Multi-modal Belief Updates in Multi-Robot Human-Robot Dialogue Interactions

Gordon Briggs[1] and **Matthias Scheutz**[2]

Abstract. Humans working in teams typically use task-based natural language dialogues to coordinate activities. And they use mental models of team mates which they update automatically based on perceived and communicated information to predict the actions of their team mates. It is thus reasonable to assume that humans will expect future robots interacting with humans in natural language as part of mixed-initiative teams to exhibit the same kinds of belief modeling exhibited by humans.

In this paper, we propose principles that robots can use to represent beliefs and goals of other agents based on task-based natural language dialogues and use automatic inference based on communicated information to update their mental models of other agents. We demonstrate the proposed principles in a simple case study involving two robots and a human operator performing simple tasks in a laboratory environment.

1 Introduction

Mixed human-robot initiatives – teams that consist of both human and robotic team members – are widely seen as an important application domain for future autonomous robots. The goal in such teams is to utilize unique strengths of both humans and robots in order to accomplish joint goals. For example, NASA envisions space robots to help astronauts with the construction of planetary space stations. Or rescue robots in disaster areas are envisioned to aid human rescue workers in finding and retrieving wounded people. From a robotics perspective, the research challenge here is twofold: to provide the robotic capabilities necessary for a given task and to provide appropriate mechanisms for human-robot interactions that are effective and natural for humans.

While human teams typically use natural language to coordinate activities (such as discussing goals, developing plans, adjusting behaviors, etc.), mixed initiative teams are severely limited by current robots' cognitive limitations. Current robotic systems do not have the necessary modeling and inferencing capabilities for extensively emulating human mental models, nor do they have the natural language capabilities to engage in natural task-based dialogues, though progress is being made on these fronts [1]. Specifically, humans in teams are capable of (1) following multi-agent dialogues, (2) automatically updating their mental models of the involved agents based on the information communicated in natural language, and (3) automatically drawing inferences from the obtained information which may prompt them to confirm, augment or correct information and communicate those updates to their interlocutors effectively [6, 8]. If we want mixed initiative teams to interact in natural human-like ways, then robots will need mechanisms (built-in or learned) for performing mental and belief modeling and updating very much like humans.

In this paper, we propose simple belief update schemes for multi-agent dialogues that can be integrated into a cognitive architecture, thus allowing artificial agents to engage in more natural dialogues for activity coordination in mixed initiative teams. Specifically, we show how robots can use information gained from listening to dialogues among other agents to make inferences about those agents' belief states and goals, and how agents can use automatic inferences applied to mental models of other agents to better understand natural language directives and arrive at explicit goal representation (of their own and other agents' goals).

The paper is organized as follows. We begin with a few motivating mixed initiative scenarios where a human commander instructs autonomous robots to perform various tasks. These scenarios are intended to isolate several of the principles humans automatically employ in the context of teams and underscore the importance of belief-modeling mechanisms and perceptual integration. Then, we formalize the principles and describe our framework for belief modeling and updating, which also includes principles for inferring belief state based on particular natural language expressions. Next, we introduce the evaluation scenario and present the implementation details of the previously introduced principles in a distributed robotic architecture run on two robots. A particular dialogue interaction between the operator and the two robots then demonstrates the operation and utility of the proposed principles and framework. The subsequent discussion section addresses some of the challenges for larger, more capable systems while the conclusion summarizes our accomplishments and briefly touches on future work.

2 Motivation

Robotic systems are often well-suited for operating along-side or in place of humans, for example, in hazardous environments such as nuclear power plants or outer space. However, for humans to work with robots effectively, the interaction and cognitive capabilities of the robot become a critical factor, in addition to its physical characteristics and behavioral repertoire. While it is possible to devise special purpose interfaces that allow for teleoperation of robots or interactions with robots capable of limited autonomy, the much more natural case – from a human perspective – is where humans interact with *autonomous* robotic teams members as they would with other human team members: *using natural language*. This is particularly important in cases where typical human-machine communication

[1] Human-Robot Interaction Lab, Tufts University, Medford, MA, 02155, USA. gbriggs@cs.tufts.edu
[2] Human-Robot Interaction Lab, Tufts University, Medford, MA, 02155, USA. mscheutz@cs.tufts.edu

modalities are impractical [7]. For example, human searchers during rescue operations in disaster zones typically coordinate their activities through spoken natural language interactions via wireless audio links, while simultaneously occupying their eyes and hands with time-critical work. To efficiently interact with human team members, robotic team members would then also have to be capable of using spoken natural language.

However, being capable of using (rudimentary) spoken language alone is not sufficient because so much other cognitive activity is triggered in humans when humans engage in even simple dialogue exchanges. For example, human team members automatically form *mental models* of the beliefs and goals other team members have based on dialogue context and use those mental models to make inferences about other states that obtain (e.g., if a searcher says that she is done searching area X and has a goal to search another area, then she will likely leave X) and to adapt their natural language interactions (e.g., A telling B that A searched X will lead allow C to assume that B knows that A left X in questions like "Do you know where A is going next?"). Moreover, if A knows that B wants to be informed if a subtask is completed (e.g., area X search), A will automatically update B ("I finished searching X"). And B can then update C on A's activity if A is temporarily not reachable when C inquires about the status of area X; or B can correct C if C makes a statement that indicates a false belief (e.g., "Area X still needs to be searched" when A previously informed B that the search of area X was completed). Working in teams thus requires agents to monitor the dialogues and employ similar mental modeling and automatic model updates as in the human case. And it also requires similar automatic application of inference principles to communicated information.

To make these types of example more concrete in a simple robotic domain (as we will later use for evaluating our proposal mechanisms), consider a simple environment consisting of three navigation-points. A human-operator (O) is charged with coordinating in natural language the goals and behaviors of two autonomous robots, a quadrotor (Q1) and a ground-transport (T1) via radio. Let us use $at(\alpha, \lambda)$ to denote that agent α is located at nav-point λ and $B(\alpha, \phi)$ to denote that agent α believes that ϕ is true. Consider the following scenario:

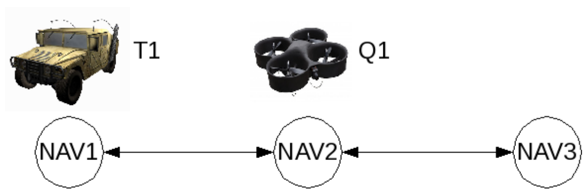

Figure 1. Sample environment for joint human-robot tasks.

```
O:  Transport 1, travel to Nav-point 3.
T1: Okay.
O:  Quadrotor 1, follow Transport 1.
Q1: Okay.
```

The scenario is illustrated in Figure 1. In order for the transport to get to Nav-point 3 in the above scenario, it must pass through Nav-point 2. Q1 can thus simply wait for T1 to show up at Nav-point 2 at which point it can follow it to Nav-point 3.

Note that autonomous agents must not only be able to update their beliefs to reflect the propositions communicated in utterances, but they must be able to compare these propositions to their own perceptions. If contradictions occur, agents must be able to ask for clarification or offer corrective statements as in the following example:

```
O:  Transport 1, are you at Nav-point 2?
T1: Yes, I am at Nav-point 2.
Q1: I do not see Transport 1.
```

In addition to the need for integrating the agent's perceptions into the dialogue system, a belief-modeling competency is necessary to generate the proper clarification or correction statement. Quadrotor 1's response would only make sense if $B(O, at(Q1, N2))$ was true. If $B(O, \neg at(Q1, N2))$, the quadrotor would have to produce a corrective utterance that implies $at(Q1, N2) \wedge \neg Sees(Q1, T1)$, such as "I am at Nav-point 2 and do not see Transport 1, " or "I do not see Transport 1 at Nav-point 2."

In sum, two important competencies must be present in a robotic agent for it to be able to communicate efficiently with human operators and team members: (1) the ability to build and maintain mental models or belief-model of the other agents (based on the synthesis of perceived, communicated, and inferred information) in order to maintain situational awareness; and (2) the ability to support the maintenance of other's mental models of oneself through communicating new information. For the rest of the paper we will use the term "mental model" in a technical sense to refer to the set of beliefs $B(\beta, \phi)$ agent α has about other agents β.

3 Belief Modeling and Dialogue

For robots to be able to engage in simple but natural sounding dialogues and automatically perform the types of belief modeling presented above, we need to add explicit rules that represent relationships among linguistic expressions as well as past and future beliefs. In particular, for task-based dialogue interactions we need to add rules that allow agents to reason about (1) the effects of perceptions, actions, and past beliefs on new/updated beliefs and (2) the effects of different utterance types (i.e., statement, questions, commands, and acknowledgments) on beliefs. The former includes all relevant properties of the agent in the world for it to be able to understand task-based dialogues; the latter includes all kinds of pragmatic implications of the employed utterance types both general (e.g., adverbial modifiers) as well as specific to the communicated context (e.g., the location predicate).

Indeed, these rules that enable belief-modeling of interlocutors are necessary to enable plan-based dialogue agents, which were first explored by Cohen and Perrault (1979) and Perrault and Allen (1980). Traum (1999) provides a review of dialogue agents enabled by the modeling of beliefs, desires, and intentions (BDI) and articulates the advantages of this approach, stating that modeling the changes communicative acts have on the mental models of agents "...allows an agent theorist or designer to place agent communication within the same general framework as agent action."

3.1 Agent Behavioral Rules

For the purposes of the employed example and the given space limitations, we make several simplifying assumptions: (1) we will not worry about employing generalizable and robust mechanisms for the translation from natural language expression to logical formulas for the simple domains employed here (e.g., which we have done elsewhere [5]), we simply use pattern-matching to convert from natural language to logical forms; (2) we assume that all agents are truth tellers and never lie; (3) we assume that all agents immediately execute the most recent order and only that order (e.g., we have previously dealt with the more complex case of giving agents multiple possibly contradictory orders in natural language [12]); and (4) we make each agent first utter its name and then the name of the addressee so that it is easy for each agent to determine the speaker and the intended listener based on the linguistic information alone.

These rules include facts about agent behaviors that other agents can use to predict the other agents' behaviors. The first rule is concerned with an agent's perceptual system which is taken to automatically generate beliefs about what it perceives. In particular, if an agent α perceives the presence of another agent β at location λ, then it generates the belief that $B(\beta, \lambda)$.

$$Perceives(\alpha, at(\beta, \lambda)) \Rightarrow B(\alpha, at(\beta, \lambda)) \quad (1)$$

The next three rules are about agent actions: If agent α has a goal to be at location λ, then α is heading there:

$$goal(\alpha, at(\alpha, \lambda)) := goingTo(\alpha, \lambda) \quad (2)$$

If agent α is supposed to follow agent β, and β is heading to location λ, α is also going to λ:

$$follow(\alpha, \beta) \land goingTo(\beta, \lambda) := goingTo(\alpha, \lambda) \quad (3)$$

The next rule pertains to triggering a notification event. If you are supposed to inform agent β when a condition ϕ is achieved, then when ϕ is achieved, generate an intention-to-know ϕ by β, which will leverage the dialogue generation capabilities of the agent:

$$Inform(\beta, \phi) \land \phi := IK(\beta, \phi) \quad (4)$$

3.2 Belief Update Rules for Utterances

We need to add rules for handling utterances both from a speaker's and a listener's perspective. Here we will build on our recently introduced formal framework [4] where we use $[[u]]_c$ to denote the "pragmatic meaning" of an utterance u in context c (which includes task, goal, belief and discourse aspects).

The first general rule (based on the above discussed simplifications) is that an agent always believes all propositions it is able to infer from the utterance of another agent:

$$([[u]]_c \Rightarrow_{\alpha}^{b} \phi) \land Heard(\alpha, u) \Rightarrow B(\alpha, \phi) \quad (5)$$

Note that the inferences here are bounded by the agent's computational and algorithmic inference limitations (indicated by the agent's inference $\Rightarrow_{\alpha}^{b}$ mechanism bounded by b). While this rule is reasonable for simple agents in limited task-domains and might allow an agent to generate all implications given by an utterance in context C, it is likely that more sophisticated agents in more complex domains will not be able to generate all implications.

The second rule is that an agent believes everything it said itself:

$$([[u]]_c \Rightarrow_{\alpha}^{b} \phi) \land Said(\alpha, u) \Rightarrow B(\alpha, \phi) \quad (6)$$

This rule comports with Gricean conversation maxims, specifically the maxim of quality, which requires one to not say what one believes is false [9]. Though it would fail in cases of intentional deception, it is assumed that in the domain of collaborative HRI, such cases are not to be expected. Such a rule would rely on feedback utterances, such as acknowledgments (e.g. "OK") and/or reiterations (e.g. "Yes, I am going to nav-point 3.") to maintain correct mental-models.

In addition to adding general rules for utterances based on speaker and listener roles, we need to add more specific rules for capturing the pragmatic implications of different utterance types such as statements, questions, commands and acknowledgments based on prior dialogue history and sentential modifiers. We present several pragmatic rules below in the form of $UtteranceType(\alpha, \beta, X, M)$, where α denotes the speaker, β denotes the audience, X denotes the surface semantics, and M denotes the set of sentential modifiers present in the utterance (which may be the empty set, denoted here as $\{\}$). Pragmatic rules for various adverbial modifiers in this domain (such as "still" and "now") were presented in [4].

3.2.1 *Statements*

If α informs β that it is at λ, then we can assume that α is indeed at that location:

$$[[Stmt(\alpha, \beta, at(\alpha, \lambda), \{\})]]_c := at(\alpha, \lambda) \quad (7)$$

If α informs β that it is going to λ, then we can assume that α is indeed going to location λ:

$$[[Stmt(\alpha, \beta, goingTo(\alpha, \lambda), \{\})]]_c := goingTo(\alpha, \lambda) \quad (8)$$

If α informs β that it is going to λ with γ, then we can assume that α is indeed going to location λ and believes γ is doing the same:

$$[[Stmt(\alpha, \beta, at(\beta, \lambda), \{with(\gamma)\})]]_c := goingTo(\alpha, \lambda) \land B(\alpha, goingTo(\gamma, \lambda)) \quad (9)$$

If α informs β that it is engaged in action θ, then we can assume that α is indeed doing that action:

$$[[Stmt(\alpha, \beta, doing(\alpha, \theta), \{\})]]_c := doing(\alpha, \theta) \quad (10)$$

3.2.2 *Questions*

If α asks β about its location in the general sense ("where are you?"), then one can infer that α has an intention to know (expressed via the "IK" operator) where β is located:

$$[[Ask_{loc}(\alpha, \beta, \{\})]]_c := IK(\alpha, at(\beta, \lambda)) \quad (11)$$

for some λ.

If α asks β about its heading in the general sense ("where are you going?"), then one can infer that α has an intention to know the location where β is traveling to:

$$[[Ask_{goto}(\alpha, \beta, \{\})]]_c := IK(\alpha, goingTo(\beta, \lambda)) \quad (12)$$

for some λ.

If α asks β about its current action in the general sense ("what are you doing?"), then one can infer that α has an intention to know the current action that β is engaged in, which is specified by the $doing()$ predicate:

$$[[Ask_{doing}(\alpha, \beta, \{\})]]_c := IK(\alpha, doing(\beta, \theta)) \quad (13)$$

for some action θ.

3.2.3 Commands

If α orders β to travel to λ, then one can infer that β has a goal to be at λ, α wishes to be informed when β reaches λ, and that α wants to know whether β heard the command:

$$
\begin{aligned}
&[[Cmd(\alpha, \beta, at(\beta, \lambda), \{\})]]_c := \\
&\quad G(\beta, at(\beta, \lambda)) \wedge Inform(\alpha, at(\beta, \lambda)) \\
&\quad \wedge IK(\alpha, Heard(\beta, G(\beta, at(\beta, \lambda))))
\end{aligned} \tag{14}
$$

If α orders β to follow γ, then one can infer that β has a goal to follow γ and α wants to know whether β heard the command:

$$
\begin{aligned}
&[[Cmd(\alpha, \beta, at(\beta, \lambda), \{\})]]_c := \\
&\quad follow(\beta, \gamma) \wedge IK(\alpha, Heard(\beta, follow(\beta, \gamma)))
\end{aligned} \tag{15}
$$

If α orders β to travel to λ, then one can infer that β has a goal to be at λ, α wishes to be informed when β reaches λ, and that α wants to know whether β heard the command:

$$
\begin{aligned}
&[[Cmd(\alpha, \beta, at(\beta, \lambda), \{\})]]_c := \\
&\quad G(\beta, at(\beta, \lambda)) \wedge Inform(\alpha, at(\beta, \lambda)) \\
&\quad \wedge IK(\alpha, Heard(\beta, G(\beta, at(\beta, \lambda))))
\end{aligned} \tag{16}
$$

3.2.4 Acknowledgments

If α utters an acknowledgment (e.g., "OK.") when the previous utterance was a positive statement of location by β, then one can infer α no longer has the intention to know β's location:

$$[[Ack(\alpha, \beta, \{\})]]_c := \neg IK(\alpha, at(\beta, \lambda)) \tag{17}$$

for some λ where for any M $Prior(Stmt(\beta, \alpha, at(\beta, \lambda), \{M\})) \in c$. If α utters an acknowledgment (e.g., "OK.") when the previous utterance was a command by β to be at λ, then one can infer that

$$
\begin{aligned}
&[[Ack(\alpha, \beta, \{\})]]_c := \\
&\quad G(\alpha, at(\alpha, \lambda)) \wedge heard(\alpha, G(\alpha, \lambda))
\end{aligned} \tag{18}
$$

where $Prior(Cmd(\beta, \alpha, at(\alpha, \lambda), \{M\})) \in c$ If α utters an acknowledgment (e.g., "OK.") when the previous utterance was a command by β to follow γ, then one can infer that

$$
\begin{aligned}
&[[Ack(\alpha, \beta, \{\})]]_c := \\
&\quad follow(\alpha, \gamma) \wedge heard(\alpha, meet(\alpha, \gamma))
\end{aligned} \tag{19}
$$

where $Prior(Cmd(\beta, \alpha, at(\alpha, \lambda), \{M\})) \in c$

3.3 Belief Updates

Each agent γ updates its beliefs whenever it hears an utterance u from speaker α addressing another agent β (which may or may not be the same agent as γ) or whenever it receives a set of perceptual updates Ψ_γ. It uses the above specified principles to determine all pragmatic implications of the utterance and also to detect any beliefs inconsistent with existing beliefs (both pragmatic implications and inconsistency detection are determined by γ's inference algorithm $\Rightarrow_\gamma^b$ and are thus subject to b – for low values of b the agent might fail to compute all implications or to derive a contradiction); the set of conflicting beliefs P_γ are then removed from the agent γ's sets of beliefs.

4 CASE STUDY

All principles and belief updates described above were implemented as a special dialogue component in the Java-based *Agent Development Environment* (ADE) (see http://ade.sourceforge.net/) which is a framework for implementing distributed architectural components for robotic architectures. A simple resolution-style inference mechanism with a shallow one-step look-ahead search limit was used. The new dialogue component (in conjunction with previous algorithms for utterance generation and response selection as detailed in [4]) was used integrated into the existing robotic DIARC architecture which comprises components for perceptual processing (using camera-based vision) and navigation (for ground-based and air-based vehicles), action planning and natural language processing and has been used extensively for human-robot interactions in natural language [13]. For the case study, we used a Videre Erratic mobile robot and Parrot AR Drone Quadricopter from ExPansys. A picture of the platforms used can be found in Figure 2, while video of the interaction can be found online.

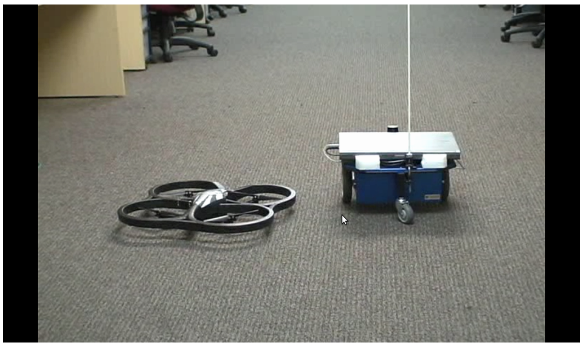

Figure 2. The quadricopter (left) and Videre (right) robotic platforms utilized for the study[4].

As illustrated in Figure 2, both the quadrotor (Q1) and the Videre transport (T1) start in the same location, which we designate S. The belief-spaces of both agents are initialized to be empty, though both are able to perceive that they are at the starting location:

$$
\begin{aligned}
\Psi_{Q1} &:= \{at(Q1, S)\} \rightarrow B_{Q1} := \{at(Q1, S)\} \\
\Psi_{T1} &:= \{at(T1, S)\} \rightarrow B_{T1} := \{at(T1, S)\}
\end{aligned}
$$

The human operator (O) then queries the quadrotor:

```
O: Drone, what are you doing?
```

Since the quadrotor is idle, and has no $doing(Q1, \theta)$ terms in its belief-space, the quadrotor replies accordingly and the operator acknowledges:

```
Q1: Commander, I am not doing anything.
O: Okay.
```

The operator then gives the transport an order to travel to location alpha:

[4] http://www.youtube.com/watch?v=40_Ee2g5ztg

```
O: Transport, go to alpha.
```

Both the transport and quadrotor hear this utterance, and they update their beliefs accordingly:

$$
\begin{array}{l}
u := parse(\text{"O: T1, go to alpha."}) \\
\rightarrow u := Cmd(T1, O, at(T1,\alpha), \{\}) \\
[[u]]_c := \{G(T1, at(T1,\alpha)), Inform(O, at(T1,\alpha)), \\
...IK(O, heard(T1, G(T1, at(T1,\alpha))))\} \\
\bar{P}_{Q1} := contradictedTerms([[u]]_c, B_{Q1}) \\
\bar{P}_{T1} := contradictedTerms([[u]]_c, B_{T1}) \\
B_{Q1} := (B_{Q1} - \bar{P}_{Q1}) + [[u]]_c \\
B_{T1} := (B_{T1} - \bar{P}_{T1}) + [[u]]_c
\end{array}
$$

The belief-spaces of both agents are consequently:

$$
\begin{array}{l}
B_{Q1} := \{at(Q1,S), G(T1, at(T1,\alpha)), \\
...Inform(O, at(T1,\alpha)), \\
...IK(O, heard(T1, G(T1, at(T1,\alpha))))\} \\
B_{T1} := \{at(T1,S), G(T1, at(T1,\alpha)), \\
...Inform(O, at(T1,\alpha)), \\
...IK(O, heard(T1, G(T1, at(T1,\alpha))))\}
\end{array}
$$

As the previous command utterance was directed at T1, the agents assume the dialogue's turn is passed to T1. The transport subsequently satisfies the operator's intention to know the command was heard by generating an acknowledgment utterance:

```
T1: Okay.
```

The transport then begins to travel to location α, adding the $doing(T1, goingTo(T1,\alpha))$ term to its belief-space. The operator then gives the quadrotor an order to follow the transport, which is followed by the transport's acknowledgment:

```
O: Drone, follow transport.
Q1: Okay.
```

Having no other goals, the quadrotor then begins to follow the transport, adding the $doing(Q1, follow(Q1, T1))$ term to its belief-space. At this point, the belief-spaces of each agent are:

$$
\begin{array}{l}
B_{Q1} := \{at(Q1,S), G(T1, at(T1,\alpha)), \\
...Inform(O, at(T1,\alpha)), follow(Q1,T1), \\
...doing(Q1, follow(Q1,T1))\} \\
B_{T1} := \{at(T1,S), G(T1, at(T1,\alpha)), \\
...Inform(O, at(T1,\alpha)), follow(Q1,T1), \\
...doing(T1, goingTo(T1,\alpha))\}
\end{array}
$$

Later, the operator queries the quadrotor:

```
O: Drone, what are you doing?
```

Retrieving the appropriate term from its beliefs, the quadrotor responds in accordance with Rule 10:

```
Q1: Commander, I am following transport.
O: Okay.
```

The operator subsequently asks the quadrotor about its destination:

```
O: Drone, where are you going?
```

Because the quadrotor has previously heard that the transport was commanded to go to α, Q1 is able to infer:

$$
\begin{array}{l}
G(T1, at(T1,\alpha)) := goingTo(T1,\alpha) \\
goingTo(T1,\alpha) \wedge follow(Q1,T1) := \\
goingTo(Q1,\alpha) \\
\rightarrow goingTo(Q1,\alpha)
\end{array}
$$

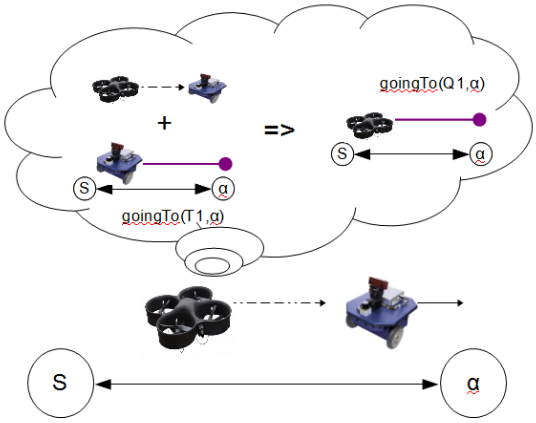

Figure 3. The quadrotor infers its destination.

The above inference is depicted in Figure 3. Based on this inference, Q1 responds accordingly:

```
Q1: Commander, I am going to alpha with
transport.
O: Okay.
```

At this point, we should clarify that there is a distinction between the robot having a explicit goal to undertake and action and having the knowledge that it is undertaking an action. Having a goal to perform an action (or achieve some state) will modulate the behavior of the robot, while just having knowledge of the current action will not necessarily affect the system's behavior. In this case the inference made in Figure 3 is only resulting in knowledge that the robot is going to the final location alpha, but not a *goal* to go to location alpha. Thus, the quadrotor will not be in danger of incorrectly following the transport by predicting the final location and traveling there (instead of following).

We should also contrast the two questions we have just examined. "What are you doing?" we have interpreted as a query to ascertain the robot's current goal, whereas "Where are you going?" seeks more specific knowledge from the robot that can only be answered by making the inference in Figure 3. We believe that, given the same level of knowledge and dialogue history, it would be plausible for a human to answer the question in a similar manner, and as such, the dialogue interaction is made more natural by enabling the robot to do the same.

Later, when the transport finally reaches location alpha, the transport is able to perceive it's new location. The transport updates its belief-space accordingly:

$$
\begin{array}{l}
\Psi_{T1} := \{at(T1,\alpha)\} \\
\rightarrow B_{T1} := \{at(T1,\alpha), Inform(O, at(T1,\alpha)), ...\}
\end{array}
$$

Inferring via Rule 4 the operator's intention to be notified of its arrival at the intended destination, the transport states:

```
T1: Commander, I am now at alpha.
```

5 DISCUSSION

The dynamics of human teams are complex and multifarious, deeply integrating and intertwining natural language exchanges and actions. Humans are extremely good at building mental models of their team mates that include general team mate characteristics as well as particular team mate beliefs, goals and intentions. And humans can effortlessly use all this knowledge to make quick inferences about the mental states of their team mates based on information gleaned from natural language interactions and the details of how that information was linguistically expressed. Most importantly, humans will expect future robots, in particular if they are otherwise very capable, to be able to perform the same kind of mental modeling and to make the same kinds of quick automatic inferences as part of task-based natural language dialogues.

We have introduced a set of principles that can form the basis of a mental modeling mechanism that is deeply integrated with the natural language dialogue mechanisms. The formalism captures perceptual and behavioral aspects of agents as well as their beliefs and intentions/goals. It also allows for different models and model updates for different agents (e.g., how an agent reacts to a particular command given by the operator) by allowing for the definition of agent-specific update rules. And it provides a natural level of abstraction where agents can introspect on their own behaviors and behavioral dispositions in an effort to model themselves and other agents.

Similar challenges involving utilizing natural language communication and maintaining situation-awareness have been investigated in [2] and in the multi-robot domain in [3]. In contrast with these approaches, our approach so far involves simple reactive agents, rather than agents with planning capabilities.

Beyond the sophistication of our agents, our current approach has additional shortcomings. First, it is unclear how far the search depth of the inference algorithm can be reasonable extended if more dialogue principles are added without losing real-time processing. Clearly, there will be limits to the set of propositions an agent can derive automatically given the number of pragmatic and agent-based rules. To curb the complexity and avoid generating thousands of irrelevant beliefs, it will become necessary to incorporate a notion of relevance that allows for targeted inference (also to derive contradictions as part of belief updates). Finally, the current version makes several simplifying assumptions (e.g., about perceptions and behavioral decisions) that will clearly be too simple for more complex tasks and agents. For instance, our communication is currently accomplished individually between single agents. Belief update rules need to be extended to account for group communication [14, 11]. The problem of collaborative planning, in which agents must work together to develop a joint plan, poses further challenges in that agents must have the ability to communicate and reason about partial candidate plans [10].

6 CONCLUSIONS

In this paper, we introduced new principles for belief modeling and updating for autonomous agents (such as robots or virtual characters) interacting with humans and other autonomous agents in mixed initiative teams through spoken natural language dialogues. We showed how we can represent beliefs and intentions of other agents to generate mental models that are rich enough to capture task-based aspects of other agents and their beliefs. We also showed how a robot can update its mental model of another robot based on task-based utterances it heard and how it can automatically apply inference-rules to the information obtained from the utterance to model and predict other agents beliefs and behaviors.

Future work will address the issues of scalability, relevance, an scope mentioned in the discussion section above. And we will conduct simple HRI evaluation experiments that will allow a human operator to command a mixed initiative team with one ground and one aerial robot as described in the case study, with and without belief modeling. This will allow us to determine whether and to what extent belief modeling as proposed in this paper can lead to objectively better task performance and subjectively better acceptance by human team mates.

ACKNOWLEDGEMENTS

This work was in part funded by ONR grants #N00014-07-1-1049 and #N00014-11-1-0493. We would also like to thank the referees for their userful comments which helped improve this paper.

REFERENCES

[1] Cynthia Breazeal, Andrew Brooks, Jesse Gray, Guy Hoffman, Cory Kidd, Hans Lee, Jeff Lieberman, Andrea Lockerd, and David Mulanda, 'Humanoid robots as cooperative partners for people.', *Journal of Humanoid Robots*, **1**, (2004).

[2] M Brenner, 'Situation-aware interpretation, planning and execution of user commands by autonomous robots', in *Proceedings of the 16th International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, (2007).

[3] M Brenner and B Nebel, 'Continual planning and acting in dynamic multiagent environments', *Autonomous Agents and Multi-Agent Systems*, **19**(3), 239–331, (2009).

[4] Gordon Briggs and Matthias Scheutz, 'Facilitating mental modeling in collaborative human-robot interaction through adverbial cues', in *Proceedings of the 12th Annual SIGdial Meeting on Discourse and Dialogue*. ACL, (2011).

[5] Rehj Cantrell, Matthias Scheutz, Paul Schermerhorn, and X Wu, 'Robust spoken instruction understanding for hri', in *Proceedings of the 6th International Conference on Human-Robot Interaction*. ACM/IEEE, (2010).

[6] P. R. Cohen and Hector Levesque, 'Teamwork', Technote 504, SRI International, Menlo Park, CA, (1991).

[7] P. R. Cohen and S. L. Oviatt, 'The role of voice input for human-machine communication', *Proceedings of the National Academy of Science, USA*, **92**, 9921–9927, (1995).

[8] Philip R. Cohen, Hector J. Levesque, Jose Nunes, and Sharon L. Oviatt, 'Task-oriented dialogue as a consequence of joint activity', in *Pacific Rim International Conference on Artificial Intelligence*, (1990).

[9] H. P. Grice, 'Logic and conversation', *Syntax and Semantics*, **3**(1), 43–58, (1975).

[10] Barbara J. Grosz and Sarit Kraus, 'Collaborative plans for complex group action', *ARTIFICIAL INTELLIGENCE*, **86**(2), 269–357, (1996).

[11] Sanjeev Kumar, Marcus J. Huber, David R. Mcgee, Philip R. Cohen, and Hector J. Levesque, 'Semantics of agent communication languages for group interaction', in *In Proceedings of the 17 th Int. Conf. on Artificial Intelligence*, pp. 42–47, (2000).

[12] Paul Schermerhorn and Matthias Scheutz, 'Using logic to handle conflicts between system, component, and infrastructure goals in complex robotic architectures', in *Proceedings of the International Conference on Robotics and Automation*, (2010).

[13] Matthias Scheutz, Paul Schermerhorn, J Kramer, and D Anderson, 'First steps toward natural human-like hri', *Autonomous Robots*, **22**(4), 411–423, (2007).

[14] Ira A. Smith and Philip R. Cohen. Toward a semantics for an agent communications language based on speech-acts, 1995.