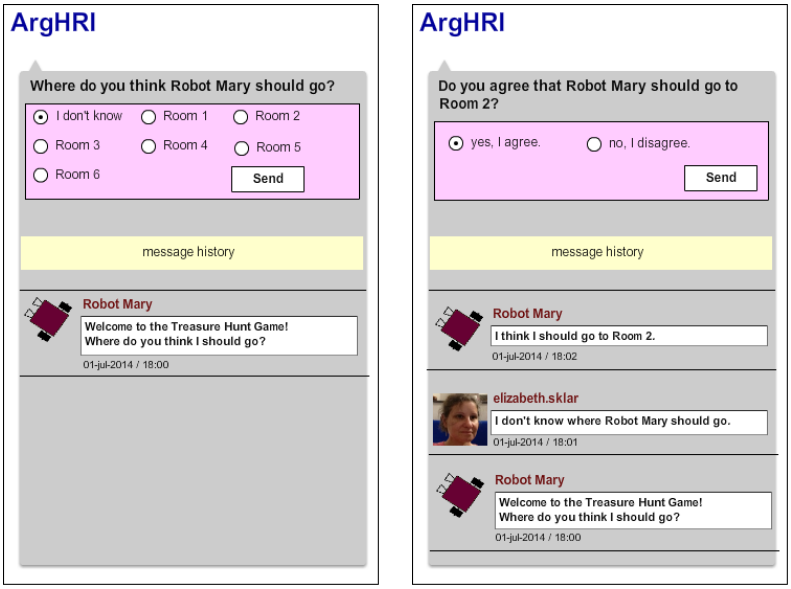
4. The process iterates back to step 1.

4.2 Human Interface

In our *ArgHRI* implementation, the human interacts with the robot using a "chat" style interface, as shown in Fig. 7. Since our work concerns the application of logical argumentation and dialogue games, we (currently) avoid natural language issues by providing the human with multiple-choice style questions for interacting with the robot. This also ensures that the human obeys the rules of the dialogue game.

The benefits of our methodology are illustrated by the ability to handle a range of options provided to the human and the flexible ways in which responses are handled. For example, the opening question in our implementation asks the human where she thinks the robot ("Robot Mary" in Fig. 7) should go. As described in Section 5, our human-robot experimental domain is the *Treasure Hunt Game*. The robot has a choice of possible rooms to explore (to search for treasures). If the human responds to the initial question by selecting one or more rooms, then her choice is compared with the robot's choice of room(s). If they have chosen the same room(s), then no dialogue is necessary (cases 2 and 4 in Table 2). However, if they have chosen different room(s), then a *persuasion dialogue* is initiated in which they can reach agreement about which room(s) the robot should visit (cases 1 and 5 in Table 2). If the human selects "I don't know," then an *information-seeking dialogue* is initiated in which the human can query the robot about its choice of room(s)[3]. In our experimental work, the robot's choices of where to go at the start of a game are determined randomly—and also include the "I don't know" option. However, the robot can be instantiated with any desired initial set

[3]Note that the interface prevents the human from selecting both "I don't know" and any room, while allowing selection of multiple rooms (without selecting on "I don't know").



(a) (b)

Figure 7. Human Interface for Treasure Hunt Game Play. The left-hand screen (a) displays the welcome message that the user sees when the game starts up. The right-hand screen (b) displays a short message history—the commitment store for human and robot—for their dialogue about defining a goal for the robot to achieve.

of beliefs. If both human and robot have selected "I don't know," then an *inquiry dialogue* ensues in which they decide together which room(s) the robot should visit.

More detail about the dialogues within the context of our experimental domain is provided in Section 6. But first, we present our experimental domain.

5. Experimental Domain: The Treasure Hunt Game

This section provides a formal description of the *Treasure Hunt Game (THG)*, which we designed for conducting experiments with human-robot teams. Our game is a variation on the *treasure hunt domain* introduced in Jones et al. (2006). The original domain was designed to assess the performance of competitive "pick up" (i.e., *ad hoc*) teams of heterogenous robots exploring an unknown environment and searching for treasure. The objective was for each team to maximize the amount of treasure collected within a fixed period of time.

Our variation frames the domain as a real-time strategy game, where a human "operator" and a robot work together to search for treasure in an environment that is accessible to the robot but not to the human. The robot moves around and collects sensor data, which is shared with the human. The human and the robot jointly make decisions, based on the data collected, about actions to take in order to win the game. The human-robot team receives points for correctly locating and identifying treasures. The human-robot team loses points for incorrectly identifying and/or locating treasures. The human-robot team expends energy for robot movement, sensing and communication.

Next, we provide a formal definition for our version of the THG, a description of the rules, and

a scoring mechanism.

5.1 Formal Description

A **THG instance** is defined by the tuple:

$$\langle map, treasures \rangle$$

The components that comprise the THG have been previously introduced and implemented by Özgelen and Sklar (2013, 2014), Azhar, Parsons, and Sklar (2013), Azhar, Schneider, Salvit, Wall, and Sklar (2013), and Sklar et al. (2012). Each component is detailed below.

A **map** is a tuple $\langle size, walls \rangle$ where:

- $size$ is the extent of the rectangular bounding box circumscribing the robot's physical 2-dimensional (2D) environment represented as an ordered pair, (w, h), where w is the width of the bounding box (along the east-west axis) and h is the height of the bounding box (along the north-south axis)[4]; and
- $walls$ is a set of $wall$ specifications.

A **wall** is defined as a tuple $\langle id, x_1, y_1, x_2, y_2 \rangle$, where:

- id is a unique identification name or number of the wall,
- (x_1, y_1) is one endpoint of the wall, with constraints $0 \leq x_1 \leq w$ and $0 \leq y_1 \leq h$, and
- (x_2, y_2) is the other endpoint of the wall, with constraints $x1 \leq x_2 \leq w$ and $y_1 \leq y_2 \leq h$ (specifying a thick wall or enclosed rectangular region when $x_1 \neq x_2$ and $y_1 \neq y_2$).

A **room** is defined as a set of walls, which collectively form a boundary surrounding a spatial region. Walls within a room may share common endpoints, but this is not required if there are doorways in the room. A room is a logical structure within which *containment* can be computed, such that a robot or object can be determined to be in a room or not in a room.

A set of **treasures** contains one or more treasure items, each represented by a tuple:

$$\langle id, type, color, value, n, x_1, y_1, \ldots, x_n, y_n \rangle$$

where:

- id is the unique identification name or number of the treasure item;
- $type$ is the type of the treasure item (e.g., "cube" or "bottle");
- $color$ is the color of the treasure item (e.g., "red" or "blue");
- $value$ is the value of the treasure item (i.e., the number of points rewarded to the human-robot team for correctly locating and identifying the treasure item);
- n is the number of points in the polygon that describes the footprint of the treasure item; and
- each (x_i, y_i) is a point in the polygon that describes the treasure item's footprint, with points ordered in a clockwise sequence and constraints $0 \leq x_i \leq w$ and $0 \leq y_i \leq h$.

A **THG mission** is an instantiated THG instance. The objective of a THG mission is for the human-robot team to locate and correctly identify as many treasures as possible, before the team runs out of energy.

5.2 Scoring

The team receives a **score** for the mission based on the number of correctly identified treasures. Each treasure item has a value associated with it (as defined above). When a treasure item is correctly located and identified, the value of that treasure is added to the team's score. Values are assigned before a game begins, generally adhering to the following heuristic. Small, ambiguous treasures

[4] The northwest (upper left) corner of the bounding box is at (0,0) and the southeast (lower right) corner of the bounding box is at (w,h).

Name	*Color*	*Footprint*	*Identifiability*	*Value*
basketball	Orange	Round & Large	Unique color, unique footprint	Low
fuzzy_die	Pink	Square & Large	Unique color, ambiguous footprint	Medium
candy_box	Green	Square & Large	Ambiguous color, ambiguous footprint	High
beer_bottle	Green	Round & Small	Ambiguous color, unique footprint	Medium

Table 3: Sample set of treasure items. *Identifiability* and *value* are computed relative to the set.

are hard to identify, so they have higher value. Big, unambiguous treasures are easy to identify, so they have lesser value. A sample set of treasure items is shown in Table 3. When a treasure item is incorrectly located or identified, a percentage of the value of that treasure is subtracted from the team's score. Here are some examples of incorrect answers that might be provided. Assume that there is one candy_box in the environment, and it is located at position $(3, 9)$. If the human-robot team decides that the candy_box is at position $(25, 2)$, then they would have *mislocated* the object. If the human-robot team decides that the beer_bottle is at position $(3, 9)$, then they would have *misidentified* the object.

5.3 Energy

The robot has a limited amount of energy, referred to as *health points*. The robot cannot simply perform an exhaustive search of the environment to find all the treasure items, because it will run out of energy before visiting the whole environment. Thus, the human-robot team must collaborate to decide how best to make use of the robot's energy and locate as many treasures as possible.

The number of health points cannot be increased during a mission. The robot starts with a maximum number of health points, and this value declines as the mission proceeds. Health points decrease when energy is expended in any of the following ways: when the robot moves; when the robot collects sensor data; or when the robot transmits information. We assume fixed values for health point computation, based on distance travelled (for motion), amount of sensor data collected and size of message transmitted.

6. Detailed Example

In this section, we demonstrate the use of argumentation-based dialogue games to facilitate flexible HRI by providing a detailed example of the Treasure Hunt Game, as played using our ArgHRI interface. As described above, in a THG instance, a human and robot work together to locate treasures in an arena that is inaccessible to the human. At the start of the game, the human and robot are given a map of the THG arena, so that they know how many rooms are in the arena and how they are connected. They know that a number of treasures are hidden in the arena, and their mission is to find these treasures. The robot does not have enough energy to perform an exhaustive search, so the robot and human have to work together to solve the mission.

For experimentation and to demonstrate the flexibility of our argumentation-based dialogue methodology, we have designed a game-play scenario that involves three types of decisions to be performed jointly between the human and the robot: (1) deciding where to look for treasures (i.e., which rooms to search); (2) deciding how the robot should travel to the rooms (i.e., which order to search the rooms); and (3) deciding what is found in each room once the robot arrives (i.e., analyzing images collected by the robot). Although there is a logical sequence, the structure of the system

allows the human-robot team to consider each decision in any order. Next, we provide examples for the decisions and demonstrate features of argumentation-based dialogues for each decision. The examples are based on an arena with 6 rooms, and a THG mission in which 4 treasures are hidden.

6.1 Deciding Where To Look

First, the human-robot team must decide which rooms the robot will visit to look for treasure. In our example, we have 6 rooms to choose from, which means that there are 64 possible options[5]. Thus, there is significant possibility for conflict in this seemingly simple decision: $64 \times 64 = 4096$ possible combinations of human and robot choices of where to look for treasure. Referring back to Table 2 and assuming that all the possibilities are equally likely, the probability of agreement is $63/4096 = 1.54\%$ (cases 1 and 5); the probability of disagreement is $3906/4096 = 95.36\%$ (cases 2 and 4); the probability of lack of knowledge (because either the human or the robot selected "I don't know") is $126/4096 = 3.08\%$; and the probability of shared lack of knowledge (because both the human and the robot selected "I don't know") is $1/4096 = 0.02\%$ (case 9). These calculations do not separate out cases where there is partial agreement. For example, if the human selects $\{\mathsf{Room1}\}$ and the robot selects $\{\mathsf{Room1}, \mathsf{Room3}\}$, then they agree about Room1 but disagree about Room3; this situation is counted in the above calculation as a case of *disagreement*. We could consider discussing each room individually as a binary decision (should the robot look in Room i or not), but there is still the question of how many rooms should be visited in total; and because of the energy constraint, some pairs of rooms (e.g., adjacent rooms) are more economical to consider than others.

The robot can perform path planning[6] and so can compute the amount of energy it will take to visit the chosen room(s). This information is used to seed the evidence component of the robot's belief base, so for each room chosen (randomly), the amount of energy required to visit the room (from the robot's starting or current location) is computed and stored in the robot's set of beliefs as supporting evidence for visiting that room. For example, if it will take 100 units of energy to reach Room1 from Room4, then the robot's belief set will contain:

$$\begin{array}{lcll} b_1 & = & selected(\mathsf{Room1}) & [b_1] \\ (S, b_1) & = & \{current\_energy\_level(1000), & [s_1] \\ & & energy\_cost(\mathsf{Room4} \rightarrow \mathsf{Room1}, 100), & [s_2] \\ & & less\_than(100, 1000)\} & [s_3] \end{array}$$

Here is a sample dialogue sequence for this example situation:

Pre-conditions:

Beliefs	*Description*
$R.\Sigma \ni selected(\mathsf{Room1}) \wedge selected(\mathsf{Room3})$	The robot believes that it should visit both rooms 1 and 3.
$R.\Gamma(H) \ni selected(\mathsf{Room1})$	The human believes that the robot should only visit room 1.

The beliefs can be represented as b_1 (as above) and $b_3 = selected(\mathsf{Room3})$. Thus we have:

$$\begin{array}{lcllcl} \langle b_1 \in R.\Sigma \rangle & \text{and} & \langle b_1 \in R.\Gamma(H) \rangle & \rightarrow & \textit{agreement} \text{ (case 1)} \\ \langle b_3 \in R.\Sigma \rangle & \text{and} & \langle \neg b_3 \in R.\Gamma(H) \rangle & \rightarrow & \textit{disagreement} \text{ (case 4)} \end{array}$$

Case 1 calls for no dialogue. Case 4 calls for *persuasion dialogue*. As indicated in Fig. 5, the control

[5] $\binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} + 1 = 64$

[6] The system currently uses the A* path planning algorithm (Hart, Nilsson, & Raphael, 1968).

layer will present a message to the human to start the dialogue, and then the persuasion dialogue, initiated by the robot, will commence:

Dialogue move	*Scripted text in chat-style interface*
Control layer	There is a conflict about selecting Room3.
$R.\mathsf{assert}(b_3)$	Robot Mary: I believe that Room3 should be selected.

Here, according to the persuasion dialogue protocol (see Fig. 3a), the human can either accept the robot's belief, challenge the robot's belief, or assert the opposite belief. So the interface will give the human these choices:

Dialogue move	*Scripted text in chat-style interface*
Control layer	Do you agree with Robot Mary? ○ Yes ○ Maybe—I want to know why Robot Mary selected Room3. ○ No! I think that Room3 should not be selected.

The dialogue will continue, based on the human's choice. If the human chooses "Yes," then the dialogue will continue as follows:

Dialogue move	*Scripted text in chat-style interface*
$H.\mathsf{accept}(b_3)$	Human: Yes, I agree that Room3 should be selected.
Control layer	*Persuasion dialogue terminates*

If the human chooses "Maybe—I want to know why," then the dialogue will continue as follows:

Dialogue move	*Scripted text in chat-style interface*
$H.\mathsf{challenge}(b_3)$	Human: Why do you think that Room3 should be selected?

As described above, the robot's belief set is seeded with information about how much energy it would expend to visit each room that is randomly selected in the experiment. This is taken as supporting evidence for the belief that the robot should visit the room. So, according to the dialogue protocol, the robot responds to the human's challenge by putting forth its supporting evidence:

Dialogue move	*Scripted text in chat-style interface*
$R.\mathsf{assert}(S, b_3)$	Robot Mary: There are three reasons why I selected Room3.
$R.\mathsf{assert}(s_1)$	Robot Mary: My current energy level is 1000 units.
Control layer	Do you agree with this statement? (yes/no)
$H.\mathsf{accept}(s_1)$	Human: Yes
$R.\mathsf{assert}(s_2)$	Robot Mary: It will cost 100 energy units to visit Room 3.
Control layer	Do you agree with this statement? (yes/no)
$H.\mathsf{accept}(s_2)$	Human: Yes
$R.\mathsf{assert}(s_3)$	Robot Mary: The energy cost (100 units) is less than my current energy level (1000 units).
Control layer	Do you agree with this statement? (yes/no)
$H.\mathsf{accept}(s_3)$	Human: Yes
Control layer	You have agreed with Robot Mary's reasons for selecting Room3, which implies agreement to select Room3. *Persuasion dialogue terminates*

If the human answers "No" to any of the above questions, then an $H.\mathsf{reject}(s_i)$ utterance will be issued and the dialogue will terminate.

If the human answers "No! I think that Room3 should not be selected" in response to the opening assertion in the persuasion dialogue, then one of two things can happen according to the rules of the persuasion dialogue protocol; either the robot can accept the human's assertion, or the robot can challenge the human's assertion. Because the robot's belief set contains the belief that Room3 should be selected (i.e., $b_3 \in R.\Sigma$), then the pre-conditions required for allowing the robot to utter $\mathsf{assert}(\neg b_3)$ would be violated (see Fig. 4). This means that the robot has to attack the human's argument $(S, \neg b_3)$, which *rebuts* (S, b_3), so the robot issues a challenge:

Dialogue move	*Scripted text in chat-style interface*
$H.\mathsf{assert}(\neg b_3)$	Human: I think that Room3 should not be selected.
$R.\mathsf{challenge}(\neg b_3)$	Robot Mary: Why do you think that Room3 should not be selected?

The human then responds by providing evidence that supports the belief that Room3 should not be visited, for example: "Because I made a random choice," "Because Room3 is too far away given our current energy level," "Because Room3 has already been visited," etc.

After the persuasion dialogue terminates, the beliefs of the robot are updated, including the robot's beliefs about the human's beliefs ($R.\Gamma(H)$). The persuasion dialogue can terminate in three possible states: (1) the robot has persuaded the human (i.e., the human accepts the robot's initial assertion, $H.\mathsf{accept}(b_3)$); (2) the human has persuaded the robot (i.e., the robot accepts the human's counter-assertion, $R.\mathsf{accept}(\neg b_3)$); or (3) nobody has been persuaded (i.e., either the initial assertion or the counter-assertion was rejected). The post-conditions for the first two termination states are listed below. For the third state, the post-conditions do not change from the pre-conditions.

Beliefs	*Description*
Post-conditions if human accepts robot's initial assertion, $H.\mathsf{accept}(b_3)$*:*	
$R.\Sigma \ni selected(\mathsf{Room1}) \wedge selected(\mathsf{Room3})$	The robot believes that it should visit both rooms 1 and 3.
$R.\Gamma(H) \ni selected(\mathsf{Room1}) \wedge selected(\mathsf{Room3})$	The human believes that the robot should visit both rooms 1 and 3.
Post-conditions if robot accepts human's counter-assertion, $R.\mathsf{accept}(\neg b_3)$*:*	
$R.\Sigma \ni selected(\mathsf{Room1})$	The robot believes that it should only visit room 1.
$R.\Gamma(H) \ni selected(\mathsf{Room1})$	The human believes that the robot should only visit room 1.

6.2 Deciding How To Get There

Next, the human and robot have to agree on how to visit the rooms—that is, the order in which the robot should visit the selected room(s). A partial order can be agreed upon by just concurring about which room to visit next, one room at a time; or a complete order can be agreed upon for all rooms selected. The control layer provides the initial multiple-choice question to the human for addressing this decision. Note that if there is only one room chosen, then this dialogue doesn't take place, because there is no question of which room to visit first, since there is only one room to visit.

Below, we continue the scenario detailed above. For ease of presenting a complete demonstration of our approach, we will assume that agreement was reached regarding visiting both rooms 1 and 3.

Thus, before addressing the "how to get there" question, we have the following *pre-conditions*:

Beliefs	*Description*
$R.\Sigma \ni selected(\mathsf{Room1}) \wedge selected(\mathsf{Room3})$	The robot believes that it should visit both rooms 1 and 3.
$R.\Gamma(H) \ni selected(\mathsf{Room1}) \wedge selected(\mathsf{Room3})$	The human believes that the robot should visit both rooms 1 and 3.

Beliefs about the order in which rooms should be visited can be represented as:

$$\begin{array}{rcl} b_{11} & = & order(1, \mathsf{Room1}) \\ b_{13} & = & order(1, \mathsf{Room3}) \\ b_{21} & = & order(2, \mathsf{Room1}) \\ b_{23} & = & order(2, \mathsf{Room3}) \end{array}$$

Assume the user indicates the following choice:

Beliefs	*Description*
$R.\Gamma(H) \supseteq \{selected(\mathsf{Room1}), selected(\mathsf{Room3}),$ $order(1, \mathsf{Room3}), order(2, \mathsf{Room1})\}$	The human believes that the robot should visit both rooms 1 and 3, and that room 3 should be visited first and room 1 should be visited second.

Since only two rooms were selected to visit, the possibility for disagreement between the human and the robot is much smaller than in the previous example (with the "where to go" decision). Here, the options for the robot are[7]:

$\langle b_{13}, b_{21} \in R.\Sigma \rangle$ and $\langle b_{13}, b_{21} \in R.\Gamma(H) \rangle$ $\rightarrow$ *agreement* (case 1)
$\langle b_{11}, b_{23} \in R.\Sigma \rangle$ and $\langle \neg b_{11}, b_{13}, b_{21}, \neg b_{23} \in R.\Gamma(H) \rangle$ $\rightarrow$ *disagreement* (case 4)

The robot will also autonomously select an order in which to visit the rooms. For example, it might compute the cost of traveling from its starting (or current) location to each of the rooms selected and determine a shortest-path order for visiting the rooms:

$$\begin{array}{lcll} b_{11} \wedge b_{23} & = & order(1, \mathsf{Room1}) \wedge order(2, \mathsf{Room3}) & [b_{11}, b_{23}] \\ (S, \langle b_{11} \wedge b_{23} \rangle) & = & \{current\_energy\_level(1000), & [s_1] \\ & & energy\_cost(\mathsf{Room4} \rightarrow \mathsf{Room1}, 100), & [s_2] \\ & & energy\_cost(\mathsf{Room4} \rightarrow \mathsf{Room3}, 250), & [s_4] \\ & & energy\_cost(\mathsf{Room1} \rightarrow \mathsf{Room3}, 250), & [s_5] \\ & & less\_than(100, 250), & [s_6] \\ & & less\_than(100 + 250, 1000)\} & [s_7] \end{array}$$

Because there is a conflict between the human's and robot's choices about the order in which to visit the rooms, another persuasion dialogue occurs, this time initiated by the robot:

[7] Note that the system performs some implicit computation with the $order$ predicate, since inherently b_{11} and b_{13} cannot both be true at the same time. These implicit predicates are listed explicitly in the "disagreement" line, above, to clearly illustrate where the disagreement occurs.

Dialogue move	*Scripted text in chat-style interface*
Control layer	Specify the order in which to visit the rooms:
The human enters her beliefs:	Room3 → Room1
$R.\mathsf{assert}(b_{11} \wedge b_{23})$	Robot Mary: I think that the rooms should be visited in this order: Room1 → Room3
$H.\mathsf{assert}(\neg b_{11} \wedge \neg b_{23})$	Human: I disagree.
$R.\mathsf{challenge}(\neg b_{11} \wedge \neg b_{23})$	Robot Mary: Why do you disagree?
$H.\mathsf{assert}(S \vdash \langle \neg b_{11} \wedge \neg b_{23} \rangle)$	Human: I have two reasons for disagreeing.
$H.\mathsf{assert}(s_{10})$	Human: I believe Room3 should be visited first, not Room1.

This last assertion by the human is not a convincing argument to present to the robot, because the robot has evidence that it is less costly to visit Room1 first. So the robot rejects the human's supporting evidence and the dialogue terminates.

In many HRI systems, the discussion would terminate in a stalemate and the THG would end in failure. However, because ArgHRI allows the robot to take the initiative and start a new dialogue, and because ArgHRI keeps track of what evidence has been presented, the THG need not fail. Instead, the robot can initiate a new persuasion dialogue to convince the human of its plan. The robot opens the new dialogue with the same utterance as the previous dialogue, which is allowed in the context of dialogue game rules because all possible responses have not been exhausted yet. The human is bound by the dialogue game rules, however, to provide a different response than was given previously. Since the human's beliefs have not changed ($R.\Gamma(H) \ni \{b_{13}, b_{21}\}$), the only option is to attack the robot's argument:

Dialogue move	*Scripted text in chat-style interface*
$R.\mathsf{assert}(b_{11} \wedge b_{23})$	Robot Mary: I think that the rooms should be visited in this order: Room1 → Room3
$H.\mathsf{challenge}(b_{11} \wedge b_{23})$	Human: Why?
$R.\mathsf{assert}(S \vdash \langle b_{11} \wedge b_{23} \rangle)$	Robot Mary: I have six reasons.
$R.\mathsf{assert}(s_1)$	Robot Mary: My current energy level is 1000 units.
$H.\mathsf{accept}(s_1)$	Human: Okay.
$R.\mathsf{assert}(s_2)$	Robot Mary: It costs 100 energy units to go from Room4 to Room1.
$H.\mathsf{accept}(s_2)$	Human: Okay.
$R.\mathsf{assert}(s_4)$	Robot Mary: It costs 250 energy units to go from Room4 to Room3.
$H.\mathsf{accept}(s_4)$	Human: Okay.
$R.\mathsf{assert}(s_6)$	Robot Mary: 100 is less than 250.
$H.\mathsf{accept}(s_6)$	Human: Okay.
$R.\mathsf{assert}(s_3)$	Robot Mary: 100 is less than 1000.
$H.\mathsf{accept}(s_3)$	Human: Okay.

At this point in the dialogue, the robot has presented all its evidence in support of b_{11}, and all the acceptance history is in the human's commitment store. So the robot could open an *embedded* dialogue in order to obtain the human's acceptance of b_{11}:

Dialogue move	*Scripted text in chat-style interface*
Control layer	*Robot initiates an embedded persuasion dialogue*
$R.\mathsf{assert}(b_{11})$	Robot Mary: Please agree that Room1 should be visited first.
$H.\mathsf{accept}(b_{11})$	Human: Okay.
Control layer	*Embedded persuasion dialogue terminates and control returns to previously unfinished dialogue*

This acceptance and embedded dialogue termination means that the robot's beliefs about the human's beliefs can be updated with the newly accepted belief (b_{11}) as well as the support for that belief (S, b_{11}):

$$
\begin{array}{rcll}
R.\Gamma(H) & \supseteq & \{\ order(1, \mathsf{Room1}), & [b_{11}] \\
 & & current\_energy\_level(1000), & [s_1] \\
 & & energy\_cost(\mathsf{Room4} \rightarrow \mathsf{Room1}, 100), & [s_2] \\
 & & energy\_cost(\mathsf{Room4} \rightarrow \mathsf{Room3}, 250), & [s_4] \\
 & & less\_than(100, 250), & [s_6] \\
 & & less\_than(100, 1000), & [s_3] \\
 & & \neg order(1, \mathsf{Room3}), & [\neg b_{13}] \\
 & & order(2, \mathsf{Room1}), & [b_{21}] \\
 & & selected(\mathsf{Room1}), & [b_1] \\
 & & selected(\mathsf{Room3})\ \} & [b_3]
\end{array}
$$

Now there are only two more pieces of evidence to support b_{23} that the robot has not yet put forth. So upon resuming the initial persuasion dialogue, the robot offers:

Dialogue move	*Scripted text in chat-style interface*
$R.\mathsf{assert}(s_5)$	Robot Mary: It costs 250 energy units to go from Room1 to Room3.
$H.\mathsf{accept}(s_5)$	Human: Okay.
$R.\mathsf{assert}(s_7)$	Robot Mary: 100 + 250 is less than 1000.
$H.\mathsf{accept}(s_7)$	Human: Okay.

This completes the presentation and acceptance of support for b_{23}, so the dialogue terminates in agreement.

Dialogue move	*Scripted text in chat-style interface*
Control layer	You have agreed with Robot Mary's reasons for going to Room3 second, which implies agreement to visit Room3 second. *Persuasion dialogue terminates*

6.3 Deciding What Is Found There

When a robot arrives in a room, it will perform a sensor-sweep task. This involves capturing 5 images in a circle, each 72° apart, and showing the images to the human. The robot can perform color segmentation and can form its own hypotheses about the contents of the images. For example, the robot might store the following beliefs:

$$
\begin{array}{lcl}
e_{100} & = & color\_found(image1, brown) \\
e_{101} & = & color\_found(image2, grey) \\
e_{102} & = & color\_found(image2, \mathbf{blue}) \\
e_{103} & = & color\_found(image3, brown) \\
e_{104} & = & color\_found(image4, yellow) \\
e_{105} & = & color\_found(image5, \mathbf{blue})
\end{array}
$$

Predicates e_{102} and e_{105} provide evidence that there is something "blue" in Room1, but there is no evidence for which blue treasure it is. So, the robot initiates an *information-seeking* dialogue, because the human can discern "shape" and the robot cannot:

Dialogue move	*Scripted text in chat-style interface*
$R.\mathsf{question}(b_{55})$	Robot Mary: Is there a bottle in image5?
$H.\mathsf{assert}(\mathcal{U})$	Human: I don't know.
Control layer	*Information-seeking dialogue terminates*

This assertion and dialogue termination means that the robot's beliefs can be updated with respect to the unknown belief (b_{55}), as well as the robot's beliefs about the human's beliefs:

$$
\begin{array}{lcl}
R.\Sigma & \ni & ?b_{55} \\
R.\Gamma(H) & \ni & ?b_{55}
\end{array}
$$

Since the robot detected blue in two images, it can start another information-seeking dialogue:

Dialogue move	*Scripted text in chat-style interface*
$R.\mathsf{question}(b_{52})$	Robot Mary: Is there a bottle in image2?
$H.\mathsf{assert}(b_{52})$	Human: Yes.
$R.\mathsf{accept}(b_{52})$	Robot Mary: Okay.
Control layer	*Information-seeking dialogue terminates*

This acceptance and dialogue termination means that the robot's beliefs can be updated:

$$
\begin{array}{lcll}
R.\Sigma & \ni & in\_image(\mathsf{bottle}, \mathsf{image2}) & [b_{52}] \\
R.\Gamma(H) & \ni & in\_image(\mathsf{bottle}, \mathsf{image2}) & [b_{52}]
\end{array}
$$

Now the human can initiate an inquiry dialogue to concur with the robot regarding the evidence it has found:

Dialogue move	*Scripted text in chat-style interface*
$H.\mathsf{propose}(a \rightarrow b)$	Human: Is there evidence that a blue bottle is in Room1?
$R.\mathsf{accept}(a \rightarrow b)$	Robot Mary: Yes there is evidence that a blue bottle is in Room1.
Control layer	*Inquiry dialogue terminates*

6.4 Discussion

The extended example has shown how each of the different types of dialogues discussed here—persuasion, information-seeking and inquiry—are implemented and can be applied to a variety of decisions encountered when playing the human-robot Treasure Hunt Game. The example has also demonstrated how multiple dialogues can be running simultaneously, particularly by including several instances of embedded dialogues.

The opening of this article highlighted three specific cases where shared human-robot decision-making would benefit from the flexibility offered by the argumentation-based dialogue approach we have demonstrated. These cases are as follows: (1) responding to discovery, (2) pre-empting failure, and (3) recovering from failure. Specific examples of the first two instances have been demonstrated in this article. The human and robot respond to discovery in Section 6.3, when the robot captures images but cannot identify the contents. The human and robot then engage in argumentation-based dialogue to analyze the images together and cooperatively discover the contents. The human and robot pre-empt failure in Section 6.2, when they disagree about the order in which rooms should be visited. They engage in argumentation-based dialogue in which the robot justifies its reasoning and is able to convince the human that the robot's plan is more energy-efficient.

Although not explicitly described in the extended example, it is easy to see how the demonstrated argumentation-based dialogue system could be used to support the human and robot recovering from failure. The complete Treasure Hunt Game includes a specification for obstacles that could appear dynamically in the robot's arena during game play. For example, a fire or gremlin may appear in the robot's path, blocking its movement, and the robot may need to interrupt its journey to report the error to the human. Then, using the argumentation-based dialogue method, the human and robot can discuss alternative goals.

7. Related Work

This section briefly outlines some work within the HRI community that discusses cooperative relationships between humans and robots.

Scholtz (2003) defines different roles that humans may undertake when operating alongside a robot: bystander, supervisor, operator, mechanic, programmer, and teammate. In the "bystander" case, the human is an observer who has no physical interaction or direct communication with the robot. In the middle cases, the human has a dominating role over the robot in which the human either tells the robot what to do ("supervisor" and "operator") or actually constructs ("mechanic") or programs ("programmer") the robot to perform a specific task. In the last case, the robot and human interact as peers, where they collaborate and discuss ideas about which task(s) to undertake 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 performs actions. As we have shown, argumentation-based dialogue is one way to achieve this.

Much of the work on human-robot cooperation involves less explicit communication than we have explored here. For example, Ogata, Sugano, and Tani (2004) experimented with a person and a mobile robot leading each other around an obstacle course. Communication was effected primarily by pushing and pulling each other. Rosenthal, Biswas, and Veloso (2010) also experimented with a robot that leads people through an office building to attend meetings. The robot can have trouble localizing, so it is programmed to ask people for help.

In work where communication is important, the focus is less on the content of the communication than the delivery. In Scheutz, Schermerhorn, and Kramer (2006), the focus is on interpreting tone of voice and facial expressions. In Chidambaram, Chiang, and Mutlu (2012), aspects such as eye contact, proximity, and vocal cues are employed to persuade a human subject to perform actions.

TeamTalk (Marge, Pappu, Frisch, Harris, & Rudnicky, 2009) is a multi-modal, natural language human-robot interface that is widely cited. The system is capable of interpreting spoken-dialogue interactions (as well as mouse clicks and pen gestures) using the *Olympus* spoken-dialog framework. The dialogue system performs the following three tasks: understanding what the user says/types (and in multimodal systems, the user's gaze, gestures, etc.); planning an appropriate response; and generating natural language to express the response to the user (again, in multimodal systems, this might

also include animating a character, displaying particular graphics, etc). In Olympus, these functions are performed by the following three components: *Phoenix*—a robust parser using context-free grammars; *RavenClaw*—a framework to build dialogue managers; and *Rosetta*—a template-based natural language generation.

None of these systems employ logical argumentation or argumentation-based dialogue. As demonstrated in the previous section, our approach produces a system that allows for flexible initiative-taking and provides a more robust approach to handling failure than other systems.

8. Conclusion

We have presented a model for HRI that supports flexible and dynamic argumentation-based dialogue. We have described our methodology for implementing dialogue protocols to support human-robot collaboration, based on theoretical models found in the literature on argumentation and argumentation-based dialogue. We also described the application of control layers required for those dialogue protocols and illustrated how these can be used to engage agents in combinations of dialogues simultaneously (or sequentially). Our methodology applies theoretical models in a real-time setting and contributes to both HRI and argumentation. In addition, we have introduced a formal model for the Treasure Hunt Game and demonstrated how the argumentation-based dialogue game implementation can be applied effectively to that domain.

We do not claim that our ArgHRI system is a complete representation of argumentation theory or argumentation-based dialogues; indeed, our model is purposely simple. Our goal is to capture the essence of argumentation as a structured means to guide practical human-robot dialogue. Argumentation theory, which provides a computational model that can support arbitrarily long chains of reasoning, is suitable for multi-agent environments, but inappropriate for systems that reason with humans—purely because humans can only keep track of a limited chain of reasoning. We also do not claim that ArgHRI is the definitive human-robot dialogue framework. Despite significant progress in recent years, particularly with commercial products[8], the area of human-robot dialogue domain is still in its infancy. The practical implementation of argumentation-based dialogues is also in its infancy. ArgHRI is an early attempt to demonstrate the utility of a proven theoretical model adapted to a dynamic physical environment to provide dialogue-based support for a robot to "challenge" or "persuade" a human collaborator, and vice versa.

Our ArgHRI system has been implemented and tested using physical and simulated robots (Azhar, 2015; Azhar, Schneider, et al., 2013). Our robot control architecture builds on a framework we designed and implemented to support experimentation in human-robot teams, called *HRTeam* (Sklar et al., 2011; Sklar, Parsons, Özgelen, et al., 2013). ArgHRI employs the *ArgTrust* engine (Tang, Cai, McBurney, Sklar, & Parsons, 2012) to compute the derivations, such as those demonstrated in the examples outlined in Section 6. A dialogue manager component facilitates the control layer and also manages the robot's belief set.

To date, we have conducted two user studies with our ArgHRI system. The first, described in Azhar (2012) and Azhar, Schneider, et al. (2013), employed only one type of dialogue game (persuasion). The second study was recently completed, with all three types of dialogues, and has produced metrics on the efficacy of playing the Treasure Hunt Game both using dialogue games and without any dialogue support, as well as user feedback providing subjective views regarding game play in both modes (Azhar, 2015). In both studies, the results comparing games played with and without argumentation-based dialogue support showed that users trusted the robot more when they played with dialogue. Detailed results and analysis of the most recent study are forthcoming.

[8] For example, Pepper, by Aldebaran, `https://www.aldebaran.com/en/a-robots/who-is-pepper`

Acknowledgements

The authors are grateful for the help and advice from Professor Simon Parsons, who provided invaluable comments on drafts of this paper, as well as Professor Peter McBurney for input on the application of argumentation-based dialogue theory.

This work was partially funded by the US National Science Foundation (NSF) under grants #IIS-1116843, #IIS-1338884 and #CNS-1117761, by the US Army Research Office under the Science of Security Lablet grant (SoSL), by the US Army Research Laboratory under the Network Science Collaborative Technology Agreement, by a University of Liverpool Research Fellowship, and by a Fulbright-King's College London Scholar Award. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of the funders.

References

Austin, J. L. (1975). *How to do things with words* (Vol. 88). Cambridge, MA: Harvard University Press.

Azhar, M. Q. (2012). Toward an argumentation-based dialogue framework for human-robot collaboration. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI)* (pp. 305–308). Santa Monica, CA.

Azhar, M. Q. (2015). *Toward an argumentation-based dialogue framework for human-robot collaboration*. Ph.D. thesis, The Graduate Center, City University of New York, New York, NY.

Azhar, M. Q., Parsons, S., & Sklar, E. I. (2013). An argumentation-based dialogue system for human-robot collaboration (Demonstration). In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1353–1354). St. Paul, MN.

Azhar, M. Q., Schneider, E., Salvit, J., Wall, H., & Sklar, E. I. (2013). Evaluation of an argumentation-based dialogue system for human-robot collaboration. In *Workshop on Autonomous Robots and Multirobot Systems (ARMS) at the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS).* St. Paul, MN.

Bohus, D., Horvitz, E., Kanda, T., Mutlu, B., & Raux, A. (2011). Introduction to the Special Issue on Dialog with Robots. *AI Magazine*, *32*(4), 15–16.

Bohus, D., Raux, A., Harris, T. K., Eskenazi, M., & Rudnicky, A. I. (2007). Olympus: An open-source framework for conversational spoken language interface research. In *Proceedings of the HLT-NAACL Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technology* (pp. 32–39). Rochester, NY.

Carbonell, J. R. (1970, May 31). *Mixed-Initiative Man-Computer Instructional Dialogues, Final Report* (Tech. Rep. No. BBN-1971). Cambridge, MA: Bolt Beranek and Newman, Inc. (BBN).

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., et al. (2013). Towards Empathic Virtual and Robotic Tutors. In *Artificial Intelligence in Education, Lecture Notes in Computer Science* (Vol. 7926, pp. 733–736).

Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). 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 (HRI)* (pp. 293–300). Boston, MA.

Cohen, A., Parsons, S., Sklar, E. I., & McBurney, P. (2014). *Support between Rule-Based Arguments* (Tech. Rep.). New York, NY: Graduate Center, City University of New York.

Habib, M. K. (2007). Humanitarian demining: Reality and the challenge of technology. *International Journal of Advanced Robotic Systems*, *4*(2), 151–172.

Hart, P., Nilsson, N., & Raphael, B. (1968). A formal basis for the heuristic determination of minimal cost paths. *IEEE Transactions on Systems Science and Cybernetics*, *4*(2), 100–107.

Horvitz, E. (1999). Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 159–166). Pittsburgh, PA.

Hulstijn, J. (2000). *Dialogue models for inquiry and transaction*. Unpublished doctoral dissertation, Universiteit Twente, Netherlands.

Jones, E. G., Browning, B., Dias, M. B., Argall, B., Veloso, M., & Stentz, A. (2006). Dynamically formed het-

erogeneous robot teams performing tightly-coordinated tasks. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (pp. 570–575). Orlando, FL.

Lemon, O., Gruenstein, A., & Peters, S. (2002). Collaborative activities and multi-tasking in dialogue systems. *Traitement Automatique des Langues, Special Issue on Dialogue*, *43*(2), 131–154.

Marge, M. R., Pappu, A. K., Frisch, B., Harris, T. K., & Rudnicky, A. I. (2009). Exploring spoken dialog interaction in human-robot teams. In *Proceedings of the ACM/IROS Workshop on Robots, Games, and Research: Success Stories in USARSim.* St. Louis, MO: ACM.

Matthews, J. T. (2002). The Nursebot Project: Developing a personal robotic assistant for frail older adults in the community. *Home Health Care Management & Practice*, *14*(5), 403–405.

McBurney, P., & Parsons, S. (2001). Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence*, *32*(1–4), 125–169.

McBurney, P., & Parsons, S. (2002). Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language, and Information*, *11*(3), 315–334.

McBurney, P., & Parsons, S. (2009). Dialogue games for agent argumentation. In G. Simari & I. Rahwan (Eds.), *Argumentation in Artificial Intelligence* (pp. 261–280). Berlin: Springer.

Modayil, J. (2010). Learning grounded communicative intent from human-robot dialog. In *Proceedings of the AAAI Fall Symposium on Dialog with Robots.* Arlington, VA: AAAI.

Modgil, S., & Prakken, H. (2013). A general account of argumentation and preferences. *Artificial Intelligence*, *195*, 361–397.

Murphy, R. R., Casper, J., & Micire, M. (2001). Potential tasks and research issues for mobile robots in RoboCup Rescue. In *Robot Soccer World Cup IV, Lecture Notes in Artificial Intelligence* (Vol. 2019, pp. 339–344). Berlin: Springer Verlag.

Nilsson, N. J. (1984). *Technical note* (Tech. Rep. No. 323). Palo Alto, CA: SRI International.

Ogata, T., Sugano, S., & Tani, J. (2004). 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* (pp. 435–444).

Özgelen, A. T., & Sklar, E. I. (2013, May). A task complexity assessment tool for single-operator multi-robot control scenarios (Demonstration). In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1393–1394). St. Paul, MN.

Özgelen, A. T., & Sklar, E. I. (2014, March). Modeling and analysis of task complexity in single-operator multi-robot teams (Late Breaking Report). In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 262–263). Bielefeld, Germany.

Parsons, S., McBurney, P., Sklar, E. I., & Wooldridge, M. (2007). On the relevance of utterances in formal inter-agent dialogues. In *Proceedings of the 6th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1002–1009). Honolulu, HI.

Parsons, S., Wooldridge, M., & Amgoud, L. (2003a). On the outcomes of formal inter-agent dialogues. In J. S. Rosenschein, M. Wooldridge, T. Sandholm, & M. Yokoo (Eds.), *2nd International Conference on Autonomous Agents and Multi-Agent Systems* (pp. 616–623). Melbourne, Australia: ACM Press.

Parsons, S., Wooldridge, M., & Amgoud, L. (2003b). Properties and complexity of formal inter-agent dialogues. *Journal of Logic and Computation*, *13*(3), 347–376.

Prakken, H. (2006). Formal systems for persuasion dialogue. *Knowledge Engineering Review*, *21*(2), 163–188.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, *1*(2), 93–124.

Rahwan, I., & Simari, G. R. (Eds.). (2009). *Argumentation in Artificial Intelligence*. Berlin: Springer Verlag.

Rosenthal, S., Biswas, J., & Veloso, M. (2010). An effective personal mobile robot agent through symbiotic human-robot interaction. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 915–922).

Santana, P. F., Barata, J., & Correia, L. (2007). Sustainable robots for humanitarian demining. *International Journal of Advanced Robotic Systems*, *4*(2), 207–218.

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). 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 (HRI)* (pp. 226–233). Salt Lake City, UT.

Scholtz, J. (2003, January). Theory and evaluation of human robot interactions. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS).* Big Island, HI.

Sklar, E. I., & Azhar, M. Q. (2011). Toward the application of argumentation to interactive learning systems. In *Proceedings of the Workshop on Argumentation in Multiagent Systems (ArgMAS) at the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 213–230). Taipei, Taiwan.

Sklar, E. I., Azhar, M. Q., Parsons, S., & Flyr, T. (2013). Enabling human-robot collaboration via argumentation (Extended Abstract). In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1251–1252). St. Paul, MN.

Sklar, E. I., Özgelen, A. T., Muñoz, J. P., Gonzalez, J., Manashirov, M., Epstein, S. L., et al. (2011, May). Designing the HRTeam framework: Lessons learned from a rough-and-ready human/multi-robot team. In *Workshop on Autonomous Robots and Multirobot Systems (ARMS) at the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS).* Taipei, Taiwan.

Sklar, E. I., & Parsons, S. (2004). Towards the application of argumentation-based dialogues for education. In *Proceedings of the 3rd International Conference of Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1420–1421).

Sklar, E. I., Parsons, S., Epstein, S. L., Özgelen, A. T., Muñoz, J. P., Schneider, E., et al. (2012, July). Demonstration: Investigating Human/Multi-Robot Team Interaction. In *Demonstration at the AAAI Robotics and Multimedia Fair.* Toronto, Canada.

Sklar, E. I., Parsons, S., Özgelen, A. T., Schneider, E., Costantino, M., & Epstein, S. L. (2013, May). HRTeam: a framework to support research on human/multi-robot interaction (Demonstration). In *Proceedings of 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1409–1410). St. Paul, MN.

Sklar, E. I., Parsons, S., & Singh, M. P. (2013, May). Towards an argumentation-based model of social interaction. In *Proceedings of the Workshop on Argumentation in Multiagent Systems (ArgMAS) at the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS).* St. Paul, MN.

Tang, Y., Cai, K., McBurney, P., Sklar, E. I., & Parsons, S. (2012). Using argumentation to reason about trust and belief. *Journal of Logic and Computation, Special Issue on Agreement Technologies*, *22*(5), 959–1018.

Torrey, C., Powers, A., Marge, M., Fussell, S. R., & Kiesler, S. (2006). Effects of adaptive robot dialogue on information exchange and social relations. In *Proceedings of the ACM Conference on Human-Robot Interaction (HRI)* (pp. 126–133). Salt Lake City, UT: ACM.

Wada, K., Shibata, T., Saito, T., & Tanie, K. (2002). Robot assisted activity for elderly people and nurses at a day service center. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (Vol. 2, pp. 1416–1421). Washington, DC.

Walton, D. N., & Krabbe, E. C. W. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. Albany, NY: SUNY Press.

Yanco, H., Baker, M., Casey, R., Keyes, B., Thoren, P., Drury, J. L., et al. (2006). Analysis of Human-Robot Interaction for Urban Search and Rescue. In *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics* (pp. 22–24). Piscataway, NJ.

Authors' names and contact information: Elizabeth I. Sklar, Department of Informatics, King's College London, UK. Email: elizabeth.sklar@kcl.ac.uk. M. Q. Azhar, Borough of Manhattan Community College, City University of New York, USA. Email: mazhar@bmcc.cuny.edu.