Decision Making in Complex Multiagent Contexts: A Tale of Two Frameworks

Prashant Doshi

■ *Decision making is a key feature of autonomous systems. It involves choosing optimally between different lines of action in various information contexts that range from perfectly knowing all aspects of the decision problem to having just partial knowledge about it. The physical context often includes other interacting autonomous systems, typically called agents. In this article, I focus on decision making in a multiagent context with partial information about the problem. Relevant research in this complex but realistic setting has converged around two complementary, general frameworks and also introduced myriad specializations on its way. I put the two frameworks, decentralized partially observable Markov decision process (Dec-POMDP) and the interactive partially observable Markov decision process (I-POMDP), in context and review the foundational algorithms for these frameworks, while briefly discussing the advances in their specializations. I conclude by examining the avenues that research pertaining to these frameworks is pursuing.*

A key feature of an autonomous system, here onwards referred to as an agent (Russell and Norvig 2010), is the capability to choose optimally between different lines of action available to it. This capability of normative decision making becomes demanding in diverse informational and physical contexts, which may range from having precise information about all aspects of the problem to partial knowledge about it and multiple interacting agents.

For illustration, consider a toy problem involving an autonomous unmanned aerial vehicle (AUAV) tasked with intercepting a fugitive in its theater of surveillance that is divided into a grid of large sectors. Interception requires the AUAV to move to the sector occupied by the fugitive. Consistent sighting reports, which may be noisy, could lead the AUAV to the sector containing the fugitive. However, a false positive intercept in a proximal sector due to the noise in the sightings would cause the alarmed fugitive to flee his true sector and reappear in some other random sector that is further away. Of course, the AUAV's objective is to succeed in as many such missions as possible over a time period.

Clearly, the AUAV faces a tough decision-making problem because of the uncertainty in sightings and the consequences of making a bad decision. Furthermore, presence of another AUAV in the theater — whether the AUAV is helpful or not — could potentially complicate the decision-making problem. If this

  ISSN 0738-4602

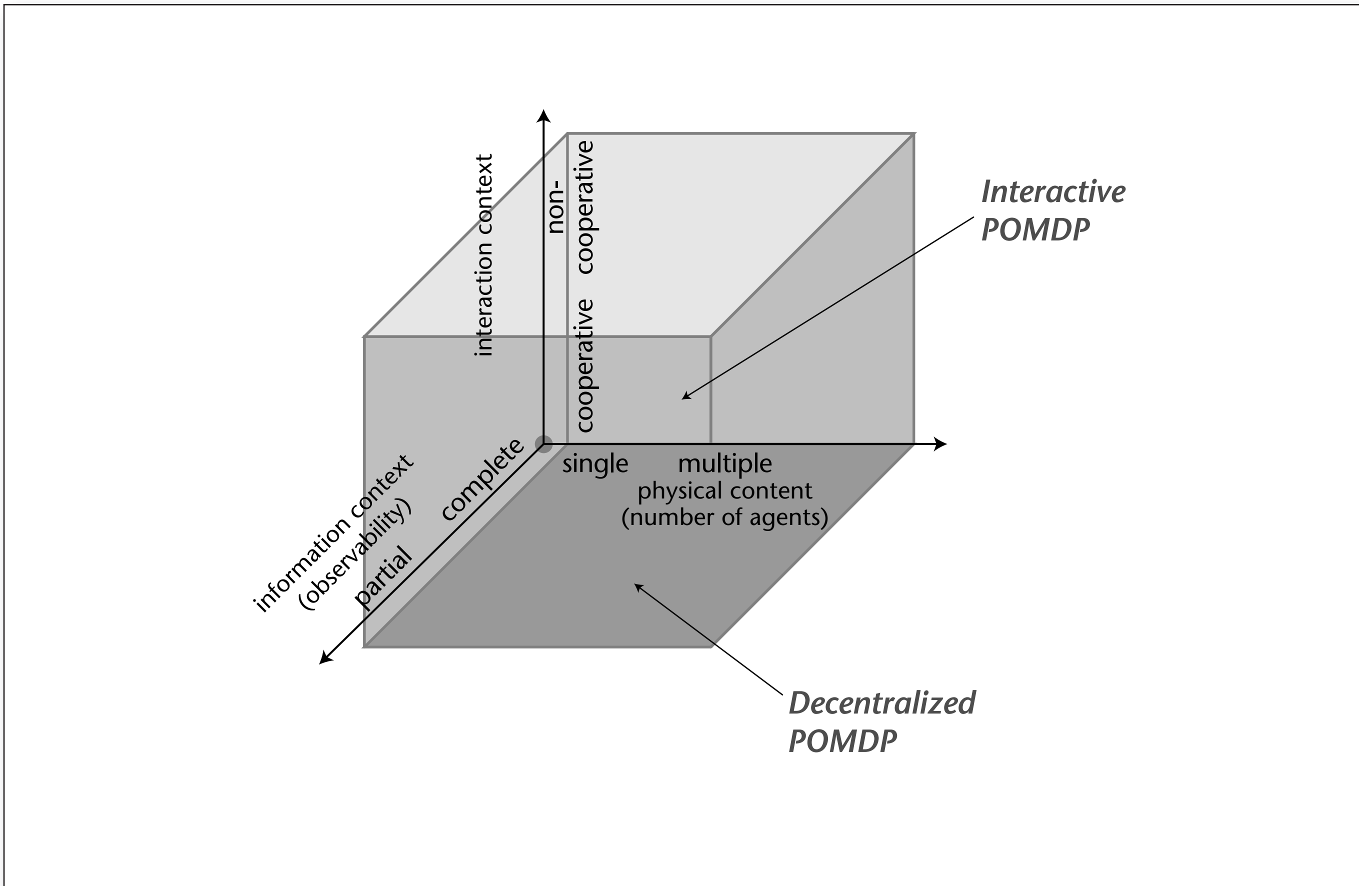

Figure 1. Comparing Two Frameworks.

Both frameworks apply in multiagent contexts with Dec-POMDPs suitable for cooperative interactions only. While I-POMDPS are applicable in both cooperative and noncooperative contexts thereby occupying the remaining volume of the problem space. The parenthesis on an axis gives analogous environment parameters used by Russell and Norvig (2010).

AUAV is helpful, then both AUAVs must jointly decide whether to move to intercept the fugitive in some sector or await more consistent sighting reports. A premature move by one of the AUAVs in the team may spoil the chances for both! This remains true when the other AUAV is neutral in its intentions toward the subject AUAV.

This illustration provides anecdotal evidence about the technical difficulty of sequential decision making in partial information contexts that are shared by multiple agents. In this article, I review two formal frameworks that occupy this problem space. Although relevant AI research has explored various decision-making models, it has gradually converged on these complementary, general frameworks. Both these frameworks are founded on decision theory as formalized by partially observable Markov decision processes (POMDPs) (Smallwood and Sondik 1973; Kaelbling, Littman, and Cassandra 1998) and generously draw inspiration from game theory. Despite the inherent complexity, decision-theoretic frameworks such as POMDPs offer a principled and theoretically sound formalism for decision making under uncertainty with guarantees of optimality of the solution, which makes them an appealing choice for extending to multiagent settings.

The first framework that I consider is the decentralized POMDP (Dec-POMDP) (Bernstein et al. 2002), which is applicable in contexts where the agents are strictly cooperative. In this framework, we seek an optimal joint behavior for all agents given an initial state of the problem that is common knowledge to all and the agents receive a common reward based on their joint actions. In contrast, the interactive POMDP (I-POMDP) (Gmytrasiewicz and Doshi 2005) models the problem from the subjective view of an individual agent in a multiagent context. The framework seeks to find agent behavior that is individually optimal and applies to both cooperative and noncooperative contexts. Figure 1 compares the two frameworks based on the problem spaces they occupy. The information and interaction contexts in figure 1 may be viewed as continuums while the physical context is discrete but could get arbitrarily large.

Optimally accomplishing tasks using a team of agents has continually fascinated AI researchers. The Dec-POMDP was preceded by substantial and distinct efforts in teamwork such as STEAM (Tambe 1997), which was based on an early formal theory of joint intentions (Cohen and Levesque 1991). An alternate theory of teamwork involved the concept of SharedPlans (Grosz 1996). These elaborate theories provided logical formalizations of requisite joint mental states and plans thereby facilitating teamwork through reasoning. Parallel advances on a myriad of frameworks that could be linked to a Dec-POMDP as specializations have also continued. Building on this past progress, Dec-POMDPs emphasize quantitative mathematical modeling of the economics that promotes cooperation and of realistic uncertainties with the benefit that precise optimality of behavior can be claimed and the computational complexity of the problem class be ascertained. However, Dec-POMDPs champion offline decision making for every contingency as opposed to the execution-centered thrust of the logical teamwork theories, which promoted efficiency.

While a focus on cooperation admits some simplifications such as common initial state and reward, adopting an agent's view of a multiagent context does the opposite. Although the focus is on individual behavior, issues such as considering what the other agent(s) thinks that the subject agent will do and about the subject agent's thought on others' actions confound the decision making in general. I-POMDPs elegantly tie recursive thinking to the process of decision making using modeling approaches initially explored in game theory (Harsanyi 1967, Brandenburger and Dekel 1993, Aumann 1999). I-POMDPs, too, build upon previous efforts in AI such as the recursive modeling method (Gmytrasiewicz and Durfee 2000), which targets individual decision making in simpler contexts.

Recent applications of the two frameworks testify to their significance and growing appeal. Dec-POMDPs are finding applications in coordinating scans in sensor networks exhibiting various configurations (Kim et al. 2006), in emerging smart energy grids, and in coordinating AUAVs in a team. I-POMDPs are being used to explore strategies for countering money laundering by terrorists (Ng et al. 2010, Meissner 2011) and enhanced to include trust levels for facilitating defense simulations (Seymour and Peterson 2009a, 2009b). They have been used to produce winning strategies for playing the lemonade stand game (Wunder et al. 2011) and even modified to include empirical models for simulating human behavioral data pertaining to strategic thought and action (Doshi et al. 2010).

In the remainder of this article, I acquaint the reader with the two frameworks while keeping in mind their important foundations. Then, I review the early algorithmic developments that built a solid platform for continuing research in each framework. Throughout this article, I compare the two frameworks, pointing out some similar and some differing facets, and conclude this article with a discussion of some research directions that this article did not touch upon and those that are ripe for pursuing.

Dual Basis: Decision and Game Theories

As I mentioned previously, both Dec- and I-POMDP are founded on decision theory, which has its own roots in the axioms of probability theory and of von Neumann and Morgenstern's utility theory (1947). In particular, the frameworks generalize an instantiation of the principles of decision theory for partial information contexts called the partially observable Markov decision process. Although POMDPs first appeared in the operations research literature (Smallwood and Sondik 1973), they were later cast into the AI spotlight (Kaelbling, Littman, and Cassandra 1998) as a framework for optimal control of single agents.

An agent, say *i,* in the POMDP framework acts and makes observations in a sequential cycle, as I show in figure 2. The partial information context results from agent *i* receiving observations that do not reveal the current physical state precisely. Nevertheless, we seek the actions that agent *i* should perform in order to maximize the reward it can get or its preferences.

A POMDP models the decision-making problem for agent *i* using the following tuple:

$$POMDP = \langle S, A_i, T_i, \Omega_i, O_i, R_i, OC_i \rangle$$

Here, S is the set of physical states of the agent and its environment that are relevant to the decision making. In the realm of my previous illustrative example, the states could be the different sectors in the AUAV's grid. A_i is the set of actions available to the agent, such as the AUAV's maneuvering and sensory actions. T_i is the transition function that models the dynamism and part of the uncertainty in the problem. It does this by mapping each transition from one state to another given the agent's action to its probability. Ω_i is the set of observations that the agent makes, such as the AUAV spotting the fugitive and receiving sighting reports. O_i is the observation function that gives the likelihood of making an observation on performing an action and moving to the subsequent state. It models another source of uncertainty in the problem, one that manifests due to noise in the agent's observation sensors. Together, A_i, T_i, Ω_i and O_i in the tuple, represent the capabilities of the agent; R_i is the quantitative reward or preference function that models the preference structure of the agent; and OC_i is the agent's optimality criterion, which typically takes the form of maximizing cumulative rewards over a finite number of time steps or maximizing cumulative discounted rewards over an infinite number of steps in the limit.

The solution of a POMDP is a function called the policy, which maps the agent's history of observations and an action to the probability with which the action should be performed. We may compress the representation of a policy by noting that a probability distribution over the states, called the agent's belief, is a sufficient statistic for the agent's observation history. While POMDPs offer principled optimality in near-real contexts, their usefulness is inhibited by the high complexity of solving them. In particular, the POMDP problem whose optimality criterion involves maximizing reward over a finite number of time steps that is less than the number of states is PSPACE-complete (Papadimitriou and Tsitsiklis 1987), otherwise it is incomputable for an infinite number of steps (Madani, Hanks, and Condon 2003). A factor that drives up the complexity is the disproportionate growth in the size of the belief space over which the solution is optimized as the number of states increase, which I call the belief space complexity. The other predominant factor is the policy space complexity, which is the exponential growth in the number of policies to search over with time, and is also affected by the action and observation spaces. Due to this, approximate solutions that make solving POMDPs feasible are crucial. Various insights have led to a multitude of approximation techniques. Notable among these is a group that utilizes a subset of belief points over which to optimize the solution instead of over the entire simplex. These point-based approximations (Pineau, Gordon, and Thrun 2006; Kurniawati, Hsu, and Lee 2008) now allow POMDPs to scale to reasonably large domains thereby facilitating applications.

Although POMDPs lay down the groundwork, the bridge to multiagent contexts is built on insights from game theory. Oddly, both cooperative and noncooperative interaction contexts are studied in the noncooperative branch of game theory. Both Dec- and I-POMDP frameworks make an analogy to the use of types while modeling the multiple agents. Specifically, agent models include sufficient information to predict the agent's actions. This is consistent with Harsanyi's stance (1967), which describes a type as an attribute vector summarizing the physical, social, and psychological attributes of the agent including its payoffs and beliefs, all of which are relevant to its actions. Efforts to concretely define the type have converged on a coherent infinite hierarchy of beliefs with topological assumptions on the state space (Mertens and Zamir 1985, Brandenburger and Dekel 1993), as representing the agent's type. These beliefs, while not computable, play an important role in individual decision making in multiagent contexts, and indeed, they find a place in the I-POMDP framework.

Surprisingly, the game-theoretic solution concept of Nash equilibrium finds little consensus as a solution for Dec- and I-POMDP frameworks. A Nash equilibrium (Fudenberg and Tirole 1991) is a profile of behaviors for the different agents such that each behavior is best for the corresponding agent given others' behaviors. While equilibria-based solution techniques for Dec-POMDPs were proposed (Nair et al. 2003), recent focus has been firmly on computing globally optimal solutions instead of settling for the locally optimal strategies that Nash equilibria provide, of which there could be many. Of course, the globally optimal solution is a Pareto-optimal Nash equilibrium in the cooperative context. Other limitations such as the possible existence of multiple Nash equilibria with no clear way to choose between them and the inability of an equilibrium-based solution to account for behaviors outside the steady state, make equilibria inappropriate for I-POMDPs as well.

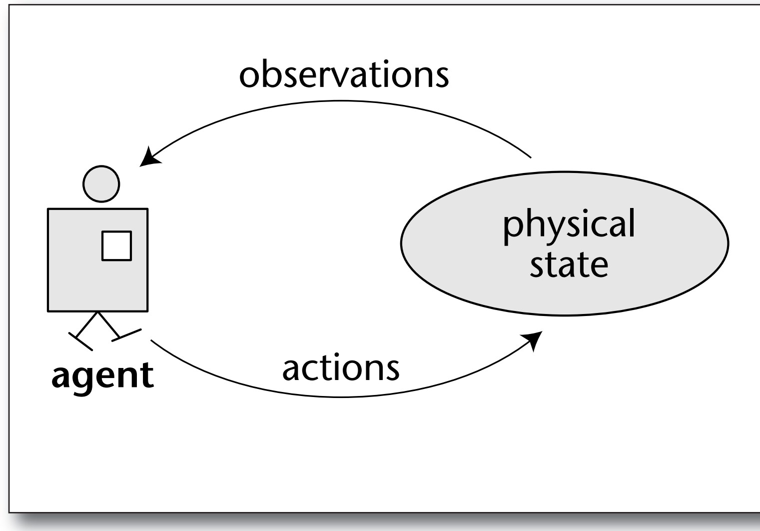

Figure 2. A Schematic Diagram of the POMDP Framework.

Actions of the agent affect its own physical state and that of its environment. As is often the case in the real world, the agent's observations do not reveal the possibly changed physical state precisely. This makes meeting the objective of finding actions that maximize the agent's reward or preferences challenging.

Decision Making in Multiagent Contexts

If it is common knowledge that the two AUAVs in my previous illustrative example are cooperative, then we would be interested in how each chooses its action given its observations such that their joint behavior is optimal in intercepting the fugitive. On the other hand, an individual AUAV's deliberation about what the other AUAV might do and its own subsequent action in the setting also interests us, and more so if cooperation is not a given or if the initial belief state of the AUAVs is uncertain. Both these directions of investigation are critical to a holistic understanding of decision making.

Two Sides of the Same Coin

Although the Dec- and I-POMDP differ in their perspectives and in the interaction contexts, both offer a framework for normative decision making in multiagent contexts and share procedural elements. Figure 3 shows schematic diagrams for both frameworks that detail how the different components of the processes interact. Each agent acts and makes an observation, which influences its belief. Expectedly, actions of both agents affect the physical state of the problem and the rewards that each agent receives. Keeping track of the physical state becomes relevant because it may influence the rewards as well. Therefore, the interaction between the agents manifests not only in the rewards that each gets but also in the dynamic physical state of the problem. Indeed, this makes for an expressive but complex interaction model between agents.

As is shown in figure 3, the I-POMDP perspective merges the physical state with how the other agent(s) acts and observes into an interactive state. An agent in this perspective keeps track of the dynamic interactive state of the problem. This offers an agent's point of view to the interaction and allows the agent to deliberate and decide at its own individual level. In comparison, the Dec-POMDP's focus is on both agents who share a common reward function.

Joint Cooperative Decision Making

Guiding a team of agents in order to accomplish tasks is a classic AI problem with implications in many fields such as in robotics. However, the problem is infamously hard in realistic contexts. A Dec-POMDP provides a framework within which we may formalize this problem in partially observable information contexts. In such contexts, even knowledge of all agents' observations is not sufficient to precisely determine the physical state. Although a Dec-POMDP models the cooperative decision-making problem expressively, it stays silent on how the team should behave when faced with adversarial agents sharing its environment.

For the sake of readability, I define the Dec-POMDP for a two-agent context, but it is applicable to more agents as well. It generalizes the POMDP outlined previously to a multiagent context using the following tuple:

$$\textit{Dec-POMDP} = \langle S, s_0, A, T, \Omega, O, R, OC \rangle$$

While S is analogous to that in a POMDP, s_0 is the initial state; A is the set of joint actions of both agents, $A = A_i \times A_j$ where A_i and A_j are the action sets of agents i and j, respectively; T is the joint transition function and maps each transition between states given the joint actions of the two agents to its probability; Ω is the set of joint observations, $\Omega = \Omega_i \times \Omega_j$ where Ω_i and Ω_j are the observation sets of agents i and j, respectively; the joint observation function, O, gives the likelihood of both agents jointly making their observations given their joint actions and the resulting state; R is the common reward function that maps joint actions and the state to a real number representing the reward shared by the agents. In some cases, the reward may also be associated with a complete state transition; and OC is the optimality criterion analogous to that for POMDPs. Given the partial observability of the state space, the initial state is often replaced with an initial belief over the state space that is known to all agents. Dec-POMDPs are usually solved in advance of the actual problem occurrence. Techniques for solving Dec-POMDPs often internally make use of a multiagent belief state, which is a distribution over the physical states and possible policies of the other agent.

Simplifications to facets of the Dec-POMDP definition motivated by the problem domains on hand abound: If observations of the agents are conditionally independent of each other given the joint action and state, as is sometimes the case, we may factor the joint observation function, although this is not a requirement for Dec-POMDPs. Each component focuses on the likelihood of an individual agent's observation given the joint action and state. Another simplification, observation independence, further trims down the observation function components by assuming that the likelihood of an agent's observation depends on its own action and the resulting state only. Analogously, an assumption of transition independence allows the transition function to be factored into components that map the local transitions of each agent — between its local states given its own action — to a probability. These simplifications are desirable because, through an ingenious reduction from the tiling problem, Bernstein et al. (2002) showed that solving the Dec-POMDP is NEXP-complete for the finite time steps case. For the first time, this result explicitly illustrates how hard it really is for agents to engage in principled and optimal cooperation in uncertain contexts.

Behavior of a team of two AUAVs tasked with

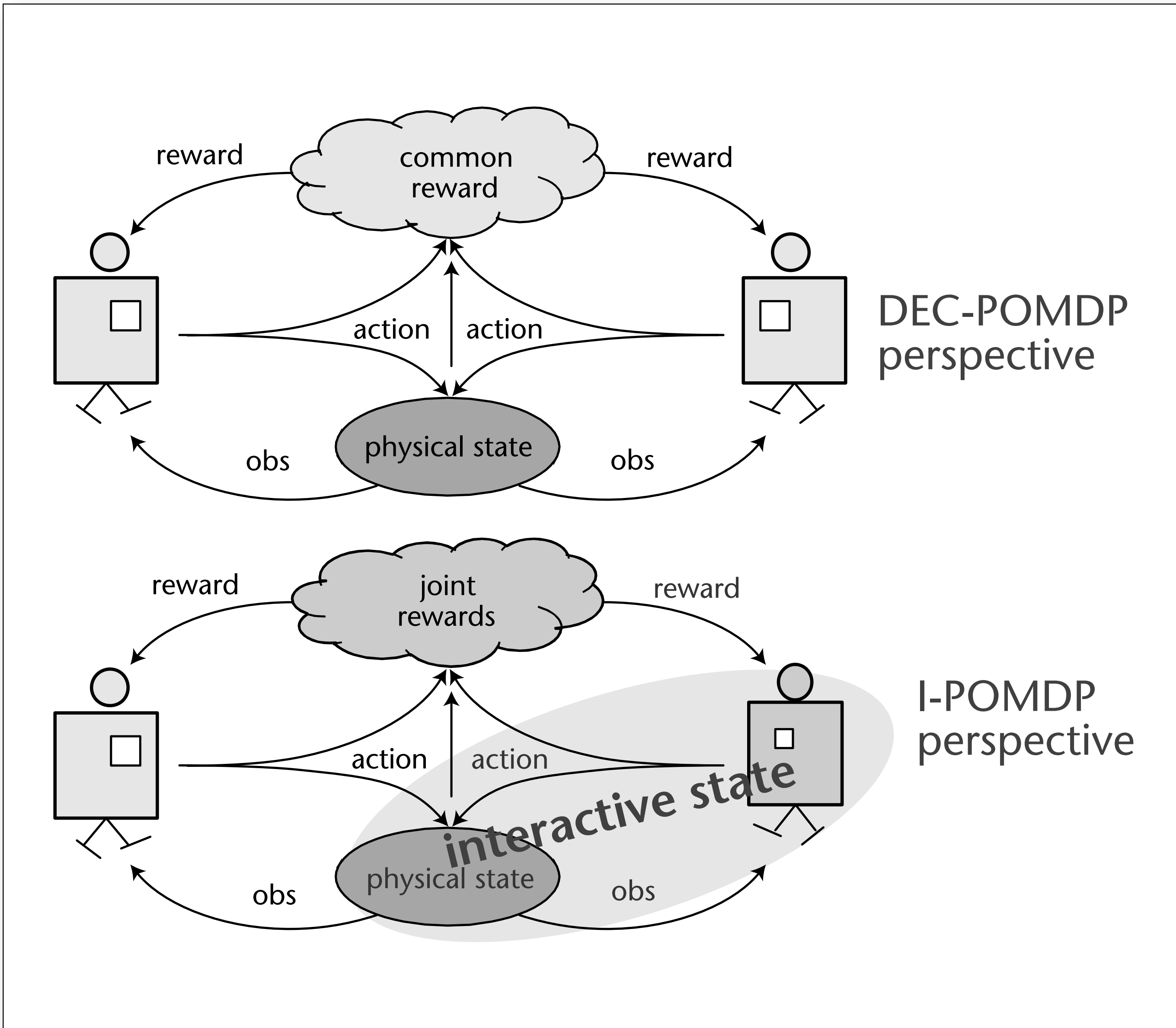

Figure 3. Schematic Diagrams for the Dec- and I-POMDP Frameworks.

This figure shows the procedural similarities in their setup and highlights the key differences between their perspectives. In each framework, decisions of the agent(s) in bold are sought.

intercepting the fugitive could be modeled as a Dec-POMDP. The physical state space would be possible combinations of the sector locations of the AUAVs and the fugitive. The joint action and observation spaces would be combinations of the two AUAVs' individual actions and observations mentioned previously in the context of POMDPs. Because a false positive intercept by any of the AUAVs causes the fugitive to reappear in a random sector, the problem interestingly exhibits neither transition nor observation independences.

An independent framework shown to be equivalent to the Dec-POMDP under the condition that agents perfectly remember their past history of actions and observations (Seuken and Zilberstein 2008) is the Markov team decision process (MTDP) (Pynadath and Tambe 2002). While the underlying condition is a de facto assumption in both the frameworks discussed in this article, its violation exacerbates the uncertainty in the information context; decision making in such situations remains as yet unexplored.

Specializations

Of course, the information context, in general, need not always be partially observable, and Goldman and Zilberstein (2004) explore various flavors of the information context ranging from joint partial observability to local full observability categorizing the complexity of the cooperative decision-making problem in each. One of these problems is an important specialization called the decentralized MDP, which applies when the physical state is jointly fully observable. This specialization is of significance because if the information context for each agent in the decentralized MDP is additionally locally fully observable, and the problem exhibits both transition and observation independences, its complexity lessens to being NP-complete (Becker et al. 2004), although it is still generally intractable.

Problem domains such as cooperative sensing by sensor networks where adjacency of the sensor motes defines interaction motivate frameworks that could be perceived as specializations of Dec-POMDPs. One such framework is the networked and distributed POMDP (ND-POMDP) (Nair et al. 2005) that exploits the locality and sparseness of the interaction between agents in solving the cooperative decision-making problem. This locality is modeled using an interaction hypergraph, and additional efficiency of the decision making is enabled by utilizing transition and observation independences between the interacting agents. While the locality allows a factorization of the reward function, the physical state space may also be factored into a set of variables and the conditional independence between the variables is exploited to represent problems more efficiently (Oliehoek et al. 2008).

Individual Decision Making

The external, objective perspective — as adopted by Dec-POMDPs — is suitable for controlling agents in cooperative contexts, but is more appropriate for an analysis rather than control of agents in noncooperative contexts. On the other side of the coin, an individual, subjective perspective as adopted by I-POMDPs offers a single approach toward control of agents in interaction contexts where the other agents are cooperative, noncooperative, or a mixture of both. In the game theory literature, this approach has been previously recognized as the decision-theoretic approach to game theory (Kadane and Larkey 1982). However, the elegance of a single approach comes with the cost of maintaining explicit belief systems that should be continually updated, and these belief systems could become noncomputable in general.

An I-POMDP for an agent *i* sharing its environment with one other agent, *j*, is formalized using the following tuple of parameters:

$$\text{I-POMDPi} = \langle IS_i, A, \Omega_i, T_i, O_i, R_i, OC_i \rangle$$

Here, IS_i is the set of interactive states: the physical states of the problem, S, augmented with models, M_j, of the other interacting agent, $IS_i = S \times M_j$. For the sake of operationalizing the framework, the models are limited to be computable. While models need not be just intentional, the focus so far has been on such models. An intentional model, $\theta_j \times \Theta_j$, is analogous to a type in game theory and encapsulates all the private information about the agent relevant to its actions. This is formalized in terms of the agent's beliefs, capabilities, and preferences: $\theta_j = \langle b_j, A, \Omega_j, T_j, O_j, R_j, OC_j \rangle$, where b_j is a distribution over IS_j, and the remaining parameters have their standard interpretations. Next, A is the set of joint actions of the agents, $A = A_i \times A_j$, where A_i and A_j are the action sets of agents *i* and *j*, respectively. Ω_i is the set of observations made by agent *i*; T_i and O_i are analogous to those defined in POMDPs, with the exception that a transition or an observation usually depends on the joint action. Despite being a part of the interactive state, models do not appear in the transition and observation functions because the framework assumes that models, especially other's beliefs, cannot be directly altered nor observed. Features that could be altered or directly observed should be included in S; R_i gives the reward of agent *i* based on the physical state and the joint actions; and OC_i is the optimality criterion analogous to that for POMDPs. I show the problem setting for the I-POMDP framework in figure 4.

Agent *i*'s belief is a probability distribution over IS_i. Because the interactive state space includes models that contain *j*'s beliefs, agent *i*'s belief is a distribution over different *j*'s beliefs, each of which itself could be a distribution over *i*'s beliefs. Figure 4 also shows this nested modeling for agent *i*. Several researchers, including Binmore (1982) and Brandenburger (2007), have pointed out the existence of self-contradicting, and therefore impossible, beliefs when complete infinite hierarchies are considered, tying the result to other well-known impossibility results such as the Russell Paradox. This, combined with the noncomputability of operating on infinitely nested beliefs, makes I-POMDPs generally not computable and motivates a logical approximation: bounding the nesting to finite levels by assuming a "level 0" belief. Not only does this make the beliefs computable, it also precludes the existence of impossible beliefs alluded to previously. However, as we shall see, this approximation is not without its own limitations. Consequently, finitely nested I-POMDPs additionally include a strategy level, l, that defines the level of nesting with a level 0 model being a POMDP containing a belief that distributes probability over the physical states only. I-POMDPs with increasing strategy levels lead to actions that are more strategic, and the general rule of thumb is

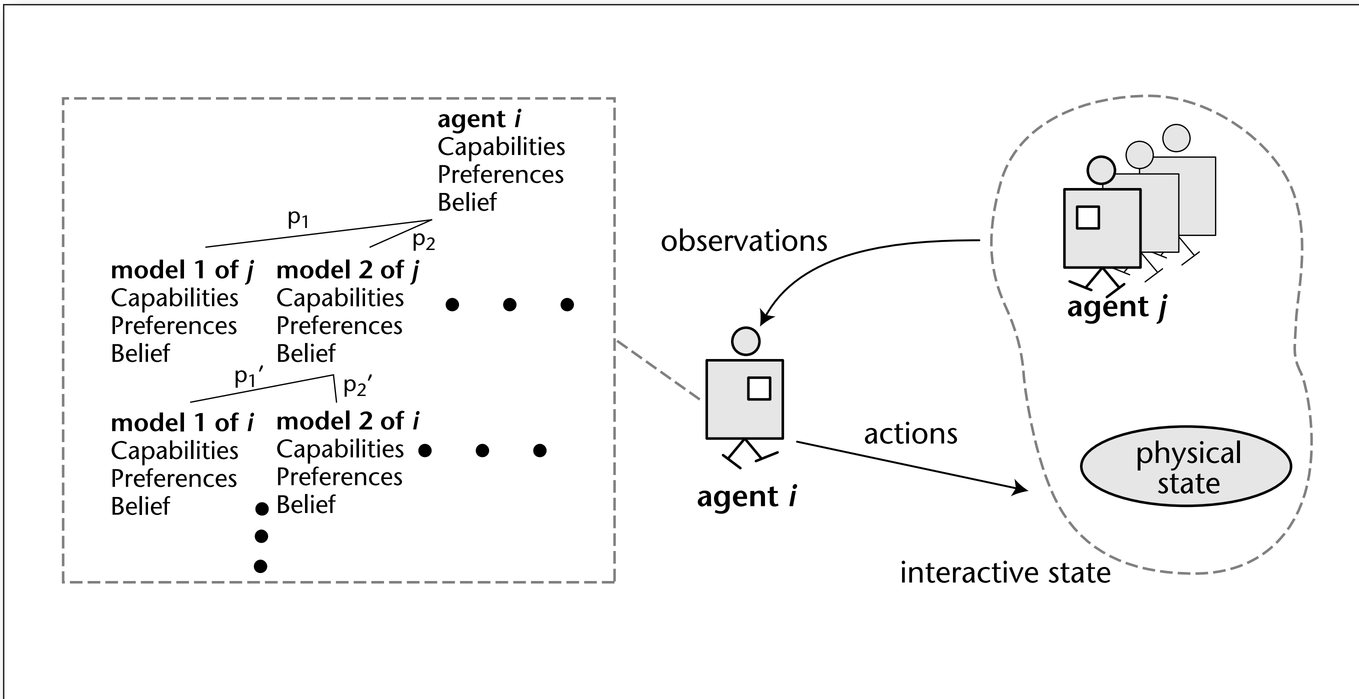

Figure 4. An Agent in the I-POMDP Framework Maintains Explicit Beliefs About the States and Models of Other Agent(s), About Other's Beliefs over States and Models of Others.

In cooperative contexts, a common initial belief as used in a Dec-POMDP collapses the initial belief hierarchy into a single branch of probability 1 at each level. Nevertheless, as the agents act and observe the candidate models grow due to the uncertainty.

to utilize as high a strategy level as we can. While the asymptotic complexity of finitely nested I-POMDPs has not been formally established as yet, it is thought that they are at least as hard to solve as Dec-POMDPs.

Two differences complicate the belief update in multiagent contexts compared to the POMDP belief update. First, since the physical state depends on the actions performed by both agents the prediction of how it changes has to be made based on the probabilities of various actions of the other agent. Probabilities of other's actions are obtained by solving its models. Thus we need not assume that actions are fully observable by others. Rather, agents can attempt to infer what actions other agents have performed by sensing their results on the environment. Second, changes in the models of other agents have to be included in the update. The changes reflect the other's observations and, if they are modeled intentionally, the update of other agent's beliefs. In this case, the agent has to update its beliefs about the other agent based on what it anticipates the other agent observes and how it updates.

The decision making of each AUAV in my illustrative example may be modeled as a finitely nested I-POMDP, particularly if the AUAV is unsure about the type of the other. In this case, M_j would include possible models that ascribe different shades of cooperation, noncooperation and neutrality to the other AUAV using appropriate reward functions in the models. Based on the computational resources available to the AUAV, it would benefit from utilizing a strategy level that is as high as possible.

Relationships

Both Dec- and I-POMDPs solve decision-making problems that are instances of partially observable stochastic games as they are referred to in the game-theoretic realm. Stochastic games are sequential games additionally involving a dynamic state of the game. While these games are well studied, their partially observable generalizations have received scant attention from game-theoretic researchers. Specifically, Dec-POMDPs focus on solving a type of the partially observable stochastic game, which involves identical payoff functions for the agents. I-POMDPs solve the general game but from an individual agent's perspective, transforming the physical state space that is a part of the game's definition into an interactive state space by including others' models. Consequently, advances in Dec- and I-POMDPs have the potential to inform game theory as well.

Because the two frameworks generalize POMDPs, decision-making problems modeled by

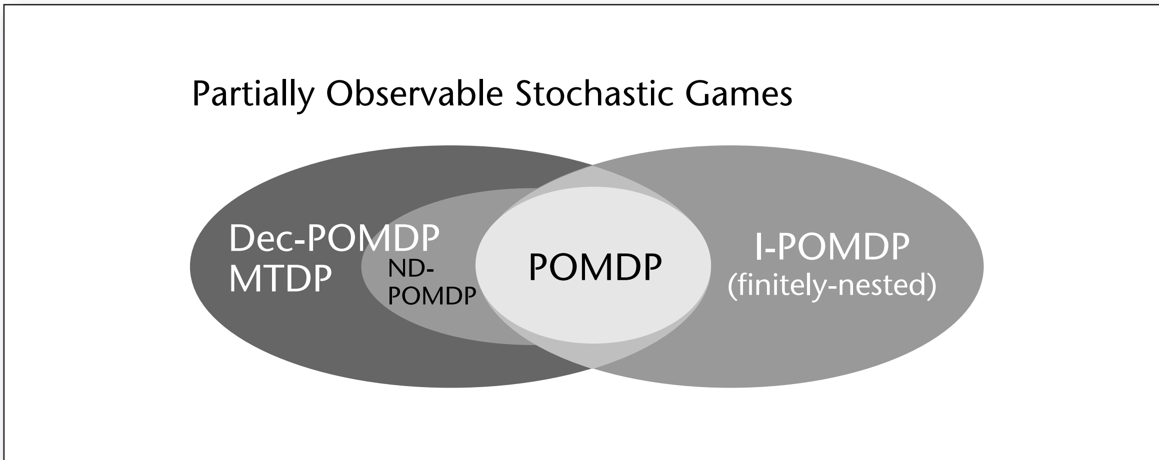

Figure 5. Each Set in the Venn Diagram Represents the Class(es) of Problems that the Corresponding Framework Can Solve.

Gray area in the intersecting region of Dec- and I-POMDP indicates the current uncertainty about whether team behavior as produced by a Dec-POMDP could be obtained from a finitely nested I-POMDP for certain problems, as well. Still, this region may not be empty and is a topic of ongoing investigations.

POMDPs are also instances of Dec- and I-POMDPs with the number of agents as one. Figure 5 shows a Venn diagram displaying the relationships between the members of the different frameworks that I have discussed so far in this article. Note that the diagram is by no means exhaustive in the frameworks that it includes. Its purpose is succinctly to align the classes of problems and perspectives modeled by the different frameworks in order to promote clarity.

Foundational Algorithmic Developments

Algorithms for solving Dec-POMDPs and finitely nested I-POMDPs have drawn heavy inspiration from past algorithms for solving POMDPs. Some of them have also looked toward game theory for direction. Recall that the solution to a Dec-POMDP is a joint policy, which is a vector of policies, one for each agent. On the other hand, solution of an I-POMDP is a single policy for the subject agent in the multiagent context. A policy in both frameworks may be represented using directed trees if the number of time steps are finite; otherwise a finite-state machine is appropriate.

Exact solutions retain the attractive guarantee of optimality. A caveat is that exact solutions of finitely nested I-POMDPs are technically computable approximations of the full I-POMDP framework. However, solving for exactness faces the barrier of high computational complexity. This complexity is partly due to the very large spectrum of policies that must be searched, which is further exacerbated in multiagent contexts where policies for all agents must be considered. Nevertheless, exact solution techniques exist for both frameworks. Analogous to POMDPs, we may define value functions that provide the expected reward for a joint policy in a Dec-POMDP or an individual policy in an I-POMDP, albeit the functions are more complex.

A dynamic programming operator introduced by Hansen, Bernstein, and Zilberstein (2004) kickstarted investigations into feasible algorithms for Dec-POMDPs. In an iterative two-step process, the operator first exhaustively generates policies of increasing length in a bottom-up manner starting with a single action. This is followed by pruning those policies for each agent that are very weakly dominated by other policies. We illustrate the dynamic programming operator in figure 6. Of course, the pruned policy should be dominated at any belief over the state space and no matter the policies of the other agents. This is accomplished by testing for dominance over all multiagent belief states, each of which, as we may recall, is a distribution over the physical state and possible policies of the other agents. Consequently, the operator enables a reduced but sufficient set of policies to be searched at each step. This is analogous to iterated elimination of very weakly dominated behavioral strategies — a well-known technique for compacting games — in the context of partially observable stochastic games. Among the remaining policies for each agent, the joint policy with the largest value is the exact solution of the Dec-POMDP. Although this approach consumes a fraction of the resources that an exhaustive search needs, it fails to solve small toy problems beyond time steps as

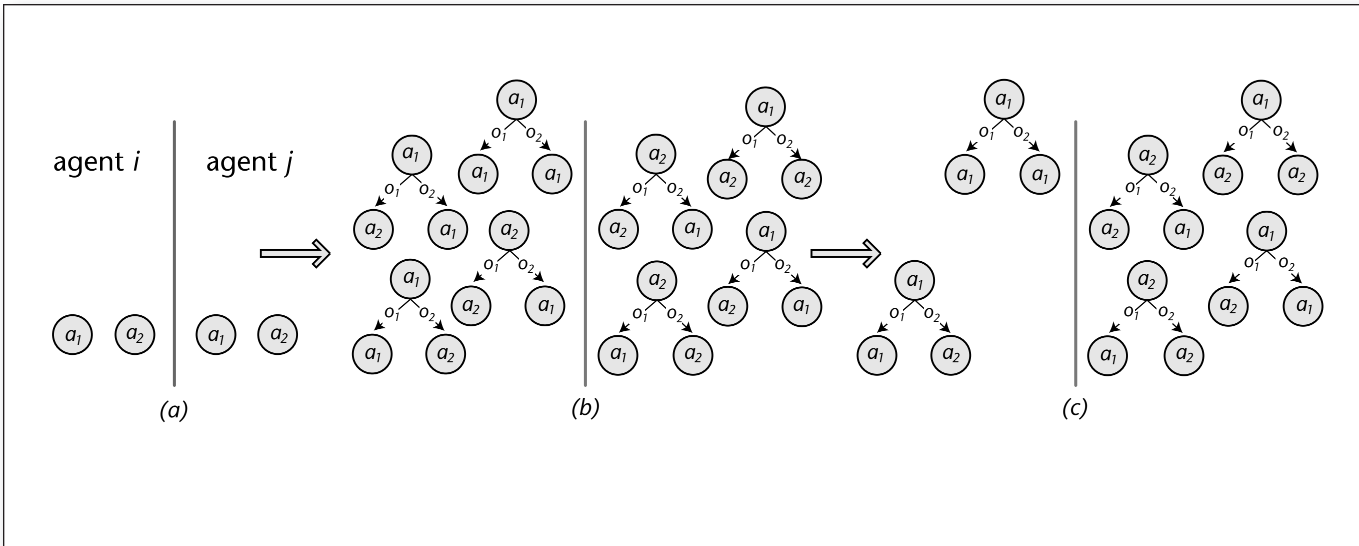

Figure 6. Generic Example of Dynamic Programming for Dec-POMDPs.

To illustrate dynamic programming for Dec-POMDPs, I consider a simple and generic example involving two agents, *i* and *j*, each with two actions, a_1 and a_2, and two observations, o_1 and o_2. (a) Beginning with single actions for each agent, which are optimal, the operator generates possible policies (trees) of increased length for each agent, as shown in (b). Here, I show just a small subset of all possible policies of increased length. Note that the policies differ in the action selected initially or on receiving the two observations, or in both. (c) In the pruning step, two of agent *i*'s policies are found to be very weakly dominated by the remaining over all of *i*'s multiagent belief states, thereby removing them from consideration. The dominance test is iterated for agent *j* as well given the residual set of *i*'s policies, and this step is iteratively repeated until no more policies get pruned for either agent.

short as four. A top-down technique, different from the previous bottom-up dynamic programming, applies A* search where subsequent nodes of the search tree represent joint policies of longer time steps (Szer, Charpillet, and Zilberstein 2005). Of course, an admissible heuristic that is optimistic and never undervalues the complete policies at any time step is needed. A simple approach is to assume reduced uncertainty in the problem such as full observability of the physical states and compute the value. Improved heuristics that make the search more efficient rely on tighter approximations of the joint policy values by, say, collapsing the joint decision problem into a POMDP with joint actions. These heuristics are plugged into a generalized version of the A* search (Oliehoek, Spaan, and Vlassis 2008).

The large model space in finitely nested I-POMDPs is the predominant barrier against exact solutions. Rather than imposing ad hoc restrictions on candidate models, a method for compacting the general space is needed. One way of doing this is to group together models that are behaviorally equivalent (Pynadath and Marsella 2007). If the models are intentional — they are POMDPs or I-POMDPs with differing beliefs — we may utilize their solutions, starting at level 0 and climbing up the strategy levels, to partition the model space into a discrete number of equivalence classes at each level (Rathnasabapathy, Doshi, and Gmytrasiewicz 2006). This provides a tractable foothold on the model space with which we may solve I-POMDPs exactly.

While exact solutions serve as important benchmarks of solution quality, their lack of scalability precludes their usefulness. Needless to say, the bulk of algorithmic development for both frameworks has focused on approximations. These trade the solution quality for greater scalability and often exhibit properties such as monotonically improving solutions as more computational resources are allocated and, less frequently, an upper bound on the loss in optimality given a parametric measure of the resources. Even so, a negative result that the lower-bound complexity of finding an approximate solution of Dec-POMDP within ε of the exact continues to remain in the NEXP class (Rabinovich, Goldman, and Rosenschein 2002) puts a theoretical damper on the benefits of approximations.

An approximation of the exact dynamic programming operator allowed Dec-POMDP-based solutions of toy problems to scale tremendously but without any quality guarantees: up to four orders of magnitude in time steps (Seuken and Zilberstein 2007). While it generates all new policies of increasing length that are possible from the solution of the previous iteration for each agent, it limits the number of new joint policies retained at each step based on the available memory and utilizes a subset of the multiagent belief space, as in point-based approximations for POMDPs, over which to test for dominance of a policy. A portfolio of different heuristics enables a focus on possibly relevant belief points. An improvement on this technique avoids generating all possible policies in the first place by heuristically estimating good quality joint policies that would then be evaluated on the belief points (Dibangoye, Mouaddib, and Chaib-draa 2009). Notice that these techniques are motivated by the efficacy of point-based algorithms in the context of POMDPs, and they affect finitely nested I-POMDPs as well. A generalized version of point-based value iteration (Doshi and Perez 2008) allows the use of select belief points over which to optimize behavior, at each level of the nested modeling. The net effect is an improvement of more than an order of magnitude in the length of the policies that are possible to compute, although the problem sizes remained limited to few physical states.

The A* search may be approximated by viewing the cooperative joint decision-making problem modeled by a Dec-POMDP as a sequence of Bayesian games, which are games generalized to condition an agent's payoff on its private information called its type (Harsanyi 1967). Here, the types of an agent are its different action-observation histories. In order to solve Bayesian games, we need a joint distribution over the types of all agents. This distribution, which is assumed to be common knowledge, is computed by solving the Bayesian games in the previous time steps to obtain the agents' actions. A Bayes-Nash equilibrium or the Pareto-optimal equilibrium that is also the optimal joint solution, could be computed given the payoffs. The sequential approach avoids backtracking because a single, optimal solution is obtained from the Bayesian games. As the general type space is exponential in the number of time steps, low probability types are pruned to reduce the type space, thereby approximating the solution (Emery-Montemerlo et al. 2004). Furthermore, the intractability of using the optimal expected rewards as payoffs motivates heuristic values, which are explored by Oliehoek, Spaan, and Vlassis (2008). Another algorithm also embracing game-theoretic solution concepts conducts a search for joint policies in Nash equilibrium (Nair et al. 2003) for a team of agents. The existence of multiple equilibria opens up the possibility that the decentralized agents in these techniques may choose policies in different equilibria, which diminishes the value of the joint behavior. This pitfall is not unique to equilibria-based solutions as multiple globally optimal joint solutions may also be present. It is avoided by solving the decision problem in advance, as is usual, and fixating on policies in a single equilibrium or optima, which are then distributed to the agents.

Approximate representations of the nested beliefs in finitely nested I-POMDPs make the interactive state space more tractable. One such representation utilizes samples, called particles, where each particle contains a physical state and the other agent's model (Doshi and Gmytrasiewicz 2009). Particles are obtained by Monte Carlo sampling of the state space based on the subject agent's belief. Other agent's belief in each particle is itself represented using a sampled set of particles resulting in a nested particle set. The agent's belief is updated approximately by projecting the entire nested particle set in time using a two-step approach of sampling the next physical state followed by using the observation likelihood, at each level of nesting. This technique takes an order of magnitude less time in solving I-POMDPs in comparison to the exact, with a controlled loss in optimality. Doshi (2007) showed how the nested particle set may be modified and propagated when the physical state space is continuous or very large.

Previous algorithms all focus on computing solutions when the optimality criterion requires maximizing the expected reward over a finite number of time steps. This allows the policies to manifest as trees of finite height. However, if the optimality criterion specifies infinite steps, trees are no longer feasible; instead the policies manifest as stochastic finite-state machines as I mentioned previously. We may progressively improve the joint controllers, one at a time and holding the other fixed, while taking care not to exceed a bound on the number of nodes, until no more improvement is possible. In cooperative contexts, Bernstein, Hansen, and Zilberstein (2005) show that a shared correlation device could also be helpful in improving the quality of the joint controller. Although such improvements of the joint controllers could get mired in local optima, the technique scales favorably to problems involving a large number of physical states and actions. Specifically, small controllers of reasonably good quality may be obtained for two-agent problem domains exhibiting up to a hundred physical states, five actions, and a similar number of observations. Iterative improvement of the joint controllers and its subsequent convergence to a local optimum could be avoided by reformulating the problem as a nonlinear program whose optimization has the potential to immediately generate the globally optimal controllers of given size (Amato, Bernstein, and Zilberstein 2010). However, practical nonlinear program solvers are unable to exclude local optima, though experiments reveal that the solution quality is high.

Developments outlined in this section provide a firm algorithmic foundation for both Dec- and I-POMDPs on which advances could take place. In table 1, I list techniques that have played a key role in this regard. Seuken and Zilberstein (2008) pro-

	Belief Space	Policy Space	Multiple Agents
Dynamic programming		✔	
Behavioral equivalence	✔		✔
Memory-boundedness		✔	
Point-based optimization		✔	
Sampling states and observations	✔	✔	
Locality of interaction			✔

Table 1. A Tabulation of Some Core Techniques and the Types of Complexity That Affect Both Frameworks.

Recall that the belief space complexity refers to the disproportionate growth in the space of beliefs as the physical states increase, the policy space complexity is due to the exponential growth in the number of policies that must be considered with time, both of which are exacerbated as the number of interacting agents increase. A ✔ indicates that the technique in that row helps in mitigating the type of complexity specified in the column.

vide more details on some of the approaches discussed here, and numerous collections of publications related to the two frameworks exist on the web, one of which is currently available at rbr.cs.umass.edu/~camato/decpomdp. Indeed, newer techniques, not discussed here, continue to push the scalability envelope for both frameworks. This is critically needed to facilitate more real-world applications.

Discussion

Automated decision making in multiagent contexts is a crucial research frontier in our quest for designing intelligent agents because of its wide applicability in problem solving. Although we may not see the Dec- or I-POMDP frameworks being a part of field trials of AUAVs anytime soon, steady progress is being made in making them scalable. These frameworks, along with their specializations that I discussed here, are model-based ways of computing optimal behavior. An equally important research direction investigates in parallel multiagent decision making without presupposing a model by relying on repeated problem simulations to acquire data; this direction generalizes single-agent reinforcement learning. Hybrid approaches that learn incompletely specified models also exist.

An important facet of multiagent contexts, which I did not discuss in this article, is communication between the agents. This is because the topic demands a detailed treatment in itself, partly due to its importance in cooperative decision making. In the context of Dec-POMDPs, relevant research has mostly focused on the potential of communication mitigating the high computational complexity of the decision-making process by reducing the interagent uncertainty (for a recent example, see Wu, Zilberstein, and Chen [2011]). Communicative extensions of Dec-POMDPs and other frameworks have been introduced, most of which model communication as a distinguished act. On the other side of the coin, an intentional agent modeled using finitely nested I-POMDP augmented with communicative acts exhibits a surprising rational behavior: it will choose not to communicate or to respond to any requests. At the heart of this observation is the simple insight that at the lowest level (level 0), an intentional agent is modeled as having no understanding of others in its environment. Consequently, such an agent does not view sending a message (to nobody in its perspective) beneficial, nor is it able to rationalize an observed request other than attributing it to noise. Reasoning that the other agent (at level 0) will not engage in communication, the level 1 agent does not see any benefit in communicating either. We may apply this argument upward in the nesting and thereby conclude that the (boundedly) rational level l agent may not engage in communication. This paradoxical result is primarily due to bottoming out the nesting — a consequence of computability. I do not think that this paradox is an artifact of the particular framework. However, it does motivate changes to how we model other agents and suggests that models that additionally incorporate learning are needed, as realized in the context of Dec-POMDPs as well (Spaan, Gordon, and Vlassis 2006).

Future Advances

While the frameworks have drawn insights from game theory, they may, in return, advance game theory as well. One way is by helping to explain the equilibration process in complex settings, which as Binmore (1982) pointed out, continues to remain a key gap in game theory. A focus on the procedural aspects of decision making by I-POMDPs allows us to understand optimal behavior, and the computational difficulty in satisfying sufficiency conditions for equilibration (Doshi and Gmytrasiewicz 2006) informs us as to why equilibrium is a perplexing concept.

As Seuken and Zilberstein (2008) note, reconciling the behaviors prescribed by Dec- and I-POMDP in cooperative contexts is worthwhile because each one represents a different approach to decision making in multiagent contexts. In particular, I-POMDPs provide a fertile ground for investigating the interaction between computability and intentionality, and its possibly negative implications on asymptotic agent behaviors (for example, see again Doshi and Gmytrasiewicz [2006] and references therein), in the backdrop of global optimality as prescribed by Dec-POMDPs.

Finally, as applications of decision-making frameworks emerge, an important one is in decision support systems that are part of mixed human-agent teams. Using models of the task at hand, the frameworks may provide normative recommendations to the human who may choose to act on them. Furthermore, the frameworks could additionally utilize descriptive models of human behavior on the task at hand and perform actions that support those of the human as a true teammate, though this application is more challenging.

Acknowledgments

This work was supported in part by NSF CAREER grant, #IIS-0845036, and in part by a grant from the U.S. Air Force, #FA9550-08-1-0429. All opinions expressed in this article are those of the author alone and do not reflect the views of the sponsoring agencies. The author would also like to acknowledge insightful discussions with Christopher Amato, Daniel Bernstein, Piotr Gmytrasiewicz, Matthijs Spaan,and Shlomo Zilberstein, and the feedback from the anonymous reviewers.

References

Amato, C.; Bernstein, D. S.; and Zilberstein, S. 2010. Optimizing Fixed-Size Stochastic Controllers for POMDPs and Decentralized POMDPs. *Journal of Autonomous Agents and Multi-Agent Systems* 21(3): 293–320.

Aumann, R. J. 1999. Interactive Epistemology I: Knowledge. *International Journal of Game Theory* 28(3): 263–300.

Becker, R.; Zilberstein, S.; Lesser, V.; and Goldman, C. 2004. Solving Ttransition Independent Decentralized Markov Decision Processes. *Journal of Artificial Intelligence Research* 22: 423–455.

Bernstein, D. S.; Givan, R.; Immerman, N.; and Zilberstein, S. 2002. The Complexity of Decentralized Control of Markov Decision Processes. *Mathematics of Operations Research* 27(4): 819–840.

Berstein, D. S.; Hansen, E. A.; and Zilberstein, S. 2005. Bounded Policy Iteration for Decentralized POMDPs. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence,* 1287–1292. San Francisco: Morgan Kaufmann Publishers.

Binmore, K. 1982. *Essays on Foundations of Game Theory.* Oxford, UK: Basil Blackwell.

Brandenburger, A. 2007. The Power of Paradox: Some Recent Developments in Interactive Epistemology. *International Journal of Game Theory* 35(4): 465–492.

Brandenburger, A., and Dekel, E. 1993. Hierarchies of Beliefs and Common Knowledge. *Journal of Economic Theory* 59(1): 189–198.

Cohen, P. R., and Levesque, H. J. 1991. Confirmations and Joint Action. In *Proceedings of the Twelfth International Joint conference on Artificial Intelligence,* 951–957. San Francisco: Morgan Kaufmann Publishers.

Dibangoye, J. S.; Mouaddib, A.-I.; and Chaib-draa, B. 2009. Point-Based Incremental Pruning Heuristic for Solving Finite-Horizon DEC-POMDPs. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems,* 569–576. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Doshi, P. 2007. Improved State Estimation in Multiagent Settings with Continuous or Large Discrete State Spaces. In *Proceedings of the Twenty Second Conference on Artificial Intelligence,* 712–717. Menlo Park, CA: AAAI Press.

Doshi, P., and Gmytrasiewicz, P. J. 2009. Monte Carlo Sampling Methods for Approximating Interactive POMDPs. *Journal of Artificial Intelligence Research* 34: 297–337.

Doshi, P., and Gmytrasiewicz, P. J. 2006. On the Difficulty of Achieving Equilibrium in Interactive POMDPs. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, 1131–1136. Menlo Park, CA: AAAI Press.

Doshi, P., and Perez, D. 2008. Generalized Point Based Value Iteration for Interactive POMDPs. In *Proceedings of the Twenty Third AAAI Conference on Artificial Intelligence,* 63–68. Menlo Park, CA: AAAI Press.

Doshi, P.; Qu, X.; Goodie, A.; and Young, D. 2010. Modeling Recursive Reasoning in Humans Using Empirically Informed Interactive POMDPs. In *Proceedings of the 9th International Autonomous Agents and Multiagent Systems Conference,* 1223–1230. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Emery-Montemerlo, R.; Gordon, G.; Schneider, J.; and Thrun, S. 2004. Approximate Solutions for Partially Observable Stochastic Games with Common Payoffs. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems,* 136–143. Los Alamitos, CA: IEEE Computer Society.

Fudenberg, D., and Tirole, J. 1991. *Game Theory.* Cambridge, MA: The MIT Press.

Gmytrasiewicz, P. J., and Doshi, P. 2005. A Framework for Sequential Planning in Multiagent Settings. *Journal of Artificial Intelligence Research* 24: 49–79.

Gmytrasiewicz, P. J., and Durfee, E. 2000. Rational Coordination in Multi-Agent Environments. *Journal of Autonomous Agents and Multiagent Systems* 3(4): 319–350.

Goldman, C. V., and Zilberstein, S. 2004. Decentralized Control of Cooperative Systems: Categorization and Complexity Analysis. *Journal of Artificial Intelligence Research* 22: 143–174.

Grosz, B. 1996. Collaborating Systems. *AI Magazine* 17(2): 67–85.

Hansen, E.; Bernstein, D.; and Zilberstein, S. 2004. Dynamic Programming for Partially Observable Stochastic Games. In *Proceedings of the Nineteenth Conference on Artificial Conference,* 709–715. Menlo Park, CA: AAAI Press.

Harsanyi, J. C. 1967. Games with Incomplete Information Played by Bayesian Players. *Management Science* 14(3): 159–182.

Kadane, J., and Larkey, P. 1982. Subjective Probability and the Theory of Games. *Management Science* 28(2): 113–120.

Kaelbling, L.; Littman, M.; and Cassandra, A. 1998. Planning and Acting in Partially Observable Stochastic Domains. *Artificial Intelligence* 101(1–2): 99–134.

Kim, Y.; Nair, R.; Varakantham, P.; Tambe, M.; and Yokoo, M. 2006. Exploiting Locality of Interaction in Networked Distributed

POMDPs. In *Distributed Plan and Schedule Management: Papers from the AAAI Spring Symposium*. Technical Report SS-06-04. Menlo Park, CA: AAAI Press.

Kurniawati, H.; Hsu, D.; and Lee, W. S. 2008. SARSOP: Efficient Point-Based POMDP Planning by Approximating Optimally Reachable Belief Spaces. Paper presented at 2008 Robotics: Science and Systems, Zurich, Switzerland, 25–27 June.

Madani, O.; Hanks, S.; and Condon, A. 2003. On the Undecidability of Probabilistic Planning and Related Stochastic Optimization Problems. *Artificial Intelligence* 147(1–2): 5–34.

Meissner, C. 2011. A Complex Game of Cat and Mouse. *Lawrence Livermore National Laboratory Science and Technology Review* (March): 18–21.

Mertens, J., and Zamir, S. 1985. Formulation of Bayesian Analysis for Games with Incomplete Information. *International Journal of Game Theory* 14(1): 1–29.

Nair, R.; Tambe, M.; Yokoo, M.; Pynadath, D.; and Marsella, S. 2003. Taming Decentralized POMDPs: Towards Efficient Policy Computation for Multiagent Settings. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence,* 705–711. San Francisco: Morgan Kaufmann Publishers.

Nair, R.; Varakantham, P.; Tambe, M.; and Yokoo, M. 2005. Networked Distributed POMDPs: A Synthesis of Distributed Constraint Optimization and POMDPs. In *Proceedings of the Twentieth AAAI Conference on Artificial Intelligence,* 133–139. Menlo Park, CA: AAAI Press.

Ng, B.; Meyers, C.; Boakye, K.; and Nitao, J. 2010. Towards Applying Interactive POMDPs to Real-World Adversary Modeling. In *Proceedings of the Twenty-Second Conference on Innovative Applications in Artificial Intelligence*, 1814–1820. Menlo Park, CA: AAAI Press.

Oliehoek, F.; Spaan, M.; Whiteson, S.; and Vlassis, N. 2008. Exploiting Locality of Interaction in Factored Dec-POMDPs. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems,* 517–524. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Oliehoek, F.; Spaan, M.; and Vlassis, N. 2008. Optimal and Approximate Q-Value Functions for Decentralized POMDPs. *Journal of Artificial Intelligence Research* 32: 289–353.

Papadimitriou, C., and Tsitsiklis, J. 1987. The Complexity of Markov Decision Processes. *Mathematical Journal of Operations Research* 12(3): 441–450.

Pineau, J.; Gordon, G.; and Thrun, S. 2006. Anytime Point-Based Value Iteration for Large POMDPs. *Journal of Artificial Intelligence Research* 27: 335–380.

Pynadath, D., and Marsella, S. 2007. Minimal Mental Models. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence,* 1038–1044. Menlo Park, CA: AAAI Press.

Pynadath, D., and Tambe, M. 2002. The Communicative Multiagent Team Decision Problem: Analyzing Teamwork Theories and Models. *Journal of Artificial Intelligence Research* 16: 389–426.

Rabinovich, Z.; Goldman, C. V.; and Rosenschein, J. S. 2002. Non-Approximability of Decentralized Control. Technical Report 2002-29, Leibniz Center for Computer Science, Hebrew University, Jerusalem, Israel.

Rathnasabapathy, B.; Doshi, P.; and Gmytrasiewicz, P. J. 2006. Exact Solutions to Interactive POMDPs Using Behavioral Equivalence. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-Agent Systems,* 1025–1032. New York: Association for Computing Machinery.

Russell, S., and Norvig, P. 2010. *Artificial Intelligence: A Modern Approach.* Engelwood Cliffs, NJ: Prentice Hall.

Seuken, S., and Zilberstein, S. 2007. Improved Memory Bounded Dynamic Programming for Decentralized POMDPs. Paper presented at the 23rd Conference in Uncertainty in Artificial Intelligence. Vancouver, BC, Canada, 19–22 July.

Seuken, S., and Zilberstein, S. 2008. Formal Models and Algorithms for Decentralized Decision Making Under Uncertainty. *Journal of Autonomous Agents and Multiagent Systems* 17(2): 190–250.

Seymour, R., and Peterson, G. L. 2009a. Responding to Sneaky Agents in Multi-Agent Domains. In *Proceedings of the 22nd Florida Artificial Intelligence Research Society Conference (FLAIRS),* 99–104. Menlo Park, CA: AAAI Press.

Seymour, R., and Peterson, G. L. 2009b. A Trust-Based Multiagent System. In *Proceedings of the 2009 IEEE International Conference on Computational Science and Engineering,* 109–116. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Smallwood, R., and Sondik, E. 1973. The Optimal Control of Partially Observable Markov Decision Processes over a Finite Horizon. *Operations Research* 21(5): 1071–1088.

Spaan, M.; Gordon, G.; and Vlassis, N. 2006. Decentralized Planning Under Uncertainty for Teams of Communicating Agents. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-Agent Systems,* 249–256. New York: Association for Computing Machinery.

Szer, D.; Charpillet, F.; and Zilberstein, S. 2005. MAA*: A Heuristic Search Algorithm for Solving Decentralized POMDPs. Paper presented at the 21st Conference in Uncertainty in Artificial Intelligence. Edinburgh, Scotland, 26–29 July.

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

von Neumann, J., and Morgenstern, O. 1947. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Wu, F.; Zilberstein, S.; and Chen, X. 2011. Online Planning for Multi-Agent Systems with Bounded Communication. *Artificial Intelligence Journal* 175(2): 487–511.

Wunder, M.; Kaisers, M.; Yaros, J.; and Littman, M. 2011. Using Iterated Reasoning to Predict Opponent Strategies. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems,* 593–600. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Prashant Doshi is an associate professor of computer science and the director of the THINC lab at the University of Georgia, USA. His research interests include decision making in single and multiagent contexts with a focus on POMDP-based frameworks and graphical models. During the past five years, he has been coconducting a highly successful tutorial on decision making in multiagent settings in the AAMAS conferences. He is also the co-organizer of a series of workshops on multiagent sequential decision making and interactive decision theory and game theory. Additionally, he regularly teaches a graduate-level course on decision making under uncertainty in his department, which is well regarded by the students. More details on his research and teaching are available at thinc.cs.uga.edu.