Anonymity-proof Shapley Value: Extending Shapley Value for Coalitional Games in Open Environments

Naoki Ohta*, Vincent Conitzer†, Yasufumi Satoh*, Atsushi Iwasaki*, Makoto Yokoo*
*Kyushu University, Fukuoka 819-0395, Japan, {ohta, satoh}@agent.is.kyushu-u.ac.jp,
{iwasaki, yokoo}@is.kyushu-u.ac.jp
†Duke University, Durham, NC 27708, USA, conitzer@cs.duke.edu

ABSTRACT

Coalition formation is an important capability for automated negotiation among self-interested agents. In order for coalitions to be stable, a key question that must be answered is how the gains from cooperation are to be distributed. Coalitional game theory provides a number of solution concepts for this. However, recent research has revealed that these traditional solution concepts are vulnerable to various manipulations in open anonymous environments such as the Internet. To address this, previous work has developed a solution concept called the anonymity-proof core, which is robust against such manipulations. That work also developed a method for compactly representing the anonymity-proof core. However, the required computational and representational costs are still huge.

In this paper, we develop a new solution concept which we call the anonymity-proof Shapley value. We show that the anonymity-proof Shapley value is characterized by certain simple axiomatic conditions, always exists, and is uniquely determined. The computational and representational costs of the anonymity-proof Shapley value are drastically smaller than those of existing anonymity-proof solution concepts.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence - Multiagent systems; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Theory, Economics

Keywords

Multi-agent systems, Coalitional game theory, Shapley value

1. INTRODUCTION

Coalition formation is an important capability for automated negotiation among self-interested agents. In order for coalitions to be stable, a key question that must be answered is how the gains from cooperation are to be distributed. Coalitional game theory provides a number of solution concepts for this, such as the Shapley value, the core, the least core, and the nucleolus. Some of these solution concepts have already been adopted in the multiagent systems literature [1, 2, 3, 4, 8, 10]. Besides being of interest to the game-theory and multiagent systems research communities, the growth of Internet and electronic commerce has expanded the application areas of coalitional game theory. For example, consider a large number of companies, some subsets of which could form profitable virtual organizations that can respond to larger or more diverse orders than an individual company can. The Internet makes forming such virtual organizations much easier, but the companies must agree on how to divide the profits among themselves.

However, the authors have pointed out that existing game-theoretic solution concepts have limitations when applied to open anonymous environments such as the Internet [9]. In such environments, an agent can use multiple identifiers (or *false names*) to pretend to be multiple agents. Alternatively, multiple agents can collude and pretend to be a single agent. Furthermore, an agent might try to hide some of its capabilities (or *skills*). These manipulations are virtually impossible to detect in open anonymous environments, and thus present a serious vulnerability in such environments.

In our prior research, we developed a new solution concept called the *anonymity-proof core*, which is robust against these manipulations [9]. However, the anonymity-proof core and other similar solution concepts have one serious limitation, i.e., the representation size of the outcome function requires space exponential in the number of agents and skills. In our more recent research [6], we have developed a method for compactly representing the anonymity-proof core (and other anonymity-proof solution concepts), given that the characteristic function is represented using a compact language that explicitly specifies only coalitions that introduce synergy [2]. However, the required representation size is still exponential in the number of skills/agents in the worst case. Also, the required computational cost of the existing anonymity-proof solution concepts remains huge.

In this paper, we propose a new solution concept, which we call the *anonymity-proof Shapley value*. The characteristics of the anonymity-proof Shapley value are as follows.

- It is based on the Shapley value, one of the most important solution concepts for coalitional games [7].
- It is characterized by simple axiomatic conditions, i.e., Pareto efficiency, the weak-null property, weak-symme-

Cite as: Anonymity-proof Shapley Value: Extending Shapley Value for Coalitional Games in Open Environments, Naoki Ohta, Vincent Conitzer, Yasufumi Satoh, Atsushi Iwasaki and Makoto Yokoo, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 927-934.

try, weak-additivity, and best-approximate-monotonicity (which is a stricter condition than anonymity-proofness). It is the only solution concept that satisfies these axiomatic conditions simultaneously.

- It always exists and is uniquely determined.
- We do not need to compute/store the entire outcome function; we need to compute/store only the value division for the grand coalition, and we can calculate the rest on demand.

The rest of this paper is organized as follows. First, we review the model of coalitional games in an open, anonymous environment (Section 2). Next, we summarize the main results of this paper (Section 3). Then, we give a series of theorems to explain why and how the axiomatic conditions of the (traditional) Shapley value should be relaxed (Section 4). Furthermore, we introduce a new axiomatic condition called best-approximate-monotonicity (Section 5). Then, we give our definition of the anonymity-proof Shapley value and show how it is axiomatically characterized (Section 6). Next, we discuss the computational/representational costs and possible variations of the anonymity-proof Shapley value (Section 7). Finally, we show experimental results that illustrate the discrepancy between the anonymity-proof Shapley value and the traditional Shapley value (Section 8).

2. MODEL

Traditionally, value division in coalition formation is studied in *characteristic function games*, where each potential coalition (that is, each subset X of the agents) has a value $w(X)$ that it can obtain. However, in an open anonymous environment, the characteristic function by itself does not give sufficient information to assess what manipulations may be performed by agents. Previous research has introduced a more fine-grained representation of what each agent brings to the table, i.e., instead of defining the characteristic function over agents, the characteristic function is defined over the *skills* that the agents possess [9].[1]

Definition 1 (characteristic function). *A characteristic function $v : 2^T \to \Re$ assigns a value to each set of skills $S \subseteq T$, where T is the set of all possible skills.*

We denote by w the characteristic function defined over agents and by v the characteristic function defined over skills. For a given set of agents X, let $S_X = \bigcup_{i \in X} S_i$, where S_i is agent i's set of skills. Thus, $w(X) = v(S_X)$. Typically, both v and w are nondecreasing: adding more skills or agents to a coalition never causes harm. We also assume that the characteristic function v is zero-normalized, i.e., the value of a single skill is zero.

We assume that the coalition and the value division (payoffs to agents) are established as follows.

- There exists a special agent whom we will call the *mechanism designer*. The mechanism designer knows T, the set of all possible skills,[2] and v, the characteristic function defined over T.
- If agent i is interested in joining a coalition, it declares the skills it has to the mechanism designer.
- The mechanism designer determines the value division among participants.

[1] The word "skills" should be interpreted broadly, e.g., they may correspond to resources that the agents possess.

[2] We do not require that each skill in T is actually possessed by some agent; the only thing that is required is that every skill that an agent possesses is indeed in T. Therefore, the mechanism designer really only needs to know an upper bound on the set of skills possessed by the agents.

For this setting, the following three types of manipulation by the agents have been identified: *hiding skills*, *using false names*, and *acting in collusion* [9]. These manipulations take the following forms, respectively:

- An agent i can declare that its skill set is $S'_i \subseteq S_i$. It is assumed that an agent cannot claim to have skills that it does not have. Such a lie is detectable because the lie will be exposed if the agent is called on to apply such skills.
- Agent i can use multiple identifiers and declare that each identifier has a subset of the skills S_i. Because the skills are unique, two different identifiers cannot declare that they have the same skill. Thus, a false-name manipulation by agent i corresponds to a partition of S_i into multiple identifiers.
- Multiple agents can collude, i.e., pretend to be a single agent. They can declare that this agent's skills are the union of their skills.

Combinations of these manipulations are also possible, e.g., hiding skills in addition to one of the last two manipulations.

To prevent these manipulations, we assume that the coalition and the value division (payoffs to agents) are established in the following way.

- The mechanism designer determines an outcome function $\pi(s, S)$ for all $S \subseteq T$ and $s \in S$. $\pi(s, S)$ determines the payoff to skill s when the total set of declared skills is S.
- The payoff to agent i, who declared a set of skills S_i, is determined as $\sum_{s \in S_i} \pi(s, S)$.

By using this method, the latter two manipulations become ineffective, since they do not change the payoffs to skills. Thus, to develop a solution concept that is robust to the manipulations under consideration, we only need to ensure that hiding skills is ineffective as well. This motivates the following condition.

Definition 2 (anonymity-proof outcome function). *An outcome function π is* anonymity-proof *iff $\forall S_i, \forall S'_i, \forall S$, subject to $S'_i \subset S_i$ and $S_i \cap S = \emptyset$, $\sum_{s \in S'_i} \pi(s, S'_i \cup S) \leq \sum_{s \in S_i} \pi(s, S_i \cup S)$ holds.*

This condition means that if agent i has a set of skills S_i, while the other agents have a combined set of skills S, then, for agent i, if it declares that it has only S'_i, which is a subset of S_i, then its payoff never increases.

The anonymity-proof core [9] is an anonymity-proof solution concept that satisfies the Pareto efficiency and *nonblocking* conditions described below.

Definition 3 (Pareto efficiency). *A solution concept is* Pareto efficient *iff the outcome function π that represents the solution concept satisfies the following condition: $\forall S, \sum_{s \in S} \pi(s, S) = v(S)$.*

Definition 4 (non-blocking condition). *A solution concept satisfies the* non-blocking condition *iff the outcome function π that represents the solution concept satisfies the following condition: $\forall S, \forall S' \subseteq S, \sum_{s \in S'} \pi(s, S) \geq v(S')$.*

Definition 5 (anonymity-proof core). *An anonymity-proof outcome function π is in the* anonymity-proof core *iff it satisfies Pareto efficiency and the non-blocking condition.*

Please note that for the second argument of the outcome function π, we need to consider *all* possible combinations (subsets) of skills. If the mechanism designer knows the set of skills possessed by agents beforehand, then it suffices to specify the value division for these skills. However, in our setting, we assume the mechanism designer knows only an upper bound on the set of skills. Thus, the mechanism designer needs to prepare value divisions for all possible subsets of skills. In general, for a game with a set of skills T, where $|T| = n$, a traditional solution concept needs to specify the value division only for T, while an anonymity-proof solution concept has to specify value divisions for all subsets of T, the number of which is $2^n - 1$. Thus, the size of the (naïve) representation of the outcome function is exponential in the number of skills.

Although we have developed a method for compactly representing the anonymity-proof core (and other anonymity-proof solution concepts), under the assumption that the characteristic function is represented using a compact language that explicitly specifies only coalitions that introduce synergy [6], the required representation size is still exponential in the number of skills/agents in the worst case. Also, the required computational cost for calculating an outcome function for these solution concepts is quite large.

3. SUMMARY OF MAIN RESULTS

In this section, we briefly describe the main technical results of this paper. The (traditional) Shapley value is characterized by the following axiomatic conditions: Pareto efficiency, null property, symmetry, and additivity. Since the Shapley value is not anonymity-proof [9], we need to introduce weaker notions of these axiomatic conditions. In Section 4, we will present a series of theorems to explain why/how these axiomatic conditions should be relaxed.

Among these conditions, we will not relax anonymity-proofness or Pareto efficiency, since these properties are basic requirements for any solution concept that is used in open, anonymous environments.

Figure 1 illustrates the main results of this paper. First, Theorem 1 shows that there exists no anonymity-proof and Pareto efficient solution concept that satisfies the null property. Thus, we need to relax the null property. We introduce a new axiomatic condition called the weak-null property (Figure 1 (i), on the left side).

However, Theorem 2 shows that there exists no anonymity-proof and Pareto efficient solution concept that satisfies the weak-null property and symmetry. Thus, we need to relax symmetry as well. We introduce a new axiomatic condition called weak-symmetry (Figure 1 (i), on the right side).

Furthermore, Theorem 3 shows that there exists no anonymity-proof and Pareto efficient solution concept that satisfies the weak-null property and additivity. Thus, we need to relax additivity. We introduce a new axiomatic condition called weak-additivity (Figure 1 (ii)).

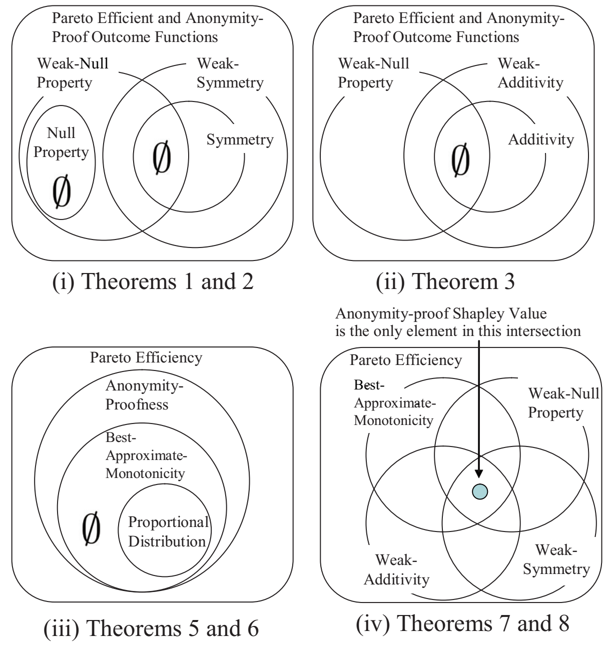

Figure 1: Summary of Main Results

In Section 5, we introduce a new axiomatic condition called best-approximate-monotonicity. Theorem 6 shows that this is a stricter condition than anonymity-proofness, i.e., any solution concept that is best-approximate-monotone is also anonymity-proof, but not vice versa (Figure 1 (iii)). Also, Theorem 5 shows that an outcome function is best-approximate-monotone if and only if the outcome function distributes the value for each skill $s \in S$ proportionally to the value divisions in the grand coalition of all skills T (Figure 1 (iii)).

In Section 6, we introduce our newly developed solution concept, i.e., the anonymity-proof Shapley value. Theorems 7 and 8 show that the anonymity-proof Shapley value is the only solution concept that satisfies Pareto efficiency, best-approximate-monotonicity, the weak-null property, weak-symmetry, and weak-additivity (Figure 1 (iv)).

4. IMPOSSIBILITY RESULTS

In this section, we present a series of theorems to explain why/how the axiomatic conditions that characterize the Shapley value should be relaxed to obtain a concept that is satisfactory for our setting.

First, we review the definition of the Shapley value [7].[3]

Definition 6 (Shapley value). *Give an ordering o of the set of skills T, let $S(o, s)$ be the set of skills in T that appear before s in the ordering o. Then, if the characteristic function of this game is v, the Shapley value for skill s is defined as*

$$sh_v(s, T) = \frac{1}{|T|!}(\sum_o v(S(o, s) \cup \{s\}) - v(S(o, s))).$$

[3]In the traditional definition, the Shapley value is defined over agents. In this definition, it is defined over skills.

The intuitive interpretation of the Shapley value is that it averages the marginal value of a skill over all possible orders in which the skills may join the coalition.

The Shapley value is the only solution concept that satisfies Pareto efficiency, the null property, symmetry, and additivity. We now show the formal definitions of these properties.

Definition 7 (null skill). *A skill s is a* null skill *in a set of skills S, iff s satisfies the following condition:*

- *$\forall S'$ for which $S' \ni s$ and $S' \subseteq S$, we have $v(S') = v(S' \setminus \{s\})$.*

Definition 8 (null property). *A solution concept satisfies the* null property *iff the outcome function π that represents the solution concept satisfies the following condition:*

- *if s is a null skill in S, then $\pi(s, S) = 0$.*

Definition 9 (symmetric skills). *Two skills $s, s' \in S$ are* symmetric *in S iff s and s' satisfy the following condition:*

- *$\forall S' \subseteq S \setminus (\{s, s'\})$, we have $v(S' \cup \{s\}) = v(S' \cup \{s'\})$.*

Definition 10 (symmetry). *A solution concept satisfies* symmetry *iff the outcome function π that represents the solution concept satisfies the following condition:*

- *if s and s' are symmetric in S, then $\forall S' \subseteq S \setminus (\{s, s'\})$, it must be that $\pi(s, S' \cup \{s\}) = \pi(s', S' \cup \{s'\})$ and $\pi(s, S' \cup \{s, s'\}) = \pi(s', S' \cup \{s, s'\})$ hold.*

Definition 11 (additivity). *A solution concept satisfies* additivity *iff for any characteristic functions v, v_1, and v_2, the corresponding outcome functions π, π_1, and π_2 that represent the solution concept satisfy the following condition:*

- *if the characteristic functions v, v_1, v_2 satisfy $v = v_1 + v_2$, then $\forall S \subseteq T$, it must be that $\forall s \in S$, $\pi(s, S) = \pi_1(s, S) + \pi_2(s, S)$.*

The Shapley value satisfies Pareto efficiency, the null property, symmetry, and additivity. However, as shown in [9] it is not anonymity-proof. Let us consider the following example.

Example 1. *Let there be three skills a, b, and c. Let the characteristic function over skills be as follows.*

- *$v(\{a, b\}) = v(\{a, c\}) = v(\{a, b, c\}) = 1$,*
- *$v(\{a\}) = v(\{b\}) = v(\{c\}) = v(\{b, c\}) = 0$.*

Let us assume π is the outcome function for the Shapley value. Then, $\pi(a, \{a, b, c\}) = 2/3$ and $\pi(b, \{a, b, c\}) = \pi(c, \{a, b, c\}) = 1/6$. On the other hand, if there are only two skills a and b, $\pi(a, \{a, b\}) = \pi(b, \{a, b\}) = 1/2$, since these two skills are symmetric in $\{a, b\}$.

Thus, the Shapley value is not anonymity-proof, since $\pi(b, \{a, b, c\}) + \pi(c, \{a, b, c\}) = 1/3 < 1/2 = \pi(b, \{a, b\})$. More specifically, let us assume there are two agents 1 and 2, and 1 has a and 2 has b and c, respectively. If agent 2 declares it has b and c, it receives $1/3$. On the other hand, if it declares it has only b, then it receives $1/2$. Thus, hiding a skill is profitable for agent 2.

Since the Shapley value is not anonymity-proof, we need to relax some of these axiomatic conditions if we are to obtain a good solution concept. We now present a series of theorems to explain why/how these axiomatic conditions should be relaxed.

Theorem 1. *There exists no solution concept that satisfies the null property, Pareto efficiency, and anonymity-proofness.*

Proof. Let us assume some outcome function π satisfies these three properties. Suppose there are four skills, $T = \{a, b, c, d\}$, and that the characteristic function is defined as follows:

- $v(a, b, c, d) = 1$,
- $v(a, b, c) = v(a, b, d) = v(a, c, d) = v(b, c, d) = 1$,
- $v(a, b) = v(c, d) = 1$,
- otherwise, $v(S) = 0$.

If only a, b and d are present, since d is a null skill in this situation, $\pi(d, \{a, b, d\}) = 0$. From Pareto efficiency, $\pi(a, \{a, b, d\}) + \pi(b, \{a, b, d\}) = 1$ must hold. Furthermore, from anonymity-proofness, the following condition must hold.

$$\begin{aligned} & \pi(a, T) + \pi(b, T) + \pi(c, T) \\ \geq \; & \pi(a, \{a, b, d\}) + \pi(b, \{a, b, d\}) = 1 \end{aligned} \tag{1}$$

Similarly, we know that:

- When a, b, c are present, c is a null skill.
- When b, c, d are present, b is a null skill.
- When a, c, d are present, a is a null skill.

By using the same argument as before in each case, we obtain the following conditions:

$$\pi(a, T) + \pi(b, T) + \pi(d, T) \geq 1 \tag{2}$$

$$\pi(a, T) + \pi(c, T) + \pi(d, T) \geq 1 \tag{3}$$

$$\pi(b, T) + \pi(c, T) + \pi(d, T) \geq 1 \tag{4}$$

By adding inequalities 1-4 together and dividing by 3, we obtain $\sum_{s \in T} \pi(s, T) \geq 4/3$, which is impossible because the payments must sum to 1 if π is Pareto efficient. □

From Theorem 1, it follows that we need to introduce a weaker version of the null property. We propose the following new axiomatic condition, called the weak-null property.

Definition 12 (weak-null property). *A solution concept satisfies the* weak-null property *iff the outcome function π that represents the solution concept satisfies the following condition:*

- *if s is a null skill in T, then $\forall S \subseteq T$, $\pi(s, S) = 0$.*

An outcome function π that satisfies the weak-null property must always assign 0 to s if s is a null skill in the grand coalition T. However, if s is a null skill in $S \subseteq T$, but s is not a null skill in T, $\pi(s, S)$ can be a positive value without violating the weak-null property.

Next, we show that we need to relax symmetry and additivity. This is demonstrated by the following two theorems.

Theorem 2. *There exists no solution concept that satisfies the weak-null property, symmetry, Pareto efficiency, and anonymity-proofness.*

Proof. Let us consider a characteristic function v that is identical to the example in Theorem 1, except that a null skill in grand coalition e is added Moreover, assume that

π satisfies the conditions in the theorem. Suppose a, b, d, e are present. Since e is a null skill in the grand coalition, $\pi(e, \{a, b, d, e\}) = 0$ must hold. Also, since d and e are symmetric in $\{a, b, d, e\}$, $\pi(d, \{a, b, d, e\}) = 0$ must hold. From these facts, and the fact that π is Pareto efficient, it follows that $\pi(a, \{a, b, d, e\}) + \pi(b, \{a, b, d, e\}) = 1$ must hold. Also, from anonymity-proofness, the following condition must hold.

$$\begin{aligned} & \pi(a, T) + \pi(b, T) + \pi(c, T) \\ \geq \quad & \pi(a, \{a, b, d, e\}) + \pi(b, \{a, b, d, e\}) = 1 \end{aligned} \tag{5}$$

Similarly, we know that:

- c and e are symmetric in $\{a, b, c, e\}$.
- b and e are symmetric in $\{b, c, d, e\}$.
- a and e are symmetric in $\{a, c, d, e\}$.

By using the same argument as before in each case, we obtain the following conditions:

$$\pi(a, T) + \pi(b, T) + \pi(d, T) \geq 1 \tag{6}$$

$$\pi(a, T) + \pi(c, T) + \pi(d, T) \geq 1 \tag{7}$$

$$\pi(b, T) + \pi(c, T) + \pi(d, T) \geq 1 \tag{8}$$

By summing inequalities 5, 6, 7, and 8, and dividing by 3, we obtain $\pi(a, T) + \pi(b, T) + \pi(c, T) + \pi(d, T) \geq 4/3$, which is impossible because the payments must sum to 1 if π is Pareto efficient. □

Theorem 3. *There exists no solution concept that satisfies the weak-null property, additivity, Pareto efficiency, and anonymity-proofness.*

Proof. Let v_1 be identical to the characteristic function in the example of Theorem 1. Also, let v_2 be defined by $v_2(\{a, b, c, d\}) = 1$, and $v_2(S) = 0$ otherwise. Also, let v_3 and v_4 be defined as follows, so that $v_1 + v_2 = v_3 + v_4$ holds.

- $v_3(\{a, b, c, d\}) = v_3(\{a, b, c\}) = v_3(\{a, b, d\}) = v_3(\{a, b\}) = 1$, and $v_3(S) = 0$ otherwise.
- $v_4(\{a, b, c, d\}) = v_4(\{a, c, d\}) = v_4(\{b, c, d\}) = v_4(\{c, d\}) = 1$, and $v_4(S) = 0$ otherwise.

Let us assume that a solution concept with the properties in the theorem exists, and let $\pi_1,\ \pi_2,\ \pi_3,\ \pi_4$ be the corresponding outcome functions for $v_1,\ v_2,\ v_3,\ v_4$, respectively. By additivity, $\pi_1 + \pi_2 = \pi_3 + \pi_4$ must hold.

We will show that π_1 does not satisfy anonymity-proofness. By Pareto efficiency, there exists at least one skill that obtains positive value in the grand coalition. Without loss of generality, we assume $\pi_1(c, \{a, b, c, d\}) > 0$ holds.

Let us consider the case where a, b, c are present. By the weak-null property, $\pi_3(c, \{a, b, c\}) = 0$ must hold. From this fact and the fact that $\pi_4(c, \{a, b, c\}) = 0$ (by Pareto efficiency), it must be that $\pi_1(c, \{a, b, c\}) = 0$ holds, because $v_1 + v_2 = v_3 + v_4$. Thus, because $\pi_1(c, \{a, b, c\}) = 0$, we have $\pi_1(a, \{a, b, c\}) + \pi_1(b, \{a, b, c\}) = 1$ by Pareto efficiency.

Now, let us consider the case where a, b, c, d are present. By anonymity-proofness, $\pi_1(a, \{a, b, c, d\}) + \pi_1(b, \{a, b, c, d\}) + \pi_1(d, \{a, b, c, d\}) \geq \pi_1(a, \{a, b, c\}) + \pi_1(b, \{a, b, c\}) = 1$ must hold. However, since we assumed $\pi_1(c, \{a, b, c, d\}) > 0$, $\pi_1(a, \{a, b, c, d\}) + \pi_1(b, \{a, b, c, d\}) + \pi_1(d, \{a, b, c, d\}) < 1$ must hold by Pareto efficiency. Thus, we have derived a contradiction. □

From theorems 2 and 3, it follows that we need to introduce weaker notions of symmetry and additivity to obtain a desirable solution concept. We propose the following new axiomatic conditions.

Definition 13 (weak-symmetry). *A solution concept satisfies* weak-symmetry *iff the outcome function π that represents the solution concept satisfies the following condition:*

- *π assigns the same value to s and s', if they are symmetric skills in T, i.e., $\forall S \subseteq T \setminus \{s, s'\}$, $\pi(s, S \cup \{s\}) = \pi(s', S \cup \{s'\})$ and $\pi(s, S \cup \{s, s'\}) = \pi(s', S \cup \{s, s'\})$ hold.*

Definition 14 (weak-additivity). *A solution concept satisfies* weak-additivity *iff for any characteristic functions v, v_1, v_2 (with corresponding outcome functions π, π_1, and π_2), subject to $v = v_1 + v_2$, $\pi(s, T) = \pi_1(s, T) + \pi_2(s, T)$ holds.*

5. BEST-APPROXIMATE-MONOTONICITY

In this section, we consider another condition called best-approximate-monotonicity. It turns out that this is a stricter condition than anonymity-proofness, i.e., any solution concept that is best-approximate-monotone is also anonymity-proof, but not vice versa.

To define best-approximate-monotonicity, we first define *α-approximate-monotonicity*.

Definition 15 (α-approximate-monotonicity). *An outcome function π is* α-approximate-monotone *for S and S', subject to $S \subseteq S'$, iff $\max_{s \in S} \pi(s, S)/\pi(s, S') = \alpha$ holds.*

Since we assume that the characteristic function v is monotone, for all $S \subseteq S' \subseteq T$, $v(S) \leq v(S')$ holds. Thus, it is intuitively natural to think that $\pi(s, S)/\pi(s, S') \leq 1$ can be made to hold (i.e., 1-approximate-monotonicity). However, the following theorem contradicts that.

Theorem 4. *There exists no Pareto efficient solution concept that satisfies 1-approximate-monotonicity for all $S \subseteq S' \subseteq T$.*

Proof. Let v be defined as follows.

- $v(\{a, b, c\}) = v(\{a, b\}) = v(\{b, c\}) = v(\{a, c\}) = 1$, and $v(S) = 0$ otherwise.

Let us assume π satisfies 1-approximate-monotonicity and Pareto efficiency. By Pareto efficiency, there exists at least one skill that obtains a positive value in the grand coalition. Without loss of generality, we assume $\pi(a, T) = \epsilon > 0$. By Pareto efficiency, we have $\pi(b, T) + \pi(c, T) = 1 - \epsilon$. By 1-approximate-monotonicity, we have $\pi(b, \{b, c\}) \leq \pi(b, T)$ and $\pi(c, \{b, c\}) \leq \pi(c, T)$. Thus, we have $\pi(b, \{b, c\}) + \pi(c, \{b, c\}) \leq \pi(b, T) + \pi(c, T) = 1 - \epsilon$. However, this contradicts the assumption that π is Pareto efficient, since it implies that $\pi(b, \{b, c\}) + \pi(c, \{b, c\})$ must be 1. □

Since there exists no solution concept that satisfies 1-approximate-monotonicity for all $S \subseteq S' \subseteq T$ as well as Pareto efficiency, for some characteristic functions, α must be greater than 1 for some $S \subseteq S'$.

Now, we define best-approximate-monotonicity.

DEFINITION 16 (BEST-APPROXIMATE-MONOTONICITY). *A Pareto efficient outcome function π is* best-approximate-monotone *iff for all $S \subseteq S' \subseteq T$, for any (other) Pareto efficient outcome function π' for which $\pi'(s,S') = \pi(s,S')$ for all $s \in S'$, $\max_{s\in S} \pi(s,S)/\pi(s,S') \leq \max_{s\in S} \pi'(s,S)/\pi'(s,S')$ holds.*

In other words, a Pareto efficient outcome function π satisfies best-approximate-monotonicity if for every $S \subseteq S' \subseteq T$, π is $\alpha_{\pi,S,S'}$-approximate-monotone, and there exists no other Pareto efficient outcome function π' that is the same as π on S' and that is α-approximate-monotone for the same $S \subseteq S' \subseteq T$ for some $\alpha < \alpha_{\pi,S,S'}$.

Now, we show that best-approximate-monotonicity implies anonymity-proofness. First, we identify a lower-bound on α for $S \subseteq S' \subseteq T$.

LEMMA 1. *If π satisfies Pareto efficiency and α-approximate-monotonicity for $S \subseteq S' \subseteq T$, then $\alpha \geq v(S)/\sum_{t\in S}\pi(t,S')$ holds.*

PROOF. Suppose π is Pareto efficient and α-approximate-monotone for $S \subseteq S' \subseteq T$. Since π is α-approximate-monotone for $S \subseteq S' \subseteq T$, for all $s \in S$, $\pi(s,S) \leq \alpha\pi(s,S')$ holds. Thus, $\sum_{s\in S}\pi(s,S) \leq \alpha\sum_{t\in S}\pi(t,S')$ holds. By Pareto efficiency, $\sum_{s\in S}\pi(s,S) = v(S)$.

Thus, $v(S)/\sum_{t\in S}\pi(t,S') \leq \alpha$ holds. □

Actually, this lower bound is tight, by the following lemma.

LEMMA 2. *There exists an outcome function π that is Pareto efficient and $\alpha_{\pi,S,S'}$-approximate-monotone for all $S \subseteq S' \subseteq T$, where $\alpha_{\pi,S,S'} = v(S)/\sum_{t\in S}\pi(t,S')$. Specifically, π is defined as follows:*

- *Choose $\pi(s,T)$ for all $s \in T$ so that Pareto efficiency for T is satisfied.*
- *Define $\pi(s,S)$, where $S \subset T$, as $v(S)\cdot\frac{\pi(s,T)}{\sum_{t\in S}\pi(t,T)}$.*

PROOF. π is Pareto efficient because: $\sum_{s\in S}\pi(s,S) = v(S)\times\frac{\sum_{s\in S}\pi(s,T)}{\sum_{t\in S}\pi(t,T)} = v(S)$. We now show that $\pi(s,S)/\pi(s,S') = v(S)/\sum_{t\in S}\pi(t,S')$ holds. From the definition of π, $\pi(s,S) = v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$, and $\pi(s,S') = v(S')\cdot\pi(s,T)/\sum_{t\in S'}\pi(t,T)$ hold. Thus, $\pi(s,S)/\pi(s,S') = v(S)\cdot\frac{\sum_{t\in S'}\pi(t,T)}{v(S')\cdot\sum_{t\in S}\pi(t,T)}$ holds. Now, we are going to show that $\sum_{t\in S}\pi(t,S') = v(S')\cdot\frac{\sum_{t\in S}\pi(t,T)}{\sum_{t\in S'}\pi(t,T)}$ holds; this is the inverse of the fraction in the previous equation, and by replacing that fraction with $1/\sum_{t\in S}\pi(t,S')$, the result is proved. By the definition of π, $\pi(t,S') = v(S')\cdot\pi(t,T)/\sum_{t'\in S'}\pi(t',T)$ holds.

Thus, $\sum_{t\in S}\pi(t,S') = v(S')\cdot\sum_{t\in S}\pi(t,T)/\sum_{t'\in S'}\pi(t',T)$ holds, and we are done. □

Also, an outcome function is best-approximate-monotone if and only if it is of the form introduced in Lemma 2.

THEOREM 5. *An outcome function π is best-approximate-monotone if and only if for all $S \subset T$, $s \in S$, $\pi(s,S) = v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$ holds.*

PROOF. First, we prove the "only if" part. Suppose π satisfies best-approximate-monotonicity (which implies π is Pareto efficient).

Now, let us assume for some s, S, where $s \in S \subseteq T$, $\pi(s,S) \neq v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$ holds.

Also, let us define π' by $\pi'(s,T) = \pi(s,T)$, and $\pi'(s,S) = v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$ for all $s \in S \subseteq T$.

If $\pi(s,S) > v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$ holds, then $\pi(s,S)/\pi(s,T) > v(S)/\sum_{t\in S}\pi(t,S)$ holds. However, this means that $\max_{s\in S}\frac{\pi'(s,S)}{\pi'(s,T)} = \frac{v(S)}{\sum_{t\in S}\pi(t,S)}$, is strictly smaller than $\frac{\pi(s,S)}{\pi(s,T)} \leq \max_{s\in S}\frac{\pi(s,S)}{\pi(s,T)}$. This contradicts the assumption that π is best-approximate-monotone, since π and π' are identical on the grand coalition T.

Also, if $\pi(s,S) < v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,S)$ holds, then by Pareto efficiency, there exists at least one $s' \in S$ for which $\pi(s',S) > v(S)\cdot\pi(s',T)/\sum_{t\in S}\pi(t,T)$ holds, which we already showed leads to a contradiction.

Next, we prove the "if" part.

Suppose $\pi(s,S) = v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$ holds for all $S \subseteq T$, $s \in S$. Now, let us assume π is not best-approximate monotone, i.e., there exists another outcome function π' and $S \subseteq S' \subseteq T$, such that $\pi'(s,S') = \pi(s,S')$ for all $s \in S'$, and $\max_{s\in S}\pi(s,S)/\pi(s,S') > \max_{s\in S}\pi'(s,S)/\pi'(s,S')$ holds. By Lemma 2, $\pi(s,S)/\pi(s,S') = v(S)/\sum_{t\in S}\pi(t,S')$ holds. By Lemma 1, $\max_{s\in S}\pi'(s,S)/\pi'(s,S') \geq v(S)/\sum_{t\in S}\pi'(t,S')$ holds. Also, from $\sum_{t\in S}\pi'(t,S') = \sum_{t\in S}\pi(t,S')$, $\max_{s\in S}\pi'(s,S)/\pi'(s,S') \geq \max_{s\in S}\pi(s,S)/\pi(s,S')$ holds. This contradicts the assumption that $\max_{s\in S}\pi(s,S)/\pi(s,S') > \max_{s\in S}\pi'(s,S)/\pi'(s,S')$ holds. □

Now, we are ready to show that best-approximate-monotonicity implies anonymity-proofness.

THEOREM 6. *If an outcome function π is best-approximate-monotone, then it is anonymity-proof.*

PROOF. Assume π is best-approximate-monotone (which implies π is also Pareto efficient). From Theorem 5, $\pi(s,S) = v(S)\cdot\pi(s,T)/\sum_{t\in S}\pi(t,T)$ holds for all $s \in S \subseteq T$.

We show that by hiding skills, the obtained value never increases, i.e., for any sets of skills S, S', S'', where $S' \subset S''$ and $S \cap S'' = \emptyset$, $\sum_{s\in S''}\pi(s,S\cup S'') \geq \sum_{s\in S'}\pi(s,S\cup S')$ holds.

Here, $\sum_{s\in S''}\pi(s,S\cup S'') = v(S\cup S'')\cdot\frac{\sum_{s\in S''}\pi(s,T)}{\sum_{s\in S\cup S''}\pi(s,T)}$. Also, $\sum_{s\in S'}\pi(s,S\cup S') = v(S\cup S')\cdot\frac{\sum_{s\in S'}\pi(s,T)}{\sum_{s\in S\cup S'}\pi(s,T)}$.

Since $v(S\cup S') \leq v(S\cup S'')$, it suffices to show that $\frac{\sum_{s\in S''}\pi(s,T)}{\sum_{s\in S\cup S''}\pi(s,T)} \geq \frac{\sum_{s\in S'}\pi(s,T)}{\sum_{s\in S\cup S'}\pi(s,T)}$.

If we denote $A = \sum_{s\in S'}\pi(s,T)$, $B = \sum_{s\in(S''\setminus S')}\pi(s,T)$, and $C = \sum_{s\in S\cup S'}\pi(s,T)$ (thus, $A+B = \sum_{s\in S''}\pi(s,T)$ and $C+B = \sum_{s\in S\cup S''}\pi(s,T)$), this condition is represented as $\frac{A+B}{C+B} \geq \frac{A}{C}$. It is clear that this condition holds when $A \leq C$ and $B \geq 0$. □

6. ANONYMITY-PROOF SHAPLEY VALUE

In this section, we introduce our newly developed solution concept, which we call the anonymity-proof Shapley value. It satisfies Pareto efficiency, the weak-null property, weak-symmetry, weak-additivity, and best-approximate-monotonicity (and hence, it automatically satisfies anonymity-proofness). Also, it is the only solution concept that satisfies these properties simultaneously.

DEFINITION 17 (ANONYMITY-PROOF SHAPLEY VALUE). *The* anonymity-proof Shapley value *is defined as follows. Let $sh_v(s,T)$ be the Shapley value for skill s in the grand coalition T for characteristic function v.*

- $\forall S \subseteq T,\ \forall s \in S$,
  $\pi(s,S) = v(S) \cdot sh_v(s,T) / \sum_{t \in S} sh_v(t,T)$

When S is a set of skills that are present, then the anonymity-proof Shapley value distributes $v(S)$ among the skills in S, and the value that each skill $s \in S$ receives is proportional to the Shapley value of s in the grand coalition T.

THEOREM 7. *The anonymity-proof Shapley value satisfies Pareto efficiency, the weak-null property, weak-symmetry, weak-additivity, and best-approximate-monotonicity.*

PROOF. Let π be the outcome function of the anonymity-proof Shapley value. From Definition 17, $\pi(s,S) = v(S) \cdot sh_v(s,T)/\sum_{t \in S} sh_v(t,T)$ holds. We have $\sum_{s \in S} \pi(s,S) = v(S) \cdot \sum_{s \in S} sh_v(s,T)/\sum_{t \in S} sh_v(t,T) = v(S)$, so Pareto efficiency is satisfied.

Next, we show that π satisfies the weak-null property. Suppose that s is a null skill in the grand coalition T. Since the Shapley value satisfies the null property, $sh_v(s,T) = 0$ holds. Thus, by Definition 17, $\pi(s,S) = 0$ holds for all $S \subseteq T$.

Then, we show that π satisfies weak-symmetry. Suppose that s_1 and s_2 are symmetric in the grand coalition T. Since the Shapley value satisfies symmetry, $sh_v(s_1,T) = sh_v(s_2,T)$ holds. Also, $v(\{s_1\} \cup S) = v(\{s_2\} \cup S)$ holds for all $S \subseteq T \setminus \{s, s'\}$. Thus, from Definition 17, $\pi(s, S \cup \{s\}) = \pi(s', S \cup \{s'\})$ and $\pi(s, S \cup \{s,s'\}) = \pi(s', S \cup \{s,s'\})$ hold for all $S \subseteq T \setminus \{s,s'\}$.

Next, we show that π satisfies weak-additivity. Suppose that $v = v_1 + v_2$ holds for three characteristic functions v, v_1, and v_2. Since the Shapley value satisfies additivity, $sh_v(s,T) = sh_{v_1}(s,T) + sh_{v_2}(s,T)$ holds. Thus, from Definition 17, $\pi(s,T) = \pi_1(s,T) + \pi_2(s,T)$ holds, where π, π_1, and π_2 are the outcome functions when the characteristic functions are v, v_1, and v_2, respectively. Thus, the anonymity-proof Shapley value satisfies weak-additivity.

Finally, by Theorem 5, π also satisfies best-approximate-monotonicity. □

THEOREM 8. *The anonymity-proof Shapley value is the only solution concept that satisfies weak-null property, weak-symmetry, weak-additivity, and best-approximate-monotonicity simultaneously.*

PROOF. Suppose π satisfies the weak-null property, weak-symmetry, weak-additivity, and best-approximate-monotonicity simultaneously. For the grand coalition T, by Pareto efficiency, the weak-null property, weak-symmetry, and weak-additivity, π must be identical to the Shapley value [7].

Also, because π satisfies best-approximate-monotonicity, by Theorem 5, $\pi(s,S) = v(S) \cdot \pi(s,T)/\sum_{t \in S} \pi(t,T)$ must hold for any $s \in S \subset T$. Thus, π is identical to the anonymity-proof Shapley value. □

7. DISCUSSIONS

7.1 Computational/Representational Cost

The anonymity-proof Shapley value is identical to the Shapley value in the grand coalition T. After we obtain the Shapley value for the grand coalition, the computational cost for obtaining the anonymity-proof Shapley value for each $S \subseteq T$ is at most $O(n)$, where $n = |T|$. Thus, it suffices to compute/store the anonymity-proof Shapley value only for the grand coalition T, and compute the rest on demand according to S that are actually present. Thus, the computational cost for the anonymity-proof Shapley value is about the same as that for the traditional Shapley value. The computational cost for finding the traditional Shapley value depends on the representation of characteristic functions. Ieong and Shoham developed a compact representation scheme of characteristic functions [5]. Using this representation, the Shapley value can be computed in $O(2^n)$. Also, if we can decompose a problem into smaller sub-problems by utilizing (weak-)additivity, we can obtain further speedup.

On the other hand, one needs to solve a linear program with $O(n \cdot 2^n)$ variables to compute one element of the anonymity-proof core. Also, the amount of information that we store for the anonymity-proof Shapley value is $O(n)$, which is much smaller than the representational cost of the anonymity-proof core, i.e., $O(n \cdot 2^n)$.

7.2 Variations of the Anonymity-proof Shapley Value

The anonymity-proof Shapley value is the only solution concept that satisfies Pareto efficiency, the weak-null property, weak-symmetry, weak-additivity, and best-approximate-monotonicity. On the other hand, if we replace best-approximate-monotonicity by anonymity-proofness, other solution concepts can satisfy these axiomatic conditions.

For example, let us consider an outcome function π, where $\pi(s,T) = sh_v(s,T)$ for $s \in T$, and for $s \in S \subseteq T$, $\pi(s,S)$ is determined as follows.

- A priority ordering among skills, where symmetric skills in the grand coalition have the same order, is determined.
- The skills with the highest priority take a share equal to $\pi(s,T)$ from $v(S)$, then the skills with the next-highest priority do the same, etc.
- If at some point the remaining amount is not enough, then the remaining amount is divided equally among the remaining skills with the highest priority.
- If there remains some amount after all skills take their shares, then the amount is distributed among skills with the highest priority.

Since π is identical to the Shapley value at the grand coalition T, it is clear that this outcome function satisfies Pareto efficiency, the weak-null property, weak-symmetry, and weak-additivity. Also, the value is distributed according to the predefined priority ordering. Thus, if agent i, who has S skills, hides a skill $s \in S$, it will miss the opportunity to obtain $sh_v(s,T)$. This might increase the share for another skill $s' \in S$, where s''s priority is lower than s. However, this increased amount never exceeds $sh_v(s,T)$. Thus, this outcome function is also anonymity-proof.

However, we believe the anonymity-proof Shapley value is more intuitively natural than the outcome function defined above. In the next section, we show that the anonymity-proof Shapley value is closer to the original Shapley value than such an outcome function.

8. EVALUATION

The anonymity-proof Shapley value is identical to the Shapley value for the grand coalition T, and, in a sense, it

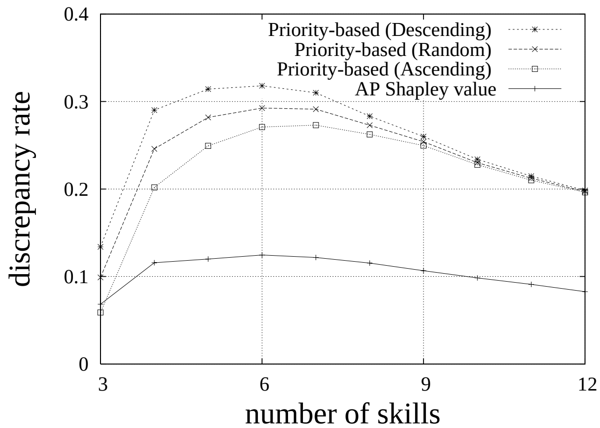

Figure 2: Comparison of Discrepancy Rate

does some kind of approximation for every smaller coalition $S \subset T$. One question we might ask is: how close is it to the traditional Shapley value on S? Also, as shown in the previous section, if we replace best-approximate-monotonicity by anonymity-proofness, there can be other solution concepts. In this section, we introduce a notion called *discrepancy rate* to measure the discrepancy between a solution concept and the traditional Shapley value.

DEFINITION 18 (DISCREPANCY RATE). *The discrepancy rate of an outcome function π is defined as follows:*

$$\frac{1}{2^{|T|}} \times \sum_{S \subseteq T} \frac{\sum_{s \in S} |\pi(s, S) - sh_v(s, S)|}{2v(S)}$$

The idea is to calculate the discrepancy rate for each $S \subseteq T$ and take the average (there are $2^{|T|}$ subsets). We do not count cases where $v(S) = 0$. The discrepancy for S is calculated as follows: for $S = \{s_1, s_2, \ldots\}$, consider vectors $\langle \pi(s_1, S), \pi(s_2, S), \ldots \rangle$ and $\langle sh_v(s_1, S), sh_v(s_2, S), \ldots \rangle$. We calculate the sum of the absolute values of the differences over the elements of these two vectors. Since $\sum_{s \in S} \pi(s, S) = \sum_{s \in S} sh_v(s, S) = v(S)$, the sum of the above differences is at most $2v(S)$. We normalize the sum by $2v(S)$, so the discrepancy rate is between 0 and 1. If the outcome function and the Shapley value are identical, then the discrepancy rate is 0. If they are totally different—for example, if the outcome function gives all of the value to a null skill—then the discrepancy rate becomes 1.

In Figure 2, we compare the discrepancy rates of the anonymity-proof Shapley value and priority-based anonymity-proof outcome functions described in the previous section, where the priority ordering is determined: (i) randomly, (ii) in descending order of the Shapley value, and (iii) in ascending order of the Shapley value. We vary $|T|$ from 3 to 12. We randomly generate the characteristic function v, in such a way that it is weakly increasing. More specifically, for $S \subset S' \subseteq T$, $v(S')$ is determined by $\max_{S \subseteq S'} v(S) + \delta$, where δ is chosen randomly from $[0, 2]$.

We generated 100 problem instances for each $|T|$; the average is shown. These results illustrate that the discrepancy rate of the anonymity-proof Shapley value is quite small (around 0.1) and is smaller than other outcome functions.

9. CONCLUSIONS

Anonymity-proof solution concepts are designed to be robust to various manipulations in open anonymous environments. However, for existing concepts, the representation size of the outcome function is exponential in the number of skills that agents declare (while that of the outcome function of conventional solution concepts is linear). Also, the required computational cost is quite large.

In this paper, we developed a new solution concept called the anonymity-proof Shapley value that drastically reduces these computational/representational costs. This reduction is possible because we need to compute/store only the value division for the grand coalition, and calculate the rest on demand. The anonymity-proof Shapley value is characterized by simple axiomatic conditions: Pareto efficiency, the weak-null property, weak-symmetry, weak-additivity, and best-approximate-monotonicity. It is the only solution concept that satisfies these axiomatic conditions simultaneously, always exists, and is uniquely determined.

10. REFERENCES

[1] T. Agotnes, W. van der Hoek, and M. Wooldridge. Quantified coalition logic. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pp. 1181–1186, 2007.
[2] V. Conitzer and T. Sandholm. Complexity of determining nonemptiness of the core. In *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pp. 613–618, 2003.
[3] V. Conitzer and T. Sandholm. Computing Shapley values, manipulating value division schemes,and checking core membership in multi-issue domains. In *Proc. of the 19th National Conf. on Artificial Intelligence (AAAI)*, pp. 219–225, 2004.
[4] E. Elkind, L. A. Goldberg, P. W. Goldberg, and M. Wooldridge. Computational complexity of weighted threshold games. In *Proc. of the 22nd Conf. on Artificial Intelligence (AAAI)*, pp. 718–723, 2007.
[5] S. Ieong and Y. Shoham. Marginal contribution nets: a compact representation scheme for coalitional games. In *Proc. of the 6th ACM Conf. on Electronic commerce (ACM-EC)*, pp. 193–202, 2005.
[6] N. Ohta, A. Iwasaki, M. Yokoo, K. Maruono, V. Conitzer, and T. Sandholm. A compact representation scheme for coalitional games in open anonymous environments. In *Proc. of the 21st Conf. on Artificial Intelligence (AAAI)*, pp. 509–514, 2006.
[7] L. S. Shapley. A value for n-person games. In *Contributions to the Theory of Games*, pp. 307–317. Princeton University Press, 1953.
[8] O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1-2):165–200, 1998.
[9] M. Yokoo, V. Conitzer, T. Sandholm, N. Ohta, and A. Iwasaki. Coalitional games in open anonymous environments. In *Proc. of the 20th National Conf. on Artificial Intelligence (AAAI)*, pp. 509–515, 2005.
[10] G. Zlotkin and J. S. Rosenschein. Coalition, cryptography and stability: Mechanisms for coalition formation in task oriented domains. In *Proc. of the 12th National Conf. on Artificial Intelligence (AAAI)*, pp. 432–437, 1994.