Team Coordination among Distributed Agents: Analyzing Key Teamwork Theories and Models

David V. Pynadath and Milind Tambe
Computer Science Department and Information Sciences Institute
University of Southern California
4676 Admiralty Way, Marina del Rey, CA 90292
Email: pynadath@isi.edu, tambe@usc.edu

Abstract

Multiagent research has made significant progress in constructing teams of distributed entities (e.g., robots, agents, embedded systems) that act autonomously in the pursuit of common goals. There now exist a variety of prescriptive theories, as well as implemented systems, that can specify good team behavior in different domains. However, each of these theories and systems addresses different aspects of the teamwork problem, and each does so in a different language. In this work, we seek to provide a unified framework that can capture all of the common aspects of the teamwork problem (e.g., heterogeneous, distributed entities, uncertain and dynamic environment), while still supporting analyses of both the optimality of team performance and the computational complexity of the agents' decision problem. Our COMmunicative Multiagent Team Decision Problem (COM-MTDP) model provides such a framework for specifying and analyzing distributed teamwork. The COM-MTDP model is general enough to capture many existing models of multiagent systems, and we use this model to provide some comparative results of these theories. We also provide a breakdown of the computational complexity of constructing optimal teams under various classes of problem domains. We then use the COM-MTDP model to compare (both analytically and empirically) two specific coordination theories (joint intentions theory and STEAM) against optimal coordination, in terms of both performance and computational complexity.

Introduction

A central challenge in the control and coordination of distributed agents, robots, and embedded systems is enabling these different, distributed entities to work together as a team, toward a common goal. Such teamwork is critical in a vast range of domains, e.g., for future teams of orbiting spacecraft, sensor teams for tracking targets, teams of unmanned vehicles for urban battlefields, software agent teams for assisting organizations in rapid crisis response, etc. Research in teamwork theory has built the foundations for successful practical agent team implementations in such domains. On the forefront so far have been theories based on belief-desire-intentions (BDI) frameworks (e.g., joint-intentions (Cohen & Levesque 1991), SharedPlans (Grosz & Kraus 1996), and others (Sonenberg *et al.* 1994)) that have provided prescriptions for coordination in practical systems. These theories have inspired the construction of practical, domain-independent teamwork models (Jennings 1995; Rich & Sidner 1997; Tambe 1997; Yen *et al.* 2001), successfully applied in a range of complex domains. While the BDI-based theories continue to be useful in practical systems, there is now a critical need for complementary foundational frameworks as practical teams scale-up towards complex, real-world domains. Such complementary frameworks would address some key weaknesses in the existing theories, as outlined below:

- Existing teamwork theories provide prescriptions that usually ignore the various uncertainties and costs in real-world environments. While such frameworks can still provide satisfactory behavior, it is difficult to evaluate the degree of *optimality* of the behavior suggested by these theories/models. For instance, the joint intentions theory (Cohen & Levesque 1991) suggests that team members commit to attainment of mutual beliefs in key circumstances, but it ignores the cost of attaining mutual belief (e.g., via communication). Coordination strategies that blindly follow such prescriptions may lead to highly suboptimal implementations. While in some cases optimality may be practically unattainable, understanding the strengths and limitations of coordination prescriptions is definitely useful.

  Of course, while existing theories have avoided costs and uncertainties, practical systems cannot do so in real-world environments. For instance, STEAM (Tambe 1997) extends the BDI-logic approach with decision-theoretic extensions. Unfortunately, while such approaches succeed in their intended domains, their very pragmatism often necessarily leads to a lack of the theoretical rigor. Furthermore, it remains unclear whether their suggested prescriptions (e.g., STEAM's prescriptions) necessarily lead to optimal team coordination.
- Existing teamwork theories have so far failed to provide insights into the computational complexity of various aspects of teamwork decisions. Such results may show intractability of some coordination strategies in specific domains and potentially justify the use of practical teamwork models as an approximation to optimality in such domains.

To answer these needs, we propose a new complementary framework, the *COMmunicative Multiagent Team Decision Problem (COM-MTDP)*, inspired by earlier work in *economic team theory* (Marschak & Radner 1971; Ho 1980), within the context of economics and operations research. While our COM-MTDP model borrows from a theory developed in another field, we are not just dressing up old wine in a new bottle. We make several contributions in applying and extending the original theory for applicability to intelligent distributed systems. First, we extend the theory to include key issues of interest in the distributed AI literature, most notably by explicitly including communication. Second, we introduce a categorization of teamwork problems based on these extensions. Third, we provide complexity analyses of these problems, providing a clearer understanding of the difficulty of teamwork problems under different circumstances. For instance, some researchers have advocated teamwork without communication (Matarić 1997). This paper clearly identifies domains where such team planning remains tractable, at least when compared to teamwork with communication. It also identifies domains where communication provides a significant advantage in teamwork by reducing the time complexity.

In addition to analyzing complexity, our framework supports the analysis of the *optimality* of different means of coordination. We isolate the conditions under which coordination based on joint intentions (Cohen & Levesque 1991) is preferred, based on the re-

Copyright © 2002, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

sulting performance of the distributed team at the desired task. We provide similar results for coordination based on the STEAM algorithm (Tambe 1997). We also analyze the complexity of the decision problem within these frameworks. The end result is the beginnings of a *well-grounded characterization of the complexity-optimality trade-off* among various means of team coordination.

Finally, we are currently conducting experiments using the domain of simulated helicopter teams, conducting a joint mission. We are evaluating different coordination strategies, and comparing them to the optimal strategy that we can compute using our framework. We will use the helicopter team as a running example in our explanations in the following sections.

Multiagent Team Decision Problems

Given a team of selfless, distributed entities, α, who intend to perform some joint task, we wish to evaluate the effectiveness of possible coordination policies. A coordination policy prescribes certain behavior, on top of the domain-level actions the agents perform in direct service of the joint task. We begin with the initial team decision model (Ho 1980). This initial model focused on a static decision, but we provide an extension to handle dynamic decisions over time. We also provide other extensions to the individual components to generalize the representation in ways suitable for representing multiagent domains. We represent our resulting *multiagent team decision problem* (MTDP) model as a tuple, $\langle S, \mathbf{A}, P, \mathbf{\Omega}, \mathbf{O}, R\rangle$. The remainder of this section describes each of these components in more detail.

World States: S

- S: a set of world states (the team's environment).

Domain-Level Actions: $\mathbf{A}$

- $\{A_i\}_{i\in\alpha}$: set of actions available to each agent $i \in \alpha$, implicitly defining a set of combined actions, $\mathbf{A} \equiv \prod_{i\in\alpha} A_i$.

These actions consist of tasks an agent may perform to change the environment (e.g., change helicopter altitude/heading). These actions correspond to the decision variables in team theory.

Extension to Dynamic Problem The original conception of the team decision problem focused on a one-shot, static problem. The typical multiagent domains require multiple decisions made over time. We extend the original specification so that each component is a time series of random variables, rather than a single one. The interaction between the distributed entities and their environment obeys the following probabilistic distribution:

- P: state-transition function, $S \times \mathbf{A} \to \Pi(S)$.

This function represents the effects of domain-level actions (e.g., a flying action changes a helicopter's position). The given definition of P assumes that the world dynamics obey the Markov assumption.

Agent Observations: $\mathbf{\Omega}$

- $\{\Omega_i\}_{i\in\alpha}$: a set of observations that each agent, i, can experience of its world, implicitly defining a combined observation, $\mathbf{\Omega} \equiv \prod_{i\in\alpha} \Omega_i$.

In a completely observable case, this set may correspond exactly to $S \times \mathbf{A}$, meaning that the observations are drawn directly from the actual state and combined actions. In the more general case, Ω_i may include elements corresponding to indirect evidence of the state (e.g., sensor readings) and actions of other agents (e.g., movement of other helicopters). In a team decision problem, the *information structure* represents the observation process of the agents. In the original conception, the information structure was a set of deterministic functions, $O_i : S \to \Omega_i$.

Extension of Allowable Information Structures Such deterministic observations are rarely present in nontrivial distributed domains, so we extend the information structure representation to allow for uncertain observations. There are many choices of models possible for representing the observability of the domain. In this work, we use a general stochastic model, borrowed from the *partially observable Markov decision process* model (Smallwood & Sondik 1973):

- $\{O_i\}_{i\in\alpha}$: an observation function, $O_i : S \times \mathbf{A} \to \Pi(\Omega_i)$, that gives a distribution over possible observations that an entity can make. This function models the sensors, representing any errors, noise, etc. Each O_i is cross product of observation functions of the state and actions of other entities: $O_{iS} \times O_{iA}$, respectively. We can define a combined observation function, $\mathbf{O} : S \times \mathbf{A} \to \Pi(\mathbf{\Omega})$, by taking the cross product over the individual observation functions: $\mathbf{O} \equiv \prod_{i\in\alpha} O_i$. Thus, the probability distribution specified by $\mathbf{O}$ forms the richer *information structure* used in our model.

We can make some useful distinctions between different classes of information structures:

Collective Unobservability This is the general case, where we make no assumptions on the observations.

Collective Observability $\forall \boldsymbol{\omega} \in \mathbf{\Omega}$, $\exists s \in S$ such that $\Pr(S^t = s|\mathbf{\Omega}^t = \boldsymbol{\omega}) = 1$. In other words, the state of the world is uniquely determined by the *combined* observations of the team of agents. The set of domains that are collectively observable is a strict subset of the domains that are collectively unobservable.

Individual Observability $\forall \omega_i \in \Omega_i$, $\exists s \in S$ such that $\Pr(S^t = s|\Omega_i^t = \omega_i) = 1$. In other words, the state of the world is uniquely determined by the observations of each individual agent. The set of domains that are individually observable is a strict subset of the domains that are collectively observable.

Policy (Strategy) Space: $\boldsymbol{\pi}$

- π_{iA}: a domain-level *policy* (or *strategy*, in the original team theory specification) to govern each agent's behavior.

Each such policy maps an agent's belief state to an action, where, in the original formalism, the agent's beliefs correspond directly to its observations (i.e., $\pi_{iA} : O_i \to A$).

Extension to Richer Strategy Space We generalize the set of possible strategies to capture the more complex mental states of the agents. Each agent, $i \in \alpha$, forms a belief state, $b_i^t \in B_i$, based on its observations seen through time t, where B_i circumscribes the set of possible belief states for the agent. We define the set of possible combined belief states over all agents to be $\mathbf{B} \equiv \prod_{i\in\alpha} B_i$. The corresponding random variable, $\mathbf{b}^t$, represents the agents' combined belief state at time t. Thus, we define the set of possible domain-level policies as mappings from belief states to actions, $\pi_{iA} : B_i \to A$. We elaborate on different types of belief states and the mapping of observations to belief states (i.e., the *state estimator function*) in a later section.

Reward Function: R

- R: reward function (or payoff function, or loss criterion), $S \times \mathbf{A} \to \mathbb{R}$.

This function represents the team's joint preferences over domain-level states (e.g., destroying enemy is good, returning to home base with only 10% of original force is bad). The function also represents the cost of domain-level actions (e.g., flying consumes fuel). Thus, all the agents share the same common interest. Following the original team-theoretic specification, we assume a Markovian property to R.

Extension for Explicit Representation of Communication: Σ

We make an explicit separation between the domain-level actions (A in the original team-theoretic model) and communication actions. In principle, A can include communication actions, but the separation is useful for the purposes of this framework. Thus, we extend our initial MTDP model to be a *communicative multiagent team decision problem* (COM-MTDP), that we define as a tuple, $\langle S, \mathbf{A}, \mathbf{\Sigma}, P, \mathbf{\Omega}, \mathbf{O}, R\rangle$, with the additional component, $\mathbf{\Sigma}$, and an extended reward function, R, as follows:

- $\{\Sigma_i\}_{i\in\alpha}$: a set of possible "speech acts", implicitly defining a set of combined communications, $\mathbf{\Sigma} \equiv \prod_{i\in\alpha}\Sigma_i$. In other words, an agent, i, may communicate an element, $x \in \Sigma_i$, to its teammates, who interpret the communication as they wish (e.g., broadcast arrival in enemy territory). We assume here perfect communication, in that each agent immediately observes the speech acts of the others without any noise or delay.
- R: reward function, $S \times \mathbf{A} \times \mathbf{\Sigma} \to \mathbb{R}$. This function represents the team's joint preferences over domain-level states and domain-level actions, as in the original team-theoretic model. However, we now extend it to also represent the cost of communicative acts (e.g., communication channels may have associated cost).

  We assume that the cost of communication and cost of domain-level actions are independent of each other, given the current state of the world. With this assumption, we can decompose the reward function into two components: a domain-level reward function, $R_A : S \times \mathbf{A} \to \mathbb{R}$, and a communication-level reward function, $R_\Sigma : S \times \mathbf{\Sigma} \to \mathbb{R}$. The total reward is the sum of the two component values: $R(s, \mathbf{a}, \boldsymbol{\sigma}) = R_A(s, \mathbf{a}) + R_\Sigma(s, \boldsymbol{\sigma})$. We assume that communication has no inherent benefit and may instead have some cost, so that for all states, $s \in S$, and messages, $\boldsymbol{\sigma} \in \mathbf{\Sigma}$, the reward is never positive: $R(s, \boldsymbol{\sigma}) \leq 0$.

With the introduction of this communication stage, the agents now update their belief states at two distinct points within each decision epoch t: once upon receiving observation Ω_i^t (producing the *pre-communication* belief state $b_{i\bullet\Sigma}^t$), and again upon observing the other agents' communications (producing the *post-communication* belief state $b_{i\Sigma\bullet}^t$). The distinction allows us to differentiate between the belief state used by the agents in selecting their communication actions and the more "up-to-date" belief state used in selecting their domain-level actions. We also distinguish between the separate *state-estimator* functions used in each update phase:

$$b_i^0 = SE_i^0()\ \ b_{i\bullet\Sigma}^t = SE_{i\bullet\Sigma}(b_{i\Sigma\bullet}^{t-1}, \Omega_i^t)\ \ b_{i\Sigma\bullet}^t = SE_{i\Sigma\bullet}(b_{i\bullet\Sigma}^t, \mathbf{\Sigma}^t)$$

where $SE_{i\bullet\Sigma} : B_i \times \Omega_i \to B_i$ is the pre-communication state estimator for agent i, and $SE_{i\Sigma\bullet} : B_i \times \mathbf{\Sigma} \to B_i$ is the post-communication state estimator for agent i. They specify functions that update the agent's beliefs based on the latest observation and communication, respectively. The initial state estimator, $SE_i^0 : \emptyset \to B_i$, specifies the agent's prior beliefs, before any observations are made. For each of these, we also make the obvious definitions for the corresponding estimators for the combined belief states: $\mathbf{SE}_{\bullet\Sigma}$, $\mathbf{SE}_{\Sigma\bullet}$, and $\mathbf{SE}^0$. Note that, although we assume perfect communication here, we could potentially use the post-communication state estimator to model any noise in the communication channel.

We extend our definition of agent's policy to include a *coordination policy*, $\pi_{i\Sigma} : B_i \to \Sigma_i$, analogous to the domain-level policy. We define the joint policies, $\boldsymbol{\pi}_\Sigma$ and $\boldsymbol{\pi}_A$, as the combined policies across all agents in α. We also define the overall policy, π_i, as the pair, $\langle \pi_{iA}, \pi_{i\Sigma}\rangle$, and the overall combined policy, $\boldsymbol{\pi}$, as the pair, $\langle \boldsymbol{\pi}_A, \boldsymbol{\pi}_\Sigma\rangle$.

Communication Subclasses

One key assumption we make is that communication has no latency, i.e., agents communication instantly reaches other agents without delay. However, as with the observability function, we parameterize the communication costs associated with message transmissions:

General Communication In this general case, we make no assumptions on the communication.

Free Communication $R(s, \mathbf{a}, \boldsymbol{\sigma}_1) = R(s, \mathbf{a}, \boldsymbol{\sigma}_2)$ for any $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathbf{\Sigma}$, $s \in S$, and $\mathbf{a} \in \mathbf{A}$. In other words, communication actions have no effect on the agents' reward.

No communication $\mathbf{\Sigma} = \emptyset$, so the agents have no *explicit* communication capabilities. Alternatively, communication may be prohibitively expensive, so that $\forall\sigma \in \mathbf{\Sigma}$, $s \in S$, and $\mathbf{a} \in \mathbf{A}$, $R(s, \mathbf{a}, \boldsymbol{\sigma}) = -\infty$. If communication has such an unbounded negative cost, then we can effectively treat the agents as having no communication capabilities when determining optimal policies.

The *free-communication* case is common in the multiagent literature, when researchers wish to focus on issues other than communication cost. We also identify the *no-communication* case because such decision problems have been of interest to researchers as well (Matarić 1997). Of course, even if $\mathbf{\Sigma} = \emptyset$, it is possible that there are domain-level actions in $\mathbf{A}$ that have communicative value. Such *implicitly* communicative actions may act as signals that convey information to the other agents. However, we still label such agent teams as having *no communication* for the purposes of the work here, where we gain extensive leverage from *explicit* models of communication when available.

Model Illustration

We can view the evolving state as a Markov chain with separate stages for domain-level and coordination-level actions. In other words, the agent team begins in some initial state, S^0, with separate features, $\xi_1^0, \ldots, \xi_m^0$. We also add initial belief states for each agent $i \in \alpha$, $b_i^0 = SE_i^0()$. Each agent, $i \in \alpha$, receives an observation Ω_i^0 drawn from Ω_i according to the probability distribution $O_i(S^0, \cdot)$, since there are no actions yet. Then, each agent updates its belief state, $b_{i\bullet\Sigma}^0 = SE_{i\bullet\Sigma}(b_i^0, \Omega_i^0)$.

Next, each agent $i \in \alpha$ selects a speech act according to its coordination policy, $\Sigma_i^0 = \pi_{i\Sigma}(b_{i\bullet\Sigma}^0)$, defining a combined speech act Σ^0. Each agent observes the communications of all of the others, and it updates its belief state, $b_{i\Sigma\bullet}^0 = SE_{i\Sigma\bullet}(b_{i\bullet\Sigma}^0, \mathbf{\Sigma}^0)$. Each then selects an action according to its domain-level policy, $A_i^0 = \pi_{iA}(b_{i\Sigma\bullet}^0)$, defining a combined action $\mathbf{A}^0$. In executing these actions, the team receives a single reward, $R^0 = R(S^0, \mathbf{A}^0, \mathbf{\Sigma}^0)$. The world then moves into a new state, S^1, according to the distribution, $P(S^0, \mathbf{A}^0)$, and the process continues.

Now, each agent i receives an observation Ω_i^1 drawn from Ω_i according to the distribution $O_i(S^1, \mathbf{A}^0)$, and it updates its belief state, $b_{i\bullet\Sigma}^1 = SE_{i\bullet\Sigma}(b_{i\Sigma\bullet}^1, \Omega_i^1)$. Each agent again chooses its speech act, $\Sigma_i^1 = \pi_{i\Sigma}(b_{i\bullet\Sigma}^1)$. The agents then incorporate the received messages into their new belief state, $b_{i\Sigma\bullet}^1 = SE_{i\Sigma\bullet}(b_{i\bullet\Sigma}^1, \mathbf{\Sigma}^1)$. The agents then choose new domain-level actions, $A_i^1 = \pi_{iA}(b_{i\Sigma\bullet}^1)$, resulting in yet another state, etc.

Thus, in addition to the time series of world states, $S^0, S^1, \ldots, S^t$, the agents themselves determine a time series of coordination-level and domain-level actions, $\mathbf{\Sigma}^0, \mathbf{\Sigma}^1, \ldots, \mathbf{\Sigma}^t$ and $\mathbf{A}^1, \mathbf{A}^1, \ldots, \mathbf{A}^t$, respectively. We also have a time series of observations for each agent i, $\Omega_i^0, \Omega_i^1, \ldots, \Omega_i^t$. Likewise, we can treat the combined observations, $\mathbf{\Omega}^0, \mathbf{\Omega}^1, \ldots, \mathbf{\Omega}^t$, as a similar time series of random variables.

Finally, the agents receive a series of rewards, $R^0, R^1, \ldots, R^t$. We can define the *value*, V, of the policies, $\boldsymbol{\pi}_A$ and $\boldsymbol{\pi}_\Sigma$, as the expected reward received when executing those policies. Over a finite horizon, T, this value is equivalent to the following:

$$V^T(\boldsymbol{\pi}_A, \boldsymbol{\pi}_\Sigma) = E\left[\sum_{t=0}^{T} R^t \middle| \boldsymbol{\pi}_A, \boldsymbol{\pi}_\Sigma\right] \tag{1}$$

	Individually Observable	Collectively Observable	Collectively Unobservable
No Comm.	P-complete	NEXP-complete	NEXP-complete
Gen. Comm.	P-complete	NEXP-complete	NEXP-complete
Free Comm.	P-complete	P-complete	PSPACE-complete

Table 1: Time complexity of COM-MTDPs under combinations of observability and communicability.

Complexity Analysis of Team Decision Problems

The problem facing these agents (or the designer of these agents) is how to construct the joint policies, $\boldsymbol{\pi}_\Sigma$ and $\boldsymbol{\pi}_A$, so as to maximize their joint utility, $R(S^t, \mathbf{A}^t, \Sigma^t)$.

Theorem 1 *The decision problem of determining whether there exist joint policies, $\boldsymbol{\pi}_\Sigma$ and $\boldsymbol{\pi}_A$, for a given team decision problem that yield a total reward at least K over some finite horizon T (given integers K and T) is NEXP-complete if $|\alpha| \geq 2$ (i.e., more than one agent).*

The proof for this theorem follows from some recent results in decentralized POMDPs (Bernstein, Zilberstein, & Immerman 2000). Due to lack of space, we will not provide the proofs for the theorems outlined in this paper.

The distributed nature of the agent team is one factor behind the complexity of the problem. The ability of the agents to communicate potentially allows sharing of observations and, thus, simplification of the agents' problem.

Theorem 2 *The decision problem of determining whether there exist joint policies, $\boldsymbol{\pi}_\Sigma$ and $\boldsymbol{\pi}_A$, for a given team decision problem with* free communication, *that yield a total reward at least K over some finite horizon T (given integers K and T) is PSPACE-complete.*

Theorem 3 *The decision problem of determining whether there exist joint policies, $\boldsymbol{\pi}_\Sigma$ and $\boldsymbol{\pi}_A$, for a given team decision problem with* free communication *and* collective observability, *that yield a total reward at least K over some finite horizon T (given integers K and T) is P-complete.*

Table 1 summarizes the above results. The columns outline different observability conditions, while the rows outline the different communication conditions. We can make the following observations:

- We can identify that difficult problems in teamwork — in the sense of providing agent teams with domain-level and coordination policies — arise under the conditions of collective observability or unobservability, and when there are costs associated with communication. In attacking teamwork problems, our energies should concentrate in these arenas.
- When the world is individually observable, communication makes little difference in performance.
- The collective observability case clearly illustrates a case where teamwork without communication is highly intractable, potentially with a double exponential. However, with communication, the complexity drops down to a polynomial running time.
- There is a reduction in complexity with the collectively unobservable case as well, but the difference is not as significant as with the collectively unobservable case.

Evaluating Coordination Strategies

While providing domain-level and coordination policies for teams in general is seen to be a difficult challenge, many systems attempt to alleviate this difficulty by having human users provide agents with domain-level plans (Tambe 1997). The problem for the agents then is to generate appropriate team coordination actions, so as to ensure that the domain-level plans are suitably executed. In other words, agents are provided with $\boldsymbol{\pi}_A$, but they must compute $\boldsymbol{\pi}_\Sigma$ autonomously. Different team coordination strategies have been proposed in the literature, essentially to compute $\boldsymbol{\pi}_\Sigma$. We illustrate that COM-MTDP can be used to evaluate such coordination strategies, by focusing on a key theory and a practical teamwork model.

Communication under Joint Intentions

Joint-intention theory provides a prescriptive framework for multi-agent coordination in a team setting. It does not make any claims of optimality in its coordination, but it provides theoretical justifications for its prescriptions, grounded in the attainment of mutual belief among the team members. Our goal here is to identify the domain properties under which joint-intention theory provides a good basis for coordination. By "good", we mean that a team of agents communicates as specified under joint-intention theory will achieve a high expected utility in performing a given task. A team following joint-intention theory may not perform optimally, but we wish to find a method for identifying precisely how suboptimal the performance will be. In addition, we have already shown that finding the optimal policy is a complex problem. The complexity of finding the joint-intention coordination policy is lower than that of the optimal policy. In many problem domains, we may be willing to incur the suboptimal performance in exchange for the gains in efficiency.

Under the joint-intention framework, the agents, α, make commitments to achieve certain joint goals. We specify a joint goal, G, as a subset of states, $G \subseteq S$, where the desired condition holds. Under joint-intention theory, when an individual privately believes that the team has achieved its joint goal, it should then attain mutual belief of this achievement, with its teammates. Presumably, such a prescription indicates that joint intentions do not apply to *individually observable* environments, since all of the agents would observe that $S^t \in G$, and they would instantaneously attain mutual belief. Instead, the joint-intention framework aims at domains with some degree of unobservability. In such domains, the agents must communicate to attain mutual belief. However, we can also assume that joint-intention theory does not apply to domains with *free communication*. In such domains, we do not need any specialized prescription for behavior, since we can simply have the agents communicate everything, all the time.

The joint-intention framework does not specify a precise communication policy for the attainment of mutual belief. Let us consider one possible instantiation of joint-intention theory: a *simple joint-intention* communication policy, π_Σ^{SJI}, that prescribes that agents communicate the achievement of the joint goal, G, in each and every belief state where they believe G to be true. Under joint intentions, we make an assumption of *sincerity* (Smith & Cohen 1996), so that the agents never select the special σ_G message in any belief state where G is not believed to be true with certainty. We make no other assumptions about what messages π_Σ^{SJI} specifies in all other belief states.

Given the assumption of sincerity, we can assume that all of the other agents immediately accept the special message, σ_G, as true, and we define our post-communication state estimator function accordingly. With this assumption, as well as our assumptions of perfect communication, the team attains mutual belief that G is true immediately upon receiving the message, σ_G.

We can define the *negated* simple joint-intention communication policy, $\pi_\Sigma^{\neg SJI}$, as an alternative to π_Σ^{SJI}. In particular, this negated policy *never* prescribes communicating the σ_G message. For all other belief state, the two policies are identical. The difference between these policies depends on many factors within our model. Given that the domain is not individually observable, certain agents may not be aware of the achievement of G when it happens. The risk with the

	Individually Observable	Collectively Observable	Collectively Unobservable
No Comm.	$\Omega(1)$	$\Omega(1)$	$\Omega(1)$
General Comm.	$\Omega(1)$	$O((\lvert S\rvert \cdot \lvert\mathbf{\Omega}\rvert)^T)$	$O((\lvert S\rvert \cdot \lvert\mathbf{\Omega}\rvert)^T)$
Free Comm.	$\Omega(1)$	$\Omega(1)$	$\Omega(1)$

Table 2: Time complexity of choosing between SJI and $\neg SJI$ policies at a single point in time.

negated policy, $\pi_\Sigma^{\neg SJI}$, is that, even if one of their teammates knows of the achievement, these unaware agents may unnecessarily continue performing actions in the pursuit of achieving G. The performance of these extraneous actions could potentially incur costs and lead to a lower utility than one would expect under the simple joint-intention policy, π_Σ^{SJI}.

To more precisely compare a communication policy against its negation, we define the following difference value, for a fixed domain-level policy, $\boldsymbol{\pi}_A$:

$$\Delta_X^T \equiv V^T(\boldsymbol{\pi}_A, \boldsymbol{\pi}_\Sigma{}^X) - V^T(\boldsymbol{\pi}_A, \boldsymbol{\pi}_\Sigma{}^{\neg X}) \qquad (2)$$

So all else being equal, we would like to see that the simple joint-intention policy dominates its negation—i.e., $\Delta_{SJI}^T \geq 0$. The execution of the two policies differs only if the agents achieve G *and* one of the agents comes to know this fact with certainty. Due to space restrictions, we omit the precise characterization of the policies, but we can use our COM-MTDP model to derive a computable expression corresponding to our condition, $\Delta_{SJI}^T \geq 0$. This expression takes the form of a page-long inequality grounded in the basic elements of our model. Informally, the inequality states that we prefer the SJI policy whenever the magnitude of the cost of execution after already achieving G outweighs the cost of communication of the fact that G has been achieved. At this high level, the result may sound obvious, but the inequality provides an operational criterion for an agent to make optimal decisions. Moreover, the details of the inequality provide a measure of the complexity of making such optimal decisions. In the worst case, determining whether this inequality holds for a given domain requires iteration over all of the possible sequences of states and observations, leading to a computational complexity of $O((|S| \cdot |\mathbf{\Omega}|)^T)$.

However, under certain domain conditions, we can reduce our derived general criterion to draw general conclusions about the SJI and $\neg SJI$ policies. Under *no communication*, the inequality is always false, and we know to prefer $\neg SJI$ (i.e., we do not communicate). Under *free communication*, the inequality is always true, and we know to prefer SJI (i.e., we always communicate). Under no assumptions about communication, the determination is more complicated. When the domain is *individually observable*, the inequality is always false (unless under *free communication*), and we prefer the $\neg SJI$ policy (i.e., we do not communicate).

When the environment is only partially observable, then agent i is unsure about whether its teammates have observed that G has been achieved. In such cases, the agent must evaluate the general criterion in its full complexity. In other words, it must consider all possible sequences of states and observations to determine the belief states of its teammates. Table 2 provides a table of the complexity required to evaluate our general criterion under the various domain properties.

STEAM

In choosing between the SJI and $\neg SJI$ policies, we are faced with two relatively inflexible styles of coordination: always communicate or never communicate. Recognizing that practical implementations must be more flexible in addressing varying communication costs, the STEAM teamwork model includes decision-theoretic communication selectivity. The algorithm that computes this selectivity uses an inequality, similar to the high-level view of our optimal criterion for communication: $\gamma \cdot C_{mt} > C_c$ Here, C_c is the cost of communication that the joint goal has been achieved; C_{mt} is the cost of miscoordinated termination, where miscoordination may arise because teammates may remain unaware that the goal has been achieved; and γ is the probability that other agents are unaware of the goal being achieved.

While each of these parameters has an corresponding expression in the optimal criterion for communication, in STEAM, each parameter is a fixed constant. Thus, STEAM's criterion is not sensitive to the particular world and belief state that the team finds itself in at the point of decision, nor does it conduct any lookahead over future states. Therefore, STEAM provides a rough approximation to the optimal criterion for communication, but it will suffer some suboptimality in domains of reasonable complexity.

On the other hand, while optimality is crucial, feasibility of execution is also important. Agents can compute the STEAM selectivity criterion in constant time, which presents an enormous savings over the worst-case complexity of the optimal decision, as shown in Table 2. In addition, STEAM is more flexible than the SJI policy, since its conditions for communication are more selective.

Summary and Future Experimental Work

In addition to providing these analytical results over general classes of problem domains, the COM-MTDP framework also supports the analysis of *specific* domains. Given a particular problem domain, we can use the COM-MTDP model to construct an optimal coordination policy. If the complexity of computing an optimal policy is prohibitive, we can instead use the COM-MTDP model to evaluate and compare candidates for approximate policies.

To provide a concrete illustration of the application of the COM-MTDP framework to a specific problem, we are investigating an example domain inspired by our experiences in constructing teams of agent-piloted helicopters. Consider two helicopters that must fly across enemy territory to their destination without detection. The first, piloted by agent A, is a transport vehicle with limited firepower. The second, piloted by agent B, is an escort vehicle with significant firepower. Somewhere along their path is an enemy radar unit, but its location is unknown (a priori) to the agents. Agent B's escort helicopter is capable of destroying the radar unit upon encountering it. However, agent A's transport helicopter is not, and it will itself be destroyed by enemy units if detected. On the other hand, agent A's transport helicopter can escape detection by the radar unit by traveling at a very low altitude (*nap-of-the-earth* flight), though at a lower speed than at its typical, higher altitude. In this scenario, agent B will not worry about detection, given its superior firepower; therefore, it will fly at a fast speed at its typical altitude.

The two agents form a top-level joint commitment, G_D, to reach their destination. There is no incentive for the agents to communicate the achievement of this goal, since they will both eventually reach their destination, at which point they will achieve mutual belief in the achievement of G_D. However, in the service of their top-level goal, G_D, the two agents also adopt a joint goal, G_R, of destroying the radar unit, since, without destroying the radar unit, agent A's transport helicopter cannot reach the destination. We consider here the problem facing agent B with respect to communicating the achievement of goal, G_R. If agent B communicates the achievement of G_R, then agent A knows that the enemy is destroyed and that it is safe to fly at its normal altitude (thus reaching the destination sooner). If agent B does *not* communicate the achievement of G_R, there is still some chance that agent A will observe the event anyway. If agent A does not observer the achievement of G_R, then it must fly nap-of-the-earth the whole distance, and the team receives a lower reward because of the later arrival. Therefore, agent B must weigh the increase in expected reward against the cost of communication.

We are in the process of conducting experiments using this example scenario with the aim of characterizing the optimality-efficiency tradeoffs made by various policies. In particular, we will compute the expected reward achievable by the agents using the optimal coordination policy vs. STEAM vs. the simple joint-intentions policy. We will also measure the amount of computational time required to generate these various policies. We will evaluate this profile over various domain parameters—more precisely, agent A's ability to observe G_R and the cost of communication. The results of this investigation will appear in a forthcoming publication (Pynadath & Tambe 2002).

References

Bernstein, D. S.; Zilberstein, S.; and Immerman, N. 2000. The complexity of decentralized control of Markov decision processes. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 32–37.

Cohen, P. R., and Levesque, H. J. 1991. Teamwork. *Nous* 25(4):487–512.

Grosz, B., and Kraus, S. 1996. Collaborative plans for complex group actions. *Artificial Intelligence* 86:269–358.

Ho, Y.-C. 1980. Team decision theory and information structures. *Proceedings of the IEEE* 68(6):644–654.

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

Marschak, J., and Radner, R. 1971. *The Economic Theory of Teams.* New Haven, CT: Yale University Press.

Matarić, M. J. 1997. Using communication to reduce locality in multi-robot learning. In *Proceedings of the National Conference on Artificial Intelligence*, 643–648.

Pynadath, D. V., and Tambe, M. 2002. Multiagent teamwork: Analyzing the optimality and complexity of key theories and models. In *Proceedings of the International Joint Conference on Autonomous Agents and Multi-Agent Systems*, to appear.

Rich, C., and Sidner, C. 1997. COLLAGEN: When agents collaborate with people. In *Proceedings of the International Conference on Autonomous Agents (Agents'97)*.

Smallwood, R. D., and Sondik, E. J. 1973. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research* 21:1071–1088.

Smith, I. A., and Cohen, P. R. 1996. Toward a semantics for an agent communications language based on speech-acts. In *Proceedings of the National Conference on Artificial Intelligence*, 24–31.

Sonenberg, E.; Tidhard, G.; Werner, E.; Kinny, D.; Ljungberg, M.; and Rao, A. 1994. Planned team activity. Technical Report 26, Australian AI Institute.

Tambe, M. 1997. Towards flexible teamwork. *Journal of Artificial Intelligence Research* 7:83–124.

Yen, J.; Yin, J.; Ioerger, T. R.; Miller, M. S.; Xu, D.; and Volz, R. A. 2001. CAST: Collaborative agents for simulating teamwork. In *Proceedings of the International Joint Conference on Artificial Intelligence*.