The Role of Models and Communication in the Ad Hoc Multiagent Team Decision Problem

Trevor Sarratt TSARRATT@SOE.UCSC.EDU
Arnav Jhala JHALA@SOE.UCSC.EDU
Department of Computer Science, University of California Santa Cruz, Santa Cruz, CA 95060 USA

Abstract

Ad hoc teams are formed of members who have little or no information regarding one another. In order to achieve a shared goal, agents are tasked with learning the capabilities of their teammates such that they can coordinate effectively. Typically, the capabilities of the agent teammates encountered are constrained by the particular domain specifications. However, for wide application, it is desirable to develop systems that are able to coordinate with general ad hoc agents independent of the choice of domain. We propose examining ad hoc multiagent teamwork from a generalized perspective and discuss existing domains within the context of our framework. Furthermore, we consider how communication of agent intentions can provide a means of reducing teammate model uncertainty at key junctures, requiring an agent to consider its own information deficiencies in order to form communicative acts improving team coordination.

1. Introduction

Effective teamwork relies on the coordination of individual team members which, in turn, requires the team to have formed a consensus not only on the task at hand but also over the expectations of the individual team members. Without the ability to correctly predict the actions of a teammate, an agent may fail to perform a joint action or to adopt a complimentary strategy to its teammates'. As discussed in (Stone et al., 2010), humans are capable of working together without prior coordination or prior knowledge of one another, forming *ad hoc* teams. Similarly, ad hoc autonomous agent teams are composed of two or more agents designed separately and with minimal shared information. In these ad hoc settings, efficient teamwork depends on an agent's capability to accurately predict and adapt to its teammate's behavior.

One of the major aims of ad hoc teams research is to provide methods of accurately modeling a teammate for improved coordination in ad hoc team domains. Typically, predictions of teammate actions can be made from identifying a model of similar behavior from a set of previously learned models, learning a model through observing the teammate, or by combining both approaches (Barrett et al., 2012). Furthermore, these ideas can be extended by considering agents of non-static behavior (i.e. occasionally changing strategy) and illustrating the effectiveness of identifying changes between a small number of simple behaviors as an approximation to an unknown model (Sarratt & Jhala, 2015).

© 2015 Cognitive Systems Foundation. All rights reserved.

However, we believe that at the core of the ad hoc team problem is a perspective perhaps uniquely provided by that of cognitive systems. When an ad hoc teammate is assumed to be drawn from a class of agents adhering to some known decision mechanism, the assumption ascribes a set of corresponding present or absent capabilities to the agent, such as having and updating beliefs regarding an aspect of the world or being unable to recursively model other teammates. This narrowing of the space of potential teammates permits the demonstration of coordination strategies requiring a feasible amount of computation, but it does not address a larger question:

> *Can a cognitive system be designed to coordinate with an arbitrary ad hoc teammate?*

Though a more directed fashion, it may be better posed as

> *How can a cognitive system identify the capabilities of a teammate such that it can coordinate effectively?*

In short, rather than require a human to specify the constraints on considered models of teammate behavior, an ideal agent should possess the capacity to reason over the task definition and observed actions of other agents in order to construct and adapt models of its associates. Humans are able to adapt to the behaviors of other humans, domesticated animals (herding dogs, for example), virtual agents, and robots. This motivates the exploration of cognitive systems that are similarly flexible.

For the purpose of coordination, once a model has been established, it may be necessary for the team to exchange information, perhaps sharing observations about the world or negotiating a plan. In much of the ad hoc teamwork literature, communication is assumed to be infeasible. However, various exchanges of information are employed in (Barrett et al., 2014; Genter, Laue, & Stone, 2015), though as a domain-specific addition. Furthermore, emergent language between agents could allow for team discourse (Steels, 2003; Roy & Reiter, 2005). Allowing for such exchanges then raises the additional question:

> *When an agent possesses little or no information regarding an ad hoc teammate, what should it communicate?*

This inquiry is related to the existing multiagent communication work (Roth, Simmons, & Veloso, 2006), yet it extends the breadth of information available for exchange as well as relaxes the assumptions regarding the nature of the teammate, often assumed to be identical to the communicating agent. It is immediately obvious that modeling and communication are intertwined, as deficiencies in a model may be compensated for by eliciting information from a teammate, yet increasingly accurate predictions will reduce the need for communication.

In order to unify the consideration of models and communication, we propose a framework for discussing general ad hoc team domains. The remainder of this paper briefly surveys existing domains as well as common approaches to multiagent communication. Finally, as a demonstration of the open questions within ad hoc teamwork, we discuss the utilization of multiagent communication—commonly given only superficial treatment in ad hoc domains—and describe its potential merits as a mechanism for complementing established agent modeling approaches.

2. A Motivating Example

Consider a scenario where two rescue robots are tasked with exploring an area and reporting the locations of humans in need of help. These robots may have heterogeneous sensing, acting, and planning capabilities. They have been designed to work in a shared space with a common goal. With no other shared information, how can the agents coordinate effectively? Humans are able to form such ad hoc teams and cooperate, adapting to the behaviors of their teammates as necessary. For illustrative purposes, we will refer to the following scenario throughout this paper:

The two rescue robots are positioned in a room with two exit doors. One door is shut, and the other is open but blocked by a desk. In order to continue their exploration, the robots must proceed either together through one door or separately through both. Robot Tom is only capable of opening doors, while robot Mary can open doors as well as move large objects, though neither robot has knowledge of the other's ability. Once a doorway is unblocked, the robot who performed the action is given an observation about the state of the next room, which we will restrict to being either another room with doors or a closet (dead-end). If both agents perform the same action, they collide and fail the attempt.

An ideal handling of the scenario may consist of discussing the robot's respective capabilities, executing a plan where Tom opens the door while Mary moves the desk, sharing observations regarding the adjacent rooms, and finally deciding whether to move to one adjacent room together or, in the possibility of two adjacent rooms, transition separately.

3. Problem Description

The task of an ad hoc collaborative agent is to adjust its policy to pursue a goal shared with other ad hoc agents under some degree of uncertainty regarding the other agents' policies. This relates to multiagent decision problems such as Dec-POMDPs (Bernstein, Zilberstein, & Immerman, 2000), communicative multiagent team decision problems (Pynadath & Tambe, 2002), and interactive POMDPs (Doshi, Zeng, & Chen, 2009). In contrast to those models, we free ad hoc agents of any assumptions regarding the space of possible intentional models, reward functions, or beliefs of teammates. Instead, predictions of teammate behavior are made from models learned from previous experience with similar agents, current observations of a given teammate, and communicated information regarding the task. Agent models, in the most simple form, are represented as policies, mappings between states in the domain and appropriate actions to be taken when those states are encountered. While it is infeasible in reasonably complex domains to learn an associated action for every possible state, an agent can still coordinate effectively by observing or communicating about agent actions specific to sections of the state space that are likely to occur. The problem representation is left purposefully broad such that existing ad hoc team domains can be mapped to our description and adopt the results from our work.

3.1 Formalization

Given an agent, d, who must coordinate with one or more ad hoc agents, $i \in I$, where I is the set of ad hoc team agents from d's perspective, we represent an ad hoc multiagent team decision

problem (AMTDP) for agent d as a tuple, $\langle P_d, J, I, M, \Sigma \rangle$. We have relaxed the specification of a joint reward function for a broader set of shared goal states. Section 3.1.2 specifies the joint information, J, for all teammates, while Section 5 discusses the potential for communication, Σ, over the shared information. We demonstrate the generality of this framework by mapping one of the existing domains to the model in Section 4.

3.1.1 Problem Specification

P_d defines the problem from agent d's perspective. Minimally, this specification will be described by its own tuple, $\langle S, A_d, T, G \rangle$, where

- S is the set of world states,
- A_d is the set of domain-level actions available to agent d,
- T is the transition function for states when an action is taken, and
- G is the set of goal states.

Commonly, world states are expressed as a cross product of separate features, $S = \Xi_1 \times ... \times \Xi_m$. We use this representation of state when discussing state abstractions for high level policy discussion in Section 5.4.2.

As agents may have heterogeneous acting capabilities that are unknown to each other, the problem description only specifies the agent's own actions. Both actions and the corresponding transition probabilities, defined by T, may be unknown to other agents.

$G \subseteq S$ is the set of goal states shared among all agent teammates. While traditional decision-theoretic multiagent decision problems assume a shared utility function, R, we utilize the less specific assumption of shared goals from joint intention theory (Cohen & Levesque, 1991). Without a shared set of goal states, the problem ceases to be a team effort and is better represented as a partially observable stochastic game (Bernstein et al., 2002). Furthermore, using G allows ad hoc agents to pursue states according to decision processes other than those described by utility theory.

The problem specification is not limited to the features discussed here, but may extend generally to include other items. One common extension for such work is partial observability, when state information is not directly observable by one or more agents. Similarly, agents may not always be able to observe each other's actions. All agents involved would likely have sensors or other mechanisms for making observations about the state. As the agents may be heterogeneous, we may expect observation types and associated likelihoods unique to each teammate. However, in current ad hoc team domains, the method of acquiring information is often shared across teammates. In the RoboCup standard platform league, for example, all players use the same model of robot, allowing for the assumption of identical observation capabilities between agents. In order to handle the sources of uncertainty in partially observable domains, agents typically hold and update beliefs regarding the world state as well as the behavior of their teammates. Beliefs can likewise be added as a component of a teammate model. Modeling beliefs of another agent can lead to the nesting of beliefs dealt with by existing frameworks, DEC-POMDPs (Bernstein et al., 2002) and I-POMDPs (Doshi, Zeng, & Chen, 2009).

3.1.2 Joint Information

J refers to the set of shared information regarding the task. Minimally, this contains the state space definition and set of goal states present in the individual problem specifications. The state descriptions for the state space implies a set of common features for modeling and perceiving the world, though agents may extend state descriptions with additional features according to their own world view.

In many instances, as discussed in Section 4, J additionally includes teammate-specific information such as full knowledge of the action capabilities and corresponding transition functions, as well as common communicative abilities.

3.1.3 Ad Hoc Teammates

The set of ad hoc teammates ideally makes minimal assumptions regarding their behavior. However, in order to discuss coordination, a few conjectured properties are necessary. First, the space of team problems implies a set of one or more goal states common among the team's members. We exclude domains where the goal states are unreachable for all team joint policies, that is $G = \{g \,|\, g \in S \wedge \exists \pi_{team} \;\; Pr(g \,|\, s^0, \pi_{team}) > 0\}$. We note that the path to $g \in G$ need only exist from s^0, allowing for the goal to be unreachable from the state s^t, $t > 1$ if it is not in the path of states leading to g. It is up to the design of the domain whether coordination is necessary to achieve a goal or if a single agent's actions are sufficient.

Following the domain restrictions, we consider only teammates whose policies do not strictly prevent the accomplishment of a goal. Specifically, $I = \{i \,|\, \exists \pi_{team} \;\; \pi_i \in \pi_{team} \wedge Pr(g \in G \,|\, s^0, \pi_{team}) > 0\}$. Teammates are not assumed to have optimal behavior.

3.1.4 Models

$M = \{m_i\}$ is the set of models agent d uses to predict the actions of its teammates. An agent model serves an approximation to the agent's policy, as dictated by its personal decision procedure.

While it is desirable to avoid erroneously ascribing cognitive capabilities such as beliefs and intentions to teammates, we are motivated to consider notions of commitment, consistency, and internal state that arise in such work (Jennings, 1995; Jennings, 1993; Wooldridge & Jennings, 1996). An ad hoc agent's policy may be static, that is, the mapping of actions to world states does not change. This is the most simple case for coordinating, as every observation of the agent can be directly used in a model for future use. Over time, as the model incorporates observations from more states, it will be able to more accurately predict the actions of the teammate. The alternative is an agent whose policy changes according to time, observations of state information, updated models of its teammates, or some other change in its internal state. It is then useful to draw parallels between intention theory and non-static state-action mappings. When the internal state of an agent specifies an action for a given world state, it forms a sort of commitment, as it will attempt to perform the action in the specified state. Consequently, a change in the agent's internal state may cause a corresponding alteration of its policy. The agent reevaluates the selection of an action according to its internal state, exercising some unknown convention for reconsideration.

Consider the example from Section 2. Tom observes Mary open the closed door then move the desk from in front of the second door. Tom may attempt to comprehend the situation in a number of ways. Immediately, Tom recognizes that Mary can open doors and move desks. He updates his model m_{Mary} with her actions, $A_{Mary} = \{open\,door, move\,desk\}$. How does he use the information to adjust his expectation of Mary's future behavior as well as the state of the world? Perhaps Mary, upon opening the first door, saw that it was a closet, then proceeded with her only other option: unblocking the second door. If so, Tom should adjust his world state estimation with an increased likelihood of the first door leading to a closet. However, Tom can also consider that Mary has a model of Tom's own capabilities, recursively. Without any prior interaction, it may be that Mary does not believe Tom capable of unblocking either door. She may expect a higher utility from dividing the search space by splitting up. This proposes a competing model with the first. Additionally, it could be that Mary obeys a set of simple commands that dictate she unblock any passageway that she is able to, such that other search robots and an eventual team of human rescuers have an easier time navigating the space. In complex domains, the large space of possible teammate policies combined with the possibility of non-observable information results in a space of potential models that is intractable to reason over exhaustively using observation-based evidence. As a result, agents in ad hoc teams are limited to using observations to learn or fit behavior to an approximate model. In Section 5.4, we discuss the potential for exchange of model mechanics during communication and the necessary capabilities of an agent in order to do so.

3.1.5 Communication

Σ defines the explicit communicative capabilities of the agents. The capabilities need not be symmetric between agents. A review of existing multiagent communication approaches as well as a discussion of the unique challenges encountered in ad hoc teams is given in Section 5.

4. A Short Survey of Existing Ad Hoc Team Domains

In order to discuss the progress and future of work in ad hoc teams, it is useful to consider the domains and agent capabilities currently used. For this purpose, we outline many of the problem domains. For illustrative purposes, we discuss aspects of the first, the multi-armed bandit problem, in relation to the AHMTDP model. The remaining domains can be mapped similarly.

4.1 Multi-Armed Bandit

The multi-armed bandit problem is a well-studied example in sequential decision making. It has been extended to ad hoc team settings to examine how information can be conferred to a teammate via observed actions (Stone & Kraus, 2010) as well as explicit communication (Barrett et al., 2014).

The initial state, S^0 of the multi-armed bandit problem is drawn from a set of possible assignments of payout distributions to k arms. In the teaching version, two agents, a teacher and a learner, are able to pull one of the $k-1$ arms per round while the teacher has the additional option of selecting the remaining arm which has the highest payout, forming the set of actions, A. The arm chosen then outputs a payout according to its hidden distribution, which serves as an observation of

the distribution, and transitions the state to a new round. After a finite set of rounds, the end state is reached, and the goal, G, is specified as a utility function where end states are preferred according to total payout.

The team shares knowledge, J, regarding the basic premise of the game as well as the complete behavior of the learner, which acts greedily according to its imperfect knowledge of the payouts. I consists solely of the learning agent. The teacher has full knowledge of the arms' payout distributions. It models the learner perfectly and must decide whether it is better to pull the highest paying arm or "teach" the learner by pulling the second best arm to demonstrate its distribution to the greedy, learning agent.

In the communicative version (Barrett et al., 2014), extensions are made to add obscurity to the teammate's model, add explicit communication options, as well as restrict payout information from all agents. Specifically, teammates either behave according to UCT (Kocsis & Szepesvári, 2006) or are drawn from a set of greedy agents. The primary agent must use observations to decide which model $m \in M$ best supports the teammate's behavior. An agent has the option to broadcast its last observation, the mean of a given arm, as well as a suggestion for which arm its teammate should pull. Each communication option $\sigma \in \Sigma$ has an associated cost.

4.2 Robot Soccer

The Standard Platform League of the RoboCup Soccer division has recently featured a drop-in player competition (MacAlpine et al., 2014; Genter, Laue, & Stone, 2015). Teams of five members are drawn from the set of submissions. This domain is unique in that it features cooperative behavior with teammates and adversarial behavior against a separate team of ad hoc agents. All agents use identical robots, unifying the types of information received across the team. Field, player, and ball location information is detected through vision sensors, while sounds such as whistles and horns are used to indicate phases of the game. Therefore, the basic state description contains the locations of the ball and players as well as the phase and time of the game. The basic actions are in the form of the controlling the robot's limbs and head, though such basic movements are often used to form higher level behaviors for moving about the field, tracking the ball, and kicking. The shared aim of the team is to score more goals than the opposing team. Models of teammates were left to the creators of each agent.

In the most recent competition, players were allowed basic communication using a standard set of messages including *want to be goalkeeper, want to play defense, want to play the ball, and I am lost* (Genter, Laue, & Stone, 2015). Not all teams utilized this ability; some teams disregarded most or all of the information received from teammates.

4.3 Flocking

Flocking has been proposed as a domain for ad hoc teams, where an ad hoc agent must attempt to coerce a group of flocking agents to adjust their orientation (Genter, Agmon, & Stone, 2013). The collection of ad hoc and flocking agents are given initial positions and orientations. Vision is restricted to a cone in the direction of the agent's orientation. The flocking agents deterministically adjust their orientation according to a function over the visible agents. The behavior of the flocking

agents is known to the directing agents, allowing perfect modeling. The flocking agents do not utilize any information regarding the goal or the capabilities of the agents attempting to achieve the goal. No communication is used.

4.4 Pursuit

Variations of the pursuit domain (Stone & Veloso, 2000) have been explored in ad hoc settings (Barrett et al., 2012; Barrett, Stone, & Kraus, 2011; Barrett et al., 2013; Sarratt & Jhala, 2015) for the purpose of exploring how teammate models can be learned or estimated using previous experience. In essence, the goal of the ad hoc team is to trap a prey agent moving through a fully observable grid. The agents have basic movement capabilities but do not communicate. The problem description is represented as a partially observable Markov decision process, which maps directly to the state, actions, and transitions of our framework. Models are either learned from a series of games with other types of teammates (Barrett et al., 2012; Barrett et al., 2013) or provided as a set of basic behaviors (Sarratt & Jhala, 2015). An agent must then infer which model best fits the observed behavior of its current teammates and coordinate accordingly.

4.5 Matrix Games

In the initial description of ad hoc autonomous teams (Stone et al., 2010), it was suggested that an ad hoc agent should be able to adapt to an unknown teammate in an iterative normal form game. These matrix games provide full game knowledge to all teammates, who must model and adapt to each other's behavior using the joint history of observed actions. Commonly, some assumptions are made regarding the type of teammate, restricting the model used to the corresponding type, such as a Markovian teammate (Chakraborty & Stone, 2013). For a more broad consideration of teammate behavior, a set of multiagent learning algorithms has been evaluated within the context of games where agents may or may not share a preferred outcome (Albrecht & Ramamoorthy, 2012).

5. Multi-Agent Communication

The exchange of information is a key component in many multiagent systems. Due to the breadth of applications of such systems, we restrict ourselves to discussing types of communication particularly relevant to coordination, namely the sharing of world state information and the negotiation of plans. Finally, we discuss extending existing methods to incorporate the exchange of modeling information across teammates.

5.1 Shared Mental Models

During their description of teamwork, Cohen and Levesque (Cohen & Levesque, 1991) mentioned the concept of a shared mental state, "the glue that binds teammates together." While they did not elucidate further on this specific notion, we can surmise through their discussion of the perils of diverging beliefs that the conceptual idea referred to what is known in related team research as *shared mental models* (Rouse, Cannon-Bowers, & Salas, 1992; Orasanu, 1994). A shared mental model is loosely a collection of joint information regarding the world state, expected transitions,

potential tasks, knowledge among teammates, and the behavior or roles of the team members. The theory, stemming from work in psychology, suggests that team members who possess information regarding the task as well as each member's individual participation can better anticipate the needs and requisite actions of their collaborators. This shared model of the process permits the team to coordinate effectively, often with reduced communication.

In recent years, the concept of shared mental models has been applied to multiagent decision systems. Yen et al. (Yen et al., 2006; Yen et al., 2003) proposed CAST—Collaborative Agents for Simulating Teamwork—as a model for teamwork among distributed, heterogeneous agents. The CAST architecture makes decisions according to the interplay of the individual and shared mental models, proactively communicating when it the expected utility of sharing information exceeds that of not sharing. The explicit concept of a shared mental model was later formalized for use by agent systems (Jonker, Van Riemsdijk, & Vermeulen, 2011).

5.2 State Information

In a partially observable environment, sharing information allows agents to sync their internal beliefs. Commonly, agents assume perfect knowledge about the decision processes of their teammates with the sole exception being the current state of beliefs regarding the world state (Pynadath & Tambe, 2002; Roth, Simmons, & Veloso, 2006). As the set of possible observations (Ω) and their associated probabilities (O) is shared information ($\Omega, O \in J$), the agents can broadcast their individual observation histories, then perform belief revision identically to update their teammate models as well as the agent's current estimation of the world state. The agents may then proceed from a point of mutual beliefs.

In domains where the observation function of another agent is unknown or when another agent is assumed to have perfect knowledge regarding some of the state information, an agent may query the necessary information directly from other agents (Roth, Simmons, & Veloso, 2007). The identification of which state information is needed can be performed traversing a factored policy tree and determining variable assignments are missing. Teaching agents in the multi-armed bandit scenario (Stone & Kraus, 2010; Barrett et al., 2014) are posed with the alternate form, where the agent has perfect information and must reason about what information it must communicate to the learning agent. The method of reasoning over communicative acts is, as we described earlier, a function of its value, which in turn is a function of the timing and content. These two characteristics refer to the *when* and *what* questions posed by Maayan Roth et al. (Roth, Simmons, & Veloso, 2006; Roth, Simmons, & Veloso, 2005).

Barrett et al. (Barrett et al., 2014) has applied similar practices, allowing agents to communicate individual observations and aggregate average payouts. For general application in ad hoc domains, an agent must determine whether a teammate estimates the current status of a hidden aspect of the world state. If the teammate has some state estimation capacity, the agent may attempt to query the teammate's estimation at a given time. Alternatively, the agent may attempt to infer or request the precise mechanism of the teammate's estimation for the corresponding teammate model.

5.3 Communicating Behavior

A key element of planning in a multiagent domain is the accurate projection of the actions of other agents. While this can be learned over time (Barrett et al., 2013; Barrett et al., 2012), it is often more direct to communicate regarding the expected behavior of coordinating agents. Stone and Veloso (Stone & Veloso, 1999) discuss the communication of roles, setplays, and formations for coordination in robotic soccer. Tan (Tan, 1993) demonstrated the enhanced performance of agent teams when learned policies are shared. We return to the idea of sharing policies in the following section.

SharedPlans (Grosz & Kraus, 1996) provides the most well-specified theory of communication and negotiation of plans in a multiagent setting. SharedPlans are constructed as hierarchical abstractions of actions required to complete a task. High level actions are then expanded into lower level actions, which can themselves be abstract or domain-level actions. A task is assigned to an agent or group of agents if the team mutually believes the assigned party can complete the task. Through the elaboration of single and multiagent sub-tasks, a full plan is realized. Communication must support the beliefs regarding partial plans for the completion of subtasks. Furthermore, an agent must have the ability to assess whether a teammate, a group of teammates, or itself *can bring about* a result. Within ad hoc team domains, teammates may have some unknown representation for acting, preventing the discussion of high level abstractions such as hierarchical plans, although a sufficiently capable agent may have the capacity to map its own behavior to such abstract concepts.

5.4 The Interplay of Models and Communication

An agent in an ad hoc team is tasked with planning its own actions toward the pursuit of a shared goal. In domains requiring coordination, the estimation of the actions of a teammate is necessary for successful planning. Relying on observations necessitates an act be performed, perhaps irreversibly, before the information can be utilized to infer or construct an accurate model. Communication, however, can provide a method of identifying an ensuing action before it occurs. An agent may reason over unknown state-action associations of its teammate's policy in the context of current and expected future states, identify key pairs that would be beneficial to communicate, and then acquire the necessary information. Similar to existing work in communicating observation data (Roth, Simmons, & Veloso, 2006), in order to be effective, an agent must reason over when communication should occur and what information needs to be broadcast.

In many of the domains described previously, it is reasonable to assume restrictions on the types of behavior of an ad hoc teammate. However, for more general application, we motivate the exploration of the exchange of information pertinent to an agent's behavior. The broad idea is that agents may, to some degree, be capable of reducing the uncertainty of their current teammate models through communicating future intended actions, allowing for the adaptation of plans in advance, resulting in improved coordination.

5.4.1 Exchanging Policy Information

While each additional constraint regarding an ad hoc teammate restricts the class of possible agents a technique can be applied to, it is necessary for communicated future behaviors to add a few

assumptions to the AHMTDP model. First, we assume each agent has the ability to consider its own action given an arbitrary state defined in S, defined within the joint information shared across teammates. This is a reasonable assumption given that the agent was designed to operate in the space of world state defined by the problem. Second, we introduce the concept of hypothetical action, such that an agent can consider its resulting action in a given state without necessarily believing it currently exists in the specified state or that its resulting action is taken currently. Third, we extend Σ to include queries between agents regarding their actions when given a specified state. We suspect that such simple associations, the foundation for behavior in agents, would be an early achievement in an emerging language between teammates. As an example in natural language, queries could take the form "*What would you currently do if found in the state ... ?*". We narrow the space of consideration through the implication of *currently*, as the internal state of an agent may change before the hypothetical state is encountered, potentially affecting the agent's choice of action. Additionally, we note that other communicative acts may provide act identically, providing sufficient information to an agent that it can reduce the uncertainty of a teammate's action when faced with a specific world state. As such, our discussion of the utility of communicated behavior is not restricted to hypotheticals.

Consider again the scenario of Tom and Mary. Before Mary acts, Tom plans his own course of action by exploring his policy tree with the simple goal of unblocking a door and transitioning to new room. Without any knowledge of Mary's plan, Tom must weigh the likelihood of achieving the goal according to her possible actions. Recognizing that a conflict can occur if both robots attempt to open the closed door, Tom can query Mary about her behavior given their current state, then choose not to open the door if she indicates that she would make the attempt. Note that it is unnecessary for Tom to query Mary regarding other potential states. For example, in the case that Mary moves the desk while Tom opens the closed door, Tom is able to pursue his attempt at moving into a new room unhindered. If Tom associated the goal states with utility value and communication with a cost, as is common in multiagent domains, Tom would need to decide if querying Mary to avoid possible conflict was worth the cost of communication. Similarly, if multiple states in the policy tree required such communication, queries could be ordered by gains in expected utility as well as immediacy of resulting changes in policy. Selection of communication instances is similar to that in (Roth, Simmons, & Veloso, 2006); however, rather than consider which of a series of observations linear with respect to time, a forward-looking agent must consider an exponential number of potential states the team may encounter, due to the branching nature of the planning process.

The unbounded nature of the ad hoc teammate's decision process of course raises a critical question: For how long is observed or communicated information consistent with teammate's policy? The question of consistency of behavior has been partially addressed in (Sarratt & Jhala, 2015), where the teammate occasionally switches which goal it is pursuing, resulting in varied behavior in otherwise identical world states. The core difficulty results from an unobserved variable within the teammate's decision process. Without a corresponding facet of the "inconsistent" agent's model used by other agents in the team, previously observed actions may suggest an incorrect course of action after the internal state of the teammate has changed. Accounting for such behavior changes is necessary for effective coordination (Sarratt & Jhala, 2015). Returning to intention theory, it has been suggested that intentions and their pursuit should be stable to some degree (Jennings, 1995), as

an agent switching its course of action too frequently cannot make progress toward any of the corresponding goal states. It is reasonable, then, to suggest that recent observations and communicated information may be considered more reliable than older instances.

5.4.2 Discussing Collections of State-Action Pairs

A direct extension to communicating state-action pairs is to consider a communicative act that addresses more than one state, allowing for multiple points of uncertainty to be reduced simultaneously. The primary motivation for this consideration is the ability to acquire information regarding entire portions of a teammate's policy. A natural form would be the utilization of abstractions of state, where all of the included states share some subset of state features. This, of course, requires further extending the communicative options, Σ, as well as the assumed capabilities of the responding agent, as it must be able to consider abstract states[1]. Following the description of states as a product of state features, we can define an abstract state as one only partially specified, leaving certain features open to a range of potential values. As an example, if the world states in the rescue robot scenario are represented by a room number, a set of exits, and the status of those exits, a fully specified state would be $s = \langle 1, \{d_1, d_2\}, \{closed, blocked\}\rangle$ while an abstract state would be represented by $s = \langle \_, \{d_1, d_2\}, \{closed, \_\}\rangle$, referring to an arbitrary room with two doors, one of which is known to be closed.

Two possible action policies may result. In the most simple case, the queried agent would perform a single action in all possible instances of the abstract state, perhaps always opening closed doors. If so, the querying agent can expect the queried teammate to open doors any time it encounters one, which informs the querying agent's planning phase. However, the queried agent's decision process may result in various actions according to the unspecified state features. In this case, the agent may return a partial policy for its actions conditioned on the remaining state features. For example, the agent may open d_1 if d_2 is blocked or closed, otherwise it would proceed through the d_2 if it is unblocked and open. This requires both the support of conditioned responses within the agents' communicative capacities as well as the integration of such responses within the agents' teammate models.

The challenge from the querying agent's perspective is constructing the question. It is possible for the agent to specify none of the state features, in essence asking what the queried agent would do in every possible state. In order to dissuade the agent from forming such open queries, a consideration should be given for received message length, which is a common metric for issuing communication cost. One reasonable restriction would be for the agent to only consider states encountered in its own planning step, as they are the most immediately reachable states in the state space. Knowing what a teammate would do in a state outside the agent's space of consideration will not affect the querying agent's resulting plan. Additionally, it may further assess the worst case cost of responses by calculating the branching factor of all unassigned state features. The querying agent can then assess the potential gains from communicating against the expected cost.

1. If a teammate can only consider fully specified states, the ability to consider abstract states can be approximated by enumerating through all possible states defined by the abstract state and querying the teammate about each, though this approach is potentially costly.

5.4.3 Exposing Hidden Model Information

While an agent's state-action policy is relatively static, the proposed communicative abilities may be sufficient for coordination on a task, particularly if the task is likely to be accomplished before the policy changes. However, in instances where a policy is in flux, perhaps when an agent's internal state dictates it changes strategy according to a hidden condition, it is useful to consider exposing such mechanisms to the agent's teammates. This requires very broad assumptions regarding the nature of the agents as well as the language used to communicate. First, the agent divulging aspects of its decision process must be able to represent the features according to some model of computation. The language must then support communication of arbitrary features from the model, including new features not explicitly shared in the team's joint information. Finally, the receiving agent must be able to incorporate the information into its model of the agent and execute the described functionality. Under free communication, an agent could transmit the entirety of its decision mechanism, allowing its teammate to have perfect knowledge of its behavior. However, under the typical case, where communication costs time and/or energy, agents would need to reason over which aspects of their behavior are worth describing.

A subset of this mode of communication is the extension of the state specification to include new variables that are not observable to all agents. These extended features can then be used as conditions on partial policies communicated as described in the previous section. Revisiting the robot scenario, if Mary only opens the closed door if she perceives that Tom cannot, she may be able to relate this aspect of her decision process through communication. First, she introduces to J her belief state regarding Tom's capability toward opening the door, its possible values $\langle can, cannot \rangle$, and its current state in her model, $cannot$. Once the concept is shared between the agents, she can return a conditioned policy, informing Tom that if her model does not support the possibility that Tom can open the door, she will open it. Furthermore, there are three possible considerations for how Tom can adjust his model once Mary believes Tom is able to open doors. When introducing the concept to the joint information, Mary could have specified the observations necessary for her model to update with the gained information regarding Tom. Alternatively, Tom could query Mary for the current value of her belief. Ultimately, Tom could use future observations of Mary's behavior to infer changes in the condition given the partial policy she specified.

6. Discussion

In this paper, we have provided an overview of the problem of ad hoc teamwork. Thus far, existing work has demonstrated the ability for agents to coordinate within ad hoc teams, though primarily through constraining the types of agents or methods of communication in advance. A truly effective system would be capable of narrowing the space of expected behaviors or constructing an explanatory model for an observed behavior by utilizing the information available to it. Furthermore, it should be capable of reasoning over the information deficiencies of its current knowledge, seeking required information from its teammates when necessary.

The development of a system capable of working in general domains and with arbitrary teammates may be an incremental task, requiring the progressive casting of wider nets, so to speak. It may require other technologies, such as emergent language (Roy & Reiter, 2005), to progress be-

yond their current states. Yet despite these limitations, we find the general case worth discussing. We provide a basic framework such that the various instances of the work can have a unified language for proposing advances, particularly in the hopes that the underlying ideas can be extracted and applied broadly. As a further point of interest, it may be that the abstract case motivates work with universal guarantees in successfully coordinating.

Our initial step in addressing the broad modeling case is the proposal of identifying key areas of uncertainty within an agent's models of its teammates. These points of interest provide incentive to initiate an exchange of information, as obtaining pieces of the team's individual plans allows for adaptation by the uncertain agent, reducing conflicting actions. This proposal is not without its limitations, however, as we merely trade restrictions on a teammate's decision process for assumptions regarding basic communicative capability. Nonetheless, we find the assumptions reasonable, particularly for application in domains where language is established, such as in human-computer or human-robot interaction. The potential to construct a reasonably approximate model of an ad hoc teammate through discourse is worth pursuing, particularly if it can address a significant portion of both theoretical and applied domains.

References

Albrecht, S. V., & Ramamoorthy, S. (2012). Comparative evaluation of mal algorithms in a diverse set of ad hoc team problems. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems* (pp. 349–356).

Barrett, S., Agmon, N., Hazon, N., Kraus, S., & Stone, P. (2014). Communicating with unknown teammates. *Proceedings of the 2014 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1433–1434).

Barrett, S., Stone, P., & Kraus, S. (2011). Empirical evaluation of ad hoc teamwork in the pursuit domain. *The 10th International Conference on Autonomous Agents and Multiagent Systems* (pp. 567–574).

Barrett, S., Stone, P., Kraus, S., & Rosenfeld, A. (2012). Learning teammate models for ad hoc teamwork. *AAMAS Adaptive Learning Agents (ALA) Workshop*.

Barrett, S., Stone, P., Kraus, S., & Rosenfeld, A. (2013). Teamwork with limited knowledge of teammates. *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI-13)*.

Bernstein, D. S., Givan, R., Immerman, N., & Zilberstein, S. (2002). The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, *27*, 819–840.

Bernstein, D. S., Zilberstein, S., & Immerman, N. (2000). The complexity of decentralized control of markov decision processes. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence* (pp. 32–37).

Chakraborty, D., & Stone, P. (2013). Cooperating with a markovian ad hoc teammate. *Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1085–1092).

Cohen, P. R., & Levesque, H. J. (1991). Teamwork. *Nous*, 487–512.

Doshi, P., Zeng, Y., & Chen, Q. (2009). Graphical models for interactive pomdps: representations and solutions. *Autonomous Agents and Multi-Agent Systems*, *18*, 376–416.

Genter, K., Agmon, N., & Stone, P. (2013). Ad hoc teamwork for leading a flock. *Proceedings of the 2013 International Conference on Autonomous Agents and Multiagent Systems* (pp. 531–538).

Genter, K., Laue, T., & Stone, P. (2015). The robocup 2014 spl drop-in player competition: Experiments in teamwork without pre-coordination.

Grosz, B. J., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence*, *86*, 269–357.

Jennings, N. R. (1993). Commitments and conventions: The foundation of coordination in multi-agent systems. *The Knowledge Engineering Review*, *8*, 223–250.

Jennings, N. R. (1995). Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artificial Intelligence*, *75*, 195–240.

Jonker, C. M., Van Riemsdijk, M. B., & Vermeulen, B. (2011). Shared mental models. In *Coordination, organizations, institutions, and norms in agent systems vi*, 132–151. Springer.

Kocsis, L., & Szepesvári, C. (2006). Bandit based monte-carlo planning. In *Machine learning: Ecml 2006*, 282–293. Springer.

MacAlpine, P., Genter, K., Barrett, S., & Stone, P. (2014). The robocup 2013 drop-in player challenges: Experiments in ad hoc teamwork. *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on* (pp. 382–387).

Orasanu, J. (1994). Shared problem models and flight crew performance. *Aviation psychology in practice*, 255–285.

Pynadath, D. V., & Tambe, M. (2002). The communicative multiagent team decision problem: Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, 389–423.

Roth, M., Simmons, R., & Veloso, M. (2005). Reasoning about joint beliefs for execution-time communication decisions. *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 786–793).

Roth, M., Simmons, R., & Veloso, M. (2006). What to communicate? execution-time decision in multi-agent pomdps. In *Distributed autonomous robotic systems*, Vol. 7, 177–186. Springer.

Roth, M., Simmons, R., & Veloso, M. (2007). Exploiting factored representations for decentralized execution in multiagent teams. *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems* (p. 72).

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *Systems, Man and Cybernetics, IEEE Transactions on*, *22*, 1296–1308.

Roy, D., & Reiter, E. (2005). Connecting language to the world. *Artificial Intelligence*, *167*, 1–12.

Sarratt, T., & Jhala, A. (2015). Rapid: A belief convergence strategy for collaborating with inconsistent agents. *AAAI Workshop on Multiagent Interaction without Prior Coordination*.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in cognitive sciences*, *7*, 308–312.

Stone, P., Kaminka, G. A., Kraus, S., Rosenschein, J. S., et al. (2010). Ad hoc autonomous agent teams: Collaboration without pre-coordination. *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI-10)*.

Stone, P., & Kraus, S. (2010). To teach or not to teach?: decision making under uncertainty in ad hoc teams. *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1* (pp. 117–124).

Stone, P., & Veloso, M. (1999). Task decomposition, dynamic role assignment, and low-bandwidth communication for real-time strategic teamwork. *Artificial Intelligence*, *110*, 241–273.

Stone, P., & Veloso, M. (2000). Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, *8*, 345–383.

Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. *Proceedings of the 10th International Conference on Machine Learning* (pp. 330–337).

Wooldridge, M., & Jennings, N. R. (1996). Towards a theory of cooperative problem solving. In *Distributed software agents and applications*, 40–53. Springer.

Yen, J., Fan, X., Sun, S., Hanratty, T., & Dumer, J. (2006). Agents with shared mental models for enhancing team decision makings. *Decision Support Systems*, *41*, 634–653.

Yen, J., Fan, X., Sun, S., Wang, R., Chen, C., Kamali, K., & Volz, R. A. (2003). Implementing shared mental models for collaborative teamwork. *The Workshop on Collaboration Agents: Autonomous Agents for Collaborative Environments in the IEEE/WIC Intelligent Agent Technology Conference, Halifax, Canada*.