Game Theory 2 - Strategic Social Choice Lectures

1 Introduction

We discovered the implications of incentive compatibility when transfers are allowed with quasi-linear utilities. In the following lectures, we will not allow transfers and see the implications of incentive compatibility. Indeed, there are many practical settings where transfers are not allowed. The most common scenario is various voting environments. Consider the following examples: agents are voting to elect a candidate; students are assigned rooms in a hostel; citizens are voting to locate a facility etc.

The usual definitions and fundamental results like revelation principle continue to hold in these settings. The private information of the agent becomes an ordinal ranking of the alternatives. One can enrich this model by using cardinal information, but we restrict attention to the ordinal model. Before we can formally describe the model, we need some basic definitions.

1.1 Basic Definitions

Let $A = \{a, b, c, \ldots, x, y, z, \ldots\}$ be a finite set of alternatives. Let $N = \{1, \ldots, n\}$ be a finite set of agents. Every agent has a preference over alternatives. The preference relation of agent i over alternatives is denoted by R_i, where aR_ib denotes that preference a is at least as good as b for agent i in preference relation R_i. The preferences can be represented in many ways. Here are two plausible ways of representing the preferences.

1. Ordering: A preference relation R_i of agent i is called an **ordering** if it satisfies the following properties:

   - Completeness: For all $a, b \in A$ either aR_ib or bR_ia.
   - Reflexivity: For all $a \in A$, aR_ia.
   - Transitivity: For all $a, b, c \in A$, $\big[aR_ib, bR_ic\big] \Rightarrow \big[aR_ic\big]$.

   We will denote the set of all orderings over A as $\mathcal{R}$. By definition, an ordering gives ordered pairs of A.

2. Binary Relation: A preference relation R_i of agent i is called a **binary relation** if it satisfies completeness and reflexivity. An ordering is a transitive binary relation.

Let Q_i be a binary relation. The **symmetric component** of Q_i is denoted by $\bar{Q}_i$, and is defined as: for all $a, b \in A$, $a\bar{Q}_ib$ if and only if aQ_ib and bQ_ia. The asymmetric component of Q_i is denoted by $\hat{Q}_i$, defined as: for all $a, b \in A$, $a\hat{Q}_ib$ if and only if aQ_ib but $\neg(bQ_ia)$, where $\neg$ is the "not" logic symbol. Informally, $\hat{Q}_i$ is the strict part of Q_i, whereas $\bar{Q}_i$ is the weak part of Q_i. We will also refer to the symmetric component of a preference relation R_i as I_i and asymmetric component as P_i.

PROPOSITION 1 *Let R_i be an ordering. Then P_i and I_i are transitive. Conversely, if R_i is a binary relation such that P_i and I_i are transitive, then R_i is an ordering.*

Proof: Consider $a, b, c \in A$ such that aP_ib and bP_ic. Assume for contradiction $\neg(aP_ic)$. Since R_i is an ordering, it is complete. Hence, aR_ic or cR_ia. Since $\neg(aP_ic)$, we get cR_ia. But aP_ib. By transitivity of R_i, we get cR_ib. This contradicts bP_ic.

Similarly, assume aI_ib and bI_ic. This implies, aR_ib and bR_ic. Also, bR_ia and cR_ib. Due to transitivity, we get aR_ic and cR_ia. This implies that aI_ic.

The converse part is equally easy. ■

DEFINITION 1 *An ordering R_i is* **anti-symmetric** *if for all $a, b \in A$ aR_ib and bR_ia implies $a = b$ (i.e., no indifference). An anti-symmetric ordering is also called a* **linear ordering**.

2 The General Impossibility in Strategic Voting

Let A be a finite set of alternatives with $|A| = m$. Let N be a finite set of individuals or agents or voters with $|N| = n$. Every agent has a preference over the set of alternatives. Let P_i denote the preference ordering of agent i. We assume that the preference ordering of every agent is a **linear ordering**. Given a preference ordering P_i we say aP_ib if and only if a is strictly preferred to b under P_i. Given a preference ordering P_i of agent i, the top ranked element of this ordering is denoted by $P_i(1)$, the second ranked element by $P_i(2)$, and so on. Let $\mathcal{P}$ be the set of all strict preference orderings over A. A profile of preference orderings (or simply a preference profile) is denoted as $P \equiv (P_1, \ldots, P_n)$. So, $\mathcal{P}^n$ is the set of all preference profiles. A **social choice function (SCF)** is a mapping $f : \mathcal{P}^n \to A$. Note that this definition of a social choice function implicitly assumes that all possible profiles of linear orderings are permissible. This is known as the **unrestricted domain** assumption in the strategic voting (social choice) literature. Later, we will study some interesting settings where domain of the social choice function is restricted.

Every agent knows his own preference ordering but does not know the preference ordering of other agents, and the mechanism designer (planner) does not know the preference orderings of agents. This is a very common situation in many voting scenarios: electing a candidate among a set of candidates, selecting a project among a finite set of projects for a company, selecting a public facility location among a finite set of possible locations, etc. Monetary transfers are precluded in these settings. This is major difference from the standard quasi-linear utility environments with money. The objective of this section is to find out which social choice functions are implementable in dominant strategies in such strategic voting scenarios.

We first describe several desirable properties of an SCF. The first property is a range condition on the social choice function.

DEFINITION **2** *A social choice function f is* **onto** *if for every $a \in A$ there exists a profile of preferences $P \in \mathcal{P}^n$ such that $f(P) = a$.*

The next property is a monotonicity property.

DEFINITION **3** *A social choice function f is* **monotone** *if for any two profiles P and P' with $f(P) = a$ such that for all $b \neq a$, we have $aP'_i b$ if $aP_i b$ for all $i \in N$, then $f(P') = a$.*

Note that in the definition of monotonicity when we go from preference profile P to P' with $f(P) = a$, whatever every agent was preferring to a in P continues to prefer it in P' also, but other relations may change. For example, the following is a valid P and P' in the definition of monotonicity with $f(P) = a$ (see Table 1).

P_1	P_2	P_3	P'_1	P'_2	P'_3
a	b	c	a	a	a
b	a	a	b	c	c
c	c	b	c	b	b

Table 1: Two valid profiles for monotonicity

The next property is an efficiency property.

DEFINITION **4** *A social choice function f is* **efficient** [1] *if for every profile of preferences P and every $b \in A$, if there exists $a \neq b$ such that $aP_i b$ for all $i \in N$, then $f(P) \neq b$.*

Efficiency requires that if an alternative a is preferred over b by all the agents, then b cannot be chosen as an outcome. The next property requires to respect unanimity.

DEFINITION **5** *A social choice function f is* **unanimous** *if for every preference profile $P \equiv (P_1, \ldots, P_n)$ with $P_1(1) = P_2(1) = \ldots = P_n(1) = a$ we have $f(P) = a$.*

Note that this version of unanimity is a stronger version than requiring that if the *preference ordering* of all agents is the same, then the top ranked alternative must be chosen. This definition requires only the top to be the same, but other alternatives can be ranked differently by different agents.

Finally, we define the strategic property of a social choice function.

[1] Such a social choice function is also called Pareto optimal or Pareto efficient or ex-post efficient.

Definition 6 *A social choice function f is* **manipulable by agent i at profile** $P \equiv (P_i, P_{-i})$ [2] **by profile** (P'_i, P_{-i}) *if* $f(P'_i, P_{-i}) P_i f(P)$. *A social choice function f is* **strategy-proof** *if it is not manipulable by any agent i at any profile P by any other profile* (P'_i, P_{-i}).

This notion of strategy-proofness is the dominant strategy requirement since no manipulation is possible for every agent for every possible profile of other agents.

2.1 Examples of Social Choice Functions

We give three examples of social choice functions.

- Constant SCF. A social choice function f^c is a constant SCF if there is some alternative $a \in A$ such that for every preference profile P, we have $f^c(P) = a$. This SCF is strategy-proof but not onto.

- Dictatorship SCF. A social choice function f^d is a **dictatorship** if there exists an agent i, called the dictator, such that for every preference profile P, we have $f^d(P) = P_i(1)$. Dictatorship is strategy-proof and onto. Moreover, as we will see later, they are also monotone, efficient, and unanimous.

- Plurality SCF (with fixed tie-breaking). Let $\succ$ be a linear ordering over alternatives A. For every preference profile P and every alternative $a \in A$, define the score of a in P as $s(a, P) = |\{i \in N : P_i(1) = a\}|$. Define $\tau(P) = \{a \in A : s(a, P) \geq s(b, P)\ \forall\ b \in A\}$ for every preference profile P, and note that $\tau(P)$ is non-empty.

  A social choice function f^p is called a plurality SCF with tie-breaking according to $\succ$ if for every preference profile P, $f^p(P) = a$, where $a \in \tau(P)$ and $a \succ b$ for all $b \in \tau(P) \setminus \{a\}$.

  Though the plurality SCF is onto, it is not strategy-proof. To see this, consider an example with three agents $\{1, 2, 3\}$ and three alternatives $\{a, b, c\}$. Let $\succ$ be defined as: $a \succ b \succ c$. Consider two preference profiles shown in Table 2. We note first that $f(P) = a$ and $f(P') = b$. Since bP_3a, agent 3 can manipulate at P by P'.

P_1	P_2	P_3	$P'_1 = P_1$	$P'_2 = P_2$	P'_3
a	b	c	a	b	b
b	c	b	b	c	a
c	a	a	c	a	c

Table 2: Plurality SCF is manipulable.

[2] We use the standard notation P_{-i} to denote the preference profile of agents other than agent i.

- Borda SCF (with fixed tie-breaking). The tie-breaking in this SCF is defined similar to Plurality SCF. Let $\succ$ be a linear ordering over alternatives A. Fix a preference profile P. For every alternative $a \in A$, the *rank* of a in P for agent i is given by $r_i(a, P) = k$, where $P_i(k) = a$. From this, the score of alternative a in preference profile P is computed as $s(a, P) = \sum_{i \in N}[|A| - r_i(a, P)]$. Define for every preference profile P, $\tau(P) = \{a \in A : s(a, P) \geq s(b, P)\ \forall b \in A\}$. A social choice function f^b is called a Borda SCF with tie-breaking according to $\succ$ if for every preference profile P, $f^b(P) = a$ where $a \in \tau(P)$ and $a \succ b$ for all $b \in \tau(P) \setminus \{a\}$.

  Like the Plurality SCF, the Borda SCF is onto but manipulable. To see this, consider an example with three agents $\{1, 2, 3\}$ and three alternatives $\{a, b, c\}$. Let $\succ$ be defined as: $c \succ b \succ a$. Consider two preference profiles shown in Table 3. We note first that $f(P) = b$ and $f(P') = c$. Since cP_1b, agent 1 can manipulate at P by P'.

P_1	P_2	P_3	P'_1	$P'_2 = P_2$	$P'_3 = P_3$
a	b	b	c	b	b
c	c	c	a	c	c
b	a	a	b	a	a

Table 3: Borda SCF is manipulable.

2.2 Implications of Properties

We now examine the implications of these properties. We start out with a simple characterization of strategy-proof social choice functions.

Theorem 1 *A social choice function $f : \mathcal{P}^n \to A$ is strategy-proof if and only if it is monotone.*

Proof: Consider social choice function $f : \mathcal{P}^n \to A$ which is strategy-proof. Consider two preference profiles P and P' such that $f(P) = a$, and for all $b \neq a$ and for all $i \in N$, we have aP'_ib if aP_ib. We define $(n-1)$ new preference profiles. Define P^1 as follows: $P^1_1 = P'_1$ and $P^1_i = P_i$ for all $i > 1$. Define P^k for $k \in \{1, \ldots, n-1\}$ as $P^k_i = P'_i$ if $i \leq k$ and $P^k_i = P_i$ if $i > k$. Set $P^0 = P$ and $P^n = P'$. Note that if we pick two preference profiles P^k and P^{k+1} for any $k \in \{0, \ldots, n-1\}$, then preference of all agents other than agent $(k+1)$ are same in P^k and P^{k+1}, and preference of agent $(k+1)$ is changing from P_{k+1} in P^k to P'_{k+1} in P^{k+1}.

We will show that $f(P^k) = a$ for all $k \in \{0, \ldots, n\}$. We know that $f(P^0) = f(P) = a$, and consider $k = 1$. Assume for contradiction $f(P^1) = b \neq a$. If bP_1a, then agent 1 can manipulate at P^0 by P^1. If aP_1b, then aP^1_1b, and agent 1 can manipulate at P^1 by P^0. This is a contradiction since f is strategy-proof.

We can repeat this argument by assuming that $f(P^q) = a$ for all $q \leq k < n$, and showing that $f(P^{k+1}) = a$. Assume for contradiction $f(P^{k+1}) = b \neq a$. If $bP_{k+1}a$, then agent $(k+1)$ can manipulate at P^k by P^{k+1}. If $aP_{k+1}b$ then $aP'_{k+1}b$. This means agent $(k+1)$ can manipulate at P^{k+1} by P^k.

Hence, by induction, $f(P^n) = f(P') = a$, and f is monotone.

Now suppose, $f : \mathcal{P}^n \rightarrow A$ is a monotone social choice function. Assume for contradiction that f is not strategy-proof. In particular, agent i can manipulate at preference profile P by preference profile $P' \equiv (P'_i, P_{-i})$. Suppose $f(P) = a$ and $f(P') = b$, and by assumption bP_ia. Consider a preference profile $P'' \equiv (P''_i, P_{-i})$, where P''_i is any preference ordering satisfying $P''_i(1) = b$ and $P''_i(2) = a$. By monotonicity, $f(P'') = f(P') = b$ and $f(P'') = f(P) = a$, which is a contradiction. ■

Theorem 1 is a strong result. The necessity of monotonicity is true in any domain - even if a subset of all possible preference profiles are permissible.

We now explore implications of other properties.

Lemma **1** *If an SCF f is monotone and onto then it is efficient.*

Proof: Consider $a, b \in A$ and a preference profile P such that aP_ib for all $i \in N$. Assume for contradiction $f(P) = b$. Since f is onto, there exists a preference profile P' such that $f(P') = a$. We construct another preference profile $P'' \equiv (P''_1, \ldots, P''_n)$ as follows. For all $i \in N$, let $P''_i(1) = a$, $P''_i(2) = b$, and $P''_i(j)$ for $j > 2$ can be set to anything. Since f is monotone, $f(P'') = f(P) = b$, and also, $f(P'') = f(P') = a$. This is a contradiction. ■

Lemma **2** *If an SCF f is efficient then it is unanimous.*

Proof: Consider a preference profile $P \equiv (P_1, \ldots, P_n)$ with $P_1(1) = P_2(1) = \ldots = P_n(1) = a$. Consider any $b \neq a$. By definition, aP_ib for all $i \in N$. By efficiency, $f(P) \neq b$. Hence, $f(P) = a$. ■

Lemma **3** *If a social choice function is unanimous then it is onto.*

Proof: Take any alternative $a \in A$ and a social choice function f. Consider a profile P such that $P_i(1) = a$ for all $i \in N$. Then $f(P) = a$ by unanimity. So, f is onto. ■

We can summarize these results in the following proposition.

Proposition **2** *Suppose $f : \mathcal{P} \rightarrow A$ is a strategy-proof social choice function. Then, f is onto if and only if it is efficient if and only if it is unanimous.*

Proof: Suppose f is strategy-proof. By Theorem 1, it is monotone. Lemmas 1, 2, and 3 establish the result. ■

2.3 The Gibbard-Satterthwaite Theorem

Theorem **2 (Gibbard-Satterthwaite Theorem)** *Suppose $|A| \geq 3$. A social choice function $f : \mathcal{P}^n \to A$ is onto and strategy-proof if and only if it is a dictatorship.*

We do the proof using induction on number of agents. We first analyze the case when $n = 2$.

Lemma **4** *Suppose $|A| \geq 3$ and $N = \{1, 2\}$. Suppose f is an onto and strategy-proof social choice function. Then for every preference profile P, $f(P) \in \{P_1(1), P_2(1)\}$.*

Proof:

P_1	P_2	P_1	P'_2	P'_1	P'_2	P'_1	P_2
a	b	a	b	a	b	a	b
$\cdot$	$\cdot$	$\cdot$	a	b	a	b	$\cdot$
$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$

Table 4: Preference profiles required in proof of Lemma 4.

Fix a preference profile $P = (P_1, P_2)$. If $P_1(1) = P_2(1)$, the claim is due to unanimity (Proposition 2). Else, let $P_1(1) = a$ and $P_2(1) = b$, where $a \neq b$. Assume for contradiction $f(P) = c \notin \{a, b\}$. We will use the preference profiles shown in Table 4.

Consider a preference ordering P'_2 for agent 2 where $P'_2(1) = b$, $P'_2(2) = a$, and the remaining ordering can be anything. By efficiency, $f(P_1, P'_2) \in \{a, b\}$. Further $f(P_1, P'_2) \neq b$ since agent 2 can then manipulate at P by (P_1, P'_2). So, $f(P_1, P'_2) = a$.

Now, consider a preference ordering P'_1 for agent 1 where $P'_1(1) = a$, $P'_1(2) = b$, and the remaining ordering can be anything. Using an analogous argument, we can show that $f(P'_1, P_2) = b$. Now, consider the preference profile (P'_1, P'_2). By monotonicity (implied by strategy-proofness - Theorem 1) , $f(P'_1, P'_2) = f(P_1, P'_2) = a$ and $f(P'_1, P'_2) = f(P'_1, P_2) = b$. This is a contradiction. ■

Lemma **5** *Suppose $|A| \geq 3$ and $N = \{1, 2\}$. Suppose f is onto and strategy-proof social choice function. Consider a profile P such that $P_1(1) = a \neq b = P_2(1)$. If $f(P) = a$, then for all preference profiles $P' = (P'_1, P'_2)$ with $P'_1(1) = c$ and $P'_2(1) = d$, we have $f(P') = P'_1(1) = c$.*

Proof: We can assume that $c \neq d$, since the claim is true due to unanimity when $c = d$. We do the proof for different possible cases.

Case 1: $c = a$, $d = b$. From Lemma 4, $f(P') \in \{a, b\}$. Assume for contradiction $f(P') = b$ (i.e., agent 2's top is chosen). Consider a preference profile $\hat{P} \equiv (\hat{P}_1, \hat{P}_2)$ such that $\hat{P}_1(1) = a, \hat{P}_1(2) = b$ and $\hat{P}_2(1) = b, \hat{P}_2(2) = a$ (See Table 5). By monotonicity, $f(\hat{P}) = f(P') = f(P)$, which is a contradiction.

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	a	b	a	b
·	·	·	·	b	a
·	·	·	·	·	·

Table 5: Preference profiles required in Case 1.

Case 2: $c \neq a$, $d = b$. Consider any profile $\hat{P} = (\hat{P}_1, \hat{P}_2)$, where $\hat{P}_1(1) = c \neq a$, $\hat{P}_1(2) = a$, and $\hat{P}_2(1) = b$ (See Table 6).

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$
a	b	$c \neq a$	$d = b$	c	b
·	·	·	·	a	·
·	·	·	·	·	·

Table 6: Preference profiles required in Case 2.

By Lemma 4, $f(\hat{P}) \in \{b, c\}$. Suppose $f(\hat{P}) = b$. Then, agent 1 can manipulate by reporting any preference ordering where his top is a, and this will lead to a as the outcome (Case 1). Hence, $f(\hat{P}) = c = \hat{P}_1(1)$. Using Lemma 5, $f(P') = c$.

Case 3: $c \notin \{a, b\}$, $d \neq b$ [3]. Consider a preference profile $\hat{P}$ such that $\hat{P}_1(1) = c, \hat{P}_2(1) = d, \hat{P}_2(2) = b$ (See Table 7).

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_1$	$\hat{P}'_2$
a	b	$c \neq \{a, b\}$	$d \neq b$	c	d	c	b
·	·	·	·	·	b	·	d
·	·	·	·	·	·	·	·

Table 7: Preference profiles required in Case 3.

By Lemma 4, $f(\hat{P}) \in \{c, d\}$. Suppose $f(\hat{P}) = d$. Then consider $\hat{P}'_2$ such that $\hat{P}'_2(1) = b$ and $\hat{P}'_2(2) = d$ (See Table 7). By Case 2, $f(\hat{P}_1, \hat{P}'_2) = c$. Since $d\hat{P}'_2 c$, agent 2 will manipulate at $(\hat{P}_1, \hat{P}'_2)$ by $\hat{P}$. Hence, $f(\hat{P}) = c$. Using Lemma 5, $f(P') = c$.

[3] This case actually covers two cases: one where $d = a$ and the other where $d \notin \{a, b\}$.

CASE 4: $c = a$, $d \neq b$. By Lemma 4, $f(P') \in \{a, d\}$. Assume for contradiction $f(P') = d$. Consider a preference ordering $\hat{P}_2$ such that $\hat{P}_2(1) = b$ and $\hat{P}_2(2) = d$ (See Table 8).

P_1	P_2	P'_1	P'_2	P'_1	$\hat{P}_2$
a	b	$c = a$	$d \neq b$	a	b
.	.	.	.	.	d
.	.	.	.	.	.

Table 8: Preference profiles required in Case 4.

Now, by Case 1, $f(P'_1, \hat{P}_2) = a$. But $d\hat{P}_2 a$. Hence, agent 2 can manipulate at $(P'_1, \hat{P}_2)$ by P', which is a contradiction. Using Lemma 5, $f(P') = a$.

CASE 5: $c = b$, $d \neq a$. By Lemma 4, $f(P') \in \{b, d\}$. Assume for contradiction $f(P') = d$. Consider a preference ordering $\hat{P}_1$ such that $\hat{P}_1(1) = b$ and $\hat{P}_1(2) = a$, and $\hat{P}_2$ such that $\hat{P}_2(1) = d$. Consider another preference ordering $\hat{P'}_1$ such that $\hat{P'}_1(1) = a$ (See Table 9).

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	$\hat{P}_2$	$\hat{P'}_1$	$\hat{P}_2$
a	b	$c = b$	$d \neq a$	b	d	a	d
.	.	.	.	a	.	.	.
.	.	.	.	.	.	.	.

Table 9: Preference profiles required in Case 5.

By Cases 1 and 4, $f(\hat{P'}_1, \hat{P}_2) = a$. But $a\hat{P}_1 d$. So, agent 1 can manipulate $(\hat{P}_1, \hat{P}_2)$ by $(\hat{P'}_1, \hat{P}_2)$. This is a contradiction. Using Lemma 5, $f(P') = b$

CASE 6: $c = b$, $d = a$. Since there are at least three alternatives, consider $x \notin \{a, b\}$. Consider a preference ordering $\hat{P}_1$ such that $\hat{P}_1(1) = b$ and $\hat{P}_1(2) = x$ (See Table 10).

P_1	P_2	P'_1	P'_2	$\hat{P}_1$	P'_2	$\hat{P'}_1$	P'_2
a	b	$c = b$	$d = a$	b	a	x	a
.	.	.	.	x	.	.	.
.	.	.	.	.	.	.	.

Table 10: Preference profiles required in Case 6.

By Lemma 4, $f(\hat{P}_1, P'_2) \in \{b, a\}$. Assume for contradiction $f(\hat{P}_1, P'_2) = a$. Consider a preference ordering $\hat{P'}_1$ such that $\hat{P'}_1(1) = x$ (See Table 10). By Case 3, $f(\hat{P'}_1, P'_2) = x$. But $x\hat{P}_1 a$. Hence, agent 1 can manipulate $(\hat{P}_1, P'_2)$ by $(\hat{P'}_1, P'_2)$. This is a contradiction. Hence, $f(\hat{P}_1, P'_2) = b$. By Lemma 5, $f(P') = b$. ■

Proposition **3** *Suppose $|A| \geq 3$ and $n = 2$. A social choice function is onto and strategy-proof if and only if it is dictatorship.*

Proof: This follows directly from Lemmas 5 and unanimity (implied by onto and strategy-proofness - Proposition 2). ■

Once we have the theorem for $n = 2$ case, we can apply induction on the number of agents. In particular, we prove the following proposition.

Proposition **4** *Let $n \geq 3$. Consider the following statements.*

(a) For all positive integer $k < n$, we have if $f : \mathcal{P}^k \to A$ is onto and strategy-proof, then f is dictatorial.

(b) If $f : \mathcal{P}^n \to A$ is onto and strategy-proof, then f is dictatorial.

Statement (a) implies statement (b).

Proof: Suppose statement (a) holds. Let $f : \mathcal{P}^n \to A$ be an onto and strategy-proof social choice function. We construct another social choice function $g : \mathcal{P}^{n-1} \to A$ from f by merging agents 1 and 2 as one agent. In particular, $g(P_1, P_3, P_4, \ldots, P_n) = f(P_1, P_1, P_3, P_4, \ldots, P_n)$ for all preference profiles $(P_1, P_3, P_4, \ldots, P_n)$. So agents 1 and 2 are "coalesced" in social choice function g, and will be referred to as agent 1 in SCF g.

We do the proof in two steps. In the first step, we show that g is onto and strategy-proof. We complete the proof in the second step, i.e., show that f is dictatorship.

Step 1: It is clear that agents 3 through n cannot manipulate in g (if they can manipulate in g, they can also manipulate in f, which is a contradiction). Consider an arbitrary preference profile of $n - 1$ agents $(P_1, P_3, P_4, \ldots, P_n)$. Suppose

$$f(P_1, P_1, P_3, P_4, \ldots, P_n) = g(P_1, P_3, P_4, \ldots, P_n) = a.$$

Consider any arbitrary preference ordering $\bar{P}_1$ of agent 1. Let

$$f(P_1, \bar{P}_1, P_3, P_4, \ldots, P_n) = b.$$

Let

$$f(\bar{P}_1, \bar{P}_1, P_3, P_4, \ldots, P_n) = g(\bar{P}_1, P_3, P_4, \ldots, P_n) = c.$$

If $a = c$, then agent 1 cannot manipulate g at $(P_1, P_3, P_4, \ldots, P_n)$ by $(\bar{P}_1, P_3, P_4, \ldots, P_n)$. So, assume $a \neq c$. Suppose $a = b \neq c$. Then, agent 1 cannot manipulate f at $(P_1, \bar{P}_1, P_3, P_4, \ldots, P_n)$

by $(\bar{P}_1, \bar{P}_1, P_3, P_4, \ldots, P_n)$. So, $a = bP_1c$. Hence, agent 1 cannot manipulate g at $(P_1, P_3, P_4, \ldots, P_n)$ by $(\bar{P}_1, P_3, P_4, \ldots, P_n)$. A similar logic works for the case when $b = c$.

Now, assume that a, b, and c are distinct. Since f is strategy-proof, agent 2 cannot manipulate f at $(P_1, P_1, P_3, P_4, \ldots, P_n)$ by $(P_1, \bar{P}_1, P_3, P_4, \ldots, P_n)$. So, aP_1b. Similarly, agent 1 cannot manipulate f at $(P_1, \bar{P}_1 P_3, P_4, \ldots, P_n)$ by $(\bar{P}_1, \bar{P}_1, P_3, P_4, \ldots, P_n)$. So, bP_1c. By transitivity, aP_1c. Hence, agent 1 cannot manipulate g at $(P_1, P_3, P_4, \ldots, P_n)$ by $(\bar{P}_1, P_3, P_4, \ldots, P_n)$. This shows that g is strategy-proof.

It is straightforward to show that if f is onto, then g is onto (follows from unanimity of f).

Step 2: By our induction hypothesis, g is dictatorship. Suppose j is the dictator. There are two cases to consider.

Case A: Suppose $j \in \{3, 4, \ldots, n\}$ is the dictator in g. We claim that j is also the dictator in f. Assume for contradiction that there is a preference profile $P \equiv (P_1, P_2, \ldots, P_n)$ such that

$$f(P) = b \text{ and } P_j(1) = a \neq b.$$

Since g is dictatorship, we get

$$f(P_1, P_1, P_3, P_4, \ldots, P_n) = g(P_1, P_3, P_4, \ldots, P_n) = a,$$
$$f(P_2, P_2, P_3, P_4, \ldots, P_n) = g(P_2, P_3, P_4, \ldots, P_n) = a.$$

We get bP_1a, since f is strategy-proof, and agent 1 cannot manipulate f at $(P_1, P_2, P_3, P_4, \ldots, P_n)$ by $(P_2, P_2, P_3, P_4, \ldots, P_n)$. Similarly, agent 2 cannot manipulate at $(P_1, P_1, P_3, P_4, \ldots, P_n)$ by $(P_1, P_2, P_3, P_4, \ldots, P_n)$. So, aP_1b. This is a contradiction.

Case B: Suppose $j = 1$ is the dictator in g. In this case, we construct a 2-agent social choice function h as follows: for every preference profile $(P_1, P_2, \ldots, P_n)$, we define

$$h^{P_{-12}}(P_1, P_2) = f(P_1, P_2, \ldots, P_n).$$

Since agent 1 is the dictator in g, $h^{P_{-12}}$ is onto. Moreover, $h^{P_{-12}}$ is strategy-proof: if any of the agents can manipulate in $h^{P_{-12}}$, they can also manipulate in f. By our induction hypothesis, $h^{P_{-12}}$ is dictatorship. But $h^{P_{-12}}$ was defined for every $n-2$ agent profile $P_{-12} \equiv (P_3, P_4, \ldots, P_n)$. We show that the dictator does not change across two $n-2$ agent profiles.

Assume for contradiction that agent 1 is the dictator for profile $(P_3, P_4, \ldots, P_n)$ but agent 2 is the dictator for profile $(\bar{P}_3, \bar{P}_4, \ldots, \bar{P}_n)$. Now, progressively change the preference profile $(P_3, P_4, \ldots, P_n)$ to $(\bar{P}_3, \bar{P}_4, \ldots, \bar{P}_n)$, where in each step, we change the preference of one agent j from P_j to $\bar{P}_j$. Then, there must exist a profile $(\bar{P}_3, \bar{P}_4, \bar{P}_{j-1}, P_j, P_{j+1}, \ldots, P_n)$ where agent 1 dictates and another profile $(\bar{P}_3, \bar{P}_4, \bar{P}_{j-1}, \bar{P}_j, P_{j+1}, \ldots, P_n)$ where agent 2 dictates with

$3 \leq j \leq n$. Consider $a, b \in A$ such that aP_jb. Pick P_1 and P_2 such that $P_1(1) = b$ and $P_2(1) = a$ with $a \neq b$. By definition,

$$f(P_1, P_2, \bar{P}_3, \bar{P}_4, \bar{P}_{j-1}, P_j, P_{j+1}, \ldots, P_n) = P_1(1) = b,$$
$$f(P_1, P_2, \bar{P}_3, \bar{P}_4, \bar{P}_{j-1}, \bar{P}_j, P_{j+1}, \ldots, P_n) = P_2(1) = a.$$

This means agent j can manipulate in SCF f at $(P_1, P_2, \bar{P}_3, \bar{P}_4, \bar{P}_{j-1}, P_j, P_{j+1}, \ldots, P_n)$ by $(P_1, P_2, \bar{P}_3, \bar{P}_4, \bar{P}_{j-1}, \bar{P}_j, P_{j+1}, \ldots, P_n)$. This is a contradiction since f is strategy-proof. This shows that f is also a dictatorship.

This completes the proof of the proposition. ■

The proof of the Gibbard-Satterthwaite theorem follows from Propositions 3 and 4, and from the fact that the proof is trivial for $n = 1$.

Note that the induction step must start at $n = 2$, and not $n = 1$, since the induction argument going from k to $k + 1$ works for $k \geq 2$ only.

2.4 Remarks on Gibbard-Satterthwaite Theorem

We make the following observations about the Gibbard-Satterthwaite (GS) theorem.

1. $|A| = 2$. The GS theorem fails when there are only two alternatives. An example of a non-dictatorial social choice function which is onto and strategy-proof is the plurality social choice function with a fixed tie-breaking. (The proof of this fact is an exercise.)

2. Indifference. Suppose every agent has a preference ordering which is not necessarily anti-symmetric, i.e., there are ties between alternatives. Let $\mathcal{R}$ be the set of all preference orderings. Note that $\mathcal{P} \subsetneq \mathcal{R}$. Now, consider a domain $\mathcal{D} \subseteq \mathcal{R}$ such that $\mathcal{P} \subseteq \mathcal{D}$. Call such a domain **admissible**. A social choice function $f : \mathcal{D}^n \to A$ is **admissible** if $\mathcal{D}$ is admissible. In other words, if the domain of preference orderings include the *all possible* linear orderings, then such a domain is admissible. The GS theorem is valid in admissible domains, i.e., if $|A| \geq 3$ and $f : \mathcal{D}^n \to A$ is admissible, onto, and strategy-proof, then it is a dictatorship. (The proof of this fact is an exercise.)

3. Restricted Domains. The GS theorem fails in various *restricted domains.* In particular, a domain $\mathcal{D} \subseteq \mathcal{R}$ is called a restricted domain if $\mathcal{P} \nsubseteq \mathcal{D}$. This will be the focus of discussion in the next section.

3 House Allocation Mechanisms

In this section, we look at an important model where the GS theorem does not hold. There is a finite set of objects $M = \{a_1, \ldots, a_m\}$ and a finite set of agents $N = \{1, \ldots, n\}$. We assume that $m \geq n$. The objects can be houses, jobs, projects, positions, candidates or students etc. Each agent has a linear order over the set of objects, i.e., a complete, transitive, and anti-symmetric binary relation. In this model, this ordering represents the preference of agents, and is the private information of agents. The preference ordering of agent i will be denoted as $\succ_i$. A profile of preferences will be denoted as $\succ \equiv (\succ_1, \ldots, \succ_n)$. The set of all preference orderings over M will be denoted as $\mathcal{M}$. The top element amongst a set of objects $S \subseteq M$ according to ordering $\succ_i$ is denoted as $\succ_i (1, S)$, and the k-th ranked object by $\succ_i (k, S)$.

The main departure of this model is that agents do not have direct preference over alternatives. We need to extract their preference over alternatives from their preference over objects. What are the alternatives? An alternative is a *feasible matching*, i.e., an injective mapping from N to M. The set of alternatives will be denoted as A, and this is the set of all injective mappings from N to M. For a given alternative $a \in A$, if $a(i) = j \in M$, then we say that agent i is assigned object j (in a).

Consider two alternatives a and b. Suppose agent 1 is is assigned the same object in both a and b (this is possible if there are at least three objects). Then, it is reasonable to assume that agent 1 will **always** be indifferent between a and b. Hence, for any preference ordering of agent 1, aP_1b and bP_1a are not *permissible*. This restriction implies that the domain of preference orderings over alternatives is not the unrestricted domain, which was the case in the GS theorem. Because of this reason, we cannot apply the GS theorem. Indeed, we will show that non-dictatorial social choice functions are strategy-proof in these settings.

A social choice function f is a mapping $f : \mathcal{M}^n \rightarrow A$. We now define a **fixed priority (serial dictatorship)** mechanism. We call this a mechanism but not a social choice function since it is not a direct revelation mechanism. A **priority** is a mapping $\sigma : N \rightarrow N$, i.e., an ordering over the set of agents. The fixed priority mechanism is defined inductively. Fix a preference profile $\succ$. We now construct an alternative a as follows:

$$
\begin{aligned}
a(\sigma(1)) &= \succ_{\sigma(1)} (1, N) \\
a(\sigma(2)) &= \succ_{\sigma(2)} (1, N \setminus \{a(\sigma(1))\}) \\
a(\sigma(3)) &= \succ_{\sigma(3)} (1, N \setminus \{a(\sigma(1)), a(\sigma(2))\}) \\
&\ldots\ldots \\
a(\sigma(i)) &= \succ_{\sigma(i)} (1, N \setminus \{a(\sigma(1)), \ldots, a(\sigma(i-1))\} \\
&\ldots\ldots \\
a(\sigma(n)) &= \succ_{\sigma(n)} (1, N \setminus \{a(\sigma(1)), \ldots, a(\sigma(n-1))\}.
\end{aligned}
$$

Now, the fixed priority mechanism (and the underlying SCF) assigns $f^\sigma(\succ) = a$.

Let us consider an example. We start with an example. The ordering over houses $\{a_1, a_2, \ldots, a_6\}$ of agents $\{1, 2, \ldots, 6\}$ is shown in Table 11. Fix a priority σ as follows: $\sigma(i) = i$ for all $i \in N$. According to this priority, the fixed priority mechanism will let agent 1 choose his best object first, which is a_3. Next, agent 2 chooses his best object

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$
a_3	a_3	a_1	a_2	a_2	a_1
a_1	a_2	a_4	a_1	a_1	a_3
a_2	a_1	a_3	a_5	a_6	a_2
a_4	a_5	a_2	a_4	a_4	a_4
a_5	a_4	a_6	a_3	a_5	a_6
a_6	a_6	a_5	a_6	a_3	a_5

Table 11: An example for housing model

among remaining objects, which is a_2. Next, agent 3 gets his best object among remaining objects $\{a_1, a_4, a_5, a_6\}$, which is a_1. Next, agent 4 gets his object among remaining objects $\{a_4, a_5, a_6\}$, which is a_5. Next, agent 5 gets his best object among remaining objects $\{a_4, a_6\}$, which is a_6. So, agent 6 gets a_4.

Note that a fixed priority mechanism is a generalization of dictatorship. We show below (quite obvious) that a fixed priority mechanism is strategy-proof. Moreover, it is efficient in the following sense.

Definition **7** *A social choice function f is* **efficient** *(in the house allocation model) if for all preference profiles $\succ$ and all matchings a, if there exists another matching $a' \neq a$ such that either $a'(i) \succ_i a(i)$ or $a'(i) = a(i)$ for all $i \in N$, then $f(\succ) \neq a$.*

Proposition **5** *Every fixed priority social choice function (mechanism) is strategy-proof and efficient.*

A word of caution here about strategy-proof notion of the fixed priority social choice function. The fixed priority mechanism is not a direct mechanism. However, using revelation principle, one can think of the associated direct mechanism - agents report their entire ordering, and the mechanism designer executes the fixed priority SCF on this ordering. Whenever, we say that the fixed priority mechanism is strategy-proof, we mean that the underlying direct mechanism is strategy-proof.

Proof: Fix a priority σ, and consider f^σ- the associated fixed priority mechanism. The strategy of any agent i is any ordering over M. Suppose agent i wants to deviate. When agent i is truthful, let M^{-i} be the set of objects allocated to agents who have higher priority than i (agent j has higher priority than agent i if and only if $\sigma(j) < \sigma(i)$). So, by being truthful, agent i get $\succ_i (1, M \setminus M^{-i})$. When agent i deviates, any agent j who has a higher priority than agent i continues to get the same object that he was getting when agent i was truthful. So, agent i gets an object in $M \setminus M^{-i}$. Hence, deviation cannot be better.

To show efficiency, assume for contradiction that f^σ is not efficient. Consider a profile $\succ$ such that $f(\succ) = a$. Let a' be another matching satisfying $a'(i) \succ_i a(i)$ or $a'(i) = a(i)$ for all $i \in N$. Then, consider the first agent j in the priority σ such that $a'(j) \succ_j a(j)$. Since agents before j in priority σ got the objects of matching a', object $a'(j)$ was still available to agent j. This is a contradiction since agent j chose $a(j)$ with $a'(j) \succ_j a(j)$. ■

But one can construct social choice functions which are strategy-proof but not a fixed priority mechanism in this model. We show this by an example. Let $N = \{1, 2, 3\}$ and $M = \{a_1, a_2, a_3\}$. The social choice function we consider is f, and is *almost* a fixed priority SCF. Fix a priority σ as follows: $\sigma(i) = i$ for all $i \in N$. Another priority is σ': $\sigma'(1) = 2, \sigma'(2) = 1, \sigma'(3) = 3$. The SCF f generates the same outcome as f^{σ} whenever $\succ_2 (1, M) \neq a_1$. If $\succ_2 (1, M) = a_1$, then it generates the same outcome as $f^{\sigma'}$. To see that this is strategy-proof, it is clear that agents 1 and 3 cannot manipulate since they cannot change the priority. Agent 2 can change the priority. But, can he manipulate? If his top ranked house is a_1, he gets it, and he cannot manipulate. If his top ranked house is $\in \{a_2, a_3\}$, then he cannot manipulate without changing the priority. If he does change the priority, then he gets a_1. But being truthful, either he gets his top ranked house or second ranked house. So, he gets a house which is either a_1 or some house which he likes more than a_1. Hence, he cannot manipulate.

3.1 Top Trading Cycle Mechanism with Fixed Endowments

The top trading cycle mechanism (TTC) with fixed endowment is a class of general mechanisms which are strategy-proof, and has some nice properties. We will study them in detail here.

We assume here $m = n$. To explain the mechanism, we start with the example in Table 11. In the first step of the TTC mechanism, agents are endowed with a house each. Suppose the *fixed endowment* for this example is a^*: $a^*(1) = a_1, a^*(2) = a_3, a^*(3) = a_2, a^*(4) = a_4, a^*(5) = a_5, a^*(6) = a_6$.

The TTC mechanism goes in steps. In each step, a set of houses are assigned to a set of agents, and they are excluded from the subsequent steps of the mechanism. Hence, the mechanism maintains a set of "remaining agents" and a set of "remaining houses" in each step.

At every step, a directed graph is constructed. The set of nodes in this directed graph is the same as the set of remaining agents. Initially, the set of remaining agents is N. Then, there is a directed edge from agent i to agent j if and only if agent j is endowed with agent i's top ranked house amongst the remaining houses (initially, all houses are remaining houses). Formally, if $H \subseteq M$ is the set of remaining houses in any step, then the directed graph in this iteration has an edge from agent i to agent j (i can be j also) if and only if $\succ_i (1, H) = a^*(i)$. Note that such a graph will have exactly one outgoing edge from every node (though possibly many incoming edges to a node). Further, there may be an edge from a node to itself (this will be treated as cycle, and called a loop). It is clear that such a graph will always have a cycle.

Figure 1 shows the directed graph for the first step of the example in Table 11. The only cycle in this graph is a loop involving agent 2. So, agent 2 gets his endowment, which is house a_3. Agent 2 is eliminated from the graph, and house a_3 is eliminated from the problem. Now, the graph for the next step is constructed. Now, every agent points to his top ranked house amongst houses remaining (which is the houses except house a_3). This graph is shown in Figure 2. Here, the only cycle is a loop involving agent 1. So, agent 1 gets his endowment a_1. Agent 1 and house a_1 is eliminated from the problem. Next, the graph for the next step is constructed, which is shown in Figure 3. There is a cycle involving agents 3 and 4. So, agent 3 gets the endowment of agent 4 (a_4) and agent 4 gets the endowment of agent 3 (a_2). These agents and houses are eliminated from the problem, and the next graph is constructed as shown in Figure 4. This graph has a loop involving agent 6. So, agent 6 gets his endowment a_6, and the only remaining house a_5 goes to agent 5.

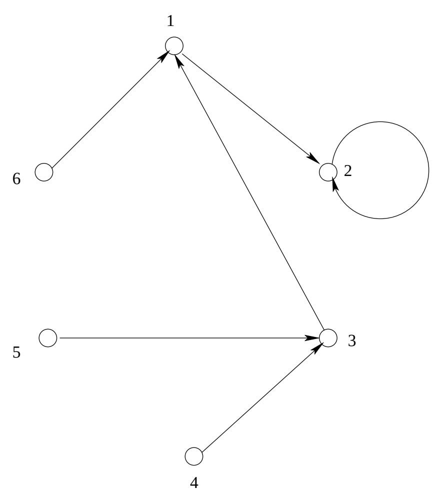

Figure 1: Cycle in Step 1 of the TTC mechanism

We now formally describe the TTC mechanism. Fix an endowment of agents a^*. The mechanism maintains the remaining set of houses M^k and remaining set of agent N^k in every Step k of the mechanism.

- Step 1: Set $M^1 = M$ and $N^1 = N$. Construct a directed graph G^1 with nodes N^1. There is a directed edge from node (agent) $i \in N^1$ to agent $j \in N^1$ if and only if $\succ_i (1, M^1) = a^*(j)$.

  Allocate houses along every cycle of graph G^1. Formally, if $(i^1, i^2, \ldots, i^p, i^1)$ is a cycle in G^1 then set $a(i^1) = a^*(i^2), a(i^2) = a^*(i^3), \ldots, a(i^{p-1}) = a^*(i^p), a(i^p) = a^*(i^1)$. Let $\hat{N}^1$ be the set of agents allocated in such cycles in G^1, and $\hat{M}^1$ be the set of houses assigned in a to N^1.

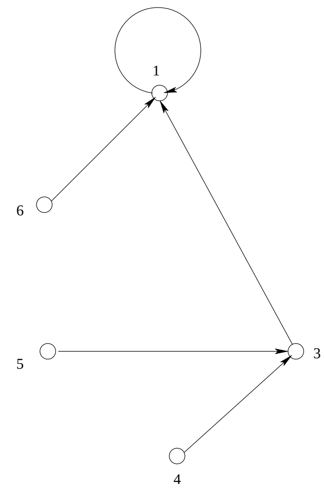

Figure 2: Cycle in Step 2 of the TTC mechanism

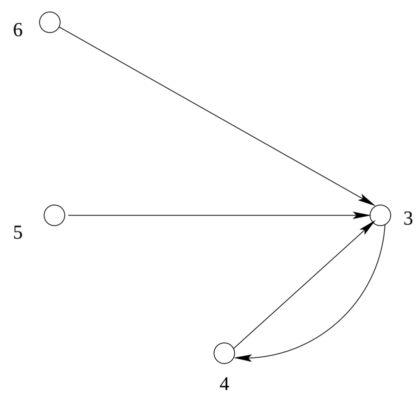

Figure 3: Cycle in Step 3 of the TTC mechanism

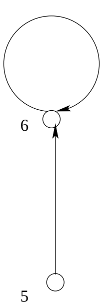

Figure 4: Cycle in Step 4 of the TTC mechanism

Set $N^2 = N^1 \setminus \hat{N}^1$ and $M^2 = M^1 \setminus \hat{M}^1$.

- STEP k: Construct a directed graph G^k with nodes N^k. There is a directed edge from node (agent) $i \in N^k$ to agent $j \in N^k$ if and only if $\succ_i (1, M^k) = a^*(j)$.

Allocate houses along every cycle of graph G^k. Formally, if $(i^1, i^2, \ldots, i^p, i^1)$ is a cycle in G^k then set $a(i^1) = a^*(i^2), a(i^2) = a^*(i^3), \ldots, a(i^{p-1}) = a^*(i^p), a(i^p) = a^*(i^1)$. Let $\hat{N}^k$ be the set of agents allocated in such cycles in G^k, and $\hat{M}^k$ be the set of houses assigned in a to N^k.

Set $N^{k+1} = N^k \setminus \hat{N}^k$ and $M^{k+1} = M^k \setminus \hat{M}^k$. If N^{k+1} is empty, STOP, and a is the final matching chosen. Else, repeat.

PROPOSITION **6** *TTC with fixed endowment mechanism is strategy-proof and efficient.*

Proof: Consider agent i who wants to deviate. Suppose agent i is getting assigned in Step k of the TTC mechanism if he is truthful. Given the preferences of the other agents, suppose agent i reports a preference ordering different from his true preference ordering. Let H^{k-1} be the set of houses assigned in Steps 1 through $k-1$ when agent i is truthful. If the deviation of agent i results in no change of his strategy (pointing to the most preferred remaining house) before Step k, then the allocation of houses in H^{k-1} will not change due to his deviation. As a result agent i will get an object from $M \setminus H^{k-1}$. Since agent i gets his most preferred object from $M \setminus H^{k-1}$ if he is truthful, this is not a successful manipulation. Hence, we focus on the case where the deviation of agent i result in a change of his strategy before Step k.

Suppose $r < k$ is the first step in the TTC mechanism where the underlying graph G^r changes due to this deviation. Notice that the only change in graph G^r in cases where agent i is truthful and where he is deviating is the outgoing edge of agent i. Consider the case when agent i is truthful. In that case since agent i is not allocated in Step r, he is not involved in any cycle in G^r. But there may be sequence of nodes of the nature $(i^1, i^2, \ldots, i^p, i)$, where i^1 has no incoming edge, but edges exist from i^1 to i^2, and i^2 to i^3, and so on. Call such sequence of nodes i-paths. Let P_i be the set of all nodes in all the i-paths - P_i includes i also.

Figure 5 gives an illustration. Here, $P_i = \{i^1, i^2, i^3, i^5, i^6, i\}$.

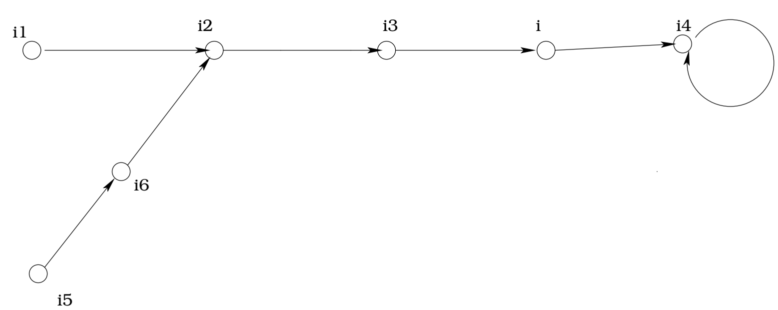

Figure 5: i-Paths in a Step

Note that if agent i's deviation does not lead agent i to point to an agent in P_i, then the allocations in Step r is unchanged because of his deviation. This follows from the fact that the only way i can change allocation in Step r is by creating a new cycle involving himself - he cannot break cycles which does not involve him. As a result, the only way to change the allocation in Step r is to deviate by pointing to an agent in P_i. In that case, a subset of agents in P_i which includes i, call them C^r, will form a cycle, and get assigned in Step r. We argue that agents in C^r must be unassigned (i.e., part of the "remaining agents") in Step k when agent i is truthful. To see this, consider any agent $i^1 \in P_i$. By definition, there is a path from i^1 from i^1 to i - say, $(i^1, i^2, \ldots, i^p, i)$. Since house of i is available till Step k, i^p will continue to point to i. Hence, the house of i^p is available till Step k. As a result, i^{p-1} will continue to point to i^p till Step k, and so on. Hence, the path $(i^1, i^2, \ldots, i^p, i)$ will continue to exist in Step k. This shows that agent in C^r are unassigned in Step k. Hence, the allocation achieved by agent i by his deviation in Step r can also achieved by deviating in Step k. But, we know that if he deviates in Step k, then it is not a successful manipulation. So, the only possibility is that he deviates by pointing to an agent not in P_i, in which case he does not alter the allocation in Step r. As a result, the cycles in subsequent rounds also do not change due to deviations.

Hence, all the agents who were assigned in Steps 1 through $(k-1)$ still get assigned the same houses. By definition, agent k gets his top ranked object amongst $M \setminus H^{k-1}$ if he is truthful. By deviating he will get an object in $M \setminus H^{k-1}$. Hence, deviation cannot be better.

Now, we prove efficiency. Let a be a matching produced by the TTC mechanism for preference profile $\succ$. Assume for contradiction that this matching is not efficient, i.e., there exists a different matching a' such that $a'(i) \succ_i a(i)$ or $a'(i) = a(i)$ for all $i \in N$. Consider the first step of the TTC mechanism where some agent i gets $a(i) \neq a'(i)$. Since all the agents get the same object in a and a' before this step, object $a'(i)$ is available in this step, and since $a'(i) \succ_i a(i)$, agent i cannot have an edge from i to the "owner" of $a(i)$ in this step. This means that agent i cannot be assigned to $a(i)$. This gives a contradiction. ■

4 Stable House Allocation with Existing Tenants

We consider a variant of the house allocation problem. In this model, each agent already has a house that he owns - if an agent i owns house j then he is called the tenant of j. Immediately, one sees that the TTC mechanism can be applied in this setting with initial endowment given by the house-tenant relationship. This is, as we have shown, strategy-proof and efficient (Proposition 6).

We address another concern here, that of *stability*. In this model, agents own resources that are allocated. So, it is natural to impose some sort of stability condition on the mechanism. Otherwise, a group of agents can break away and trade their houses amongst themselves.

Consider the example in Table 11. Let the existing tenants of the houses be given by matching a^*: $a^*(1) = a_1, a^*(2) = a_3, a^*(3) = a_2, a^*(4) = a_4, a^*(5) = a_5, a^*(6) = a_6$. Consider a matching a as follows: $a(i) = a_i$ for all $i \in N$. Now consider the coalition of agents $\{3, 4\}$. In the matching a, we have $a(3) = a_3$ and $a(4) = a_4$. But agents 3 and 4 can reallocate the houses they own among themselves in a manner to get a better matching for themselves. In particular, agent 3 can get a_4 (house owned by agent 4) and agent 4 can get a_2 (house owned by agent 3. Note that $a_4 \succ_3 a_3$ and $a_2 \succ_4 a_4$. Hence, both the agents are better off trading among themselves. So, they can potentially *block* matching a. We formalize this idea of blocking below.

Let a^* denote the matching reflecting the initial endowment of agents. We will use the notation a^S for every $S \subseteq N$, to denote a matching of agents in S to the houses owned by agents in S. Whenever we write a matching a without any superscript we mean a matching of all agents. Formally, a coalition (group of agents) $S \subseteq N$ can **block** a matching a at a preference profile $\succ$ if there exists a matching a^S such that $a^S(i) \succ_i a(i)$ or $a^S(i) = a(i)$ for all $i \in S$ with $a^S(j) \succ_j a(j)$ for some $j \in S$. A matching a is in the **core** at a preference profile $\succ$ if no coalition of agents can block a at $\succ$. A social choice function f is **stable** if for all preference profile $\succ$, $f(\succ)$ is in the core at preference profile $\succ$.

We will now analyze if the TTC mechanism is stable. Note that when we say a TTC mechanism, we mean the TTC mechanism where the initial endowment is the endowment given by the house-tenant relationship.

Proposition **7** *The TTC mechanism is stable. Moreover, there is a unique core matching for every preference profile.*

Proof: Assume for contradiction that the TTC mechanism is not stable. Then, there exists a preference profile $\succ$, where the matching a produced by the TTC mechanism at $\succ$ is not in the core. Let coalition S block this matching a at $\succ$. This means there exists another matching a^S such that $a^S(i) \succ_i a(i)$ or $a^S(i) = a(i)$ for all $i \in S$, with equality not holding for all $i \in S$. Let $T = \{i \in S : a^S(i) \succ_i a(i)\}$.

To remind notation, we denote $\hat{N}^k$ to be the set of agents allocated houses in Step k of the TTC mechanism, and $\hat{M}^k$ be the set of these houses. Clearly, agents in $S \cap \hat{N}^1$ are getting their respective top ranked houses. So, $(S \cap \hat{N}^1) \subseteq (S \setminus T)$. Now, agents in $S \cap \hat{N}^2$ are getting their respective top ranked houses amongst houses in $M \setminus \hat{M}^1$. Given that agents in $S \cap \hat{N}^1$ get the same set of houses in a^S and a, any agent in $S \cap \hat{N}^2$ cannot be getting a better house in a^S than his house in a. Hence, again $(S \cap \hat{N}^2) \subseteq (S \setminus T)$. Continuing this way, we get $S \subseteq (S \setminus T)$ or $T = \emptyset$, which is a contradiction.

Finally, we show that the core matching returned by the TTC mechanism is the unique one. Suppose the core matching returned by the TTC mechanism is a, and let a' be another core matching for preference profile $\succ$. Note that (a) in every Step k of the TTC mechanism agents in $\hat{N}^k$ get allocated to houses owned by agents in $\hat{N}^k$, and (b) agents in $\hat{N}^1$ get their top ranked houses. Hence, if $a(i) \neq a'(i)$ for any $i \in \hat{N}^1$, then agents in N^1 will block.

Given this, agents in $\hat{N}^2$ get their highest ranked house from $M \setminus \hat{M}^1$. So, given that agents in $\hat{N}^1$ get the same houses in a and a', if agents in $\hat{N}^2$ get different houses in a and a', then they will block. Continuing this way, we conclude that the the matchings a and a' must be the same. ■

The TTC mechanism with existing tenants has another nice property. Call a mechanism f **individually rational** if at every profile $\succ$, the matching $f(\succ) \equiv a$ satisfies $a(i) \succ_i a^*(i)$ or $a(i) = a^*(i)$ for all $i \in N$, where a^* is the matching given by the initial endowment or existing tenants.

Clearly, the TTC mechanism is individually rational. To see this, consider a profile $\succ$ and let $f(\succ) = a$. Note that the TTC mechanism has this property that if the house owned by an agent i is matched in Step k, then agent i is matched to a house in Step k too. If $a(i) \neq a^*(i)$ for some i, then agent i must be part of a trading cycle where he is pointing to a house better than $a^*(i)$. Hence, $a(i) \succ_i a^*(i)$.

In the model of house allocation with existing tenants, the TTC mechanism satisfies three compelling properties along with stability - it is strategy-proof, efficient, and individually rational. Remarkably, these three properties characterize the TTC mechanism. We skip the proof.

THEOREM **3** *A mechanism is strategy-proof, efficient, and individually rational if and only if it is the TTC mechanism.*

Note that the serial dictatorship with a fixed priority is strategy-proof and efficient but not individually rational. The "status-quo mechanism" where everyone is assigned the houses they own is strategy-proof and individually rational but not efficient. So, the properties of individual rationality and efficiency are crucial for the characterization of Theorem 3.

5 The Marriage Market Model

The house allocation model is a model of one-sided matching - only agents (one side of the market) had preference over the houses. In many situations, the matching market can be partitioned into two sides, and an agent on one side will have preference over agents on the other side. For instance, consider the scenario where students are matched to schools. It is plausible that not only students have a preference ordering over the schools but schools also have a preference over students. Other applications of two-sided matching include job applicants matched to firms, doctoral students matched to faculty etc.

Let M be a set of men and W be a set of women. For simplicity, we will assume that $|M| = |W|$ - but this is not required to derive the results. Every man $m \in M$ has a *strict* preference ordering $\succ_m$ over the set of women W. So, for $x, y \in W$, $x \succ_m y$ will imply that m ranks x over y. A matching is a bijective mapping $\mu : M \to W$, i.e., every man is assigned to a unique woman. If μ is a matching, then $\mu(m)$ denotes the woman matched to man m and $\mu^{-1}(w)$ denotes the man matched to woman w. This model is often called the "marriage market" model or "two-sided matching" model. We first discuss the stability aspects of this model, and then discuss the strategic aspects.

5.1 Stable Matchings in Marriage Market

As in the house allocation model with existing tenants, the resources to be allocated to agents in the marriage market model are owned by agents themselves. Hence, stability becomes an important criteria for designing any mechanism.

We consider an example with three men and three women. Let $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3\}$. Their preferences are shown in Table 12.

$\succ_{m_1}$	$\succ_{m_2}$	$\succ_{m_3}$	$\succ_{w_1}$	$\succ_{w_2}$	$\succ_{w_3}$
w_2	w_1	w_1	m_1	m_3	m_1
w_1	w_3	w_2	m_3	m_1	m_3
w_3	w_2	w_3	m_2	m_2	m_2

Table 12: Preference orderings of men and women

Consider the following matching μ: $\mu(m_1) = w_1, \mu(m_2) = w_2, \mu(m_3) = w_3$. This matching is *unstable* in the following sense. The pair $(m_1, \mu(m_2) = w_2)$ will *block* this matching (ex post) since m_1 likes w_2 over $\mu(m_1) = w_1$ and w_2 likes m_1 over $\mu^{-1}(w_2) = m_2$. So, (m_1, w_2) will break away, and form a new market. This motivates the following definition of stability.

Definition 8 *A matching μ is* **unstable** *if there exists $m, m' \in M$ such that (a) $\mu(m') \succ_m \mu(m)$ and (b) $m \succ_{\mu(m')} m'$. The pair $(m, \mu(m'))$ is called a* **blocking pair**. *If a matching μ has no blocking pairs, then it is called a* **stable matching**.

The following matching μ' is a stable matching: $\mu'(m_1) = w_1, \mu'(m_2) = w_3, \mu'(m_3) = w_2$ for the example in Table 12. The question is: Does a stable matching always exist? The answer to this question is remarkably yes, as we will show next.

One can imagine a stronger requirement of stability, where groups of agents block instead of just pairwise blocking. We say that a coalition $S \subseteq (M \cap W)$ **blocks** a matching μ at a profile $\succ$ if there exists another matching μ' such that (i) for all $m \in M \cap S$, $\mu'(m) \in S$ and for all $w \in M \cap S$, $\mu'^{-1}(w) \in S$, and (ii) for all $m \in M \cap S$, $\mu'(m) \succ_m \mu(m)$ and for all $w \in M \cap S$, $\mu'^{-1}(w) \succ_w \mu^{-1}(w)$. We say a matching μ is in core at a profile $\succ$ if no coalition can block μ at $\succ$. The following theorem suggests that this notion of stability is equivalent to the pairwise notion of stability we have initially defined.

Theorem 4 *A matching is stable at a profile if and only if it belongs to the core at that profile.*

Proof: Consider a matching μ which is stable. Assume for contradiction that μ is not in the core. Then, there must exist $S \subseteq (M \cup W)$ and a matching $\hat{\mu}$ such that for all $m \in M \cap S$ and for all $w \in W \cap S$ with $\hat{\mu}(m), \hat{\mu}^{-1}(w) \in S$ we have $\hat{\mu}(m) \succ_m \mu(m)$ and $\hat{\mu}^{-1}(w) \succ_w \mu^{-1}(w)$. This means for some $m \in S$ we have $\hat{\mu}(m) \in S$. Let $\hat{\mu}(m) = w$. We know $w \succ_m \mu(m)$. Then, we have $m \succ_w \mu^{-1}(w)$. Hence, (m, w) is a blocking pair. This implies that μ is not stable, which is a contradiction.

The other direction of the proof is trivial. ■

5.2 Deferred Acceptance Algorithm

In this section, we show that a stable matching always exists in the marriage market model. The fact that a stable matching always exists is proved by constructing an algorithm to find such a matching (this algorithm is due to David Gale). There are two versions of this algorithm. In one version men propose to women and women either accept or reject the proposal. In another version, women propose to men and men either accept or reject the proposal. We describe the men-proposal version.

- S1. First, every man proposes to his top ranked woman.
- S2. Then, every woman who has at least two proposals keeps (tentatively) the top man amongst these proposals and rejects the rest.
- S3. Then, every man who was rejected in the last round, proposes to the top woman amongst those women who have not rejected him in earlier rounds.
- S4. Then, every woman who has at least two proposals, including any proposal tentatively kept from earlier rounds, keeps (tentatively) the top man amongst these proposals and rejects the rest. The process in then repeated from Step S3 till each woman has a proposal at which point, the tentative proposal accepted by a woman becomes permanent.

Since each woman is allowed to keep only one proposal in every round, no woman will be assigned to more than one man. Since a man can propose only one woman at a time, no man will be assigned to more than one woman. Since the number of men and women are the same, this algorithm will terminate at a matching. Also, the algorithm will terminate finitely since in every round, the set of women a man can propose does not increase and decreases for at least one man.

We illustrate the algorithm for the example in Table 12. A proposal from $m \in M$ to $w \in W$ will be denoted by $m \to w$.

- In the first round, every man proposes to his best woman. So, $m_1 \to w_2, m_2 \to w_1, m_3 \to w_1$.
- Hence, w_1 has two proposals: $\{m_2, m_3\}$. Since $m_3 \succ_{w_1} m_2$, w_1 rejects m_2 and keeps m_3.
- Now, m_2 is left to choose from $\{w_2, w_3\}$. Since $w_3 \succ_{m_2} w_2$, m_2 now proposes to w_3.
- Now, every woman has exactly one proposal. So the algorithm stops with the matching μ_m given by $\mu_m(m_1) = w_2, \mu_m(m_2) = w_3, \mu_m(m_3) = w_1$.

It can be verified that μ_m is a stable matching. Also, note that μ_m is a different stable matching than the stable matching μ' which we discussed earlier. Hence, there can be more than one stable matching.

One can also state a women proposing version of the deferred acceptance algorithm. Let us run the women proposing version for the example in Table 12. As before, a proposal from $w \in W$ to $m \in M$ will be denoted by $w \to m$.

- In the first round, every woman proposes to her top man. So, $w_1 \to m_1, w_2 \to m_3, w_3 \to m_1$.
- So, m_1 has two proposals: $\{w_1, w_3\}$. We note that $w_1 \succ_{m_1} w_3$. Hence, m_1 rejects w_3 and keeps w_1.
- Now, w_3 is left to choose from $\{m_2, m_3\}$. Since $m_3 \succ_{w_3} m_2$, w_3 proposes to m_3.
- This implies that m_3 has two proposals: $\{w_2, w_3\}$. Since $w_2 \succ_{m_3} w_3$, m_3 rejects w_3 and keeps w_2.
- Now, w_3 is left to choose only m_2. So, the algorithm terminates with the matching μ_w given by $\mu_w(m_1) = w_1, \mu_w(m_2) = w_3, \mu_w(m_3) = w_2$.

Note that μ_w is a stable matching and $\mu_m \neq \mu_w$.

5.3 Stability and Optimality of Deferred Acceptance Algorithm

Theorem **5** *The Deferred Acceptance Algorithm terminates in a stable matching.*

Proof: Consider the Deferred Acceptance Algorithm where men propose (a similar proof works if women propose). Let μ be the final matching of the algorithm. Assume for contradiction that μ is not a stable matching. This implies that there exists a pair $m \in M$ and $w \in W$ such that (m, w) is a blocking pair. By definition $\mu(m) \neq w$ and $w \succ_m \mu(m)$. This means that w rejected m earlier in the algorithm (else m would have proposed to w at the end of the algorithm). But a woman rejects a man only if she gets a better proposal, and her proposals improve in every round. This implies that w must be assigned to a better man than m, i.e., $\mu^{-1}(w) \succ_w m$. This contradicts the fact that (m, w) is a blocking pair. ■

The men-proposing and the women-proposing versions of the Deferred Acceptance Algorithm may output different stable matchings. Is there a reason to prefer one of the stable matchings over the other? Put differently, should we use the men-proposing version of the algorithm or the women-proposing version?

To answer this question, we start with some notations. A matching μ men-dominates another matching μ' if for all $m \in M$ either $\mu'(m) \succ_m \mu(m)$ or $\mu'(m) = \mu(m)$ with the strict preference holding for at least one m. A stable matching μ is **men-optimal stable** if there does not exist a stable matching μ' such that μ' men-dominates μ. Similarly, we can define a **women-optimal stable** matching.

Theorem **6** *The men-proposing version of the Deferred Acceptance Algorithm terminates at a men-optimal stable matching and the women-proposing version of the Deferred Acceptance Algorithm terminates at a women-optimal stable matching.*

Proof: We do the proof for men-proposing version of the algorithm. The proof is similar for the women-proposing version. Let $\hat{\mu}$ be the stable matching obtained at the end of the men-proposing Deferred Acceptance Algorithm. Assume for contradiction that $\hat{\mu}$ is not men-optimal. Then, there exists a stable matching μ such that for all $m \in M$, either $\mu(m) \succ_m \hat{\mu}(m)$ or $\mu(m) = \hat{\mu}(m)$, with strict inequality holding at least for one men. Let $M' = \{m \in M : \mu(m) \succ_m \hat{\mu}(m)\}$. Hence, $M' \neq \emptyset$.

Now, for every $m \in M'$, since $\mu(m) \succ_m \hat{\mu}(m)$, we know that m is rejected by $\mu(m)$ in some round of the algorithm. Denote the round in which $m \in M'$ is rejected by $\mu(m)$ by t_m. Choose $m' \in \arg\min_{m \in M'} t_m$, i.e., choose a man m' who is the first to be rejected by $\mu(m')$ among all men in M'. Since $\mu(m')$ rejects m', she must have got a better proposal from some other man m'', i.e.,

$$m'' \succ_{\mu(m')} m'. \tag{1}$$

Now, consider $\mu(m')$ and $\mu(m'')$. If $m'' \notin M'$, then $\hat{\mu}(m'') = \mu(m'')$. Hence, m'' proposes to $\mu(m')$ before proposing to $\mu(m'')$. If $m'' \in M'$, then, since $t_{m''} > t_{m'}$, m'' has not been rejected by $\mu(m'')$ till round $t_{m'}$. This means, again, m'' proposed to $\mu(m')$ before proposing to $\mu(m'')$. In both cases, we conclude that

$$\mu(m') \succ_{m''} \mu(m''). \tag{2}$$

By Equations 1 and 2, $(m'', \mu(m'))$ forms a blocking pair. Hence, μ is not stable. This is a contradiction. ■

The natural question is then whether there exists a stable matching that is optimal for both men and women. The answer is no. The example in Table 12 has two stable matchings, one is optimal for men but not for women and one is optimal for women but not for men. Also, there is a unique men-optimal stable matching and a unique women-optimal stable matching (the proof of this fact is skipped).

5.4 Strategic Issues in Deferred Acceptance Algorithm

We next turn to strategic properties of the Deferred Acceptance Algorithm. We first consider the men-proposing version. We define the notion of strategyproofness informally here. Consider the strategy faced by an agent in this mechanism. A man proposes to some agent from a *feasible set* of agents in every round. A woman accepts some agent from the set of existing proposals in every round. We say a man is truthful is he proposes to the top ranked woman from his feasible set in every round. Similarly, we say a woman is truthful if she accepts the top ranked proposal in every round. We say that the Deferred Acceptance Algorithm is strategyproof for men (women) if in every profile of preference ordering, the best strategy for every man (woman) is to be truthful.

We first show that the men-proposing version of the Deferred Acceptance Algorithm is not strategyproof for women. Let us return to the example in Table 12. We know if everyone is truthful, then the matching is: $\mu(m_1) = w_2, \mu(m_2) = w_3, \mu(m_3) = w_1$. We will show that w_1 can get a better outcome by not being truthful. We show the steps here.

- In the first round, every man proposes to his best woman. So, $m_1 \to w_2, m_2 \to w_1, m_3 \to w_1$.
- Next, w_2 only has one proposal (from m_1) and she accepts it. But w_1 has two proposals: $\{m_2, m_3\}$. If she is truthful, she should accept m_3. We will see what happens if she is not truthful. So, she accepts m_2.
- Now, m_3 has two choices: $\{w_2, w_3\}$. He likes w_2 over w_3. So, he proposes to w_2.
- Now, w_2 has two proposals: $\{m_1, m_3\}$. Since she likes m_3 over m_1, she accepts m_3.
- Now, m_1 has a choice between w_1 and w_3. Since he likes w_1 over w_3, he proposes to w_1.

- Now, w_1 has two proposal: $\{m_1, m_2\}$. Since she prefers m_1 over m_2 she accepts m_1.
- So, m_2 is only left with $\{w_2, w_3\}$. Since he likes w_3 over w_2 he proposes to w_3, which she accepts. So, the final matching $\hat{\mu}$ is given by $\hat{\mu}(m_1) = w_1, \hat{\mu}(m_2) = w_3, \hat{\mu}(m_3) = w_2$.

Hence, w_1 gets m_1 in $\hat{\mu}$ but was getting m_3 earlier. The fact that $m_1 \succ_{w_1} m_3$ shows that not being truthful helps w_1. However, the same result does not hold for men.

Theorem 7 *The men-proposing version of the Deferred Acceptance Algorithm is strategyproof for men. The women-proposing version of the Deferred Acceptance Algorithm is strategyproof for women.*

Proof: Suppose there is a profile $\pi = (\succ_{m_1}, \ldots, \succ_{m_n}, \succ_{w_1}, \ldots, \succ_{w_n})$ such that man m_1 can misreport his preference to be $\succ_*$, and obtain a better matching. Let this preference profile be π'. Let μ be the stable matching obtained by the men-proposing deferred acceptance algorithm when applied to π. Let ν be the stable matching obtained by the men-proposing algorithm when applied to π'. We show that if $\nu(m_1) \succ_{m_1} \mu(m_1)$, then ν is not stable at π', which is a contradiction.

Let $R = \{m : \nu(m) \succ_m \mu(m)\}$. Since $m_1 \in R$, R is not empty. We show that $\{w : \nu^{-1}(w) \in R\} = \{w : \mu^{-1}(w) \in R\}$. Take any $\nu^{-1}(w) \in R$, we will show that $\mu^{-1}(w) \in R$, and this will establish the claim. If $\mu^{-1}(w) = m_1$, then we are done by definition. Else, let $w = \nu(m)$ and $m' = \mu^{-1}(w)$. Since $w \succ_m \mu(m)$, stability of μ at π implies that $m' \succ_w m$. Stability of ν at π' implies that $\nu(m') \succ_{m'} w$. Therefore, $m' \in R$. Let $S = \{w : \nu^{-1}(w) \in R\} = \{w : \mu^{-1}(w) \in R\}$.

By definition $\nu(m) \succ_m \mu(m)$ for any $m \in R$. By stability of μ, we then have $\mu^{-1}(w) \succ_w \nu^{-1}(w)$ for all $w \in S$. Now, pick any $w \in S$. By definition, $w \succ_{\nu^{-1}(w)} \mu(\nu^{-1}(w))$. This implies that during the execution of the men-proposing deferred acceptance algorithm at π, $\nu^{-1}(w) \in R$ must have proposed to w which she had rejected. Let $m \in R$ be the last man in R to make a proposal during the execution of the men-proposing deferred acceptance algorithm at π. Suppose this proposal is made to $w = \mu(m) \in S$. As argued, w rejected $\nu^{-1}(w)$ earlier. This means that when m proposed to w, she had some proposal, say from m', which she rejected. By definition, m' cannot be in R. This means that $m' \neq \nu^{-1}(w)$, and hence, $m' \succ_w \nu^{-1}(w)$. Since $m' \notin R$, $\mu(m') \succ_{m'} \nu(m')$ or $\mu(m') = \nu(m')$. Also, since w rejects m', $w \succ_{m'} \mu(m')$. This shows that $w \succ_{m'} \nu(m')$. This shows that (m', w) form a blocking pair for ν at π'. ■

Does this mean that no mechanism can be both stable and be strategyproof to all agents? The answer is yes.

Theorem 8 *No mechanism which gives a stable matching can be strategy-proof for both men and women.*

Proof: If mechanism is strategy-proof, then for every instance, it must be non-manipulable by both men and women. Hence, it is sufficient to show that any stable mechanism is manipulable for some instance (example). We consider the example in Table 12.

One can verify that there are exactly two stable matchings for this example - one is men-optimal, and can be obtained using the men-proposing version of the deferred acceptance algorithm, and the other is women-optimal, and can be obtained using the women-proposing version of the deferred acceptance algorithm. So, any stable mechanism is equivalent (in terms of direct revelation) to either the men-proposing or the women proposing version of the deferred acceptance algorithm. But we have seen that in the men-proposing version woman w_1 can manipulate. Similarly, we can show that in the women-proposing version, man m_1 can manipulate by rejecting w_1 instead of w_3 in the first step. So, any stable mechanism is either manipulable by a man or a woman. ■

However, one can trivially construct strategy-proof mechanisms for both men and women. Consider a mechanism which ignores all men (or women) orderings. Then, it can run a fixed priority mechanism for men (or women) or a TTC mechanism with fixed endowments for men (or women) to get a strategy-proof mechanism.

5.5 Extensions with Quotas and Individual Rationality

The deferred acceptance algorithm can be suitably modified to handle some generalizations. One such generalization is used in school choice problems. In a school choice problem, a set of students (men) and a set of schools (women) have preference ordering over each other. Each school has a quota, i.e., the maximum number of students it can take. In particular, every school i has a quota of $q_i \geq 1$.

Students, on the other hand, have a set of schools that are acceptable and another set which is not acceptable, i.e., on top of the usual linear order over the set of schools, each student also has a *cut-off school*, below which he prefers to not attend any school. The preferences of agents are handled by adding a **dummy** school 0, whose quota is the number of students (so this school can admit possibly all students). An admission in the dummy school indicates that the student is not assigned any school. Now, each student has a preference ordering over the set of schools and the dummy school. All the schools below dummy school are never preferred by the student.

The deferred acceptance algorithm can be modified in a straightforward way in these settings. Each student proposes to its favorite remaining school. A proposal to the dummy school is always accepted. Any other school k evaluates the set of proposals it has, and accepts the top $\min(q_k,$ number of proposals$)$. The procedure is repeated as was described earlier. One can extend the stability, student-optimal stability, and strategy-proofness results of previous section to this setting in a straightforward way.

Another important property of a mechanism in such a set up is **individual rationality**. Individual rationality says that no student should get a school lower than the dummy school. It is clear that the deferred acceptance algorithm produces an individually rational matching.

6 Applications of Various Matching Models

The matching theory is one of those theories which have been applied extensively in practice. We give some examples.

- Deferred Acceptance Algorithm. Deferred acceptance algorithm (DAA) has been successfully used in assigning students to schools in New York City (high school) and Boston (all public schools). It is also used in assigning medical interns (doctors) to hospitals in US medicine schools. The US medical community has been at the forefront of implementing DAA - it is used in residents matching, doctor assignments to jobs, and many other markets.

- Versions of Serial Dictatorship. Some (random) version of serial dictatorship (priority) mechanism is widely used in US Universities like Yale, Princeton, CMU, Harvard, Duke, Michigan to allocate graduate housing to graduate students. The version that is used is called *random serial dictatorship with squatting rights*. In this version, first existing tenants are given the option of entering the mechanism or going away with their existing house. After everyone announces their willingness to participate in the mechanism, an ordering of (participating) students is done uniformly at random. Then, serial dictatorship is applied on this ordering.

- Kindney Exchange. The kidney exchange problem can be modeled as a house allocation problem with existing tenants. In a kidney exchange problem, each patient (agent) can come with an incompatible donor agent (house which is endowed to him), and there is a set of donor agents (vacant houses). Patients have preference over donors (houses). A matching in this case is an assignment of patients to donors. There are two major differences from the model of house allocation with existing tenants: (i) not all houses have tenants (ii) number of houses is more than the number of agents. Variants of top trading cycle algorithm has been proposed, and run in US hospital systems to match kidney patients to donors.

7 Single Peaked Domain of Preferences

We will now study another important setting where the Gibbard-Satterthwaite theorem does not apply. This is the setting where preferences of agents exhibit *single-peaked* property. To understand single-peaked preferences, consider an election with several candidates (possibly infinite). Candidates are ordered on a line so that candidate on left is the most leftist, and candidates become more and more right wing as we move to right. Now, it is natural to assume that every voter has an ideal political position. As one moves away from his ideal political position, either to left or to right, his preference decreases.

To be more precise, let $\{a, b, c\}$ be three candidates, with a to extreme left, b in the center, and c to extreme right. Now, suppose a voter's ideal position is b. Then, he likes b over a and b over c, but can have any preference over a and c. On the other hand, suppose a voter likes a the most. Then, the only possible ordering is a better than b better than c. Hence, when a is on top, c cannot be better than b. This restriction shows that this is a domain which is restricted, and the Gibbard-Satterthwaite theorem does not apply.

We now formally define the single-peaked preferences. Let $N = \{1, \ldots, n\}$ be the set of agents. Let A be a set of alternatives. All the results we state will hold for A finite or infinite. We assume $A = [0, 1]$. Consider the linear order $<$ induced by less than relation on $[0, 1]$. A preference ordering P_i (a linear order over the set of alternatives A) of agent i is **single peaked** with respect to $<$ if there exists an alternative $p_i \in A$, called the **peak**, such that

- for all $b, c \in A$ with $b < c < p_i$ we have $p_i P_i c$ and $c P_i b$, and
- for all $b, c \in A$ with $p_i > b > c$ we have $p_i P_i b$ and $b P_i c$.

So, preferences away from peak decreases, but no restriction is put for comparing alternatives when one of them is on the left to the peak, but the other one is on the right of the peak. We show some preference relations in Figure 6, and color the single-peaked ones in green.

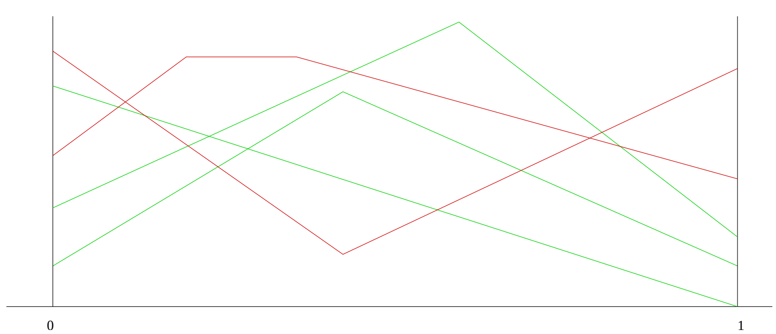

Figure 6: Examples of single-peaked preferences

Since we fix the order $<$ on $[0, 1]$, we will just say single-peaked preferences instead of single-peaked with respect to $<$. Note that if we have some finite set of alternatives, then single-peaked preference first maps each of the alternative onto $[0, 1]$, which induces the ordering $<$, and then single-peaked preferences can be defined in the usual way. We illustrate the idea with four alternatives $A = \{a, b, c, d\}$. Let us put the alternatives on $[0, 1]$ such that $a < b < c < d$. With respect to $<$, we give the permissible single peaked preferences in Table 13. There are sixteen more preference orderings that are not permissible here. For example, $bP_i dP_i aP_i c$ is not permissible since c, d are on the same side of peak, and in that case c is nearer to b than d is to b. So, $cP_i d$, which is not the case here.

a	b	b	b	c	c	c	d
b	a	c	c	d	b	b	c
c	c	d	a	b	a	d	b
d	d	a	d	a	d	a	a

Table 13: Single-peaked preferences

We now give some more examples of single-peaked preferences.

- An amount of public good (number of buses in the city) needs to be decided. Every agent has an optimal level of public good that needs to be consumed. The preferences decrease as the difference of the decided amount and optimal level increases.

- If we are locating a facility along a line then agents can have single-peaked preferences. For every agent, there is an optimal location along a line where he wants the facility, and the preference decreases as the distance from the optimal location increases in one direction.

- Something as trivial as setting the room temperature of a building by a group of agents exhibit single-peaked preferences. Everyone has an ideal temperature, and as the difference from the ideal temperature increases, the preference decreases.

Let $\mathcal{S}$ be the set of all single-peaked preferences. A social choice function f is a mapping $f : \mathcal{S}^n \rightarrow A$. An SCF f is manipulable by i at (P_i, P_{-i}) if there exists another single-peaked preference $\hat{P}_i$ such that $f(\hat{P}_i, P_{-i})P_i f(P_i, P_{-i})$. An SCF is strategy-proof if it is not manipulable by any agent at any preference profile.

7.1 Possibility Examples in Single-Peaked Domains

We start with an example to illustrate that many non-dictatorial social choice functions are strategy-proof in this setting. For any single-peaked preference ordering P_i, we let $P_i(1)$ to denote its peak. Now, consider the following SCF f: for every preference profile P, $f(P)$ is the minimal element with respect to $<$ among $\{P_1(1), P_2(1), \ldots, P_n(1)\}$. First, this is not a dictatorship since at every profile, a different agent can have its peak to the left. Second, it is strategy-proof. To see this, note that the agent whose peak coincides with the chosen alternative has no incentive to deviate. If some other agent deviates, then the only way to change the outcome is to place his peak to the left of chosen outcome. But that will lead to an outcome which is even more left to his peak, which he prefers less than the current outcome. Hence, no manipulation is possible.

One can generalize this further. Pick an integer $k \in \{1, \dots, n\}$. In every preference profile, the SCF picks the k-th lowest peak. Formally, $f(P_1, \dots, P_n)$ chooses among $\{P_1(1), \dots, P_n(1)\}$ the k-th lowest alternative according to $<$. To understand why this SCF is manipulable, note that those agents whose peak coincides with the k-th lowest peak have no incentive to manipulate. Consider an agent i, which lies to the left of k-th lowest peak. The only way he can change the outcome is to move to the right of the k-th lowest peak. In that case, an outcome which is even farther away from his peak will be chosen. According to single-peaked preferences, he prefers this less. A symmetric argument applies to the agents who are on to the right of k-th lowest peak.

7.2 Median Voter Result

We now define the notion of a *median voter*. Consider any sequence of points $(x_1, \dots, x_{2k+1})$ such that for all $j \in \{1, \dots, 2k+1\}$, we have $x_j \in A$. Now $b \in B$ is the median if $|\{x \in B : x < b \text{ or } x = b\}| \geq k+1$ and $|\{x \in B : x > b \text{ or } x = b\}| \geq k+1$. The median of a sequence of points B will be denoted as $med(B)$. Also, for any profile $(P_1, \dots, P_n)$, we denote the set of peaks as $peak(P) \equiv \{P_1(1), \dots, P_n(1)\}$.

Definition **9** *A social choice function $f : \mathcal{S}^n \to A$ is a* **median voter** *social choice function if there exists $B = \{y_1, \dots, y_{n-1}\} \subseteq A$ such that $f(P) = med(B \cup peak(P))$ for all preference profiles P. The alternatives in B are called the peaks of* **phantom voters**.

Note that by adding $(n-1)$ phantom voters, we have $(2n-1)$ (odd) peaks, and a median exists. We give an example to illustrate the ideas. Figure 7 shows the peak of 4 agents (in green). Then, we add 3 phantom voters, whose peaks are shown (in brown). The median voter SCF chooses the median of this set, which is shown to be the peak of the 3rd phantom voter in Figure 7.

Of course, the median voter SCF is a class of SCFs. A median voter SCF must specify the peaks of the phantom voters (it cannot change across profiles). We can simulate the k-th lowest peak social choice function that we described earlier by placing the phantom voters suitably. In particular, place peaks of $(n-k)$ phantom voters at 0 (or lowest alternative according to $<$) and the remaining $(k-1)$ peaks of phantom voters at 1 (or highest alternative according to $<$). It is clear that the median of this set lies at the kth lowest peak of agents.

Proposition **8** *Every median voter social choice function is strategy-proof.*

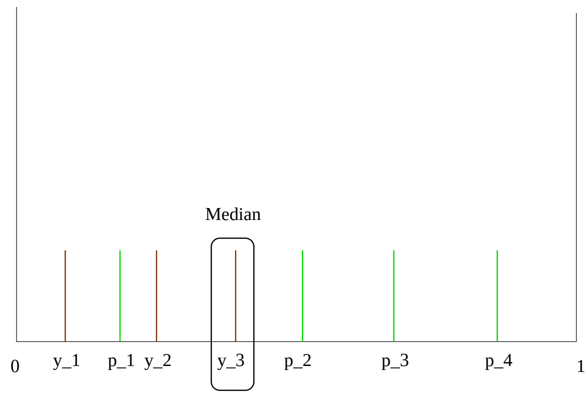

Figure 7: Phantom voters and the median voter

Proof: Consider any profile of single-peaked preferences $P = (P_1, \ldots, P_n)$. Let f be a median voter SCF, and $f(P) = a$. Consider agent i. Agent i has no incentive to manipulate if $P_i(1) = a$. Suppose agent i's peak is to the left of a. The only way he can change the outcome is by changing the median, which he can only do by changing his peak to the right of a. But that will shift the median to the right of a which he does not prefer to a. So, he cannot manipulate. A symmetric argument applies if i's peak is to the right of a. ■

One may wonder if one introduces an arbitrary number of phantom voters. Will the corresponding social choice function be still strategy-proof? We assume that whenever there are even number of agents (including the phantom voters), we pick the minimum of two medians. Along the lines of proof of Proposition 8, one can show that even this social choice function is strategy-proof.

We then ask what is unique about the median voter social choice function (where we take $n - 1$ phantom voters). We next intend to characterize the median voter social choice function.

7.3 Properties of Social Choice Functions

We first define some desirable properties of a social choice function. Most of these properties have already been discussed earlier for the Gibbard-Satterthwaite result.

Definition **10** *A social choice function $f : \mathcal{S}^n \rightarrow A$ is* **onto** *if for every $a \in A$, there exists a profile $P \in \mathcal{S}^n$ such that $f(P) = a$.*

Onto rules out constant social choice functions.

Definition **11** *A social choice function $f : \mathcal{S}^n \rightarrow A$ is* **unanimous** *if for every profile P with $P_1(1) = P_2(1) = \ldots = P_n(1) = a$ we have $f(P) = a$.*

Definition 12 *A social choice function $f : \mathcal{S}^n \rightarrow A$ is* **efficient** *if for every profile of preferences P and every $b \in A$, if there exists $a \neq b$ such that aP_ib for all $i \in N$, then $f(P) \neq b$.*

Denote by $[a, b]$, the set of all alternatives which lie between a and b (including a and b) according to $<$.

Lemma 6 *For every preference profile P, let p^{min} and p^{max} denote the smallest and largest peak (according to $<$) respectively in P. A social choice function $f : \mathcal{S}^n \rightarrow A$ is efficient if and only if for every profile P, $f(P) \in [p^{min}, p^{max}]$.*

Proof: Suppose f is efficient. Fix a preference profile P. If $f(P) < p^{min}$, then choosing p^{min} is better for all agents. Similarly, if $f(P) > p^{max}$, then choosing p^{max} is better for all agents. Hence, by efficiency, $f(P) \in [p^{min}, p^{max}]$. For the converse, if $f(P) \in [p^{min}, p^{max}]$, then any alternative other than $f(P)$ will move it away from either p^{min} or p^{max}. Hence, f is efficient. ■

Median voting with arbitrary number of phantom voters may be inefficient. Consider the median voting with $(3n - 1)$ phantom voters. Suppose we put all the phantoms at zero, and consider the instance where the peaks of the agents are arbitrarily close to 1. The outcome in this case is zero. But choosing one of the agents' peaks make every agent better off.

Definition 13 *A social choice function $f : \mathcal{S}^n \rightarrow A$ is* **monotone** *if for any two profiles P and P' with $f(P) = a$ and for all $b \neq a$, aP'_ib if aP_ib we have $f(P') = a$.*

Like in the unrestricted domain, strategy-proofness implies monotonicity.

Lemma 7 *If a social choice function $f : \mathcal{S}^n \rightarrow A$ is strategy-proof, then it is monotone.*

Proof: The proof is exactly similar to the necessary part of Theorem 1. We take two preference profiles $P, P' \in \mathcal{S}^n$ such that $f(P) = a$ and aP'_ib if aP_ib for all $b \neq a$. As in the proof of Theorem 1, we can consider P and P' to be different in agent j's preference ordering *only* (else, we construct a series of preference profiles each different from the previous one by just one agent's preference). Assume for contradiction $f(P') = b \neq a$.

If bP_ja, then agent j can manipulate at P by P'. Hence, aP_jb. But that means aP'_jb. In such a case, agent j will manipulate at P' by P. This is a contradiction. ■

Like in the unrestricted domain, some of these properties are equivalent in the presence of strategy-proofness.

Proposition 9 *Suppose $f : \mathcal{S}^n \rightarrow A$ is a strategy-proof social choice function. Then, f is onto if and only if it is unanimous if and only if it is efficient.*

Proof: Consider a strategy-proof social choice function $f : \mathcal{S}^n \to A$. We do the proof in three steps.

UNANIMITY IMPLIES ONTO. Fix an alternative $a \in A$. Consider a single peaked preference profile P where every agent has his peak at a. By unanimity, $f(P) = a$.

ONTO IMPLIES EFFICIENCY. Consider a preference profile P such that $f(P) = b$ but there exists a $a \neq b$ such that aP_ib for all $i \in N$. Since f is onto, there exists a profile P' such that $f(P') = a$. Consider another preference profile P'' such that the peaks of every agent is a, but the second ranked alternative is b - such a preference is possible in a single-peaked domain. By Lemma 7, f is monotone. By monotonicity, we get $f(P'') = f(P') = a$ and $f(P'') = f(P) = b$. This is a contradiction.

EFFICIENCY IMPLIES UNANIMITY. In any profile, where peaks are the same, efficiency will imply that the peak is chosen. ■

We now define a new property which will be crucial. For this, we need some definitions. A permutation of agents is denoted by a bijective mapping $\sigma : N \to N$. We apply a permutation σ to a profile P to construct another profile as follows: the preference ordering of agent i goes to agent $\sigma(i)$ in the new preference profile. We denote this new preference profile as P^σ.

Table 14 shows a pair of profiles, one of which is obtained by permuting the other. We consider $N = \{1, 2, 3\}$ and σ as $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$.

P_1	P_2	P_3	P_1^σ	P_2^σ	P_3^σ
a	b	b	b	a	b
b	a	c	c	b	a
c	c	a	a	c	c

Table 14: Example of permuted preferences

DEFINITION **14** *A social choice function $f : \mathcal{S}^n \to A$ is* **anonymous** *if for every profile P and every permutation σ such that $P^\sigma \in \mathcal{S}^n$, we have $f(P^\sigma) = f(P)$.*

Anonymous social choice functions require that the identity of agents are not important, and does not discriminate agents on that basis. Dictatorial social choice functions are not anonymous (it favors the dictator). Any social choice function which *ignores* the preferences of some agent is not anonymous.

7.4 Characterization Result

We show now that the only strategy-proof social choice function which is onto and anonymous is the median voter.

Theorem 9 *A strategy-proof social choice function is onto and anonymous if and only if it is the median voter social choice function.*

We discuss the necessity of all the properties. First, a dictatorial social choice function is onto and strategy-proof. So, anonymity is crucial in the characterization. Second, putting arbitrarily large number of phantoms at the lowest alternative according to $<$, and then taking the median is anonymous and strategy-proof, but it is not onto - it always selects the lowest alternative according to $<$. Hence, all the conditions are necessary in the result. We now give the proof.

Proof: It is clear that the median voter social choice function is strategy-proof (Proposition 8), onto (all the peaks in one alternative will mean that is the median), and anonymous (it does not distinguish between agents). We now show the converse.

Suppose $f : \mathcal{S}^n \rightarrow A$ is a strategy-proof, onto, and anonymous social choice function. The following two preference orderings are of importance for the proof:

- P_i^0: this is the unique preference ordering where the peak of agent i is at the lowest alternative according to $<$.
- P_i^1: this is the unique preference ordering where the peak of agent i is at the highest alternative according to $<$.

Finding the phantoms. For any $j \in \{1, \ldots, n-1\}$, define y_j as follows:

$$y_j = f(P_1^0, \ldots, P_{n-j}^0, P_{n-j+1}^1, \ldots, P_n^1).$$

So, y_j is the chosen alternative, when $(n-j)$ agents have their peak at the lowest alternative and the remaining j agents have their peak at the highest alternative. Notice that which of the j agents have their peaks at the highest alternative does not matter due to anonymity of f.

Further, we show that $y_j = y_{j+1}$ or $y_j < y_{j+1}$ for any $j \in \{1, \ldots, n-1\}$. To see this consider two profiles $P = (P_1^0, \ldots, P_{n-j}^0, P_{n-j+1}^1, \ldots, P_n^1)$ and $P' = (P_1^0, \ldots, P_{n-j-1}^0, P_{n-j}^1, \ldots, P_n^1)$. Only preference ordering of agent $k \equiv n-j$ is changing from P to P'. Note that $f(P) = y_j$ and $f(P') = y_{j+1}$. Since f is strategy-proof, $y_j P_k^0 y_{j+1}$. But the peak of agent k in P_k^0 is at the lowest alternative according to $<$. So, either $y_j = y_{j+1}$ or $y_j < y_{j+1}$.

Now, we consider a preference profile $P = (P_1, \ldots, P_n)$, where $P_i(1) = p_i$. We wish to show that

$$f(P) = med(p_1, \ldots, p_n, y_1, \ldots, y_{n-1}).$$

Assume without loss of generality (due to anonymity) that $p_1 \leq p_2 \leq \ldots \leq p_n$. We let $a = med(p_1, \ldots, p_n, y_1, \ldots, y_{n-1})$, and consider two possible cases.

Median is phantom peak. Suppose $a = y_j$ for some $j \in \{1, \ldots, n-1\}$. Since we are taking median of $2n-1$ points, and exactly $j-1$ phantom voters are to left of a and $n-j-1$ phantom voters to right (monotonicity of y_js), we must have $n-j$ agent peaks to the left and the remaining to the right. This means, $p_{n-j} \leq a = y_j \leq p_{n-j+1}$ due to monotonicity of agent peaks.

Now, we consider two preference profiles where preference ordering of agent 1 is different: $P' = (P^0_1, \ldots, P^0_{n-j}, P^1_{n-j+1}, \ldots, P^1_n)$ and $P'' = (P_1, P^0_2, \ldots, P^0_{n-j}, P^1_{n-j+1}, \ldots, P^1_n)$. Note that $f(P') = y_j$. Let $f(P'') = b$. Since f is strategy-proof, $y_j P^0_1 b$ or $y_j \leq b$. Also, strategy-proofness implies that $bP_1 y_j$. But $p_1 \leq y_j$. This implies that $b \leq y_j$. Hence, $b = y_j$.

Now, we consider another preference profile $P''' = (P_1, P_2, P^0_3, \ldots, P^0_{n-j}, P^1_{n-j+1}, \ldots, P^1_n)$, and repeat the previous argument for P'' and P'''. Repeating this way, we get $y_j = f(P_1, \ldots, P_{n-j}, P^1_{n-j+1}, \ldots, P^1_n)$.

Now, let $P' = (P_1, \ldots, P_{n-j}, P^1_{n-j+1}, \ldots, P^1_n)$ and $P'' = (P_1, \ldots, P_{n-j}, P^1_{n-j+1}, \ldots, P_n)$. By assumption, $y_j = f(P')$. Let $f(P'') = b$. Since f is strategy-proof, $y_j P^1_n b$, which implies that $y_j \geq b$. Again, applying f to be strategy-proof, we get that $bP_n y_j$. But, by assumption, $y_j \leq p_n$. This implies that $y_j \leq b$. This shows that $b = y_j$. Repeating this argument for all agents greater than j, we conclude that $f(P) = y_j$.

Median is agent peak. We do this part of the proof for two agents. Suppose $N = \{1, 2\}$. We first show a claim that shows *peaks-only* property of a strategy-proof and efficient social choice function.

Claim **1** *Suppose $N = \{1, 2\}$ and f is a strategy-proof and efficient social choice function. Let P and P' be two profiles such that $P_i(1) = P'_i(1)$ for all $i \in N$. Then, $f(P) = f(P')$.*

Proof: Consider preference profiles P and P' such that $P_1(1) = P'_1(1) = a$ and $P_2(1) = P'_2(1) = b$. Consider the preference profile (P'_1, P_2), and let $f(P) = x$ but $f(P'_1, P_2) = y$. By strategy-proofness, $xP_1 y$ and $yP'_1 x$. This implies, if x and y belong to the same side of a, then $x = y$. Then, the only other possibility is x and y belong to the different sides of a. We will argue that this is not possible. Assume without loss of generality $x < a < y$. Suppose, without loss of generality, $b < a$. Then, by efficiency (Lemma 6) at profile P'_1, P_2, we must have $y \in [b, a]$. This is a contradiction since $a < y$. Hence, it is not possible that x and y belong to the different sides of a. Thus, $x = y$ or $f(P_1, P_2) = f(P'_1, P_2)$.

Now, we can replicate this argument by going from (P'_1, P_2) to (P'_1, P'_2). This will show that $f(P'_1, P'_2) = x = f(P_1, P_2)$. ■

Now, consider a profile (P_1, P_2) such that $P_1(1) = a$, $P_2(1) = b$, and y_1 is the phantom peak. By our assumption, the median of (a, b, y_1) is an agent peak. Suppose that peak is a. Let $f(P_1, P_2) = c$. By efficiency, $c \in [a, b]$. Assume for contradiction that $c > a$. Consider another single-peaked preference ordering P'_1 for agent 1 such that $P'_1(1) = a = P_1(1)$ and $y_1 P'_1 c$ - this is possible since c and y_1 are on different sides of a. By Claim 1, $f(P'_1, P_2) = c$. Now, consider the preference profile (P^0_1, P_2). By definition, the median of $P^0_1(1)$, $P_2(1) = b$, and y_1 is y_1. By the earlier case of the proof, $f(P^0_1, P_2) = y_1$. Since $y_1 P'_1 c$, agent 1 will manipulate at (P'_1, P_2) via P^0_1. This is a contradiction since f is strategy-proof. ■

The peaks of the phantom voters reflect the degree of compromise the social choice function has when agents have *extreme* preferences. If j agents have the highest alternative as the peak, and the remaining $n - j$ agents have the lowest alternative as the peak, then which alternative is chosen? A true median will pick the peak which has more agents, but the median voter social choice function may do something intermediate.

7.5 Generalized Median Voter

We now define an extension of the median voter SCF, which may violate anonymity. It is called the *generalized median voter* social choice function. We assume that $A = [0, 1]$ as before.

Definition 15 *A social choice function f is a* **generalized median voter** *social choice function if there exists weights y_S for every $S \subseteq N$ satisfying*

1. $y_\emptyset = 0$, $y_N = 1$ *and*
2. $y_S \leq y_T$ *for all* $S \subseteq T$

such that for all preference profile P, $f(P) = \max_{S \subseteq N} z(S)$, where $z(S) = \min\{y_S, P_i(1) : i \in S\}$.

I leave it an exercise to find the weights which ensure that a generalized median voter SCF is a median voter SCF. We now discuss what y_S for any coalition reflects. Consider the following preference profile where all agents in S have their peaks at 1 and all agents in $N \setminus S$ have their peaks at 0. Fix a generalized median voter SCF with weights y_S for all $S \subseteq N$. Note that $z(T) = 0$ if $T \cap (N \setminus S) \neq \emptyset$. Further $z(T) = y_T$ for all $T \subseteq S$. Hence, the outcome at this profile is $\max_{T \subseteq S} z(T) = \max_{T \subseteq S} y_T = y_S$. In this sense, y_S reflects the degree of flexibility at extreme preference profiles.

An example to illustrate this SCF is given for $N = \{1, 2, 3\}$. Suppose $a = 0, b = 1, c = \frac{1}{2}$ and $p_1 = a$ but $p_2 = p_3 = b$. Suppose $y_{\emptyset} = 0$, $y_N = 1$, $y_1 = y_2 = y_3 = \frac{1}{3}$ and $y_{12} = \frac{1}{2}$, $y_{13} = \frac{3}{4}$, and $y_{23} = \frac{2}{3}$. Then $f(P)$ can be computed as follows: for every coalition S, we first compute $z(S) = \min\{y_S, P_i(1) : i \in S\}$. This is done as: $z(\emptyset) = 0$, $z(1) = 0$, $z(2) = \frac{1}{3}$, $z(3) = \frac{1}{3}$, $z(12) = 0$, $z(13) = 0$, $z(23) = \frac{2}{3}$, $z(123) = 0$. So, $f(P) = z(23) = y_{23} = \frac{2}{3}$.

Though, a generalized median voter SCF looks very different from median voter SCF, it is a generalization of the median voter SCF. First, note that if we impose anonymity on a generalized median voter SCF, then the weights have to respect $y_S = y_T$ if $|S| = |T|$. We prove the following.

Proposition **10** *Every median voter SCF is an anonymous generalized median voter SCF. Further, every anonymous generalized median voter SCF is a median voter SCF.*

Proof: Consider a median voter SCF f with phantom voters at $y_1 \leq y_2 \leq \ldots \leq y_{n-1}$. Now, define an anonymous generalized median voter SCF f' with weights such that $y'_S = y_{|S|}$. Note that f' is anonymous. Consider a profile $P \equiv (P_1, \ldots, P_n)$. We will show that $f(P) = f'(P)$. Note that $f(P)$ equals the median of $(y_1, \ldots, y_{n-1})$ and the peaks of the agents $(P_1(1), \ldots, P_n(1))$. We consider two possible cases.

Case 1: The median of $(y_1, \ldots, y_{n-1})$ and the peaks of the agents $(P_1(1), \ldots, P_n(1))$ is a phantom peak y_k. In that case, there are $(k-1)$ phantom peaks to the left of y_k and $(n-k-1)$ phantom peaks to the right of y_k. As a result, there are $(n-k)$ agent-peaks to the left of y_k and k agent-peaks to the right of y_k. Figure 8 shows this scenario.

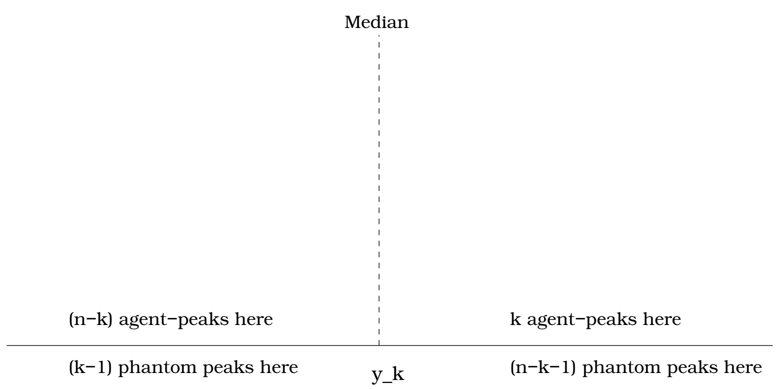

Figure 8: Median voter and generalized median voter - Case 1

Now consider a coalition S. If $|S| > k$, there is at least one agent in S whose peak is to the left of y_k. As a result, $z'(S) = \min(y'_S, \min_{i \in S} P_i(1))$ must be to the left of y_k. If $|S| < k$, $y'_S = y_{|S|} \leq y_k$. Hence, $z'(S)$ must be to the left of y_k. Now, consider a coalition S of k agents whose peaks are to the right of y_k - such a coalition exists by definition. Clearly,

$z'(S) = y_k$. Hence, $\max_{S \subseteq N} z'(S) = y_k$. So, $f'(P) = y_k = f(P)$.

CASE 2: The median of $(y_1, \ldots, y_{n-1})$ and the peaks of the agents $(P_1(1), \ldots, P_n(1))$ is an agent-peak $P_k(1)$. In that case, there are $(k-1)$ agent-peaks to the left of $P_k(1)$ and $(n-k)$ agent-peaks to the right of $P_k(1)$. As a result, there are $(n-k)$ phantom peaks to the left of $P_k(1)$ and $(k-1)$ agent-peaks to the right of $P_k(1)$. Figure 9 shows this scenario.

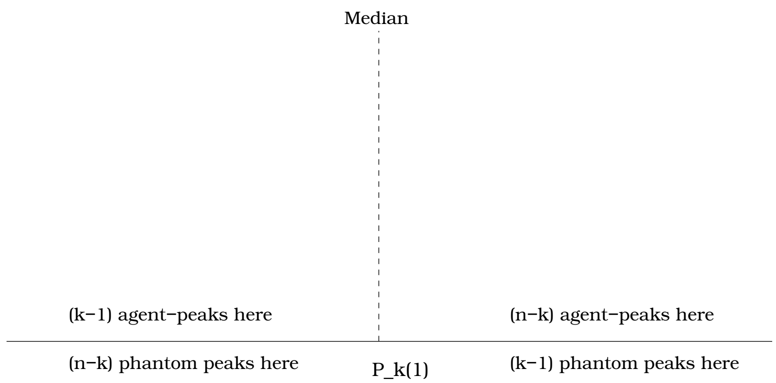

Figure 9: Median voter and generalized median voter - Case 2

Now consider a coalition S. If $|S| > (n-k+1)$, there is at least one agent in S whose peak is to the left of $P_k(1)$. As a result, $z'(S) = \min(y'_S, \min_{i \in S} P_i(1))$ must be to the left of $P_k(1)$. If $|S| < (n-k+1)$, then $y'_S = y_{|S|} \leq y_{n-k+1}$. Hence, $z'(S)$ must be to the left of $P_k(1)$. Now, consider a coalition S of $(n-k+1)$ agents whose peaks are to the right of $P_k(1)$ - such a coalition exists by definition. Clearly, $z'(S) = P_k(1)$. Hence, $\max_{S \subseteq N} z'(S) = P_k(1)$. So, $f'(P) = P_k(1) = f(P)$.

Consider an anonymous generalized voter SCF f'. Since f' is anonymous its weights must satisfy $y'_S = y'_T$ if $|S| = |T|$. From this, construct the median voter SCF f with weight of the k-th phantom as $y_k = y'_S$ for any coalition S. Using our previous argument, we can now conclude that for any preference profile P, $f'(P) = f(P)$. ■

We next establish that every generalized median voter SCF is strategy-proof and unanimous.

PROPOSITION **11** *Every generalized median voter SCF is strategy-proof and unanimous.*

Proof: Consider a generalized median voter SCF f. Consider a preference profile P with peaks of agents $(p_1, \ldots, p_n)$. Fix an agent $i \in N$. If $p_i = f(P) = z(S)$ for some $S \subseteq N$, then agent i has no incentive to manipulate. We consider the case when $p_i \neq f(P)$. Suppose $f(P) = z(S)$ for some $S \subseteq N$. Hence, $z(S) \geq z(T)$ for all $T \subseteq N$. Note that f only depends on the peaks of the agents. We consider two cases.

- Suppose $p_i = z(T)$ for some $T \subseteq N$. Since $p_i \neq z(S)$, we have $p_i = z(T) < z(S)$. This implies that $i \notin S$. By reporting a lower peak $p'_i < p_i$, agent i cannot change the outcome. By reporting a higher peak, the outcome will become $> z(S)$, which the agent prefers less than $z(S)$ due to single-peakedness.

- Suppose $p_i \neq z(T)$ for all $T \subseteq N$ such that $i \in T$. Then, $p_i > z(T)$ for all $T \subseteq N$ such that $i \in T$. Then, by reporting a higher peak $p'_i > p_i$, agent i cannot change the outcome. By reporting a lower peak $p'_i < p_i$, if the outcome changes, then it must change to p'_i. In that case, $i \in S$ - because if $i \notin S$, by reporting lower peak, $z(S) \geq z(T)$ for all $T \subseteq N$, and hence the outcome does not change. If the outcome changes to p'_i, then $p'_i < z(S)$. Since $i \in S$, we have $z(S) < p_i$. Hence, i likes $z(S)$ to p'_i due to single-peakedness.

To see unanimity, consider a preference profile P, where every agent has the peak at p. For any $S \subseteq N$, we have $z(S) = \min(y_S, p)$ and $z(N) = p$. But $z(N) \geq z(S)$. So, $p = z(N) \geq z(S)$ for all $S \subseteq N$. Hence, $f(P) = p$. ■

Note that the proof of strategy-proofness of generalized median voter SCF did not use the properties in its definition. Hence, the strategy-proofness result holds for even more general SCFs which do not necessarily satisfy these properties. However, such SCFs will violate unanimity (or onto property).

It is easy to see that these rules are not anonymous. Some coalitions have more weight than others. In particular, a singleton coalition may have more weight than other singleton coalition. However, the following theorem is known to be true. The proof of this theorem is skipped.

Theorem **10** *A social choice function is strategy-proof and onto if and only if it is a generalized median voter social choice function.*

7.6 Extension - Private Good Allocation

Here, we consider an extension of single-peaked domain. The single-peaked domain can be considered to be an instance of a domain where a public good is being allocated. Consider a problem where an infinitely divisible private good has to be allocated among n agents.

Suppose each agent (country) i has to be allocated a share $s_i \in [0, 1]$ of a total quota of carbon permits. It is reasonable to assume that the utility of a country increases till a certain point of carbon emission and falls there after. For example, country i may have utility $k_i\sqrt{s_i}$ but cost $\theta_i s_i$ from emitting s_i amount of carbon to its environment. In such a case, the net utility function is $k_i\sqrt{s_i} - \theta_i s_i$, which is single-peaked. In that case, each country has a preference ordering over the set $[0, 1]$ which exhibits a single-peaked structure. However, as we show next, this does not translate to a single-peaked preference ordering over the set of alternatives. Hence, earlier results cannot be applied.

Here, every alternative is a vector $s = (s_1, \ldots, s_n)$ such that $\sum_{i \in N} s_i = 1$ and $s_i \geq 0$ for all $i \in N$. So, the set of alternatives is

$$A = \{(s_1, \ldots, s_n) : s_i \geq 0 \ \forall \ i \in N, \ \sum_{i \in N} s_i = 1\}.$$

Suppose agents preferences are not known (but only known to be single-peaked), and agents care only about their own shares. If agents only care about their own shares, then the preferences over A cannot be single-peaked because two alternatives with same share to an agent must be same for that agent. Hence, the earlier results on single-peaked domains do not apply.

However, for various possible shares of agent i, that agent has a preference ordering $\succ_i$ over $[0, 1]$ with a peak at $p_i(\succ_i)$, and single-peaked. Denote by $\mathcal{S}$ the set of all single-peaked preferences over $[0, 1]$. A social choice function is a mapping $f : \mathcal{S}^n \to A$. The share allocation to agent i at preference profile $\succ$ is denoted by $f_i(\succ) \in [0, 1]$.

We first look at the implication of *efficiency* in this setting. Without formally defining it, an allocation is efficient if there does not exist another allocation which makes everyone better off with at least one agent getting strictly better. If $\sum_{i \in N} p_i(\succ_i) = 1$, then efficiency directs us to allocate $p_i(\succ_i)$ to agents i for all $i \in N$. If $\sum_{i \in N} p_i(\succ_i) > 1$, then for some agent $k \in N$, $f_k(\succ) < p_k(\succ_k)$. In that case, no agent $j \neq k$ must be getting $f_j(\succ) > p_j(\succ_j)$. Because in that case, decreasing agent j's share and increasing agent k's share makes both of them better off. So, in this case, we must have $f_j(\succ) \leq p_j(\succ_j)$ for all $j \in N$. Similarly, if $\sum_{i \in N} p_i(\succ_i) < 1$, we must have $f_j(\succ) \geq p_j(\succ_j)$ for all $j \in N$.

Consider the **uniform rule** social choice function f^u as defined below. For any profile $\succ$, we define for every $i \in N$,

$$\begin{aligned}
f_i^u(\succ) &= p_i(\succ_i) \qquad \text{if } \sum_{i \in N} p_i(\succ_i) = 1 \\
f_i^u(\succ) &= \max(p_i(\succ_i), \mu(\succ)) \qquad \text{if } \sum_{i \in N} p_i(\succ_i) < 1 \\
f_i^u(\succ) &= \min(p_i(\succ_i), \lambda(\succ)) \qquad \text{if } \sum_{i \in N} p_i(\succ_i) > 1,
\end{aligned}$$

where $\mu(\succ)$ solves $\sum_{i \in N} \max(p_i(\succ_i), \mu(\succ)) = 1$ in the second case and $\lambda(\succ)$ solves $\sum_{i \in N} \min(p_i(\succ_i), \lambda(\succ)) = 1$ in the third case. It can be verified that these quantities have a unique solution.

The uniform rule SCF has a nice interpretation. Every agent has a bucket of 1 unit capacity. There is a mark at $p_i(\succ_i)$ in every bucket i. If the sum of these marks are equal to 1, we fill the buckets with water till their marks. If the sum of these marks are greater than 1, we fill water in the buckets at equal rate, till one of the buckets hits the mark. We stop filling that bucket, but fill the other buckets at equal rate, till we hit another mark, and so on till the sum of water in the buckets is 1. The water level in the buckets at the end indicate the final allocation.

When sum of the marks is less than 1, we fill the buckets completely and empty them uniformly till the sum is equal to 1. We stop filling a bucket once we hit the mark. Then, we continue emptying the other buckets at a uniform rate, and so on till the sum of water in the buckets is 1. The water level in the buckets at the end indicate the final allocation.

PROPOSITION **12** *The uniform rule social choice function is efficient, anonymous, and strategy-proof.*

Proof: The uniform rule social choice function is anonymous since only the peaks of agents matter but not the identity of the "owners" of peaks. Consider a preference profile $\succ$. Efficiency is equivalent to verifying the following two cases.

- When $\sum_{i \in N} p_i(\succ_i) < 1$, then $f^u_i(\succ) \geq p_i(\succ_i)$ for all $i \in N$. This is true because in this case, we empty the buckets uniformly, and stop as soon as a bucket hits the peak.
- When $\sum_{i \in N} p_i(\succ_i) > 1$, then $f^u_i(\succ) \leq p_i(\succ_i)$ for all $i \in N$. This is true because in this case, we fill the buckets uniformly, and stop as soon as a bucket hits the peak.
- When $\sum_{i \in N} p_i(\succ_i) = 1$, then $f^u_i(\succ) = p_i(\succ_i)$ for all $i \in N$. This is true by definition of f^u.

To verify strategy-proofness, consider agent j at a preference profile $\succ$. We consider three cases separately.

- If $\sum_{i \in N} p_i(\succ_i) = 1$, then $f^u_j(\succ) = p_j(\succ_j)$. Hence, agent j has no incentive to manipulate.
- If $\sum_{i \in N} p_i(\succ_i) < 1$, then $f^u_j(\succ) \geq p_j(\succ_j)$. He will like to manipulate if $f^u_j(\succ) > p_j(\succ_j)$. Since f^u only depends on the peaks of agents, the only way to manipulate is to change the peak. Suppose agent j reports $\succ'_j$ with peak $p_j(\succ'_j)$. If $p_j(\succ'_j) \leq f^u_j(\succ)$, then we will still have $\sum_{i \neq j} p_i(\succ_i) + p_j(\succ'_j) < 1$. Since we will again empty the buckets, the outcome will not change. If $p_j(\succ'_j) > f^u_j(\succ)$, the share of j will only increase, which he prefers less. To see why j's share will increase we consider two cases.
  - $\sum_{i \neq j} p_i(\succ_i) + p_j(\succ'_j) < 1$. In this case, $f^u_j(\succ'_j, \succ_{-j}) \geq p_j(\succ'_j) > f^u_j(\succ)$.
  - $\sum_{i \neq j} p_i(\succ_i) + p_j(\succ'_j) > 1$. In this case, we fill the buckets. If we hit the peak of agent j, then clearly $f^u_j(\succ'_j, \succ_{-j}) = p_j(\succ'_j) > f^u_j(\succ)$. If we do not hit the peak, then it is the highest share amongst all agents in profile $(\succ'_j, \succ_{-j})$. More specifically, $1 = \sum_{i \neq j} f^u_i(\succ'_j, \succ_{-j}) + f^u_j(\succ'_j, \succ_{-j}) \leq n f^u_j(\succ'_j, \succ_{-j})$. On the other

hand, since $f^u_j(\succ) > p_j(\succ_j)$, it is the lowest share amongst all agents in profile $(\succ)$. More specifically, $1 = \sum_{i \neq j} f^u_i(\succ) + f^u_j(\succ) \geq n f^u_j(\succ)$. Hence, we get $f^u_j(\succ'_j, \succ_{-j}) \geq f^u_j(\succ)$.

- If $\sum_{i \in N} p_i(\succ_i) > 1$, then $f^u_i(\succ) \leq p_i(\succ_i)$ for all $i \in N$. Using a symmetric argument as the previous case, we can show that no agent j can manipulate.

■

The converse of Proposition 12 is known to be true also. In particular, the following theorem is true, whose proof is skipped.

THEOREM **11** *A social choice function is strategy-proof, efficient, and anonymous if and only if it is the uniform rule.*

8 Randomized Social Choice Function

Randomization is a way of expanding the set of possible strategy-proof social choice function. Lotteries are also common in practice. So, it makes sense to study the effects of randomization on strategy-proofness. Here, we discuss the model in the Gibbard-Satterthwaite theorem but with randomization.

As before let $A = \{a, b, c, \ldots\}$ be a finite set of alternatives with $|A| = m$ and $N = \{1, \ldots, n\}$ be the set of agents. Let $\mathcal{L}(A)$ denote the set of all probability distributions over A. We will refer to this set as the set of **lotteries** over A. A particular element $\lambda \in \mathcal{L}(A)$ is a probability distribution over A, and λ_a denotes the probability of alternative a. Of course $\lambda_a \geq 0$ for all $a \in A$ and $\sum_{a \in A} \lambda_a = 1$.

As before, every agent i has a linear order over A, which is his preference ordering, and $\mathcal{P}$ is the set of all linear orders over A. A **randomized social choice function (RSCF)** f is a mapping $f : \mathcal{P}^n \to \mathcal{L}(A)$. We let $f_a(P)$ to denote the probability of alternative a being chosen at profile P. To avoid confusion, we refer to $f : \mathcal{P}^n \to A$ as a **deterministic** social choice function (DSCF).

8.1 Defining Strategy-proof RSCF

There are several meaningful ways to define strategy-proofness in this setting. We follow one of the first-proposed approaches (by Gibbard). It requires that an RSCF be non-manipulable for every *utility representation* of linear orders.

A utility function $u : A \to \mathbb{R}$ represents a preference ordering $P_i \in \mathcal{P}$ if for all $a, b \in A$, $u(a) > u(b)$ if and only if $a P_i b$.

An RSCF is strategy-proof if for every possible representation of orderings, the expected utility of telling the truth is not less than the expected utility of lying.

Definition **16** *An RSCF $f : \mathcal{P}^n \to \mathcal{L}(A)$ is* **strategy-proof** *if for all $i \in N$, all $P_{-i} \in \mathcal{P}^{n-1}$, for all $P_i \in \mathcal{P}$, and for all utility functions $u : A \to \mathbb{R}$ representing P_i, we have*

$$\sum_{a \in A} u(a) f_a(P_i, P_{-i}) \geq \sum_{a \in A} u(a) f_a(P'_i, P_{-i}) \qquad \forall\ P'_i \in \mathcal{P}.$$

For the strategy-proofness of DSCF, we did not require this utility representation. However, it is easy to verify that a DSCF is strategy-proof in the sense of Definition 6 if and only if it is strategy-proof in the sense of Definition 16.

It is well known that such a notion is equivalent to first-order stochastic dominance. To define this formally, let $B(a, P_i) = \{b \in A : b = a \text{ or } bP_ia\}$.

Definition **17** *An RSCF $f : \mathcal{P}^n \to \mathcal{L}(A)$ is* **strategy-proof** *if for all $i \in N$, all $P_{-i} \in \mathcal{P}^{n-1}$, for all $P_i \in \mathcal{P}$, and for all $a \in A$, we have*

$$\sum_{b \in B(a,P_i)} f_b(P_i, P_{-i}) \geq \sum_{b \in B(a,P_i)} f_b(P'_i, P_{-i}) \qquad \forall\ P'_i \in \mathcal{P}.$$

To understand this definition a little better let us take an example with two agents $\{1, 2\}$ and three alternatives $\{a, b, c\}$. The preference of agent 2 is fixed at P_2 given by aP_2bP_2c. Let us consider two preference orderings of agent 1: $P_1 : bP_1cP_1a$ and $P'_1 : cP_1aP_1b$. Denote $P = (P_1, P_2)$ and $P' = (P'_1, P_2)$. Suppose $f_a(P) = 0.6$ and $f_b(P) = 0.1$ and $f_c(P) = 0.3$. First order stochastic dominance requires the following.

$$f_b(P) = 0.1 \geq f_b(P')$$
$$f_b(P) + f_c(P) = 0.4 \geq f_b(P') + f_c(P').$$

We now introduce the notion of unanimity in this model. It is the exact version of unanimity we used in the deterministic social choice functions.

Definition **18** *An RSCF $f : \mathcal{P}^n \to \mathcal{L}(A)$ satisfies* **unanimity** *if for all $i \in N$, all $P \in \mathcal{P}^n$ such that $P_1(1) = P_2(1) = \ldots = P_n(1) = a$, we have $f_a(P) = 1$.*

As in the deterministic SCF case, we can see that the constant social choice function is not unanimous. But there is even a bigger class of RSCFs which are strategy-proof but not unanimous.

Definition **19** *An RSCF f is a* **unilateral** *if there exists an agent i and $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_{|A|}$ with $\alpha_j \in [0, 1]$ and $\sum_{j=1}^{|A|} \alpha_j = 1$ such that for all P we have $f_{P_i(j)} = \alpha_j$ for all $j \in \{1, \ldots, |A|\}$.*

In a unilateral RSCF, there is a **weak dictator** i such that top ranked alternative of i gets probability α_1, second ranked alternative of i gets probability α_2, and so on. Notice that every unilateral is strategy-proof, but not unanimous.

We now define another broad class of RSCFs which are strategy-proof and unanimous.

Definition **20** *An RSCF $f : \mathcal{P}^n \to \mathcal{L}(A)$ is a* **random dictatorship** *if there exists weights $\beta_1, \ldots, \beta_n \in [0,1]$ with $\sum_{i \in N} \beta_i = 1$ such that for all $P \in \mathcal{P}^n$,*

$$f_a(P) = \sum_{i \in N : P_i(1) = a} \beta_i.$$

If a particular agent i has $\beta_i = 1$, then such a random dictatorship is the usual dictatorship. A random dictatorship can be thought to be a randomization over deterministic dictatorships, where β_i reflects the probability with which agent i is a dictator. For example, if $N = \{1,2,3\}$ and $A = \{a,b,c\}$ and $\beta_1 = \frac{1}{2}$, $\beta_2 = \beta_3 = \frac{1}{4}$, then at a profile P where $P_1(1) = a, P_2(1) = a, P_3(1) = c$, the output of this random dictatorship will be $f_a(P) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ and $f_c(P) = \frac{1}{4}$.

Random dictatorship can be thought of as a convex combination of dictatorships, where β_i is the probability with which agent i is the dictator. Since dictatorship is strategy-proof, one can show that random dictatorship is also strategy-proof. We show a general result on strategy-proofness RSCFs which can be expressed as a convex combination of other strategy-proof RSCFs.

Proposition **13** *Let $f^1, f^2, \ldots, f^k$ be a set of k strategy-proof RSCFs. Let $f : \mathcal{P}^n \to \mathcal{L}(A)$ be defined as: for all $P \in \mathcal{P}^n$ and for all $a \in A$, $f_a(P) = \sum_{j=1}^{k} \lambda_j f^j_a(P)$, where $\lambda_j \in [0,1]$ for all $j \in \{1, \ldots, k\}$ and $\sum_{j=1}^{k} \lambda_j = 1$. Then, f is strategy-proof.*

Proof: Fix an agent i and a profile P_{-i}. For some preference P_i consider a utility representation $u : A \to \mathbb{R}$. Then, for any P'_i,

$$\begin{aligned}
\sum_{a \in A} u(a) f_a(P) &= \sum_{a \in A} u(a) \sum_{j=1}^{k} \lambda_j f^j_a(P) = \sum_{j=1}^{k} \lambda_j \sum_{a \in A} u(a) f^j_a(P) \\
&\geq \sum_{j=1}^{k} \lambda_j \sum_{a \in A} u(a) f^j_a(P'_i, P_{-i}) = \sum_{a \in A} u(a) \sum_{j=1}^{k} \lambda_j f^j_a(P'_i, P_{-i}) \\
&= \sum_{a \in A} u(a) f_a(P'_i, P_{-i}).
\end{aligned}$$

■

Another way to interpret Proposition 13 is that the set of strategy-proof RSCFs form a convex set. As a corollary of Proposition 13, we get the following.

Corollary 1 *Every random dictatorship is strategy-proof.*

Proof: A random dictatorship is a convex combination of dictatorships. Hence, it is strategy-proof by Proposition 13. ■

We are now ready to state the counterpart of the Gibbard-Satterthwaite theorem for RSCFs. This was proved by Gibbard.

Theorem **12** *Suppose* $|A| \geq 3$*. An RSCF is unanimous and strategy-proof if and only if it is a random dictatorship.*

The proof of this theorem is more involved than the Gibbard-Satterthwaite theorem. We only do the case with two agents.

Proof: We have already shown that a random dictatorship is strategy-proof (Corollary 1). It is also unanimous - if all agents have the same alternative as top ranked, βs will sum to 1 for that alternative. We now prove that any RSCF which is unanimous and strategy-proof must be a random dictatorship for $n = 2$ case. We do the proof by showing two claims. Let f be a strategy-proof and unanimous RSCF.

Claim **2** *Let* $P \in \mathcal{P}^2$ *be a preference profile such that* $P_1(1) \neq P_2(1)$*. If* $f_a(P) > 0$ *then* $a \in \{P_1(1), P_2(1)\}$*.*

Proof: Consider a preference profile P such that $P_1(1) = a \neq b = P_2(1)$. Let $f_a(P) = \alpha$ and $f_b(P) = \beta$. Consider a preference ordering P_1' such that $P_1'(1) = P_1(1) = a$ and $P_1'(2) = P_2(1) = b$. Similarly, consider a preference ordering P_2' such that $P_2'(1) = P_2(1) = b$ and $P_2'(2) = P_1(1) = a$.

Strategy-proofness implies that $f_a(P_1', P_2) = \alpha$. Also, by unanimity the outcome at (P_2, P_2) is b. So, strategy-proofness implies that $f_a(P_1', P_2) + f_b(P_1', P_2) \geq f_a(P_2, P_2) + f_b(P_2, P_2) = 1$. Hence, $f_a(P_1', P_2) + f_b(P_1', P_2) = 1$.

Using a symmetric argument, we can conclude that $f_b(P_1, P_2') = \beta$ and $f_a(P_1, P_2') + f_b(P_1, P_2') = 1$.

Strategy-proofness implies that $f_b(P_1', P_2') = f_b(P_1', P_2) = 1 - \alpha$. and $f_a(P_1', P_2') = f_a(P_1, P_2') = 1 - \beta$. But $f_a(P_1', P_2') + f_b(P_1', P_2') \leq 1$ implies that $\alpha + \beta \geq 1$ and $f_a(P) + f_b(P) \leq 1$ implies $\alpha + \beta \leq 1$. Hence, $\alpha + \beta = 1$. ■

Claim **3** *Let* $P, \bar{P} \in \mathcal{P}^2$ *be such that* $P_1(1) = a \neq b = P_2(1)$ *and* $\bar{P}_1(1) = c \neq d = \bar{P}_2(1)$*. Then* $f_a(P) = f_c(\bar{P})$ *and* $f_b(P) = f_d(\bar{P})$*.*

Proof: We consider various cases.

Case 1: $c = a$ and $d = b$. Strategy-proofness implies that $f_a(P_1, P_2) = f_a(\bar{P}_1, P_2)$. By Claim 2, $f_a(P_1, P_2) + f_b(P_1, P_2) = f_a(\bar{P}_1, P_2) + f_b(\bar{P}_1, P_2) = 1$. Hence, $f_b(P_1, P_2) = f_b(\bar{P}_1, P_2)$. Repeating this argument for agent 2 while going from $(\bar{P}_1, P_2)$ to $(\bar{P}_1, \bar{P}_2)$, we get that $f_a(\bar{P}) = f_a(P)$ and $f_b(\bar{P}) = f_b(P)$.

Case 2: $c = a$ or $d = b$. Suppose $c = a$. Consider a preference profile $(P_1, \hat{P}_2)$ such that $\hat{P}_2(1) = d \notin \{a, b\}$ and $\hat{P}_2(2) = b$. Assume without loss of generality that $P_2(1) = b$ and $P_2(2) = d$. Then, strategy-proofness implies that $f_b(P_1, \hat{P}_2) + f_d(P_1, \hat{P}_2) = f_b(P) + f_d(P)$. By Claim 2, $f_b(P_1, \hat{P}_2) = f_d(P) = 0$. Hence, $f_b(P) = f_d(P_1, \hat{P}_2)$. This further implies that $f_a(P) = f_a(P_1, \hat{P}_2)$. By Case 1, $f_a(P) = f_a(\bar{P})$ and $f_b(P) = f_d(\bar{P})$. An analogous proof works if $d = b$.

Case 3: $c = b$ and $d \notin \{a, b\}$. Let $\hat{P} = (P_1, \bar{P}_2)$. By Case 2, $f_a(P) = f_a(\hat{P})$ and $f_b(P) = f_d(\hat{P})$. Again, applying Case 2, we get $f_a(P) = f_a(\hat{P}) = f_b(\bar{P})$ and $f_b(P) = f_d(\hat{P}) = f_d(\bar{P})$.

Case 4: $c \notin \{a, b\}$ and $d = a$. A symmetric argument to Case 3 can be made.

Case 5: $c = b$ and $d = a$. Since there are at least three alternatives there is a $x \notin \{a, b\}$. We construct a profile $\hat{P} = (\hat{P}_1, \bar{P}_2)$ such that $\hat{P}_1(1) = x$. By Case 4, $f_x(\hat{P}) = f_a(P)$ and $f_b(P) = f_a(\hat{P})$. Now, applying Case 2, we can conclude that $f_x(\hat{P}) = f_b(\bar{P})$ and $f_a(\hat{P}) = f_a(\bar{P})$.

Case 6: $c \notin \{a, b\}$ and $d \notin \{a, b\}$. Consider a profile $\hat{P} = (\hat{P}_1, P_2)$ such that $\hat{P}_1(1) = c$. By Case 2, $f_c(\hat{P}) = f_a(P)$ and $f_b(\hat{P}) = f_b(P)$. Applying Case 2 again, we get $f_c(\bar{P}) = f_c(\hat{P}) = f_a(P)$ and $f_d(\bar{P}) = f_b(\hat{P}) = f_b(P)$. ■

Claims 2 and 3 establishes that f is a RSCF. ■

As we have seen a unilateral SCF is not unanimous but strategy-proof. Hence, unanimity is a crucial assumption in Theorem 12.

References