Bayesian Markov Games with Explicit Finite-Level Types

Muthukumaran Chandrasekaran[1] and **Yingke Chen**[2] and **Prashant Doshi**[3]
539G Boyd GSRC, THINC Lab, University of Georgia
Athens, Georgia 30602
{mkran[1],pdoshi[3]}@uga.edu, yke.chen@gmail.com[2]

Abstract

We present a new game-theoretic framework where Bayesian players engage in a Markov game and each has private but imperfect information regarding other players' types. Instead of utilizing Harsanyi's abstract types and a common prior distribution, we construct player types whose structure is explicit and induces a finite level belief hierarchy. We characterize equilibria in this game and formalize the computation of finding such equilibria as a constraint satisfaction problem. The effectiveness of the new framework is demonstrated on two ad hoc team work domains.

Introduction

A plethora of empirical findings in strategic games (Camerer, Ho, and Chong 2004; 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 a 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.

Are there characterizations of equilibrium between players engaging in finite levels of inter-personal reasoning? Recently, Kets (2014) generalized the standard Harsanyi framework for games of incomplete information to allow players to have 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.

We generalize the single-stage framework of Kets to allow Bayesian players to play an incomplete-information *Markov game* (Littman 1994). Each player 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. Within this new framework for Bayesian Markov games (BMG) with *explicit types*, we generalize the constraint satisfaction algorithm introduced by Soni et al. (2004) 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 others' actions. This makes the types dynamic. Contextual to types that induce finite belief hierarchies, we define a Markov-perfect finite-level equilibrium, and present a method for solving BMGs to obtain this equilibrium. Motivated by behavioral equivalence (Zeng and Doshi 2012), we use equivalence between types in order to speed up computation of the equilibrium.

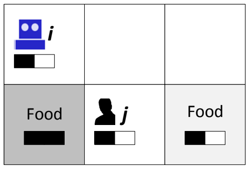
BMGs with explicit types are well-suited toward formalizing impromptu or ad hoc teamwork because of their emphasis on the uncertainty over types of others and computations that provide an individual's best response to its bounded beliefs. Consequently, we model the multi-access broadcast channel and foraging problems – well-known evaluation domains for such teamwork – as a BMG and solve it. Obtained equilibria offer locally-optimal solutions that serve as points of comparison to the value of previous solutions in these domains. The framework also offers a promising departure point for modeling empirical data on strategic interactions betwen humans. This further motivates its study and forms an important avenue of future work.

Background

Consider a 2-player single-stage foraging problem (Albrecht and Ramamoorthy 2013) 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 coordinate 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 $i(j)$. [1] The pre-

[1] This interpretation is often considered naive because knowing a player's payoff function also implies perfectly knowing its beliefs



(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'.

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

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

(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 in X. Rows correspond to player i and columns to j; and (c) Conditional beliefs in the explicit Harsanyi type spaces of players i and j.

vailing theoretical interpretation (Mertens and Zamir 1985; Brandenburger and Dekel 1993) introduces 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 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; $\Sigma_i : \Theta_i \to \mathcal{S}_j$ 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 define and illustrate a Bayesian game (BG), and the induced belief hierarchy next:

Definition 1 (Bayesian game)**.** *A BG 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 at least two states; A_i is the set of player i's actions; $R_i : X \times \prod_{i \in N} A_i \to \mathbb{R}$ is i's payoff function; and type space Θ_i^H is as defined previously.*

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 whose $\beta_j(\theta_j^1)$ 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 player j.

Recently, Kets (2014) introduced a way to formalize the insight that i's level l belief assigns a probability to all events that can be expressed by j's belief hierarchies up to level $l-1$. Furthermore, 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. We explain the formalization using an example.

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

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

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

Example 2 (Kets type spaces with finite-level beliefs)**.** *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 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 $\{\theta_i^1, \theta_i^2\}$, while corresponding $\beta_i(\theta_i^1)$ and $\beta_i(\theta_i^2)$ differ in their 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 $\beta_j(\theta_j^b)$. Consequently,

over other's types from the common prior.

the type spaces in Fig. 2 induces a finite-level belief hierarchy of the same depth of 1 for both agents.

Let us denote finite-level type spaces of player i using Θ_i^k, where each type for i induces a belief hierarchy of depth k.

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 Kets type spaces:

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

However, the expected utility may not be well defined in the context of Kets 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). Obviously, lack of comprehensibility does not arise in the context of Harsanyi type spaces.

Finally, we define an equilibrium profile of strategies:

Definition 2 (Equilibrium)**.** *A profile of strategies, $(\pi_i)_{i \in N}$, is in equilibrium for a BG $\mathcal{G}$ if for every type, $\theta_i \in \Theta_i$, $i \in N$, the following holds:*

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 2 is that of the standard Bayes-Nash equilibrium. Otherwise, if $\mathcal{G}$ contains Kets 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.

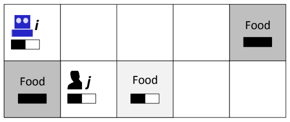
We define a Bayesian Markov game (BMG) as follows:

Definition 3 (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;*
- *$R_i : S \times X \times \prod_{i \in N} A_i \to \mathbb{R}$ is i's reward function;*
- *Θ_i^k is the finite-level Kets type space of depth k;*
- *$T : S \times \prod_{i \in N} A \to \Delta(S)$ is a stochastic physical state transition function of the Markov game; and*
- *OC is the optimality criterion in order to optimize over finite or infinite 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.



	Mv-W	Mv-E	Ld-W	Ld-N	Ld-E	Ld-S
Mv-W	0,-1	0,-1	0,0	0,0	0,10	0,0
Mv-E	-1,-1	-1,-1	-1,0	-1,0	-1,1	-1,0
Ld-W	0,-1	0,-1	0,0	0,0	0,10	0,0
Ld-N	0,-1	0,-1	0,0	0,0	0,10	0,0
Ld-E	0,-1	0,-1	0,0	0,0	0,10	0,0
Ld-S	0,-1	0,-1	15,15	0,0	0,10	0,0

Figure 3: Extended foraging game on a larger grid and example payoffs for a combination of physical and payoff states (s_1, x). Each physical state represents the location of both players.

Example 3 (Extended foraging problem)**.** *We illustrate an extended foraging problem in Fig. 3. Players i and j may move to adjacent cells in addition to loading food as before. However, movement is not free and incurs a small cost. Thus, the game is now sequential progressing from one physical state to another, where a physical state denotes the joint position of players. Payoffs now depend on both the physical state of the game and the state of nature.*

Dynamic Type Update

As we mentioned previously, players i and j engaging in a BMG receive observations of the state and other's previous action in subsequent steps, $o_i^{t+1} = \langle s^{t+1}, a_j^t \rangle$. 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 observation o_i^{t+1}. This is a simple example of using a current step observation to smooth past belief. This is given by:

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

In Eq. 2, term $Pr(a_j^t | x, \theta_j)$ is obtained from j's strategy in the profile under consideration and indexed by θ_j as outlined in the next subsection. Term $\beta_i^t(\theta_i)(x, \theta_j)$ is the prior.

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 an equilibrium that we label as *Markov-perfect finite-level equilibrium*. This equilibrium generalizes the FLE formalized in Def. 2 of the previous section. Prior to defining the equilibrium, define a strategy of player i as a vector of horizon-indexed strategies, $\boldsymbol{\pi}_i^h \triangleq \langle \pi_i^h, \pi_i^{h-1}, \ldots, \pi_i^1 \rangle$. 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 vector is a mapping from the current physical state, player's type space, and states of nature; this satisfies the Markov property. We define the equilibrium next.

Definition 4 (Markov-perfect finite level equilibrium)**.** *A profile of strategies, $\boldsymbol{\pi}_k^h = \langle \boldsymbol{\pi}_{i,k}^h \rangle_{i \in N}$ is in Markov-perfect finite-level equilibrium (MPFLE) of level k if the following holds:*

1. *Each player has a Kets type space of level k;*
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. 4 gives the Markov-perfect Bayes-Nash equilibrium.

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 level k Kets type space, this function is:

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

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

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

Utility function, U_i^*, extends U_i in Eq. 1 to the single stage of a BMG.

Next, we focus on the term $Pr(o_i' | s, \pi_{i,k}^h, \pi_{j,k}^h; \theta_i)$:

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

where $\hat{\beta}_i(\theta_i)$ is the marginal of measure $\beta_i(\theta_i)$ on $\Sigma_i(\theta_i)$ only and $o_i = \langle s', a_j \rangle$. Subsequently, $\boldsymbol{\pi}_{i,k}^h$ that optimizes Q_i in Eq. 3 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.

Finally, 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.

Algorithm for finding MPFLE

We formally proposed a BMG above and showed how dynamic player types are updated along with a characterization of equilibrium in these games. In this section, we investigate how to find such equilibria for a given BMG.

Constraint Satisfaction Problem

Vickrey and Koller present a way to compute Nash equilibrium in single-shot graphical games with complete information using constraint satisfaction (Vickrey and Koller 2002). Later, Soni et al. (Soni, Singh, and Wellman 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. 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 $R_i \subseteq \times_{i \in N} D_i$[2]. 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 are the tuples 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.

[2] We assume a fully-connected interaction graph. However, this representation can be easily generalized to any graphical game with arbitrary local neighborhoods.

In general, solving a CSP involves pruning the domain of each variable. If at any stage, any variable's domain becomes empty upon application of constraints, it indicates that the CSP is unsatisfiable. In other words, there is no solution for this CSP. Note that this method can be used to find all solutions to the CSP. Overall, once we represent the game as a CSP, we can easily apply any standard CSP solvers to compute equilibria. We implement the generic procedure described in an efficient arc consistency algorithm called MAC3 (Liu 1998) to solve our CSP. Further, we take advantage of the *sub-game perfection* condition in MPFLE by going bottom-up from a 1-step strategy to an H-step strategy in the consistency-checking phase to ensure additional savings. The intuition is that, if a 1-step strategy profile is not an equilibrium, then the 2-step strategy profile that includes this 1-step non-equilibrium strategy profile is also not going to be an equilibrium and so on.

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. 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 by 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. Therefore, the discretization may introduce error and motivates relaxing the exact equilibrium to ϵ-MPFLE. Interestingly, an upper bound on ϵ may be obtained for a given value of τ (Soni, Singh, and Wellman 2007). We derive this bound for a BMG next.

Lemma 1. *Let $\boldsymbol{\pi}_k$, $\boldsymbol{\pi}'_k$ be two profiles of level k strategies, and for any player $i \in N$, let strategies $\boldsymbol{\pi}_{i,k}$ and $\boldsymbol{\pi}'_{i,k}$ be present in the above profiles, respectively. Let horizon-indexed level k strategies $\pi^h_{i,k}$ and $\pi'^h_{i,k}$ from the above vectors, for each combination of state and type, be τ-close: $|\pi^h_{i,k}(s,\theta_i)(a_i) - \pi'^h_{i,k}(s,\theta_i)(a_i)| \leq \tau,\ \forall\ i, s, \theta_i, a_i$. Then, the probability of an action profile, $\langle a^i, ..., a^z\rangle$, $\forall\ s, \theta_i, a_i$, is bounded:*

$$| \prod_{i,j,\ldots,z} \pi^h_{i,k}(s,\theta_i)(a^i) - \prod_{i,j,\ldots,z} \pi'^h_{i,k}(s,\theta_i)(a^i)| \leq |N|\tau$$

Note that this upper bound can be refined in a straightforward way to $\underline{|N|}\tau$ if we know that strategies of $\underline{|N|} < |N|$ may deviate only.

Given Lemma 1, the induced difference in the immediate ex-interim expected utility for a player may also be bounded if its strategies are τ-close.

Lemma 2. *Let $\boldsymbol{\pi}_k$ and $\boldsymbol{\pi}'_k$ be strategy profiles that are τ-close as defined in Lemma 1. The corresponding induced difference in immediate ex-interim expected utility is bounded:*

$$|U^*_i(s, \pi^h_{i,k}, \pi^h_{-i,k}; \theta_i) - U^*_i(s, \pi'^h_{i,k}, \pi'^h_{-i,k}; \theta_i)| \leq R_{max}|A||N|\tau,\ \forall\ i, s, \theta_i$$

Next, we may also bound the difference in the expected future payoff induced by two profiles that are τ-close.

Lemma 3. *Let $\boldsymbol{\pi}_k$ and $\boldsymbol{\pi}'_k$ be strategies that are τ-close as defined in Lemma 1. Let $\phi^T_i(\pi^h_{i,k}(s,\theta_i); s, \theta_i, \boldsymbol{\pi}^h_{-i,k})$ denote i's expected future payoff. The corresponding difference in expected future payoff is recursively bounded:*

$$|\phi^T_i(\pi^h_{i,k}(s,\theta_i); s, \theta_i, \boldsymbol{\pi}^h_{-i,k}) - \phi^T_i(\pi'^h_{i,k}(s,\theta_i); s, \theta_i, \boldsymbol{\pi}'^h_{-i,k})| \leq \gamma(T-1)R_{max}|A||N|\tau + \gamma|\phi^{T-1}_i(\pi^h_{i,k}(s', \theta'_i); \cdot) - \phi^{T-1}_i(\pi'^h_{i,k}(s', \theta'^h_{i,k}); \cdot)|$$

Combining Lemmas 2 and 3, we obtain the following:

Proposition 1. *Let $\boldsymbol{\pi}_k$ and $\boldsymbol{\pi}'_k$ be strategy profiles that are τ-close as defined in Lemma 1. Then the difference in the expected value due to the difference in the profiles, $|Q^T_i(\pi^h_{i,k}(s,\theta_i); s, \theta_i, \boldsymbol{\pi}^h_{-i,k}) - Q^T_i(\pi'^h_{i,k}(s,\theta_i); s, \theta_i, \boldsymbol{\pi}'^h_{-i,k})|$ for all $s \in S, \theta_i \in \Theta_i$ is bounded by $\epsilon^T = |A||N|\tau R_{max} \left(T\frac{\gamma^T - 1}{\gamma - 1} + \gamma^{T-1}\right)$.*

Proposition 1 bounds the loss suffered by any player in moving to the adjacent joint strategy on the τ-grid. Now, we can show that a relaxed MPFLE is preserved by the discretization.

Proposition 2 (ϵ-MPFLE)**.** *Let the joint strategy $\boldsymbol{\pi}_k$ be an MPFLE for a given BMG. Let $\boldsymbol{\pi}'_k$ be the nearest strategy profile on the τ-grid. Then $\boldsymbol{\pi}'_k$ is a ϵ-MPFLE for the BMG, where ϵ is as defined previously in Proposition 1.*

Experimental Results

We implemented the MAC3 algorithm for obtaining MPFLE as discussed earlier. We show the applicability of BMGs toward two benchmark problem domains used in the ad hoc team work literature: *n-agent multiple access broadcast channel* (nMABC) (Hansen, Bernstein, and Zilberstein 2004) and *level-based foraging* ($m \times m$ Foraging) (Albrecht and Ramamoorthy 2013) illustrated previously; ties are broken randomly if more than one food can be loaded. Table 1 summarizes the domain statistics and parameter settings.

Domain	Specifications
nMABC	$\|N\|$ = 2 to 5; H = 2 to 5; $\|S\|$ = 4; $\|A\|$ = 4; $\|X_{i\in N}\|$ = up to 4; $\prod_{i\in N}\|\Theta_i\|$ up to 1024
3 × 3 Foraging	$\|N\|$ = 2; T = 1 to 3; $\|S\|$ = 81; $\|A\|$ = 25; $\|X_{i\in N}\|$ = 2; $\prod_{i\in N}\|\Theta_i\|$ = 16

Table 1: Specifications of the different domains. 3 × 3 Foraging is one of the largest benchmarks in ad hoc teamwork.

Validation First, we focus on generating MPFLE in games of N Bayesian players. We begin by noting that the Pareto-optimal MPFLE generated by our CSP coincides with the optimal joint policy obtained from a decentralized POMDP using DP-JESP (Nair et al. 2003) for the 2MABC problem. This empirically verifies the correctness of our approach. Multiple equilibria with pure and mixed comprehensible strategies were found for level-1 Kets 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.

Next, to enhance significance and position BMGs (and MPFLE) better, we present empirical results on the 2-agent 3 $\times$ 3 Foraging domain in Table 2 comparing converged (i.e., until all food is loaded) *ex-interim* values with the *Harsanyi-Bellman Ad Hoc Coordination* algorithm (HBA) introduced by Albrecht et al. (Albrecht and Ramamoorthy 2013) and two popular multiagent reinforcement learning (MARL) approaches: JAL (Claus and Boutilier 1998) learns the action frequencies of each player in each state and uses them to compute expected action payoffs; and CJAL (Banerjee and Sen 2007) is similar to JAL but learns the frequencies conditioned on its own actions. Notice that BMG's level-1 equilibrium is Pareto-efficient implying that level-1 reasoning is sufficient in the Foraging domain and corresponding equilibrium brings perspective to performances of other methods.

Method	**Reward**		**Terminated**
	i	j	**Timestep**
BMG	2.98	1.98	3
HBA Vs HBA	2.97	1.96	5
JAL Vs HBA	1.95	2.96	6
JAL Vs JAL	1.95	2.96	6
CJAL Vs HBA	2.67	1.66	35
CJAL Vs CJAL	2.67	1.66	35

Table 2: Comparison with HBA and MARL algorithms on the 2-agent 3 $\times$ 3 Foraging domain.

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. 4 (top), we report the average time to compute the first 10 equilibria for a 3-horizon $n = 2$ and 3MABC with $|X|$ = 2 and $|\Theta|$ = 16 and 64 types, respectively (4 types for each player). The bound on ϵ given τ in Proposition 1 is loose. Therefore, we consider various values of ϵ that are well within this bound. An example 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. 4 (bottom-left), increasing times for 5MABC for increasing horizons as we may expect with the exact subgame-perfect equilibrium taking just under 4 hours to compute for $H = 3$ and $\epsilon = 0.1$. Notice that this time increases considerably if we compute profiles in exact equilibria.

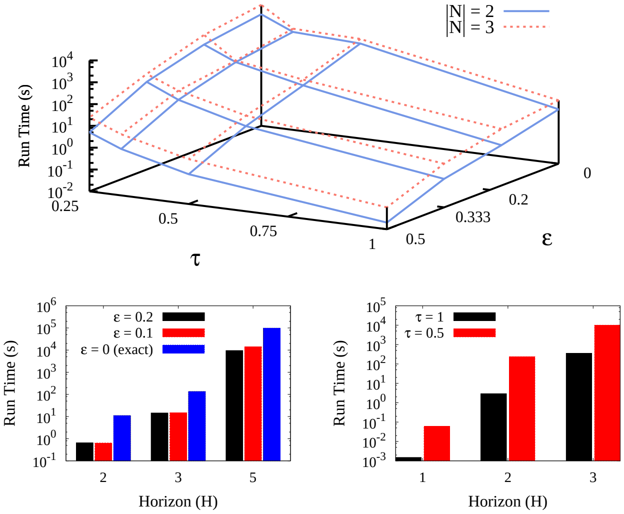


Figure 4: **Impact of parameters on performance**. Time taken to compute: (**top**) MPFLE in 2MABC and 3MABC for varying τ and ϵ at $H = 3$, (**bottom-left**) a pure-strategy MPFLE in 5MABC for varying ϵ and H *showing scalability in agents*, and (**bottom-right**) MPFLE in 2-agent 3$\times$3 Foraging for varying τ and H with $\epsilon = 0.1$ *showing scalability in domain size (in $|A|$ and $|S|$)*.

To scale in the number of states, we experimented on the larger 3 $\times$ 3 Foraging and illustrate empirical results in Fig. 4 (bottom-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 $\times$ 3 Foraging exhibited ex-interim values [1.98, 0.98] for players i and j, respectively. In general, we point out that 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	Without TE		With TE	
	$\lvert\Theta^{k=1}\rvert$	**Time (s)**	$\lvert\Theta^{k=1}\rvert$	**Time (s)**
3	16	0.07	4	0.11
	36	0.78	9	0.54
	64	1335.59	16	27.2
4	16	1.01	4	0.96
	36	42.6	12	14.3
	64	1481.06	16	311.2
5	16	1.56	4	1.26
	36	31.24	16	26.7
	64	>1 day	25	3161.7

Table 3: Computational savings due to TE while computing a pure-strategy MPFLE in 2MABC for Kets level 1 type spaces.

Type equivalence Rathnasabapathy et al. (2006) show how we may systematically and exactly compress large type spaces using exact behavioral equivalence. This *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 3 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, one of which is large in the context of ad hoc teaming 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 ad hoc coordination.

Discussion

There is growing interest in game-theoretic frameworks and their solutions that can model more pragmatic types of players. To the best of our knowledge, BMG is the first formalization of incomplete-information Markov games played by Bayesian players, which integrates types that induce finite-level beliefs into an operational framework. As repeated games are Markov games collapsed into a single state, the properties and solution of BMG presented in this paper are applicable to repeated Bayesian games with incomplete information as well. Importantly, BMGs are better suited for modeling ad hoc coordination in comparison to previous game-theoretic frameworks.

In conclusion, we ask the following questions as we further investigate BMGs. Does increasing the depth of reasoning get MPFLE "closer" to Bayes-Nash equilibria, and can we formalize the closeness? Are there profiles in MPFLE which do not occur in the set of Bayes-Nash equilibria 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 Bayes-Nash equilibrium with increasing levels. Kets (2014) establishes the presence of Bayes-Nash equilibria 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 Bayes-Nash equilibrium.

References

Banerjee, D., and Sen, S. 2007. Reaching pareto-optimality in prisoners dilemma using conditional joint action learning. *Autonomous Agents and Multi-Agent Systems* 15(1):91–108.

Claus, C., and Boutilier, C. 1998. The dynamics of reinforcement learning in cooperative multiagent systems.

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

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.

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

Camerer, C.; Ho, T.-H.; and Chong, J.-K. 2004. A cognitive hierarchy model of games. *Quarterly Journal of Economics* 119(3):861–898.

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

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.

Kets, W. 2014. Finite depth of reasoning and equilibrium play in games with incomplete information. Technical Report 1569, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

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.

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.

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.

Singh, S.; Soni, V.; and Wellman, M. 2004. Computing approximate bayes-nash equilibria in tree-games of incomplete information. In *5th ACM Conference on Electronic Commerce*, 81–90.

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.

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