Towards a Unifying Characterization for Quantifying Weak Coupling in Dec-POMDPs

Stefan J. Witwicki and Edmund H. Durfee
Computer Science and Engineering
University of Michigan
Ann Arbor, MI 48109
{witwicki,durfee}@umich.edu

ABSTRACT

Researchers in the field of multiagent sequential decision making have commonly used the terms "weakly-coupled" and "loosely-coupled" to qualitatively classify problems involving agents whose interactions are limited, and to identify various structural restrictions that yield computational advantages to decomposing agents' centralized planning and reasoning into largely-decentralized planning and reasoning. Together, these restrictions make up a heterogeneous collection of facets of "weakly-coupled" structure that are conceptually related, but whose purported computational benefits are hard to compare evenhandedly. The contribution of this paper is a unified characterization of weak coupling that brings together three complementary aspects of agent interaction structure. By considering these aspects in combination, we derive new bounds on the computational complexity of optimal Dec-POMDP planning, that together quantify the relative benefits of exploiting different forms of interaction structure. Further, we demonstrate how our characterizations can be used to explain why existing classes of decoupled solution algorithms perform well on some problems but poorly on others, as well as to predict the performance of a particular algorithm from identifiable problem attributes.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*

General Terms

Theory, Performance

Keywords

Multiagent Planning, Coordination, Weak Coupling, Loose Coupling, Locality of Interaction, Policy Abstraction, Influence, Decentralized Markov Decision Processes, POMDPs

1. INTRODUCTION

The Decentralized Partially-Observable Markov Decision Process (Dec-POMDP) has emerged as a popular theoretical model for planning coordinated decisions for teams of agents under uncertainty, but its well-established general NEXP hardness (reviewed by Goldman [6]) has raised concerns about its practical applicability beyond small toy problems. In response, researchers have defined a variety of subclasses amenable to efficient, scalable solution methods, but that impose various constraints on problem structure [2, 11, 15]. In this paper, we endeavor to illuminate significant aspects of Dec-POMDP interaction structure that make one problem easier than another, and to quantify the computational advantages of exploiting these aspects in concert.

Authors in the multiagent sequential decision making literature commonly use the terms "weakly-coupled" or "loosely-coupled" to classify problems involving agents whose interactions are limited (e.g., [8, 12, 13, 22]). Intuitively, weakly-coupled interaction structure engenders conditional independencies among individual agents' decisions that allow for efficient decomposition of joint planning and reasoning. However, different authors' uses of these terms refer to slightly different structural conditions, and often frame the consequences of the structure in slightly different algorithmic contexts. Given the heterogeneity of structural conditions for weak coupling, and the diverse contexts of published results, it is difficult to ascertain the computational advantages of the various structures in relation to one another.

Here, we generalize and synthesize several elements of problem structure into a more unified characterization of weak coupling. In particular, we highlight three complementary aspects of weakly-coupled problem structure: *agent scope size*, *state factor scope domain size*, and *degree of influence*. Not only does our characterization highlight useful relationships between these three aspects, but it also concretely quantifies the relative computational benefits of exploiting each. By considering these three aspects in concert, we derive new bounds on the worst-case complexity of optimal Dec-POMDP planning (the context of which we describe in Section 2). After presenting our characterization and theoretical results in Section 3, we illustrate the usefulness of this contribution in Section 4 by demonstrating that our theory helps (1) to better explain trends observed in past work, and (2) to predict the performance of solution algorithms based on the degree to which test problems are weakly coupled. As a case study, we illustrate how our theoretical results can be used in conjunction with empirical analysis to extrapolate the relative performance of a particular algorithm, Optimal Influence-space Search [22], on 4-agent problems. In Section 5, we relate our characterization to foundational and alternative analyses from past work, and conclude with a discussion of our results and future work in Section 6.

Cite as: Towards a Unifying Characterization for Quantifying Weak Coupling in Dec-POMDPs, S.J. Witwicki and E.H. Durfee, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 29-36.

2. CONTEXT

We are considering the problem of planning optimal policies for a group of agents whose behavior is modeled as a Dec-POMDP. To illustrate the interaction structure that we are characterizing, we refer to several examples expressed using a particular Dec-POMDP model called a Transition-Decoupled POMDP (TD-POMDP) [22], whose specification emphasizes conditional independencies among agents.

Figure 1a shows a simple problem involving a three-agent team whose objective is to plan the coordinated executions of *tasks*. Here, task interdependencies called *enablements* each represent the constraint that one task must have completed with positive outcome quality before another task can begin. The execution of each task, by the agent that owns that task, is also constrained by a *window* that expresses the task's earliest start time and latest finish time (after which the task will "fail" by achieving zero outcome quality). The uncertainty of each task's execution is conveyed by its *outcome distribution*, which assigns a probability (P) to each possible duration (D) and quality (Q). In the past, Dec-POMDP researchers have studied examples of this flavor that were geared towards application domains such as Mars rover control [13, 22] and disaster response [11].

2.1 Decentralized POMDP

A (finite-horizon) Dec-POMDP [1, 6, 16, 22] is specified with a tuple $\mathcal{M} = \langle \mathcal{N}, S, A, \Omega, O, R, P, T \rangle$, where $\mathcal{N}$ is a set of n agents, S is a finite set of world states, with a distinguished initial state, and $A = \times_{i \in \mathcal{N}} A_i$ is the joint action space, each component of which refers to the set of actions of an agent $i \in \mathcal{N}$. The transition function $P(s'|s,a)$ specifies the probability distribution over next states given that joint action $a = \langle a_1, a_2, \ldots, a_n \rangle \in A$ is taken in state $s \in S$. The reward function $R(s,a,s')$ specifies the immediate value of taking joint action a in state s and arriving in state s'. Observation function $O(o'|a,s')$ specifies the probability of the joint observation $o \in (\Omega = \times_{i \in \mathcal{N}} \Omega_i)$, observed upon taking a and transitioning into s'.

The team's behavior is specified with a *joint policy* $\pi = \langle \pi_1, \ldots, \pi_n \rangle$, where each component π_i (agent i's *local policy*) maps agent i's observation history $\vec{o}_i^{\,t}$ to an action a_i, thereby encoding a deterministic decision rule for any sequence of observations that each agent might encounter. The set of possible joint policies Π denotes the *policy space*, and Π_i agent i's local policy space. The *value* $V(\pi)$ of a joint policy $\pi \in \Pi$ is the expected cumulative reward received by the team (from times 1 to T) when executing that policy from the initial state. Finally, Dec-POMDP planning is the problem of computing the *optimal joint policy* π^*, which can be expressed as a maximization: $\pi^* = \arg\max_{\pi \in \Pi} V(\pi)$.

2.2 Transition-Decoupled POMDP

The Dec-POMDP is an extremely general representation of joint behavior, allowing for arbitrary dependencies among agents' observations and action consequences. As such, the naïve Dec-POMDP specification is oblivious to any interaction structure that may exist among agents. To express some of the structure (which we find useful in Section 3) that is present in problems like the one depicted in Figure 1a, we turn to the TD-POMDP model [22].

The TD-POMDP assumes an inherent decomposition of the joint model into transition-dependent local agent models. The world state s is *factored* into individual state features that are distributed among agents' *local states* $\langle s_1, ..., s_n \rangle \in S = \times_{i \in \mathcal{N}} S_i$; however, each feature does not necessarily reside exclusively in a single agent's local state s_i. Figure 1b depicts the decomposition of world state for the example problem as a two-stage dynamic Bayesian network, wherein feature "D*en*" (encoding whether or not Task D is enabled) is shared among agent 1's and agent 2's local states, and *time* is shared among all agents' local states. These features are referred to as *mutually-modeled features*, comprising set $\bar{m}$.

Similarly, the TD-POMDP specifies decomposable observation functions and decomposable local rewards [22]. In this paper, we consider TD-POMDP problems for which the joint value function decomposes into local value functions $V(\pi) = \sum_{i \in \mathcal{N}} V_i(\pi)$, each of which, $V_i(\pi)$, is equal to one agent i's expected cumulative *local* reward. In the example problem, local rewards account for the quality accrued whenever an agent finishes one of its tasks. The TD-POMDP explicitly distinguishes features in an agent j's local state that are controlled by agent i as a *nonlocal features* $\bar{n}_j$, wherein each serves as an attribute through which i can affect j's local state, observations, and subsequent local transitions [22]. For instance, in Figure 1b, when agent 1 completes task C, nonlocal feature F*en* (F*enabled*) changes from $false$ to $true$, thereby altering how agent 3's actions can affect feature F.

A TD-POMDP's transition-dependent interactions may be illustrated graphically using an *agent interaction digraph* [22], examples of which are shown in Figure 2. The interaction digraph contains a vertex for each agent, and an edge for each nonlocal feature that connects the controlling agent with the affected agent. For any two agents i and j, there may be more than one edge leading from i to j, one for each nonlocal feature controlled by i and affecting j. For the purpose of our analysis we shall denote the *digraph ancestors* of an agent j as $\Lambda_j = \{i \neq j |$ there is a directed path from i to $j\}$, and the set of *digraph descendants* of agent i as $\Psi_i = \{j \neq i |$ there is a directed path from i to $j\}$. In contrast, we shall use the word *peer* to refer to "some other agent" in $\mathcal{N}$ without the implication of any particular graphical relationship.

2.3 Relationship Between Dec-POMDP Planning and Constraint Optimization

Our analysis in Section 3 makes use of a reformulation of the Dec-POMDP planning problem into a *constraint optimization problem* (COP). The reformulation is a slight generalization of that explored in past work [2, 15]. In review, a classical COP [4] is specified as a tuple $\mathcal{C} = \langle X, D, C \rangle$, where $X = \{x_1, ..., x_n\}$ is a set of n *variables* with possible assignments $\bar{a} = \langle a_1, ..., a_n \rangle \in D = \{D_1, ..., D_n\}$, and C is a set of *constraints*. Each constraint represents a cost function C_k with a restricted (variable) scope $Q_k \subseteq \{1, ..., n\}$, such that $C_k : [\times_{i \in Q_k} D_i] \mapsto \{\mathbb{R}, \infty\}$. The restricted scopes of COP constraints constitute graphical structure that is naturally expressed using a *constraint graph* $\mathcal{G}$. Illustrated in Figure 2, there is a hyperedge for each constraint C_k that connects those vertices (which we refer to as *neighbors*) corresponding to the variables indexed by Q_k.

In solving a COP, and obtaining solution $\bar{a}^*$, the objective is to minimize the summation of cost values of the variable assignments: $\bar{a}^* = \arg\min_{\bar{a}} \sum_{k=1}^{\|C\|} C_k(\bar{a})$. Analogously, the objective of Dec-POMDP planning is to maximize the expected utility of the joint policy, which can often be decomposed into component value functions [11, 15, 16, 22], one for each agent in the case of the TD-POMDP.

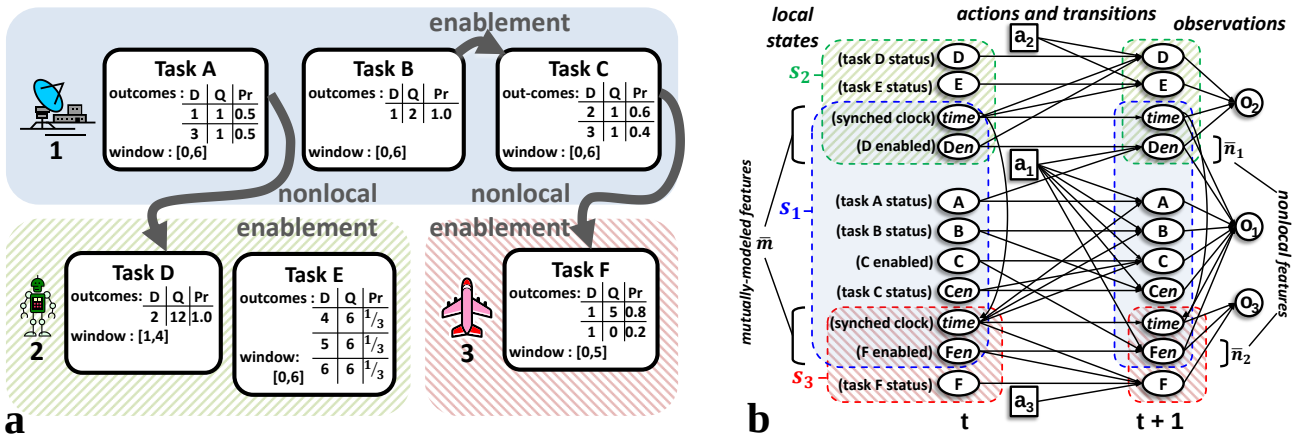

Figure 1: An example problem (a) and corresponding TD-POMDP specification (b).

OBSERVATION 1. *A Dec-POMDP $\mathcal{M}$, whose value function $V(\pi = \langle \pi_1, ..., \pi_n \rangle)$ decomposes into a summation of component value functions $V(\pi) = \sum_k V_k(\pi)$, reduces to a COP $\mathcal{C}_\mathcal{M} = \langle X, D, C \rangle$ structured as follows:*

- *X contains exactly one variable x_i for each agent i,*
- *the domain of x_i is agent i's local policy space: $D_i \equiv \Pi_i$,*
- *an assignment $\bar{a} = \langle \pi_i, ..., \pi_n \rangle$ is a joint policy,*
- *C consists of a single constraint $C_k(\bar{a} \equiv \pi) = -V_k(\pi)$ for each component V_k of $\mathcal{M}$'s value function,*
- *and the solution $\bar{a}^*$ is an optimal joint policy π^*.*

This reformulation gives rise to a decoupled joint policy search methodology (such as that employed in past work [2, 14, 22]). In contrast to centralized methods that optimize all components of the optimal joint policy at once, a decoupled joint policy search is an iterative process where, at each step, an each agent i computes one possible local policy, $\pi_i^*(\bar{\pi}_{\neq i}) = \arg\max_{\pi_i} V(\pi_i, \bar{\pi}_{\neq i})$, referred to as a **best response** to proposed local policies $\bar{\pi}_{\neq i}$ of i's peers. Equation 1 below reduces the computation of an optimal joint policy for a three agent Dec-POMDP problem (such as in Figure 1) to a series of best response calculations.

$$\begin{aligned} \pi^* &= \arg\max_{\langle \pi_1, \pi_2, \pi_3 \rangle} V(\pi_1, \pi_2, \pi_3) \\ &= \arg\max_{\langle \pi_1, \pi_2 \rangle} V(\pi_1, \pi_2, \pi_3^*(\pi_1, \pi_2)) \\ &= \arg\max_{\pi_1} V(\pi_1, \pi_2^*(\pi_1), \pi_3^*(\pi_1, \pi_2^*(\pi_1))) \\ &(\text{where } \pi_2^*(\pi_1) = \arg\max_{\pi_2} V(\pi_1, \pi_2, \pi_3^*(\pi_1, \pi_2))) \end{aligned} \quad (1)$$

The search implied by the arg max invocations in Eq. 1, which enumerates all combinations of local policies, serves as the basis for more advanced decoupled solution methods cited in the next section.

3. DIMENSIONS OF WEAKLY COUPLING

Intuitively, the computational benefit of solving a problem with a decoupled solution method instead of a centralized method depends upon the presence of problem structure that renders agents more or less independent of each other. Here, we generalize and formally characterize three different previously-studied aspects of weakly-coupled problem structure whose exploitation has been shown to be beneficial in past work (which we review in Section 5). Our characterization takes the form of a three-dimensional landscape that can be used to quantify the advantage gained through exploiting a problem's interaction structure and, ultimately, to predict the amount of computation needed to solve the problem. We describe each dimension in Sections 3.1–3.3, over the course of which we gradually refine a bound on the computational complexity of Dec-POMDP planning, and then in Section 3.4 we bring these terms together into a unified characterization.

3.1 Agent Scope Size

The first aspect that we examine lies in the graphical structure present in a Dec-POMDP $\mathcal{M}$'s equivalent COP constraint graph $\mathcal{G}_\mathcal{M}$. The connectivity of each hyperedge is dictated by the scope Q_k of a constraint C_k, which is equal to the *agent scope* [7, 16] of component value function V_k.

Definition 1. The **agent scope**, denoted Q_k, of a (component) value function $V_k()$ is the subset of agents on whose policies its value depends, such that $V_k : [\times_{i \in Q_k} \Pi_i] \mapsto \mathbb{R}$.

In a TD-POMDP agent, for instance, the scope Q_i of an agent i's local value function contains all agents that can affect i's rewards through their actions, which are i and its interaction digraph ancestors: $Q_i = \{i\} \cup \Lambda_i$ [21]. In a Network-Distributed POMDP (ND-POMDP), the size of the agent scope corresponds to *local neighborhood size* [15].

At one extreme of the weak coupling spectrum, agents are *uncoupled*: they do not interact, so the constraint graph consists of n unconnected vertices. In this case, the optimal joint policy is simply the combination of independently-computed optimal local policies: $\pi^* = \langle \arg\max_{\pi_i} V_i(\pi_i), \forall i \in \mathcal{N} \rangle$, and agent scope $Q_i = \{i\}, \forall i$. At the opposite extreme, all agents' decisions are affected by all other agents, and hence no agent can optimize its local policy without considering the potential policies of all other agents. In between these two extremes, there exist conditional independencies that allow agents to plan independently of some peers but not others.

Definition 2. An agent i is **conditionally decision-independent** of agent j conditioned on peer agents $K \subseteq (\mathcal{N} - \{i, j\})$ if: $\forall \{\pi_j^x, \pi_j^y\} \subseteq \Pi_j, \forall \bar{\pi}_K \in (\times_{k \in K} \Pi_k)$,
$\arg\max_{\pi_i \in \Pi_i} V(\pi_i, \pi_j^x, \bar{\pi}_K) = \arg\max_{\pi_i \in \Pi_i} V(\pi_i, \pi_j^y, \bar{\pi}_K)$.

In the example from Figure 1, agents 2 and 3 are *conditionally decision-independent* of each other conditioned on

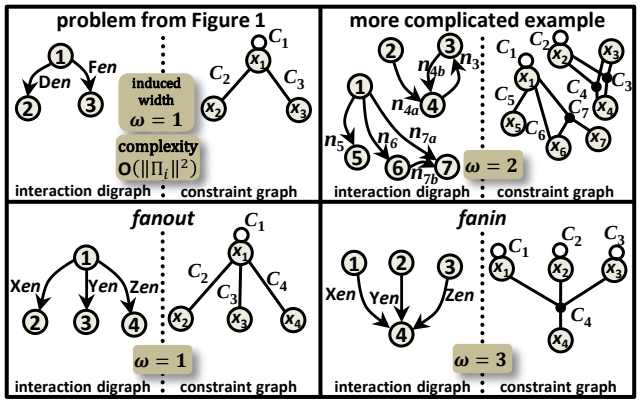

Figure 2: Examples of constraint graphs (right) derived from TD-POMDP interaction digraphs (left).

agent 1. Intuitively, this is because agent 2 cannot affect the values of agent 3's local state features nor *vice versa*, no matter what actions they decide to take. Mathematically, it is because, from the constraint graph in Figure 2 (upper-left), agent 3's best response $\pi_3^*(\pi_1, \pi_2)$ is independent of π_2:

$$\begin{aligned}\pi_3^*(\pi_1,\pi_2) &= \arg\max_{\pi_3} V(\pi_1,\pi_2,\pi_3)\\ &= \arg\max_{\pi_3} [V_1(\pi_1)+V_2(\pi_1,\pi_2)+V_3(\pi_1,\pi_3)]\\ &= \arg\max_{\pi_3} V_3(\pi_1,\pi_3) \equiv \pi_3^*(\pi_1)\end{aligned}$$

and *vice versa*. The practical benefit of this *conditional decision independence* relationship is a simpler optimal solution computation. Substituting $\pi_3^*(\pi_1)$ (as well as the equivalently-reduced $\pi_2^*(\pi_1)$) into Equation 1 significantly reduces the combinations of local policies that need be considered:

$$\pi^* = \arg\max_{\pi_1} V(\pi_1, \pi_2^*(\pi_1), \pi_3^*(\pi_1)) \quad (2)$$

Whereas Equation 1 required $\|\Pi_1\|\|\Pi_2\|\|\Pi_3\|$ evaluations of V(), Equation 2 requires only $(\|\Pi_1\|\|\Pi_2\| + \|\Pi_1\|\|\Pi_3\| + \|\Pi_1\|)$ evaluations due to the reduction in the number of best responses, given the restricted *agent scopes* of $V_2()$ and $V_3()$.

Using COP theory, we can generalize the computational reduction as well as the methodology of exploiting graphical structure beyond our simple example problem. The computation performed in Equation 2 is an instance of *bucket elimination* (essentially, nonserial dynamic programming) [4], a general solution methodology that performs well on sparsely-connected constraint graphs. Dechter has proven that the worst-case time and space complexity of bucket elimination is $O\left(\|C\|\|D_i^{max}\|^{\omega+1}\right)$, where $\|C\|$ is the number of cost functions, D_i^{max} is the largest domain size of any COP variable and ω is the *induced width* of the constraint graph [4].

Observation 2. *The worst-case time and space complexity of optimal planning for Dec-POMDP $\mathcal{M}$ is bounded by $O\left(n \cdot \|\Pi_i^{max}\|^{\omega+1}\right)$, where n is the number of component value functions, Π_i^{max} is the largest local policy space, and ω is the induced width of the equivalent constraint graph $\mathcal{G}_{\mathcal{M}}$.*

By Observation 2, a lower induced width implies an exponential reduction in worst-case computation time. However, note that the local policy space size $\|\Pi_i^{max}\|$, at the base of the exponent, is itself exponentially dependent on the number of local observations histories: $\|\Pi_i^{max}\| = O\left(\|A_i\|^{\|O_i\|^T}\right)$ [14].

By Dechter's definition [4], the induced width ω may be calculated by taking the minimum, over all possible orderings of vertices in $\mathcal{G}_{\mathcal{M}}$, of the following measure: process vertices in order from last to first, for each vertex connecting its earlier-ordered neighbors, then return the largest number of earlier-ordered neighbors of any vertex. Alternatively, we can estimate ω using *agent scope size*. While in general, $\omega \geq (\max_k \|Q_k\| - 1)$ (which follows from the definitions of ω and Q_k), for a wide variety of TD-POMDP interaction digraph topologies (some of which are shown in Figure 2), $\omega = (\max_k \|Q_k\| - 1)$.

3.2 State Factor Scope Domain Size

The theoretical results presented thus far assume a naïve algorithm for performing best response calculations: enumeration of all local policies $\pi_i \in \Pi_i$ and explicit evaluation of $V(\pi_i, \bar{\pi}_{\neq i})$ for each. In a classical COP, enumeration of variable domains would be the only way to compute a best response. However, the COP that we are solving involves policy variables with structured domains. To exploit this structure, several algorithms have been developed for computing a best response by solving a single-agent POMDP model seeded with peers' policy information [14, 15, 22]. Aside from harnessing the efficiencies of state-of-the-art POMDP solvers, a best-response model can also exploit weakly-coupled problem structure. In particular, a best-response model does not necessarily need to represent all world state features [15, 22].

Intuitively, there may be features that have no bearing on the value ascribed to the agent's own behavior. For instance, in Figure 1b, the enabling of Task F (encoded by feature F*en*, appearing in agent 1's local state, but unobservable to agent 2) is inconsequential to agent 2 as it plans its best response policy $\pi_2^*(\pi_1)$. Using Definition 3, which we have adapted from previous work [7, 16] to fit this context, feature F*en* is not in agent 2's state factor scope.

Definition 3. An agent i's **state factor scope** $\mathcal{X}_i$ is the minimal[1] set of features sufficient for modeling the (belief) state used to compute i's optimal best response.

Becker *et al.* have derived that a Transition-Independent Dec-MDP (TI-Dec-MDP) agent i's best response may be calculated with an *augmented MDP* whose state space includes *only* i's local state features (but whose rewards are modified to account for peer agent j's proposed policy) [2]. In this case, even though the joint utility is dependent upon features from both agents' local state representations, it suffices for agent i to reason over a greatly-reduced space of features when computing a best response. In earlier work [22], we have developed a POMDP for computing the best response for a TD-POMDP agent i that includes (at most) the features from i's local state s_i and the histories of mutually-modeled features $\vec{m}_i$. A TD-POMDP agent i's state factor scope is thus $\mathcal{X}_i \subseteq \{s_i, \vec{m}_i\}$.

Intuitively, the smaller the portion of the world state that an agent observes and interacts with, the smaller its state factor scope, and the easier its local planning and reasoning becomes. Accounting for the sizes of the domains of features in the state factor scope, denoted $Dom(\mathcal{X}_i)$, we can refine our bound on computational complexity as follows.

[1]Given that multiple flavors of best response model may be applicable [14, 22], we are most interested in those that exploit weakly-coupled problem structure by reducing their modeled set of features as much as possible.

THEOREM 1. *The worst-case complexity of Dec-POMDP planning is* $O\left(n \cdot \text{EXP}(\|Dom(\mathcal{X}_i^{max})\|) \cdot \|\Pi_i^{max}\|^{\omega}\right)$, *where* $\|Dom(\mathcal{X}_i^{max})\| = \max_{i \in \mathcal{N}} \|Dom(\mathcal{X}_i)\|$.

PROOF SKETCH. The derivation of the complexity result from Observation 2 entails every best response computation requiring an $\arg\max_{\pi_i}$ to be taken, enumerating the local policy space bounded by $\|\Pi_i^{max}\|$, for all combinations of policies of ω peers, yielding complexity $\|\Pi_i^{max}\|\|\Pi_i^{max}\|^{\omega}$ for each of the n agents. By replacing each best response calculation with one POMDP solution, we can substitute the first term $\|\Pi_i^{max}\|$ in our complexity computation with the complexity of solving a finite-horizon POMDP, which is $O(\text{EXP}(\|S\|) = \text{EXP}(\|Dom(\mathcal{X}_i^{max})\|)$ given that the state space is bounded by $\|Dom(\mathcal{X}_i^{max})\|$ [21]. □

3.3 Degree Of Influence

Making use of Definition 2, for any two agents i and j, i is either decision-dependent on j or decision-independent of j (possibly conditioned on some other agents). Considering the rich space of dependencies that may exist between the two agents, a binary relation such as *decision-independent* lacks the precision to characterize weakly-coupled problems satisfactorily. For instance, in the example problem from Figure 1a, agent 2 is decision-dependent on agent 1, but only dependent on those decisions relating to the execution of Task A. Whether agent 1 executes Task B after completing Task A, or simply idles, cannot impact agents 2's decisions in any way. Moreover, any two of agent 1's possible policies, π_1^x and π_1^y that differ only in the decisions made after completing Task A induce the same best response from agent 2.

Definition 4. Two policies, π_i^x and π_i^y, of agent i are **impact-equivalent**, denoted $\pi_i^x \overset{I}{\equiv} \pi_i^y$, if π_i^x and π_i^y result in the same peer best responses: $\forall j \neq i, \forall \bar{\pi}_{(K=\mathcal{N}-\{i,j\})}$, $\pi_i^x \overset{I}{\equiv} \pi_i^y \Leftrightarrow \left[\arg\max_{\pi_j} V(\pi_i^x, \pi_j, \bar{\pi}_K) = \arg\max_{\pi_j} V(\pi_i^y, \pi_j, \bar{\pi}_K)\right]$.

Definition 5. An **impact equivalence class** $E_{i,x}$ is a set of impact-equivalent policies: $\forall \{\pi_i^x, \pi_i^y\} \in E_{i,x}, \pi_i^x \overset{I}{\equiv} \pi_i^y$.

In essence, an agent i's local policy space can be partitioned into disjoint equivalence classes, each of which may impact other agents in the system in a different way, thereby (potentially) inducing a different combination of best responses from i's peers. Definitions 4–5 elicit a spectrum of varying degrees of agent dependence. At one end of the spectrum, all of agent i's policies are grouped into a single impact equivalence class, indicating that any given peer j's behavior is unaffected by i's decisions. At the opposite end of the spectrum, agent j's best response is highly sensitive to the policy that i adopts, such that no two policies of agent i are impact-equivalent, and the number of i's impact equivalence classes is equal to the size of its policy space $\|\Pi_i\|$.

There are several Dec-POMDP planning algorithms that take advantage of this kind of weak agent coupling [2, 22]. Each algorithm employs what we shall call a *partitioning scheme* that implicitly partitions each agent i's local policy space into a set of impact equivalence classes $\mathbb{P}_i = \{E_{ix}\}$, parameterizing each class with information representative of the policies from that class. The key is that to compute the optimal joint policy, it suffices for each agent j to compute a best response to just one of the local policies from each of agent i's $\|\mathbb{P}_i\|$ impact equivalence classes. This set of best responses is referred to as agent j's *optimal coverage set* [2] because it sufficiently covers every possible policy of agent i.

Definition 6. For a given problem, the **degree of influence** $d_{\mathbb{P}}$, afforded by a partitioning $\mathbb{P}$, is the maximum ratio of impact equivalence classes to local policies:

$$d_{\mathbb{P}} = \max_{i \in \mathcal{N}} \frac{\|\mathbb{P}_i\|}{\|\Pi_i\|}. \quad (3)$$

By Definition 6, if a very coarse partitioning is found for a particular problem, wherein partitions contain large numbers of local policies, a low *degree of influence* $d_{\mathbb{P}}$ has been achieved. All else being equal, problems with a low degree of influence should be easier to solve than problems with a high degree of influence because of the reduction, from $\|\Pi_i\|$ to $\|\mathbb{P}_i\|$, in the number of necessary best responses per step of a distributed joint policy search. However, the computational benefit of a coarse partitioning could be offset by the computational overhead required to partition each agent's local policy space, whose worst case we denote $C_{\mathbb{P}}$.

3.4 Unified Characterization

In the past three subsections, we have quantified three problem characteristics associated with the degree of coupling. Conceptually, *agent scope size* refers to the number of agents in the system that are affecting each others' decisions, *state factor scope domain size* refers to the portion of state feature values that must be considered by each individual agent when coordinating its decisions, and *degree of influence* refers to the proportion of unique ways that agents can impact each others' decisions (subject to a given partitioning scheme). Each aspect manifests itself in a different set of problem attributes, and each affects the overall complexity in a different manner. However, we can formally characterize the combination of their effects as follows:

THEOREM 2. *The worst-case time and space complexity of Dec-POMDP planning, using a decoupled solution method that partitions agents' policy spaces, is bounded by:*

$$O\left(n \cdot \text{EXP}(\|Dom(\mathcal{X}_i^{max})\|) \cdot (d_{\mathbb{P}}\|\Pi_i^{max}\|)^{\omega} + n \cdot C_{\mathbb{P}}\right) \quad (4)$$

where n denotes the number of component value functions, $Dom(\mathcal{X}_i^{max})$ denotes the largest domain of any agent's state factor scope (Def. 3), $d_{\mathbb{P}}$ denotes the degree of influence (Def. 6) given partitioning $\mathbb{P}$, $C_{\mathbb{P}}$ is the worst-case complexity of computing $\mathbb{P}$, Π_i^{max} denotes the largest local policy space, and ω is the induced width.

PROOF SKETCH. Equation 4 is straightforwardly derived by manipulating the bound from Theorem 1. The base of the exponent is replaced with $d_{\mathbb{P}}$ due to the worst-case reduction in the number of best response computations afforded by $\mathbb{P}$. Next, a second term is added accounting for the accumulation of computation required by the partitioning process. □

Parameters $\{\omega, \mathcal{X}_i^{\max}, d_{\mathbb{P}}\}$ can be thought of as separate dimensions whose combination provides a concrete measure of the weakness (or strength) of coupling of a problem. A problem's worst-case complexity depends on where it lies along the spectra of *agent scope size*, *state factor scope domain size*, and *degree of influence*. For any two problems, we can now compare their worst case complexities by estimating the values of the three parameters and positioning each in the

3-dimensional space. Given a factored representation of Dec-POMDP problem structure (such as a TD-POMDP [22], ND-POMDP [15], or a more general factored Dec-POMDP [16]), the first two parameters, ω and $\mathcal{X}_i^{\max}$, can be evaluated directly from the problem specification. The third dimension, $d_{\mathbb{P}}$, is not readily assessable from any Dec-POMDP problem specification that we are aware of. Moreover, $d_{\mathbb{P}}$ is inherently tied to the partitioning scheme used by the solution algorithm. Thus, for any given algorithm, we propose to estimate $d_{\mathbb{P}}$ through empirical profiling of the partitioning scheme (as we demonstrate in Section 4.2).

4. EVALUATION

A significant contribution of the theory we presented in Section 3 lies in its explanatory and predictive power, a claim which we now defend (anecdotally in Section 4.1, and empirically in Section 4.2).

4.1 Explaining Trends

Researchers have developed a number of different algorithms for exploiting the kinds weakly-coupled problem structured formalized above [2, 9, 10, 13, 14, 15, 16, 22]. Our unified characterization can explain some of the trends observed in the performance of these algorithms that are not easily explained without considering combinations of dimensions of agent coupling.

For instance, the successes of a family of ND-POMDP algorithms [9, 15] in scaling to many agents has been attributed to the reduced agent scope associated with ND-POMDP agents' *local neighborhoods* [9, 10, 15]. That is, as long as the agent scope remains small, these algorithms are expected to be practical. However, a generalized version of one of these algorithms (JESP [14]) has recently been reported as intractable for a test set of Distributed POMDPs with Coordination Locales (DPCLs) containing *just two agents*, even when generating an approximate solution [20]. A likely explanation for this phenomenon is contained within Equation 4, which suggests that it was not the agent scope of the problems that foiled JESP but instead the cost of JESP's best response calculation. Whereas ND-POMDP problems have an inherently restricted state factor scope due to the strict separation of transition-independent agents' local states, DPCL problems involve transition-dependent agents that need to reason about each others' state variables in order to compute optimal best responses (which JESP employs in computing approximate solutions), making the DPCL more strongly coupled even in its two-agent incarnation.

Transition dependence alone does not make a problem strongly coupled, however. In earlier work, we demonstrated the capability of an algorithm inspired by JESP, Optimal Influence-Space search (OIS), to scale optimally to sets of TD-POMDP problems with four transition-dependent agents [22]. At the time, little was understood about the structure of these problems that OIS was exploiting, especially given the wide range of OIS runtimes reported for a single set of problems. The theory developed in Section 3 leads us to attribute the successes of OIS to a low degree of influence in the test problems, a claim that is supported by empirical evidence presented in Section 4.2.1.

4.2 Predicting Performance of OIS

Our theory can also be used to make detailed predictions about the computational overhead of algorithms such as OIS that exploit weakly-coupled problem structure. Instead of presenting a comprehensive empirical analysis of all the dimensions of weak coupling, we use the limited space here to illustrate how to use Equation 4 to predict the relative computation time taken by OIS to solve variations of two example problems, named *fanout* and *fanin*, whose interaction digraphs are shown in Figure 2 (bottom). The two examples differ in their topology, but both include four agents connected by three *enablement* features, each linking randomly-selected tasks from the task sets of the corresponding agents. Each agent's task set contains three tasks, each with three randomly-selected outcomes whose *qualities* are random integers $\in (1, 10)$ and whose *durations* are random integers $\in (1, 5)$ selected without replacement.

Aside from demonstrating how to make empirically-guided theory-driven predictions, this experiment serves to elucidate the relationship between the *window* sizes of agents' tasks (examples of which appear in Figure 1a) and the computation time of OIS that we observed in an earlier analysis [22]. As such, we have generated sets of problems (of both the *fanin* and *fanout* flavors) whose task window sizes were fixed at {1, 2, 3, 4, and 5} and whose earliest start time and latest end times were selected so as to position the task's fixed-size window uniformly randomly in the interval $(0, 5)$.

4.2.1 Profiling Partitioner and Best-Response Solver

Whereas Equation 4 conveys a general bound, OIS employs a specific form of impact equivalence partitioning. In order to estimate the degree of influence $d_{\mathbb{P}}$ and partitioning complexity $C_{\mathbb{P}}$ that OIS will achieve on *fanin* and *fanout*, we profile OIS's partitioner and best response solver on two sets of smaller 2-agent enabler-enablee problems, one in which the enabler controls a single enablement (as do the enablers in *fanin*), and one in which the enabler controls three enablements (as does the enabler in *fanout*).

Figure 3 (top-left) shows $d_{\mathbb{P}}\|\Pi_i^{max}\|$ plotted as a function of *window size*. Each point represents the mean value of 20 randomly-generated profiling problems. Both flavors of problems exhibit an exponential trend in the number of equivalence classes, which explains why, in previously-reported results [22], OIS appeared to compute optimal solutions in exponentially less time as the window size was decreased. However, a striking feature of these plots is the high variance[2] (indicated by the error bars). This suggests that, in addition to window size, there are other factors at play that have a significant effect on the degree of influence. We ran an additional experiment in which we held the window size constant at 3 and varied the *earliest start time* of one enabling task, the results of which are shown in Figure 3 (top-right). From this plot, it appears that the temporal placement of agents' interactions is also a good predictor of the degree of influence, though analysis beyond the scope of this paper is required to verify this supposition.

Upon measuring the computation time taken by the enabler to perform its impact equivalence partitioning (represented as $C_{\mathbb{P}}$ in Eq. 4), we observed that it did not consume more than 70 percent of the total solve time on any given problem. We thereby deduced that for this particular suite of problems, the first term of Equation 4, $n \cdot \text{EXP}(\|Dom(\mathcal{X}_i^{max})\|) \cdot (d_{\mathbb{P}}\|\Pi_i^{max}\|)^{\omega}$, would be just as strong a predictor of OIS's

[2]To verify that the means were not simply driven up by outliers, we also examined the medians (not shown here), and observed the same exponential trend.

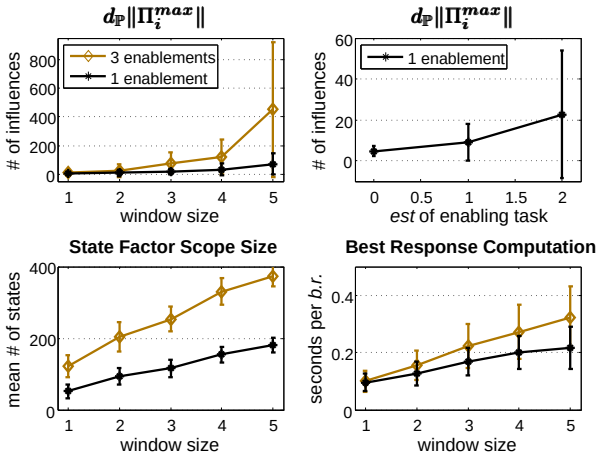

Figure 3: Profiling: $d_{\mathbb{P}}\|\Pi_i\|$ (top-left), $d_{\mathbb{P}}\|\Pi_i\|$ vs. earliest start time (top-right), $\|Dom(\mathcal{X}_i^{max})\|$ (bottom-left), and best-response comp. time (bottom-right).

relative computation time as the sum of both terms.

Whereas $\text{EXP}(\|Dom(\mathcal{X}_i^{max})\|)$ in Equation 4 is a worst-case bound on the complexity that makes no assumptions about the POMDP solver, OIS uses a particular solver whose computation time we measured on the problems in our profiling set. Figure 3 (bottom-left) plots the state factor scope domain size ($\|Dom(\mathcal{X}_i)\|$), measured as the actual size of the state space of the best-response model, as well as the mean computation time per best response (bottom-right). Note that, due to their topologies, the "3 enablement" problem profiles the best response computation of *fanin*, whereas the "1 enablement" problem profiles that of *fanout*. For both flavors, we observe a slight increase in both state factor scope domain size and best response computation time (which for these problems appear to be linearly, not exponentially, related in practice).

4.2.2 Predicting Relative Computation Time

Next, we evaluated $n \cdot \text{EXP}(\|Dom(\mathcal{X}_i^{max})\|) \cdot (d_{\mathbb{P}}\|\Pi_i^{max}\|)^{\omega}$ for both *fanin* and *fanout* across all window sizes, replacing $\text{EXP}(\|Dom(\mathcal{X}_i^{max})\|)$ with the profiled best response computation time shown in Figure 3 (bottom-right) and $d_{\mathbb{P}}\|\Pi_i^{max}\|$ with its profiled value (top-left). For *fanin* problems, ω was set to 3, and for *fanout*, 1; for all problems, $n = 4$. Figure 4 shows the computation time predicted by our theory (left), and the mean values of actual runtime of OIS (right) on 20 random *fanin* and *fanout* problems per window size.

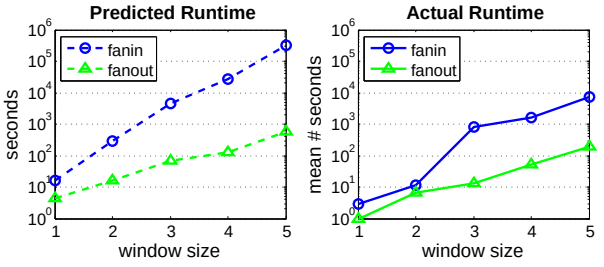

Figure 4: Predicted (left) and actual (right) OIS computation time vs. window size.

At a high level, the trend we predicted using our theoretical characterization matches the trend observed in actual data: for weakly-coupled problems that have a low degree of influence (because of their small task windows), the agent scope size has little effect on the computation time. However, as the window size increases, the size of the agent scope makes exponentially more difference. Such a prediction could not have been made, nor this trend well understood, without taking into account two different aspects of weak coupling: *degree of influence* and *agent scope size*. While our predicted runtimes appear to overestimate the actual runtimes in both cases, we do not expect the actual runtimes to precisely match predictions made using worst-case complexity bounds. For instance, Equation 4 does not account for the fact that in *fanin* problems, only a single agent is computing a best response. (In general, $\omega = 3$ topologies would require that all agents compute best responses.)

5. RELATED WORK

The first author's dissertation [21] includes a more rigorous, though less general, treatment of the theory presented in Sections 2–3. Compared to other treatments of weak coupling in multiagent sequential decision making, the primary distinctions of our analysis are (1) its synthesis of three aspects of weak coupling into a single unified characterization, and (2) its quantification of the computational benefits of exploiting weakly-coupled structure in a *general* context of optimal Dec-POMDP planning. Each of these aspects has appeared in the literature in some shape or form. For example, the work of Guestrin *et al.* [7] on exploiting restricted state factor scope and agent scope (both of which have also been referred to under the heading *locality of interaction* [9, 15, 16]) in factored value functions, though limited in context to approximate solution computation, plays a foundational role in our analysis.

The effect of restricted agent scope on problem hardness has been previously explored in MMDPs [5] (assuming fully-jointly-observable state), and also in ND-POMDPs [9, 10, 15] (assuming transition and observation independence). In the latter, Kumar and Zilberstein [10] make an explicit connection between induced width and complexity. Oliehoek *et al.* [16] analyze the stage-by-stage dynamics of state factor scope and agent scope in factored Dec-POMDPs, treating the planning problem as a series of collaborative graphical Bayesian games. Degree of influence, is grounded in the work of Becker *et al.* [2], who identify the *coverage set* as a subset of local policies that need be considered in a joint policy search, and the works of Rathnasabapathy *et al.* [18] and Pynadath and Marsella [17], who define *behavior*(*al*) *equivalence* over candidate models of an agent's peers. Our earlier work [22] on influence-based abstraction can be viewed as a means of partitioning the policy space into impact equivalence classes, each of which is summarized by an *influence*.

Aside from the three aspects on which our characterization concentrates, researchers have analyzed the relationships between other forms of interaction structure and problem hardness. For instance, Goldman and Zilberstein [6] characterize the complexity of various Dec-POMDP subclasses by classifying agents' direct communication and indirect sharing of information through observation. Shen *et al.* [19] characterize complexity of optimal Dec-MDP planning according to the complexity of the minimal encoding of agents' local policies. Allen and Zilberstein [1] develop an information-

theoretic metric, *influence gap*, that quantifies the difference in the degree to which each agent can affect world state transitions and joint rewards in a two-agent Dec-POMDP. Though the semantics of *influence* in Allen and Zilberstein's work differ substantially from ours, their results express the same general sentiment that varying levels of impact result in varying problem hardness.

Lastly, there is a strong connection between our analysis and that of Brafman and Domshlak [3], who transform the classical planning problem into one of *constraint satisfaction*. As in our analysis of joint policy computation as constraint optimization, they incorporate a parameter ω corresponding to the induced width of the constraint graph.

6. CONCLUSIONS AND FUTURE WORK

This paper takes an important step towards gaining a better understanding of what makes some Dec-POMDP problems so much harder to solve than others. We have jointly characterized three aspects of the weakly-coupled problem structure that, when exploited, accommodate quantifiable computational gains. By studying these aspects in a unified context, we have derived new bounds on the complexity of optimal multiagent planning.

Our theoretical results complement the abundance of recent algorithmic development geared towards solving problems with structured agent interactions [2, 9, 10, 13, 14, 15, 16, 22], by providing a gauge of problem difficulty based upon the degree to which a problem is weakly-coupled. We have demonstrated that our theory can explain observations about algorithm performance, as well as predict the relative computational overhead of algorithms that exploit some or all of the elements of weakly-coupled problem structure that we have characterized. These explanations and predictions could not have been formed without considering the combination of different aspects of weakly-coupled structure.

In the process of predicting the computation time of one such algorithm, OIS [22], our empirical analysis illustrated that OIS's past success was due, in part, to its exploitation of structure in problems with a low degree of influence. However, our analysis also exposed the need for a better understanding of the identifiable problem attributes that affect *degree of influence*, whose underlying structure is less discernible than that of *state factor scope domain size* and *agent scope size*. Moreover, in the future, we hope to find other advantageous structural aspects that would improve the predictive power of our characterization. Such a pursuit could not only expand our understanding of the performance of existing algorithms, but guide the design of better algorithms.

7. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their thoughtful comments, and Frans Oliehoek for his constructive feedback. This work was supported, in part, by NSF grants IIS-0534280 and IIS-0964512, and by AFOSR grant FA9550-07-1-0262.

8. REFERENCES

[1] M. Allen and S. Zilberstein. Agent influence as a predictor of difficulty for decentralized problem-solving. In *AAAI*, pages 688–693, 2007.

[2] R. Becker, S. Zilberstein, V. Lesser, and C. Goldman. Solving transition independent decentralized Markov Decision Processes. *JAIR*, 22:423–455, 2004.

[3] R. Brafman and C. Domshlak. From one to many: Planning for loosely coupled multi-agent systems. In *ICAPS*, pages 28–35, 2008.

[4] R. Dechter. *Constraint Processing*. Morgan K., 2003.

[5] D. Dolgov and E. Durfee. Graphical models in local, asymmetric multi-agent Markov decision processes. In *AAMAS*, pages 956–963, 2004.

[6] C. Goldman and S. Zilberstein. Decentralized control of cooperative systems: Categorization and complexity analysis. *JAIR*, 22:143–174, 2004.

[7] C. Guestrin, D. Koller, and R. Parr. Multiagent planning with factored MDPs. In *NIPS*, pages 1523–1530, 2001.

[8] A. Guo and V. Lesser. Planning for weakly-coupled partially observable stochastic games. In *IJCAI*, 2005.

[9] Y. Kim, R. Nair, P. Varakantham, M. Tambe, and M. Yokoo. Exploiting locality of interaction in networked distributed POMDPs. In *AAAI Spring Symp. on Distributed Planning and Scheduling*, 2006.

[10] A. Kumar and S. Zilberstein. Constraint-based dynamic programming for decentralized pomdps with structured interactions. In *AAMAS*, pages 561–568, 2009.

[11] J. Marecki and M. Tambe. Planning with continuous resources for agent teams. In *AAMAS*, pages 1089–1096, 2009.

[12] N. Meuleau, M. Hauskrecht, K.-E. Kim, L. Peshkin, L. P. Kaelbling, T. Dean, and C. Boutilier. Solving very large weakly coupled Markov decision processes. In *AAAI '98/IAAI '98*, pages 165–172, 1998.

[13] H. Mostafa and V. Lesser. Offline Planning For Communication By Exploiting Structured Interactions In Decentralized MDPs. In *Proceedings of IAT*, 2009.

[14] R. Nair, M. Tambe, M. Yokoo, D. V. Pynadath, and S. Marsella. Taming decentralized POMDPs: Towards efficient policy computation for multiagent settings. In *IJCAI*, pages 705–711, 2003.

[15] R. Nair, P. Varakantham, M. Tambe, and M. Yokoo. Networked distributed POMDPs: A synthesis of distributed constraint optimization and POMDPs. In *AAAI*, pages 133–139, 2005.

[16] F. A. Oliehoek, M. T. J. Spaan, S. Whiteson, and N. A. Vlassis. Exploiting locality of interaction in factored Dec-POMDPs. In *AAMAS*, pages 517–524, 2008.

[17] D. V. Pynadath and S. Marsella. Minimal mental models. In *AAAI*, pages 1038–1044, 2007.

[18] B. Rathnasabapathy, P. Doshi, and P. Gmytrasiewicz. Exact solutions of interactive pomdps using behavioral equivalence. In *AAMAS*, pages 1025–1032, 2006.

[19] J. Shen, R. Becker, and V. Lesser. Agent Interaction in Distributed MDPs and its Implications on Complexity. In *AAMAS*, pages 529–536, 2006.

[20] P. Varakantham, J. young Kwak, M. Taylor, J. Marecki, P. Scerri, and M. Tambe. Exploiting coordination locales in distributed POMDPs via social model shaping. In *ICAPS*, pages 313–320, 2009.

[21] S. Witwicki. *Abstracting Influences for Efficient Multiagent Coordination Under Uncertainty*. PhD thesis, University of Michigan, January 2011.

[22] S. Witwicki and E. Durfee. Influence-based policy abstraction for weakly-coupled Dec-POMDPs. In *ICAPS*, pages 185–192, 2010.