Simple Coalitional Games with Beliefs

Georgios Chalkiadakis* and **Edith Elkind*,†** and **Nicholas R. Jennings***
{gc2,ee,nrj}@ecs.soton.ac.uk
* School of Electronics and Computer Science, University of Southampton, UK
† Nanyang Technological University, Singapore

Abstract

We introduce *coalitional games with beliefs (CGBs)*, a natural generalization of coalitional games to environments where agents possess private beliefs regarding the capabilities (or types) of others. We put forward a model to capture such agent-type uncertainty, and study coalitional stability in this setting. Specifically, we introduce a notion of the *core* for CGBs, both with and without coalition structures. For *simple* games *without* coalition structures, we then provide a characterization of the core that matches the one for the full information case, and use it to derive a polynomial-time algorithm to check core non-emptiness. In contrast, we demonstrate that in games *with* coalition structures allowing beliefs increases the computational complexity of stability-related problems. In doing so, we introduce and analyze weighted voting games with beliefs, which may be of independent interest. Finally, we discuss connections between our model and other classes of coalitional games.

1 Introduction

In many multi-agent scenarios, agents have individual goals, but need to form teams (coalitions) in order to achieve those goals. When agents are *selfish*, i.e., aim to optimize their own utility, stability of the resulting coalitions becomes a major issue: if agents constantly switch coalitions to improve their payoffs, no task will ever be completed. Such settings are often modeled using coalitional games [Myerson, 1991], which provide a rich framework for the study of cooperation (see, e.g., [Shehory and Kraus, 1995]). In particular, coalitional stability can be studied using the notion of the *core*, which is one of the main solution concepts in coalitional games.

While providing a solid theoretical foundation for the analysis of multi-agent scenarios, traditional models of coalitional game theory fail to capture certain important aspects of interaction between intelligent agents. In particular, real-world agents often embark on the task of forming coalitions while having incomplete information or possessing private beliefs about the capabilities (or *types*) of potential partners. Hence, in multi-agent applications, we would like to be able to reason about coalition formation and coalitional stability under uncertainty. Nevertheless, until recently, the latter did not receive much attention from the game theory community (an important exception is a paper of Myerson [1984], which, however, approaches this issue from a mechanism design perspective). The last few years have seen some progress on this problem, with a number of models for coalition formation under incomplete information being proposed [Chalkiadakis and Boutilier, 2004; Myerson, 2007; Ieong and Shoham, 2008]. However, all of these papers assume that agent beliefs about other agents' types are represented as probability distributions over possible types. This assumption is not realistic in scenarios where agents may not be well-informed or sophisticated enough to describe their beliefs in this way. Moreover, as the number of agents grows, such models can quickly become computationally intractable. For example, in disaster management or multi-robot exploration scenarios, agents usually have beliefs about others' capabilities, but may not have time and/or bandwidth to check if these beliefs are correct and update them properly.

To address these issues, in this paper we propose a simplified approach to modeling incomplete information. Our model is based on the following idea. Sometimes, an agent's belief about another agent's type can be best described by a single element of the type space, which represents the former agent's best guess about the latter agent's type. While the agents realize that their information may be imprecise, they are unwilling or unable (e.g., due to time and space constraints) to estimate the chances of other agents having a different type, so they operate based on these guesses. In this paper, we provide a formal model for this type of scenario. While it can be viewed as a special case of the probabilistic model mentioned above, our model provides important advantages from both cognitive and computational perspective. Indeed, as argued above, it may be easier for an agent to formulate a single guess about another agent's capabilities than to come up with a distribution that fully describes them. Also, in our model each agent's beliefs can be compactly represented as a vector of types, one for each other agent. Finally, we are able to prove a number of characterization and complexity results for our setting that seem to be difficult to state—and prove—in the general probabilistic model.

The rest of this paper is organized as follows. After introducing the nesessary notation and definitions, we describe our model. We then generalize the well-known notion of a *simple* game to our setting, and provide a characterization of the core of such games in the setting where *coalition structures* are not allowed, i.e., the agents are required to form a single team. Simple games are ones in which the value of each coalition is known (or, in our setting, is believed) to be either 0 or 1 (rather than any real number). While less expressive than general games, they still capture a wide range of realistic scenarios and have been studied in the context of multi-agent systems (see, e.g., [Bachrach and Rosenschein, 2007; Elkind *et al.*, 2007]). Our characterization result matches the existing characterization result for the core of simple games without beliefs, and allows us to derive a polynomial-time algorithm for checking whether a given outcome is in the core, or whether the core is non-empty. Thus, we show that introducing beliefs in simple games without coalition structures does not increase the complexity of core-related problems.

In contrast, we then demonstrate that when coalition structures are allowed, i.e., agents can split into several teams, the complexity of core-related problems can increase dramatically. Specifically, we focus on weighted voting games with beliefs, a natural generalization of the classic weighted voting games (WVGs) [Taylor and Zwicker, 1999] to our setting. The computational complexity of WVGs when coalition structures are allowed has been studied in [Elkind *et al.*, 2008a]. In particular, [Elkind *et al.*, 2008a] show that when the weights of all players are polynomially bounded, there is a polynomial-time algorithm for checking if a given outcome is in the *CS-core* (a generalization of the concept of the core to games with coalition structures), and leave open the complexity of checking the non-emptiness of the CS-core. In this paper, we demonstrate that both of these problems become computationally hard in the setting with beliefs. These results provide a complexity-theoretic separation between games with and without beliefs in the setting with coalition structures. Moreover, they contribute to our understanding of weighted voting games with beliefs. Such games provide perhaps the simplest possible model for joint task execution by a team of selfish agents under incomplete information, and are therefore interesting in their own right, constituting an important base case in this area. We conclude the paper by positioning our model with respect to other classes of coalitional games, and pointing out directions for future research.

2 Preliminaries and Notation

A (transferable-utility) coalitional game $G = (N; u)$ is given by a set of *players* $N = \{1, \ldots, n\}$ and a *utility function* $u : 2^N \to \mathbb{R}$. The function u maps every subset, or *coalition*, of players S to a real number $u(S)$, which is called the *value* of S. Intuitively, $u(S)$ is the profit that the members of S can attain by working together.

This definition can be extended to the setting where each player may have a *type*, and the payoff of a coalition is determined not only by the list of its members, but also by their types. Formally, a *coalitional game with types* $G = (N; \mathcal{T}_1, \ldots, \mathcal{T}_n; u)$ is described by: (a) a set of players $N = \{1, \ldots, n\}$; (b) for each player i, a set $\mathcal{T}_i$ of this player's possible *types*; we require $\perp \notin \mathcal{T}_i$ for all $i \in N$; and (c) a *utility function* $u : \mathcal{T}_1 \cup \{\perp\} \times \ldots \times \mathcal{T}_n \cup \{\perp\} \to \mathbb{R}$. Given an input vector $(T_1, \ldots, T_n)$, $T_i \in \mathcal{T}_i \cup \{\perp\}$, the function u outputs the value of the coalition $S = \{i \mid T_i \neq \perp\}$ when each $i \in S$ has type $T_i \in \mathcal{T}_i$. Note that $T_i = \perp$ is used to denote the fact that player i does not appear in a given coalition. We write $\vec{\mathcal{T}}$ to denote $(\mathcal{T}_1, \ldots, \mathcal{T}_n)$.

As an example, consider the setting where each player has either skill A or skill B, but not both, and a coalition has value 1 if it has at least k_1 players with skill A and at least k_2 players with skill B, and 0 otherwise. This scenario can be modeled by a coalitional game with types where $\mathcal{T}_i = \{A, B\}$ for all $i \in N$, and $u(T_1, \ldots, T_n) = 1$ if and only if $|\{i \mid T_i = A\}| \geq k_1$ and $|\{i \mid T_i = B\}| \geq k_2$.

A coalitional game is called *monotone* if $u(S) \geq u(S')$ for all $S \subseteq N$ and all $S' \subset S$. A monotone game is called *simple* if $u(S) \in \{0, 1\}$ for all $S \subseteq N$. In such games, we say that S *wins* if $u(S) = 1$ and S *loses* otherwise.

An important special class of simple games is that of *weighted voting games (WVGs)*. A WVG is described by its set of players N, a vector of players' *weights* $\mathbf{w} = (w_1, \ldots, w_n)$, $w_i \in \mathbb{R}$ for $i \in N$, and a *threshold* $q \in \mathbb{R}$; we write $G = (N; \mathbf{w}; q)$. The utility function $u(S)$ of a game $G = (N; \mathbf{w}; q)$ is given by $u(S) = 1$ iff $\sum_{i \in S} w_i \geq q$.

In some cases, the only acceptable outcome of a game is the formation of the *grand coalition*, i.e., the coalition of all players N. However, in many multi-agent scenarios it is more natural for agents to split into groups so that each group performs its own task. This is captured by the notion of a *coalition structure*, which is a partition $CS = \{C^1, \ldots, C^k\}$ of the set of agents N, i.e., (a) $C^i \cap C^j = \emptyset$ for all $i, j = 1, \ldots, k$, $i \neq j$, and (b) $\cup_{i=1}^k C^i = N$. Note that our description of a coalitional game as $G = (N; u)$ does not prescribe whether the grand coalition should form; rather, each coalitional game induces two different games: one where coalition structures are not permitted and one where they are.

The utility function u does not specify how the members of a coalition should divide the value of this coalition. This is captured by the notion of an imputation. In games without coalition structures, an imputation is simply a way to distribute the value of the grand coalition. That is, we say that a vector $\mathbf{p} = (p_1, \ldots, p_n)$ is an *imputation* in a game without coalition structures if (a) $p_i \geq 0$ for all $i \in N$, and (b) $\sum_{i \in N} p_i = u(N)$. In a game with coalition structures, an imputation should distribute the value of each coalition in the coalition structure. That is, we say that $\mathbf{p} = (p_1, \ldots, p_n)$ is an *imputation for a coalition structure* $CS = \{C^1, \ldots, C^k\}$ if (a) $p_i \geq 0$ for all $i \in N$, and (b) $\sum_{i \in C^j} p_i = u(C^j)$ for all $j = 1, \ldots, k$. Note that the value of each coalition in the coalition structure has to be distributed among its members, i.e., inter-coalitional transfers are not allowed. We write $p(S)$ to denote $\sum_{i \in S} p_i$, and use similar notation for other n-dimensional vectors throughout the paper.

An important consideration in coalitional games is that of stability, which is usually captured by the notion of the core. Roughly speaking, an outcome of a game is stable if no set of players wants to deviate from it. In games without coalition structures, an outcome of a game can be identified with an imputation. Hence, in such games, the *core* of a game $G = (N; u)$ is defined as the set of all imputations $\mathbf{p}$ such that $p(S) \geq u(S)$ for any $S \subseteq N$. In games with coalition structures, an outcome is a pair $(CS, \mathbf{p})$, where CS is a coalition structure and $\mathbf{p}$ is an imputation for CS. In this setting, the *core* of a game $G = (N; u)$ consists of all pairs $(CS, \mathbf{p})$ such that CS is a partition of N, $\mathbf{p}$ is an imputation for CS, and $p(S) \geq u(S)$ for any $S \subseteq N$. In what follows, we refer to the core of a game with coalition structures as its *CS-core*, and reserve the term *core* for games without coalition structures.

3 Coalitional Games with Beliefs

In this section, we introduce our model for *coalitional games with beliefs*. These games are based on coalitional games with types. However, in our model the players' types are not public information. Instead, each player knows his own type and, in addition, has a *belief* about each other player's type.

Definition 1. *A* coalitional game with beliefs $G = (N; \vec{\mathcal{T}}; u; \mathbf{b}_1, \ldots, \mathbf{b}_n)$ *is given by a coalitional game with types* $G' = (N; \vec{\mathcal{T}}; u)$ *and, for each player* $i \in N$*, a vector* $\mathbf{b}_i = (b_i(1), \ldots, b_i(n))$ *of that player's* beliefs*, where* $b_i(j) \in \mathcal{T}_j$ *denotes* i*'s belief about player* j*'s type. We denote* $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ *by* $\vec{\mathbf{b}}$*. We refer to* G' *as the* associated game with types *of the game* G *and write* $G = (G', \vec{\mathbf{b}})$.

Note that we do not explicitly specify the players' true types; rather, we assume that i's belief about its own type is correct, i.e., $b_i(i)$ is player i's true type. In fact, all of our subsequent results hold true in the case where agents do not know their own true types but only have beliefs about them.

A player i can form a belief about the value of every possible coalition S based on his beliefs about individual players' types by computing $u(T_1, \ldots, T_n)$, where $T_j = b_i(j)$ for $j \in S$ and $T_j = \perp$ for $j \notin S$. We will denote i's belief about the value of a coalition S by $u_i(S)$.

As argued above, this model is inspired by Bayesian coalitional games with agent-type uncertainty [Chalkiadakis and Boutilier, 2004; Chalkiadakis *et al.*, 2007], but with an important difference. In contrast to that work, we do not assume here that the agents' beliefs are probabilistic. Instead, the agents have "point" beliefs about the type of others. This is a natural assumption in many real-world situations, where probabilistic beliefs are simply not available to the agents, or where reasoning about the full joint distribution over the possible type vectors of all potential coalitions is a practical impossibility for computational reasons.

Now, to describe an outcome of a coalitional game with beliefs, we need to generalize the notion of imputation to this setting. This task is complicated somewhat by the fact that the players may have different beliefs about the value of the coalition to be formed. To tackle this, we adopt an approach used in previous work on coalitional games with uncertainty [Suijs and Borm, 1999; Chalkiadakis and Boutilier, 2004; Chalkiadakis *et al.*, 2007]. Namely, the players have to agree on the *shares* of the total payoff that each of them is going to get rather than the actual payoff amounts.

Definition 2. *We say that* $\mathbf{d} = (d_1, \ldots, d_n)$ *is a* demand vector *for a coalition structure* $CS = \{C^1, \ldots, C^k\}$ *in a game* $G = (N; \vec{\mathcal{T}}; u; \vec{\mathbf{b}})$ *if* $d_i \geq 0$ *for* $i \in N$ *and* $d(C^j) = 1$ *for each* $C^j \in CS$*. Given a demand vector* $\mathbf{d}$*, the* expected payoff p_i *of a player* $i \in C^j$ *is given by* $d_i u_i(C^j)$*; we refer to* $\mathbf{p} = (p_1, \ldots, p_n)$ *as the* expected payoff vector *for* $\mathbf{d}$.

Similarly to the case of coalitional games with types, we can identify the outcome of a game without coalition structures with a demand vector for $\{N\}$. In a game with coalition structures, an outcome is a pair of the form $(CS, \mathbf{d})$, where CS is a partition of N and $\mathbf{d}$ is a demand vector for CS. We are now ready to define the core and the CS-core of a coalitional game with beliefs.

Definition 3. *Given a game* $G = (N; \vec{\mathcal{T}}; u; \vec{\mathbf{b}})$ *the* core *of* G *consists of all demand vectors* $\mathbf{d}$ *such that for any coalition* S *and any demand vector* $\mathbf{d}'$ *for* $\{S, N \setminus S\}$ *there exists an* $i \in S$ *such that* $d'_i u_i(S) \leq p_i$*. The* CS-core *of* G *consists of all pairs* $(CS, \mathbf{d})$ *such that* CS *is a partition of* N*,* $\mathbf{d}$ *is a demand vector for* CS*, and for any coalition* S *and any demand vector* $\mathbf{d}'$ *for* $\{S, N \setminus S\}$ *there exists an* $i \in S$ *such that* $d'_i u_i(S) \leq p_i$.

Intuitively, this definition says that a demand vector $\mathbf{d}$ (respectively, a pair $(CS, \mathbf{d})$) is stable if there is no set S and a demand vector for S which would provide each player in S with a higher expected payoff than $\mathbf{d}$ (respectively, $(CS, \mathbf{d})$).

4 Simple Games with Beliefs

In this section, we define simple coalitional games with beliefs and show that the core of such games can be characterized in essentially the same way as the core of (standard) simple games. Moreover, we demonstrate that this characterization can be used to construct a polynomial-time algorithm for checking whether the core is non-empty or whether a given demand vector is in the core.

For the rest of the section, we only consider games without coalition structures. Recall that in such settings an outcome of a game is simply a demand vector $\mathbf{d}$ that satisfies $d(N) = 1$.

We start by generalizing the notions of monotone and simple games to games with types and games with beliefs. Namely, we say that a game $G = (N; \vec{\mathcal{T}}; u)$ is *monotone* if $u(T_1, \ldots, T_n) \leq u(T'_1, \ldots, T'_n)$ for any two vectors $(T_1, \ldots, T_n)$, $(T'_1, \ldots, T'_n)$ such that for all $i \in N$ either $T_i = T'_i$ or $T_i = \perp$. That is, in a monotone game adding players to a coalition (while keeping the types of the other coalition members unchanged) cannot lower the value of a coalition. Furthermore, we say that G is *simple* if it is monotone and $u(T_1, \ldots, T_n) \in \{0, 1\}$ for all $T_i \in \mathcal{T}_i \cup \{\perp\}$ and all $i \in N$. Finally, we say that a coalitional game with beliefs $G = (G'; \vec{\mathbf{b}})$ is monotone (respectively, simple) if its associated game with types G' is monotone (respectively, simple).

It turns out that for simple games with beliefs, the definition of the core can be simplified considerably.

Proposition 1. *A demand vector* $\mathbf{d}$ *is in the core of a simple game with beliefs* G *if and only if any coalition* $S \subseteq N$ *such that* $u_i(S) = 1$ *for all* $i \in S$ *satisfies* $p(S) = 1$.

Proof. Suppose that $\mathbf{d}$ is in the core of G. Consider any coalition S such that $u_i(S) = 1$ for all $i \in S$. Clearly, $p(S) \leq p(N) = \sum_{i \in N} d_i u_i(N) \leq 1$; we will show that $p(S) \geq 1$. Suppose that $p(S) < 1$ and set $\varepsilon = \frac{1 - p(S)}{|S|}$. Construct a demand vector $\mathbf{d}'$ for $\{S, N \setminus S\}$ by setting $d'_i = p_i + \varepsilon$

for $i \in S$, $d'_i = \frac{1}{|N \setminus S|}$ for $i \in N \setminus S$; clearly, we have $d'(S) = d'(N \setminus S) = 1$. Moreover, we have $d'_i > p_i$ for all $i \in S$. As $u_i(S) = 1$ for all $i \in S$, this implies $u_i(S)d'_i > p_i$ for all $i \in S$, a contradiction with the definition of the core. Hence, $p(S) \geq 1$, as required.

Conversely, suppose that $\mathbf{d}$ is not in the core of G. Then, according to Definition 3 there exists a coalition S and a demand vector $\mathbf{d}'$ for $\{S, N \setminus S\}$ such that $p_i < u_i(S)d'_i$ for all $i \in S$. For this to be the case, it has to be true that $u_i(S) = 1$ for all $i \in S$. Therefore, we have $p_i < d'_i$ for all $i \in S$. As $d'(S) = 1$, this implies $p(S) < 1$. That is, S satisfies $p(S) < 1$, $u_i(S) = 1$ for all $i \in S$, as required. □

In simple games *without* beliefs, there is a very natural characterization of the core that relies on the notion of veto players. A player i is said to be a *veto* player in a game $G = (N; u)$ if any $S \subseteq N$ such that $u(S) = 1$ satisfies $i \in S$. By monotonicty, this is equivalent to requiring that $u(N \setminus \{i\}) = 0$. One can then describe the core as follows.

Theorem 1. *[Taylor and Zwicker, 1999] The core of a simple game $G = (N; u)$ is non-empty if and only if G has a veto player. Moreover, a payoff vector $\mathbf{p}$ is in the core of G if and only if $p_i = 0$ for any i that is not a veto player in G.*

We will now show that this result can be extended to simple games with beliefs. First, we have to generalize the notion of a veto player to our setting.

Definition 4. *A player i is said to be a* veto *player in a simple game with beliefs $G = (N; \vec{\mathcal{T}}; u; \vec{\mathbf{b}})$ if for any $S \subseteq N \setminus \{i\}$ we have $u_j(S) = 0$ for some $j \in S$.*

That is, i is a veto player if it impossible to form a coalition not involving i so that all of its members believe that this coalition will succeed. It turns out that one can efficiently verify whether a given player is a veto player.

Theorem 2. *Consider a simple game with beliefs $G = (N; \vec{\mathcal{T}}; u; \vec{\mathbf{b}})$ such that for all $S \subseteq N$ and all $i \in N$ one can compute $u_i(S)$ in time at most B. Then the algorithm $A(G, i)$ given in Fig. 1 correctly decides whether a player i is a veto player in G and runs in time $O(n^2B)$.*

Proof. Our algorithm starts with the coalition $S = N \setminus \{i\}$ and successively removes players who do not believe that the current coalition can succeed. By monotonicity, they would not believe that any subset of the current coalition can succeed either, so no such player can be a member of a coalition $S \subseteq N \setminus \{i\}$ that satisfies $u_j(S) = 1$ for all $j \in S$. After all such players have been deleted, we either obtain an empty coalition (in which case there is no coalition not involving i that believes in its own success, i.e., i is a veto player), or the remaining coalition S satisfies $u_j(S) = 1$ for all $j \in S$, i.e., i is not a veto player. The bound on the running time is immediate from the description of the algorithm. □

Note that, unlike in games without beliefs, it is not sufficient to check whether $N \setminus \{i\}$ is a winning coalition: it can happen that $N \setminus \{i\}$ contains a player who is sceptical about its success and therefore does not provide a witness that i is not a veto player, while a smaller coalition $S \subset N \setminus \{i\}$ does.

We now prove an analogue of Theorem 1 in our setting.

```
S = N \ {i}; flag = ⊥;
while flag = ⊥ and S ≠ ∅ do
   flag = ⊤; k_0 = 0;
   for each k ∈ S
       if flag = ⊤ and u_k(S) = 0
          k_0 = k; flag = ⊥;
       endif
   endfor
   if k_0 ≠ 0 then S = S \ {k_0};
endwhile
if S = ∅ output "i is a veto player";
else output "i is not a veto player";
```

Figure 1: $A(G, i)$ checks whether i is a veto player in G.

Theorem 3. *The core of a simple coalitional game with beliefs $G = (N; \vec{\mathcal{T}}; u; \vec{\mathbf{b}})$ is non-empty if and only if G has a veto player. Moreover, a demand vector $\mathbf{d}$ is in the core of G if and only if $d_i = 0$ for any i that is not a veto player in G.*

Proof. First, suppose that G has a veto player i. Consider a demand vector $\mathbf{d}$ that satisfies $d_i = 1$, $d_j = 0$ for $j \neq i$. Consider any coalition S such that $u_j(S) = 1$ for all $j \in S$. By definition of a veto player, we have $i \in S$. By monotonicity, we have $u_i(N) = 1$, so $p(S) \geq p_i = d_i u_i(N) = 1$. Hence, $\mathbf{d}$ is in the core.

Conversely, suppose that G has no veto players. Consider a demand vector $\mathbf{d}$. We have $d(N) = 1$, so $\mathbf{d}$ has at least one non-zero coordinate. Fix any i with $d_i > 0$. Since i is not a veto player, there is a coalition $S \subseteq N \setminus \{i\}$ such that $u_j(S) = 1$ for all $j \in S$. Moreover, we have $p(S) \leq p(N \setminus \{i\}) = 1 - p_i < 1$. Hence, $\mathbf{d}$ is not in the core. As this is true for any demand vector $\mathbf{d}$, G has an empty core.

The argument above also demonstrates that any demand vector $\mathbf{d}$ with $d_i > 0$ for a non-veto player i is not in the core, thus proving the second part of the theorem. □

Theorem 3 provides an efficient algorithm for checking whether the core is non-empty or whether a given demand vector is in the core. Indeed, we can first check, for each player $i \in N$, if i is a veto player (by Theorem 2, this can be done in polynomial time). The core is non-empty if the answer is "yes" for at least one $i \in N$. Furthermore, to check if $\mathbf{d}$ is in the core, we can check if $d_i > 0$ for any non-veto player i; if the answer is "yes", $\mathbf{d}$ is not in the core. We summarize these observations in the following theorem.

Theorem 4. *Given a simple coalitional game with beliefs G and a demand vector $\mathbf{d}$, one can check in polynomial time if G has an empty core, or if $\mathbf{d}$ is in the core of G.*

5 Computational Complexity: The Case of Weighted Voting Games with Beliefs

In this section, we consider the complexity of computing the CS-core in simple games with beliefs. It is known that even in full-information settings the CS-core is computationally more demanding than the core [Elkind *et al.*, 2008a]. We will now show that introducing beliefs adds another layer of complexity: for weighted voting games some natural CS-core-related problems that are polynomial-time solvable in the absense of beliefs become computationally intractable when beliefs are introduced.

We start by specializing the concept of coalitional games with beliefs to WVGs. For these games, it is natural to identify a player's type with his weight. Players may also have beliefs about the threshold. However, there is no loss of generality in assuming that the threshold is publicly known: a game $(N; \mathbf{w}; q)$ has the same set of winning coalitions as the game $(N; \alpha\mathbf{w}; \alpha q)$, and therefore we can rescale all beliefs so that the players have the same belief about the threshold.

Hence, we say that a *weighted voting game with beliefs* is given by a set of players $N = \{1, \ldots, n\}$, n belief vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$, where $\mathbf{b}_i = (b_i(1), \ldots, b_i(n))$ describes the beliefs of player i about the weights of the other players, and a threshold q; we write $\vec{\mathbf{b}} = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ and $G = (N; \vec{\mathbf{b}}; q)$. We will abuse notation by writing $b_i(S) = \sum_{j \in S} b_i(j)$. The utility function of a WVG with beliefs $G = (N; \vec{\mathbf{b}}; q)$ satisfies $u_i(S) = 1$ if $b_i(S) \geq q$, and $u_i(S) = 0$ otherwise. In the computational problems considered in the rest of this section, we assume that $b_i(j) \in \mathbb{Z}$ for all $i, j \in N$.

[Elkind *et al.*, 2008a] shows that many natural problems related to the CS-core are computationally hard when the weights are given in binary. However, for the important special case where all weights are polynomial in the number of agents, it gives a polynomial-time algorithm for checking if a given outcome is in the CS-core. Interestingly, it turns out that in our setting this problem is hard even if all weights are believed to be at most n (the number of players). This illustrates that introducing beliefs results in a quantifiable increase in the computational complexity of the problem.

Theorem 5. *Given a game $G = (N; \vec{\mathbf{b}}; q)$ and an outcome $(CS, \mathbf{d})$, it is coNP-complete to decide whether $(CS, \mathbf{d})$ is in the CS-core of G, even if $b_i(j) \leq n$ for all $i, j \in N$.*

Proof. Let $\mathbf{p}$ be the expected payoff vector for $(CS, \mathbf{d})$. Clearly, if an outcome is not in the CS-core of G, one can easily prove this by exhibiting a coalition S such that $p(S) < 1$ but $b_i(S) \geq q$ (and hence $u_i(S) = 1$) for all $i \in S$. It follows that our problem is in coNP. To show coNP-hardness, we will demonstrate that the complementary problem of checking whether $(CS, \mathbf{d})$ is not in the CS-core of G is NP-hard. To this end, we provide a reduction from a classic NP-hard problem CLIQUE [Garey and Johnson, 1990].

An instance of CLIQUE is given by an undirected graph $\mathcal{G} = (V, E)$, $|V| = t$, and an integer k. It is a "yes"-instance if $\mathcal{G}$ contains a clique (a complete subgraph) of size at least k and a "no"-instance otherwise. Given an instance of CLIQUE, we construct an instance of our problem as follows.

Suppose that $V = \{v_1, \ldots, v_t\}$. We create $n = 2t$ players $u_1, \ldots, u_t$, $v_1, \ldots, v_t$ and set $q = k$. The players' beliefs about each other's weights are defined as follows. For $i = 1, \ldots, t$, we set $b_{u_i}(v_i) = k$, $b_{u_i}(v_j) = 0$ for $j = 1, \ldots, t, j \neq i$, $b_{u_i}(u_j) = 0$ for $j = 1, \ldots, t$. Furthermore, for $i = 1, \ldots, t$ we set $b_{v_i}(u_i) = k - 1$, $b_{v_i}(u_j) = 0$ for $j = 1, \ldots, t, j \neq i$, and $b_{v_i}(v_j) = 1$ if $i = j$ or $(v_i, v_j) \in E$ and $b_{v_i}(v_j) = 0$ otherwise.

Let $CS = \{C^1, \ldots, C^t\}$, where for $i = 1, \ldots, t$ we set $C^i = \{u_i, v_i\}$. Observe that we have $b_{u_i}(C^i) = k$, $b_{v_i}(C^i) = k$, i.e., each player believes that he belongs to a winning coalition. Finally, for $i = 1, \ldots, t$ set $d_{u_i} = 1 - 1/(k+1)$, $d_{v_i} = 1/(k+1)$. Clearly, $\mathbf{d}$ is a demand vector for CS, and the corresponding expected payoff vector $\mathbf{p}$ satisfies $p_{u_i} = d_{u_i}$, $p_{v_i} = d_{v_i}$ for $i = 1, \ldots, t$.

One can show that $(CS, \mathbf{d})$ is in the CS-core of our game if and only if the graph $\mathcal{G}$ does not have a clique of size at least k. The proof is omitted due to space restrictions. □

Remark 1. *We can show that the problem remains coNP-complete even if $b_i(j) \in \{0, 1\}$ for all $i, j \in N$.*

Another computational problem considered in [Elkind *et al.*, 2008a] is that of checking whether the CS-core of a WVG is non-empty. Specifically, [Elkind *et al.*, 2008a] shows that this problem is hard if the weights are given in binary. However, it leaves open the question of whether this problem remains intractable if weights are polynomial in the number of players. We will now show that the corresponding problem for WVGs with beliefs is NP-hard even if all weights are at most linear in the number of players.

Theorem 6. *It is NP-hard to check if a game $G = (N; \vec{\mathbf{b}}; q)$ has a non-empty CS-core, even if $b_i(j) \leq n$ for all $i, j \in N$.*

Proof. The reduction is from PARTITION INTO TRIANGLES (PT) [Garey and Johnson, 1990]. An instance of PT is given by a graph $\mathcal{G} = (V, E)$, $|V| = 3t$. It is a "yes"-instance if $\mathcal{G}$ can be partitioned into triangles, i.e., there is a family of sets $V_1, \ldots, V_t \subset V$ such that (1) $|V_i| = 3$ for all $i = 1, \ldots, t$; (2) $\cup_{i=1}^{t} V_i = V$; (3) each V_i is a triangle, i.e., for any $i = 1, \ldots, t$ and any $u, w \in V_i$ we have $(u, w) \in E$. Otherwise, it is a "no"-instance.

Fix an instance $\mathcal{G} = (V, E)$ of PT with $V = \{v_1, \ldots, v_{3t}\}$. Let $N' = \{1, \ldots, 3t\}$, $N'' = \{3t+1, 3t+2, 3t+3\}$, and set $N = N' \cup N''$. Let $q = 6t+1$. For $i, j \in N'$, set $b_i(j) = 3t$ if $i = j$ or $(v_i, v_j) \in E$ and $b_i(j) = 0$ otherwise. Furthermore, set $b_i(j) = 3t$ and $b_j(i) = 1$ for all $i \in N'$, $j \in N''$. Finally, set $b_i(j) = 3t$ for all $i, j \in N''$. Let $G = (N; \vec{\mathbf{b}}; q)$.

Suppose that $\mathcal{G}$ is a "yes"-instance of PT, and let $\{V_1, \ldots, V_t\}$ be the corresponding partition. Consider the outcome $(CS, \mathbf{d})$, where $CS = \{C^1, \ldots, C^t, C^{t+1}\}$, $C^i = \{j \mid v_j \in V^i\}$ for $i = 1, \ldots, t$, $C^{t+1} = N''$, and $d_i = \frac{1}{3}$ for all $i \in N$. It is not hard to see that $(CS, \mathbf{d})$ is in the CS-core of G. Conversely, if $\mathcal{G}$ cannot be partitioned into triangles, G has an empty CS-core; the proof is omitted for brevity. □

6 Discussion

Recall that in transferable utility games (TU games), the worth of each coalition is given by a single number, which is the total payoff available to the members of this coalition; this payoff can be freely distributed between the members of that coalition. In contrast, in *non-transferable utility games (NTU games)*, the worth of a coalition C, $|C| = k$, is described by a vector of length k, which, for each player $i \in C$, describes the utility that i derives from participating in C. In such games, players cannot make payments to each other.

We constructed coalitional games with beliefs on the basis of TU games. Perhaps surprisingly, it turns out that they share certain features with both TU games and NTU games. Indeed, in the spirit of TU games, the players have to agree on how to distribute the profits achieved by a coalition; this agreement is captured by the demand vector. On the other hand, since each player in our games assigns his own utility to a coalition to be formed, he may not be able to transfer some of this utility to other players, especially if these players are much more pessimistic about the value of this coalition. This is of course in the spirit of NTU games. To give a simple example, consider a game in which $u_1(\{1\}) = 0$, $u_2(\{2\}) = 1$, $u_1(\{1,2\}) = 10$, $u_2(\{1,2\}) = 0$. Player 1 is eager to form a coalition with player 2, and would gladly offer a lion's share of its value (believed to be 10) to him. However, there is no demand vector that will make this proposal attractive for 2.

In fact, if the demand vectors for each coalition were fixed in advance (i.e., for each potential coalition S and each member of that coalition i player i's share of S's profit was part of the game description rather than subject to negotiation between the members of S), coalitional games with beliefs would be equivalent to (hedonic) NTU games: player i's value for a coalition S would be given by $d_i u_i(S)$. In contrast, in this paper we allow the players to negotiate the payoffs resulting from an expected outcome. This provides the players more flexibility and is therefore more natural in incomplete information scenarios.

We now describe two well-known classes of games that bear certain similarities to the weighted voting games with beliefs considered in the previous section. We hope that this discussion will help the reader appreciate the properties of our model, and its differences from existing related models.

Vector weighted voting games. In *k-vector weighted voting games* [Taylor and Zwicker, 1999; Elkind *et al.*, 2008b], the value of each coalition is determined using a set of k weight vectors $\mathbf{w}^1, \ldots, \mathbf{w}^k$ and k thresholds $q^1, \ldots, q^k$: a coalition S is considered to be winning if it wins in each of the component games, i.e., we have $w^i(S) \geq q^i$ for $i = 1, \ldots, k$. The similarity with WVGs with beliefs is obvious: in those games, too, the value of each coalition is determined using a set of weight vectors (one for each player). However, while in vector WVGs the number of weight vectors used to determine the value of each coalition is given exogenously (i.e., it is fixed to k independently of the coalition), in WVGs with beliefs the number (and the actual list) of weight vectors is determined endogenously: only the weight vectors (i.e., beliefs) of the coalition's members have to be taken into account.

Additively separable hedonic games. *Additively separable hedonic games* are described by a weighted directed graph whose vertices are players, and the weight w_{ij} of an edge (i, j) corresponds to the utility i extracts from being in a coalition with j; note that we may have $w_{ij} \neq w_{ji}$. Given a coalition S, the payoff of a player $i \in S$ is given by $p_i = \sum_{j \in S} w_{ij}$. While the weight w_{ij} appears to play a similar role to the belief $b_i(j)$ in the WVGs with beliefs, there are two important differences between these classes of games. First, as argued above, additively separable games are NTU games, while WVGs with beliefs are not. Second, there is no notion of a threshold in additively separable games.

A related class of TU games is defined by Deng and Papadimitriou [1994]: the games considered in this paper are also represented by a weighted (undirected) graph, and the value of a coalition S is determined as $u(S) = \sum_{i,j \in S} w_{ij}$. WVGs with beliefs are different from those games as well: the games of [Deng and Papadimitriou, 1994] allow for arbitrary transfers of utility between players, and do not involve any notion of a threshold.

7 Conclusions

In this paper, we developed a model for coalitional games with beliefs, which allows us to reason about coalition formation under uncertainty while being significantly less complex than the more general Bayesian models. We proved a characterization result for the core of simple games with beliefs, which matches the known result for the setting without beliefs, and can be used to derive polynomial-time algorithms for core-related problems. On the other hand, we showed that incomplete information can increase the complexity of the CS-core-related problems. In doing so, we introduced and analyzed weighted voting games with beliefs, which provide a convenient model for simple task allocation scenarios under uncertainty, and are therefore of independent interest.

We believe that our results provide a natural framework for modeling the behavior of selfish agents under incomplete information. Also, we hope they will prove useful for studying more sophisticated incomplete information scenarios, such as the full Bayesian setting.

Acknowledgements This research is supported by the ALADDIN (Autonomous Learning Agents for Decentralised Data and Information Networks) project, which is jointly funded by a BAE Systems and EPSRC strategic partnership (EP/C548051/1), as well as by EPSRC (GR/T10664/01) and ESRC (ES/F035845/1).

References

[Bachrach and Rosenschein, 2007] Y. Bachrach and J. S. Rosenschein. Computing the Banzhaf power index in network flow games. In *Proc. of AAMAS'07*, 2007.

[Chalkiadakis and Boutilier, 2004] G. Chalkiadakis and C. Boutilier. Bayesian reinforcement learning for coalition formation under uncertainty. In *AAMAS'04*, 2004.

[Chalkiadakis *et al.*, 2007] G. Chalkiadakis, E. Markakis, and C. Boutilier. Coalition formation under uncertainty: Bargaining equilibria and the Bayesian core stability concept. In *Proc. of AAMAS'07*, 2007.

[Deng and Papadimitriou, 1994] X. Deng and C. H. Papadimitriou. On the complexity of cooperative solution concepts. *Math. Oper. Res.*, 19(2):257–266, 1994.

[Elkind *et al.*, 2007] E. Elkind, L. A. Goldberg, P. W. Goldberg, and M. Wooldridge. Computational complexity of weighted threshold games. In *Proc. of AAAI'07*, 2007.

[Elkind *et al.*, 2008a] E. Elkind, G. Chalkiadakis, and N. R. Jennings. Coalition structures in weighted voting games. In *Proc. of ECAI'08*, 2008.

[Elkind *et al.*, 2008b] E. Elkind, L. A. Goldberg, P. W. Goldberg, and M. Wooldridge. On the dimensionality of voting games. In *Proc. of AAAI'08*, 2008.

[Garey and Johnson, 1990] M. Garey and D. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., N. York, 1990.

[Ieong and Shoham, 2008] S. Ieong and Y. Shoham. Bayesian coalitional games. In *Proc. of AAAI'08*, 2008.

[Myerson, 1984] R. Myerson. Cooperative games with incomplete information. *International Journal of Game Theory*, 1984.

[Myerson, 1991] R. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

[Myerson, 2007] R. Myerson. Virtual utility and the core for games of incomplete information. *Journal of Economic Theory*, 136:260–285, 2007.

[Shehory and Kraus, 1995] O. Shehory and S. Kraus. Task allocation via coalition formation among autonomous agents. In *Proc. of IJCAI'95*, 1995.

[Suijs and Borm, 1999] J. Suijs and P. Borm. Stochastic cooperative games: superadditivity, convexity and certainty equivalents. *Games and Economic Behavior*, 27:331–345, 1999.

[Taylor and Zwicker, 1999] A. Taylor and W. Zwicker. *Simple Games: Desirability Relations, Trading, Pseudoweightings*. Princeton University Press, Princeton, 1999.