

# Bayesian Coalitional Games

Samuel Yeung and Yoav Shoham

Computer Science Department

Stanford University

{sieung, shoham}@cs.stanford.edu

## Abstract

We introduce *Bayesian Coalitional Games*<sup>1</sup> (BCGs), a generalization of classical coalitional games to settings with uncertainties. We define the semantics of BCG using the partition model, and generalize the notion of payoffs to contracts among agents. To analyze these games, we extend the solution concept of the core under three natural interpretations—*ex ante*, *ex interim*, and *ex post*—which coincide with the classical definition of the core when there is no uncertainty. In the special case where agents are risk-neutral, we show that checking for core emptiness under all three interpretations can be simplified to linear feasibility problems similar to that of their classical counterpart.

## 1 Introduction

In typical multiagent systems, individuals have limited capability and information. Agents often have to cooperate with one another to perform the desired tasks. As agents are ultimately interested in their own welfare, the question of payoff division is central to the formation of successful cooperative partnerships. Coalitional game theory (CGT) provides guidance as to how to divide the payoffs to achieve stability and fairness, and has been used in AI as means of achieving coordination (Sandholm and Lesser 1997).

However, most work in CGT to date has made two crucial assumptions. First, it assumes that the payoff to each coalition is given by a fixed, deterministic value. Second, it assumes that these values are common knowledge among all agents. Both assumptions often fail to hold for real-world problems. Consider the following example, adapted from (Chalkiadakis, Markakis, and Boutilier 2007).

Suppose a carpenter, a painter, and a stone mason are interested in forming a partnership for building houses. The revenue they can make depends on the skills of the individuals, and the jobs that come along. While each of them knows how skillful he is, he does not know how skillful the others are. To them, therefore, the revenue to the partnership is an uncertain payoff that depends on the true state of the world;

Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>This term has appeared in (Chalkiadakis and Boutilier 2007) as an abbreviation for Bayesian coalition formation problems. We employ the same term as it is the most descriptive of our work—a direct generalization of coalitional games to Bayesian settings.

further, they may have different beliefs about its value. To divide the benefits, the three of them may want to arrange with one another how to divide the revenue under different scenarios. Our interest lies in analyzing the stability and fairness of such “arrangements.”

In order to study these problems of cooperation under uncertainty, we generalize coalitional games to a Bayesian framework using the information partition model, which we call *Bayesian Coalitional Games* (BCGs). In BCGs, agents have a common prior over the set of possible coalitional games, and some private information about the true state of the world captured by their information partitions. To divide the possibly uncertain payoffs, agents enter into *contracts* with one another. As in CGT, we do not worry about how agents come to agree on a contract; rather, we are interested in properties of contracts. One such property that we study is *stability*. Specifically, we generalize the solution concept of the *core* and give conditions under which no group of agents would prefer to form a separate coalition. We also show that these conditions can be simplified to sets of linear constraints when agents are risk neutral.

There has been some work in Economics (Suijs 1999; Myerson 2005) and in AI (Chalkiadakis 2007) that addresses uncertainty in coalitional games in different ways. We discuss how our work is related to and yet different from these after presenting our model.

## 2 Coalitional Game Theory

CGT is the study of payoff division within groups of agents. A game assigns to each group of agents, called a *coalition*, a set of possible payoffs. Throughout this paper, we assume that the payoff to a coalition can be freely redistributed among its members. This is known as the *transferable utility* assumption, and is commonly made in CGT.

**Definition 1.** A *coalition game* (with transferable utility) (CG) is given by  $\langle N, v \rangle$ , where

- $N$  is a set of agents (the *grand coalition*); and
- $v : 2^N \mapsto \mathbb{R}$  is a function that maps each group of agents  $S \subseteq N$  to a real-valued payoff.

An *outcome* or *payoff vector* in a CG specifies how to divide the payoff of the grand coalition among the agents. A *solution concept* assigns to each CG a set of “reasonable”

outcomes. In this paper, we focus on stable outcomes as captured by the solution concept of the *core*.

Intuitively, the core attempts to characterize when an outcome is stable with respect to coalitional deviations. Stability under the core means that no set of players can jointly deviate to improve their payoffs.

**Definition 2.** An outcome  $x \in \mathbb{R}^N$  is in the *core* of CG  $\langle N, v \rangle$  if for all  $S \subseteq N$ ,

$$\sum_{i \in S} x_i \geq v(S)$$

For a given game, the core game may be *empty*, i.e., there may be no payoff vector that satisfies the stated condition.

### 3 Bayesian Coalitional Games

In this section we define Bayesian coalitional games and explain how they are analyzed. We also give a simple example to illustrate the key concepts.

#### 3.1 Semantics

We define Bayesian coalitional games (BCGs) using the partition model, similar to how (non-cooperative) Bayesian games are defined (Osborne and Rubinstein 1994). The two predominant approaches to modeling these non-cooperative games are based on possible worlds and on types. While the latter is mathematically most elegant, we find the former better suited here, since its semantics are more transparent. This turns out particularly useful for disambiguating solution concepts in the coalitional case. We note that (Myerson 2005; Chalkiadakis and Boutilier 2007) both use type-based formulations. We discuss this further in Section 4.

**Definition 3.** A *Bayesian coalitional game* is given by  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$  where

- $N = \{1, 2, \dots, n\}$  is a set of agents;
- $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$  is a set of *possible worlds*, where each world specifies a coalitional game defined over  $N$ ;
- $\mathbb{P}$  is a *common prior* over the worlds  $\Omega$ ;

and for each agent  $j$ ,

- $\mathcal{I}_j$  is agents  $j$ 's partition of the worlds  $\Omega$ ;
- $\succsim_j$  describes agent  $j$ 's preference over *distributions of payoffs*.

The interpretation of BCGs is as follows. There are a set of possible worlds  $\Omega$  in which the coalitional game may take place, drawn according to some probability distribution  $\mathbb{P}$ , *commonly known* among all agents. For each world, in addition to the common prior, each agent knows that the world lies in a subset of worlds that are indistinguishable from his point of view; these subsets are known as *information sets*, and together they form a partition of the worlds, called an *information partition*. The information partitions of all agents are also common knowledge. Each agent has some preference over the distribution of payoffs as captured by the relation  $\succsim$ . We are interested in the “agreements” that the agents might make; we formalize this in Subsection 3.2.

As mentioned in Section 2, we make the transferable utility assumption in this paper. In BCGs, this assumption applies to the *payoffs* in the individual worlds. In other words, the payoffs in each world is transferable. This does not apply to the game as a whole, as agents have individual preferences over distributions of payoffs. This will be made clear when we speak about coalitional deviation in Subsection 3.2.

**Notation.** We overload the notation  $\omega$  to denote the coalitional game in world  $\omega \in \Omega$ , i.e.,  $\omega(S)$  denotes the value of coalition  $S$  in world  $\omega$ . We use  $\mathcal{I}(\omega)$  to denote the set of worlds to which  $\omega$  belongs under partition  $\mathcal{I}$ . We use  $\mathbb{P}(\omega|I)$  to denote the conditional probability of the true world being  $\omega$  when the set of possible worlds is  $I$ .

#### 3.2 Contracts

In CGs, we specify a single number for each agent to be interpreted as his payoff. In BCGs, however, since agents may not know the exact value of a coalition, one cannot specify a precise payoff to an agent in advance before uncertainty is resolved. Instead, we assume that agents enter into agreements about how to divide the values of the coalitions. We call these *contracts*. They specify how payoffs should be divided *after* the true world is made *known* to all agents.

**Definition 4.** A *contract* among agents of coalition  $S$  ( $S$ -contract) is a mapping from the set of worlds to payoff vectors,  $\mathbf{c}^S : \Omega \mapsto \mathbb{R}^S$ , such that  $\mathbf{c}_j^S(\omega)$  denotes the payoff to agent  $j \in S$  in world  $\omega$ . A contract is *feasible* if for all worlds  $\omega$ ,  $\sum_{j \in S} \mathbf{c}_j^S(\omega) \leq \omega(S)$ .

Since contracts assign distributions of payoffs to agents, agents have preferences over contracts, induced by their preferences over distribution of payoffs,  $(\succsim_j)_{j \in N}$ .

In BCGs, we are interested in comparing different  $N$ -contracts (*grand contracts*). A solution concept specifies the conditions that grand contracts should satisfy, and captures desirable properties such as stability and fairness. Just as we do not worry about where payoff vectors come from in simple coalitional games, we do not worry about where contracts come from in BCGs; we take them as exogenously given. In this paper, we focus only on feasible contracts, and study the question of whether stable grand contracts exist.

#### 3.3 Stable Contracts—Core of BCGs

A natural question in BCGs is whether a grand contract is stable with respect to coalitional deviation. In other words, whether any group of agents is dissatisfied with it and prefers to divide the payoff to their coalition instead. To speak about this kind of deviation, we must first discuss what it means for a coalition to prefer one division of payoff to another.

**Definition 5.** Let  $\mathbf{X} = (X_j)_{j \in S}$  and  $\mathbf{Y} = (Y_j)_{j \in S}$  be two distributions of payoffs for agents in coalition  $S$ . Let the preferences of the agents be  $(\succsim_j)_{j \in S}$ . We say that  $S$  *weakly prefers*  $\mathbf{X}$  to  $\mathbf{Y}$ ,  $\mathbf{X} \succsim_S \mathbf{Y}$ , if for all agents  $j$ ,  $X_j \succsim_j Y_j$ . We say that  $S$  *strictly prefers*  $\mathbf{X}$  to  $\mathbf{Y}$ ,  $\mathbf{X} \succ_S \mathbf{Y}$ , if the preferences are strict for all agents.

Note that an agent only cares about the distribution of payoffs he receives, governed by the contract that specifies how payoffs to a coalition is to be divided. As a result, even if the

preferences of the agents can be described by utility functions, utilities are not transferable as they depend directly on how the contracts distribute the payoffs.

Given a contract, an agent's perception of its desirability depends on when it is evaluated. When it is evaluated before a world is drawn, we call the situation *ex ante*; when it is evaluated after a world is drawn, and each agent is made aware of the information set to which the world belongs in his partition of the worlds, but before the world itself is made known, we call it *ex interim*; when it is evaluated after the true world is made known, we call it *ex post*. These notions are borrowed from non-cooperative Bayesian games. To illustrate the differences, let  $G = \langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$  be a BCG and  $c^S$  be some  $S$ -contract.

*Ex ante*, the distribution of payoffs to agent  $j$  equals  $c_j^S$ .

*Ex interim*, the distribution of payoffs to agent  $j$  is conditional on the information set to which the true world  $\omega^*$  belongs in his partition of the worlds, namely

$$c_j^S(\omega) \text{ with probability } \mathbb{P}(\omega | \mathcal{I}_j(\omega^*)). \quad (1)$$

We denote this by  $c_j^S | \mathcal{I}_j(\omega^*)$ .

*Ex post*, no uncertainty remains. If the true world is  $\omega^*$ , the payoff to agent  $j$  is simply  $c_j^S | \omega^*$ .

For coalition  $T \subseteq S$ , we write  $c_T^S$  (resp.  $c_T^S | I_j$ ) to denote the set of distributions of payoffs to agents in  $T$  under contract  $c^S$  (resp. when  $j$ 's payoff is conditional on worlds  $I_j$ ).

We are now ready to define the notion of *blocking*, i.e., when a coalition is dissatisfied with a grand contract.

**Definition 6** (*Ex-Ante Blocking, Ex-Post Blocking*). Given a BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$  and a grand contract  $c^N$ , a coalition  $S$  *ex ante blocks*  $c^N$  if there exists an  $S$ -contract  $c^S$  such that

$$c_S^S \succ_S c_S^N. \quad (2)$$

A coalition  $S$  *ex post blocks*  $c^N$  if there exists world  $\omega^* \in \Omega$  and an  $S$ -contract  $c^S$  such that

$$c_S^S | \omega^* \succ_S c_S^N | \omega^* \quad (3)$$

We defer the discussion of *ex interim* blocking to Subsection 3.4 as there are additional subtleties with its definition.

The core of BCGs is defined analogously to the core of coalitional games using the more general notion of blocking.

**Definition 7** (*Core*). A grand contract  $c^N$  is in the (*ex ante*, *ex interim*, *ex post*) core of a BCG if no coalition  $S \subseteq N$  (*ex ante*, *ex interim*, *ex post*) blocks  $c^N$ .

In the special case where there is no uncertainty, i.e.,  $|\Omega| = 1$ , all three definitions coincide with the classical definition of the core in CG. As a consequence, as BCGs generalize CGs, all three kinds of core may be empty.

### 3.4 Ex-Interim Blocking

Let us start by examining the following intuitive-looking but *myopic* definition: a coalition  $S$  *ex-interim blocks* a contract  $c^N$  if for some world  $\omega^*$  and contract  $c^S$ ,

$$c_S^S | \mathcal{I}_j(\omega^*) \succ_S c_S^N | \mathcal{I}_j(\omega^*). \quad (4)$$

The problem with this definition is that for some contracts  $c^N$  and  $c^S$ , an agent may be able to *learn* more about the true state of the world if all agents in  $S$  prefer  $c^S$  to  $c^N$ . Consequently, he may refine his information set when evaluating this  $S$ -contract. Consider this example.

**Example 1.** Suppose there are two agents  $\{1, 2\}$  and two worlds  $\{\omega_1, \omega_2\}$ . Let the information partitions be  $\mathcal{I}_1 = \{\{\omega_1, \omega_2\}\}$ , and  $\mathcal{I}_2 = \{\{\omega_1\}, \{\omega_2\}\}$ . Consider some grand contract  $c^N$ . Suppose for some  $c^S$ ,

$$\begin{aligned} c_2^S | \mathcal{I}_2(\omega_1) &\succ_2 c_2^N | \mathcal{I}_2(\omega_1) \\ c_2^N | \mathcal{I}_2(\omega_2) &\succ_2 c_2^S | \mathcal{I}_2(\omega_2) \end{aligned}$$

Suppose the true world is  $\omega_1$ . In principle, agent 1 should not be able to tell whether the true world is  $\omega_1$  or  $\omega_2$  because of his information partition. However, he can reason that if both agents believe  $c^S$  is better than  $c^N$ , then the true world must be  $\omega_1$ , since agent 2 would not have preferred  $c^S$  otherwise. Notice that this reasoning does not require agent 1 to know agent 2's realized information set; it simply requires agent 1 to know agent 2's information partition.

Therefore, when evaluating  $c^S$ , agent 1 should only condition his distribution of payoffs on worlds where  $c^S$  is beneficial to both players. In this instance, he should compare  $c_1^S | \omega_1$  with  $c_1^N | \omega_1$ . This is because if the true world is  $\omega_2$ ,  $c^S$  will be rejected by agent 2 anyway, so it should not factor into agent 1's evaluation.  $\square$

This example suggests that when an agent evaluate a contract at the *ex-interim* stage, he should ignore worlds where other agents, given their information partitions, will find unattractive. To reason about this formally, we first specify the choices of the agents at the *ex-interim* stage.

**Definition 8.** Given a contract  $c^S$ , the *response*  $r_j$  of an agent  $j \in S$  is a mapping from  $\Omega$  to  $\{0, 1\}$ , with the constraint that  $r_j(\omega) = r_j(\omega')$  if  $\omega$  and  $\omega'$  belong to the same information set under partition  $\mathcal{I}_j$ . A contract  $c^S$  is *agreed* in world  $\omega$  if  $r_j(\mathcal{I}_j(\omega)) = 1$  for all agents  $j \in S$ .

We also need to define the *finest common coarsening* of a set of partitions, also known as the *common knowledge* among agents with those partitions (Fagin et al. 1995).

**Definition 9.** For a set of partitions  $(\mathcal{I}_j)_{j \in S}$  over worlds  $\Omega$ , a *common coarsening* is a partition  $\mathcal{I}$  over  $\Omega$  such that for all  $I \in \mathcal{I}$ , for all  $j$  and  $I_j \in \mathcal{I}_j$ , either  $I_j \subseteq I$ , or  $I_j \cap I = \emptyset$ . A *common coarsening* is the *finest* one if any further partitioning will fail to be a common coarsening. We denote the finest common coarsening of  $(\mathcal{I}_j)_{j \in S}$  by  $\mathcal{I}_S^\wedge$ .

We can now formalize “unattractiveness.”

**Definition 10.** Given BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$ , contracts  $c^N$  and  $c^S$ , and some subset of worlds  $\Omega' \subseteq \Omega$ . For coalition  $T \subseteq S$ , an information set  $I \in \mathcal{I}_T^\wedge$  is *T-dominated* in  $\Omega'$  if for all responses by agents in  $S$ , if contract  $c^S$  will be agreed in some subset of worlds  $I' \subseteq (I \cap \Omega')$ , then for all worlds  $\omega \in I'$ ,

$$c^S | \mathcal{I}_j(\omega) \cap I' \not\succeq_T c^N | \mathcal{I}_j(\omega) \cap I'.$$

In other words, for coalition  $T$ , when the set of worlds is restricted to  $\Omega'$ , for some information set  $I$  in partition  $\mathcal{I}_T^\wedge$ ,

over all responses by agents  $S \supseteq T$ , agents in  $T$  cannot all strictly prefer contract  $\mathbf{c}^S$ . The set of worlds  $I$  can therefore be eliminated from consideration, because the contract will never be agreed there. The reason why  $I$  has to be in  $\mathcal{I}_T^\wedge$  is because all agents in  $T$  needs to know the others know which set of worlds is under consideration, and that they know the other know that they know, etc.; this is captured by *common knowledge of I*.

After some worlds are eliminated, with respect to the remaining worlds, some previously undominated information set may now be dominated, and the worlds in this set can be eliminated. Formally,

**Definition 11.** Given BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$ , contracts  $\mathbf{c}^N$  and  $\mathbf{c}^S$ . A set of worlds  $\Omega^*(\mathbf{c}^N, \mathbf{c}^S)$  survives iterated elimination of dominated information sets if there exists a sequence of worlds,  $(\Omega_t)_{t=0}^K$ , such that

- $\Omega_0 = \Omega$  and  $\Omega_K = \Omega^*(\mathbf{c}^N, \mathbf{c}^S)$ .
- $\Omega_{t+1} = \Omega_t \setminus (I \cap \Omega_t)$ , where  $I \in \mathcal{I}_T^\wedge$  is  $T$ -dominated in  $\Omega_t$  for some coalition  $T \subseteq S$ .
- In  $\Omega_K$ , for all coalitions  $T \subseteq S$  and  $I \in \mathcal{I}_T^\wedge$ ,  $I$  is not  $T$ -dominated in  $\Omega_K$ .

This process can be likened to that of iterated elimination of dominated strategies in non-cooperative games. Note that the set of worlds that survives iterated elimination is unique.

**Lemma 1.** For BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$  and any contracts  $\mathbf{c}^N, \mathbf{c}^S$ ,  $\Omega^*(\mathbf{c}^N, \mathbf{c}^S)$  is unique.

The proof is omitted for brevity.

We can now define *ex-interim blocking*.

**Definition 12** (Ex-interim blocking). Given a BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succsim_j) \rangle$  and grand contract  $\mathbf{c}^N$ , a coalition  $S$  *ex-interim blocks*  $\mathbf{c}^N$  if there exists an  $S$ -contract  $\mathbf{c}^S$  such that for some  $\omega \in \Omega^*(\mathbf{c}^N, \mathbf{c}^S)$ ,

$$\mathbf{c}_S^S | \mathcal{I}_j(\omega) \cap \Omega^*(\mathbf{c}^N, \mathbf{c}^S) \succ_S \mathbf{c}_S^N | \mathcal{I}_j(\omega) \cap \Omega^*(\mathbf{c}^N, \mathbf{c}^S).$$

### 3.5 Illustration of Main Concepts

We now consider a simple example of BCG. Let  $N = \{1, 2\}$ , and consider five possible worlds with the following probability (prior) and values:

$\omega =$	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$
$\mathbb{P}(\omega)$	0.1	0.2	0.2	0.3	0.2
$\omega(\{1\})$	5	5	2	2	3
$\omega(\{2\})$	2	6	1	3	3
$\omega(\{1, 2\})$	6	12	4	6	7

Let the information partitions be

$$\begin{aligned} \mathcal{I}_1 &= \{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5\}\}; \\ \mathcal{I}_2 &= \{\{\omega_1\}, \{\omega_2, \omega_3\}, \{\omega_4\}, \{\omega_5\}\}. \end{aligned}$$

Finally, let  $\succsim_1$  be induced by the utility function  $u_1(X) = \mathbb{E}[X]$ , i.e., agent 1 is risk-neutral and only cares about his expected payoff, and  $\succsim_2$  be induced by the utility function  $u_2(X) = \mathbb{E}[X] - 0.2 \text{Var}[X]$ , modeling risk averseness.

To analyze this game, consider the following grand contract. (Payoffs appeared in order of the worlds.)

$$\mathbf{c}^N = [(3, 3), (9, 3), (1, 3), (3, 3), (4, 3)]$$

For individual agents, contracts play no role—they simply receive their own payoff in each world. Hence,  $\mathbf{c}_1^1$  and  $\mathbf{c}_2^2$  are simply distributed according to rows 1 and 2 in the table.

To determine if  $\mathbf{c}^N$  is in the *ex-ante* core, we compute

$$\begin{aligned} u_1(\mathbf{c}_1^N) &= \mathbb{E}[\mathbf{c}_1^N] = 4.0 \\ u_1(\mathbf{c}_1^1) &= \mathbb{E}[\mathbf{c}_1^1] = 3.1 \\ u_2(\mathbf{c}_2^N) &= \mathbb{E}[\mathbf{c}_2^N] - 0.2 \text{Var}[\mathbf{c}_2^N] = 3.0 - 0.0 = 3.0 \\ u_2(\mathbf{c}_2^2) &= \mathbb{E}[\mathbf{c}_2^2] - 0.2 \text{Var}[\mathbf{c}_2^2] = 3.1 - 0.5 = 2.6 \end{aligned}$$

Hence,  $u_1(\mathbf{c}_1^N) > u_1(\mathbf{c}_1^1)$  and  $u_2(\mathbf{c}_2^N) > u_2(\mathbf{c}_2^2)$ . Note that although  $\mathbf{c}_2^2$  has higher expectation, agent 2 prefers  $\mathbf{c}_2^N$  due to risk averseness. For the grand coalition, no contract can increase utilities for one agent without hurting the other since  $\mathbf{c}^N$  minimizes the variance for agent 2. Thus,  $\mathbf{c}^N$  is in the *ex ante* core of the game.

To decide if  $\mathbf{c}^N$  is in the *ex interim* core, we examine the distribution of payoffs conditional on the agents' respective information sets. For agent 1,

$\Omega' =$	$\{\omega_1, \omega_2\}$	$\{\omega_3, \omega_4\}$	$\{\omega_5\}$
$u_1(\mathbf{c}_1^1   \Omega')$	5.0	2.0	3.0
$u_1(\mathbf{c}_1^N   \Omega')$	7.0	2.2	4.0

And for agent 2,

$\Omega' =$	$\{\omega_1\}$	$\{\omega_2, \omega_3\}$	$\{\omega_4\}$	$\{\omega_5\}$
$u_2(\mathbf{c}_2^2   \Omega')$	2.0	2.3	3.0	3.0
$u_2(\mathbf{c}_2^N   \Omega')$	3.0	3.0	3.0	3.0

Hence, both agents do worse by themselves. We also need to check to see if the grand coalition can do better with some other contracts. Consider the following alternative

$$\mathbf{c}'^N = [(6, 0), (7.5, 4.5), (1.5, 2.5), (2.8, 3.2), (4, 3)]$$

Suppose the true world is  $\omega_4$ . If we naively apply the myopic *ex-interim* criterion of Equation (4), we find that

$$u_1(\mathbf{c}'^N | \{\omega_3, \omega_4\}) = 2.3, \quad u_2(\mathbf{c}'^N | \{\omega_4\}) = 3.2,$$

and it would appear that both agents are better off. However, if we apply the process of iterated elimination,  $\omega_1$  is first eliminated since agent 1 knows agent 2 would prefer  $\mathbf{c}^N$ . After that,  $\omega_2$  is eliminated since, without  $\omega_1$ , agent 1 is worse in information set  $\{\omega_2\}$ . Continuing this process, all worlds are eventually eliminated, except  $\omega_5$ . However, no agent is strictly better off in  $\omega_5$  hence  $\mathbf{c}'^N$  does not *ex-interim* block  $\mathbf{c}^N$ . Note that iterated elimination is necessary, as  $\omega_3$  is not eliminable initially in agent 1's view because agent 2 is also better off in information set  $\{\omega_3, \omega_4\}$  with utility of 3.3. The contract  $\mathbf{c}^N$  is in fact in the *ex-interim* core of the game, since  $\mathbf{c}^N$  minimizes the variance for agent 2 in each of her information sets.

On the other hand, it is also easy to verify that  $\mathbf{c}^N$  is not in the *ex post* core of the game.

## 4 Related Work

There are three strands of work that are most relevant to this paper, in which uncertainty is modeled probabilistically.

They are CGs with random payoffs, incentive-compatible CGs under uncertainty, and Bayesian coalition formation problems. We discuss them in turn.

CGs where payoffs may be random were first analyzed in (Charnes and Granot 1973). In (Suijs et al. 1998), the authors generalized the model and introduced risk preferences and coalitional actions. Both are important for modeling cooperation in financial and insurance applications. The authors studied properties of payoff divisions of the form  $(\mathbf{d}, \mathbf{r})$ , where  $\mathbf{d}$  gives an agent a deterministic share of the payoff, and  $\mathbf{r}$  a *relative* share. See (Suijs 1999) for a survey.

Assuming that each coalition has only one action, BCG is a strict generalization of Suijs' model. Their model restricts the information partition of each agent to be the set of all worlds  $\Omega$ . The payoff divisions that they are interested in places a restriction on the space of contracts. Their generalization of the core is equivalent to our *ex-ante* core, and incidentally also the *ex-interim* core, as the two concepts are identical when the information partitions of all agents equal  $\Omega$ . They also consider coalitional action; our model can be generalized to take that into account.

Incentive-compatible CGs under uncertainty are presented in (Myerson 2005). In his model, agents have uncertain types, and payoffs to coalitions depend on the actions of their members and the types of all agents. The outcome of a game is determined by *mechanisms* that take reported agent types as input, pick actions for the agents, and redistribute the payoffs. Mechanisms are required to be *truthful*, i.e., agents should have no incentives to misreport their types. An outcome, induced by a mechanism, ultimately assigns payoffs to agents as a function of their types; this can be interpreted as a contract in our model. An outcome is in the core defined by Myerson (*M-core*) if it is not *blocked*, where an outcome is blocked if there exists a *blocking mechanism* that selects a *random* coalition according to some known distribution, asks agents their types, picks actions, and reallocates payoffs such that no agent in the selected coalition is worse off compared to the blocked outcome, and the mechanism makes a profit *in expectation*.

The primary difference between the M-core and ours is that blocking is caused by some blocking mechanism that can choose from a *distribution* of coalitions, rather than caused by a single coalition, as in ours or classical CGT. The M-core requires contracts to be feasible only in expectation, whereas ours requires feasibility in all worlds. Finally, the requirement that mechanisms are truthful constrains payoff distributions across different worlds.

The work closest to our model is the Bayesian coalitional formation problems introduced by Chalkiadakis *et al.* in (2004; 2007; 2007). They investigate a host of problems, including learning, bargaining, and coalitional stability. Full discussion of their work and its connection to ours will appear in the full version of the paper, but the following will help explain the main similarity and differences.

In their model, agents have uncertain types, and payoffs to coalitions depend on types and actions. The outcome of a game is specified by three components  $\langle CS, \mathbf{a}, \mathbf{d} \rangle$ : a coalition structure  $CS$  that specifies the coalitions formed, the actions  $\mathbf{a}$  chosen by the coalitions, and the agents' demands

$\mathbf{d}$  of relative share of payoffs. For comparison, we focus on the case when the grand coalition is formed in their model, i.e.,  $CS = \{N\}$ . An outcome,  $\langle \{N\}, \mathbf{a}, \mathbf{d} \rangle$ , specifies the payoffs to the agents as a function of their types. This can be interpreted as a contract, with constraints over the payoff distribution across different worlds governed by  $\mathbf{d}$ .

Chalkiadakis *et al.* propose three notions of stability. The weak Bayesian core (*weak C-core*) defines blocking as the existence of a coalition of which its members each believes he is better off in the deviating coalition. The strict Bayesian core coincides with the weak one under the transferable utility assumption made in this paper. The strong Bayesian core (*strong C-core*) defines blocking as having an agent who believes there is an alternative outcome in which he is better off and no one else in the deviating coalition is worse off. These three cores differ from our *ex-interim* core with respect to how blocking is defined.

Let us call the set of  $S$ -contracts, for some deviating coalition  $S$ , that an agent  $i \in S$  considers preferable to the current outcome his *defection set*. One can view each notion of blocking as characterized by three considerations: (a) Does the defection set of an agent take into account the other agents' reasoning at all? (b) If it does, at what level of depth does it do it, and in particular, does it consider the other agents' own modeling of other agents? And (c) how are the agents' individual defection sets aggregated to define blocking? Do all agents in the deviating coalition have to prefer to defect, i.e., one should take the *intersection* of these sets, or just some agent prefers to defect, i.e., take the *union* of these sets? The differences are summarized as follows:

Concept	(a)	(b)	(c)
Weak C-core	No	N/A	$\cap$
Strong C-core	Yes	One level of modeling (with respect to other agents' expected payoff), but no mutual modeling	$\cup$
Ex-interim core	Yes	Infinite level of mutual modeling	$\cap$

Given that the M-core, the C-cores, and the ex-interim core can all be interpreted as contracts in our model (in some cases with transformations), a natural question is whether these sets of contracts that constitute these cores satisfy any set relationship, such as subset or non-empty intersection. We conjecture that they do not satisfy any relationship in general, but may be related in special cases.

Finally, other forms of uncertainty in coalitional games have been studied in (Yamamoto and Sycara 2001; Blankenburg, Klusch, and Shehory 2003; Li et al. 2003; Yokoo et al. 2005). Their models are different from BCG as uncertainty is not explicitly modeled as probabilities in these papers.

## 5 Risk-Neutral Agents

We now explore an important special case of BCGs when agents are risk-neutral. We show that checking for whether a grand contract belongs to any of the three kinds of core can be simplified to sets of linear constraints. Let us start by defining risk-neutrality.

**Definition 13.** An agent is *risk-neutral* if his preference over distributions of payoffs is such that  $X \succcurlyeq Y$  if and only if the

expectation of  $X$  is at least that of  $Y$ ,  $\mathbb{E}[X] \geq \mathbb{E}[Y]$ .

For risk-neutral agents, checking whether the *ex-ante* core of a BCG is empty is equivalent to checking whether the core is empty in a related simple CG.

**Theorem 1.** *Given a BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succ_j) \rangle$  with risk-neutral agents, a grand contract  $\mathbf{c}^N$  is in the ex-ante core of the game if and only if for all  $S \subseteq N$ ,*

$$\mathbb{E}[\omega(S)] \leq \sum_{j \in S} \mathbb{E}[\mathbf{c}_j^N]$$

The proof is omitted for brevity.

Theorem 1 reduces the checking of *ex-ante* core membership of BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succ_j) \rangle$  to checking the core membership of the CG with value of coalition  $S$  equals the expected value of the coalition  $S$  in the BCG,  $\mathbb{E}[\omega(S)]$ . A similar reduction is noted in (Suijs and Borm 1996).

The situation is more complicated for the *ex-interim* core. For coalition  $S$ , let us denote by  $\mathcal{I}_S^\vee$  the set  $\{I : I = \bigcap_{j \in S} I_j \text{ for all } I_j \in \mathcal{I}_j\}$ . Note that  $\mathcal{I}_S^\vee$  is a partitions over the worlds  $\Omega$ . We establish the following lemma.

**Lemma 2.** *Given BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succ_j) \rangle$  with risk-neutral agents, a coalition  $S$  can ex-interim block a grand contract  $\mathbf{c}^N$  if and only if there exists  $I^* \in \mathcal{I}_S^\vee$  such that*

$$\mathbb{E}[\omega(S)|I^*] > \sum_{j \in S} \mathbb{E}[\mathbf{c}_j^N | I^*] \quad (5)$$

The proof is omitted for brevity. Using this lemma, checking for membership in *ex-interim* core can also be simplified.

**Theorem 2.** *Given BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succ_j) \rangle$  with risk-neutral agents, a grand contract  $\mathbf{c}^N$  is in the ex-interim core of the game if and only if for all  $S \subseteq N$ , for all information sets  $I \in \mathcal{I}_S^\vee$ ,*

$$\mathbb{E}[\omega(S)|I] \leq \sum_{j \in S} \mathbb{E}[\mathbf{c}_j^N | I]$$

Checking whether a grand contract belongs to the *ex-post* core of a BCG is conceptually simple.

**Theorem 3.** *Given BCG  $\langle N, \Omega, \mathbb{P}, (\mathcal{I}_j), (\succ_j) \rangle$  with risk-neutral agents, a grand contract  $\mathbf{c}^N$  is in the ex-post core of the game if and only if for all  $S \subseteq N$ , for all worlds  $\omega$ ,*

$$\omega(S) \leq \sum_{j \in S} \mathbb{E}[\mathbf{c}_j^N | \omega] \left( = \sum_{j \in S} \mathbf{c}_j^N(\omega) \right)$$

In other words,  $\mathbf{c}^N$  is in the *ex-post* core of BCG if and only if it is in the core of each constituent game. The proof is straight-forward and omitted for brevity.

While determining whether the core is empty in all three cases simplify to sets of linear constraints, this does *not* mean that the problems are necessarily easy to solve, as the number of constraints are large (exponential in the number of agents, polynomial in the number of information sets and worlds). Nonetheless, this result may help to design efficient algorithms that takes advantage of linearity of constraints.

As a consequence of the three characterization theorems, we obtain the following interesting corollary.

**Corollary 1.** *For BCG with risk-neutral agents,*

$$\text{Ex-ante core} \supseteq \text{Ex-interim core} \supseteq \text{Ex-post core}$$

Note that this relationship is not necessarily true for general BCGs; it fails to hold for our example in Subsection 3.5.

## 6 Concluding Remarks

We introduce BCGs, a generalization of classical CGs to a Bayesian setting. We study payoff division in the form of contracts, and generalize the core in three ways to capture different notions of coalitional stability. We also show that checking for core emptiness reduces to linear feasibility problems when agents are risk-neutral.

There are many open questions regarding BCGs. Chief among them is about representation. Direct specification of CGs as a set function takes space exponential in the number of agents. For BCGs, there is the additional complication of having to specify a CG in each world. We are interested in designing compact representation scheme for BCGs and efficient algorithms for finding contracts in the core.

## 7 Acknowledgments

We acknowledge the helpful feedback from the anonymous reviewers. This research is funded by NSF ITR-0205633. Samuel Ieong is funded by a Stanford Graduate Fellowship.

## References

- Blankenburg, B.; Klusch, M.; and Shehory, O. 2003. Fuzzy kernel-stable coalitions between rational agents. In *2nd AAMAS*.
- Chalkiadakis, G., and Boutilier, C. 2004. Bayesian reinforcement learning for coalition formation under uncertainty. In *3rd AAMAS*.
- Chalkiadakis, G., and Boutilier, C. 2007. Coalitional bargaining with agent type uncertainty. In *Proc. 20th IJCAI*.
- Chalkiadakis, G.; Markakis, E.; and Boutilier, C. 2007. Coalition formation under uncertainty: Bargaining equilibria and the bayesian core stability concept. In *Proc. 6th AAMAS*.
- Chalkiadakis, G. 2007. *A Bayesian Approach to Multiagent Reinforcement Learning and Coalition Formation under Uncertainty*. Ph.D. Dissertation, University of Toronto.
- Charnes, A., and Granot, D. 1973. Prior solutions: Extensions of convex nucleolus solutions to change constrained games. In *Proc. Comp. Sci. and Stat. 7th Symp. Iowa State Univ.*
- Fagin, R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning About Knowledge*. The MIT Press.
- Li, C.; Rajan, U.; Chawla, S.; and Sycara, K. 2003. Mechanisms for coalition formation and cost sharing in an electronic marketplace. In *Proc. ICEC'03*.
- Myerson, R. B. 2005. Virtual utility and the core for games with incomplete information. Available at <http://home.uchicago.edu/~rmyerson/research/>.
- Osborne, M. J., and Rubinstein, A. 1994. *A Course in Game Theory*. Cambridge, Massachusetts: The MIT Press.
- Sandholm, T., and Lesser, V. 1997. Coalitions among computationally bounded agents. *Artificial Intelligence* 94(1):99–137.
- Suijs, J., and Borm, P. 1996. Cooperative games with stochastic payoffs: Deterministic equivalents. FEW Memo. 713, Tilburg U.
- Suijs, J.; Borm, P.; De Waegenaere, A.; and Tijs, S. 1998. Cooperative games with stochastic payoffs. *Euro OR* 113:193–205.
- Suijs, J. 1999. *Cooperative Decision-Making under Risk*. Kluwer.
- Yamamoto, J., and Sycara, K. 2001. A stable and efficient buyer coalition formation scheme for e-marketplaces. In *AGENTS '01*.
- Yokoo, M.; Conitzer, V.; Sandholm, T.; Ohta, N.; and Iwasaki, A. 2005. Coalitional games in open anonymous environments. In *Proc. 20th AAAI*.