Proceedings of the Twenty-First International FLAIRS Conference (2008)

Stability of Coalitions in Belief-Based Non-Transferable Utility Games

Chi-Kong Chan and Ho-fung Leung

Dept of Computer Science and Engineering,
The Chinese University of Hong Kong
{chanck, lhf}@cse.cuhk.edu.hk

Abstract

Coalition stability is an important concept in coalition formation. One common assumption in many stability criteria in non-transferable utility games is that the preference of each agent is publicly known, so that a coalition is said to be stable if there is no objections by any sub-group of agents according to the publicly known preferences. However, in many applications including some software agent applications, this assumption is not realistic. Instead, intelligent agents are modeled as individuals with private belief and decisions are made according to those beliefs instead of common knowledge. Such belief based agents architectures have impacts on the coalition's stability which are not reflected in the current stability criteria. In this paper, we extend the classic stability concept of the core by proposing a new belief based stability criterion which we labeled the belief-based core, and give examples to illustrate how the new concept can be used to provide both ex-ante and ex-post analysis of coalition formation mechanism.

1 Introduction

Intelligent agents in semi-competitive environments often need to form coalitions in order to achieve tasks that cannot be done alone, or to maximize their own utility via mutually benefiting agreements. Two research directions can be identified. First, we have the works in mechanism design, where many coalition formation mechanisms have been developed (Blankenburg and Klusch 2004, Sandholm 1999, Ketchpel 1994). In order to analyze the stability and efficiency of the output of such mechanisms, or to provide ex-ante prediction for the possible outcomes, we need the models from the other research direction which is cooperative game theory (Scarf 1967, Osborne and Rubinstein 1994), where various coalition stability solution concepts have been proposed. So far, most of these stability concepts have a *common knowledge* assumption, meaning that various characteristics of the game, including each individual agent's preference, are known to all agents. However, this assumption is not realistic in belief-based agents systems, where the agents' decisions are based on private beliefs instead of common knowledge. For this reason, we focus on the second research direction in this paper and discuss a belief based stability concept.

One general class of coalition formation games are the non-transferable utility (NTU) games, where agents come together to form non-overlapping coalitions, and each coalition is associated with a set of feasible consequences, which are the outcomes of the coalition as a result of the agents' joint action. In game theory, one classic solution concept for NTU games (and TU games also) is the core (Gillies 1959), which requires the specification of all agents' preferences regarding each possible consequence obtainable by the coalition, and these preferences are supposed to be publicly known for certain, meaning that each agent knows not just its preferences, but also (accurately) the preferences of each other as well.

In many applications, we are not able to provide such publicly known preferences, and instead, agents often have to rely on their own internal belief during the coalition formation process. For instance, consider a typical distributed propose-and-evaluate type mechanism (*e.g.*, Kraus, Shehory and Taase 2003), where coalitions are formed in steps, and in each steps, agents are allowed to send proposals to others for forming new coalitions. Time constraints and other limitations (*e.g.*, problem size) often mean that, in practice, the agents can only make those proposals that have a reasonable chance of being accepted, according to the beliefs of the proposing agent. Thus, during such a process, if a point is reached, such that each agent believes there is no better alternative solution than the current arrangement, for both himself and his partners, then the current solution should be regarded as stable, no matter whether those beliefs are accurate or not.

To model this situation, we are proposing a new stability criterion, the belief-based core that also takes into accounts the beliefs of the agents. We believe the proposed concepts can provide useful solution concepts for this emerging type of coalition games, which we call non-transferable utility games with private beliefs.

2 Motivating Examples

In this section we illustrate the ideas by studying examples of games that are stable in practice, despite they are not in the core. The existence of these examples suggests that such belief-based games are not well handled by the core-based approaches.

Example 1: a belief-based dating game

We first look at an example dating game involving three agents, *a*, *b*, and *c*, who are considering to go to a movie. Their preferences are that each of them prefers to go in pairs (coalition of size 2) if possible. Failing that, their next choice is to go in a group of all three, and their last choice is to go alone. Furthermore, among the coalitions of size 2, agent *a* prefers to go with agent *b,* but agent *b* prefers to go with agent *c*, and finally, agent *c* prefers to go with agent *a*. In summary, their preferences are given by:

Agent *a*'s preference:

$$(\{a,b\},\text{movie}) \succ_a (\{a,c\},\text{movie}) \succ_a (\{a,b,c\},\text{movie}) \succ_a (\{a\},\text{movie})$$

Agent *b*'s preference:

$$(\{b,c\},\text{movie}) \succ_b (\{a,b\},\text{movie}) \succ_b (\{a,b,c\},\text{movie}) \succ_b (\{b\},\text{movie})$$

Agent *c*'s preference:

$$(\{a,c\},\text{movie}) \succ_c (\{b,c\},\text{movie}) \succ_c (\{a,b,c\},\text{movie}) \succ_c (\{c\},\text{movie})$$

Of course, such preferences are private information, not common knowledge. However, since the agents know each other and have interacted before, each of them also has a belief of the other two's preference:

Agent *b*'s belief of *a*'s preference:

$$(\{a,b,c\},\text{movie}) \succ_a (\{a,b\},\text{movie}) \succ_a (\{a,c\},\text{movie}) \succ_a (\{a\},\text{movie}))$$

Agent *c*'s belief of *a*'s preference:

$$(\{a,b,c\},\text{movie}) \succ_a (\{a,b\},\text{movie}) \succ_a (\{a,c\},\text{movie}) \succ_a (\{a\},\text{movie})$$

Agent *a*'s belief of *b*'s preference:

$$(\{a,b,c\},\text{movie}) \succ_b (\{b,c\},\text{movie}) \succ_b (\{a,b\},\text{movie}) \succ_b (\{b\},\text{movie})$$

Agent *c*'s belief of *b*'s preference:

$$(\{a,b,c\},\text{movie}) \succ_b (\{b,c\},\text{movie}) \succ_b (\{a,b\},\text{movie}) \succ_b (\{b\},\text{movie})$$

Agent *a*'s belief of *c*'s preference:

$$(\{a,b,c\},\text{movie}) \succ_c (\{a,c\},\text{movie}) \succ_c (\{b,c\},\text{movie}) \succ_c (\{c\},\text{movie})$$

Agent *b*'s belief of *c*'s preference:

$$(\{a,b,c\},\text{movie}) \succ_c (\{a,c\},\text{movie}) \succ_c (\{b,c\},\text{movie}) \succ_c (\{c\},\text{movie})$$

In short, each of them wrongly believes that the others prefer a coalition of size three to a coalition of size two. The game is denoted by Figure 1, where each node represents a possible coalition structure. For example, the top left node $\{(\{a\},\text{movie}), (\{b\},\text{movie}), (\{c\},\text{movie})\}$ represents the outcome where each agent forms their own singleton coalition and go to the movie on his own, which is also the default outcome without any negotiation, and the bottom right node $\{(\{a,c\},\text{movie}), (\{b\},\text{movie})\}$ represents agents *a* and *c* forming a coalition of size two and going to the movies together. Starting form the default outcome, the agents are allowed to make stepwise improvements by proposing alternative coalition structures. A proposal is considered successful if it is accepted by all the members of at least one coalition in the alternative outcome. In game theoretic terms, each successful proposal is called an *objection* to the original outcome. Objections are shown by both solid and dotted edges in Figure 1. The left most edge, for example, says that the outcome $\{(\{a\},\text{movie}), (\{b\},\text{movie}), (\{c\},\text{movie})\}$, is objected by the outcome $\{(\{a,b\},\text{movie}), (\{c\},\text{movie})\}$, which is because there exists a coalition ($\{a,b\}$ in this case) in the latter outcome such that each of its member (agent *a* and agent *b*) prefers the latter to the former. As seen in Figure 1, each outcome is objected by at least one objection, so the core is empty in this case.

However, if the agents make the proposals according to their beliefs, that is, if each agent only proposes alternatives such that (i) he is better off in the alternative, and (ii) he thinks the proposal can be accepted by his new partners in the alternative outcome (perhaps in order to avoid the embarrassment of being rejected and to speed up the coalition formation process). Then there is actually a stable outcome for this game. Consider the dotted edge that leads from $\{(\{a, b, c\}, \text{movie})\}$ to $\{(\{a,b\},\text{movie}), (\{c\},\text{movie})\}$. This edge is no longer an objection in the belief-based game because each agent (wrongly) believes that the others would prefer a coalition of size 3 to a coalition of size 2. The same is true for the other two objections which are represented by dotted lines. So in this case, we can expect that the outcome $\{(\{a, b, c\},\text{movie})\}$, once reached, would in fact be stable: although there are in fact better outcomes according to the concept of the core, no agents realize this and they are happy to stay in the original node, making it stable in practice.

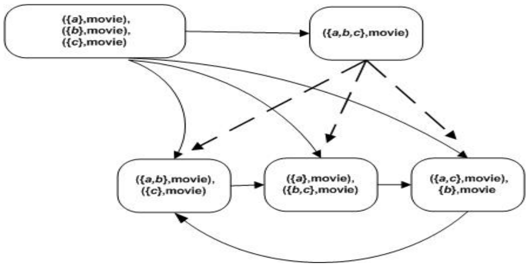

Figure 1. A belief based dating game

Example 2: Randomized mechanisms

In most distributed coalition formation mechanisms that employ propose-and-evaluate type protocols, one decision problem faced by the agents is what to propose to other agents. This can be problematic in situations where agents process only private beliefs instead of common knowledge, because the agent has no way of knowing whether his proposal is acceptable to other agents or not. Naturally, one strategy for the agents is to only make proposals that are consistent with their belief regarding other agent's preferences. Such a strategy will have an impact on the coalition's stability and speed of convergent to a stable solution. Suppose that, in order to investigate the effect on stability caused by such belief-based strategy, we decide to implement and test the following mechanism as illustrated in figure 2, which is a typical randomized approach for

```
A Typical Randomized Coalition Formation Mechanism
Repeat until the pre-designated number of rounds is reached, or no new proposal
is accepted in a pre-designated consecutive number of rounds:
1 Randomly select an agent as the next proposing agent. Let C be the coalition
where he is  currently a member of.
2 The proposing agent finds and proposes an alternative coalition.
3 If such an alternative coalition is found, the proposing agent then send proposal
   messages to each    members of C'.
4 The recipient of a proposal message responses as follows: accepts it if it prefers
   the proposed coalition C' to the coalition that it is currently a member of,
   rejects it otherwise.
5 If the proposal is accepted by all recipients, form the new coalition structure as
   follows:
  5.1 Each member of C' leaves their previous coalition and become member of
        C'.
  5.2 All other agents (that is not in C') remain in the same coalition as before.
 6 Go to step 1.
Strategy of type-b (belief-based) agents
In step 2 of the mechanism, the agent finds and proposes an alternative coalition
such that:
     i) The proposing agent is a member of a coalition C'
     ii) The proposing agent prefers the coalition C' to C.

Strategy of  type-n (non-belief-based) agents
In step 2 of the mechanism, the agent finds and proposes an alternative coalition
such that:
     i) The proposing agent is a member of a coalition C'
     ii) The proposing agent prefers the coalition C' to C.
     iii) The proposing agent believes that all member of C' would prefer the
           coalition C'   to C.
```

Figure. 2. A Typical Randomized coalition formation mechanism

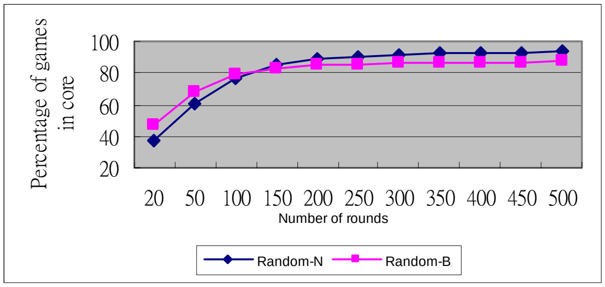

Figure 3a. Comparison of games populated with type-b and type-a agents, using stability concept of the core.

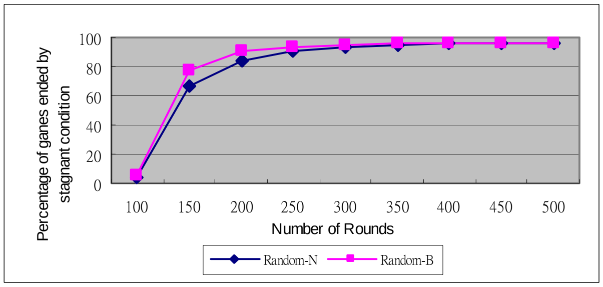

Figure 3b. Comparison of games populated with type-b and type-a agents, using the stagnant criterion.

coalition formation. The mechanism, beginning with some initial outcome, are divided into rounds, and in each round, one of the agents is randomly selected. The randomly selected agent is then allowed to make proposal for changes to the profile such that i) a new coalitional is formed in the alternative outcome, where the proposing agent is a member of the corresponding coalition ii) the proposing agent is better off in the new coalition than previously. The new outcome becomes effective if the proposal is accepted by each member of the new coalition. The mechanism terminates after a pre-defined number of rounds is reached (termination by time) or if no new proposals are accepted for a pre-defined consecutive number of rounds (termination by stagnant criterion). For simplicity, we assume there is only one possible action per coalition in this example.

In order to investigate the effect of agent beliefs, we assume there are two types of agents, namely type-b and type-n agents, with different strategies. Type-n agents, when selected to make proposals, do not consider his beliefs about the other member's preference (*i.e.,* he does not care whether his proposal is likely to be accepted or not, he just proposes it as long as he thinks it is beneficial to himself), while the second type of agents (type-b) carefully propose only the ones that are consistent with their beliefs (*i.e.*, the agent only proposes those proposals which he thinks will be accepted).

We perform an experiment involving six agents. The agents' preferences regarding the candidate coalitions are randomly ordered. Each agent also has a belief of each other's preference, but subjected to a 25% error rate: for any two agents *i* and *j*, and for any two coalitions that involves both agents, there is a 25% chance that agent *i* is wrong about agent *j*'s preference order regarding those two coalitions. Each game is repeated twice: first with all agents employing type-n strategy (labeled as Random-N games), then with all agents employing type-b strategy (labeled as Random-B games).

The results are shown in figure 3a and 3b. In Figure 3a, we see the percentage of core-stable outcomes, out of 1000 repetition after various numbers of rounds. The result seems to suggest that, according to the concept of the core, games that are populated with belief-based type-b agents (Random-b games) obtain less stable results than games populated with non-belief-based type-n agents (Random-n games) in the long run, despite some gain in the early rounds. However, to the contrary, closer examination of the experiment data suggests that, in the long run, almost all Random-B games terminates because no agent is able to make any more proposals, which is a suggestion that the solutions *are* in fact stable. To verify this, we also measure the number of games that terminate by the stagnant criterion and the result is shown in Figure 3b, which confirms that the outcomes of Random-B games should be at least as stable as Random-N games in the long run, while outperforming Random-N in the short run. Thus, the analysis as suggested by the core is in contrast to the real stability of the outcomes and we see that concepts like the core are insufficient in describing the stability of games such as this one. ■

What these two examples suggest is that the traditional stability criteria, which assume all preferences to be common knowledge, are inadequate in scenarios where private beliefs are important factors in determining the behavior of the agents. In example 1, the core based concepts fail to predict a stable outcome. In example 2, the analytical result does not reflect the real stability of the solutions achieved by the mechanisms. The reason is that we are facing a new type of games where the stability is based on private information instead of common knowledge. For these reasons, we are proposing a solution concept that is suitable for this new type of games.

3 NTU Games with Private Beliefs

The games depicted in examples 1 and 2 are examples of what we shall call non-transferable utility games with private beliefs (NTUPB games), which can be represented by a tuple $g = (N,\ A,\ (\succ_i)\ , B)$ as follows. Let $N = \{1,\ldots,n\}$ be a set of agents and let any subset $C \subseteq N$ be called a coalition. The goal of the game is to partition the set of agents into a coalition structure of exhaustive and non-overlapping coalitions. There is a set A of possible actions that are available to the agents so that each member of the same coalition can jointly choose one of the actions in the set. We assume that the outcome of a coalition is decided only by the coalition itself (i.e., who its members are) and the joint action chosen. For this reason, we define a *coalitional act* by a couple $\alpha = (C, a)$, where $C \subseteq N$ and $a \in A$, which represents the possible outcomes achievable by the members of the coalition C performing joint-action a. The preference of each agent i is represented by a total ordered preference relation $\succ_i$ on the set of possible coalitional acts, so that for any two coalitional acts $\alpha_1 = (C_1, a_1)$ and $\alpha_2 = (C_2, a_2)$, $i \in C_1 \cap C_2$, we have $\alpha_1 \succ_i \alpha_2$ if agent i prefers α_1 to α_2 . We define a *coalitional action profile* (or *profile* for short) to be a set S of coalitional acts that corresponds to a coalition structure, *i.e.,* $S = \{\alpha_1, \ldots, \alpha_k\}$ where k is the number of coalitions in the coalition structure, and each $\alpha_i = (C_i, a_i)$ represents the coalitional act of the i^{th} coalition. We use $C_i(S)$ to denote the coalition in S which the agent j is a member of, and $\alpha_j(S)$ to denote its corresponding coalitional act. That is, $\alpha_j(S) = (C_i, a_i) \in S$ such that $j \in C_i$.

The core of a NTUPB game is defined as in the traditional NTU game:

Definition 1 (Core of NTUPB game). The core of an NTU game is defined as the set of coalitional action profile $S = \{(C_1, a_1), \ldots, (C_k, a_k)\}$ such that no subset of agents $C' \subseteq N$ can deviate from their corresponding coalitions by finding an alternative coalitional act where each member of the coalition C' would prefer to their current coalitional act. That is, a profile S is in the core if there does not exist a coalition $C' \subseteq N$ and an alternative coalitional act $\alpha = (C', a)$ such that $\alpha \succ_i \alpha_i(S), \forall\ i \in C'$. ▪

Given an NTUPB game g, we use the notation core(g) to represent the set of coalitional action profiles that is in the core of the game.

However, as discussed in the examples above, there are situations in NTUPB games that are not well handled by the core. The reason is that, unlike in traditional NTU games, the agents' preferences in NTUPB games are private information, and this is not reflected by the traditional concepts. To handle this, we also assume each agent i maintains beliefs regarding other agents' preferences which is represented by a relation bel_i, so that for two agents i and j, we write $bel_i(\alpha_1 \succ_j \alpha_2)$ if agent i believes that agent j prefers coalitional act α_1 to α_2 .

Definition 2 (Agent's Beliefs)). Given two agents i and j, and two coalitional act α_1 and α_2 , we write $bel_i(\alpha_1 \succ_j \alpha_2)$ if agent i believes agent j prefers α_1 and α_2 . ■

The set of all beliefs of all agents in a NTUPB game is represented by a belief profile $B = \{bel_1, bel_{2,\ldots,} bel_n\}$ where bel_i is the private beliefs of the i^{th} agent.

Before we define our main stability criterion, we first define two more concepts:

Definition 3 (Domination Relation, dom*).* Given any two coalitional action profiles S_1 and S_2, we say S_1 is dominated by S_2 through a coalitional act $\alpha = (C, a) \in S_2$, written S_2 $\mathtt{dom}_\alpha$ S_1, if , for each agent $i \in C$, we have $\alpha \succ_i \alpha_i(S_1)$.

Definition 4 (Belief-based Domination Relation, B-dom*).* Given any two coalitional action profiles S_1 and S_2, we say S_1 is dominated by S_2 through a coalitional act $\alpha = (C, a) \in S_2$ based on beliefs, written S_2 $\mathtt{B\text{-}dom}_\alpha$ S_1, if there exists an agent $j \in C$ such that, for each agent $k \in C, k \neq j$, we have $bel_j(\alpha \succ_k \alpha_k(S_1))$ ■

Intuitively, definition 3 simply says that a profile S_1 is dominated by another profile S_2 if we can find a coalition and a corresponding joint action in S_2 such that every member of that coalition would prefer S_2 to S_1. Definition 4 says that profile S_1 is dominated based on belief by profile S_2 if at least one member of that coalition believes that every member of that coalition would prefer S_2 to S_1.

Now, we are ready to define a new belief based stability criterion, the B-core:

Definition 5 (B-core of NTUPB game). A coalitional action profile $S = \{(C_1, a_1), \ldots, (C_k, a_k)\}$ is in the B-core of an NTUPB game if there does not exist any alternative coalitional action profile S_2, so that we have both S_2 $\mathtt{dom}_\alpha$ S and S_2 $\mathtt{B\text{-}dom}_\alpha$ S, for any $\alpha \in S_2$. ■

Intuitively, we say a profile is in the B-core of a game if there does not exist any alternative profiles that satisfies the following two conditions: 1) every member of at least one coalition in the alternative prefers the alternative to the original profile and 2) at least one agent in that coalition correctly believes that point 1 is the case.

Given an NTUPB game g, we use the notation B-core(g) to represents the set of coalitional action profile that is in the B-core of the game.

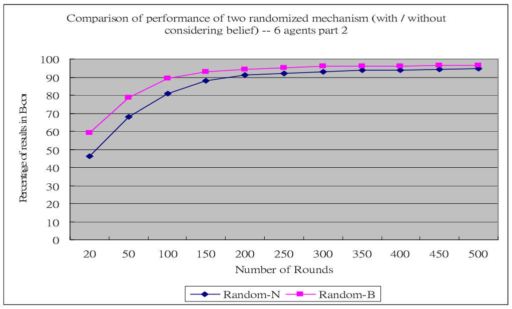

Figure 4. Stability of the mechanism in example 2, according to the B-core.

Example 3. Consider again example 1, the B-core consists of one coalitional action profiles, namely {({*a*, *b*, *c*}, movie)} ■

Example 4. Consider again example 2. This time, we measure the percentage of stable outcomes according to the concept of B-core, out of 1000 repetitions, achieved after various numbers of rounds. The result, as shown in Figure 4, suggests that belief-based games (Random-B) achieves larger number of stable results than the non-belief-based games (Random-N) throughout the execution of the mechanism, which is consistent with our previous observation in Figure 3b. In fact, by comparing Figure 3a and Figure 4, we now know that Random-B games tends to converge to a result in the B-core, whereas Random-N games tends to converge to the traditional core. ■

Theorem 1. The core of a NTUPB game is a subset of the B-core.

Proof In definition 5, part of the requirement for a coalitional action profile S to be in the B-core of an NTUPB game is that there does not exist any alternative coalitional action profile S_2, so that S_2 $\mathtt{dom}_\alpha$ S, which implies that S is also in the core.

4 The Core, B-core And Belief Accuracy

The reason that, in general, the B-core of an NTUPB game differs from the core is that the private beliefs of the agents are often inaccurate. To understand the effects of such inaccuracy on stability, and the relation between the core and B-core in general, we have the following definitions and theorems.

Definition 6 (Accuracy relation of agents private beliefs) Given two private belief relations bel_1 and bel_2, we say bel_1 is *more accurate* than bel_2, if for all agents $j \in N$, and any coalitional acts α_1 and α_2, we have the followings:

i) If $bel_1(\alpha_1 \succ_j \alpha_2)$ holds but $bel_2(\alpha_1 \succ_j \alpha_2)$ does not hold, then $\alpha_1 \succ_j \alpha_2$ holds.

ii) If $bel_1(\alpha_1 \succ_j \alpha_2)$ does not hold but $bel_2(\alpha_1 \succ_j \alpha_2)$ holds, then $\alpha_1 \succ_j \alpha_2$ does not hold.

Definition 7 (Accuracy relation of beliefs profiles) Given two external beliefs profiles $B = \{bel_1, bel_2, ..., bel_n\}$ and $B' = \{bel_1', bel_2', ..., bel_n'\}$, we say B is *more accurate* than B' if there exists $i \in N$ such that bel_i is more accurate then bel_i', and either $bel_j \equiv bel_j'$ or bel_j *is more accurate than* bel_j' for all $j \in N - \{i\}$.

Theorem 2. Given two NTUPB games $g = (N, A, (\succ_i), B)$ and $g' = (N, A, (\succ_i), B')$,we have B-core(g) $\subseteq$ B-core(g') if B is *more accurate than* B'.

Proof (Sketch) If B is more accurate than B', then there exists coalitional action profiles S_1 and S_2 such that S_1 $\mathtt{B\text{-}dom}_\alpha$ S_2 in g but not in g', which follows that the b-core of g is a subset of the b-core of g'. ■

Theorem 2 suggests that inaccuracies in the agents' belief actually lead to *more* stable results. Intuitively, we can understand this as follows. Recall that according to the definition of B-core, an objection (*i.e.*, an alternative coalitional act) to a coalitional action profile in an NTUPB game need to satisfy two conditions. First, all members of the deviating coalition must actually prefer the alternative coalitional act to the one that they are currently in. Second, at least one agent of the coalition correctly believes that all fellow members will also prefer the alternative. Thus, any inaccuracies in the beliefs of the agents may cause *less* objections to be raised, because some of the agents may wrongly think that an otherwise valid objection to be invalid (which cause the objection to fail the second requirement), which in turns means that there are *more* stable results as there as less objections. (Another way to think about this is: the inaccurate beliefs can lead to mutual-misunderstanding amongst the agents, making some otherwise sub-optimal solution stable because some better solution cannot be reached because of the misunderstandings).

Theorem 3. The core of a NTUPB is the same as the B-core if all agents' beliefs are accurate. That is, given an NTUPB games $g = (N, A, (\succ_i), B)$, we have B-core(g) $\equiv$ core(g) if $bel_i(\alpha_1 \succ_j \alpha_2) \Rightarrow \alpha_1 \succ_j \alpha_2$ for all agents i, j, and any coalitional acts α_1 and α_2. ■

Proof (Sketch). If all agents' beliefs are accurate, then for any two coalitional action profiles S_1 and S_2 we have S_1 $\mathtt{B\text{-}dom}_\alpha$ S_2 ⇔ S_1 $\mathtt{dom}_\alpha$ S_2, which implies the B-core is the same as the core. ■

5 Discussion

There are two possible interpretations to most game theoretical stability concepts. In an *ex-ante* interpretation, we try to predict beforehand, as an omniscient observer, which outcomes are stable and achievable via some protocol by the participants of a game. An example of this type of analysis is the dating game in examples 1 and 3. This interpretation not only requires the observer to know the preference of each agent, but each agent's private beliefs also. Thus we see that the core is not really suitable to this type of interpretation as long as beliefs are involved in the decision making process, and that the B-core is a better tool for these tasks.

The other type of interpretations is the *ex-posts* analysis, which attempts to answer the question, "Given this particular coalition structure and coalition action profile, what can we say about its stability?" Thus, while the ex-ante analysis tries to predict which stable outcomes can be reached, the ex-post analysis tries to decide whether any arbitrary outcome is stable. An example of ex-post analysis is the result analysis of the randomized mechanisms in examples 2 and 4. We see that in ex-post analysis, both the core and the B-core can be used, but they correspond to two different stability concepts. The core is really asking the question "Assuming that the agents are now told about the real preferences of the other agents, will any of them now change their mind and deviate and form new coalitions?" Whereas the B-core is still asking the original question "Will any agents change their mind and deviate, given what they always believe". With both tools in hand, we can now provide additional analysis on coalition formation mechanisms. For example, a mechanism that produces a low percentage of results in the B-core (and hence also the core) suggests that the results are not stable and not satisfactory, and the mechanism should be improved. A high percentage of results in B-core and a low percentage of result in core suggest that the results are stable, but there are errors or mutual misunderstanding in the beliefs of the agents which prohibit any better results from occurring, and one way to tackle this is to allow the agents to communicate regarding their beliefs. Finally a high percentage of results in the core (and hence also B-core) suggest that the results are stable and there is no mutual misunderstanding between the agents.

6 Related Works

We discuss some related works in this section. A Bayesian-core concept is proposed in Chalkiadakis and Boutilier 2007 where the agents are assumed to belong to various types which are unknown to other agents. The agents are required to estimate the value of potential coalitions by maintaining a Bayesian belief system regarding the possible types of their potential partners. Our work differs from theirs in that our model assumes the more general problem of non-transferable utilities games, instead of transferable ones.

A solution concept for coalition game with stochastic payoff is presented in (Suijs *et al.* 1999). In this approach, the coalitional payoffs are assumed to be stochastic variables, and agents preferences over those stochastic variables are used to determine the stability. Thus, their work is on stochastic games, whereas our focus is on a more general class of non-transferable utility games that are not necessarily probabilistic in nature.

7 Conclusion

Most classical solution concepts in non-transferable utility coalitional game theory rely on a public information assumption. That is, they assume the agents' preferences to be publicly known. However this assumption is not practical in many software agent applications where intelligent agents have to rely on their private beliefs during decision making. In this paper, we propose a new type of game which we label non-transferable utility games with private belief, and provide a new concept for describing the stability of coalitions these games, namely, the B-core. By doing so, we are able to provide useful stability concepts for this new type of game which otherwise cannot be analyzed properly using the classic public information based approaches. We believe our model provide a useful tool in evaluating coalition formation algorithms for agent based cooperative games, for the purpose of both ex-ante and ex-post analysis.

References

Blankenburg, B. and Klusch, M., *On safe kernel stable coalition forming among agents*. Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, volume 2, pg 580–587. (2004).

Scarf, H., *The core of n person game.* In: Econometrica, 35(1) pg 50-69. (1967)

Osborne, M. J. and Rubinstein, A, *A Course in Game Theory*. The MIT Press. (1994).

Sandholm, T, *Distributed rational decision making.* In: Weiss, G., editor, Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. The MIT Press. (1999).

Ketchpel, S, *Forming coalitions in the face of uncertain rewards.* In: Proceedings of National Conference on Artificial Intelligence (AAAI-94*)*, pg 414–419. (1994).

Gillies, D. B., *Solutions to general non-zero-sum games.* In: Tucker, A.W. and Luce, R. D., editors, Contributions to the Theory of Games Volume IV. Princeton University Press. (1959).

Kraus, S., Shehory, O., and Taase, G., *Coalition formation with uncertain heterogeneous information.* Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, pgs 1–8. (2003).

Chalkiadakis, G. and Boutilier, C. *Coalition Formation under Uncertainty: Bargaining Equilibria and the Bayesian Core Stability Concept.* In: Proceedings of the 2007 International Joint Conference on Autonomous Agents and Multiagent Systems, pg 1090–1097. (2007).

Suijs, J., Borm, P., De Waegenaere, A. and Tijs, S., *Cooperative Games with Stochastic Payoffs.* In: European Journal of Operational Reseach 133. (1999).

Dieckmann**,** T and Schwalbe, U., *Dynamic Coalition Formation and the Core.* Econometric Society World Congress 2000 Contributed Papers (2000).