On the Role Of Interactive Epistemology in Multiagent Planning

Prashant Doshi
Department of Computer Science
University of Georgia
Athens, GA 30602
pdoshi@cs.uga.edu

Abstract

This paper focuses on the foundational role of interactive epistemology in the problem of generating plans for rational agents in multiagent settings. Interactive epistemology deals with the logic of knowledge and belief when there is more than one agent. In multiagent settings, we are interested in not only the agent's knowledge of the state of the world, but also its belief over the other agents' beliefs and their beliefs over others'. We adopt a probabilistic approach for formalizing the epistemology. This paper attempts to answer the question of why we should study the interactive epistemology of agents within the context of multiagent planning. In doing so, it motivates the need for a more detailed examination of the epistemological foundations of multiagent planning. We conclude this paper with a framework for multiagent planning that explicitly constructs and reasons with nested belief structures.

1 Introduction

Agents that plan are often only partially aware of the state of the world they inhabit. Perfect knowledge of the state though plausible is realistically rare. However, in the event that we do have perfect knowledge of the state, planning involves deciding on a sequence of steps from the current state that optimize the agent's preferences or rewards; this is not inordinately expensive and may be performed in polynomial time [16]. More relevant is the case where the agent is uncertain of the state of its environment and must plan a sequence of steps that optimize its preferences.

We may adopt a Bayesian approach and assume that the agent has an *a'priori belief* about the uncertain state of the environment. An agent's belief is a probability distribution over the space of states (assumed discrete) and indicates how likely is it that a particular state is the current true state of the environment. For example, if an agent has no information about the state, its belief may be a uniform distribution. As the agent acts and makes observations, its beliefs are updated and evolve over time. In order to act optimally, the agent decides on a sequence of steps that optimize its expected preferences given its current beliefs. Because the beliefs are a sufficient statistic for the history of the agent's observations, they are adequate for computing the optimal actions. While realistic, the planning problem is computationally complex – PSPACE-Complete [16] for a finite sequence of actions and undecidable if there is no limit on the number of lookahead steps. This type of single agent planning is usually studied within the framework of partially observable Markov decision processes (POMDPs) [13, 18].

The problem of planning in multiagent settings is more complex. This is because the outcomes in terms of the state and rewards are determined by the joint actions of all agents. [1] Without assumptions on the behaviors of the other agents, optimizing the agent's beliefs over the physical state is insufficient for guaranteeing an optimal plan. [2] One such assumption – that forms the cornerstone of much of the research in game theory – is that of *rationality*. In particular, the assumption is that all agents are assumed to be rational, everybody knows that everybody is rational, everybody knows that everbody knows that all agents are rational and so on, *ad infinitum*. [3] This assumption provides the required epistemic condition for the adoption of plans that are in Nash equilibrium by the agents. In a general setting where agents could be of different *types* [11], an additional epistemic condition – the common knowledge of a prior probability distribution over the types of all agents – is necessary for arriving at a (Bayesian) Nash equilibrium. In other words, given that we have common knowledge of rationality (and a prior over the agent types), we may solve

[1] While situations may exist in which the state and the reward are individually determined, these may be seen as special cases of the more general problem that we study here.

[2] One reason for this is that in a multiagent setting, an agent's belief over just the state is not a sufficient statistic.

[3] This assumption is also called the common knowledge of rationality. See [1, 8] for a formalization of common knowledge.

the multiagent planning problem by computing plans that are in (Bayesian) Nash equilibrium. Indeed, Nash equilibrium allows us to cut through the recursive reasoning that an agent's optimal action depends on what the other agents' actions are, which in turn depend on the other agents' actions and so on.

Despite the obvious advantages of equilibrium, adopting it as a solution concept has several drawbacks: First, the self-referential nature of the necessary epistemic condition makes it impossible to model it computationally. Second, there could be multiple Nash equilibria with no general way to choose between them. Third, Nash equilibrium does not prescribe what to do if the others fail to follow their part of the equilibrium. In light of these arguments, we adopt the decision-theoretic approach that an agent in a multiagent setting should decide on a sequence of actions that optimize its preferences given its beliefs. In addition to having beliefs about the state (which, from here onwards we will refer to as the physical state for clarity), the agent has beliefs over the possible behavioral models of the other agents as well. As the agent acts and receives observations, its beliefs over both the physical state and the others' models are updated over time. This approach, which has been previously referred to as subjective rationality [14] and a decision-theoretic approach to game theory [12], has however received scant attention in the game theoretic literature.

Rest of this paper is structured as follows. In the next section, we uncover the epistemology that is inherent in interactions in multiagent settings. In Section 3 we briefly review a framework for individual planning in multiagent settings that explicitly constructs and deals with recursive belief structures that naturally occur in multiagent settings. Finally, in Section 4, we discuss the epistemological commitments that are often assumed of agents and the role they play in planning.

2 Interactive Epistemology

While our approach appears intuitively simple, admittedly it hides the complexity. In the next subsection, we illustrate how assumptions about others' behaviors affects rational choice. In Section 2.2, we uncover the epistemology that is crucial to interactions and mathematically formalize it. We briefly review supporting work thereafter.

2.1 An Example: The Centipede Game

The impact of recursive reasoning on an agent's rational action is aptly demonstrated in the two-agent Centipede game, first introduced by Rosenthal [17]. We show a simple version of the game in Fig. 1. Agents A and B alternate in choosing, at each decision node, whether to defect by moving down or cooperate by moving across the game. If agent A defects, then the game stops at that point and the rewards to the two agents are shown in the parenthesis. Whenever an agent moves across, it incurs a penalty of 1 unit while the other agent gains a reward of 10 units. Notice that both agents stand to gain large rewards if they cooperate.

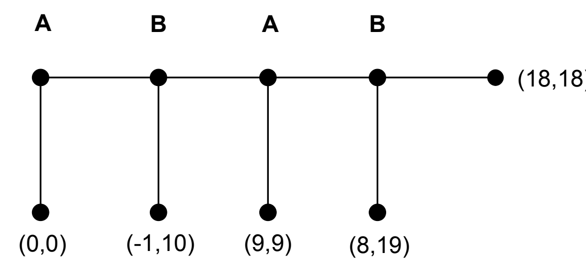

Figure 1. The two-agent Centipede game.

We are interested in the rational choice of agent A. If agent A believes that B will defect during its turn, then A will choose defection over cooperation. However, for B to choose defection, it must believe that A will defect in its turn, who in turn must believe that B being a rational player will choose defection in the final move of the game. Thus, if the depth of A's recursive reasoning is identical to the length of the Centipede game and A believes that B is rational, A will choose defection over cooperation.

On the other hand, if agent A is unsure of B's rationality and believes that B is equally likely to cooperate or defect, then it will choose cooperation. This is because in its second move, it will choose to cooperate since its expected reward from cooperation is 13 as compared to 9 from defection. As B is equally likely to defect or cooperate, A's expected reward from cooperating in the initial move is 6 compared to -1 from defection, thereby choosing cooperation.

2.2 A Recursive Characterization of the State Space

In our decision-theoretic approach, though a wide variety of behavioral models could be ascribed to agents, let us consider the particular case in which an agent ascribes *intentional* models to the others. These models ascribe beliefs, capabilities, preferences and Bayesian rationality to the other agents, and are therefore analogous to POMDPs. For the sake of clarity, let us consider a two-agent setting that is populated by agents i and j. Formally, let Θ_j be the set of intentional models of j where $\theta_j \in \Theta_j$ is defined as: $\theta_j = \langle\, b_j, \hat{\theta}_j \,\rangle$. Here, b_j is agent j's belief which is a probability distribution over the states, $b_j \in \Delta(S)$, $\Delta(\cdot)$ denotes the set of probability distribution and S is the set of states, and $\hat{\theta}_j$ denotes the remaining parameters of the model including the capabilities and preferences [4].

As we mentioned before, we expand the state space to include the intentional models as well, and call the new state space as the *interactive* states of the agent i, $IS_i = S \times \Theta_j$, where S is the physical state space. [5] Agent i's belief is now a distribution over the interactive state space, $b_i \in \Delta(IS_i)$. Notice that because the interactive state space consists of the other's intentional models which in turn include the other's beliefs, agent i's belief is a distribution over j's beliefs. Also notice that j could itself be reasoning in a similar way, therefore, j's belief could itself be a probability distribution over i's beliefs. This argument may be applied recursively. We formalize this recursive description of the interactive state space next.

Consider a space of physical states S. Agent j's belief over S is then a probability distribution over S. We will call this a 0^{th} level belief, $b_{j,0}$. Agent j's 0^{th} level model is then $\theta_{j,0} = \langle b_{j,0}, \hat{\theta}_j \rangle$. The 0^{th} level models allow us to cap the otherwise self-referential epistemology of the agents. These models could be POMDPs with the other agent's actions folded in as *noise* into the T, O, and R functions. Agent i's first level beliefs are defined over the physical states of the world and the 0^{th} level models of agent j. The second level beliefs are distributions over the physical state and agent j's 1^{st} level models. We give the recursive bottom-up construction of the state spaces below:

$$\begin{array}{lll} IS_{i,0} & = S, & \Theta_{j,0} = \{\langle b_{j,0}, \widehat{\theta}_j \rangle : b_{j,0} \in \Delta(IS_{j,0})\}, \\ IS_{i,1} & = S \times \Theta_{j,0}, & \Theta_{j,1} = \{\langle b_{j,1}, \widehat{\theta}_j \rangle : b_{j,1} \in \Delta(IS_{j,1})\}, \\ IS_{i,2} & = S \times \Theta_{j,1}, & \Theta_{j,2} = \{\langle b_{j,2}, \widehat{\theta}_j \rangle : b_{j,2} \in \Delta(IS_{j,2})\}, \\ \vdots & & \vdots \end{array} \tag{1}$$

Possible infinite nestings of agents' beliefs in intentional models present an obvious obstacle to computing the belief updates and plans. Models with infinite nestings correspond to agent functions that are not computable. [6] Therefore, it is only natural that we look toward bounding the nestings to finite levels. Let l be the finite level of recursion, and $IS_{i,l}$ and $\Theta_{j,l}$ be the l^{th} level interactive state space and set of intentional models, respectively. We define it as:

$$IS_{i,l} = S \times \Theta_{j,l-1}, \quad \Theta_{j,l} = \{\langle b_{j,l}, \widehat{\theta}_j \rangle : b_{j,l} \in \Delta(IS_{j,l})\}.$$

We observe that limiting the agent's recursive beliefs to a finite nesting level is an approximation of the full optimality (or rationality) of the agent. However, this approximation becomes necessary in light of the computational ramifications.

We address the remaining issue of focusing solely on intentional models. While settings populated by rational agents have, by far, received the greater attention from researchers, the possibility that the other agent could be subintentional cannot be ruled out. [7] Let $M_{j,l}$ be the general space of all computable models inclusive of the level l intentional models. In our previous analysis, $M_{j,l} = \Theta_{j,l}$. However, if SM_j is the set of subintentional models, then $M_{j,l} = \{ \Theta_{j,l} \cup SM_j \}$. We may replace Θ_j in Eq. 1 with M_j to obtain a recursive characterization of the interactive state space that includes subintentional models as well. Considering subintentional models has the additional benefit that it allows the epistemology to transcend from a normative one to being descriptive as well.

2.3 Related Work

The foundational role of epistemology in multiagent interactions and the need to study it has been recognized previously. Recursive characterizations of state spaces analogous to the one in Eq. 1 have appeared previously in the game theoretic literature [3, 7, 15] where they have led to the definitions of *hierarchical belief systems*. Hierarchical beliefs have been offered as mathematical formalizations of the agent *types* that were introduced by Harsanyi [11] to make incomplete information games tractable. This is because under the assumption of coherency of the beliefs, a hierarchical belief system induces a distribution over the possible types of the other agents. Prior to this line of work, Aumann [1, 2] formally studied the interactive epistemology of agents and provided a set-theoretic formalization of the knowledge and beliefs of agents using the concept of information partitions.

3 Multiagent Planning Using Interactive POMDPs

In this section, we briefly review a framework called the Interactive POMDP (I-POMDP) [9, 10] that generalizes POMDPs to multiagent settings. I-POMDPs explicitly model and reason with the nested belief systems shown in Eq. 1 made computable by considering a finite nesting level, l. I-POMDPs do not assume the common knowledge of rationality or the common knowledge of a prior probability distribution over the types of all agents. Instead, they adopt a Bayesian approach according to which an agent maintains a belief over the physical state as well as the candidate models of the other interacting agents.

---

[4] If the model is a POMDP, then $\hat{\theta}_j = \langle A_j, \Omega_j, T_j, O_j, R_j, OC_j \rangle$, where the parameters have their usual meanings.

[5] If there are more agents interacting with agent i, $K > 2$, then $IS_i = S \times_{j=1}^{K-1} \Theta_j$

[6] An agent function is a mapping from the agent's observation history to a distribution over its actions: $\Omega_j^* \rightarrow \Delta(A_j)$. Because there are uncountably infinite of these of which only countably infinite are computable, a large proportion of agent functions are not computable.

[7] This is especially relevant since we no longer assume the common knowledge of rationality.

3.1 Definition

We will limit the discussion to intentional models, analogous to *types* in Bayesian games, which include all private information influencing an agent's behavior. However, I-POMDPs may be easily extended to include subintentional models as well. For simplicity of presentation, let us consider an agent, i, that is interacting with one other agent, j.

I-POMDP A *finitely nested interactive POMDP* of agent i, *I-POMDP*$_{i,l}$, is:

$$\textit{I-POMDP}_{i,l} = \langle IS_{i,l}, A, T_i, \Omega_i, O_i, R_i \rangle$$

where:

• $IS_{i,l}$ denotes a set of interactive states defined as, $IS_{i,l} = S \times \Theta_{j,l-1}$, for $l \geq 1$, and $IS_{i,0} = S$, where S is the set of states of the physical environment, and $\Theta_{j,l-1}$ is the set of $(l-1)^{th}$ level *intentional models* of agent j: $\theta_{j,l-1} = \langle b_{j,l-1}, A, \Omega_j, T_j, O_j, R_j, OC_j \rangle$. $b_{j,l-1}$ is the agent j's belief nested to the level $(l-1)$ and OC_j is j's optimality criterion. Rest of the notation is standard. For the sake of convenience, let us rewrite $\theta_{j,l-1}$ as, $\theta_{j,l-1} = \langle b_{j,l-1}, \widehat{\theta}_j \rangle$, where $\widehat{\theta}_j \in \widehat{\Theta}_j$ includes all elements of the intentional model other than the belief and is called the agent j's *frame*. A detailed inductive definition of the interactive state space was shown in Section 2.

• $A = A_i \times A_j$ is the set of joint moves of all agents, where A_i and A_j are the actions of agents i and j respectively.

• T_i is a transition function, $T_i : S \times A \times S \rightarrow [0, 1]$ which describes results of agents' actions

It is assumed that actions can change only the physical state of the world, see [9].)

We assume that actions do not directly change the other agent's models.

• Ω_i is the set of agent i's observations.

• O_i is an observation function, $O_i : S \times A \times \Omega_i \rightarrow [0, 1]$ which gives the likelihood of receiving an observation from a particular state given the joint moves of all agents.

• R_i is defined as, $R_i : IS_i \times A \rightarrow \mathbf{R}$. While an agent is allowed to have preferences over physical states and models of other agents, usually only the physical state will matter.

A more detailed definition of this framework and its properties appears in [9]. Henceforth, in this paper, we will refer to finitely nested I-POMDPs as simply I-POMDPs.

3.2 Belief Update

There are two differences that complicate a belief update in multiagent settings, when compared to single-agent ones. First, since the state of the physical environment depends on the actions performed by both agents, the prediction of how the physical state changes has to be made based on the predicted actions of the other agent. The probabilities of other's actions are obtained based on its models. Second, changes in the models of the other agent – update of the other agent's beliefs due to its new observation – has to be included. In other words, the agent has to update its beliefs based on what it anticipates that the other agent observes and how it updates. Formally, we have:

$$\begin{aligned} & Pr(is^t|a_i^{t-1}, b_{i,l}^{t-1}) = \beta \textstyle\sum_{IS^{t-1}:\widehat{m}_j^{t-1}=\widehat{\theta}_j^t} b_{i,l}^{t-1}(is^{t-1}) \\ & \times \textstyle\sum_{a_j^{t-1}} Pr(a_j^{t-1}|\theta_{j,l-1}^{t-1}) O_i(s^t, a_i^{t-1}, a_j^{t-1}, o_i^t) \\ & \times T_i(s^{t-1}, a_i^{t-1}, a_j^{t-1}, s^t) \textstyle\sum_{o_j^t} O_j(s^t, a_i^{t-1}, a_j^{t-1}, o_j^t) \\ & \times \delta_D(SE_{\widehat{\theta}_j^t}(b_{j,l-1}^{t-1}, a_j^{t-1}, o_j^t) - b_{j,l-1}^t) \end{aligned} \tag{2}$$

where β is the normalizing constant, δ_D is the Dirac-delta function, $SE(\cdot)$ is an abbreviation for the belief update, and $Pr(a_j^{t-1}|\theta_{j,l-1}^{t-1})$ is the probability that a_j^{t-1} is Bayes rational for the agent described by model $\theta_{j,l-1}^{t-1}$. [8] The proof is in [9].

For better understanding of the belief update, we decompose the I-POMDP belief update (Eq. 2) into two steps:

• *Prediction:* When an agent, say i, performs an action a_i^{t-1}, and agent j performs a_j^{t-1}, the predicted belief state is,

$$\begin{aligned} & Pr(is^t|a_i^{t-1}, a_j^{t-1}, b_{i,l}^{t-1}) = \textstyle\sum_{IS^{t-1}:\widehat{\theta}_j^{t-1}=\widehat{\theta}_j^t} b_{i,l}^{t-1}(is^{t-1}) \\ & \times Pr(a_j^{t-1}|\theta_{j,l-1}^{t-1}) T_i(s^{t-1}, a_i^{t-1}, a_j^{t-1}, s^t) \\ & \times \textstyle\sum_{o_j^t} O_j(s^t, a_i^{t-1}, a_j^{t-1}, o_j^t) \\ & \times \delta_D(SE_{\widehat{\theta}_j^t}(b_{j,l-1}^{t-1}, a_j^{t-1}, o_j^t) - b_{j,l-1}^t) \end{aligned}$$

• *Correction:* When agent i perceives an observation, o_i^t, the corrected belief state is a weighted sum of the predicted belief states for each possible action of j,

$$\begin{aligned} & Pr(is^t|o_i^t, a_i^{t-1}, b_{i,l}^{t-1}) = \beta \textstyle\sum_{a_j^{t-1}} O_i(s^t, a_i^{t-1}, a_j^{t-1}, o_i^t) \\ & \times Pr(is^t|a_i^{t-1}, a_j^{t-1}, b_{i,l}^{t-1}) \end{aligned}$$

where β is the normalizing constant.

If j is also modeled as an I-POMDP, then i's belief update invokes j's belief update (via the term $SE_{\widehat{\theta}_j^t}(b_{j,l-1}^{t-1}, a_j^{t-1}, o_j^t)$), which in turn invokes i's belief update and so on. This recursion in belief nesting bottoms out at the 0^{th} level. At this level, belief update of the agent reduces to a POMDP belief update. [9] For additional details on I-POMDPs, and how they compare with other multiagent planning frameworks, see [9].

---

[8] When j's model is sub-intentional, the integration is over $IS^{t-1} : \widehat{m}_j^{t-1} = \widehat{m}_j^t$, $Pr(a_j^{t-1}|\theta_{j,l-1}^{t-1})$ is replaced with $Pr(a_j^{t-1}|m_{j,l-1}^{t-1})$, and $\delta_D(SE_{\widehat{\theta}_j^t}(b_{j,l-1}^{t-1}, a_j^{t-1}, o_j^t) - b_{j,l-1}^t)$ is replaced with $\delta_K(APPEND(h_j^{t-1}, o_j^t) - h_j^t)$. δ_K is the Kronecker delta, and $APPEND$ returns a string with the second argument appended to the first.

[9] The 0^{th} level model is a POMDP: other agent's actions are treated as exogenous events and folded into T, O, and R.

The fact that Eq. 2 expresses the belief in terms of parameters of the previous time step only shows that an agent's belief over the physical state and the other agent's models is a sufficient statistic (Proposition 1).

Proposition 1. (Sufficiency) *In a finitely nested I-POMDP of agent i, i's current belief, i.e., the probability distribution over the set $S \times \Theta_{j,l-1}$, is a sufficient statistic for the past history of i's observations.*

The proof of Proposition 1 is given in [9]. As we mentioned before, Proposition 1 implies that the updated beliefs are sufficient for deciding on an optimal sequence of actions. A way of doing this is shown next.

3.3 Solution

Each level l belief state in I-POMDP has an associated value reflecting the maximum payoff the agent can expect in this belief state:

$$V^t(\langle b_{i,l}, \widehat{\theta}_i \rangle) = \max_{a_i \in A_i} \left\{ \sum_{is} ER_i(is, a_i) b_{i,l}(is) + \gamma \sum_{o_i \in \Omega_i} Pr(o_i | a_i, b_{i,l}) V^{t-1}(\langle SE_{\widehat{\theta}_i}(b_{i,l}, a_i, o_i), \widehat{\theta}_i \rangle) \right\} \quad (3)$$

where, $ER_i(is, a_i) = \sum_{a_j} R_i(is, a_i, a_j) Pr(a_j | \theta_{j,l-1})$ (since $is = (s, \theta_{j,l-1})$). Eq. 3 is a basis for value iteration in I-POMDPs, and may be implemented using dynamic programming.

Agent i's optimal action, a_i^*, for the case of finite horizon with discounting, is an element of the set of optimal actions for the belief state, $OPT(\theta_i)$, defined as:

$$OPT(\langle b_{i,l}, \widehat{\theta}_i \rangle) = \underset{a_i \in A_i}{argmax} \left\{ \sum_{is} ER_i(is, a_i) b_{i,l}(is) + \gamma \sum_{o_i \in \Omega_i} Pr(o_i | a_i, b_{i,l}) V^{t-1}(\langle SE_{\widehat{\theta}_i}(b_{i,l}, a_i, o_i), \widehat{\theta}_i \rangle) \right\}$$

4 Discussion

Interactive epistemology is the study of knowledge and beliefs of agents participating in an interaction. Because the outcomes of the interactions in terms of the physical state and the reward depend on the joint actions of all agents, assumptions about the behaviors of the agents assume significance. In this regard, a common assumption has been the common knowledge of rationality of all agents, which is also one of the necessary epistemic conditions for adopting Nash equilibrium as a solution concept. The other is the knowledge of the rewards of the other agents (or the distribution over possible reward functions in imperfect information settings). If the agents are Bayesian, their rational behaviors are conditioned on their beliefs. Thus knowledge about the beliefs of others gains importance in interactive settings. Because other agents' beliefs are private and therefore not perfectly observable, agents would have beliefs over others' beliefs. This argument may be applied recursively resulting in the so called hierarchical belief systems.

While the importance of belief systems is well established, their computability has long been suspect. [10] We have sought to make them tractable by limiting our attention to finitely nested beliefs which precludes the self-references that have formed the basis of the arguments of non-computability. However, this makes planning using finitely nested belief systems necessarily suboptimal. Indeed, there is some evidence that *optimal* decision-making in interactive settings is computationally impossible [4].

The I-POMDP framework for individual planning in multiagent settings is a step in this direction. By showing how we may revise an agent's belief at all levels of the finitely nested belief system, it offers an approach for deciding on a sequence of actions that is optimal given the agent's nested belief system. The approach contributes to a growing focus on dynamic interactive epistemology [5] which studies not only agents' static knowledge and beliefs but also how the knowledge and beliefs are revised as the agents act and observe.

Acknowledgments

This research is supported by a grant from UGA's research foundation. The work has benefited from past discussions with Piotr Gmytrasiewicz.

References

[1] R. J. Aumann. Interactive epistemology i: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

[2] R. J. Aumann. Interactive epistemology ii: Probability. *International Journal of Game Theory*, 28:301–314, 1999.

[3] P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1):188–230, 1999.

[4] K. Binmore. *Essays on Foundations of Game Theory*. Pittman, 1982.

[5] O. Board. Dynamic interactive epistemology. *Journal of Games and Economic Behavior*, 49:49–80, 2004.

[6] A. Brandenburger. The power of paradox: Some recent developments in interactive epistemology. discussion paper, 2002.

[7] A. Brandenburger and E. Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, 59:189–198, 1993.

[8] R. Fagin, J. Geanakoplos, J. Halpern, and M. Vardi. The hierarhical approach to modeling knowledge and common knowledge. *International Journal of Game Theory*, 28, 1999.

[10] See [6] for an impossiblity result on the existence of complete belief systems.

[9] P. Gmytrasiewicz and P. Doshi. A framework for sequential planning in multiagent settings. *JAIR*, 24:49–79, 2005.

[10] P. J. Gmytrasiewicz and P. Doshi. A framework for sequential planning in multi-agent settings. In *AMAI*, Ft. Lauderdale, Florida, 2004.

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

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

[13] L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence Journal*, 2, 1998.

[14] E. Kalai and E. Lehrer. Rational learning leads to nash equilibrium. *Econometrica*, 61(5):1019–1045, 1993.

[15] J. Mertens and S. Zamir. Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.

[16] C. Papadimitriou and J. Tsitsiklis. The complexity of markov decision processes. *Mathematical Journal of Operations Research*, 12(3):441–450, 1987.

[17] R. Rosenthal. Games of perfect information, predatory pricing, and the chain store paradox. *Journal of Economic Theory*, 25:92–100, 1981.

[18] R. Smallwood and E. Sondik. The optimal control of partially observable markov decision processes over a finite horizon. *Operations Research*, 21:1071–1088, 1973.