Deciding When to Commit To Action During Observation-based Coordination

Marcus J. Huber and Edmund H. Durfee*

Distributed Intelligent Agents Group (DIAG)
The University of Michigan
marcush@engin.umich.edu durfee@engin.umich.edu
http://ai.eecs.umich.edu/people

Abstract

We have developed a multiagent scheme which utilizes plan recognition as its primary means of acquiring the information necessary to coordinate the activities of agents. Preliminary research has demonstrated that the plan recognition system developed makes coordination of multiple agents possible. An important issue that arises when observation is the primary means of information acquisition is the introduction of uncertainty into the coordination process. We have explored the issue of early versus late commitment to the uncertain information thus gained and the resulting tradeoff between time and effort as the commitment level is changed. Our results show that while in some situations it is worthwhile delaying commitment until uncertainty is reduced, in other situations it is important to act even when uncertainty is high. The long–term goal of the research is to develop the notion of *coordination through observation*, where agents utilize plan recognition to acquire coordination information.

Introduction

To coordinate its plans with those of others, an agent must have a model of the plans of each of the other agents. It may then compare these plans to its own plans and goals and determine the potential for conflict or cooperation. Traditional techniques in multi-agent planning and coordination typically allow agents to explicitly communicate about their intentions, plans, and the relationships between them (Conry, Meyer, & Lesser 1988; Durfee, Lesser, & Corkill 1987). Some environments might not admit to explicit communication, however. This may arise because the communication medium between two agents that *can* communicate with each other might be unreliable (e.g. RF interference). Or, perhaps, using the communication channel might introduce new risks to one or more of the agents involved in the communications, such as may occur when operating in an environment in which there are hostile agents. And in situations where agents are antagonistic, the agents will certainly *not* wish to communicate plans and goals with other agents, and may have to rely on other means to ascertain the intentions of these other agents. In addition, traditional DAI techniques rely on strong assumptions of common communication language, communication protocol, and plan and goal representation (Cohen & Perrault 1979; Durfee & Montgomery 1990).

Communication-poor coordination techniques do exist, including social conventions (Shoham & Tennenholtz 1992), focal points (Kraus & Rosenschein 1991), decision-theoretic (Genesereth, Ginsberg, & Rosenschein 1984), and game-theoretic recursive modeling (Gmytrasiewicz, Durfee, & Wehe 1991). In general, these techniques emphasize implicitly or explicitly infering others' actions based on established norms for behavior or on beliefs about the preferences or interests of others. Thus, social conventions constrain behavior to make others predictable, but can be overconstraining in variable environments. Meanwhile, using focal points depends on agents identifying choices that are somehow universally distinctive, and so agents must see the world very similarly for this technique to work. Decision and game-theoretic techniques allow agents to model others as being utility-maximizing, and thus require agents to model the preferences over (utilities of) alternative choices of others. Therefore, communication-poor techniques generally assume either that agents are constrained in their actions in some commonly known way, or that agents can anticipate each others' actions based on knowledge of what others consider relevant or desirable.

Another alternative, and the approach that we have taken, is for the agents to infer the plans of each other by observing the actions or behaviors of the other agents (Charniak & Goldman 1991), rather than either actively communicating plans, goals, and intentions, or inferring these through knowledge of what others con-

*This research was sponsored in part by the NSF under grants IRI-9015423, IRI-9010645, and IRI-9158473, and by DARPA under contract DAAE-07-92-C-R012.

sider relevant or desirable. Moreover, the same coordination techniques may be applied to the recognized plans that were employed when explicit communication about plans was the means used for acquiring coordination information. Agents coordinating through the use of plan recognition do not even need to have a common communications medium or language. In fact, nothing has to be in common among the agents other than the knowledge of the observed agent's possible goals, and the actions it will exhibit while pursuing those goals. Perception capable of making observations of other agents' behavior is also necessary, but is not dependent upon the other agents in the world. The various agents, then, may be heterogeneous, having different representations of models and plans, perceptual capabilities, communication means, etc.

We recognize, however, that there are costs associated with the plan recognition process. These costs include the introduction of uncertainty due to imperfect observation and inferencing (Huber & Durfee 1993), additional computation time and effort to perform plan and goal inferencing, and the relative difficulty (compared to direct communication methods) of recognizing and incorporating new goals and behaviors into the models of other agents (Huber, Durfee, & Wellman 1994). Cooperating agents will need to weigh the costs and risks associated with explicit communication against the costs associated with plan recognition. If explicit communication is relatively inexpensive and risk free it is probably advantageous for the agent to utilize it. On the other hand, as explained above, there will be circumstances that may force an agent to use observation and plan recognition to cooperate with other agents. We think that by experimenting with plan recognition in a variety of domains, we will gain a deeper understanding of the situations in which plan recognition is, and is not, a useful process.

In this paper, we present preliminary research results showing that multiple agents employing plan recognition can successfully coordinate their activities. We have further explored one of the significant issues identified during the preliminary research, that of the uncertainty introduced by inferencing during the plan recognition process, and we present the results of this research. We are working towards a more formal examination of *coordination through observation*, the paradigm in which multiple agents can coordinate their behavior through the use of plan recognition for plan and goal recognition.

Probabilistic Plan Recognition

In this section, we first discuss our experimentation environment and agent capabilities. We then describe our plan recognition system and the integration of the plan recognition system into agents performing navigation tasks in the environment.

The environment in which our agents act is a discrete time, two-dimensional, simulated grid world. This environment was created using MICE (Montgomery & Durfee 1990), a testbed designed explicitly for experimentation with multiple agent coordination techniques. The agents in this environment were limited to the primitive actions of motion (movement north, south, east, or west) or no motion. Two types of agents were implemented: observed agents ("bounding agents" or "bounders") and observing agents ("overwatching agents" or "overwatchers"). The bounding agents were given a goal to occupy a certain designated location, and would head directly for that location throughout an experiment. The overwatching agents observed the actions of the bounding agents, tried to infer the goals of the observed agents (the particular goal destination), and acted accordingly. The behavior of the overwatching agents will be explained in more detail later.

The bounding agent's heuristics were quite simple. Placed in some initial starting location, and given a goal location, the agent would plan the shortest, straight-line path to the goal and then start moving toward it. Once the bounding agent arrived at its goal, it would then remain motionless for the remainder of the experiment.

The overwatching agent's heuristics were variable. In all cases the agent would react to its beliefs by moving to a particular destination location. This destination could be the same location, a nearby location, or a location opposite to that which the bounding agent was moving towards, depending upon the heuristics specified at the beginning of the experiment. The first heuristic, moving to the same location as the observed agent, might be used if the agents were supposed to rendezvous, enabling them to do this without having to explicitly communicate with each other. The second heuristic, moving to a nearby destination location, might be used if the two agents were supposed to perform surveillance, where the "destinations" might be hilltops and the coordination results in a larger coverage area. The third type of heuristic, moving to an opposite location, might be used if the agents were adversaries and the overwatching agent wanted to avoid being noticed by the other agent.

Goal Inference

The overwatching agent's primary objective was to accurately infer the goal (the final destination) of the bounding agent. To do this, each overwatching agent employed a belief network implemented using the IDEAL system (Srinivas & Breese 1990). The IDEAL system consists of a collection of LISP functions that permit the construction and maintenance of belief networks, the accumulation of observations, and the subsequent propagation of the beliefs throughout the network in order to perform plan inferencing. At each simulated time step, the overwatching agent would observe the actions of the bounding agent. This evidence was added to the history of previously observed actions and the beliefs of the Bayesian network would then be updated. With consistent and reinforcing evidence, belief in a possible destination would increase. Conversely, contradictory evidence would cause a belief in a destination to decrease, possibly to zero.

The belief network architecture is quite simple, and is shown in Figure 1(b). In this figure, nodes are probabilistic variables and arcs represent causality or dependence. Each of the nodes contains a set of possible states, representing the possible values that each value may take. Root nodes (i.e. the *Location* and *Destination* nodes) contain values conditionally independent of any other node. The dependent nodes (i.e. the *Move* node) are conditionally dependent upon its immediate predecessors. The belief network shown represents the idea that a move is dependent upon both the goal destination and the particular location the agent is in.

The state spaces of each of the nodes varies. The *Move* node contains the states of {North, South, East, West, Stay}. The *Location* and *Destination* nodes contain a mutually exclusive and exhaustive list of regions of the environment, (a hierarchical, multi-resolution representation similar in form to a quad-tree representation.

For the belief network shown, we had to determine the prior probabilities for each of the possible states of the independent nodes (the *Location* and *Destination* nodes), and the conditional probabilities for the dependent nodes (the *Move* node). The probabilities were computed with the implicit assumption that the agent will move in a straight line to its destination from its current location, so the network is "fooled" in situations where the watched agent is trying to trick the watching agent by making feinting maneuvers, or is taking a path that takes it around some obstacle that is in its way.

For the *Location* node, the state that represents the current position of the agent is set to a probability of 1.0, since the observing agent can tell where the other agent is at all times and sensors are assumed to be perfect in this implementation.[1] The *Destination* node's states are initially equiprobable, as the observing agent starts without any bias toward which goal the bounding agent will be heading toward.

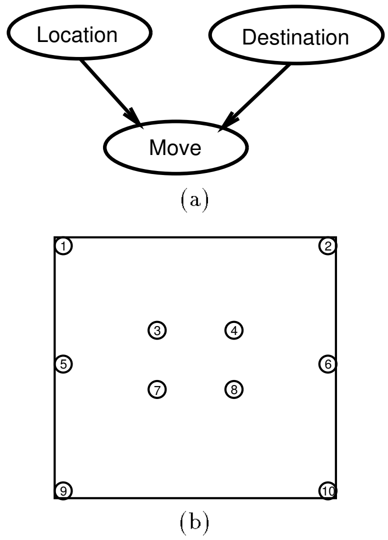


Figure 1: (a) Initial belief network architecture and (b) grid-world plan recognition environment.

Once the belief network is constructed, an agent can make observations of another agent and add this "evidence" to nodes in the belief network. This evidence can then be propagated throughout the network to the other nodes, resulting in posterior probabilities (the observing agent's "beliefs" in the observed agent's goals and intentions) given the evidence. In the next section, we discuss the issue of when to commit to this uncertain information concerning other agents' intentions.

Commitment

In this section, we discuss the tradeoffs an agent faces when dealing with the uncertain information generated by the plan recognition system. For example, suppose the overwatching agent observes the behavior of another agent, uses the plan recognition system to determine the most likely plans of the agent, and then acts "appropriately" based upon this inferred information. If the observing agent is always in motion, moving toward whichever destination location that was most appropriate for the highest probability plan believed to be held by the other agent, the agent would possibly be performing unnecessary actions due to errors in its beliefs of the other agent's goals. These extra actions would also be performed regardless of the level of belief

[1] Further experiments, motivated by experiments with actual robots (Huber & Durfee 1993), where uncertainty of position is added, demonstrates that the system undergoes graceful degradation of performance with increased uncertainty.

held for the most probable goal.

Early commitment, then, was shown to result in extra effort on the part of the watching agent. It is easy to see that this extra effort could be eliminated by simply having the watching agent wait until the other agent had reached its final destination and then plan an optimal, least-effort path to its own destination. The cost of doing this, however, is the extra time taken waiting for the watching agent to reach its goal.

With respect to building intelligent agents, we were interested in the tradeoff between these two factors to determine the situations in which it was preferable to commit early or late. We also wanted to get a feel for belief level thresholds and intervals that were significant in their impact upon the time and movement costs of the agents. With this meta-level knowledge then, agents finding themselves in a particular environment can more wisely decide upon belief thresholds and their courses of action. To explore these questions, we performed a series of experiments, described below.

In the following simulation experiments, we measured several parameters that we felt would reveal the tradeoff discussed above. These were:

Ave End Time: The average time step at which the simulation ended. The criterion for this was the time at which all agents arrived at their final destinations.

Ave Last Move Time: The average time step at which the observing agent performed its last movement action. The observing agent stopped moving once it had arrived at the destination determined by the heuristics described earlier (the same, nearby, or "opposite" location as that of the observed agent.)

Ave Total Moves: The average number of moves made by the observing agent throughout the experiment.

Symmetric Destination Distribution

Our first experiments were within the environment shown in Figure 1(a), in which the destinations were distributed fairly evenly and symmetrically throughout the 32x32 grid "world". In each experiment, the bounding agent's initial location was picked at random, and the agent always moved to Destination 1. Three variations of the overwatching agent's behavior were experimented with, corresponding to the three heuristics described earlier: moving to the same, nearby, or opposite location as that of the watched agent. For each experiment we measured the total time taken for the simulated run, the last time step in which the observing agent moved, and the total number of moves taken by the observing agent.

Because we wanted to get an understanding of the tradeoffs of time versus effort as commitment changed, we established a threshold of belief. Only when the overwatching agent held belief in a potential destination location above this threshold would it commit to the appropriate destination location. In the experiments where the overwatching agent moved to the same location which it believed the other agent was moving toward, shown in Figure 2, the tradeoff between time and effort is clearly seen.

In Figure 2 it is evident that, as the watching agent waits for higher levels of belief before committing to action, it saves in the total number of moves that it has to make, on the average. In this scenario, the agent saves approximately two moves if it waits until it is absolutely sure which destination the agent that it is watching is going to. The extra time spent waiting, six time steps, might be quite costly relative to the cost of motion, however, and would have to be considered within the constraints of the domain (i.e. some utility function). Another observation from these results is that there seems to be little effect of having a threshold below 0.3 or 0.4. It appears that beliefs below this level indicate that the overwatching agent is still pretty confused about the final destination of the other agent.

The results of experiments in which the overwatching agent moved to a "nearby" destination location is shown in Figure 3. The "near" destination for each destination was predefined at the beginning of the simulation, and was never changed (e.g. Destination 3 was "near" to Destination 4, Destination 5 was "near" to Destination 1, etc.) In this scenario, the tradeoff between effort and time is reversed. As the threshold level increases, the effort saved by the watching agent is six moves, while the time saved is only two time steps. This is due in part to the nonlinear relationships between the "near" destinations, resulting in moves that are more likely to take the watching agent further from the final destination of the watched robot. Using the "same" heuristic resulted in moves that, although headed toward the wrong destination, were pretty much always in the right general direction.

The results of the experiments in which the overwatching agent moved to an "opposite" destination location show a great deal of similarity with the "near" destination heuristic, with the only significant difference being the relatively high values for the "opposite" heuristic (about eight points higher for all the curves).

Asymmetric Destination Distribution

Three scenarios were explored in which the goals (the destinations) were not arranged symmetrically throughout the environment as they were above. We

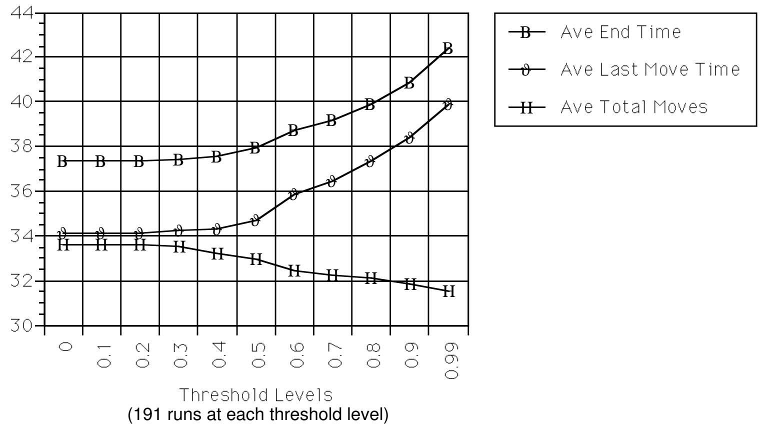


Figure 2: Results of symmetric distribution, "Same" complimentary destination.

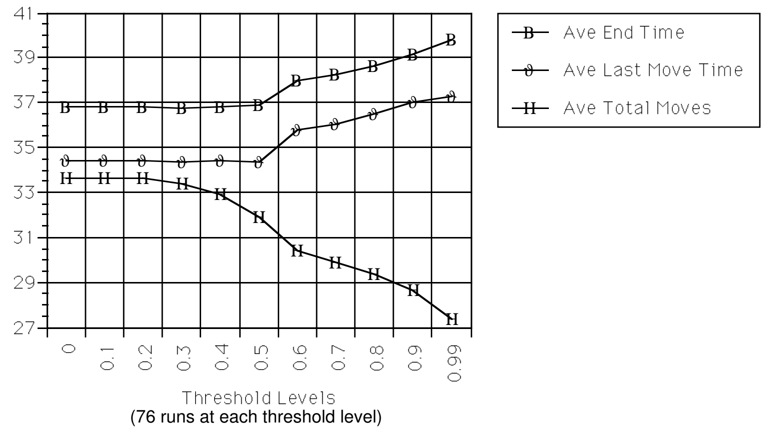


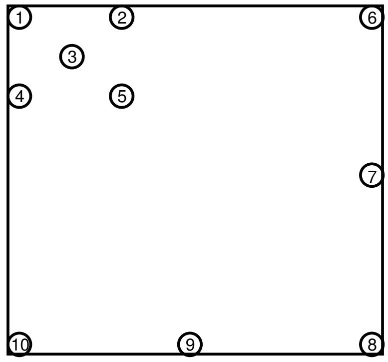
Figure 3: Results of symmetric distribution, "Near" complimentary destination.

thought that this would be more realistic and reveal other interesting phenomena. Two of the environments were designed to accentuate the tradeoffs present in the environments above by placing the destinations in such a manner as to make it costly for the watching agent to make a mistake. These two environments are shown in Figure 4(a) and Figure 5(a). The compact distribution of destinations was designed to make it difficult for the observing agent to distinguish the final destination until very late in the simulation. The second environment was designed to be a worst case situation, where switching between two probable destinations would result in a complimentary destination a great distance away from the previous complimentary destination.

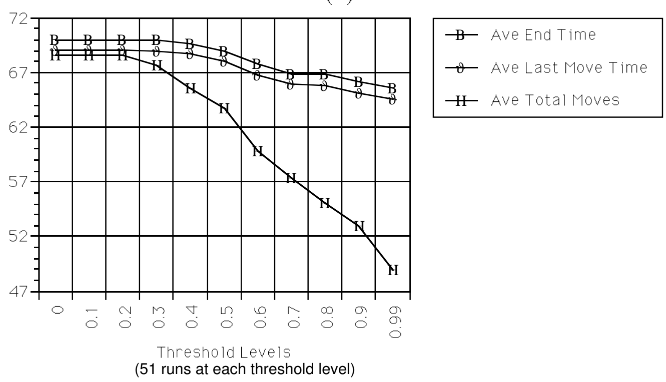
In the experiments in the compact destination distribution, the results of which are shown in Figure 4(b), the graphs indicate distinct changes in the behavior of the observing agent. Notable is the nearly constant number of moves made, regardless of the threshold value. As the threshold increases, however, the time cost becomes quite large, approaching eight time steps at the upper limit of the threshold range. The threshold level at which this becomes apparent (0.2) is significantly lower than in other environments, also indicating that the observing agent can make decisions with quite a higher level of uncertainty in its beliefs.

In this environment, it does not benefit the overwatching agent to wait for the emergence of a highly probable destination, but to commit itself to whichever destination is most likely at each time step. The environment characteristic that results in this behavior is the compact distribution of the potential destination locations. All of the destinations are in a very small region of the environment, so the observing agent's best course of action is to start moving toward the grouping, regardless of the particular goal destination of the other agent. Otherwise, the agent must wait quite a while until the watched agent's behavior distinguishes the particular destination to which it is going.

In the experiments in the worst–case destination distribution, the results of which are shown in Figure 5(b), we see a phenomenon not seen in any of the other environments. Here we see a sharp and very large decrease (approximately twenty) in the total number of moves. Most notable, however, is the *decrease* in the simulation times. This "worst–case" scenario brought out the behavior that we were looking for. Namely, that early commitment to an incorrect final destination, resulting in subsequently switching to the correct destination later in the simulation, led to movement on the part of the observing agent that took it pretty much in the opposite direction of where it eventually wanted to go. The decrease in the simulation times for large



(a)



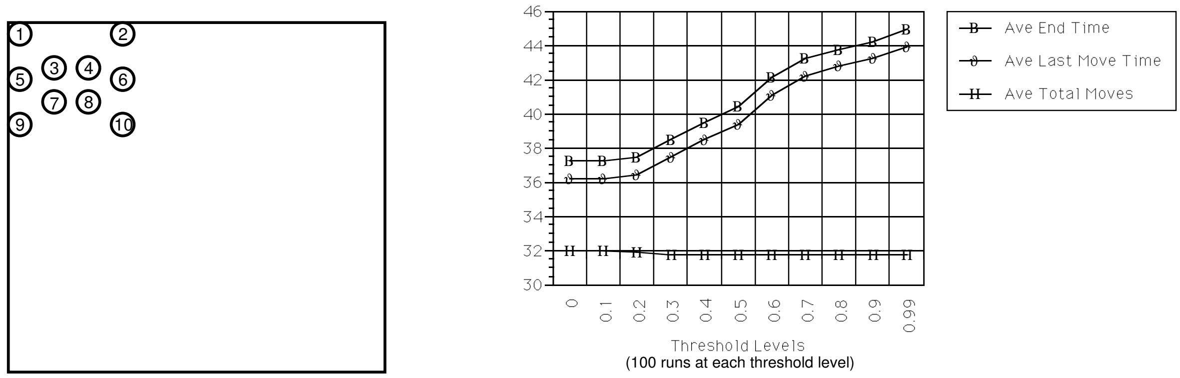
(b)

Figure 5: (a) Worst-case destination distribution, and (b) experimental results, "Worst-case opposite" complimentary destination.

thresholds results from not moving in the wrong direction, which would require later backtracking. And, of general note, is the relatively high values for all of the variables in this environment. This is an indication that it is a very difficult environment in which to operate, requiring a high level of effort for all random placements of the agents.

Discussion

It is apparent that there is a distinct tradeoff between time and effort as a function of commitment. In almost all of the experiments, there was a noticeable increase in the time of completion of the simulations, indicating that the observing agent waited for some amount of time until it was certain of the intentions of the watched agent. And, while this generally resulted in a decrease in the total effort expended by the observing agent, it was also apparent that the magnitude of this reduction in effort depended markedly upon the environment. In some scenarios, the small decrease in expended effort may not be enough of a savings to the agent to compensate for the extra time taken. In one particular scenario, both the time *and* the effort



(a) (b)

Figure 4: (a) Compact destination distribution and (b) experimental results, "Same" complimentary destination.

decreased as the level of belief before commitment was increased, showing that there may be domains in which the cost of waiting may indeed be justified.

The implications of these experiments is that an agent will have to be able to recognize certain characteristics of the environment, and the particular situation that it is in, in order to determine the relative cost of acting versus perceiving. It appears that in the domain specified here, pretty much regardless of the environment setup, that a threshold of below 0.30 was useless. Committing to beliefs below this value resulted in little or no improvement in either coordination time or effort. On the contrary, while in general an upper threshold of approximately 0.50 to 0.70 seems be the point at which benefits of reduced effort are first realized, it is not true for all environments; the observing agent must first categorize the environment (perhaps by determining how costly it is to commit to an incorrect final destination) before deciding upon its belief threshold.

For the sake of simplicity, our analysis in all of these experiments assumed the agent had a very simple "cost", or "utility" function, which simply compared the raw values of elapsed time and number of moves. Agents may certainly have more complex cost functions, of course, to reflect the relative importance of time, movement, and other factors.

Summary

We have developed a multiagent scheme which utilizes plan recognition as its primary means of acquiring the information necessary to coordinate the activities of agents. The research domain to which we have applied our scheme is coordinated motion and navigation among multiple robots, both in simulation and in the real world. Preliminary research has demonstrated that the plan recognition system developed makes coordination of multiple agents possible. It has also identified an important issue: observation is the primary means of information acquisition, and is responsible for the introduction of uncertainty into the coordination process. We have, therefore, explored the issue of early versus late commitment to the information thus gained and the resulting tradeoff between time and effort as the commitment level is changed. The long–term goal of the proposed research is to develop the notion of *coordination through observation*, whereby agents can successfully utilize plan recognition processes to acquire coordination information. We believe that by developing such a system in real–world domains, it will allow us to identify and begin solving critical issues in plan recognition and coordination.

Extensions that we have been working on (see (Huber & Durfee 1993)) include modeling the uncertainty inherent in sensor-based observations into the plan recognition architecture, a necessity for real-world environments. Realistic domains will also require an extension of the spatial representation beyond that of the simple quad-tree representation used thus far. An agent will also need to be able to observe multiple agents and infer (potentially multiple) plans and goals for each of these individuals, and possibly for a group of agents as well. Furthermore, the assumption that the observing agent has a complete and accurate model of the observed agents' plan hierarchy is unrealistic and will have to be relaxed.

Acknowledgements

We would like to thank Mike Wellman for many discussions regarding belief networks and reasoning under uncertainty. We would also like to thank the reviewers for their constructive comments.

References

Charniak, E., and Goldman, R. 1991. A probabilistic model of plan recognition. In *Proceedings Ninth National Conference on Artificial Intelligence*, 160–165. Anaheim, CA: American Association for Artificial Intelligence.

Cohen, P. R., and Perrault, C. R. 1979. Elements of a plan-based theory of speech acts. *Cognitive Science* 3(3):177–212.

Conry, S. E.; Meyer, R. A.; and Lesser, V. R. 1988. Multistage negotiation in distributed planning. In Bond, A. H., and Gasser, L., eds., *Readings in Distributed Artificial Intelligence*. Morgan Kaufman. 367–384.

Durfee, E. H., and Montgomery, T. A. 1990. A hierarchical protocol for coordinating multiagent behaviors. In *Proceedings of the National Conference on Artificial Intelligence*, 86–93.

Durfee, E. H.; Lesser, V. R.; and Corkill, D. D. 1987. Coherent cooperation among communicating problem solvers. *IEEE Transactions on Computers* C-36(11):1275–1291. (Also published in *Readings in Distributed Artificial Intelligence*, Alan H. Bond and Les Gasser, editors, pages 268–284, Morgan Kaufmann, 1988.).

Genesereth, M.; Ginsberg, M.; and Rosenschein, J. 1984. Cooperation without communications. Technical Report 84-36, Stanford Heuristic Programming Project, Computer Science Department, Stanford University, Stanford, California 94305. (Also published in *Readings in Distributed Artificial Intelligence*, Alan H. Bond and Les Gasser, editors, pages 220–226, Morgan Kaufmann, 1988.).

Gmytrasiewicz, P. J.; Durfee, E. H.; and Wehe, D. K. 1991. A decision-theoretic approach to coordinating multiagent interactions. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*.

Huber, M. J., and Durfee, E. H. 1993. Observational uncertainty in plan recognition among interacting robots. In *Working Notes: Workshop on Dynamically Interacting Robots*, 68–75.

Huber, M. J.; Durfee, E. H.; and Wellman, M. P. 1994. The automated mapping of plans for plan recognition. In *Proceedings of the 1994 Distributed AI Workshop*, 137–152.

Kraus, S., and Rosenschein, J. S. 1991. The role of representation in interaction: Discovering focal points among alternative solutions. In Demazeau, Y., and Muller, J.-P., eds., *Decentralized AI*. North Holland.

Montgomery, T. A., and Durfee, E. H. 1990. MICE users guide. Technical Report CSE-TR-64-90, Computer Science and Engineering Division, EE and CS Department, University of Michigan, Ann Arbor, MI 48109. (Revised 1992).

Shoham, Y., and Tennenholtz, M. 1992. On the synthesis of useful social laws for artificial agents societies (preliminary report). In *Proceedings of the National Conference on Artificial Intelligence*.

Srinivas, S., and Breese, J. 1990. Ideal: A software package for analysis of influence diagrams. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, 212–219.