# Representing Bayesian Games Without a Common Prior

Dimitrios Antos[1] and Avi Pfeffer[2]
[1]Harvard University
[2]Charles River Analytics
[1]antos@fas.harvard.edu
[2]apfeffer@rca.com

## ABSTRACT

Game-theoretic analyses of multi-agent systems typically assume that all agents have full knowledge of everyone's possible moves, information sets and utilities for each outcome. Bayesian games relax this assumption by allowing agents to have different "types," representing different beliefs about the game being played, and to have uncertainty over other agents' types. However, applications of Bayesian games almost universally assume that all agents share a common prior distribution over everyone's type. We argue, in concord with certain economists, that such games fail to accurately represent many situations. However, when the common prior assumption is abandoned, several modeling challenges arise, one of which is the emergence of complex belief hierarchies. In these cases it is necessary to specify which parts of other agents' beliefs are relevant to an agent's decision-making (or need be known by that agent). We address this issue by suggesting a concise way of representing Bayesian games with uncommon priors. Our representation centers around the concept of a *block*, which groups agents' view of (a) the game being played and (b) their posterior beliefs. This allows us to construct the *belief graph*, a graphical structure that allows agents' knowledge of other agents' beliefs to be carefully specified. Furthermore, when agents' views of the world are represented by extensive form games, our block structure places useful semantic constraints on the extensive form trees. Our representation can be used to naturally represent games with rich belief structures and interesting predicted behavior.

## Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

## General Terms

Economics, Game Theory, Human Factors

## Keywords

Bayesian games, common prior assumption, equilibrium, belief hierarchy, formalism, knowledge requirements

## 1. INTRODUCTION

Game theory has recently gained wide acceptance as an analysis tool for multi-agent systems. The theory allows intelligent agents to reason strategically about their behavior, as well as the behavior of other agents in the environment. It also offers powerful solution concepts, like Nash equilibrium. Yet game theory seems to sometimes make excessive and unreasonable assumptions about what agents in a system might know. For instance, it usually assumes that the choices available to every player, the set of outcomes of the game, and each player's utility for every such outcome, are all common knowledge.

Many situations in multi-agent systems are not adequately described by such a framework. Other agents' preferences and utility functions cannot generally be known with any certainty, while sometimes even the choices available to the players, or the observations they receive, might not be fully known. To an extent, Bayesian games (BGs) allow uncertainty over these aspects to be formally captured and reasoned about. In a BG, each player has some private information (e.g., her utility function or available moves), all of which are signified by her "type;" moreover, every player is assumed to know her type, but might be uncertain over other players' types. However, even BGs usually assume that the joint distribution over players' types is common knowledge. This assumption is known as the *common prior assumption* (CPA).

In many practical situations, the use of the CPA is unacceptable or misleading. The arguments in favor of the CPA typically maintain that the assumption is a consequence of the agents' rationality, or that it can be justified by agents' learning in the environment over time. Moreover, games in which the CPA holds are usually easier to represent. On the contrary, when the CPA is relaxed, representational difficulties emerge and the analysis of a game becomes harder (for three thoughtful treatises of the CPA, see [3], [1] and [6]).

This paper addresses two challenges. First, to make BGs with uncommon priors easier to model, it presents a novel technique for representing such games. This technique offers several advantages: Defining a game is conceptually easy, as the agents' private information and beliefs are conveniently grouped into *blocks*. Blocks may contain any representation for a game (e.g., extensive form). Moreover, the beliefs across the various blocks can be used to construct the *belief graph*, a structure that describes which parts of other agents' priors are relevant to the decision-making of a particular agent. The belief graph therefore reveals how much an agent needs to know about others' belief hierarchies to compute a solution (equilibrium) to the game.

Second, the paper provides a middle road between the two undesirable alternatives discussed above. Rather than having to choose between the priors being common knowledge or completely subjective, our formalism allows precise control of which priors a player of a particular type knows (or needs to know). It is possible in our formalism for one type of a player to be more knowledge-

**Cite as**: The title of your paper should be written here, Dimitrios Antos and Avi Pfeffer, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. XXX-XXX.

able about other players' priors than another type. This is especially useful when some types of players are "boundedly rational," and hence have limited knowledge of the game. Such a representational capability is achieved through the belief graph.

(A full version of the paper is available at the authors' website.)

## 2. THE REPRESENTATION

The main idea behind our formalism is that modeling a game becomes simpler if the agents' types and beliefs are captured in a conceptually appealing and graphical way. We therefore introduce the concept of a "block," and define a game as a *collection* of blocks $B$. A block $b \in B$ consists of two elements: (1) the model $m(b)$ an agent has about the world, and (2) the beliefs $\beta(b)$ the agent assumes, and believes others to assume, in that block. The agent's model is a complete game in normal or extensive form, with everyone's information sets, available moves and utilities fully specified. The beliefs in block $b$ consist of $n(n-1)$ probability distributions over $B$, indexed $p_{ij}^{b}$ for all $i, j \in N$, $i \neq j$, where $N$ is the set of agents ($|N| = n$). The distribution $p_{ij}^{b}$ captures agent $i$'s beliefs over which block agent $j$ might be using. Also, let us denote by $p_{ij}^{b}(b')$ the probability assigned to block $b'$ by the distribution $p_{ij}^{b}$.

It is straightforward to map this construct onto a Bayesian game. For each agent $i$, her typeset $T_i$ is equivalent to the set of blocks $B$. When agent $i$ is of a particular type, say $b \in B$, then agent $i$'s private information (utility, observations, etc.) are fully captured by the game $m(b)$. Moreover, $i$'s posterior distribution over the beliefs of all other agents given her type, $p(\mathbf{T}_{-i} = (t_j)_{j \neq i} \mid T_i = b)$, is given by the product $\prod_{j \neq i} p_{ij}^{b}(t_j)$ of the distributions in $\beta(b)$.

Notice how in our formalism the modeling is performed in terms of the posterior distributions $p_i(T_j \mid T_i = b) \equiv p_{ij}^{b}$, not the priors $p_i(T)$. Given these posteriors, *any* prior that is consistent with them will be essentially expressing the same game.

In each block $b$, the set of pure strategies for player $i$ contains all her pure strategies in the model $m(b)$. For the game as a whole, a pure strategy for $i$ is then a choice of pure strategy for every block $b \in B$. Moreover, if the models $m(b)$ are represented in extensive (tree) form, a pure strategy for $i$ for the whole game is a mapping from all information sets of all trees $m(b)$ to an action available to her in every such information set. Similarly, *mixed* strategies are probability distributions over pure strategies. Finally, a strategy profile $\sigma$ denotes, for every agent $i$ and every type $b \in B$, a choice of mixed (or behavioral) strategy $\sigma_{i,b}$ .

The main solution concept for a Bayesian game is a Bayes-Nash equilibrium. A strategy profile $\sigma$ is a Bayes-Nash equilibrium if, for all agents $i$ and for all types $b$, the strategy $\sigma_{i,b}$ maximizes $i$'s expected utility against strategies $\sigma_{j,b'}$ , where each is weighted according to the posterior distribution $p_{ij}^{b}(b')$. If the CPA is adopted, all agents agree on the game being played and therefore the equilibrium represents an optimal solution to it. Replacing the common prior with commonly known, differing priors maintains the agents' belief that this equilibrium is an optimal solution, but each of them thinks that only *her* utility is maximized in expectation under the equilibrium. Others' utilities are not necessarily maximized; they only *think*, using their erroneous priors, that their utilities are maximized. Hence equilibria are in a sense subjective solutions. On the other hand, if priors are also private, then it need not necessarily hold that agents even *agree* on what the equilibria of the game are. If the prior of agent $i$ is very different from the prior agent $j$ assumes for $i$, then clearly the equilibria of the game, as computed by the two agents, might be completely unrelated.

## 3. THE BELIEF GRAPH

One useful property of our formalism is that it allows for belief dependencies to be uncovered easily. In particular, it can help a modeler answer the question "Which of the beliefs of other agents are relevant to agent $i$'s decision-making?" This is performed by constructing the game's *belief graph*. The belief graph is constructed as follows: Its nodes are the set of blocks $B$. Then, we add an edge $(b, b')$ and we label it "$i.j$" if $p_{ij}^{b}(b') > 0$. In other words, the edge $(b, b')$ denotes that agent $i$ in block $b$ assumes that $j$ might be using block $b'$ as his model of the world. The destination block $b'$ may be the same as the source $b$ (self-edge). Next, we define a path $\pi = (b_1, \ldots, b_m)$ such that, for every node $b_k$ , where $k \in [1, m-1]$, there is an edge $(b_k, b_{k+1})$ and, for each consecutive edge pair $\{(b_k , b_{k+1}), (b_{k+1} , b_{k+2})\}$, where $k \in [1, m-2]$, the label of the first edge is "$i.j$" and the label of the second is "$j.k$" for some agents $i$, $j$ and $k$. (A path may very well contain self-edges.) We say that a block-agent pair $(b', j)$ is *reachable* from pair $(b, i)$ if there is a path from $b$ to $b'$ in which the first agent is $i$ and the last agent is $j$. The set of reachable blocks from $(b, i)$ is denoted by $R(b, i)$. The belief graph captures which distributions an agent needs to take into account in its decision-making. Only those posterior distributions $p_{l}^{b'''}$, where $(b''', l) \in R(b, i)$, are relevant to agent $i$'s decision-making, when that agent is in block $b$.

## 4. RELATED WORK

Other representations for games have been proposed as a solution to representational and conceptual aspects of game-theoretic analysis. Some of the most relevant are games of awareness [2], networks of influence diagrams [1], and I-POMDPs [5].

## 5. REFERENCES


[1] Ya'akov Gal and Avi Pfeffer. Networks of influence diagrams: Reasoning about agents' beliefs and decision-making processes. *Journal of Artificial Intelligence*, 33:109–147, 2008.

[2] Joseph Y. Halpern and Leandro Chaves Rêgo. Extensive games with possibly unaware players. In *AAMAS*, pages 744–751, 2006

[3] Stephen Morris. The common prior assumption in economic theory. *Economics and Philosophy*, 11:227–253, 1995.

[4] Roger B. Myerson. Harsanyi's Games with Incomplete Information, In *Management Science*, 2004 (special issue).

[5] Prashant Doshi and Piotr Gmytrasiewicz, On the Difficulty of Achieving Equilibria in Interactive POMDPs, *in AAAI*, 2006.

[6] Giacomo Bonanno, Klaus Nehring. How to make sense of the common prior assumption under incomplete information. In TARK, 3:133–146, Evanston, IL, 1998.