On Decision-Theoretic Approach to Game Theory

Piotr J. Gmytrasiewicz
Department of Computer Science
University of Illinois at Chicago, Chicago, IL 60607-7053
piotr@cs.uic.edu

Abstract

In our work we adopted the decision-theoretic principle of expected utility maximization as a paradigm for designing autonomous rational agents operating in multi-agent environments. Our approach differs from techniques based on game theory; we are not looking for equilibria, and we do not have to assume that the agents have arrived at the state of common knowledge. Instead, we endow an agent with a representation that captures the agent's knowledge about the environment and about the other agents, including its knowledge about their states of knowledge, which can include what they know about the other agents, and so on. This approach has been called the decision-theoretic approach to game theory. It avoids some of the drawbacks of game-theoretic equilibria that may be nonunique and do not capture off-equilibrium behaviors, but it does so at the cost of having to represent, process and continually update the nested state of agent's knowledge.

Introduction

In systems involving multiple agents, system builders have traditionally analyzed the task domain of interest and, based on their analyses, imposed upon the agents certain rules (laws, protocols) that constrain the agents into interacting and communicating according to patterns that the designer deems desirable.

The fundamental problem we address in this paper, on the other hand, is how agents should make decisions about interactions in cases where they have no common pre-established protocols or conventions to guide them.[1] Our argument is that an agent should rationally apply whatever it does know about the environment and about the capabilities, desires, and beliefs of other agents to choose (inter)actions that it expects will maximally achieve its own goals. This kind of agent description adheres to the knowledge-level view, articulated, for example by Newell (Newell 1981), and is a cornerstone of artificial intelligence, but operationalizing it is a complex design process.

In our work, we use the normative decision-theoretic paradigm of rational decision-making under uncertainty, according to which an agent should make decisions so as to maximize its expected utility (Coles *et al.* 1975; Doyle 1992; Feldman & Sproull 1977; Haddawy & Hanks 1990; Jacobs & Kiefer 1973; Russell & Norvig 1995). Decision theory is applicable to agents interacting with other agents because of uncertainty: The abilities, sensing capabilities, beliefs, goals, preferences, and intentions of other agents clearly are not directly observable and usually are not known with certainty. In decision theory, expected utility maximization is a theorem that follows from the axioms of probability and utility theories (Fishburn 1981; Myerson 1991). In other words, if an agent's beliefs about the uncertain environment conform to the axioms of probability theory, and its preferences obey the axioms of utility theory (see, for example, (Russell & Norvig 1995) page 474), then the agent should choose its actions so as to maximize its expected utility.[2]

[0]This research was supported by ONR grant N00014-95-1-0775, and by the National Science Foundation CAREER award IRI-9702132. This paper is partially based on (Gmytrasiewicz & Durfee 2000)

[1]We would like to stress that our approach does not forbid that agents interact based on protocols. However, since the protocols specify the agent's action the agent does not need to deliberate about what to do and our approach is not applicable. If the protocol is not applicable or leave a number of alternatives open then the agent needs to choose, and should do so in a rational manner.

[2]Some authors have expressed reservations as to the

The expected utilities of alternative courses of action are generally assessed based on their expected results. Intuitively, an agent is attempting to quantify how much better off it would be in a state resulting from it having performed a given action. In a multi-agent setting, however, an agent usually cannot anticipate future states of the world unless it can hypothesize the actions of other agents. Therefore, it may be beneficial for the agent to model other agents influencing its environment to assess the outcomes and the utilities of its own actions. We say that an agent is *coordinating* with other agents precisely when it considers the anticipated actions of others as it chooses its own action.

An agent that is trying to determine what the other agents are likely to do may model them as rational as well, thereby using expected utility maximization as a descriptive paradigm.[3] This, in turn, leads to the possibility that they are similarly modeling other agents in choosing their actions. In fact, depending on the available information, this nested modeling could continue on to how an agent is modeling other agents that are modeling how others are modeling, and so on.

Thus, to rationally choose its action in a multi-agent situation, an agent should represent the, possibly nested, information it has about the other agent(s), and utilize it to solve its own decision-making problem. This line of thought, that combines decision-theoretic expected utility maximization with reasoning about other agent(s) that may reason about others, leads to a variant of game theory that has been called a decision-theoretic approach to game theory (DTGT) (Aumann & Brandenburger 1995; Brandenburger 1992; Kadane & Larkey 1982; Raiffa 1982).

At least some of the comparison of DTGT to traditional equilibrium analysis has to deal with the notion of *common knowledge* (Aumann 1976). A proposition, say p, is common knowledge if and only if everyone knows p, and everyone knows that everyone knows p, and everyone knows that everyone knows that everyone knows p, and so on *ad infinitum*. In their well-known paper (Halpern & Moses 1990), Halpern and Moses show that, in situations in which agents use realistic communication channels which can lose messages or which have uncertain transmission times common knowledge is not achievable in finite time unless agents are willing to "jump to conclusions," and assume that they know more than they really do.[4]

In other related work in game theory, researchers have investigated the assumptions and limitations of the classical equilibrium concept (Binmore 1982; Geanakoplos 1992; Kadane & Larkey 1982; Reny 1988; Tan & Werlang 1988). Unlike the outside observer's point of view in classical equilibrium analysis, DTGT takes the perspective of the individual interacting agent, with its current subjective state of belief. This coincides with the subjective interpretation of probability theory used in much of AI (see (Cheeseman 1985; Neapolitan 1990; Pearl 1988) and the references therein). Its distinguishing feature seems best summarized by Myerson ((Myerson 1991), Section 3.6):

> The decision-analytic approach to player i's decision problem is to try to predict the behavior of the players other than i first, and then to solve i's decision problem last. In contrast, the usual game-theoretic approach is to analyze and solve the decision problems of all players together, like a system of simultaneous equations in several unknowns.

Binmore (Binmore 1982) and Brandenburger (Brandenburger 1992) both point out that unjustifiability of common knowledge leads directly to the situation in which one has to explicitly model the decision-making of the agents involved given their state of knowledge, which we are advocating. This modeling is not needed if one wants to talk only of the possible equilibria. Binmore points out that the common treatment in game theory of equilibria without any reference to the equilibriating process that achieved the equilibrium[5] accounts for the inability of predicting which particular equilibrium is the right one and will actually be realized, if there happens to be more than one candidate.[6]

justifiability of these axioms. See the discussions in (Malmnas 1994) and the excellent overview of descriptive aspects of decision theory in (Camerer 1995).

[3]The use of expected utility maximization to predict and explain human decision making is widely used in economics. See the overview in (Camerer 1995).

[4]Halpern and Moses consider the concepts of epsilon common knowledge and eventual common knowledge. However, in order for a fact to be epsilon or eventual common knowledge, other facts have to be common knowledge within the, so called, view interpretation. See (Halpern & Moses 1990) for details. Also, it has been argued that common knowledge can arise due to the agents' copresence, and, say, visual contact. These arguments are intuitive, but turn out to be difficult to formalize, so we treat the issue here as open.

[5]Binmore compares it to trying to decide which of the roots of the quadratic equation is the "right" solution without reference to the context in which the quadratic equation has arisen.

[6]Binmore (Binmore 1994), as well as others in game theory (Kandori, Mailath, & Rob 1991; Kandori & Rob 1991; Choo & Matsuri 1992a; 1992b) and related fields

The notion of nested beliefs of agents is also closely related to interactive belief systems considered in game theory (Aumann & Brandenburger 1995; Aumann 1999a; Harsanyi 1967; Mertens & Zamir 1985). In our own approach we decided to use a representation that is somewhat more expressive, since it also includes models of others that do not assume their rationality. Thus, they are able to express a richer spectrum of the agents' decision making situations, including their payoff functions, abilities, and information they have about the world, but also the possibility that other agents should be viewed not as intentional utility maximizers, but as mechanisms or simple objects. Somewhat related to nested belief states is also the familiar minimax method for searching game trees (Nilsson 1971). However, game tree search assumes turn taking on the part of the players during the course of the game and it bottoms out when the game terminates or at some chosen level.

The issue of nested knowledge has also been investigated in the area of distributed systems (Fagin, Halpern, & Vardi 1991) (see also (Fagin *et al.* 1995)). In (Fagin, Halpern, & Vardi 1991) Fagin and colleagues present an extensive model-theoretic treatment of nested knowledge which includes a no-information extension to handle the situation where an agent runs out of knowledge at a finite level of nesting.

Modeling Agent's Knowledge in Multi-agent Environments

We are interested in a representation capable of expressing the uncertain state of agent's knowledge about its environment, and reflect the agent's uncertainty as to the other agents' intentions, abilities, preferences, and sensing capabilities. On a deeper level of nesting, the agents may have information on how other agents are likely to view them, how they themselves think they might be viewed, and so on.

Representation

One possible representation could be based on the framework of Markov decision processes (MDP) (Boutilier, Dean, & Hanks 1999; Hauskrecht 2000). A (partially observable) MDP for an agent i is defined as

$$MDP_i =< S, A_i, \Theta_i, T_i, O_i, R_i > \quad (1)$$

where:

(Smith 1982), suggest the evolutionary approach to the equilibrating process. The centerpiece of these techniques lies in methods of belief revision, which we investigated in (Gmytrasiewicz, Noh, & Kellogg 1998; Suryadi & Gmytrasiewicz 1999).

- S is a set of possible states of the environment,
- A_i is a set of actions agent i can execute.
- T_i is a transition function $T_i : S \times A_i \times S \rightarrow [0,1]$ which describes results of agent i's actions, and.
- Θ_i is the set of observations that the agent i can make.
- O_i is the agent's observation function - O_i : $\Theta_i \times S \times A \rightarrow [0,1]$ which specifies probabilities of observations if agent executes various actions in different states.
- R_i is the reward function representing the agent i's preferences; $R_i : S \rightarrow \mathbf{R}$.

MDP's can be extended to involve multiple agents. Multi-agent MDP's have been proposed by Boutilier (Boutilier 1999) and by Milch and Koller (Koller & Milch 2002), but both of these extensions are not expressive enough to represent the agents' possibly different states of knowledge about their environment, their knowledge about their states of knowledge, and so on.

To define an MDP capable of expressing the agent's private information about other agents we consider an agent i that is interacting with $N-1$ other agents numbered 1, 2, ..., $i-1$, $i+1$, ..., N. We define a **recursive MDP** (RMDP) as:

$$RMDP_i =< RS, A, \Theta_i, T_i, O_i, R_i > \quad (2)$$

where:

- RS is a set of **augmented** possible worlds. Each augmented possible world is a pair $(s, P(\mathbf{M}_{-i}))$, such that $s \in S$ and $P(\mathbf{M}_{-i})$ is a probability distribution over possible models of other agents. Each model is a list of models of the other agents: $M_{-i} = (M_1, M_2, ..., M_{i-1}, M_{i+1}, ..., M_N)$. Each model of a an agent j can be represented in three possible forms:

$$M_j = \begin{cases} RMDP_j & \text{- the intentional model,} \\ No-Info_j & \text{- the no-information model,} \\ Sub-Int_j & \text{- the sub-intentional model.} \end{cases} \quad (3)$$

- $A = \times A_j$ is the set of joint moves of all agents.
- T_i is a transition function $T_i : S \times A \times S \rightarrow [0,1]$ which describes results of all agent's actions.
- Θ_i, O_i and R_i are defined as above.

The fact that an intentional model of another agent is part of an agent's RMDP gives rise to nesting of models. The no-information model can also be represented as an MDP (say one in which $R_i = 0$), and the sub-intentional model can be represented as a probability distribution over actions

(Gmytrasiewicz & Durfee 2000). Recursive MDP allows an agent to represent its uncertainty as to the state of the "physical" world (as a possible element of S), and also what are the possible and likely states of other agents' knowledge about the world, their preferences and available actions, their states of knowledge about others', and so on.

Our definition above is fairly general; RMDPs are related to the knowledge hierarchies considered in (Aumann 1999a), which in turn are similar to recursive Kripke structures defined in (Gmytrasiewicz & Durfee 1992). If one retains only the probability distribution $P_i(RS)$, the knowledge-belief hierarchies defined in (Aumann 1999b) obtain. If one omits the other agents' models from augmented possible worlds, stochastic games (Fudenberg & Tirole 1991) similar to ones investigated in (Boutilier 1999) and (Koller & Milch 2002) obtain.

Values and Optimality

The recursive MDPs give rise to agent's information states, just as POMDPs do (Hauskrecht 2000; Russell & Norvig 1995). The information state summarizes all of the agent's previous observations and actions. Under some conditions it turns out that the agent's belief states, $b(S)$, i.e. probability distributions over states, are sufficient and compact representations of information states.

Agent's beliefs evolve with time and they form a belief-state MDP. The new belief state, $b_t(s)$ is a function of the previous state, $b_{t-1}(s)$, the last action, and the new observation, in the predict-act-observe cycle (Russell & Norvig 1995), section 17.4, and (Hauskrecht 2000):

$$b_t(s) = \frac{O(o_t, s, a_{t-1})}{P(o_t|b_{t-1}, a_{t-1})} \sum_{s' \in S} T(s|a_{t-1}, s')b_{t-1}(s') \quad (4)$$

For an infinite horizon case, each belief state has an associated optimal value reflecting the maximum discounted payoff the agent can expect in this belief state:

$$V^*(b) = max_{a \in A_i}[\sum_{s \in S}\sum_{s' \in S} R(s,a,s')P(s'|s,a)b(s) + \gamma \sum_{o \in \Theta}\sum_{s \in S} O(o,s,a)b(s)V^*(b)]$$

and the optimal action policy $\mu^* : b \rightarrow A_i$ is:

$$\mu^*(b) = argmax_{a \in A_i}[\sum_{s \in S}\sum_{s' \in S} R(s,a,s')b(s) + \gamma \sum_{o \in \Theta}\sum_{s \in S} O(o,s,a)b(s)V(b)]$$

. The above is further complicated by the presence of other agents since the $P(s|a_i, s')$ is also dependent on these agents' actions. We have:

$$P(s|a_i, s') = \sum_{a_{-i} \in A_{-i}} T(s', (a_1, ..., a_i, ...a_N), s)P(a_{-i}|M_{-i})$$

The above makes explicit the fact that the probability of changing the state of the system depends not only on the agent's i action, a_i, but also the joint action of the other agents, a_{-i}. To predict the actions of the other agents, i can use their models, which contributes the $P(a_{-i}|M_{-i})$ factor above.

The fact that the agent i has to use the models of the other agents to predict their likely actions, and only then compute its own optimal action is the essence of the decision-theoretic approach, as expressed above by the quote from Myerson (Myerson 1991).

Alternative Representation

If the information contained in the RMDP's is compiled into payoff matrices then recursive model structures, defined in (Gmytrasiewicz & Durfee 2000), result. As we discussed in (Gmytrasiewicz & Durfee 2000), the RMDP's are infinite, i.e., they accommodate infinite nesting of agents' beliefs, but they could be terminated by no-information models if the knowledge of the agent(s) is nested only to a finite level. In Figure 1 we have depicted a finite recursive model structure, depicting a state of knowledge of agent R_1 interacting with an agent R_2 as depicted in Figure 2 (see (Gmytrasiewicz & Durfee 2000) for details.)

The no-information models that terminate the recursive nesting in our example are at the leafs of the recursive model structure in Figure 1. These models represent the limits of the agents' knowledge: The model No-Info^2 represents the fact that, in the case when R_2 cannot see P2, R_1 knows that R_2 has no knowledge that would allow it to model R_1. Thus, the uncertainty is associated with R_2, and the model's superscript specifies that the state of no information is associated with its ancestor on the second level of the structure in Figure 1. The No-Info^1 model terminating the middle branch of the recursive structure represents R_1's own lack of knowledge (on the first level of the structure) of how it is being modeled by R_2, if R_2 can see through the trees. In general, the no-information models can represent knowledge limitations on any level; the limitations of R_1's own knowledge,[7] R_1's knowing the knowledge limitations of other agents, and so on.

[7]Note that we assume the agent can introspect. This amounts to the agent's being able to detect the lack of

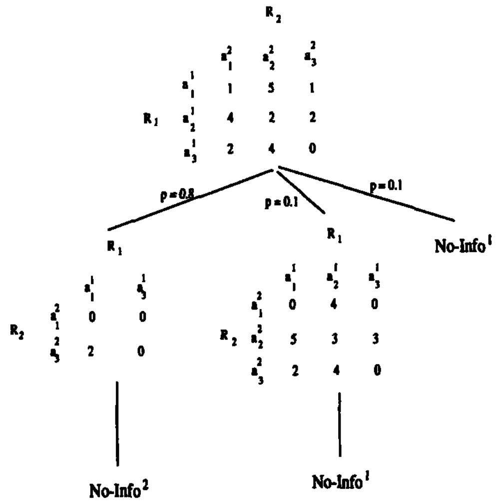

Figure 1: Recursive Model Structure depicting R_1's Decision-Making Situation in Example 1.

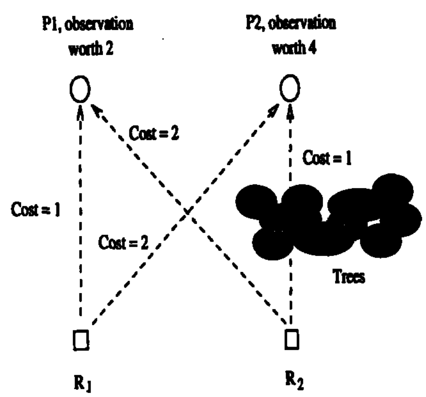

Figure 2: Example Scenario of Interacting Agents.

The no-information models are related to the problem of "only knowing", discussed in (Halpern 1993; Lakemeyer 1993) and related references.

In (Gmytrasiewicz, Noh, & Kellogg 1998; Suryadi & Gmytrasiewicz 1999) we show how the models and their probabilities, contained in the recursive model structure, can be updated based on the other agents' observed behavior using Bayes rule. Using the notation in the paper, this update is one that modifies the probability distribution over models of other agents, $P(\mathbf{M}_{-i})$, given observation of their behavior, o_{-i}, as:

$$P(M_{-i}|o_{-i}) = P(o_{-i}|M_{-i})\frac{P(M_{-i})}{P(o_{-i})} \qquad (5)$$

statements in its knowledge base that describe beliefs nested deeper than the given level.

In (Gmytrasiewicz & Durfee 2000) we described how dynamic programming can be used to solve an agent's i recursive model structure yielding the agent's optimal action. In (Gmytrasiewicz & Durfee 2000) we called the recursive model structure with the DP solution and Recursive Modeling Method (RMM). Dynamic programming is applicable because it is possible to express the solution to the problem of choice that maximizes expected utility on a given level of modeling in terms of the solutions to choices of the agents modeled on deeper levels. Thus, to solve the optimization problem on one level requires solutions to subproblems on the lower level. This means that the problem exhibits *optimal substructure* (Bellman 1957; Cormen, Leiserson, & Rivest 1990), and that a solution using dynamic programming can be formulated. The solution traverses the recursive model structure propagating the information bottom-up. The result is an assignment of expected utilities to the agent's alternative actions, based on all of the information the agent has at hand about the decision-making situation. The rational agent can then choose an action with the highest expected utility.

Clearly, the bottom-up dynamic programming solution requires that the recursive model structure be finite and terminate. Thus, we have to make the following assumption:

Assumption 1: *The recursive model structure, defined in Equation 1, is finite.*

The assumption above complements an assumption that the agents possess infinitely nested knowledge, called common knowledge or mutual knowledge, frequently made in AI and in traditional game theory. As we mentioned, these two assumptions lead to two solution concepts; one used in our work, which is decision-theoretic and implemented with dynamic programming, the other one based on the notion of equilibria (seen as fixed points of an infinite hierarchy of nested models.)

As we describe in (Gmytrasiewicz & Durfee 2000), the DP solution of the hierarchy depicted in Figure 1 yields a unique solution: The best choice for R_1 is to move toward point P2 and make an observation from there. It is the rational coordinated action given R_1's state of knowledge, since the computation included all of the information R_1 has about agent R_2's expected behavior. Intuitively, this means that R_1 believes that R_2 is so unlikely to go to P2 that R_1 believes it should go there itself.

If the interaction depicted in Figure 2 were to be sought using equilibria and assuming common knowledge, it would turn out that there are two equilibria, one corresponding to two possible assignments of tasks to agents. Were the number of tasks (or agents) larger, as in the air defense scenario below, the number of equilibria would be large, and the agents would be unable to decide which equilibrium to choose and what to do. This situation mirrors the problems described by Binmore in (Binmore 1994).

Some Experiments

Our air defense domain consists of some number of anti-air units whose mission is to defend a specified territory from a number of attacking missiles (see Figure 3). The defense units have to coordinate and decide which missiles to intercept, given the characteristics of the threat, and given what they can expect of the other defense units. The utility of the agents' actions in this case expresses the desirability of minimizing the damage to the defended territory. The threat of an attacking missile was assessed based on the size of its warhead and its distance from the defended territory. Further, the defense units considered the hit probability, $P(H)$, with which their interceptors would be effective against each of the hostile missiles. The product of this probability and a missile threat was the measure of the expected utility of attempting to intercept the missile.

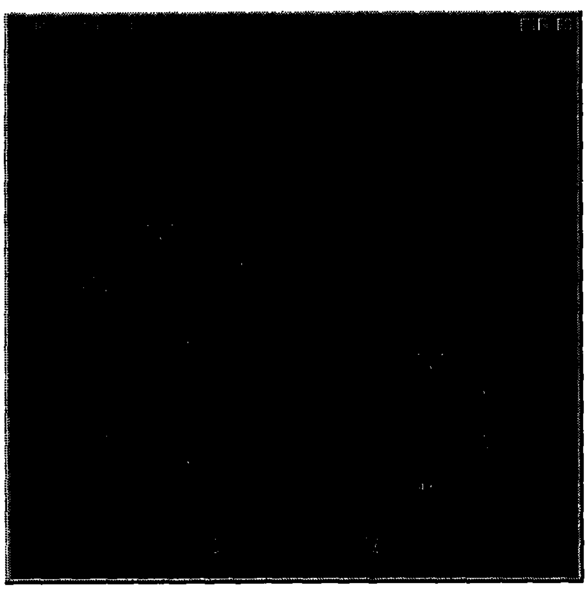

Figure 3: MICE Simulation of the Air Defense Domain.

As we mentioned, it is easy to see the advantage of using decision-theoretic approach to game theory as implemented in RMM vs. the traditional game-theoretic solution concept of equilibria. Apart from the need for common knowledge the agents have to share to justify equilibria, the problem is that there may be many equilibria and no clear way to choose the "right" one to guide the agent's behavior.

In all of the experiments we ran[8], each of two defense units could launch three interceptors, and were faced with an attack by six incoming missiles.

Our experiments was aimed at determining the quality of modeling and coordination achieved by the RMM agents in a team, when paired with human agents, and when compared to other strategies. To evaluate the quality of the agents' performance, the results were expressed in terms of (1) the number of intercepted targets, i.e., targets the defense units attempted to intercept, and (2) the total expected damage to friendly forces after all six interceptors were launched.

The target selection strategies are as follows:

- Random: selection randomly generated.
- Independent, no modeling: selection of arg $\max_j\{P(H_{ij}) \times T_j\}$ for agent i.
- Human:[9] selection by human.
- RMM: selection by RMM.

As shown in Figure 4 and Figure 5, we found that the all-RMM team outperformed the human and independent teams.

We found that the human performance was very similar to the performance of independent agents. The most obvious reason for this is that humans tend to depend on their intuitive strategies for coordination, and, in this case, found it hard to engage in deeper, normative, decision-theoretic reasoning. Sometimes the ways human subjects choose a missile were different and quite arbitrary. Some of them attempted to intercept the 3 left-most or right-most missiles, depending whether they were in charge of the left or the right defense battery. This led to difficulties when the missiles were clustered at the center area and to much duplicated effort. Others tended to choose missiles with the largest missile size. Still others tried to consider the multiplication of the missile size and the hit probability, but did not model the other agent appropriately. The performance of the RMM team was not perfect, however, since the agents were equipped with limited and uncertain knowledge of each other. The performance

[8] For an on-line demonstration of the air defense domain refer to the Web page http://dali.uta.edu/Air.html.

[9] We should remark that our human subjects were CSE and EE graduate students who were informed about the criteria for target selection. We would expect that anti-air specialists, equipped with a modern defense doctrine, could perform better than our subjects. However, the defense doctrine remains classified and was not available to us at this point.

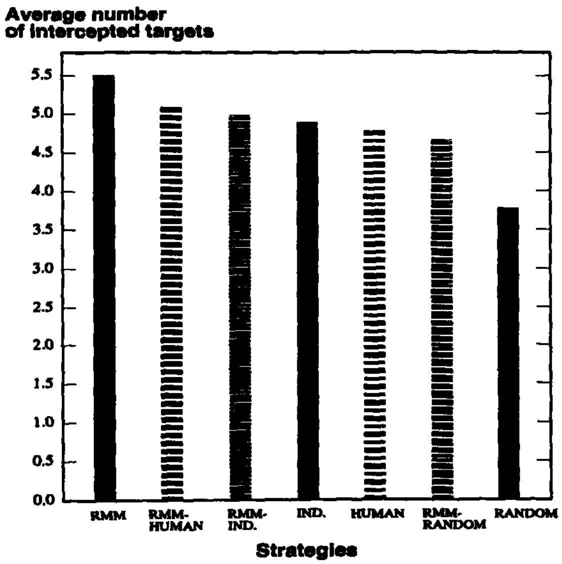

Figure 4: Average number of intercepted targets (over 100 runs).

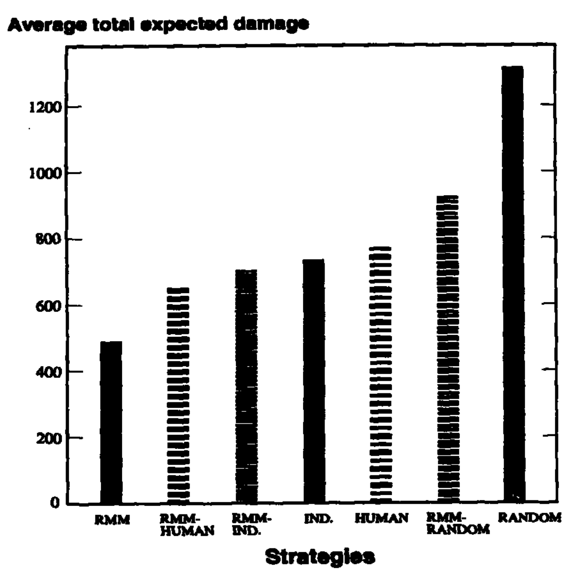

Figure 5: Average total expected damage (over 100 runs).

of the heterogeneous teams again also suggests the favorable quality of coordination achieved by RMM agents.

Conclusions

This paper proposed a decision-theoretic approach to game theory (DTGT) as a paradigm for designing agents that are able to intelligently interact and coordinate actions with other agents in multi-agent environments. We defined a general multi-agent version of Markov decision processes, called recursive MDP's, and illustrated assumptions, solution method, and an application of our approach. We argued that the DTGT approach is a viable alternative to the traditional game-theoretic approach based on equilibria.

References

Aumann, R. J., and Brandenburger, A. 1995. Epistemic conditions for Nash equilibrium. *Econometrica*.

Aumann, R. J. 1976. Agreeing to disagree. *Annals of Statistics* 4(6):1236–1239.

Aumann, R. J. 1999a. Interactive epistemology i: Knowledge. *International Journal of Game Theory* (28):263–300.

Aumann, R. J. 1999b. Interactive epistemology i: Probability. *International Journal of Game Theory* (28):301–314.

Bellman, R. E. 1957. *Dynamic Programming*. Princeton University Press.

Binmore, K. 1982. *Essays on Foundations of Game Theory*. Pitman.

Binmore, K. 1994. Rationality in the centipede. In *Proceedings of the Conference on Theoretical Aspects of Reasoning about Knowladge*, 150–159. Morgan Kaufman.

Boutilier, C.; Dean, T.; and Hanks, S. 1999. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial intelligence Research* 11:1–94.

Boutilier, C. 1999. Sequential optimality and coordination in multiagent systems. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 478–485.

Brandenburger, A. 1992. Knowledge and equilibrium in games. *Journal of Economic Perspectives* 6:83–101.

Camerer, C. 1995. Individual decision making. In Kagel, J. H., and Roth, A. E., eds., *The Handbook of Experimental Economics*. Princeton University Press.

Cheeseman, P. 1985. In defense of probability. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1002–1009.

Choo, I.-K., and Matsuri, A. 1992a. Induction and bounded rationality in repeated games. Technical report, CARESS Working Paper 92-16, University of Pennsylvania.

Choo, I.-K., and Matsuri, A. 1992b. Learning and the ramsey policy. Technical report, CARESS Working Paper 92-18, University of Pennsylvania.

Coles, L. S.; Robb, A. M.; Sinclar, P. L.; Smith, M. H.; and Sobek, R. R. 1975. Decision analysis for an experimental robot with unreliable sensors. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*.

Cormen, T. H.; Leiserson, C. E.; and Rivest, R. L. 1990. *Introduction to Algorithms.* The MIT Press.

Doyle, J. 1992. Rationality and its role in reasoning. *Computational Intelligence* 8:376–409.

Fagin, R. R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning About Knowledge.* MIT Press.

Fagin, R. R.; Halpern, J. Y.; and Vardi, M. Y. 1991. A model-theoretic analysis of knowledge. *Journal of the ACM* (2):382–428.

Feldman, J. A., and Sproull, R. F. 1977. Decision theory and Artificial Intelligence II: The hungry monkey. *Cognitive Science* 1(2):158–192.

Fishburn, P. 1981. *The Foundations of Expected Utility.* D. Reidel Publishing Company.

Fudenberg, D., and Tirole, J. 1991. *Game Theory.* MIT Press.

Geanakoplos, J. 1992. Common knowledge. In *Proceedings of the Conference on Theoretical Aspects of Reasoning about Knowladge*, 255–315. Morgan Kaufman.

Gmytrasiewicz, P. J., and Durfee, E. H. 1992. Logic of knowledge and belief for recursive modeling: Preliminary report. In *Proceedings of the National Conference on Artificial Intelligence*, 628–634.

Gmytrasiewicz, P. J., and Durfee, E. H. 2000. Rational coordination in multi-agent environments. *Autonomous Agents and Multiagent Systems Journal* 3(4):319–350.

Gmytrasiewicz, P. J.; Noh, S.; and Kellogg, T. 1998. Bayesian update of recursive agent models. *International Journal of User Modeling and User-Adapted Interaction, Special Issue on Machine Learning for User Modeling.*

Haddawy, P., and Hanks, S. 1990. Issues in decision-theoretic planning: Symbolic goals and numeric utilities. In *Proceedings of the 1990 DARPA Workshop on Innovative Approaches to Planning, Scheduling, and Control*, 48–58.

Halpern, J. Y., and Moses, Y. 1990. Knowledge and common knowledge in a distributed environment. *Journal of the ACM* 37(3):549–587.

Halpern, J. Y. 1993. Reasoning about only knowing with many agents. In *Proceedings of the National Conference on Artificial Intelligence*, 655–661.

Harsanyi, J. C. 1967. Games with incomplete information played by 'Bayesian' players. *Management Science* 14(3):159–182.

Hauskrecht, M. 2000. Value-function approximations for partially observable markov decision processes. *Journal of Artificial Intelligence Research* (13):33–94.

Jacobs, W., and Kiefer, M. 1973. Robot decisions based on maximizing utility. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, 402–411.

Kadane, J. B., and Larkey, P. D. 1982. Subjective probability and the theory of games. *Management Science* 28(2):113–120.

Kandori, M., and Rob, R. 1991. Evolution of equilibria in the long run: A general theory and applications. Technical report, CARESS Working Paper 91-01R, University of Pennsylvania.

Kandori, M.; Mailath, G. J.; and Rob, R. 1991. Learning, mutation and long run equilibria in games. Technical report, CARESS Working Paper 91-01R, University of Pennsylvania.

Koller, D., and Milch, B. 2002. Multi-agent influence diagrams for representing and solving games. In *Seventeenth International Joint Conference on Artificial Intelligence*, 1027–1034.

Lakemeyer, G. 1993. All they know about. In *Proceedings of the National Conference on Artificial Intelligence.*

Malmnas, P.-E. 1994. Axiomatic justification of the utility principle: A formal investigation. *Synthese* (99):233–249.

Mertens, J.-F., and Zamir, S. 1985. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory* 14:1–29.

Myerson, R. B. 1991. *Game Theory: Analysis of Conflict.* Harvard University Press.

Neapolitan, R. E. 1990. *Probabilistic Reasoning in Expert Systems.* John Wiley and Sons.

Newell, A. 1981. The knowledge level. *AI Magazine* 2(2):1–20.

Nilsson, N. J. 1971. *Problem-Solving Methods in Artificial Intelligence.* McGraw-Hill.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman.

Raiffa, H. 1982. *The Art and Science of Negotiation.* Harvard University Press.

Reny, P. J. 1988. Extensive games and common knowledge. In *Proceedings of the Conference on Theoretical Aspects of Reasoning about Knowladge*, 395. Morgan Kaufman.

Russell, S., and Norvig, P. 1995. *Artificial Intelligence: A Modern Approach.* Prentice Hall.

Smith, J. M. 1982. *Evolution and the theory of games.* Cambrigde University Press.

Suryadi, D., and Gmytrasiewicz, P. J. 1999. Learning models of other agents using influence diagrams. In *Preceedings of the 1999 International Conference on User Modeling*, 223–232.

Tan, T. C., and Werlang, S. R. 1988. A guide to knowledge and games. In *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge.*