Coalitional unanimity versus strategy-proofness in coalition formation problems

Koji TAKAMIYA[*]
Economics, Niigata University

April 24, 2012

Abstract

This paper examines *coalition formation problems* from the viewpoint of mechanism design. We consider the case where (i) the list of *feasible coalitions* (those coalitions which are permitted to form) is given in advance; and (ii) each individual's preference is a ranking over those feasible coalitions which include this individual. We are interested in requiring the mechanism to guarantee each coalition the "right" of forming that coalition at least when every member of the coalition ranks the coalition at the top. We name this property *coalitional unanimity*. We examine the compatibility between coalitional unanimity and incentive requirements, and prove that if the mechanism is *strategy-proof* and respects coalitional unanimity, then for each preference profile, there exists at most one *strictly core stable partition*, and the mechanism chooses such a partition whenever available. Further, the mechanism is *coalition strategy-proof* and respects coalitional unanimity if, and only if, the strictly core stable partition uniquely exists for every preference profile.

JEL Classification— C71, C72, C78, D02, D71, D78.

Keywords— coalition formation problems, strict core stability, strategy-proofness, coalitional unanimity.

1 Introduction

1.1 Motivation

This paper examines *coalition formation problems* from the viewpoint of mechanism design. In our version of coalition formation problems, the list of the coalitions which are permitted to form are given a priori. These coalitions are called *feasible coalitions*. And each individual is assumed to have a preference ranking over those feasible coalitions which contain this individual. To our knowledge, our model of coalition formation described above was introduced by Pápai (2004). Our model is general enough to include several important models such as the *marriage problem* and the *roommate*

[*] Contact. Faculty of Economics, Niigata University, 8050 2-no-cho Ikarashi Nishi-ku Niigata JAPAN, 950-2181, tel: +81-25-262-6498, E-mail: takamiya@econ.niigata-u.ac.jp.

problem (Gale and Shapley, 1962), and the model of *hedonic coalition formation* introduced independently by Banerjee, Konishi and Sönmez (2001), Bogomolnaia and Jackson (2002) and Cechlárová and Romero-Medina (2001).

Our theme is to identify conclusions from requiring mechanisms to guarantee *group rights*. This embodies the idea that *each group of people can make a decision on their own at least when the decision is an* internal *concern of the group*. This idea is an natural extension of the idea of "property right" from the individual level to the group level. This study considers a specific form of group right requirement in the following sense: We search for those mechanisms which guarantee that *each feasible coalition can form that coalition at least when every member of the coalition ranks the coalition at the top*. We name this property of mechanisms *coalitional unanimity*. This property has been studied in the context of the *marriage problem* by Takagi and Serizawa (2010). We recognized the importance of this property from an ethical viewpoint and were motivated to extend this property to a general coalition formation setting.[1]

From the viewpoint of mechanism design, those mechanisms not only satisfy the group right requirement described above but also have to be incentive compatible. Thus our concern is the compatibility between these two kinds of requirements, namely *group rights* and *incentive*. In this paper, we concentrate on *direct mechanisms*. That is, a *mechanism* is a single-valued function which specifies a feasible partition of the grand coalition for each profile of preferences. Further, we adopt *strategy-proofness* and *coalition strategy-proofness* as the relevant incentive requirements. In summing up, our problem is to see what mechanisms are *(coalition) strategy-proof* and respects *coalitional unanimity* at the same time.

1.2 Results

We postulate a class of preference domains which generalizes the strict preference domain. We prove that (i) if the mechanism is *strategy-proof* and respects *coalitional unanimity*, then *for each preference profile the mechanism chooses a strictly core stable partition whenever available, and there exists at most one strictly core stable partition.* And we show that (ii) if the requirement of strategy-proofness in the above result is strengthened to that of *coalition strategy-proofness*, then *for every preference profile there exists only one strictly core stable partition, and the mechanism chooses this partition.* Further, the converse of the second result holds true: (iii) *if for each preference profile, there exists only one strictly core stable partition, and the mechanism chooses this partition, then the mechanism is coalition strategy-proof and respects coalitional unanimity.*

Although our results are not impossibility results, they yield impossibility in many special cases because the requirement that there exists only one (or at most one) strictly

[1] After the first version of this paper was submitted to this journal, we were pointed out that an axiom essentially similar to our *coalitional unanimity* had been independently introduced in Rodríguez-Álvarez (2009). There the similar axiom was called *top coalition* and formulated as an auxiliary axiom to be used in the proofs of his results, not as one of his main axioms. We will mention the paper of Rodríguez-Álvarez in Sec.4. His axiom is named after the concept of "top coalitions" of Banerjee, Konishi and Sönmez (2001). We will mention the relationship between coalitional unanimity and their top coalition concept after the axiom is formally defined. (See Sec.2.2.)

core stable partition for every preference profile is very strong. An example where the impossibility obtains is the *marriage problem* (Takagi and Serizawa, 2010). So our results reveal a severe limitation on the possibility of strategy-proof mechanisms which respect coalitional unanimity.

In a broader perspective, our results exhibit the *conflict* between *strategy-proofness* and the *guarantee of group rights*. The conclusions of our results are very similar to that of the well-known result by Sönmez (1999): In a general model of discrete allocations which contains the present class of problems as a special case, *strategy-proofness*, *Pareto efficiency* and *individual rationality* together imply that any two strict core allocations are *Pareto indifferent* (which means the strict core is regarded as *unique* in terms of welfare) for those preference profiles for which the strict core is not empty, and that the mechanism chooses a strict core allocation for those profiles. Thus in the context of coalition formation, the conflict between strategy-proofness and group rights shows a similar quality as the well-known conflict between *strategy-proofness* and *Pareto efficiency* (plus *individual rationality*), which has been central to the study of strategy-proof mechanisms.

1.3 Related works

We note three papers highly related to the present work. Pápai (2004) introduces the coalition formation model which we study in this paper, and also studies unique strict cores for the case of *strict preferences.* She provides a necessary and sufficient condition, called the *single-lapping condition*, that the set of feasible coalitions is to satisfy for the coalition formation problem to have a unique core stable partition for all the preference profiles.[2] Further, she shows that under this condition, the mechanism which chooses the strict core stable partition for every preference profile is the unique mechanism which is *strategy-proof*, *individually rational* and *Pareto efficient.*

Takagi and Serizawa (2010) study strategy-proof mechanisms for *marriage problems* (Gale and Shapley, 1962). They introduce a property called "pairwise unanimity," which is the version of coalitional unanimity in the context of the marriage problem. They prove that there does not exist any mechanism which is strategy-proof and respects pairwise unanimity. This result can be derived from the first one of our three results. Their work has directly motivated the present work.

Toda (2006) studies *set-valued rules* for marriage problems, and independently introduces a set-valued version of the pairwise unanimity of Takagi and Serizawa (2010).[3] He proves that any rule which is Maskin monotonic and respects coalitional unanimity is a subcorrespondence of the stable correspondence. He also obtains characterizations of the stable correspondence using Maskin monotonicity and additional properties.

[2] Pápai (2004) actually deals with the core stability rather than strict core stability. But these two concepts are equivalent when preferences are strict, as is assumed in her paper.

[3] Toda (2006) calls this property *mutually best.*

2 Preliminaries

2.1 Model of coalition formation

Let $N = \{1, 2, \cdots, n\}$ with $n \geq 2$ be the set of **individuals**. A **coalition** is a nonempty subset of N. A **coalition formation problem** is a list $(N, \mathscr{F}, \succsim)$. Here $\mathscr{F}$ is the set of **feasible coalitions**. $\mathscr{F}$ is a nonempty subset of the set of all coalitions, $\{S \mid \emptyset \neq S \subset N\}$. For each $i \in N$, $\mathscr{F}(i)$ denotes the set of feasible coalitions that contain i, that is, $\{S \mid i \in S \in \mathscr{F}\}$. We assume for any $i \in N$, $\{i\} \in \mathscr{F}$.

A partition of N is called a **feasible partition** if the partition consists only of feasible coalitions. Let x be a feasible partition, and let $i \in N$. Then $x(i)$ denotes the coalition in x which contains i. Let us denote by $X(\mathscr{F})$ the set of feasible partitions. In the following, as long as there is no ambiguity, we refer to them simply "partitions." And we also call them "outcomes" depending on the context.

$\succsim = (\succsim_i)_{i \in N}$ is a **preference profile**. For each $i \in N$, $\succsim_i$ is a weak ordering (complete and transitive binary relation) over $\mathscr{F}(i)$. As usual, $\succ_i$ denotes the asymmetric part, and $\sim_i$ denotes the symmetric part of $\succsim_i$.

Note that we are assuming that preferences are *hedonic*, that is, for any individual, preferences depend only on the composition of the coalition of which that individual is a member.[4] Let $x, y \in X(\mathscr{F})$. Then abusing notation, let us denote

$$x \succsim_i y \tag{1}$$

if and only if

$$x(i) \succsim_i y(i). \tag{2}$$

Likewise, we often regard preference relations $\succsim_i$ defined on $\mathscr{F}(i)$ as if they are defined on $X(\mathscr{F})$.

As far as we know, the present model was first studied by Pápai (2004). This model generalizes the model of *hedonic coalition formation* independently introduced by Banerjee, Konishi and Sönmez (2001), Bogomolnaia and Jackson (2002), and Cechlárová and Romero-Medina (2001), where all coalitions are assumed to be feasible. Also our model includes as a special case the well-known *marriage problems* (two-sided one-to-one matching problems) and *roommate problems* (one-sided one-to-one matching problems) as in Gale and Shapley (1962). On the other hand, although our *model* also includes *college admission problems* (two-sided many-to-one matchings), our *results* are not applicable to those problems. This is because in those matching problems, preferences are assumed to have some special structures (such as "responsiveness" or "separability" (see Roth and Sotomayor (1990) for their definitions)), which are not compatible with our domain assumptions to be stated later.

2.2 Properties of mechanisms

For each $i \in N$, D_i is the (nonempty) set of the preferences relations admissible to i. Denote $D := D_1 \times D_2 \times \cdots \times D_n$.

[4] The term "hedonic" in this context was coined by Dréze and Greenberg (1980).

Let a list $(N, \mathscr{F}, D)$ be given. A **mechanism** f is a single-valued function $f : D \to X(\mathscr{F})$.[5] Let $i \in N$, and let $\succeq_i \in D_i$, $\succeq_{-i} \in D_{-i}$ and $\tilde{\succeq}_i \in D_i$. We say that i **manipulates** f **at** $(\succeq_{-i}, \succeq_i)$ **by** $\tilde{\succeq}_i$ if

$$f(\succeq_{-i}, \tilde{\succeq}_i) \succ_i f(\succeq_{-i}, \succeq_i). \tag{3}$$

We say that f is **strategy-proof** if there exists no i who manipulates f at any $(\succeq_{-i}, \succeq_i)$ by any $\tilde{\succeq}_i$.

Let S be a coalition, and let $\succeq_S \in D_S$, $\succeq_{-S} \in D_{-S}$ and $\tilde{\succeq}_S \in D_S$. We say that S **manipulates** f **at** $(\succeq_{-S}, \succeq_S)$ **by** $\tilde{\succeq}_S$ if

$$\forall i \in S,\ f(\succeq_{-S}, \tilde{\succeq}_S) \succeq_i f(\succeq_{-S}, \succeq_S), \text{and} \tag{4}$$

$$\exists i \in S,\ f(\succeq_{-S}, \tilde{\succeq}_S) \succ_i f(\succeq_{-S}, \succeq_S). \tag{5}$$

We say that f is **coalition strategy-proof** if there exists no S who manipulates f at any $(\succeq_{-S}, \succeq_S)$ by any $\tilde{\succeq}_S$.

We say that f is **individually rational** if for any $i \in N$, and any $\succeq \in D$,

$$f(\succeq)(i) \succeq_i \{i\}. \tag{6}$$

We introduce our main axiom. A mechanism is said to respect **coalitional unanimity** if it satisfies the property described as follows: *For any feasible coalition, if the coalition is top-ranked for every member of this coalition, then the mechanism recommends the formation of this coalition.* The formal description is as follows: We say that f respects **coalitional unanimity** if the following is satisfied: For any $\succeq \in D$, and any $S \in \mathscr{F}$, if

$$\forall i \in S,\ (\forall T \in \mathscr{F}(i),\ S \succeq_i T), \tag{7}$$

then

$$S \in f(\succeq) \tag{8}$$

Note that coalitional unanimity may not be well-defined when preferences include indifferences. Because with indifferences it may be the case that two coalitions which are both top-ranked by their members have a nonempty intersection. In this case, it is obviously impossible that these two coalitions present themselves in one partition so this axiom is contradictory. We will impose an assumption (Assumption 2 in Sec.3.1) on the preference domain to exclude such cases and make the axiom well-defined.

Coalitional unanimity is a generalization of the axiom "pairwise unanimity" introduced by Takagi and Serizawa (2010) in the context of *marriage problems* (two-sided one-to-one matching problems), a special case of the present model. Pairwise unanimity is identical with coalitional unanimity in this class of problems.[6]

[5] In this paper, we consider *direct* mechanisms only.

[6] Takagi and Serizawa (2010) defines an analogous axiom in the context of *college admission problems* (many-to-one matching problems) under the same name "pairwise unanimity." This axiom requires that the mechanism respects the unanimity by a college-student pair. We note that this axiom is *not* a special case of our coalitional unanimity. Because a college-student pair does not necessarily form a coalition by themselves since one college can be matched with many students.

It is worth mentioning that the concept of "top coalitions" of Banerjee, Konishi and Sönmez (2001) is related to coalitional unanimity. Let $\emptyset \neq S \subset V \subset N$. Then S is said to be a *top coalition* of V if for any $i \in S$ and any $T \subset V$ with $T \in \mathscr{F}(i)$, $S \succeq_i T$.[7] Given this definition, f satisfies coalitional unanimity if and only if for any $S \in \mathscr{F}$ and any $\succeq \in D$, if S is a top coalition of N, then $S \in f(\succeq)$.

2.3 Strict core stability

Let a problem $(N, \mathscr{F}, \succeq)$ be given. And let $x \in X(\mathscr{F})$, and $S \in \mathscr{F}$. Then we say that S **blocks** x if

$$\Big(\forall i \in S,\ S \succeq_i x(i)\Big)\ \&\ \Big(\exists j \in S : S \succ_j x(j)\Big). \tag{9}$$

A partition x is said to be **strictly core stable** if no feasible coalition blocks x.

The concept of strict core stability is a refinement of *core stability* which is defined by a weaker notion of blocking obtained by replacing the formula (9) in the above with the following:

$$\forall i \in S,\ S \succ_i x(i). \tag{10}$$

Note that these two core concepts are equivalent if preferences are all *strict*, i.e. $S \sim_i T$ implies $S = T$.

The **strict core stable correspondence** is the set-valued function which specifies the set of strict core stable partitions for each preference profile. Let us denote the strict core stable correspondence by $\mathscr{C}$.

3 Results

3.1 Domain assumptions

We define a class of domains by two assumptions. In the following, we fix the elements $(N, \mathscr{F}, D)$. Correspondingly, let us abbreviate $X(\mathscr{F})$ to X.

In the definitions in the sequel, we consider for each $i \in N$, a partition $\mathscr{P}_i$ of the set X. For any $x \in X$, let $\mathscr{P}_i(x)$ denote the cell of the partition $\mathscr{P}_i$ which contains x.

Assumption 1 *D is defined such that for any $i \in N$, there exists a partition $\mathscr{P}_i$ of X such that*

$$D_i = \Big\{ \succeq_i \mid \forall x, y \in X,\ \big(x \in \mathscr{P}_i(y) \Leftrightarrow x(i) \sim_i y(i)\big)\Big\}, \tag{11}$$

and

$$\forall x, y \in X,\ \exists i \in N : x \notin \mathscr{P}_i(y). \tag{12}$$

[7] Here we are giving one of possible definitions of top coalitions extended to the present setting. The original definition by Banerjee *et al.* is provided for the coalition formation model where any coalitions are feasible. Our definition reduces to theirs if we assume $\mathscr{F} = 2^N \setminus \{\emptyset\}$.

Note that by Assumption 1, for each individual i, the partition $\mathscr{P}_i$ constitutes the *indifference class.* That is, in any of his admissible preferences, any two outcomes are indifferent for individual i if, and only if, these two outcomes belong to the same cell of the partition $\mathscr{P}_i$.

Assumption 1 means that indifferences are incorporated in the domain D in the following way: For each individual i, the indifference class $\mathscr{P}_i$ is given *a priori*, and this individual ranks the outcomes as if he *strictly* ranks these *indifference sets*, i.e. the *cells* of $\mathscr{P}_i$. And the set of admissible preferences D_i for individual i *exactly* consists of *all* such rankings. Additionally, the profile of indifference classes $(\mathscr{P}_i)_{i \in N}$ is required to be such that for any two distinct outcomes, there is at least one individual who is not indifferent between the two outcomes.

Obviously, the indifference class $\mathscr{P}_i$ of X uniquely corresponds to the indifference class $\tilde{\mathscr{P}}_i$ of $\mathscr{F}(i)$ in the way that $x \in \mathscr{P}_i(y)$ if and only if $x(i) \in \tilde{\mathscr{P}}_i(y(i))$. Thus in the following, as long as no ambiguity arises, let us equate these two indifference classes and denote them by the same "$\mathscr{P}_i$."

The same domain condition as Assumption 1 is found in Takamiya (2007), which calls a domain satisfying this assumption an "essentially strict preference domain." Preferences which satisfy Assumption 1 naturally arise when some individual cares only about a part of the composition of the coalition which this individual belongs to.

Let a profile of indifference classes $(\mathscr{P}_i)_{i \in N}$ of X be given, and let D satisfy Assumption 1 with this $(\mathscr{P}_i)_{i \in N}$. Then we impose on $(\mathscr{P}_i)_{i \in N}$ the following assumption. This assumption is in order to make *coalitional unanimity* well-defined.

Assumption 2 *For any $i \in N$ and any $S,\ T \in \mathscr{F}$ with $S \neq T$,*

$$\big(S \in \mathscr{P}_i(T)\big) \Rightarrow \big(\exists j \in S \cap T : S \notin \mathscr{P}_j(T)\big). \tag{13}$$

Note that Assumption 2 implies the following fact which will be used in the proofs of our theorems.

$$\forall i \in N,\ \big\{\{i\}\big\} \in \mathscr{P}_i. \tag{14}$$

This means that no singleton is indifferent to any other coalition.

Example 1 We present examples for which Assumptions 1 and 2 are both satisfied.

(1) *The strict preference domain* is the domain such that each D_i consists *exactly* of those preferences which satisfy

$$\forall i \in N,\ \forall S,\ T \in \mathscr{F}(i),\ S \sim_i T \Rightarrow S = T. \tag{15}$$

If the domain is the strict preference domain, then Assumption 1 and 2 are satisfied: In this case, for each individual i, the indifference class $\mathscr{P}_i$ is set to be the finest partition $\{\{S\} \mid S \in \mathscr{F}(i)\}$. Then the two assumptions are satisfied trivially. The strict preference domain is common in the literature of matching and coalition formation.

(2) Since the strict preference domain rules out indifferences in preferences, it is worthwhile to have a natural example which incorporates indifferences. We give the following one.

- *The set of individuals* N is such that $N = N_1 \cup N_2 \cup \cdots \cup N_m$, where (i) $m \geq 1$, (ii) for any $t = 1, 2, \cdots m$, $N_t \neq \emptyset$, and (iii) for any $t, s = 1, 2, \cdots m$ with $t \neq s$, $N_t \cap N_s = \emptyset$. That is, N is partitioned into m subgroups.

- *The set of feasible coalitions* $\mathscr{F}$ is such that

$$\mathscr{F} = \Big\{ S \subset N \mid \forall t, |N_t \cap S| = 1 \Big\} \cup \Big\{ \{i\} \mid i \in N \Big\}. \tag{16}$$

  That is, each feasible coalition is (i) a coalition consisting of m components, and each t-th component is an individual from the set N_t, or (ii) a singleton.

- For each t and each $i \in N_t$, the indifference class $\mathscr{P}_i$ is set as follows:

$$\forall S,\ T \in \mathscr{F}(i),\ \Big(S \in \mathscr{P}_i(T) \Leftrightarrow S \cap N_{t+1} = T \cap N_{t+1} \Big), \tag{17}$$

  where $N_{m+1} = N_1$. That is, an individual in the t-th group N_t has "strict preferences" over the $(t+1)$-th group N_{t+1}.

It is easy to check that Assumptions 1 and 2 are satisfied. This example is interpreted as *m-sided* one-to-one matching problems with *"circular" preferences* (in the sense that individuals in N_t have preferences over N_{t+1}). In the case $m = 2$ these problems are *two-sided* matching problems, i.e. well-known *marriage problems* (Gale and Shapley, 1962). And the case $m = 3$ has been raised by Knuth (1976) and studied in Ng and Hirschberg (1991) and subsequent literature.

3.2 Statements of results

In all of our results and proofs in the sequel, it is postulated that the domain D satisfies Assumptions 1 and 2.

Theorem 1 *If the mechanism f is strategy-proof and respects coalitional unanimity, then for any $\succeq \in D$, either $\mathscr{C}(\succeq) = \{f(\succeq)\}$ or $\mathscr{C}(\succeq) = \emptyset$.*

Remark 1 From Theorem 1, an impossibility result follows: *If for some preference profile there are two or more strictly core stable partitions, then there is no mechanism which is strategy-proof and respects coalitional unanimity.* The case of the *marriage problem* (Gale and Shapley, 1962), which is dealt with by Takagi and Serizawa (2010), is one of such cases. Two other examples are the *roommate problem* (Gale and Shapley, 1962) and the *hedonic coalition formation problem* (Banerjee, Konishi and Sönmez 2001, Bogomolnaia and Jackson 2002, and Cechlárová and Romero-Medina 2001).

Next, we strengthen the requirement of *strategy-proofness* in the above theorem to that of *coalition strategy-proofness.* Then the *nonemptiness* of the strict core follows.

Theorem 2 *If the mechanism f is coalition strategy-proof and respects coalitional unanimity, then for any $\succeq \in D$, $\mathscr{C}(\succeq) = \{f(\succeq)\}$.*

The converse of Theorem 2 also holds true.

Theorem 3 *If for any* $\succeq \in D$, $\mathscr{C}(\succeq) = \{f(\succeq)\}$, *then the mechanism* f *is coalition strategy-proof and respects coalitional unanimity.*

Remark 2 Pápai (2004) gives a necessary and sufficient condition which *the set of feasible coalitions is to satisfy* for the strictly core stable partition to be unique for all *strict* preference profiles. This condition is called the "single-lapping property." Pápai also proves that *if the set of feasible coalitions satisfies the single-lapping property*, then the mechanism which chooses strictly core stable partitions is the only mechanism which is *strategy-proof*, *individually rational* and *Pareto efficient.* The significance of our Theorems 2 and 3 is that they provide another characterization of the single-valued strict core.

3.3 Proofs

First of all, we confirm that *coalitional unanimity is well-defined.*

Lemma 1 *For any* $\succeq \in D$, *and any* $S^1, S^2 \in \mathscr{F}$, *if for each* $k = 1, 2$, *and any* $i \in S^k$, $\big(\forall T \in \mathscr{F}(i),\ S^k \succeq_i T\big)$, *then* $S^1 \cap S^2 = \emptyset$.

Proof. Suppose the contrary, that is, there are some $S^1, S^2 \in \mathscr{F}$ for which every member of each coalition ranks that coalition at the top, and $S^1 \cap S^2 \neq \emptyset$. Then for each $i \in S^1 \cap S^2$, $S^1 \sim_i S^2$. This clearly contradicts Assumption 2. □

To proceed further, we need to introduce some notations.

- Let $\succeq$ be a preference profile. Let $x \in X$ and $i \in N$. Then let us define the preference relation " $\succeq_i^x$ " as follows: If $x(i) = \{i\}$, then $\succeq_i^x = \succeq_i$; otherwise, $\succeq_i^x$ is such that it satisfies the following three conditions:

  (i) $x(i) \succ_i^x \{i\}$,

  (ii) $\nexists S \in \mathscr{F}(i) \setminus \{\{i\}\},\ x(i) \succ_i^x S \succ_i^x \{i\}$,

  (iii) $\forall S,\ T \in \mathscr{F}(i) \setminus \{\{i\}\},\ (S \succeq_i T) \Leftrightarrow (S \succeq_i^x T)$.

  That is, $\succeq_i^x$ is the preference ranking which is obtained from $\succeq_i$ by moving $\{i\}$ to the position immediately below $x(i)$ and leaving other positions the same.

- Further, let us define the preference relation " $\succeq_i^{\uparrow(x)}$ " so as to satisfy the following two conditions:

  (i) $\forall S \in \mathscr{F}(i) \setminus \mathscr{P}_i(x(i)),\ x(i) \succ_i^{\uparrow(x)} S$,

  (ii) $\forall S,\ T \in \mathscr{F}(i) \setminus \mathscr{P}_i(x(i)),\ (S \succeq_i T) \Leftrightarrow (S \succeq_i^{\uparrow(x)} T)$.

  That is, $\succeq_i^{\uparrow(x)}$ is the preference ranking which is obtained from $\succeq_i$ by moving up the indifference set including $x(i)$ to the top and leaving other positions the same.

- Finally, let " $\succeq_i^{x\uparrow}$ " be a shorthand for $(\succeq_i^x)^{\uparrow(x)}$. That is, $\succeq_i^{x\uparrow}$ is obtained by performing the first and the second operations defined above sequentially on $\succeq_i$.

3.3.1 Proof of Theorem 1

Our proof starts with picking up $x \in \mathscr{C}(\succeq)$ for an arbitrary $\succeq \in D$ such that $\mathscr{C}(\succeq)$ is not empty, and ends with establishing $x = f(\succeq)$. This is done by induction in the two steps described as follows:

Step 1. Prove (i) and (ii).

(i) $f(\succeq^{x\uparrow}) = x$.

(ii) for any $i \in N$, $f(\succeq^{x\uparrow}_{-i}, \succeq_i) = x$.

Step 2. Prove that (iii) implies (iv) for any given m such that $0 < m < n$.

(iii) for any $T \subset N$ with $|T| = m$, $f(\succeq^{x\uparrow}_{-T}, \succeq_T) = x$.

(iv) for any $T \subset N$ with $|T| = m+1$, $f(\succeq^{x\uparrow}_{-T}, \succeq_T) = x$.

In the following, we complete Step 1 via a sequence of lemmas. In the sequel, until the end of the proof of this theorem, we assume that f *is strategy-proof and respects coalitional unanimity.*

Lemma 2 *f is individually rational.*

Proof. Suppose f is strategy-proof and respects coalitional unanimity, but is not individually rational. Then there are some $\succeq \in D$ and $i \in N$ for which

$$\{i\} \succ_i f(\succeq)(i) \tag{18}$$

Let $\succeq^\star_i$ be a preference relation of i such that

$$\forall S \in \mathscr{F}(i) \setminus \{\{i\}\},\ \{i\} \succ^\star_i S. \tag{19}$$

Then the coalitional unanimity of f implies

$$f(\succeq_{-i}, \succeq^\star_i)(i) = \{i\}. \tag{20}$$

This implies

$$f(\succeq_{-i}, \succeq^\star_i) \succ_i f(\succeq), \tag{21}$$

which means the violation of strategy-proofness. □

Lemma 3 *For any $x \in X$ and any $\succeq \in D$, $f(\succeq^{x\uparrow}) = x$.*

Proof. Immediate from the coalitional unanimity of f. □

Lemma 4 *For any $\succeq \in D$, if $x \in \mathscr{C}(\succeq)$, then for any $i \in N$, $f(\succeq^{x\uparrow}_{-i}, \succeq_i) \sim_i x$.*

Proof. Let us denote $f(\succeq_{-i}^{x\uparrow}, \succeq_i)(i)$ by S. Suppose the conclusion of the lemma does not hold, i.e., $x(i) \not\sim_i S$. Then either $x(i) \succ_i S$ or $S \succ_i x(i)$ is true.

Case 1. Suppose $x(i) \succ_i S$. That is,

$$f(\succeq_{-i}^{x\uparrow}, \succeq_i^{x\uparrow}) \succ_i f(\succeq_{-i}^{x\uparrow}, \succeq_i), \tag{22}$$

which violates strategy-proofness, a contradiction.

Case 2. Suppose $S \succ_i x(i)$. Note that $x(i) \succeq_i \{i\}$ since $x \in \mathscr{C}(\succeq)$. Then $S \succ_i x(i)$ implies $S \succ_i \{i\}$. Thus $S \neq \{i\}$.

Pick up $j \in S$ with $j \neq i$. Then by Lemma 2 and the fact (14), we have $S \succ_j^{x\uparrow} \{j\}$. Then this and the way $\succeq^{x\uparrow}$ is defined together imply

$$S \succeq_j x(j). \tag{23}$$

(Because if $x(j) \succ_j S$ on the contrary, then the definition of $\succeq_j^x$ implies $\{j\} \succ_j^x S$. And since $\succeq_j^{x\uparrow}$ means $(\succeq_j^x)^{\uparrow(x)}$, this implies also $\{j\} \succ_j^{x\uparrow} S$. This is a contradiction.) Note that this holds true for all $j \in S \setminus \{i\}$. Thus S blocks x under $\succeq$, which contradicts our supposition $x \in \mathscr{C}(\succeq)$. □

Lemma 5 *For any $\succeq \in D$ and any $T \subset N$ with $T \neq \emptyset$, if $x \in \mathscr{C}(\succeq)$ and for any $i \in T$, $f(\succeq_{-T}^{x\uparrow}, \succeq_T) \sim_i x$, then $f(\succeq_{-T}^{x\uparrow}, \succeq_T) = x$.*

Proof. Let $x \in \mathscr{C}(\succeq)$ and $y = f(\succeq_{-T}^{x\uparrow}, \succeq_T)$. And assume

$$\forall i \in T,\ y(i) \sim_i x(i). \tag{24}$$

Suppose that for some $i^\star \in N \setminus T$,

$$y(i^\star) \not\sim_{i^\star}^{x\uparrow} x(i^\star). \tag{25}$$

Note that $x(i^\star) \cap T \neq \emptyset$ because otherwise by the coalitional unanimity of f we would have $y(i^\star) = x(i^\star)$. Then since f is individually rational (Lemma 2),

$$\forall i \in x(i^\star) \setminus T,\ y(i) \succ_i^{x\uparrow} \{i\}. \tag{26}$$

This and the way $\succeq^{x\uparrow}$ is constructed together imply

$$\forall i \in x(i^\star) \setminus T,\ y(i) \succeq_i x(i). \tag{27}$$

Then (25) and (27) imply

$$y(i^\star) \succ_{i^\star} x(i^\star). \tag{28}$$

(24), (27) and (28) together imply that $y(i^\star)$ blocks x under $\succeq$. But this contradicts our supposition $x \in \mathscr{C}(\succeq)$. Therefore we have $\forall i \in N \setminus T,\ y(i) \sim_i^{x\uparrow} x(i)$. This and (24) imply $\forall i \in N,\ y(i) \sim_i x(i)$. Further, this and Assumption 1 imply that $y = x$, the desired conclusion. □

By Lemmas 4 and 5, we conclude that for any $\succeq \in D$, any $i \in N$ and any $x \in \mathscr{C}(\succeq)$,

$$f(\succeq^{x\uparrow}_{-i}, \succeq_i) = x. \tag{29}$$

This completes Step 1.

Now we proceed to Step 2, which as we have stated is to be done by induction. Let $x \in \mathscr{C}(\succeq)$. Our induction hypothesis is in the below.

Induction Hypothesis (IH) Let m be a natural number such that $0 < m < n$. Then for any $T \subset N$ with $|T| = m$,

$$f(\succeq^{x\uparrow}_{-T}, \succeq_T) = x. \tag{30}$$

Now under (IH), we prove that for any $T \subset N$ with $|T| = m + 1$, (30) is true.

Let $T \subset N$ with $|T| = m + 1$. Let $j \in T$. Then by (IH),

$$f(\succeq^{x\uparrow}_{-T}, \succeq_{T\setminus\{j\}}, \succeq^{x\uparrow}_j) = x. \tag{31}$$

Let us denote

$$f(\succeq^{x\uparrow}_{-T}, \succeq_{T\setminus\{j\}}, \succeq_j) = y. \tag{32}$$

In the following, we prove that $x(j) \sim_j y(j)$.

Case 1. First, note that $x(j) \not\succ_j y(j)$. Because otherwise j manipulates the outcome by reporting $\succeq^{x\uparrow}_j$ under the true preference $\succeq_j$, which violates the strategy-proofness of f.

Case 2. Next, suppose

$$y(j) \succ_j x(j). \tag{33}$$

Then since $x \in \mathscr{C}(\succeq)$, $x(j) \succeq_j \{j\}$. This implies $y(j) \succ_j \{j\}$. Thus $y(j)$ contains some member other than j.

Now we prove that the above supposition $y(j) \succ_j x(j)$ implies that there is some $k \in (y(j) \setminus \{j\}) \cap T$ for which $x(k) \succ_k y(j)$. Suppose the contrary. That is,

$$\forall k \in (y(j) \setminus \{j\}) \cap T,\ y(j) \succeq_k x(k). \tag{34}$$

Then the individual rationality of f (Lemma 2) implies for any $l \in (y(j)\setminus\{j\}) \cap (N\setminus T)$, we have

$$y(j) \succ^{x\uparrow}_l \{l\}. \tag{35}$$

This and the construction of $\succeq^{x\uparrow}$ together imply

$$\forall l \in (y(j) \setminus \{j\}) \cap (N \setminus T),\ y(j) \succeq_l x(l). \tag{36}$$

Then (33), (34) and (36) imply that $y(j)$ blocks x under $\succeq$. But this contradicts $x \in \mathscr{C}(\succeq)$. Thus we conclude that for some $k \in (y(j) \setminus \{j\}) \cap T$,

$$x(k) \succ_k y(j). \tag{37}$$

However, (IH) implies

$$f(\succeq_{-T}^{x\uparrow}, \succeq_{T\setminus\{j,k\}}, \succeq_j, \succeq_k^{x\uparrow}) = x. \tag{38}$$

Recall that we have defined

$$f(\succeq_{-T}^{x\uparrow}, \succeq_{T\setminus\{j,k\}}, \succeq_j, \succeq_k) = y. \tag{39}$$

Thus (37), (38) and (39) together imply that k manipulates the outcome by reporting $\succeq_k^{x\uparrow}$ under the true preference $\succeq_k$, which violates the strategy-proof of f. Thus we conclude $y(j) \not\succ_j x(j)$. This completes Case 2.

By the above arguments (Case 1 and Case 2) we have $x(j) \sim_j y(j)$. And since j has been taken arbitrarily from T, we have

$$\forall j \in T,\ f(\succeq_{-T}^{x\uparrow}, \succeq_T) \sim_j x. \tag{40}$$

Then by (40) and Lemma 5 together imply

$$f(\succeq_{-T}^{x\uparrow}, \succeq_T) = x. \tag{41}$$

This completes Step 2. □

3.3.2 Proof of Theorem 2

By Theorem 1, for any $\succeq \in D$, if $\mathscr{C}(\succeq) \neq \emptyset$, then $\mathscr{C}(\succeq) = \{f(\succeq)\}$. Thus to prove Theorem 2, it suffices to prove that for any $\succeq \in D$, $\mathscr{C}(\succeq) \neq \emptyset$ if f is coalition strategy-proof and respects coalitional unanimity.

Suppose that for some $\succeq \in D$, $\mathscr{C}(\succeq) = \emptyset$. Let us denote $f(\succeq) = x$. Then there is some $S \in \mathscr{F}$ which blocks x. That is,

$$\Big(\forall i \in S,\ S \succeq_i x(i)\Big)\ \&\ \Big(\exists j \in S,\ S \succ_j x(j)\Big) \tag{42}$$

For each $i \in S$, let $\succeq_i^\star$ be a preference relation of i such that

$$\forall T \in \mathscr{F}(i),\ S \succeq_i^\star T. \tag{43}$$

Then the coalitional unanimity of f implies for any $i \in S$

$$f(\succeq_{-S}, \succeq_S^\star)(i) = S. \tag{44}$$

Then (42) and (44) together imply

$$\Big(\forall i \in S,\ f(\succeq_{-S}, \succeq_S^\star) \succeq_i f(\succeq)\Big)\ \&\ \Big(\exists j \in S,\ f(\succeq_{-S}, \succeq_S^\star) \succ_j f(\succeq)\Big) \tag{45}$$

which means the violation of coalition strategy-proofness. □

3.3.3 Proof of Theorem 3

Let f be such that for any $\succeq \in D$, $\{f(\succeq)\} = \mathscr{C}(\succeq)$. It is immediate that such f respects coalitional unanimity. We prove such f is coalition strategy-proof. Suppose the contrary. Then for some $S \subset N$ with $S \neq \emptyset$, some $\succeq \in D$ and some $\succeq_S^\star \in D_S$,

$$\Big(\forall i \in S,\ f(\succeq_{-S}, \succeq_S^\star) \succeq_i f(\succeq)\Big) \ \&\ \Big(\exists j \in S,\ f(\succeq_{-S}, \succeq_S^\star) \succ_j f(\succeq)\Big) \tag{46}$$

Denote $x = f(\succeq)$ and $y = f(\succeq_{-S}, \succeq_S^\star)$.

By assumption (i.e. $\{f\} = \mathscr{C}$), we have $\{x\} = \mathscr{C}(\succeq)$. Let us consider $\succeq_S^{\uparrow(y)}$. Then clearly the change of preferences from $\succeq_S$ to $\succeq_S^{\uparrow(y)}$ does not alter the strict core stable partition because the relative position of x to y does not alter in this preference change. Thus we have

$$\{x\} = \mathscr{C}(\succeq_{-S}, \succeq_S^{\uparrow(y)}). \tag{47}$$

Similarly by assumption, we have $\{y\} = \mathscr{C}(\succeq_{-S}, \succeq_S^\star)$. Then again evidently the change of preferences $\succeq_S^\star$ to $\succeq_S^{\uparrow(y)}$ does not alter the strict core, that is,

$$\{y\} = \mathscr{C}(\succeq_{-S}, \succeq_S^{\uparrow(y)}). \tag{48}$$

(47) and (48) imply $x = y$, which contradicts our supposition (46). □

4 Concluding Remarks

Here we conclude our paper mentioning two directions of extension or variation. The first direction is as follows. Our results have been proved under some domain assumptions which are not compatible with special preference structures commonly assumed in *two-sided many-to-one matching problems* (college admission problems) such as "responsiveness" and "separability." (See Roth and Sotomayor (1990) for their definitions.) So one of our interests in the future research is to see whether similar results hold for some domains which permit such preference structures. Takagi and Serizawa (2010) have already shown that in the context of many-to-one matching problems an impossibility theorem similar to that in the marriage problem holds true. Rodríguez-Álvarez (2009) considers a coalition formation model on the separable preference domain and other domains, and characterizes a certain coalition formation rule by strategy-proofness and additional properties other than coalitional unanimity. His coalition formation model is different from ours in the point that it does not explicitly incorporate feasibility constraints on coalitions. However, the characterized rule is defined using the single-lapping property (Pápai, 2004) and closely related to the core stability in the presence of feasibility constraints.

The second direction is to consider the weakening of incentive requirements. In this direction, we announce that Takamiya (2008) deals with the case of *Nash implementation.* There a version of *coalitional unanimity* for *social choice correspondences* (i.e. multi-valued rule) is defined, and the following result is shown: Under some regular assumptions, *a social choice correspondence is Nash implementable and respects* *coalitional unanimity if and only if the correspondence is the strictly core stable correspondence.*

Acknowledgments I am very thankful to the editors in charge and two anonymous referees for their various suggestions which have improved this paper. I thank Nobuo Akai, Yuji Fujinaka, Junichiro Ishida, Kazuhiko Hashimoto, Kentaro Hatsumi, Yoshio Kamijo, Kazuharu Kiyono, Takumi Kongo, Koichi Suga, Shohei Takagi, Katsuya Takii, Shiro Yabushita and other participants in the seminars at Waseda University and Osaka University for various comments and discussions. This research is partially supported by Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan, No.21730155 and No.23730188. All errors are my own responsibility.

References

[1] Banerjee S, Konishi H, and Sönmez T (2001) Core in a simple coalition formation game. *Social Choice and Welfare* 18: 135–53.

[2] Bogomolnaia A and Jackson MO (2002) The stability of hedonic coalition structures. *Games and Economic Behavior* 38: 201–30.

[3] Cechlárová K and Romero-Medina A (2001) Stability in coalition formation games. *International Journal of Game Theory* 29: 487–94.

[4] Dréze J and Greenberg J (1980) Hedonic coalitions: optimality and stability. *Econometrica* 48: 987–1003.

[5] Gale D and Shapley L (1962) College admissions and the stability of marriage. *American Mathematical Monthly* 69: 9–15.

[6] Knuth D (1976) *Mariages stables et leurs relations avec d'autre problèmes combinatoires,* Les Presses de l'université de Montréal.

[7] Ng C and Hirschberg D (1991) Three-dimensional stable matching problems. *SIAM Journal of Discrete Mathematics* 4: 245–52.

[8] Pápai S (2004) Unique stability in simple coalition formation games. *Games and Economic Behavior* 48: 337–54.

[9] Rodríguez-Álvarez C (2009) Strategy-proof coalition formation. *International Journal of Game Theory* 38: 431–52.

[10] Roth A and Sotomayor M (1990) *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*, Econometric Society Monograph, vol. 18, Cambridge Univ Press.

[11] Sönmez T (1999) Strategy-proofness and essentially single-valued cores. *Econometrica* 67: 677–89.

[12] Takagi S and Serizawa S (2010) An impossibility theorem in matching problems. *Social Choice and Welfare* 35: 245–66.

[13] Takamiya K (2007) Domains of social choice functions on which coalition strategy-proofness and Maskin monotonicity are equivalent. *Economics Letters* 95: 348–54.

[14] Takamiya K (2008) Maskin monotonic coalition formation rules respecting group rights. Mimeo. Niigata University.

[15] Toda M (2006) Monotonicity and consistency in matching markets, *International Journal of Game Theory* 34: 13–31.