On Markov Games Played by Bayesian and Boundedly-Rational Players

Muthukumaran Chandrasekaran
THINC Lab, University of Georgia
Athens, GA, USA
mkran@uga.edu

Yingke Chen
Sichuan University
Chengdu, China
yke.chen@gmail.com

Prashant Doshi
THINC Lab, University of Georgia
Athens, GA, USA
pdoshi@cs.uga.edu

Abstract

We present a new game-theoretic framework in which Bayesian players with bounded rationality engage in a Markov game and each has private but incomplete information regarding other players' types. Instead of utilizing Harsanyi's abstract types and a common prior, we construct intentional player types whose structure is explicit and induces a *finite-level* belief hierarchy. We characterize an equilibrium in this game and establish the conditions for existence of the equilibrium. The computation of finding such equilibria is formalized as a constraint satisfaction problem and its effectiveness is demonstrated on two cooperative domains.

Introduction

A plethora of empirical findings in strategic games (Stahl and Wilson 1995; Hedden and Zhang 2002; Wright and Leyton-Brown 2010; Goodie, Doshi, and Young 2012) strongly suggest that humans reason about others' beliefs to finite and often low depths. In part, this explains why a significant proportion of participants do not play Nash equilibrium profiles of games (Camerer 2003) because reasoning about Nash play requires thinking about the other player's beliefs and actions, and her reasoning about other's, and so on *ad infinitum*. Such reasoning is generally beyond the cognitive capacity of humans. This has motivated models of finitely-nested reasoning such as the cognitive hierarchy and others (Ho and Su 2013).

Are there characterizations of equilibrium between players engaging in finite levels of inter-personal reasoning? Aumann (1999) introduced the information partition as a basic element for representing a player's knowledge. Recently, Kets (2014) elegantly generalized the standard Harsanyi framework for single-stage games of incomplete information to allow for players' partitions that lead to finite-level beliefs. Any found equilibrium in this framework is also a Bayes-Nash equilibrium (BNE) in a Harsanyi framework. However, as we may expect, not every BNE for the game is also an equilibrium between players with finite-level beliefs because some BNE necessitate infinite-level beliefs.

We generalize the single-stage framework of Kets to allow Bayesian players to play an incomplete-information *Markov game* (Littman 1994; Maskin and Tirole 2001). These sophisticated games are finding appeal as a theoretical framework for situating pragmatic interactions such as impromptu or ad hoc teamwork (Albrecht and Ramamoorthy 2013). Each player in our framework may have one of many types – *explicitly* defined unlike the abstract ones in the Harsanyi framework – and which *induces* a belief hierarchy of finite depth. Contextual to such finite-level types in this new framework of Bayesian Markov games (BMG) with explicit types, we formally define a Markov-perfect finite-level equilibrium, establish conditions for its existence, and present a method for obtaining this equilibrium. We formulate finding equilibrium in a BMG as a constraint satisfaction problem. For this, we generalize the constraint satisfaction algorithm introduced by Soni et al. (2007) for finding BNE in Bayesian graphical games. Key challenges for the generalization are that the space of types is continuous and the beliefs in each type must be updated based on the observed actions of others. This makes the types dynamic. Because strategies may be *mixed* and standard constraint satisfaction algorithms do not operate on continuous domains, we discretize the continuous space of mixed strategies analogously to Soni et al. (2007). Finally, motivated by behavioral equivalence (Zeng and Doshi 2012), we use equivalence between types for speed up.

In addition to presenting a new framework and discussing its properties, we demonstrate it on two cooperative tasks. Equilibrium has traditionally served as an important baseline behavior for agents engaged in cooperation, providing a locally-optimal solution (Nair et al. 2003; Roth, Simmons, and Veloso 2006). BMGs with explicit types are particularly well suited toward analyzing interactions between agents that are boundedly rational and are uncertain about each other's type.

Background

Inspired by the real-world transport logistics domain, we motivate our framework using the level-based foraging problem (Albrecht and Ramamoorthy 2013). Consider a 2-player single-stage foraging problem illustrated in Fig. $1(a)$. Robot i and human j must *load* food found in adjacent cells. Players can load if the sum of their *powers* is greater than or equal to the power of the food. Thus, i or j individually cannot load the food in the bottom-left corner, but they can co-

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ordinate and jointly load it. Human j by himself can load the food to the right of him. There is a possibility that the human is robophobic and derives less benefit from the food when loading it in cooperation with the robot.

Harsanyi's framework (1967) is usually applied to such games of incomplete information (human above could be robophobic or not thereby exhibiting differing payoffs) by introducing payoff-based *types* and a common prior that gives the distribution over joint types, $\Theta = \Theta_i \times \Theta_j$, where $\Theta_{i(j)}$ is the non-empty set of types of player $i(j)$. However, this interpretation that a player type is synonymous with payoffs only is now considered naive and restrictive (Qin and Yang 2013) because knowing a player's payoff function also implies perfectly knowing its beliefs over other's types from the common prior. The prevailing theoretical interpretation (Mertens and Zamir 1985; Brandenburger and Dekel 1993) decouples the player's belief from its payoffs by introducing fixed states of the game as consisting of states of nature X and the joint types Θ (where X would be the set of payoff functions), and a common prior p over $X \times \Theta$. This allows an *explicit* definition of a Harsanyi type space for player i as, $\Theta_i^{\mathcal{H}} = \langle \Theta_i, \mathcal{S}_i, \Sigma_i, \beta_i \rangle$, where Θ_i is as defined previously; $\mathcal{S}_i$ is the collection of all sigma algebras on Θ_i; $\Sigma_i : \Theta_i \rightarrow \mathcal{S}_j$ maps each type in Θ_i to a sigma algebra in $\mathcal{S}_j$; and β_i gives the belief associated with each type of i, $\beta_i(\theta_i) \in \triangle(X \times \Theta_j, \mathcal{F}_X \times \Sigma_i(\theta_i))$, $\mathcal{F}_X$ is a sigma algebra on X. Notice that $\beta_i(\theta_i)$ is analogous to $p(\cdot|\theta_i)$ where p is the common prior on $X \times \Theta$. We illustrate a Bayesian game (BG), and the induced belief hierarchy next:

Definition 1 (Bayesian game) *A BG between N players is a collection, $\mathcal{G} = \langle X, (A_i, R_i, \Theta_i^{\mathcal{H}})_{i \in N} \rangle$, where X is the non-empty set of payoff-relevant states of nature with two or more states; A_i is the set of player i's actions; $R_i : X \times \prod_{i \in N} A_i \rightarrow \mathbb{R}$ is i's payoff; and $\Theta_i^{\mathcal{H}}$ is Harsanyi type space.*

Example 1 (Beliefs in Harsanyi type spaces) *Consider the foraging problem described previously and illustrated in Fig. 1(a). Let each player possess 4 actions that load food from adjacent cells in the cardinal directions: Ld-W, Ld-N, Ld-E, Ld-S. Let $X = \{x, x'(\neq x)\}$ and the corresponding payoff functions are as shown in Fig. 1(b). Player i has 4 types, $\Theta_i = \{\theta_i^1, \theta_i^2, \theta_i^3, \theta_i^4\}$, and analogously for j. $\Sigma_i(\theta_i^a)$, $a = 1 \ldots |\Theta_i|$ is the sigma algebra generated by the set $\{\theta_j^1, \theta_j^2, \theta_j^3, \theta_j^4\}$. Finally, example belief measures, $\beta_i(\cdot)$ and $\beta_j(\cdot)$, are shown in Fig. 1(c).*

Distributions β induce higher-level beliefs as follows: Player i with type θ_i^1 believes with probability 1 that the state is x, which is its zero-level belief, $b_{i,0}$. It also believes that j believes that the state is x because $\beta_i(\theta_i^1)$ places probability 1 on θ_j^1 and $\beta_j(\theta_j^1)$ places probability 1 on state x. This is i's first-level belief, $b_{i,1}$. Further, i's second-level belief $b_{i,2}$ induced from $\beta_i(\theta_i^1)$ believes that the state is x, that j believes that the state is x, and that j believes that i believes that the state is x. Thus, $b_{i,2}$ is a distribution over the state and the belief hierarchy $\{b_{j,0}(\theta_j), b_{j,1}(\theta_j) : \theta_j = \theta_j^1, \ldots, \theta_j^4\}$. This continues for higher levels of belief and gives the belief hierarchy, $\{b_{i,0}(\theta_i^1), b_{i,1}(\theta_i^1), \ldots\}$ generated by $\beta_i(\theta_i^1)$. Other types for player i also induce analogous infinite belief hierarchies, and a similar construction induces for j.

Example 1 also suggests a path to formally defining the induced infinite belief hierarchies from types. This definition is well known (Mertens and Zamir 1985; Brandenburger and Dekel 1993) and is not reproduced here due to lack of space.

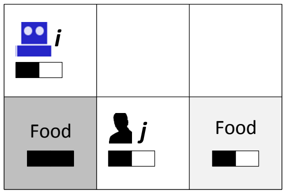
Recently, Kets (2014) introduced a way to formalize the insight that i's level l belief assigns a probability to all events that are expressed by j's belief hierarchies up to level $l-1$. Further, beliefs with levels greater than l assign probabilities to events that are expressible by j's belief hierarchies of level $l-1$ only; this is a well-known definition of finite-level beliefs. The construction involves an information partition (Aumann 1999) of other player's types, representing the cognitively-limited player's ambiguous knowledge.

Example 2 (Beliefs in depth-1 type spaces) *Let $\Sigma_i(\theta_i^1)$ be the sigma algebra generated by the partition, $\{\{\theta_j^1, \theta_j^3\}, \{\theta_j^2, \theta_j^4\}\}$. Recall that belief $\beta_i(\theta_i^1)$ is a probability measure over $\mathcal{F}_X \times \Sigma_i(\theta_i^1)$. We may interpret this construction as i's type θ_i^1 distinguishes between the events that j's type is θ_j^1 or θ_j^3 and that the type is θ_j^2 or θ_j^4 only. We illustrate example $\beta_i(\theta_i^a)$, $a = 1, \ldots, 4$ and $\beta_j(\theta_j^b)$, $b = 1, \ldots, 4$ in Fig. 2. Notice that $\beta_i(\theta_i^1)$ induces a level 0 belief $b_{i,0}$ that believes that the state of nature is x with probability 1. It also induces a level 1 belief $b_{i,1}$ that believes j believes with probability 1 that the state is x (it places probability 1 on $\{\theta_j^1, \theta_j^3\}$; both $\beta_j(\theta_j^1)$ and $\beta_j(\theta_j^3)$ place probability 1 on x). However, $\beta_i(\theta_i^1)$ does not induce a level 2 belief because $\beta_j(\theta_j^1)$ places probability 1 on $\{\theta_i^1, \theta_i^2\}$ who each, in turn, place a probability 1 on x, whereas $\beta_j(\theta_j^3)$, analogously, places a probability 1 on x'. Therefore, agent j's corresponding level 1 beliefs $\beta_j(\theta_j^1)$ and $\beta_j(\theta_j^3)$ differ in what they believe about agent i's belief about the state of nature. Consequently, $\beta_i(\theta_i^1)$ induces a belief that is unable to distinguish between differing events expressible by j's level 1 belief hierarchies. The reader may verify that the above holds true for all $\beta_i(\theta_i^a)$ and $\beta_j(\theta_j^b)$. Thus, the type spaces in Fig. 2 induces a finite-level belief hierarchy of the same depth of 1 for both agents.*

Beliefs need not always concentrate all probability mass on a single event. For example, we may replace $\boldsymbol{\beta_i(\theta_i^1)}$ in Fig. 2 with a distribution that places probability 0.5 on $\{\theta_j^1, \theta_j^3\}$ and 0.5 on $\{\theta_j^2, \theta_j^4\}$ both under column x. Yet, both agents' continue to exhibit belief hierarchies of level 1. A formal definition of an induced finite-level belief hierarchy simply modifies the definition of an infinite-level hierarchy to consider sigma algebras on the partitions of Θ_i and Θ_j.

Let us denote depth-k type spaces of player i using Θ_i^k, where each type for i induces a belief hierarchy of depth k. Let the strategy of a player i be defined as, $\pi_i : \Theta_i \rightarrow \Delta(A_i)$. Computation of the *ex-interim* expected utility of player i in the profile, (π_i, π_j) given i's type proceeds identically for both Harsanyi and depth-k type spaces:

$$U_i(\pi_i, \pi_j; \theta_i) = \int\limits_{\mathcal{F}_X \times \Sigma_i(\theta_i)} \sum_{A_i, A_j} R_i(a_i, a_j, x)\; \pi_i(\theta_i)(a_i) \times \pi_j(\theta_j)(a_j)\; d\beta_i(\theta_i) \qquad (1)$$



(a) Players i and j seek to *load* food. Sum of *powers* of players $\geq$ *power* level of the food to load it.

x	**Ld-W**	**Ld-N**	**Ld-E**	**Ld-S**
Ld-W	0,0	0,0	0,1	0,0
Ld-N	0,0	0,0	0,1	0,0
Ld-E	0,0	0,0	0,1	0,0
Ld-S	1.5,1.5	0,0	0,1	0,0

$x' \neq x$	**Ld-W**	**Ld-N**	**Ld-E**	**Ld-S**
Ld-W	0,0	0,0	0,1	0,0
Ld-N	0,0	0,0	0,1	0,0
Ld-E	0,0	0,0	0,1	0,0
Ld-S	1.5,1	0,0	0,1	0,0

(b) Payoff tables for states x and robophobic x'.

$$\begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^1)} & x & x' \\ \theta_j^1 & 1 & 0 \\ \theta_j^2 & 0 & 0 \\ \theta_j^3 & 0 & 0 \\ \theta_j^4 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^2)} & x & x' \\ \theta_j^1 & 0 & 0 \\ \theta_j^2 & 0 & 0 \\ \theta_j^3 & 0 & 1 \\ \theta_j^4 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^3)} & x & x' \\ \theta_j^1 & 0 & 0 \\ \theta_j^2 & 1 & 0 \\ \theta_j^3 & 0 & 0 \\ \theta_j^4 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^4)} & x & x' \\ \theta_j^1 & 0 & 0 \\ \theta_j^2 & 0 & 0 \\ \theta_j^3 & 0 & 0 \\ \theta_j^4 & 1 & 0 \end{array}$$

$$\begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^1)} & x & x' \\ \theta_i^1 & 1 & 0 \\ \theta_i^2 & 0 & 0 \\ \theta_i^3 & 0 & 0 \\ \theta_i^4 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^2)} & x & x' \\ \theta_i^1 & 0 & 0 \\ \theta_i^2 & 0 & 0 \\ \theta_i^3 & 0 & 1 \\ \theta_i^4 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^3)} & x & x' \\ \theta_i^1 & 0 & 0 \\ \theta_i^2 & 1 & 0 \\ \theta_i^3 & 0 & 0 \\ \theta_i^4 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^4)} & x & x' \\ \theta_i^1 & 0 & 0 \\ \theta_i^2 & 0 & 0 \\ \theta_i^3 & 0 & 0 \\ \theta_i^4 & 1 & 0 \end{array}$$

(c) Conditional beliefs of player i over the payoff states and types of j (top) and analogously for j (below).

Figure 1: (a) Single-step foraging on a 2×3 grid; (b) Payoffs corresponding to states of nature in X. Rows correspond to actions of player i and columns to actions of j; and (c) Conditional beliefs in the explicit Harsanyi type spaces of players i and j.

$$\begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^1)} & x & x' \\ \{\theta_j^1,\theta_j^3\} & 1 & 0 \\ \{\theta_j^2,\theta_j^4\} & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^2)} & x & x' \\ \{\theta_j^1,\theta_j^3\} & 0 & 0 \\ \{\theta_j^2,\theta_j^4\} & 1 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^3)} & x & x' \\ \{\theta_j^1,\theta_j^3\} & 0 & 0 \\ \{\theta_j^2,\theta_j^4\} & 0 & 1 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_i(\theta_i^4)} & x & x' \\ \{\theta_j^1,\theta_j^3\} & 0 & 1 \\ \{\theta_j^2,\theta_j^4\} & 0 & 0 \end{array}$$

$$\begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^1)} & x & x' \\ \{\theta_i^1,\theta_i^2\} & 1 & 0 \\ \{\theta_i^3,\theta_i^4\} & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^2)} & x & x' \\ \{\theta_i^1,\theta_i^2\} & 0 & 1 \\ \{\theta_i^3,\theta_i^4\} & 0 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^3)} & x & x' \\ \{\theta_i^1,\theta_i^2\} & 0 & 0 \\ \{\theta_i^3,\theta_i^4\} & 1 & 0 \end{array} \quad \begin{array}{c|cc|} \boldsymbol{\beta_j(\theta_j^4)} & x & x' \\ \{\theta_i^1,\theta_i^2\} & 0 & 0 \\ \{\theta_i^3,\theta_i^4\} & 0 & 1 \end{array}$$

Figure 2: Player i's and j's conditional beliefs on payoff states and partitions of the other agent's type set.

However, the expected utility may not be well defined in the context of depth-k type spaces. Consider Example 2 where $\Sigma_i(\theta_i^1)$ is a partition of $\{\{\theta_j^1, \theta_j^3\}, \{\theta_j^2, \theta_j^4\}\}$. U_i is not well defined for θ_i^1 if j's strategy in its argument has distributions for θ_j^1 and θ_j^3 that differ, or has differing distributions for θ_j^2 and θ_j^4. More formally, such a strategy is not *comprehensible* for type θ_i^1 (Kets 2014).

Definition 2 (Comprehensibility) *A strategy π_j is comprehensible for type θ_i^1 if it is measurable with respect to $\Sigma_i(\theta_i^1)$ (and the usual sigma algebra on set A_j).*

Obviously, lack of comprehensibility does not arise in Harsanyi type spaces because each player's belief is over a partition of the other player's types whose elements are of size 1. Finally, we define an equilibrium profile of strategies:

Definition 3 (Equilibrium) *A profile of strategies, $(\pi_i)_{i \in N}$, is in equilibrium for a BG $\mathcal{G}$ if for every type, $\theta_i \in \Theta_i$, $i \in N$,*

1. *Strategy π_j, $j \in N, j \neq i$, is comprehensible for θ_i;*
2. *Strategy π_i gives the maximal ex-interim expected utility, $U_i(\pi_i, \ldots, \pi_z; \theta_i) \geq U_i(\pi_i', \ldots, \pi_z; \theta_i)$ where $\pi_i' \neq \pi_i$ and U_i is as defined in Eq. 1.*

Condition 1 ensures that others' strategies are comprehensible for each of i's type so that the expected utility is well defined. Condition 2 is the standard best response requirement. If the type spaces in $\mathcal{G}$ are the standard Harsanyi ones, then Definition 3 is that of the standard Bayes-Nash equilibrium. Otherwise, if $\mathcal{G}$ contains depth-k type spaces, then the profile is in *finite-level equilibrium* (FLE).

BMG with Finite-Level Types

Previously, we reviewed a framework that allows characterizing equilibrium given belief hierarchies of finite depths. A key contribution in this paper is to generalize this framework endowed with finite-level type spaces to an incomplete-information Markov game played by Bayesian players. In this setting, types are now dynamic and a challenge is to identify a way of updating the types. Thereafter, we introduce an equilibrium that is pertinent for these games. We define a Bayesian Markov game (BMG) with explicit types:

Definition 4 (BMG) *A Bayesian Markov game with finite-level type spaces (BMG) is a collection:*

$$\mathcal{G}^* = \langle S, X, (A_i, R_i, \Theta_i^k)_{i \in N}, T, OC \rangle$$

- *S is the set of physical states of the game;*
- *X and A_i are as defined in the previous section for a BG;*
- *$R_i : S \times X \times \prod_{i \in N} A_i \to \mathbb{R}$ is i's reward function; it generalizes the reward function in a BG to also include the physical state;*
- *Θ_i^k is the depth-k type space of some finite depth k;*
- *$T : S \times \prod_{i \in N} A \to \Delta(S)$ is a Markovian and stochastic physical state transition function of the game; and*
- *OC is the optimality criterion which could be to optimize over a finite number of steps or over an infinite number of steps with discount factor, $\gamma \in (0, 1)$.*

A BMG between two agents i and j of some type θ_i and θ_j respectively, proceeds in the following way: both agents initially start at state s^t that is known to both and perform actions a_i^t and a_j^t according to their strategies, respectively. This causes a transition of the state in the next time step to some state s^{t+1} according to the stochastic transition function of the game, T. Both agents now receive observations, $o_i^{t+1} = \langle s^{t+1}, a_j^t \rangle$ and $o_j^{t+1} = \langle s^{t+1}, a_i^t \rangle$ respectively, that perfectly inform them about current state and other's previous action. Based on these observations, their next actions, a_i^{t+1} and a_j^{t+1}, are selected based on their strategies.

Dynamic Type Update

As we mentioned previously, players i and j engaging in a BMG observe the initial state, $o_i^0 \triangleq s^0$, followed by receiving observations of the state and other's previous action in subsequent steps, $o_i^{t+1} \triangleq \langle s^{t+1}, a_j^t \rangle$. An observation of j's action provides information that i may use to update its belief $\beta_i(\theta_i)$ in its type. Recall that $\beta_i(\theta_i)$ is a distribution over $(X \times \Theta_j, \mathcal{F}_X \times \Sigma_i(\theta_i))$. Consequently, the type gets updated. We are interested in obtaining updated distributions, $\beta_i^{t+1}(\theta_i)$ for each $\theta_i \in \Theta_i$, given the history of observations $\mathbf{o}_i^{0:t+1} \triangleq \langle o_i^0, o_i^1, \ldots, o_i^{t+1} \rangle$. This is a simple example of using a current step observation to smooth past belief.

$$\beta_i^{t+1}(\theta_i)(x, \theta_j | \mathbf{o}_i^{0:t+1}) \propto Pr(a_j^t | \theta_j, s^t)\ \beta_i^t(\theta_i)(x, \theta_j) \quad (2)$$

In Eq. 2, $Pr(a_j^t|\theta_j, s^t)$ is obtained from j's strategy in the profile under consideration and indexed by θ_j and state s^t as outlined in the next subsection. Term $\beta_i^t(\theta_i)(x, \theta_j)$ is the prior. Because of the Markovian assumption, observation history until state s^t is a sufficient statistic for the update.

Solution

Types defined using belief hierarchies limited to finite levels may not yield equilibria that coincide precisely with Bayesian-Nash equilibrium (Kets 2014), which requires that the level be infinite. We define the solution of a BMG with explicit finite-level types to be a profile of mixed strategies in *FLE* that is *Markov perfect* (Maskin and Tirole 2001); it generalizes the FLE formalized in Def. 3. Prior to defining the equilibrium, define a strategy of player i as a sequence of horizon-indexed strategies, $\boldsymbol{\pi}_i^h \triangleq (\pi_i^h, \pi_i^{h-1}, \ldots, \pi_i^1)$. Here, $\pi_i^h : S \times \Theta_i \rightarrow \Delta(A_i)$ gives the strategy that best responds with h steps left in the Markov game. Notice that each strategy in the sequence is a mapping from the current physical state and player's type; this satisfies the Markov property. Further recall that a player's type θ_i is associated with a belief measure $\beta_i(\theta_i)$. We define the equilibrium and specify conditions for its existence next.

Definition 5 (Markov-perfect finite level equilibrium) *A profile of strategies, $\boldsymbol{\pi}_k^h = (\boldsymbol{\pi}_{i,k}^h)_{i \in N}$ is in Markov-perfect finite-level equilibrium (MPFLE) of level k if the following holds:*

1. *Each player has a depth-k type space;*
2. *Strategy $\boldsymbol{\pi}_{j,k}^h$, $j \in N$, $j \neq i$ and at every horizon is comprehensible for every type of player i;*
3. *Each player's strategy for every type is a best response to all other players' strategies in the profile and the equilibrium is subgame perfect.*

Notice that if, instead of condition 1 above, players possess the standard Harsanyi type space, then Def. 5 gives the Markov-perfect Bayes-Nash equilibrium. Definition 2 characterizes a comprehensible strategy.

Strategy $\boldsymbol{\pi}_{i,k}^h$ is a best response if its value is the largest among all of i's strategies given the profile of other players' strategies. To quantify the best response, we define an *ex-interim* value function for the finite horizon game that assigns a value to each level strategy of a player, say i, given the observed state, i's own type and profile of other players' strategies. For a two player BMG $\mathcal{G}^*$, each player endowed with a depth-k type space, this function is:

$$Q_i(s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i) = U_i^*(s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i) + \quad (3)$$
$$\gamma \sum\nolimits_{o_i} Pr(o_i' | s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i)\ Q_i(s', \pi_{i,k}^{h-1}, \pi_{j,k}^{h-1}; \theta_i')$$

where o_i' denotes $\langle s', a_j \rangle$, θ_i' is the updated type of i due to a_j (Eq. 2), and $Q_i(s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i)$ reduces to U_i^* when $h = 1$. Here,

$$U_i^*(s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i) = \int\limits_{\mathcal{F}_X \times \Sigma_i(\theta_i)} \sum_{A_i, A_j} R_i(s, x, a_i, a_j)$$
$$\times\ \pi_{i,k}^h(s, \theta_i)(a_i)\ \pi_{j,k}^h(s, \theta_j)(a_j)\ d\beta_i(\theta_i)$$

Utility function U_i^* extends U_i in Eq. 1 to the single stage of a BMG. Next, we focus on the term $Pr(o_i'|s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i)$:

$$Pr(o_i' | s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i) = \int\limits_{\Sigma_i(\theta_i)} \sum_{A_i} T(s, a_i, a_j, s')$$
$$\times\ \pi_{i,k}^h(s, \theta_i)(a_i)\ \pi_{j,k}^h(s, \theta_j)(a_j)\ d\hat{\beta}_i(\theta_i)$$

where $\hat{\beta}_i(\theta_i)$ is the marginal of measure $\beta_i(\theta_i)$ on $\Sigma_i(\theta_i)$ only. This equation is derived in the supplement. Subsequently, $\boldsymbol{\pi}_{i,k}^h$ that optimizes Q_i is a best response to given $\boldsymbol{\pi}_{j,k}^h$. When the horizon is infinite, each player possesses a single strategy that is not indexed by horizon. Note that Eqs. 1 and 2 can be easily generalized to $|N|$ agents.

We define an ϵ-MPFLE which relaxes the strict requirement of the exact equilibrium allowing a player in approximate equilibrium to deviate if her loss due to deviating to some other strategy is not more than ϵ. Finally, a MPFLE may not always exist for a BMG of level k because the given depth-k type space of a player may not admit any comprehensible and best response strategy as required by conditions 2 and 3 of Def. 5. We do not view the nonexistence of equilibrium for all games as particularly limiting, but simply as a consequence of the fact that some equilibria are too complicated to reason with finite cognitive capabilities.

Finding MPFLE using constraint satisfaction

Vickrey and Koller (2002) present a way to compute Nash equilibrium in single-shot graphical games with complete information using constraint satisfaction. Later, Soni et al. (2007) extend their work and model the problem of finding a Bayes-Nash equilibrium in single-shot graphical games with incomplete information and repeated graphical games also as a constraint satisfaction problem (CSP). We further adapt their methods toward finding MPFLE in BMGs.

First, we transform the BMG into an *extended* Bayesian game by defining strategy for player $i \in N$ as a vector of horizon-indexed strategies as elucidated previously. Next, we formalize the CSP represented as a 3-tuple: $\mathbb{P}_E = \langle V, D, C \rangle$. Here, V is a set of variables, $V = \{v_1, \ldots, v_{|N|}\}$, where each variable corresponds to a player in the BMG; D is the set of domains for the variables, $D = \{D_1, \ldots, D_{|N|}\}$. The domain D_i for a variable v_i ($i \in N$) is the space of *comprehensible* strategies for player i. Comprehensibility limits the size of the domain, which in turn translates to significant computational savings in time while also ensuring that the expected utility is well-defined. C is a set of $|N|$ $|N|$-ary constraints. Each constraint $C_{i \in N}$

has the scope V which is the set of all variables, and the relation $R_i \subseteq \times_{i \in N} D_i$. A tuple $r_i \in R_i$ is considered legal if the corresponding strategy of player i is a best response to the strategy profile of others specified in r_i. The relation R_i only constitutes legal tuples. Next, we generate the *dual* CSP from the original CSP formalized above. The variables of the dual CSP are the constraints of the original CSP. Thus, the dual variables are $C = \{C_1, \ldots, C_{|N|}\}$. The domain of each dual variable is the tuple of the corresponding relation in the original CSP. Thus, the dual variable $C_{i \in N}$ has $|R_i|$ values. Finally, we add an $|N|$-ary equality constraint on the dual variables. This constraint essentially performs an intersection across the domains of each of the dual variables. This guarantees that all players play a mutual best response strategy and hence, commit to the same Nash equilibrium which is in turn an MPFLE for the BMG.

In general, solving a CSP involves pruning the domain of each variable. If at any stage, any variable's domain becomes empty on application of constraints, it indicates that the CSP is unsatisfiable. On modeling the game as a CSP, we may apply any standard CSP solver to compute equilibria. We used the generic procedure described in an efficient arc consistency algorithm called MAC3 (Liu 1998) to solve the CSP. The complexity of the best-response constraint-checking step is exponential in the total number of agents and the planning horizon H. If all agents interact with each other, this step runs in time $\mathcal{O}(|N|(\frac{1}{\tau})^{|A||\Theta||N|H})$ where τ is the granularity of agents' strategy space. Furthermore, we take advantage of *sub-game perfection* in MPFLE by going bottom-up from a 1-step equilibrium strategy to an H-step strategy in the consistency checking phase for savings.

Approximation for Mixed Strategies

Recall that a possible value of each variable is a profile of strategies. As the level strategies may be *mixed* allowing distributions over actions, the domain of each variable is continuous. Algorithms such as MAC3 are unable to operate on continuous domain spaces. Soni et al. (2007) point out this problem and suggest discretizing the continuous space of mixed strategies using a τ-grid on the simplex. In the context of a BMG, given the τ-grid and player i's strategy $\boldsymbol{\pi}^h_{i,k}$, the probability of taking an action $a_i \in A_i$, $\pi^h_{i,k}(\cdot, \cdot)(a_i) \in \{0, \tau, 2\tau, \ldots, 1\}$. Compared to uncountably many possibilities for each strategy before, we now consider $1/\tau^2$ entries on the τ-grid. Subsequently, discretizing the continuous space of mixed strategies on the τ-grid becomes a part of initializing the domain of each variable.

However, a profile of strategies in equilibrium may not lie on the τ-grid if the discretization is too coarse. Thus, the discretization may introduce error and motivates relaxing the exact equilibrium to ϵ-MPFLE. Interestingly, we can bound the loss suffered by any player in moving to the adjacent joint strategy on the τ-grid, which in turn allows us to show that a relaxed MPFLE is preserved by the discretization. *We present this bound and related proofs in the supplement.*

Unfortunately, the bound is usually loose and therefore, a small ϵ could lead to unreasonably fine τ-grids and we may end up having an intractably large mixed-strategy space. Inversely, if we fix the granularity τ of the grid to be small, we may end up approximating ϵ to an extent that the solution becomes meaningless. In both cases, the risk of not finding an equilibrium is still probable because of finite-level reasoning. In the empirical results we present next, we attempt to find a reasonable trade off while ensuring the existence of at least one MPFLE on two standard domains.

Empirical Evaluation

We implemented the MAC3 algorithm for obtaining MPFLE as discussed earlier. We experiment with two benchmark problem domains: *n-agent multiple access broadcast channel* (nMABC) (Hansen, Bernstein, and Zilberstein 2004) ($|N|$ = 2 to 5; H = 1 to 5; $|S|$ = 4; $|A|$ = 4; $|X_{i \in N}|$ up to 4; $\prod_{i \in N} |\Theta_i|$ up to 1024) and sequential *level-based foraging*, which involves players performing *move* actions in cardinal directions and just one *load* action ($m \times m$ Foraging) (Albrecht and Ramamoorthy 2013) (m=3; $|N|$ = 2; H = 1 to 3; $|S|$ = 81; $|A|$ = 25; $|X_{i \in N}|$ = 2; $\prod_{i \in N} |\Theta_i|$ = 16).

In our experiments on both domains, for each agent, we manually created partitions of the other agents' types with a maximum size of 2 with as many payoff states as there are partitions and ensured that the construction induced a level-1 belief hierarchy for all participating agents. For example, in the 2MABC problem, say each agent $\{i, j\}$ has a total of 4 types. Then, the type-set for each agent is divided into 2 partitions containing 2 types each: $\{\{\theta_i^1, \theta_i^2\}, \{\theta_i^3, \theta_i^4\}\}$ and $\{\{\theta_j^1, \theta_j^3\}, \{\theta_j^2, \theta_j^4\}\}$; and let there be a total of 2 states of nature: x and x'. We assume that the belief β_i continues to assign a point probability mass of either 1 or 0 on any particular partition and state of nature. These beliefs were manually assigned such that the type spaces induced a finite-level belief hierarchy of depth 1 for both agents. Figure 2 shows one such configuration for the 2MABC problem.

Validation First, we focus on generating MPFLE in games of N Bayesian players. Multiple equilibria with pure and mixed comprehensible strategies were found for depth-1 type spaces. For example, at $H = 2$, we found 3 pure-strategy exact MPFLE. We also found 12 and 17 ϵ-MPFLE for ϵ = 0.17 and 0.33 respectively. We begin by noting that all computed ϵ-MPFLE coincide with ϵ-BNE for the 2MABC problem. We obtained BNE by considering a common prior and a unit size of the partitions of the other agent's type set. As expected, there were additional BNEs as well. Specifically, we found 6 pure-strategy exact BNE, and 21 and 36 ϵ-BNE for ϵ = 0.17 and 0.33 respectively. This empirically verifies the correctness of our approach.

Run time for finding equilibrium Next, we explore the run time performance of BMG and investigate how varying the different parameters, N, H, X, Θ, τ, and ϵ, impacts the performance and scalability in the two domains. Our computing configuration included an Intel Xeon 2.67GHz processor, 12 GB RAM and Linux.

In Fig. 3 (top), we report the average time to compute the first 10 equilibria for a 3-horizon 2MABC and 3MABC with $|X|$ = 2 and $|\Theta|$ = 16 and 64 types, respectively (4 types for each player). The bound on ϵ given τ shown in Proposition 1 (see supplement) is loose. Therefore, we consider various values of ϵ that are well within this bound. An example

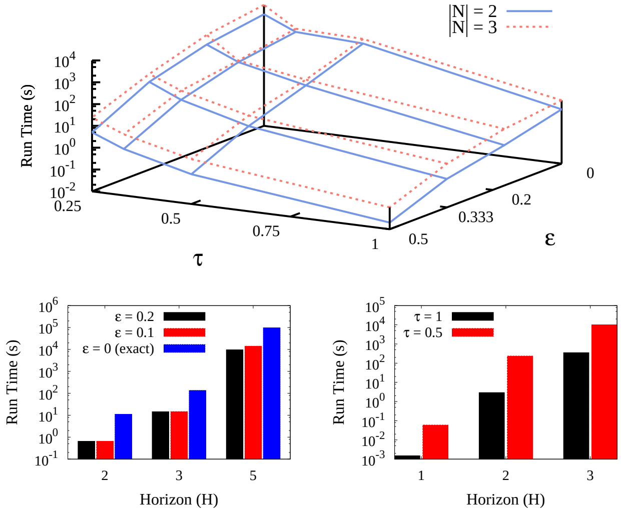


Figure 3: **Impact of parameters on performance**. Time taken to compute: (top) MPFLE in 2MABC and 3MABC for varying τ and ϵ at $H = 3$, $(left)$ a pure-strategy MPFLE in 5MABC for varying ϵ and H showing scalability in agents, and $(right)$ MPFLE in 2-agent 3×3 Foraging for varying τ and H with $\epsilon = 0.1$ showing scalability in domain size (in $|A|$ and $|S|$).

pure-strategy profile for two players in exact equilibrium in 2MABC exhibited ex-interim values [1.9,1.52] for players i and j, respectively.

Scalability We scale in the number of agents and illustrate in Fig. 3 $(left)$, times for 5MABC (5 agents) for increasing horizons with the subgame-perfect equilibrium taking just under 4 hours to compute for $H = 5$ and $\epsilon = 0.1$. Notice that this time increases considerably if we compute profiles in exact equilibria. To scale in the number of states, we experimented on the larger 3×3 Foraging and illustrate empirical results in Fig. 3 $(right)$. The time taken to compute the first ϵ-MPFLE for varying horizons and two coarse discretizations is shown. Run time decreases by about two orders of magnitude as the discretization gets coarser for $H = 2$. A pure-strategy profile for two players in exact equilibrium in 3×3 Foraging exhibited ex-interim values [1.98, 0.98] for players i and j, respectively. In general, as the approximation increases because the discretization gets coarser, the time taken to obtain strategy profiles in ϵ-equilibria decreased by multiple orders of magnitude.

H		**3**			**4**			**5**		
Without TE	$\lvert\Theta^{k=1}\rvert$	16	36	64	16	36	64	16	36	64
	Time (s)	**0.07**	**0.8**	**1335.6**	**1.01**	**42.6**	**1481.1**	**1.6**	**31.2**	**>1 day**
With TE	$\lvert\Theta^{k=1}\rvert$	4	9	16	4	12	16	4	16	25
	Time (s)	**0.11**	**0.54**	**27.2**	**0.96**	**14.3**	**311.2**	**1.3**	**26.7**	**3161.7**

Table 1: Computational savings due to TE in computing a pure-strategy MPFLE in 2MABC for level-1 types.

Type equivalence A complication in solving the CSP is that the type space is continuous because it is defined in terms of beliefs over payoff states and others' type-set partitions. This makes strategy a continuous function due to which the variables in the CSP are infinite dimensional; an additional challenge not present in Soni et al. (2007), which uses discrete types. Rathnasabapathy et al. (2006) show how we may systematically and exactly compress large type spaces using exact *behavioral equivalence*. Its manifestation here as *type equivalence (TE)* preserves the quality of the solutions obtained, which we verified experimentally as well. The reduced size of player type spaces in turn reduces the number of strategy profiles that need to be searched for finding an equilibrium. This helps lower the time complexity by several orders of magnitude as we demonstrate. Table 1 illustrates the reduction in the type space due to TE in 2MABC for varying horizons. It also shows the time savings in generating one pure-strategy profile in equilibrium. Note the overhead in computing the equivalence classes which is prominent for smaller horizons. However, savings due to TE compensate for this overhead at longer horizons and larger type spaces.

In summary, our CSP finds multiple pure and mixed-strategy profiles in MPFLE that are exact or approximate. Feasible run times are demonstrated for two domains, and we reported on scaling along various dimensions. The equilibria that we have found serve as optimal points of references for current and future methods related to coordination. The equilibrium computation could benefit from a more efficient CSP algorithm; one that potentially takes advantage of the structure of interpersonal interactions among players in BMGs.

Concluding Remarks

BMGs generalize Markov games to include imperfect information about players' types. BMGs take significant steps beyond Kets' single-shot games and Markov games by introducing sequential reasoning to the former and bounded-depth reasoning to the latter, both of which are non-trivial. They construct a type space that is founded on Aumann's concept of information partitions as a way of formalizing (imperfect) knowledge. BMG is the first formalization of incomplete-information Markov games played by Bayesian players, which integrates types that induce bounded reasoning into an operational framework.

BMGs are related to stochastic Bayesian games introduced by Albrecht and Ramamoorthy (2013) for formalizing ad hoc coordination but they exhibit key differences: 1) Types in BMG are explicitly defined while those defined in Albrecht et al. are abstract. Importantly, the latter are coupled with a prior distribution over types that is common knowledge. 2) Furthermore, we allow for a *continuous* type space (with intentional types) while Albrecht et al. relies on *arbitrarily* picking a discrete set of hypothesized user-defined (sub-intentional) types for agents. BMGs share similarities with interactive POMDPs (Gmytrasiewicz and Doshi 2005) that also allow individual agents to intentionally (or sub-intentionally) model others using a finite belief hierarchy that is constructed differently. However, the focus in I-POMDPs is to compute the best response to subjective beliefs and not to compute equilibria. Indeed, converging to equilibria in I-POMDPs is difficult (Doshi and Gmytrasiewicz 2006).

There is growing interest in game-theoretic frameworks and their solutions that model pragmatic types of players. This paper provides a natural starting point for a shared conversation about realistic and computationally feasible models. We ask the following questions as we further investigate BMGs. Does increasing the depth of reasoning get MPFLE "closer" to BNE, and can we formalize the closeness? Are there profiles in MPFLE which do not occur in the set of BNE even if the Harsanyi type space reflects the finite-level beliefs? In response, we observe that higher levels of beliefs would require increasingly fine partitions of the types. Therefore, MPFLE is hypothesized to coincide with BNE with increasing levels. Kets (2014) establishes the presence of BNE that are not present in any finite-level equilibria. However, it is always possible to construct a Harsanyi extension of the finite-level type space such that any FLE is also a BNE.

Acknowledgments

This research is supported in part by a grant from NSF IIS-0845036 and a grant from ONR N-00-0-141310870. We thank Willemien Kets and Berthe Choueiry for discussions that significantly benefited this paper.

References

Albrecht, S. V., and Ramamoorthy, S. 2013. A game-theoretic model and best response learning method for ad hoc coordination in multiagent systems. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 1155–1156.

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

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

Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Doshi, P., and Gmytrasiewicz, P. J. 2006. On the difficulty of achieving equilibrium in interactive POMDPs. In *Twenty-First Conference on Artificial Intelligence (AAAI)*, 1131–1136.

Gmytrasiewicz, P. J., and Doshi, P. 2005. A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research (JAIR)* 49–79.

Goodie, A. S.; Doshi, P.; and Young, D. L. 2012. Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making* 24:95–108.

Hansen, E.; Bernstein, D.; and Zilberstein, S. 2004. Dynamic programming for partially observable stochastic games. In *Nineteenth Conference on Artificial Conference (AAAI)*, 709–715.

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

Hedden, T., and Zhang, J. 2002. What do you think I think you think?: Strategic reasoning in matrix games. *Cognition* 85:1–36.

Ho, T.-H., and Su, X. 2013. A dynamic level-k model in sequential games. *Management Science* 59(2):452–469.

Kets, W. 2014. Finite depth of reasoning and equilibrium play in games with incomplete information. Technical Report Discussion 1569, Northwestern University.

Littman, M. 1994. Markov games as a framework for multiagent reinforcement learning. In *International Conference on Machine Learning*.

Liu, Z. 1998. Algorithms for constraint satisfaction problems (CSPs). Math thesis, Department of Computer Science, University of Waterloo.

Maskin, E., and Tirole, J. 2001. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory* 100(2):191–219.

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

Nair, R.; Tambe, M.; Yokoo, M.; Pynadath, D.; and Marsella, S. 2003. Taming decentralized POMDPs : Towards efficient policy computation for multiagent settings. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 705–711.

Qin, C.-Z., and Yang, C.-L. 2013. Finite-order type spaces and applications. *Journal of Economic Theory* 148(2):689–719.

Rathnasabapathy, B.; Doshi, P.; and Gmytrasiewicz, P. J. 2006. Exact solutions to interactive POMDPs using behavioral equivalence. In *Autonomous Agents and Multi-Agent Systems Conference (AAMAS)*, 1025–1032.

Roth, M.; Simmons, R.; and Veloso, M. 2006. What to communicate? execution-time decision in multiagent POMDPs. In *International Symposium on Distributed Autonomous Robotic Systems (DARS)*.

Soni, V.; Singh, S.; and Wellman, M. P. 2007. Constraint satisfaction algorithms for graphical games. In *Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 423–430.

Stahl, D., and Wilson, P. 1995. On player's models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10:218–254.

Vickrey, D., and Koller, D. 2002. Multi-agent algorithms for solving graphical games. In *AAAI/IAAI*, 345–351.

Wright, J. R., and Leyton-Brown, K. 2010. Beyond equilibrium: Predicting human behavior in normal-form games. In *Twenty-Fourth Conference on Artificial Intelligence (AAAI)*, 901–907.

Zeng, Y., and Doshi, P. 2012. Exploiting model equivalences for solving interactive dynamic influence diagrams. *Journal of Artificial Intelligence Research* 43:211–255.