Coalitional Games in Open Anonymous Environments*

Makoto Yokoo†, Vincent Conitzer‡, Tuomas Sandholm‡, Naoki Ohta†, Atsushi Iwasaki†

†Faculty of Information Science
and Electrical Engineering
Kyushu University
6-10-1 Hakozaki, Higashi-ku,
Fukuoka, 812-8581 Japan
yokoo/iwasaki@is.kyushu-u.ac.jp
oota@lang.is.kyushu-u.ac.jp

‡Computer Science Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
conitzer/sandholm@cs.cmu.edu

Abstract

Coalition formation is a key aspect of 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. Various solution concepts (such as the Shapley value, core, least core, and nucleolus) have been proposed. In this paper, we demonstrate how these concepts are vulnerable to various kinds of manipulations in open anonymous environments such as the Internet. These manipulations include submitting false names (one acting as many), collusion (many acting as one), and the hiding of skills. To address these threats, we introduce a new solution concept called the *anonymity-proof core*, which is robust to these manipulations. We show that the anonymity-proof core is characterized by certain simple axiomatic conditions. Furthermore, we show that by relaxing these conditions, we obtain a concept called the least anonymity-proof core, which is guaranteed to be non-empty. We also show that computational hardness of manipulation may provide an alternative barrier to manipulation.

Introduction

Coalition formation is a key capability in automated negotiation among self-interested agents. In order for coalition formation to be successful, 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 multi-agent systems literature (Zlotkin & Rosenschein 1994; Yagodnick & Rosenschein 1998; Ketchpel 1994; Shehory & Kraus 1998; Conitzer & Sandholm 2003; 2004).

Besides being of interest of the game-theory and multi-agent systems research community, the growth of the Internet and e-commerce has expanded the application areas of coalitional game theory. For example, consider a large number of companies, some subset of which could form profitable virtual organizations that can respond to larger or more diverse orders than an individual company can. Due to the advance of the Internet, forming such virtual organizations becomes much easier, but the companies must agree on how to divide the profit among themselves.

However, existing solution concepts have limitations when applied to open anonymous environments such as the Internet. In such environments, a single agent can use multiple identifiers (or *false names*), pretending to be multiple agents, and distribute its ability (skills) among these identifiers. Alternatively, multiple agents can collude and pretend to be a single agent that combines all of their skills. Furthermore, an agent might try to hide some of its skills. Finally, complex combinations of these manipulations are also possible—for example, three agents could collude and use two identifiers, distributing some of each agent's skills to each identifier, while hiding other skills.

These manipulations are virtually impossible to detect in open anonymous environments, and have thus become an issue in such environments specifically. That is also the reason why the gamut of these manipulations, in particular, false-name manipulations, has not received much research attention previously. (An important exception is the work on pseudonymous bidding in combinatorial Internet auctions. For that setting, false-name proof protocols have been developed (Yokoo, Sakurai, & Matsubara 2004).)

In this paper, we provide a range of examples that demonstrate the vulnerability of the existing coalitional solution concepts to these new types of manipulations.[1] We then develop a new solution concept for coalitional games called the *anonymity-proof core*, which is robust to the manipulations described above. We show that the anonymity-proof core is characterized by certain axiomatic conditions (including that an agent does not have an incentive to use the manipulations mentioned above). Furthermore, we show that by relaxing some of these conditions, we obtain a concept called the *least anonymity-proof core*, which is guaranteed to be non-empty. Also, as an alternative protection from manipulation, we show that manipulating the traditional solution concepts may be computationally hard.

*Yokoo and Iwasaki are partially supported by Japan Society for the Promotion of Science under Grant-in-Aid for Scientific Research 17300049. Conitzer and Sandholm are supported by the National Science Foundation under ITR grants IIS-0121678 and IIS-0427858, and a Sloan Fellowship.

[1] A similar claim has been made regarding allocation rules in a pure exchange economy (Sertel & Yildiz 2004).

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. This assumes that utility is transferable (for example, utility can be transferred using side payments), and that a coalition's value is independent of what non-members of the coalition do.

The characteristic function by itself does not give us sufficient information to assess what manipulations may be performed by agents in an open anonymous environment. For example, if an agent decides to use false names and to split itself into two different identifiers, then what is the new characteristic function over the new set of agent identifiers? Presumably, this depends on *how* the agent splits itself into multiple agents—but the characteristic function does not contain any information on different ways in which the agent may split itself up. Because of this, we need a more fine-grained representation of what each agent brings to the table. Instead of defining the characteristic function over agents, we define it over *skills* that the agents possess. (The word "skills" should not be interpreted too strictly—while the skills may indeed correspond to the abilities of the agents, they may also correspond to, for example, resources that the agents possess.)

Definition 1 (skills and agents) *Assume the set of all possible skills is T. Each agent i has one or multiple skills $S_i \subset T$. For simplicity, we assume each skill is unique, that is, $\forall i \neq j, S_i \cap S_j = \emptyset$ holds.*[2]

Definition 2 (characteristic function defined over skills) *A characteristic function $v : 2^T \rightarrow \Re$ assigns a value to each set of skills.*

We will 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 $S_X = \bigcup_{i \in X} S_i$. Then, we have $w(X) = v(S_X)$. The characteristic function over skills is a more fine-grained representation than the characteristic function over agents (the latter can be derived from the former but not vice versa). Typically, both v and w are weakly increasing: adding more skills or agents to a coalition never hurts.

We assume that the coalition and the value division (payoffs to the 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,[3] 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.

We assume the following three types of manipulation (or any combination of them) are possible for agents.

Definition 3 (hiding skills) *If agent i has a set of skills S_i, for any $S'_i \subseteq S_i$, it can declare that it has only S'_i.*

On the other hand, we assume that an agent cannot declare it has a skill that it does not have in reality—we assume that such a lie is detectable (because the lie will be exposed once the agents in the coalition are called on to apply their skills).

Definition 4 (false names) *Agent i can use multiple identifiers and declare that each identifier has a subset of skills S_i. Since we assume each skill is unique, two different identifiers cannot declare they have the same skill.*[4] *Thus, a false-name manipulation by agent i corresponds to a partition of S_i into multiple identifiers. (If the manipulation is combined with a skill-hiding manipulation, only a subset of S_i is partitioned.)*

Definition 5 (collusion) *Multiple agents can collude and pretend to be a single agent. They can declare the skills of this agent to be the union of their skills (or a subset of this union, in case we combine the manipulation with a skill-hiding manipulation).*

We can combine the various manipulations to obtain more complex manipulations. We already described how to combine the latter two manipulations with a skill-hiding manipulation. As another example, an agent with skills a and b, and another agent with only skill c, can collude and submit one identifier with only skill a, and another with skills b and c. This can be seen as a combination of the latter two manipulations: 1) the first agent splits into two false names (one with skill a and one with skill b), 2) the second false name (with skill b) colludes with the agent with skill c. More generally, the following result shows that all manipulations can be seen as a combination of these three basic manipulations.

Theorem 1 *Letting S be the union of a coalition's skills, any manipulation in which this coalition submits several identifiers with non-overlapping subsets of S as their skills can be achieved by a combination of the previous three manipulations.*

We omit the proof due to space constraint.

Traditional Solution Concepts

So far, we have not yet discussed how the value of the coalition should be divided. In this section, we briefly review some of the traditional solution concepts for doing so. First, we review a well-known solution concept known as the *Shapley value* (Shapley 1953). The Shapley value aims to distribute the gains from cooperation in a fair manner. It has many equivalent characterizations; we will review one that gives a formula in closed form for it.

[2] This assumption is just for making the notation simple; even if there are two identical skills, we can set different names on them.

[3] 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.

[4] Alternatively, we can consider a case where agents can declare that they have multiple "copies" of a single skill. We hope to address this model in our future works.

Definition 6 (Shapley value) *Give an ordering o of the set of agents W in the coalition, let $X(o, i)$ be the set of agents in W that appear before i in ordering o. Then the Shapley value for agent i is defined as*

$$Sh(W, i) = \frac{1}{|W|!} \sum_o (w(X(o, i) \cup \{i\}) - w(X(o, i))).$$

The intuitive interpretation of the Shapley value is that it averages an agent's marginal value over all possible orders in which the agents may join the coalition.

Next, we show another well-known (perhaps the best known) solution concept called the *core* (Gillies 1953; von Neumann & Morgenstein 1947).

Definition 7 (core) *Given a set of agents W, an outcome, that is, a value division $c^W = (c_1^W, c_2^W, \ldots)$ among agents is in the core if the following two conditions hold:*

1. $\forall X \subset W, \sum_{i \in X} c_i^W \geq w(X)$,
2. $\sum_{i \in W} c_i^W = w(W)$.

The first condition is called the *non-blocking* condition: if for some set of agents X, this condition does not hold, then the agents in X have an incentive to collectively deviate from the mechanism and to divide $w(X)$ among themselves. The second condition is called the *feasibility* condition: if $\sum_{i \in W} c_i^W > w(W)$, this outcome is infeasible.[5] Thus, an outcome is in the core if it is blocked by no coalition and feasible. In general, the core can be empty. Also, the core can contain a large set of outcomes.

Next, we show a solution concept called the *least* core, which can be either a relaxation or a tightening of the core. We first define the ϵ-core, which is used in the later definition.

Definition 8 (ϵ-core) *Given a set of agents W and value ϵ, an outcome $c^W = (c_1^W, c_2^W, \ldots)$ is in the ϵ-core if the following two conditions hold:*

1. $\forall X \subset W, \sum_{i \in X} c_i^W \geq w(X) - \epsilon$,
2. $\sum_{i \in W} c_i^W = w(W)$.

If $\epsilon = 0$, this definition coincides with the definition of the core. If ϵ is positive (negative), the non-blocking condition is relaxed (resp. tightened). It is obvious that for any $\epsilon < \epsilon'$, if an outcome is in the ϵ-core, it also in the ϵ'-core.

Now, we can define the least core as follows.

Definition 9 (least core) *Given a set of agents W, an outcome $c^W = (c_1^W, c_2^W, \ldots)$ is in the least core if the following two conditions hold.*

- *c^W is in the ϵ-core,*
- *$\forall \epsilon' < \epsilon$, the ϵ'-core is empty.*

The least core is non-empty for any characteristic function, but it may contain multiple outcomes. The solution concept known as *nucleolus* (Schmeidler 1969) is a refinement of the least core. It is guaranteed to be in the least core and uniquely determined for any characteristic function. Due to limited space, we omit the formal definition of the nucleolus.

Manipulability of Traditional Solution Concepts

In this section, we show a number of ways in which traditional solution concepts may fail.

Vulnerability to False Names

Example 1 *Let there be three skills a, b, and c. Let all three skills be necessary, that is, let the characteristic function over skills be as follows:*

- *$v(\{a, b, c\}) = 1$,*
- *For any proper subset $S \subset \{a, b, c\}$, $v(S) = 0$.*

Let agent 1 have skill a and let agent 2 have skills b and c. Then, the characteristic function over agents is as follows.

- $w(\{1\}) = w(\{2\}) = 0$,
- $w(\{1, 2\}) = 1$.

In this example, there are only two possible orderings of the agents and in each of those, the second agent in the ordering has marginal contribution 1. Therefore, the Shapley value for each agent is $1/2$. Any outcome (c_1, c_2) that satisfies $c_1 \geq 0, c_2 \geq 0$, and $c_1 + c_2 = 1$ is in the core. The least core has only one outcome, which is identical to the Shapley value. Hence, the nucleolus is identical to the Shapley value.

Example 2 *Let the skills and the function v be the same as in Example 1. Let there be three agents $1, 2$, and 3 who have skills a, b, and c, respectively. Then, the characteristic function over agents is as follows.*

- $w(\{1, 2, 3\}) = 1$,
- *For any proper subset $X \subset \{1, 2, 3\}$, $w(X) = 0$.*

In this example, there are six possible orderings of the agents, and the last agent has marginal contribution 1. Therefore, the Shapley value for each agent is $1/3$. Any outcome (c_1, c_2, c_3) that satisfies $c_1 \geq 0, c_2 \geq 0, c_3 \geq 0$, and $c_1 + c_2 + c_3 = 1$ is in the core. The least core has only one outcome, which is identical to the Shapley value. Hence, the nucleolus is identical to the Shapley value.

Now, we can see that the Shapley value, the least core, and nucleolus are all vulnerable to false-name manipulations: in Example 1, agent 2 can use two identifiers 2 and 3 and split its skills over these identifiers. Then, the situation becomes identical to Example 2. Thus, agent 2 can increase the value awarded to it from $1/2$ to $2/3 = 1/3 + 1/3$ using false-names. In fact, this example proves the following result:

Theorem 2 *There exists no payoff division function that 1) equally rewards the agents that are symmetric with respect to w, 2) distributes all the value, and 3) is false-name proof.*

Proof: Assume a function satisfies 1) and 2). Then, this function should coincide with the Shapley value (or nucleolus) on both Examples 1 and 2. However, we have just shown that such a payoff division is not false-name proof. □

Vulnerability to Collusion

Example 3 *Let there be three skills a, b, and c. Let the characteristic function over skills be as follows.*

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

[5] Usually, the feasibility condition is represented as $\sum_{i \in W} c_i^W \leq w(W)$. From the non-blocking condition, the equality must hold.

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

Let there be three agents $1, 2,$ and 3 with skills $a, b,$ and c, respectively. Then, the characteristic function over agents is as follows.

- $w(\{1, 2\}) = w(\{1, 3\}) = w(\{1, 2, 3\}) = 1$,
- $w(\{1\}) = w(\{2\}) = w(\{3\}) = w(\{2, 3\}) = 0$.

In this example, there are six possible orderings of the agents. The marginal contribution of agent 1 is 0 only if it is the first agent in the ordering, and 1 otherwise. Hence, the Shapley value of agent 1 is $2/3$, and the Shapley value of each of agents 2 and 3 is $1/6$ (if two agents are symmetric, their Shapley values must be identical). In this example, there is only one outcome in the core, namely outcome $(1, 0, 0)$. This is because if agent 2 (or 3) obtains any value, then the non-blocking condition is violated because agent 1 and agent 3 (or 2) have an incentive to deviate from the mechanism and form their own coalition. This is also the only outcome of the least core, and hence the nucleolus also gives this outcome.

Example 4 *Let the skills and the function v be the same as in Example 3. Let there be two agents, and let agent 1 have skill a and let agent 2 have skills b and c. Then, the characteristic function over agents is identical to the one in Example 1.*

Since the characteristic function over agents is identical to the one in Example 1, the Shapley value, the core, the least core, and nucleolus, are also identical to those for Example 1.

Now, we can see that the Shapley value, the least core, and the nucleolus are all vulnerable to collusion: in Example 3, agent 2 and 3 can collude and pretend to be a single agent 2, who has skills b and c. Then, the situation becomes identical to Example 4. Thus, if the Shapley value is used, agent 2 and 3 can increase the total value awarded to them from $1/3 = 1/6 + 1/6$ to $1/2$. Also, if the least core or the nucleolus is used, agent 2 and 3 can increase the total value awarded to them from 0 to $1/2$.

Applying the Solution Concepts to Skills

The examples from the previous section show that the characteristic function w defined over agents is too coarse-grained to represent the relative importance of agents in an open anonymous environment. In this section, we take a different approach: what if we apply the traditional solution concepts directly to the (finer-grained) characteristic function v over skills? That is, treat each submitted skill as an imaginary agent, and use the characteristic function v over these imaginary agents to compute the value division over them. Then, give each (real) agent the sum of the payoffs to the skills that it submitted. For example, in Example 1, the Shapley value, least core, and nucleolus would all give a value of $1/3$ to each skill; therefore, agent 1 receives $1/3$, and agent 2 (having submitted two skills) receives $2/3$.

Theorem 3 *Applying any solution concept to the skills directly is robust to false names, collusion, and any combinations thereof.*

Proof: Because solution concepts applied to the skills directly are indifferent to which agent submitted which skills, changing the identifiers under which skills are submitted never changes the payoffs to those skills. □

However, there may still be incentives to hide skills, as we demonstrate next. Consider again Example 4. If we calculate the solution concepts over the skills directly, the payoffs to these skills (for any one of the solution concepts) are the same as the payoffs to agents in Example 3. Thus, agent 1 receives $2/3$ and agent 2 receives $1/3$ if we use the Shapley value applied to the skills, and agent 1 receives 1 and agent 2 receives 0 if we use the core/least core/nucleolus applied to the skills. Now, consider the following example.

Example 5 *Let there be two skills a and b. Let the characteristic function over skills be as follows:*

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

Let agent 1 have skill a and let agent 2 have skill b.

It is easy to see that both the Shapley value and the least core/nucleolus will give $1/2$ to each agent in this example. Now, we can see that the Shapley value, the least core, and the nucleolus are all vulnerable to the hiding of skills when applied directly to the skills. In Example 4, agent 2 can hide skill c. Then, the situation becomes identical to Example 5. Hence the agent 2 increases its payoff from $1/3$ to $1/2$ for the Shapley value, and from 0 to $1/2$ for the least core/nucleolus.

Anonymity-Proof Core

In this section, we develop a new solution concept for our setting which we call *anonymity-proof core*. As we will show, the anonymity-proof core can be characterized by certain axiomatic conditions. Again, we assume that the only knowledge that the mechanism designer has is T, that is, the set of all possible skills, and v, that is, the characteristic function defined over T. The mechanism designer does not know the number of agents, or the skills that each agent has. The mechanism designer must define an outcome function π that decides, for all possible reports by the agents of their skills, how to divide the value generated by these skills.

We require that the outcome function π is *anonymous*, that is, the payoff to an agent does not depend on the identifiers of the agents; it depends only on the skills of the agent and the distribution of skills over other agents.

More specifically, given an agent i and a set of other agents Y, let S_i be the set of skills that agent i has, and let $SS_Y = \{S_j \mid j \in Y\}$, where S_j is the set of skills that agent j has. Then, the outcome function $\pi(S_i, SS_Y)$ takes S_i and SS_Y as arguments, and returns the payoff to agent i, when agent i declares its skills as S_i and the other agents declare their skills as SS_Y.

Let the set of agents who joined the mechanism be W, and let the profile of the skills that agents declared be $k = (k_1, k_2, \ldots)$, where k_i is the set of skills that agent i declared. Let $S_X = \bigcup_{i \in X} k_i$, that is, S_X is the union of the skills declared by a set of agents X; let $S = S_W$; and let $SS_X = \{k_i \mid i \in X\}$. Also, let $SS_{\sim i} =$

$\{k_1, \ldots, k_{i-1}, k_{i+1}, \ldots\}$, that is, a set, each of whose elements is the set of skills corresponding to agent j ($j \neq i$).

We now give six axiomatic conditions that the outcome function π should satisfy.

1. The outcome function π is anonymous.
2. π is never blocked by any coalition, that is, $\forall k, \forall X \subseteq W$, $\sum_{i \in X} \pi(k_i, SS_{\sim i}) \geq v(S_X)$ holds.
3. π is always feasible and always distributes all of the value, that is, $\forall k, \sum_{i \in W} \pi(k_i, SS_{\sim i}) = v(S)$ holds.
4. π is robust against hiding skills, that is, $\forall S', S''$, where $S'' \subseteq S', \forall SS, \pi(S'', SS) \leq \pi(S', SS)$ holds.
5. π is robust against false-name manipulations, that is, $\forall k, \forall X \subseteq W$, $Y = W \setminus X$, $\sum_{i \in X} \pi(k_i, SS_{\sim i}) \leq \pi(S_X, SS_Y)$ holds.
6. π is robust against collusions, that is, $\forall k, \forall X \subseteq W, Y = W \setminus X$, $\sum_{i \in X} \pi(k_i, SS_{\sim i}) \geq \pi(S_X, SS_Y)$ holds.

In order to define the anonymity-proof core, we first formally define the core for skills. For a set of skills $S = \{s_1, s_2, \ldots\}$, we define a set of core outcomes for skills $Core(S)$ as follows.

Definition 10 (core for skills) *$c^S = (c^S_{s_1}, c^S_{s_2}, \ldots)$ is in $Core(S)$ if it satisfies the following two conditions.*

- *$\forall S' \subset S, \sum_{s_j \in S'} c^S_{s_j} \geq v(S')$,*
- *$\sum_{s_j \in S} c^S_{s_j} = v(S)$.*

Now we are ready to define the anonymity-proof core.

Definition 11 (anonymity-proof core) *We say the outcome function π_{ap} is in the anonymity-proof core if π_{ap} satisfies the following two conditions.*

1. *For any set of skills $S \subseteq T$, there exists a core outcome for S, that is, some $c^S = (c^S_{s_1}, c^S_{s_2}, \ldots) \in Core(S)$, such that for any skill profile $k = (k_1, k_2, \ldots,)$ with $\bigcup_i k_i = S$, $\pi_{ap}(k_i, SS_{\sim i}) = \sum_{s_j \in k_i} c^S_{s_j}$ holds.*
2. *$\forall S', S''$, where $S'' \subseteq S'$, $\forall SS$, $\pi_{ap}(S'', SS) \leq \pi_{ap}(S', SS)$ holds.*

The first condition says that for any set of skills reported by the agents, some outcome in the core for that set of skills should be used to distribute the value. The second condition says that an agent has no incentive to hide its skills.

Example 6 *Let the skills and the function v be identical to those in Example 3. Since $c^{\{a,b,c\}} = (c_a^{\{a,b,c\}}, c_b^{\{a,b,c\}}, c_c^{\{a,b,c\}}) \in Core(\{a,b,c\})$ if $c_a^{\{a,b,c\}} = 1, c_b^{\{a,b,c\}} = 0, c_c^{\{a,b,c\}} = 0$, the following outcome function is in the anonymity-proof core:*

- *$\pi_{ap}(\{a\}, \{\{b, \ldots\}, \ldots\}) = \pi_{ap}(\{a\}, \{\{c, \ldots\}, \ldots\}) = \pi_{ap}(\{a, b\}, \{\ldots\}) = \pi_{ap}(\{a, c\}, \{\ldots\}) = \pi_{ap}(\{a, b, c\}, \{\}) = 1$,*
- *$\pi_{ap} = 0$ everywhere else.*

First, we show that outcome functions in the anonymity-proof core satisfy the axioms.

Theorem 4 *Any outcome function π_{ap} in the anonymity-proof core satisfies the six axioms.*

Proof: Axiom 1 holds because π_{ap} only considers which skills were reported and not by which agent they were reported. Axiom 4 holds by the second condition of the anonymity-proof core. Also, using the first condition of the anonymity-proof core, for any set of skills $S \subseteq T$, there exists a core outcome for S, that is, some $c^S = (c^S_{s_1}, c^S_{s_2}, \ldots) \in Core(S)$, such that for any skill profile $k = (k_1, k_2, \ldots,)$ with $\bigcup_i k_i = S$, $\pi_{ap}(k_i, SS_{\sim i}) = \sum_{s_j \in k_i} c^S_{s_j}$ holds. Therefore, $\forall k$ with $\bigcup_i k_i = S$, $\forall X \subseteq W$, $\sum_{i \in X} \pi_{ap}(k_i, SS_{\sim i}) = \sum_{i \in X} \sum_{s_j \in k_i} c^S_{s_j} = \sum_{s_j \in S_X} c^S_{s_j} \geq v(S_X)$. Thus, Axiom 2 holds. Also, $\sum_{i \in W} \pi_{ap}(k_i, SS_{\sim i}) = \sum_{i \in W} \sum_{s_j \in k_i} c^S_{s_j} = \sum_{s_j \in S} c^S_{s_j} = v(S)$. Thus, Axiom 3 holds. Also, $\forall k$ with $\bigcup_i k_i = S, \forall X \subseteq W, Y = W \setminus X, \sum_{i \in X} \pi_{ap}(k_i, SS_{\sim i}) = \sum_{i \in X} \sum_{s_j \in k_i} c^S_{s_j} = \pi_{ap}(S_X, SS_Y)$ holds. Therefore, Axiom 5 and Axiom 6 hold. □

Next, we prove that any outcome function that satisfies the above six axioms is actually in the anonymity-proof core. To prove this, we use the following lemma.

Lemma 1 *If an outcome function π satisfies the six axioms, then for any skill profile $k = (k_1, k_2, \ldots,)$ where $\bigcup_i k_i = S, \pi(k_i, SS_{\sim i}) = \pi(k_i, \{S \setminus k_i\})$ holds.*

Proof: Using Axiom 3, we obtain the following:

$$\pi(k_i, SS_{\sim i}) + \sum_{j \neq i} \pi(k_j, SS_{\sim j}) \quad = v(S) = \\ \pi(k_i, \{S \setminus k_i\}) + \pi(S \setminus k_i, \{k_i\}).$$

Also, from Axiom 5 and Axiom 6, $\sum_{j \neq i} \pi(k_j, SS_{\sim j}) = \pi(S \setminus k_i, \{k_i\})$ holds. Thus, $\pi(k_i, SS_{\sim i}) = \pi(k_i, \{S \setminus k_i\})$ holds. □

Hence, the outcome of agent i is determined by k_i and $S \setminus k_i$, that is, the skills of agent i and the union of skills of other agents, i.e., how the skills in $S \setminus k_i$ are distributed among other agents does not affect the payoff to agent i.

Theorem 5 *Any outcome function π that satisfies the six axioms is in the anonymity-proof core.*

Proof: If outcome function π satisfies the six axioms, then π satisfies the second condition of the anonymity-proof core, since Axiom 4 is identical to the second condition. All that remains to show is that π also satisfies the first condition of an anonymity-proof core. For a set of skills $S = \{s_1, s_2, \ldots\}$, let us denote $SS = \{\{s_j\} \mid s_j \in S\}$. From Axiom 2 and Axiom 3, considering the case where each agent has exactly one skill in S, the following two conditions hold.

- $\forall S' \subset S, \sum_{s_j \in S'} \pi(\{s_j\}, SS \setminus \{\{s_j\}\}) \geq v(S')$,
- $\sum_{s_j \in S} \pi(\{s_j\}, SS \setminus \{\{s_j\}\}) = v(S)$.

These conditions are identical to the conditions in Definition 10. Therefore, if we let $c^S_{s_j} = \pi(\{s_j\}, SS \setminus \{\{s_j\}\})$, this constitutes an element of $Core(S)$. Moreover, using Lemma 1, $\pi(\{s_j\}, SS \setminus \{\{s_j\}\}) = \pi(\{s_j\}, \{S \setminus \{s_j\}\}) = c^S_{s_j}$ holds. Now, using Lemma 1, Axiom 5, and Axiom 6, for any skill profile $k = (k_1, k_2, \ldots,)$, where

$\bigcup_i k_i = S$, and $SS_{\sim i} = \{k_1, k_2, \ldots, k_{i-1}, k_{i+1}, \ldots\}$, $\pi(k_i, SS_{\sim i}) = \pi(k_i, \{S \setminus k_i\}) = \sum_{s_j \in k_i} \pi(\{s_j\}, \{S \setminus \{s_j\}\}) = \sum_{s_j \in k_i} c^S_{s_j}$ holds. Thus, the first condition of the anonymity-proof core is satisfied. □

It is clear that if for some $S \subseteq T$, $Core(S)$ is empty, then there exists no function π_{ap} in the anonymity-proof core. The following theorem shows that the inverse is not true.

Theorem 6 *Even if for all $S \subseteq T$, $Core(S)$ is non-empty, it may be the case that the anonymity-proof core is empty.*

We use a counterexample that involves four skills to prove this theorem. Also, it turns out that if there are only three skills, no counterexamples exist. (We omit the proof due to space constraint.)

As in the case of the traditional core, there are conditions on the characteristic function v, such as convexity, under which the anonymity-proof core can be guaranteed to be non-empty. We omit this section due to limited space.

Least Anonymity-proof Core

We first define the ϵ-core for skills.

Definition 12 (ϵ-core for skills) *For given ϵ, $c^S = (c^S_{s_1}, c^S_{s_2}, \ldots)$ is in ϵ-$Core(S)$ if it satisfies the following two conditions.*

- $\forall S' \subset S, \sum_{s_j \in S'} c^S_{s_j} \geq v(S') - \epsilon$,
- $\sum_{s_j \in S} c^S_{s_j} = v(S)$.

By replacing $Core(S)$ to ϵ-$Core(S)$ in Definition 11, we obtain the definition of an ϵ-anonymity-proof core. An ϵ-anonymity-proof core satisfies all axioms except Axiom 2. Therefore, although a group of agents might have an incentive to deviate from the mechanism, they don't have an incentive to use other manipulations.

A least anonymity-proof core can be defined as follows.

Definition 13 (least anonymity-proof core) *We say the outcome function π_{ap} is in the least anonymity-proof core if π_{ap} satisfies the following conditions.*

- *π_{ap} is in the ϵ-anonymity-proof core.*
- *$\forall \epsilon' < \epsilon$, the ϵ'-anonymity-proof core is empty.*

The following theorem holds.

Theorem 7 *$\forall T, v$, there always exists an outcome function that is in the least anonymity-proof core.*

This theorem holds since if we set ϵ large enough, we can choose an ϵ-core so that the second condition in Definition 11 can be satisfied. We omit the detail due to space constraint.

Computational Manipulation Protection

Even when a value division scheme that is vulnerable against manipulations is used, it may be computationally hard to find a beneficial manipulation. This barrier of computational hardness may prevent the manipulation from occurring.

Theorem 8 *When the Shapley value over agents is used to distribute the value, it is NP-complete to determine whether an agent can benefit by submitting false names.*

The proof is by reduction from satisfiability. We omit the details due to limited space.

Conclusions

We demonstrated that traditional solution concepts of coalitional games—namely the Shapley value, core, least core, and nucleolus—are vulnerable to various kinds of manipulations in open anonymous environments such as the Internet. Specifically, we showed that they can be vulnerable to the submission of false-name identifiers and collusion. We showed that the problems of false-name identifiers and collusion can be prevented by applying the solution concepts to the skills directly rather than to the agents, but this is still vulnerable to the hiding of skills. We then introduced a solution concept called the *anonymity-proof core*, which is robust to these manipulations. We characterized the anonymity-proof core by certain simple axioms. Also, we introduced another concept called the least anonymity-proof core, which is guaranteed to be non-empty. Finally, we showed that computational hardness may provide an alternative barrier to manipulation.

References

Conitzer, V., and Sandholm, T. 2003. Complexity of determining nonemptiness of the core. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, 613–618.

Conitzer, V., and Sandholm, T. 2004. Computing Shapley values, manipulating value division schemes, and checking core membership in multi-issue domains. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 219–225.

Gillies, D. 1953. *Some theorems on n-person games*. Ph.D. Dissertation, Princeton University, Department of Mathematics.

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

Schmeidler, D. 1969. The nucleolus of a characteristic function game. *Society for Industrial and Applied Mathematics Journal of Applied Mathematics* 17:1163–1170.

Sertel, M., and Yildiz, M. 2004. Core is manipulable via segmentation. *Journal of Economic Theory* 118:103–117.

Shapley, L. S. 1953. A value for n-person games. In Kuhn, H. W., and Tucker, A. W., eds., *Contributions to the Theory of Games*, volume 2 of *Annals of Mathematics Studies, 28*. Princeton University Press. 307–317.

Shehory, O., and Kraus, S. 1998. Methods for task allocation via agent coalition formation. *Artificial Intelligence* 101(1–2):165–200.

von Neumann, J., and Morgenstein, O. 1947. *Theory of games and economic behavior*. Princeton University Press.

Yagodnick, R., and Rosenschein, J. S. 1998. Lies in multiagent subadditive task oriented domains. In *The International Workshop on Multi-Agent Systems*.

Yokoo, M.; Sakurai, Y.; and Matsubara, S. 2004. The effect of false-name bids in combinatorial auctions: New fraud in Internet auctions. *Games and Economic Behavior* 46(1):174–188.

Zlotkin, G., and Rosenschein, J. S. 1994. Coalition, cryptography and stability: Mechanisms for coalition formation in task oriented domains. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 432–437.