Notes on Social Choice Theory

Arunava Sen*

February 21, 2017

1 Binary Relations and Orderings

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. It is conventional to require R_i to satisfy the following assumptions.

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 $a, b, c \in A$, $\big[aR_ib \text{ and } bR_ic\big] \Rightarrow \big[aR_ic\big]$.

   We will denote the set of all orderings over A as $\mathbf{R}$.

2. Binary Relation: A preference relation R_i of agent i is called a **binary relation** if it satisfies completeness and reflexivity. Hence, a binary relation gives unordered pairs of A. 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 $\sim (bQ_ia)$. Informally, $\hat{Q}_i$ is the strict part of Q_i, whereas $\bar{Q}_i$ is the weak part of Q_i. Sometimes, we will refer to the symmetric component of a preference relation R_i as I_i and asymmetric

*Planning Unit, Indian Statistical Institute, 7 Shahid Jit Singh Marg, New Delhi 110016, India. I am grateful to Debasis Mishra for help in preparing these notes.

component as P_i. We define transitivity of $\hat{Q}_i$ and $\bar{Q}_i$ in the usual way, i.e. $\hat{Q}_i$ is transitive if for all $a, b, c \in A$, $\left[a\hat{Q}_i b \text{ and } b\hat{Q}_i c\right] \Rightarrow \left[a\hat{Q}_i c\right]$. Similarly for $\bar{Q}_i$.

The asymmetric and symmetric components of an ordering R_i will be denoted by P_i and I_i respectively.

PROPOSITION **1** *Let R_i be an ordering. Then P_i and I_i are transitive. Conversely, suppose Q_i is a binary relation such that $\hat{Q}_i$ and $\bar{Q}_i$ are transitive. Then Q_i is an ordering.*

Proof: Consider $a, b, c \in A$ and and ordering R_i such that aP_ib and bP_ic. Assume by way of contradiction that $\sim (aP_ic)$. Since R_i is an ordering, it is complete. Hence, aR_ic or cR_ia holds. Since $\sim (aP_ic)$, 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.

Now consider $a, b, c \in A$ and a binary relation Q_i such that aQ_ib and bQ_ic. We have to show that aQ_ic. If $a\hat{Q}_ib$ and $b\hat{Q}_ic$, then $a\hat{Q}_ic$ holds because of the transitivity of $\hat{Q}_i$. Hence aQ_ic. The argument for the case where $a\bar{Q}_ib$ and $b\bar{Q}_ic$ is analogous. The two remaining cases are (i) $a\hat{Q}_ib$ and $b\bar{Q}_ic$ and (ii) $a\bar{Q}_ib$ and $b\hat{Q}_ic$. Suppose (i) holds but $\sim (aQ_ic)$, i.e cQ_ia. If $c\hat{Q}_ia$, then the transitivity of $\hat{Q}_i$ implies $c\hat{Q}_ib$ which contradicts the assumption that $b\bar{Q}_ic$. If $c\bar{Q}_ia$, then the transitivity of $\bar{Q}_i$ implies $b\bar{Q}_ia$ which contradicts the assumption that $a\hat{Q}_ib$.

Case (ii) can be dealt with analogously. ■

DEFINITION **1** *A* **quasi-ordering** *is a binary relation Q_i whose asymmetric component is transitive.*

REMARK: The symmetric component of a quasi-ordering need not be transitive. Hence, a quasi-ordering is not an ordering. Indeed, in many situations it is natural to regard the "indifference" relation to be intransitive - for instance, an agent may be indifferent between Rs x and Rs $x + \epsilon$ ($\epsilon > 0$ and ϵ very small). Transitivity would imply the agent is indifferent between x and $x + \Delta$ for arbitrarily large Δ which is implausible.

DEFINITION **2** *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**.

REMARK: If R_i is anti-symmetric then its asymmetric component P_i is complete.

2 Arrovian Social Welfare Functions

DEFINITION **3** *An* **Arrovian social welfare function** *(ASWF) F is a mapping $F : \mathbf{R}^n \to \mathbf{R}$.*

A typical element of the set $\mathbf{R}^n$ will be denoted by $R \equiv (R_1, \ldots R_n)$ and will be referred to as a **preference profile**.

We give several examples of well-known social welfare functions.

2.1 Scoring Rules

For simplicity assume that individual orderings R_i are linear. Let $\#A = p$ and $s = (s_1, \ldots, s_p)$, where $s_1 \geq \ldots \geq s_p \geq 0$ and $s_1 > s_p$. The vector s is called a scoring vector. For all $i \in N$, $R_i \in \mathbf{R}$, $a \in A$, define the rank of a in R_i as

$$r(a, R_i) = \#\{b \in A \setminus \{a\} : bP_i a\} + 1$$

The score of rank $r(a, R_i)$ is $s_{r(a,R_i)}$. For every profile $R \in \mathbf{R}^n$ compute the score of alternative $a \in A$ as

$$s(a, R) = \sum_{i \in N} s_{r(a,R_i)}$$

The **scoring rule** F^s is defined as for all $a, b \in A$, for all $R \in \mathbf{R}^n$ we have $aF^s(R)b$ if and only if $s(a, R) \geq s(b, R)$. It is easy to see that F^s defines an ordering. Here are some special cases of the scoring rule.

- Plurality Rule: This is the scoring rule when $s = (1, 0, 0, \ldots, 0)$.
- Borda Rule: This is the scoring rule when $s = (p - 1, p - 2, \ldots, 1, 0)$.
- Anti-plurality Rule: This is the scoring rule when $s = (1, 1, \ldots, 1, 0)$.

2.2 Majority Rules

For every $R \in \mathbf{R}^n$ define the binary relation $Q^{maj}(R)$ as follows: for all $a, b \in A$ we have $aQ^{maj}(R)b$ if and only if $\#\{i \in N : aR_ib\} \geq \#\{i \in N : bR_ia\}$.

Proposition **2 (Condorcet Paradox)** *There exists R for which $Q^{maj}(R)$ is not a quasi-ordering, and hence not an ordering.*

Proof: Let $N = \{1, 2, 3\}$ and $A = \{a, b, c\}$. Consider the preference profile in Table **??**, where every agent has a linear ordering. Verify that $\{i \in N : aR_ib\} = \{1, 2\}$, $\{i \in N : bR_ic\} = \{1, 3\}$, and $\{i \in N : cR_ia\} = \{2, 3\}$. Hence, $a\hat{Q}^{maj}(R)b$, $b\hat{Q}^{maj}(R)c$, and $c\hat{Q}^{maj}(R)a$. This means that $\hat{Q}^{maj}(R)$ is not an ordering.

■

The proposition above demonstrates that the majority rule procedure (the map which associates $Q^{maj}(R)$ with every profile R) is not a ASWF.

R_1	R_2	R_3
a	c	b
b	a	c
c	b	b

Table 1: Condorcet Cycle

2.3 Oligarchies

Let $R \in \mathbf{R}^n$ be a preference profile and let $\emptyset \neq G \subseteq N$ be a group of agents. The binary relation $Q_G^{OL}(R)$ is defined as: for all $a, b \in A$ we have $aQ_G^{OL}(R)b$ if and only if there exists $i \in G$ such that aR_ib. In other words, $a\hat{Q}_G^{OL}(R)b$ if and only if for all $i \in G$ we have aP_ib and $a\bar{Q}_G^{OL}(R)b$ otherwise.

Proposition **3** *For all profiles R, the binary relation $Q_G^{OL}(R)$ is a quasi-ordering. Moreover, when $\#G = 1$, $Q_G^{OL}(R)$ is an ordering.*

Proof: Consider a preference profile R and $a, b, c \in A$. Let $a\hat{Q}_G^{OL}(R)b$ and $b\hat{Q}_G^{OL}(R)c$. By definition, aP_ib and bP_ic for all $i \in G$. Since P_i is transitive (Proposition **??**) we have aP_ic for all $i \in G$. This immediately implies that $a\hat{Q}_G^{OL}(R)c$. Hence, $\hat{Q}_G^{OL}(R)$ is transitive. This implies that $Q_G^{OL}(R)$ is a quasi-ordering.

When $G = \{i\}$, $a\hat{Q}_G^{OL}(R)b$ if and only if aP_ib and $a\bar{Q}_G^{OL}(R)b$ if and only if aI_ib. This means $aQ_G^{OL}(R)b$ if and only if aR_ib. Since R_i is transitive, $Q_G^{OL}(R)$ is transitive. Hence, $Q_G^{OL}(R)$ is an ordering. ■

Remark: The quasi-ordering $Q_G^{OL}(R)$ is not an ordering if $\#G \geq 2$. As an example, consider the preference profile (linear orderings) of two agents with three alternatives in Table **??**. Let $G = N = \{1, 2\}$. Then $a\hat{Q}_G^{OL}(R)b$, $b\bar{Q}_G^{OL}(R)c$ and $c\bar{Q}_G^{OL}(R)a$. Transitivity would imply that $a\bar{Q}_G^{OL}(R)b$, which is not true.

R_1	R_2
a	c
b	a
c	b

Table 2: Oligarchy is not an ordering if $\#G \geq 2$

3 Arrow's Impossibility Theorem

This section states and proves Arrow's impossibility theorem. In what follows, $F(R)$ is social ordering induced by F at the profile R and $\hat{F}(R)$ and $\bar{F}(R)$ denote its asymmetric and symmetric components respectively.

3.1 The Axioms

The following axioms are used in Arrow's impossibility theorem.

Definition 4 *The ASWF F satisfies the* **Weak Pareto** *(WP) axiom if for all profiles R, $a, b \in A$ we have aP_ib for all $i \in N$ implies that $a\hat{F}(R)b$.*

For the next axiom, we need some notation. Let R, R' be profiles and let $a, b \in A$. We say that R and R' agree on $\{a, b\}$ if

$$aP_ib \Leftrightarrow aP'_ib \;\; \forall i \in N$$
$$aI_ib \Leftrightarrow aI'_ib \;\; \forall i \in N.$$

We denote this by $R\mid_{a,b} = R'\mid_{a,b}$.

Definition 5 *The ASWF F satisfies* **Independence of Irrelevant Alternatives** *(IIA) axiom if for all $R, R' \in \mathbf{R}^n$ and for all $a, b \in A$, if $R\mid_{a,b} = R'\mid_{a,b}$ then $F(R)\mid_{a,b} = F(R')\mid_{a,b}$.*

Proposition 4 *Scoring rules violate IIA.*

Proof: We show it for Plurality rule and Borda rule. Let $A = \{a, b, c\}$ and $N = \{1, 2, 3\}$. Consider the linear orderings in Table **??**. Observe that $R\mid_{a,b} = R'\mid_{a,b}$. By IIA, we should

R_1	R_2	R_3	R'_1	R'_2	R'_3
a	c	b	a	c	c
b	a	c	b	a	b
c	b	a	c	b	a

Table 3: Scoring rules violate IIA

have $F(R)\mid_{a,b} = F(R')\mid_{a,b}$. Also, $a\bar{F}(R)b$ but $a\hat{F}(R')b$ in Plurality and Borda. This proves the claim. ■

Definition 6 *The ASWF F is* **dictatorial** *if there exists an agent $i \in N$ such that for all $a, b \in A$ and for all profiles R we have $[aP_ib \Rightarrow a\hat{F}(R)b]$. Voter i is called a* **dictator** *in this case.*

Remark: Notice that if F is dictatorial, it is not the case that there exists a voter i such that $F(R) = R_i$ for all profiles R. For example, the following rule is still dictatorial. For all $R \in \mathbf{R}^n$, there exists an agent i such that aP_ib implies $a\hat{F}(R)b$. But if aI_ib then $a\hat{F}(R)b$ if aP_jb for some $j \neq i$. But $F(R) = R_i$ is true if R_i is anti-symmetric. Check that $F(R)$ is an ordering for all profiles R.

3.2 Arrow's Theorem

Arrow's theorem demonstrates that the consequence of requiring ASWFs to satisfy WP and IIA is extremely restrictive.

Theorem **1 (Arrow's Impossibility Theorem)** *Suppose* $\#A \geq 3$*. A ASWF which satisfies IIA and WP must be dictatorial.*

Proof: Consider an ASWF F that satisfies IIA and WP. We say a group of agents $\emptyset \neq G \subseteq N$ is **decisive** for $a, b \in A$ (denoted by $D_G(a,b)$) if for all $R \in \mathbf{R}^n$

$$\big[aP_ib \ \forall \ i \in G\big] \Rightarrow \big[a\hat{F}(R)b\big].$$

We say a group of agents $\emptyset \neq G \subseteq N$ is **almost decisive** for $a, b \in A$ (denoted by $\bar{D}_G(a,b)$) if for all $R \in \mathbf{R}^n$

$$\big[aP_ib \ \forall \ i \in G, \ bP_ia \ \forall \ i \in N \setminus G\big] \Rightarrow \big[a\hat{F}(R)b\big].$$

Clearly, $D_G(a,b) \Rightarrow \bar{D}_G(a,b)$ for all $\emptyset \neq G \subseteq N$ and for all $a, b \in A$. We prove the following two important lemmas.

Lemma **1 (Field Expansion)** *For all* $\emptyset \neq G \subseteq N$ *and for all* $a, b, x, y \in A$

$$\bar{D}_G(a,b) \Rightarrow D_G(x,y).$$

Proof: We consider seven possible cases.

C1 Suppose $x \neq y \neq a \neq b$. Consider $R' \in \mathbf{R}^n$ such that xP'_iy for all $i \in G$ and $R \in \mathbf{R}^n$ such that $xP_iaP_ibP_iy$ for all $i \in G$. Also, for all $i \in N \setminus G$, impose xP_ia, bP_iy, bP_ia, $R_i \mid_{x,y} = R'_i \mid_{x,y}$.

Now, $\bar{D}_G(a,b) \Rightarrow a\hat{F}(R)b$. By WP, $x\hat{F}(R)a$ and $b\hat{F}(R)y$. By transitivity, we get $x\hat{F}(R)y$. But $R \mid_{x,y} = R' \mid_{x,y}$. By IIA, $x\hat{F}(R')y$. Hence, $D_G(x,y)$.

C2 Suppose $x \neq a \neq b$ but $y = b$. Consider $R' \in \mathbf{R}^n$ such that xP'_ib for all $i \in G$ and $R \in \mathbf{R}^n$ such that xP_iaP_ib for all $i \in G$. Also, for all $i \in N \setminus G$, impose xP_ia, bP_ia and $R_i \mid x, b = R'_i \mid_{x,b}$.

Now, $\bar{D}_G(a,b) \Rightarrow a\hat{F}(R)b$. Pareto gives $x\hat{F}(R)a$. By transitivity, $x\hat{F}(R)b$. By IIA, $x\hat{F}(R')b$. Hence, $D_G(x,b)$.

C3 Suppose $x = a$ and $y \neq a \neq b$. Consider $R' \in \mathbf{R}^n$ such that aP'_iy for all $i \in G$ and $R \in \mathbf{R}^n$ such that aP_ibP_iy for all $i \in G$. Also, for all $i \in N \setminus G$, impose bP_iy, bP_ia, and $R_i \mid_{a,y} = R'_i \mid_{a,y}$.

Now, $\bar{D}_G(a,b) \Rightarrow a\hat{F}(R)b$. Pareto give $b\hat{F}(R)y$. By transitivity, $a\hat{F}(R)y$. By IIA, $a\hat{F}(R')y$. Hence, $D_G(a,y)$.

C4 Suppose $x = b$ and $y \neq a \neq b$. From (C3), we get $\bar{D}_G(a,b) \Rightarrow D_G(a,y) \Rightarrow \bar{D}_G(a,y)$. From (C2), we get $\bar{D}_G(a,y) \Rightarrow D_G(b,y)$.

C5 Suppose $y = a$ and $x \neq a \neq b$. From (C2), we get $\bar{D}_G(a,b) \Rightarrow D_G(x,b) \Rightarrow \bar{D}_G(x,b)$. From (C3), we get $\bar{D}_G(x,b) \Rightarrow D_G(x,a)$.

C6 Suppose $x = a$ and $y = b$. Consider some $y \neq a \neq b$ (since $\#A \geq 3$, this is possible). From (C3) $\bar{D}_G(a,b) \Rightarrow D_G(a,y) \Rightarrow \bar{D}_G(a,y)$. Apply (C3) again to get $\bar{D}_G(a,y) \Rightarrow D_G(a,b)$.

C7 Suppose $x = b$ and $y = a$. Consider some $y \neq a \neq b$. From (C5), we get $\bar{D}_G(a,b) \Rightarrow D_G(y,a) \Rightarrow \bar{D}_G(y,a)$. From (C2), we get $\bar{D}_G(y,a) \Rightarrow D_G(b,a)$.

■

As a consequence of Field Expansion Lemma, we can speak of a decision group of agents without reference to any pair of alternatives. We now prove the other important lemma.

Lemma **2 (Group Contraction)** *Suppose $\emptyset \neq G \subseteq N$ is decisive. If $\#G \geq 2$, then there exists a proper non-empty subset of G which is also decisive.*

Proof: Let $G = G_1 \cup G_2$ with $G_1 \cap G_2 = \emptyset$ and $G_1, G_2 \neq \emptyset$. Let $a, b, c \in A$ and let $R \in \mathbf{R}^n$ be a preference profile as in Table **??**. Since aP_ib for all $i \in G$ and G is decisive, we get that $a\hat{F}(R)b$. We consider two possible cases.

G_1	G_2	$N \setminus G$
a	c	b
b	a	c
c	b	a

Table 4: A preference profile

C1 Suppose $a\hat{F}(R)c$. But aP_ic for all $i \in G_1$ and cP_ia for all $i \in N \setminus G_1$. Hence $\bar{D}_{G_1}(a,c)$. By Field Expansion Lemma, G_1 is decisive.

C2 Suppose $cF(R)a$. Since $a\hat{F}(R)b$, transitivity implies $c\hat{F}(R)b$. But cP_ib for all $i \in G_2$ and bP_ic for all $i \in N \setminus G_2$. Hence, $\bar{D}_{G_2}(c,b)$. By Field Expansion Lemma, we get that G_2 is decisive.

■

By WP, the grand coalition N is decisive. Repeated application of Group Contraction Lemma gives us that there exists an agent $i \in N$ such that i is decisive. By definition an ASWF is dictatorial if there is a single agent who is decisive. ■

4 Relaxing the Weak Pareto Axiom: Wilson's Theorem

We follow Malawski-Zhou (SCW 1994).

Definition **7** *The ASWF F satisfies* **Non-Imposition** *or NI if for all $a, b \in A$, there exists a profile R such that $aF(R)b$.*

An example of a ASWF violating NI is the following: for all profiles R, the social ordering $F(R)$ is a fixed ordering $\bar{R}_i$. Note that it trivially satisfies IIA.

Remark: If a ASWF satisfies WP, it satisfies NI.

Definition **8** *The ASWF F is* **anti-dictatorial** *if there exists a voter i such that for all $a, b \in A$ and all profiles R, we have $[aP_i b \Rightarrow b\hat{F}(R)a]$.*

The **null** ASWF F^n is defined as follows: for all $a, b \in A$ and for all profiles R, $a\bar{F}^n(R)b$.

Theorem **2 (Wilson's Theorem)** *Assume $|A| \geq 3$. A ASWF which satisfies IIA and NI must be null or dictatorial or anti-dictatorial.*

Proof: Let F be a SWF satisfying IIA and NI.
For all $a, b \in A$, we write $PO(a, b)$ if for all profiles R, $[aP_i b$ for all $i \in N \Rightarrow a\hat{F}(R)b]$.
For all $a, b \in A$, we write $APO(a, b)$ if for all profiles R, $[aP_i b$ for all $i \in N \Rightarrow b\hat{F}(R)a]$.

Lemma 1: For all $a, b, x, y \in A$ we have $PO(a, b) \Rightarrow PO(x, y)$.

Proof: There are several cases to consider like in the Field Expansion Lemma. We only prove the case $PO(a, b) \Rightarrow PO(a, y)$ where $b \neq y$. Pick an arbitrary profile R where $aP_i y$ for all $i \in N$. We will show that $a\hat{F}(R)y$.

Since F satisfies NI, there exists a profile R' such that $bF(R)y$. Construct the profile $\tilde{R}$ as follows: for all $i \in N$, $a\tilde{P}_i b$, $a\tilde{P}_i y$ and $\tilde{R}\mid_{b,y} = R'\mid_{b,y}$. This is clearly feasible. Since $PO(a, b)$ we have $a\hat{F}(\tilde{R})b$. On the other hand, IIA implies $bF(\tilde{R})y$. Since $F(\tilde{R})$ is transitive, we have $a\hat{F}(\tilde{R})y$. Now IIA implies $a\hat{F}(R)y$. This completes the proof of Lemma 1.

Lemma 2: For all $a, b, x, y \in A$ we have $APO(a, b) \Rightarrow APO(x, y)$.

Proof: Once again there are several cases to consider. We only prove the case $APO(a, b) \Rightarrow APO(a, y)$ where $b \neq y$. Pick an arbitrary profile R where $aP_i y$ for all $i \in N$. We will show that $y\hat{F}(R)a$.

Since F satisfies NI, there exists a profile R' such that $yF(R)b$. Construct the profile $\tilde{R}$ as follows: for all $i \in N$, $a\tilde{P}_i b$, $a\tilde{P}_i y$ and $\tilde{R}\mid_{b,y} = R'\mid_{b,y}$. This is clearly feasible. Since $PO(a, b)$ we have $b\hat{F}(\tilde{R}a)$. On the other hand, IIA implies $yF(\tilde{R})b$. Since $F(\tilde{R})$ is transitive, we have $y\hat{F}(\tilde{R})a$. Now IIA implies $y\hat{F}(R)a$. This completes the proof of Lemma 2.

LEMMA 3: One of the following statements must hold
(i) F is null.
(ii) $PO(a, b)$ holds for some pair a, b.
(iii) $APO(a, b)$ holds for some pair a, b.

Proof: Suppose that neither (i) nor (ii) nor (iii) hold. Since (i) does not hold, there exists a pair x, y and a profile R such that $x\hat{F}(R)y$ holds. Pick $z \neq x, y$ and let R' be a profile such that xP'_iz, yP'_iz for all $i \in N$ and $R' \mid_{x,y} = R \mid_{x,y}$. Again this is clearly feasible. Since neither $PO(x, z)$ nor $APO(x, z)$ hold, we must have $x\bar{F}(R')z$. Similarly, since neither $PO(y, z)$ nor $APO(y, z)$ hold, we must have $y\bar{F}(R')z$. Since $F(R')$ is transitive, we have $x\bar{F}(R')y$. Applying IIA, we have $x\bar{F}(R)y$. But this contradicts our assumption that $x\hat{F}(R)y$ and completes the proof of Lemma 3.

Suppose F is not null. Applying Lemma 3, either $PO(a, b)$ must hold for some a, b or $APO(a, b)$ must hold for some pair a, b. Suppose the former holds. Then WP holds and the existence of a dictator follows from Arrow's Theorem. If the latter holds, then the proof of Arrow's Theorem can be modified in a straightforward manner to show that F is anti-dictatorial. ■

5 Existence of Maximal Elements

Let Q_i be a binary relation over the elements of the set A. Let $B \subset A$.

DEFINITION **9** *The set of* **maximal elements** *of B according to Q denoted by $M(B, Q_i)$ is the set* $\{x \in B | \nexists y \in B \text{ and } y\hat{Q}_ix\}$.

REMARK: Since Q_i is complete, we can define the set of maximal elements equivalently as $M(B, Q_i) = \{x \in B | xQ_iy \text{ for all } y \in B\}$.

DEFINITION **10** *The binary relation Q_i is* **acyclic** *if for all $a_1, a_2, ..., a_K \in A$, we have* $[a_1\hat{Q}_ia_2, a_2\hat{Q}_ia_3, ..., a_{K-1}\hat{Q}_ia_K] \Rightarrow a_1Q_ia_K$.

REMARK: Q_i is transitive $\Rightarrow$ Q_i is quasi-transitive $\Rightarrow$ Q_i is acyclic.

PROPOSITION **5** *Let Q_i be a binary relation over a finite set A. Then* $[M(B, Q_i) \neq \emptyset] \Rightarrow$ [*Q_i is acyclic*].

Proof: $\Rightarrow$ Suppose not, i.e there exists $a_1, ..., a_K$ such that $a_1\hat{Q}_ia_2, ..., a_{K-1}\hat{Q}_ia_K$ and $a_K\hat{Q}_ia_1$. Let $B = \{a_1, ..., a_K\}$. Clearly $M(B, Q_i) = \emptyset$ which contradicts our hypothesis.

$\Leftarrow$ Suppose Q_i is acyclic and let B be an arbitrary subset of A. Pick an arbitrary element $a_1 \in B$. If $a_1 \in M(B, Q_i)$, we are done. Suppose $a_1 \notin M(B, Q_i)$. There must exist $a_2 \in B$ such that $a_2 \hat{Q}_i a_1$. If $a_2 \in M(B, Q_i)$, we are done again. Otherwise there exists a_3 such that $a_3 \hat{Q}_i a_2$. Note that acyclicity implies $a_3 Q_i a_1$, i.e $a_3 \neq a_1$. If $a_3 \in M(B, Q_i)$ our algorithm stops; otherwise we find an element a_4 such that $a_4 \hat{Q}_i a_3$. Critically acyclicity implies $a_4 \neq a_2, a_1$. In general, acyclicity implies that the sequence $a_1, ..., a_k....$ constructed in the manner above contains no repetitions. Since B is finite, the algorithm must stop, i.e $M(B, Q_i) \neq \emptyset$. ■

Remark: Acyclicity over triples is not sufficient for maximal elements to exist. Consider the following example: $A = \{a_1, a_2, a_3, a_4\}$ and $a_1 \hat{Q}_i a_2$, $a_2 \hat{Q}_i a_3$, $a_3 \hat{Q}_i a_4$, $a_4 \hat{Q}_i a_1$, $a_1 \bar{Q}_i a_3$ and $a_2 \bar{Q}_i a_4$. Then acyclicity over triples is satisfied but $M(B, Q_i) = \emptyset$.

Remark: Acyclicity does not guarantee the existence of maximal elements if A is not finite. For example, let Q_i be the natural ordering of the real numbers and let $A = [0, 1)$. Then $M(A, Q_i) = \emptyset$.

6 Domain Restrictions: Single-Peaked Preferences

We endow A with additional structure.

Let $\geq$ be a linear order over A. For instance A could be the unit interval and $\geq$ the natural ordering over the reals.

Definition **11** *The ordering R_i is* **single-peaked** *if there exists $a^* \in A$ (called the* **peak** *of R_i) such that for all $b, c \in A$*

$$[a^* \geq b > c \text{ or } c > b \geq a^*] \Rightarrow bP_i c$$

Let $\mathcal{R}^{SP}(\geq)$ be the set of all single-peaked preferences with respect to the ordering $\geq$. Throughout the this section we shall keep $\geq$ fixed so that we shall refer to the set of single-peaked preferences simply as $\mathcal{R}^{SP}$. We shall denote the peak of a single-peaked (or any other, for that matter) ordering R_i as $\tau(R_i)$.

Example **1** Let $A = [0, 1]$ denote the fraction of the Central Government's budget that is spent on education. According to voter i the optimal fraction is 0.1. If her preferences are single-peaked, she strictly prefers 0.2 over 0.3 and 0.08 over 0.05. Note that single-peakedness places no restrictions on alternatives on different "sides" of the peak, i.e. the voter can either prefer 0.05 to 0.2 or vice-versa.

Remark: Let $A = \{a, b, c\}$ and consider the set of linear orders (R_1, R_2, R_3) which constitute the following Condorcet in Table 1. It is easy to check that there does not exist an ordering $\geq$ over A such that (R_1, R_2, R_3) are single-peaked with respect to $\geq$. Suppose for instance $a > b > c$. Then R_2 is not single-peaked because if c is the peak, then b must be strictly better than a.

Remark: Let $|A| = m$. Then $|\mathcal{R}^{SP}| = 2^{m-1}$.

Definition **12** *Let $R \in \mathcal{R}^{SP}$ be a profile of single-peaked preferences. The* **median voter** *in the profile R is the voter h such that $|\{i \in N : \tau(R_h) \geq \tau(R_i)\}| \geq \frac{n}{2}$ and $|\{i \in N : \tau(R_i) \geq \tau(R_h)\}| \geq \frac{n}{2}$.*

Remark: The median voter exists for all profiles although she may not be unique. However if n is odd, the median peak $\tau(R_h)$ will be unique.

Theorem **3 (Median Voter Theorem)** *Let $R \in \mathcal{R}^{SP}$ be a profile of single-peaked preferences. Then $M(A, Q^{maj}) \neq \emptyset$. In particular $\tau(R_h) \in M(A, Q^{maj})$.*

Proof: Pick an arbitrary profile $R \in \mathcal{R}^{SP}$. We will show that $\tau(R_h)Q^{maj}b$ for all $b \neq \tau(R_h)$. We consider two cases.

Case 1. $\tau(R_h) > b$. Let $i \in N$ be such that $\tau(R_i) \geq \tau(R_h)$. Since R_i is single-peaked and $\tau(R_i) \geq \tau(R_h) > b$, we have $\tau(R_h)P_ib$. Since $|\{i \in N : \tau(R_i) \geq \tau(R_h)\}| \geq \frac{n}{2}$ since h is a median voter, it follows that $\tau(R_h)Q^{maj}b$.

Case 2. $b > \tau(R_h)$. Let $i \in N$ be such that $\tau(R_h) \geq \tau(R_i)$. Since R_i is single-peaked and $b > \tau(R_h) \geq \tau(R_i)$, we have $\tau(R_h)P_ib$. Since $|\{i \in N : \tau(R_h) \geq \tau(R_i)\}| \geq \frac{n}{2}$ since h is median voter, it follows that $\tau(R_h)Q^{maj}b$.

This covers all possible cases. ■

Is $Q^{maj}(R)$ transitive for all single-peaked profiles? No, as the following example shows.

Example **2** Let $A = [0, 1]$, $N = \{1, 2\}$. Let R_1 and R_2 be the following single-peaked orderings:

- $\tau(R_i) = 0.4$ and xP_iy whenever $0.4 > x$ and $y > 0.4$, i.e voter 1 prefers all alternatives to the "left" of 0.4 to everything on the "right" of 0.4.

- $\tau(R_i) = 0.5$ and xP_iy whenever $x > 0.5$ and $0.5 > y$, i.e voter 1 prefers all alternatives to the "right" of 0.5 to everything on the "left" of 0.5.

Now consider the alternatives $a = 0.1$, $b = 0.2$ and $c = 0.6$. Note that bP_1c and cP_2b so that $b\bar{Q}^{maj}c$. Similarly, aP_1c and cP_2a so that $a\bar{Q}^{maj}c$. However single-peakedness of R_1 and R_2 imply bP_1a and bP_2a so that $b\hat{Q}^{maj}a$. Clearly Q^{maj} is not transitive. Note that all alternatives in the interval $[0.4, 0.5]$ are maximal according to Q^{maj} in A.

The binary relation Q^{maj} defined over single-peaked preferences is transitive in special cases.

PROPOSITION **6** *Assume that n is odd and that voter preferences are linear and single-peaked. Then for all profiles R, $Q^{maj}(R)$ is an ordering.*

Proof: We only need to show that for all profiles R, $Q^{maj}(R)$ is transitive. Since n is odd and voter preferences do not admit indifference, $Q^{maj}(R)$ admits no indifferences, i.e. for all $a, b \in A$, either $a\hat{Q}^{maj}(R)b$ or $b\hat{Q}^{maj}(R)a$ holds. Now pick $a, b, c \in A$ and a profile R and assume w.l.o.g. that $a\hat{Q}^{maj}(R)b$ and $b\hat{Q}^{maj}(R)c$. Observe that for for all voters i, R_i induces single-peaked preferences over $\{a, b, c\}$ (prove!). Applying Theorem **??** to the set $\{a, b, c\}$, it follows that $M(\{a, b, c\}, R) \neq \emptyset$. Therefore $c\hat{Q}^{maj}a$ is impossible, i.e $a\hat{Q}^{maj}c$ holds and Q^{maj} is transitive. ■

7 Interpersonal Comparability

We now turn our attention to models where voters are endowed with "richer information" which can be used for aggregation.

Voter i will be assumed to have a utility function $u_i : A \to \Re$. We shall let $\mathcal{U}$ denote the set of all such utility functions. A utility profile u is an n-tuple $(u_1, ..., u_n) \in \mathcal{U}^n$.

DEFINITION **13** *A Social Welfare Functional (SWFL) F is a mapping $F : \mathcal{U}^n \to \mathcal{R}$.*

Let F be a SWFL. For all utility profiles u we shall let R_u denote the social ordering $F(u)$.

We now restate some axioms that we had introduced earlier for this environment and also introduce some new ones.

DEFINITION **14** *A SWFL F satisfies* **Binary Independence of Irrelevant Alternatives** *(BIIA) if for all profiles u, u' and $a, b \in A$,*

$$[u_i(a) = u'_i(a) \text{ and } u_i(b) = u'_i(b) \ \ \forall i \in N] \Rightarrow [R_u \mid_{a,b} = R_{u'} \mid_{a,b}]$$

let $a, b, c, d \in A$ and let R_i be an ordering. We say $R_i \mid_{a,b} = R_i \mid_{c,d}$ if $[aP_ib \Leftrightarrow cP_id]$ and $[aI_ib \Leftrightarrow cI_id]$.

A stronger version of BIIA is Strong Neutrality defined below.

DEFINITION **15** *A SWFL F satisfies* **Strong Neutrality** *(SN) if for all profiles u, u' and $a, b, c, d \in A$,*

$$[u_i(a) = u'_i(c) \text{ and } u_i(b) = u'_i(d) \ \ \forall i \in N] \Rightarrow [R_u \mid_{a,b} = R_{u'} \mid_{c,d}]$$

In other words, if the utilities associated with a and b in profile u agree with those of c and d respectively in profile u', then a and b must be ranked in exactly the same way under R_u as c and d under $R_{u'}$. Note that while a and b are distinct and c and d are also distinct, it may be the case that $b = c$ and $a = d$ etc.

We introduce some some Pareto type axioms.

DEFINITION **16** *The SWFL F satisfies* **Pareto Indifference** *(PI) if, for all $a, b \in A$ and profiles u, $[u_i(a) = u_i(b)$ for all $i \in N] \Rightarrow aI_ub$.*

DEFINITION **17** *The SWFL F satisfies* **Strong Pareto** *(SP) if, for all $a, b \in A$ and profiles u, $[u_i(a) \geq u_i(b)$ for all $i \in N] \Rightarrow aR_ub$. Moreover if there exists $k \in N$ such that $u_k(a) > u_k(b)$, then aP_ub.*

7.1 Measurability and Comparability Axioms

Let $\phi \equiv (\phi_1, ..., \phi_n)$ be an n-tuple of strictly increasing functions $\phi_i : \Re \rightarrow \Re$. Let Φ be an arbitrary set of such n-tuples.

Let u be a profile. The profile $\phi.u$ denote the profile $(\phi_1.u_1, ..., \phi_n.u_n)$, i.e the utility for alternative a for voter i is $\phi_i(u_i(a))$.

DEFINITION **18** *The SWFL F satisfies invariance with respect to Φ if for all profiles u, $F(u) = F(\phi.u)$.*

The idea is as follows. Divide the set of all profiles $\mathcal{U}^n$ into equivalence classes. Two profiles u, u' belong to the same equivalence class if there exists $\phi \in \Phi$ such that $u = \phi.u$. A SWFL F which is invariant with respect to Φ if $f(u) = f(u')$. In other words, two profiles in the same equivalence class have the same "information" permissible for aggregation from the viewpoint of F. Observe that the finer the partition of $\mathcal{U}^n$ into equivalence classes or partitions, the greater is the information that is being allowed for aggregation.

We now consider various assumptions on ϕ.

DEFINITION **19** *A SWFL satisfies Ordinally Measurable, Non-Comparable Utilities (OMNC) if Φ consists of all n-tuples of increasing functions $(\phi_1, ..., \phi_n)$.*

REMARK: In the OMNC, only ordinal information is being allowed for aggregation. This is the Arrovian case.

DEFINITION **20** *A SWFL satisfies Cardinally Measurable, Non-Comparable Utilities (CMNC) if $\phi \in \Phi$ if for all $i \in N$, $\phi_i(t) = \alpha_i + \beta_i t$ with $\beta_i > 0$.*

REMARK: In CMNC we allow for independent affine transformations of utilities for voters.

Definition **21** *A SWFL satisfies Ordinally Measurable, Fully-Comparable Utilities (OMFC) if $\phi_i \in \Phi$ if, for all $i \in N$, $\phi_i = \phi_0$ for some increasing function $\phi_0 : \Re \to \Re$.*

Definition **22** *A SWFL satisfies Cardinally Measurable, Fully-Comparable Utilities (CMUC) if $\phi_i \in \Phi$ if, for all $i \in N$, $\phi_i = \alpha + \beta t$ with $\beta > 0$.*

Question: What are the SWFLs which satisfy a certain class of measurability and comparability restriction together with the classical Arrovian assumptions?

7.2 Welfarism

Our goal in this subsection is to show that the questions raised in the previous subsection can be reduced to problems of ranking vectors in $\Re^n$.

Proposition **7 (Welfarism)** *SN $\Rightarrow$ BIIA. If $|A| \geq 3$, then BIIA + PI $\Rightarrow$ SN.*

Proof: The first proof of the proposition is trivial. There are several cases to deal with like in the Field Expansion Lemma. Consider the case where $a, b, c \in A$, and profiles u, u' are such that $u_i(a) = u'_i(a)$ and $u_i(b) = u'_i(c)$ for all $i \in N$. We have to show that $R_u \mid_{a,b} = R_{u'} \mid_{a,c}$. Construct a profile $\tilde{u}$ such that $\tilde{u}_i(a) = u_i(a) = u'_i(a)$ and $\tilde{u}_i(b) = \tilde{u}_i(c) = u_i(b) = u'_i(c)$ for all $i \in N$. By BIIA, $R_u \mid_{a,b} = R_{\tilde{u}} \mid_{a,b}$ and $R_{u'} \mid_{a,c} = R_{\tilde{u}} \mid_{a,c}$. By PI, $bI_{\tilde{u}}c$ so that the transitivity of $R_{\tilde{u}}$ implies $R_{\tilde{u}} \mid_{a,b} = R_{\tilde{u}} \mid_{a,c}$. Hence $R_u \mid_{a,b} = R_{u'} \mid_{a,c}$.

Similar arguments can be used to prove all cases. Note that in the case where u, u' are such that $u_i(a) = u'_i(b)$ and $u_i(b) = u'_i(a)$ for all $i \in N$ we need a third alternative c, i.e we need to use the assumption that $|A| \geq 3$. ■

We shall often use the following notation: for all $a \in A$ and profile u, $u(a) \equiv (u_1(a), ..., u_n(a))$.

Proposition **8** *Assume $|A| \geq 3$. A SWFL satisfies PI and BIIA if and only if there exists an ordering $\succeq$ on $\Re^n$ such that for all $a, b \in A$ and for all profiles u, $R_u \mid_{a,b} = \succeq\mid_{\alpha,\beta}$ where $u(a) = \alpha$ and $u(b) = \beta$.*

Proof: Let $\succeq$ be an ordering on $\Re^n$. Construct a SWFL F as follows: for all profiles u and $a, b \in A$, $R_u \mid_{a,b} = \succeq\mid_{u(a),u(b)}$. The transitivity of R_u is a direct consequence of the transitivity of $\succeq$ while BIIA and PI of F follows directly from its definition.

Let F satisfy BIIA and PI. Define $\succeq$ as follows: for all $\alpha, \beta \in \Re^n$, $\succeq\mid_{\alpha,\beta} = R_u \mid_{a,b}$ for some $a, b \in A$ and profile u such that $u(a) = \alpha$ and $u(b) = \beta$. Since F satisfies PI and BIIA, it satisfies SN (Proposition **??**). This implies that the ranking of vectors $\alpha, \beta \in \Re^n$ according to $\succeq$ does not depend on the alternatives a, b and profile u chosen in the construction (i.e. so that $u(a) = \alpha$ and $u(b) = \beta$). In other words, $\succeq$ is well-defined. It is transitive because R_u is transitive for all u. ■

Figure 1: Arrow's Theorem

Proposition **??** reduces the problem of finding an SWFL satisfying PI and BIIA to the problem of finding an appropriate ordering of utility vectors. We only need to reinterpret the measurability and comparability requirement in this environment.

Let F be an SWFL satisfying PI, BIIA and invariance with respect to Φ. Let $\succeq$ be the ordering over $\Re^n$ induced by F. Let $\alpha, \beta \in \Re^n$ and $\phi \in \Phi$. Let $\phi.\alpha$ and $\phi.\beta$ denote the n-tuples $(\phi_i(\alpha_1), ..., \phi_n(\alpha_n))$ and $(\phi_i(\beta_1), ..., \phi_n(\beta_n))$ respectively. By invariance on F, we have $R_u \mid_{a,b} = R_{\phi.u} \mid_{a,b}$. From the construction of $\succeq$ we know that $R_u \mid_{a,b} \;=\; \succeq\mid_{\alpha,\beta}$ and $R_{\phi.u} \mid_{a,b} \;=\; \succeq\mid_{\phi.\alpha,\phi.\beta}$. Therefore $\succeq\mid_{\alpha,\beta} = \succeq\mid_{\phi.\alpha,\phi.\beta}$. This motivates the following definition.

DEFINITION **23** *The ordering $\succeq$ over $\Re^n$ satisfies invariance with respect to Φ, if for all $\alpha, \beta \in \Re^n$ and $\phi \in \Phi$, we have $\succeq\mid_{\alpha,\beta} \;=\; \succeq\mid_{\phi.\alpha,\phi.\beta}$.*

PROPOSITION **9** *Assume $|A| \geq 3$. Let F be a SWFL satisfying PI and BIIA and invariance with respect to Φ. Then the induced ordering $\succeq$ over $\Re^n$ satisfies invariance with respect to Φ.*

Proposition **??** follows from our earlier discussion.

7.3 Arrow's Theorem: A Geometrical Approach

We restate Arrow's Theorem in this environment.

THEOREM **4 (Arrow's Theorem for Social Welfare Functionals)** *Assume $|A| \geq 3$. If a SWFL satisfies PI, WP, BIIA and OMNC, then it must be dictatorial.*

Proof: We will only do the case of $n = 2$. Let F satisfy PI, WP, BIIA and OMNC. Applying Proposition **??**, we will show that the induced ordering $\succeq$ over $\Re^2$ has the following property: there exists $i = \{1, 2\}$ such that for all $\alpha, \beta \in \Re^2$, we have $\alpha \succ \beta$ only if $\alpha_i > \beta_i$.

Refer to Figure 1. Let α be an arbitrary point in $\Re^2$. We will try to draw an "indifference curve" through α. Consider Regions I, II, III and IV which do not include the dotted lines.

Step 1: All vectors in region II must be strictly better than α according to $\succeq$. In other words $\beta \succ \alpha$ for all $\beta \in$ Region I. This follows from WP. Similarly all vectors in Region IV must be worse than α by WP.

Step 2: Let $\beta, \gamma \in$ Region I. Then $\succeq\mid_{\alpha,\beta} \;=\; \succeq\mid_{\alpha,\gamma}$.

Let $\phi_1 : \Re \rightarrow \Re$ be a linear function such that $\phi_1(\beta_1) = \gamma_1$ and $\phi_1(\alpha_1) = \alpha_1$. Since $\beta_1, \gamma_1 < \alpha_1$ it follows that ϕ_1 is strictly increasing. Similarly let $\phi_2 : \Re \rightarrow \Re$ be such that $\phi_1(\beta_2) = \gamma_2$ and $\phi_1(\alpha_2) = \alpha_2$. Since $\beta_2, \gamma_2 > \alpha_2$ ϕ_2 is also increasing. Observe the

$\phi(\beta) = \gamma$ and $\phi(\alpha) = \alpha$. Since ϕ_1, ϕ_2 are increasing and $\succeq$ satisfies OMNC, we must have $\succeq|_{\alpha,\beta} \;=\; \succ|_{\alpha,\gamma}$.

Step 3: Let $\beta, \gamma \in$ Region III. Then $\succeq|_{\alpha,\beta} \;=\; \succeq|_{\alpha,\gamma}$.

The arguments here are identical to those in Step 2.

Step 4: Let $\beta \in$ Region I. Then either $\beta \succ \alpha$ or $\alpha \succ \beta$ must hold.

Suppose that the claim above is false, i.e. $\beta \sim \alpha$. Since Region II is an open set, we can find $\gamma \in$ Region II (sufficiently close to β) such that $\gamma > \beta$. From Step 2, we must have $\gamma \sim \alpha$, so that $\beta \sim \gamma$ by transitivity of $\succeq$. However $\gamma \succ \beta$ by WP. Contradiction.

Step 5: Let $\beta \in$ Region III. Then either $\beta \succ \alpha$ or $\alpha \succ \beta$ must hold.

The arguments here are identical to those in Step 4.

Step 6: Let $\beta \in$ Region I and $\gamma \in$ Region III. Then $\beta \succ \alpha \Rightarrow \alpha \succ \gamma$. Similarly $\alpha \succ \beta \Rightarrow \gamma \succ \alpha$.

Suppose $\beta \succ \alpha$. Consider the following functions: $\phi_1(t) = t + (\alpha_1 - \beta_1)$ and $\phi_2(t) = t - (\beta_2 - \alpha_2)$. Note that ϕ_1 and ϕ_2 are strictly increasing. Also $\phi(\beta) = \alpha$. Since $\alpha_1 - \beta_1 > 0$, we have $\phi_1(\alpha_1) > \alpha_1$. Since $\beta_2 - \alpha_2 > 0$, we have $\phi_2(\alpha_2) < \alpha_2$. Hence $\phi(\alpha) \in$ Region III. Since $\beta \succ \alpha$, invariance implies $\phi(\beta) \succ \phi(\alpha)$, i.e. $\alpha \succ \gamma$ where $\gamma \in$ Region III.

Step 7: Let $\beta \in$ Region I. If $\beta \succ \alpha$. Let γ be a point on the boundary of Regions I and II and let $\gamma^{'}$ be a point on the boundary of Regions III and IV. Then $\gamma \succ \alpha$ and $\alpha \succ \gamma^{'}$. This follows from Step 6 and WP. By an identical argument, if $\alpha \succ \beta$ where $\beta \in$ Region I, then all points in on the boundary of Regions I and IV are strictly worse than α according to $\succ$ and all points on the boundary of Regions III and IV are strictly better than α according to $\succ$.

Summary: Steps 1 through 7 imply that there are exactly two possibilities: (i) Regions I and II are better than α and Regions III and IV are worse than α according to $\succ$ (ii) Regions II and III are better than α and Regions I and IV are worse than α according to $\succ$. We say that the *pseudo-indifference curve* through α is *horizontal* if possibility (i) holds and *vertical* if possibility (ii) holds.

Step 8: If the pseudo-indifference curve is horizontal (resp. vertical) for some α, it must be horizontal (resp. vertical) *for all* $\alpha \in \Re^2$. If this was false, the two pseudo-indifference curves would intersect, contradicting the transitivity of $\succ$.

We can now complete the proof of the theorem. If all pseudo-indifference curves are horizontal, voter 2 is the dictator; if they are horizontal, voter 1 is the dictator. ■

REMARK: The ordering $\succ$ that we have constructed above is not *complete*. For instance if all the pseudo-indifference curves are vertical, we know the following: for $\alpha, \beta \in \Re^2$ such that $\beta_1 > \alpha_1$, we have $\beta \succ \alpha$. But we say nothing in the case $\beta_1 = \alpha_1$. In order to characterize $\succ$, we need additional axioms.

Definition 24 *The ordering $\succeq$ satisfies continuity, if for all $\alpha \in \Re^n$ the sets $\{\beta : \beta \succeq \alpha\}$ and $\{\beta : \alpha \succeq \beta\}$ are closed.*

Definition 25 *The ordering $\succeq$ is strongly dictatorial, if there exists a voter i such that for all $\alpha, \beta \in \Re^n$, $[\alpha_i \geq \beta_i] \Leftrightarrow [\alpha \succeq \beta]$*

Suppose $\succeq$ is strongly dictatorial and that voter i is the dictator. Then for all $\alpha, \beta \in \Re^n$, $[\alpha_i > \beta_i] \Rightarrow [\alpha \succ \beta]$, $[\beta_i > \alpha_i] \Rightarrow [\beta \succ \alpha]$ and $[\alpha_i = \beta_i] \Rightarrow [\alpha \sim \beta]$.

Definition 26 *The ordering $\succeq$ is lexicographic, if there exists an ordering of voters $i_1, i_2, ..., i_n$ such that for all $\alpha, \beta \in \Re^n$, $\alpha \succ \beta$ implies that there exists an integer K lying between 1 and n such that*

- $\alpha_{i_k} = \beta_{i_k}$ *for all* $k = 1, ...K-1$
- $\alpha_{i_K} > \beta_{i_K}$.

Corollary 1 *Assume $|A| \geq 3$. If a SWFL satisfies PI, WP, BIIA and OMNC and the induced ordering $\succeq$ satisfies continuity, then it must be strongly dictatorial.*

Corollary 2 *Assume $|A| \geq 3$. If a SWFL satisfies PI, SP, BIIA and OMNC then it must be lexicographic.*

8 Mechanism Design: Complete Information

8.1 The King Solomon Problem: A Motivating Example

Two women, referred to as 1 and 2 both claim to be the mother of a child. King Solomon has to decide whether (i) to give the child to 1 (which we shall call outcome a) (ii) to give the child to 2 (outcome b) or (iii) to cut the child in half (outcome c).

There are two "states of the world", θ and ϕ. In state θ, 1 is the real mother while 2 is the impostor; the reverse is true in state ϕ. The preferences of the two women over the outcomes $\{a, b, c\}$ depend on the state. We assume that the following holds.

State θ		State ϕ	
1	2	1	2
a	b	a	b
b	c	c	a
c	a	b	c

Table 5: Preferences in states θ and ϕ

The best choice for each woman is to get the child in both states. However the true mother would rather see the child be given to the other mother rather than cut in half; the opposite is true for the false mother.

King Solomon's objectives are specified by a social choice function $f : \{\theta, \phi\} \to \{a, b, c\}$ where $f(\theta) = a$ and $f(\phi) = b$. The key difficulty of course, is that King Solomon does not know which state of the world has occurred. He might therefore devise a "mechanism" of the following kind. Both women are asked to reveal the state (i.e. the identity of the true mother). If both women agree that state is θ, outcome a is enforced; if both agree that it is state ϕ, outcome b is enforced; if they disagree outcome c is enforced. This s shown in Table **??** below where 1's messages are shown along the rows and 2's along the columns.

	θ	ϕ
θ	a	c
ϕ	c	b

Table 6: King Solomon's mechanism

Does this work? Unfortunately not. Suppose the state is θ. Observe that the mechanism together with the preferences specified in θ constitutes a game in normal form. The unique pure strategy Nash equilibrium of this game is for both women to announce ϕ leading to outcome b. Similarly, the equilibrium in state ϕ is for both women to announce θ leading to the outcome a. This mechanism gives the baby to wrong mother in each state!

QUESTION: Does there exist a better mechanism?

8.2 A General Formulation

- $A = \{a, b, c, \ldots\}$: set of outcomes or alternatives.
- $I = \{1, 2, \ldots N\}$: set of agents or players.
- $\Theta = \{\theta, \phi, \psi, \ldots\}$: set of states.
- $R_i(\theta)$: preference ordering of agent i of the elements of A in state θ, i.e $R_i(\theta)$ is a complete, reflexive and antisymmetric binary relation defined on the elements of A.

Definition **27** *A Social Choice Correspondence (scc) F associates a non-empty subset of A denoted by $F(\theta)$ with every state $\theta \in \Theta$.*

A *Social Choice Function* (scf) is a singleton-valued scc.

The SCC specifies the objectives of the planner/principal/mechanism designer.

Definition **28** *A mechanism G is an $N+1$ tuple $(M_1, M_2 \ldots, M_N; g)$ where M_i, $i = 1, 2 \ldots, N$ is the message set of agent i and g is a mapping $g : M_1 \times \ldots \times M_N \to A$.*

For all $\theta \in \Theta$, the pair (G, θ) constitutes a game in normal form. We let $\text{NE}(G, \theta)$ denote the set of pure strategy Nash equilibria in (G, θ), i.e.

$$\bar{m} \in \text{NE}(G, \theta) \Rightarrow g(\bar{m}) R_i(\theta) g(m_i, \bar{m}_{-i}) \text{ for all } m_i \in M_i \text{ and } i \in I.$$

Definition **29** *The mechanism $G \equiv (M_1, M_2, \ldots, M_N)$ implements the scc F if*

$$g(NE(G, \theta)) = F(\theta) \text{ for all } \theta \in \Theta.$$

We require all Nash equilibria to be optimal according to F. However we restrict attention to pure strategy equilibria.

QUESTION: What are the scc's that can be implemented?

8.3 The Information Structure

An important feature of the formulation above is that, once a state is realized, agents are assumed to play Nash equilibrium. For them to be able to do so, it must be the case that the payoffs are common knowledge to the players. This means that we are assuming that once a state is realized, it is common knowledge to all except the mechanism designer. In the earlier example, both women know who the real mother is but King Solomon does not. This is the basis for the classification of this model as a complete information model. In incomplete information models, agents have residual uncertainty about others even after receiving their private information.

The complete information structure applies to two situations of particular interest.

1. The private information (common knowledge to all agents) is not *verifiable* by an outside party (for instance, by a judge or arbiter), as in the King Solomon case. These models are important bilateral contracting tc.

2. The mechanism has to be put in place before (in the chronological sense) the realization of the state. We may, for example have to design electoral procedures or a constitution which once decided upon, will remain in place for a while.

Some classical questions in equilibrium theory can also be addressed within the framework of this model. For example, can Walrasian equilibrium be attained when there is asymmetric information and a small number of agents? Questions of this nature motivated Leonid Hurwicz who is the founder of the theory of mechanism design.

Observation **1** The physical presence of a planner or mechanism designer is not an issue. The scc could reflect the common goals of all agents.

8.4 Ideas behind Maskin's Mechanism

Suppose the mechanism designer wishes to implement the scc F. For every θ and $a \in F(\theta)$, there must exist a message vector which is a Nash equilibrium under θ whose outcome is a. Assume w.l.o.g that this message vector is labeled (a, θ), i.e. when everybody sends the message (a, θ), the outcome is a. Since this is an equilibrium in θ, any deviation by player i must lead to an outcome in the set $L(a, i, \theta) = \{b \in A | aR_i(\theta)b\}$. If $N \geq 3$, it is easy to identify the deviant and "punish" him (i.e. pick an outcome in $L(a, i, \theta)$). The mechanism designer has already ensured that $F(\theta) \subset g(\text{NE}(G, \theta))$ for all $\theta \in \Theta$. Now he must try to ensure that there are no other equilibria in (G, θ). Consider the following candidate equilibrium message vectors.

1. Message vectors that are non-unanimous, i.e some agents send (a, θ), others (b, ϕ) and yet others (c, ψ) etc. This situation is relatively easy to deal with because the mechanism designer knows that such a message vector does not need to be an equilibrium in any state of the world. He can therefore attempt to "destroy" it as an equilibrium by by allowing all agents to deviate and get any alternative in A.

2. Message vectors that are unanimous. Suppose everyone sends the message (a, θ). The true state is however ϕ. The planner must however continue to behave as though the true state is θ because there is no way for him to distinguish this situation from the one where the true state is θ. In particular the outcome will be a and deviations by player i must lead to an outcome in $L(a, i, \theta)$. But this implies that if $L(a, i, \theta) \subset L(a, i, \phi)$, then (a, θ) will be an equilibrium under ϕ. If F is implementable, it must be the case that $a \in F(\phi)$. This is the critical restriction imposes on F and is called Maskin-Monotonicity.

8.5 Maskin's Theorem

Definition **30** *The scc F satisfies Maskin-Monotonicity (MM)if, for all $\theta, \phi \in \Theta$ and $a \in A$,*

$$[a \in F(\theta) \textit{ and } L(a, i, \theta) \subset L(a, i, \phi) \textit{ for all } i \in I] \Rightarrow [a \in F(\phi)]$$

Suppose a is F-optimal in state θ. Suppose also that for all agents, all alternatives that are worse than a in θ are also worse than a in ϕ. Then a is also F-optimal in ϕ. The MM condition can also be restated as follows.

Definition **31** *The scc F satisfies MM if, for all $\theta, \phi \in \Theta$ and $a \in A$,*

$$[a \in F(\theta) - F(\phi)] \Rightarrow [\exists i \in I \textit{ and } b \neq a \textit{ s.t. } aR_i(\theta)b \textit{ and } bP_i(\phi)a]$$

Suppose a is F-optimal in θ but not in ϕ. Then there must exist an agent and an alternative b such that a *preference reversal* takes place over a and b between θ and ϕ.

DEFINITION **32** *The scc F satisfies No Veto Power (NVP if, for all $a \in A$ and $\theta, \phi \in \Theta$,*

$$[\#\{i \in I | aR_i(\theta)b \text{ for all } b \in A\} \geq N-1] \Rightarrow [a \in F(\theta)]$$

If at least $N-1$ agents rank an alternative as maximal in a state, then that alternative must be F-optimal in that state. NVP is a weak condition. It is trivially satisfied in environments where there is a private good in which agent preferences are strictly increasing.

THEOREM **5 (Maskin 1977, 1999)** *1. If F is implementable, then F satisfies MM.*

2. Assume $N \geq 3$. If F satisfies MM and NVP, then it is implementable.

Proof: 1. Let $G \equiv (M_1, M_2 \ldots, M_N, g)$ implement F. Let $\theta, \phi \in \Theta$ and $a \in A$ be such that $L(a, i, \theta) \subset L(a, i, \phi)$ for all $i \in I$. There must exist $\bar{m} \in M_1 \times \ldots, M_N$ such that $g(\bar{m}) = a$ and $\bar{m} \in \text{NE}(G, \theta)$ i.e $\{g(m_i, \bar{m}_{-i}), m_i \in M_i\} \subset L(a, i, \theta)$ for all $i \in I$. Therefore $\{g(m_i, \bar{m}_{-i}), m_i \in M_i\} \subset L(a, i, \phi)$ for all $i \in I$. This implies $\bar{m} \in \text{NE}(G, \phi)$. Since G implements F, $a = g(\bar{m}) \in g(\text{NE}(G, \phi)) = F(\phi)$.

2. Assume $N \geq 3$ and let F satisfy MM and NVP. We explicitly construct the mechanism that implements F.

Let $M_i = \{(a_i, \theta_i, n_i, b_i, c_i) \in A \times \Theta \times \mathbb{N} \times A \times A | a_i \in F(\theta_i)\}, \;\; i \in I$. [1]

The mapping $g : M_1 \times \ldots, M_N \to A$ is described as follows:

(i) if $m_i = (a, \theta, ., ., .)$ for all $i \in I$, then $g(m) = a$.

(ii) if $m_i = (a, \theta, ., ., .)$ for all $i \in I - \{j\}$ and $m_j = (a_j, \phi, n_j, b_j, c_j)$, then

$$g(m) = \begin{cases} b_j & \text{if } b_j \in L(a, i, \theta) \\ a & \text{otherwise} \end{cases}$$

Observe that in order for (ii) to be well-defined, we require $N \geq 3$.

(iii) if (i) and (ii) do not hold, then $g(m) = c_k$ where k is the lowest index in the set of agents who announce the highest integer , i.e $k = \arg\min\{i \in I | n_i \geq n_j \text{ for all } j \neq i\}$.

Let $\phi \in \Theta$ be the true state of the world and let $a \in F(\phi)$.

[1] Here $\mathbb{N}$ is the set of all integers.

Claim **1** $F(\phi) \subset g(NE(G,\phi))$.

Proof: Consider $\bar{m}_i = (a, \phi, ., ., .)$ for all $i \in I$. Then $g(\bar{m}) = a$. Observe that if i deviates, she gets an outcome in the the set $L(a, i, \phi)$. Hence $\bar{m} \in \mathrm{NE}(G, \phi)$ which establishes the Claim. ■

Claim **2** $g(NE(G,\phi)) \subset F(\phi)$.

Proof: Let $\bar{m} \in \mathrm{NE}(G, \phi)$. We will show that $g(\bar{m}) \in F(\phi)$. We consider two cases.

Case 1: $\bar{m}_i = (a, \theta, ., ., .)$ for all $i \in I$. Therefore $g(\bar{m}) = a$. By construction, $\{g(m_i, \bar{m}_{-i}) | m_i \in M_i\} = L(a, i, \theta)$ for all $i \in I$. Since $\bar{m}$ is a Nash equilibrium in (G, ϕ), we must have $L(a, i, \theta) \subset L(a, i, \phi)$ for all $i \in I$. Since $a \in F(\theta)$, MM implies $a \in F(\phi)$.

Case 2: Case 1 does not hold. Let $g(\bar{m}) = a$. By construction $\{g(m_i, \bar{m}_{-i}) | m_i \in M_i\} = A$ for all i except perhaps some j (the exception occurs when all agents i other than j announce $\bar{m}_i = (a, \theta, ., ., .)$) Since $\bar{m}$ is a Nash equilibrium in (G, ϕ), it must be the case that $aR_i(\phi)b$ for all $b \in A$ for all $i \in I - \{j\}$. Since F satisfies NVP, $a \in F(\phi)$.

The two cases above exhaust all possibilities and complete the proof of the Claim. ■

Claims**??** and **??** complete the proof of the result. ■

8.6 Understanding Maskin Monotonicity

The following sccs are Maskin Monotonic.

1. The Pareto Efficient Correspondence.
2. The Walrasian Correspondence in exchange economies provided that all Walrasian allocations are interior.
3. The Individually Rational Correspondence in exchange economies.
4. The Pareto Efficient and Individually Rational Correspondence. In general, the intersection of two MM sccs is also MM, i.e if F and G are MM sccs and $F(\theta) \cap G(\theta) \neq \emptyset$ for all $\theta \in \Theta$, then $F \cap G$ also satisfies MM.
5. The dictatorship scc. There exists an agent, say i such that for all $\theta \in \Theta$, $d(\theta) = \{a \in A | aR_i(\theta)b \text{ for all } b \in A\}$.

The following are examples of sccs that violate MM.

State θ					State ϕ				
1	2	3	4	5	1	2	3	4	5
a	a	b	c	d	a	a	b	b	b
b	b	c	b	b	b	b	c	c	c
c	c	d	d	c	c	c	d	d	d
d	d	a	a	a	d	d	a	a	a

Table 7: Preferences in states θ and ϕ

1. The King Solomon scc.
2. Scoring methods, for example, the plurality rule. Let $A = \{a, b, c, d\}$ and $I = \{1, 2, 3, 4, 5\}$. Note that $a = F(\theta)$ and $b = F(\phi)$. However $aR_i(\theta)x \rightarrow aR_i(\phi)x$ for all $x \in A$ and for all $i \in I$. Hence MM is violated.
3. The class of scfs satisfying MM over "large domains" is small. For instance, if one considers scfs defined over the domain of all strict orderings, the only ones which satisfy MM and the "full range" condition are the dictatorial ones. Over the domain of all orderings, only the constant scf satisfies MM.

8.7 The case of $N = 2$

In this case an extra condition is required to ensure that equilibria can be sustained. Suppose agent 1 sends the message (a, θ) while 2 sends the message (b, ϕ) where $a \in F(\theta)$ and $b \in F(\phi)$. It could be that 1 is deviating unilaterally from the Nash equilibrium which supports (b, ϕ) or that 2 is deviating unilaterally from the Nash equilibrium which supports (a, θ). Clearly the resulting outcome must not upset either equilibrium, i.e. it must be an alternative in *both* $L(b, 1, \phi)$ and $L(a, 2, \theta)$. Hence a necessary condition (which does not appear in the $N \geq 3$ case) is that for all $\theta, \phi \in \Theta$ and $a, b \in A$ such that $a \in F(\theta)$ and $b \in F(\phi)$

$$L(b, 1, \phi) \cap L(a, 2, \theta) \neq \emptyset$$

Other conditions are also required for implementation.

8.8 Subgame Perfect Implementation

Here a mechanism is specified in extensive form. It consists of a (finite) game tree, a player partition, an information partition and a mapping which associates elements of A with every terminal node of the tree. A mechanism Γ together with a state $\theta \in \Theta$ is a game in extensive form. Let SPE(Γ, θ) denote the set of subgame pefect equilibrium outcomes of (Γ, θ). We say that the SCC F can be implemented if there exists a mechanism Γ such that $F(\theta) = \text{SPE}(\Gamma, \theta)$ for all $\theta \in \Theta$.

It is clear that an scc that can be implemented in Nash equilibrium can be implemented in subgame perfect equilibrium by simply using the Maskin mechanism which is a simultaneous move game and has n proper subgames. But can more be implemented by using extensive-form mechanisms and the notion of subgame perfect Nash equilibrium? The following example answers this question in the affirmative.

Example **3** $A = \{a, b, c\}$, $I = \{1, 2, 3\}$ and $\Theta = \{$ all strict orderings over $A\}$. consider the following scf: $f(\theta) = \arg\max_{R_1(\theta)}\{a,$ majority winner over $\{b, c\}\}$. We claim that f is not monotonic.

State θ			State ϕ		
1	2	3	1	2	3
b	c	c	b	b	b
a	b	b	a	c	c
c	a	a	c	a	a

Table 8: Preferences in states θ and ϕ

Observe that $f(\theta) = a$ and $L(a, i, \theta) \subset L(a, i, \phi)$ for all $i \in I$. However $f(\phi) = b$. Therefore f does not satisfy MM and is not implementable.

However f can be implemented in subgame perfect equilibrium by the extensive mechanism in Figure **??**. Observe that the backwards induction outcome at the node z is always the majority winner of $\{b, c\}$.

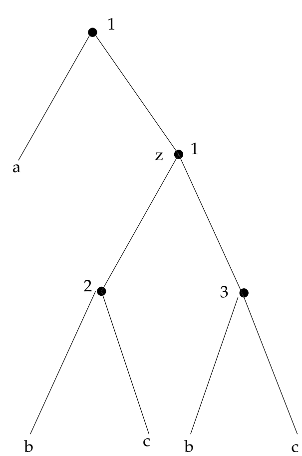

Figure 2: Tree implementing f

QUESTION: How much does the class of implementable sccs expand when we consider subgame perfect rather than Nash equilibrium?

The answer is, surprisingly quite considerably as illustrated by the result below. First we make some simplifying assumptions on the set Θ.

A.1 For all $\theta \in \Theta$ and $i, j \in I$, $\max(R_i(\theta), A) \cap \max(R_j(\theta), A) = \emptyset$.

A.2 For all $\theta, \phi \in \Theta$, there exists $k \in I$ and $x, y \in A$ such that (i) $xR_k(\theta)y$ (ii) $yP_k(\phi)x$ and (iii) $x, y \notin \max(R_i(\psi), A)$ for any $i \in I$ and $\psi \in \Theta$.

According to $A.1$, no two agents have a common maximal element in any state. This rules out the possibility of having any equilibria in the "integer game". According to $A.2$, there must exist an agent whose preferences are "reversed" over a pair of outcomes in two distinct states of the world. Moreover, neither of these outcomes is maximal for any agent in any state of the world.

Both the assumptions above are satisfied in any environment where there is a transferable good which agents' prefer monotonically (money?). This implies that the maximal alternative for an agent will be one where she gets the entire amount of this good. Clearly, maximal elements of different agents must be distinct in every state of the world.

DEFINITION **33** *The scc F is interior if there does not exist a, $\theta, \phi \in \Theta$ and $i \in I$ such that $a \in F(\theta)$ and $a \in \max(R_i(\phi), A)$.*

In exchange economies, an interior scc never gives all resources to a single agent.

THEOREM **6 (Moore-Repullo (1988), Abreu-Sen (1990))** *Assume $N \geq 3$. Let F be any Pareto-efficient, interior scc defined over an environment satisfying $A.1$ and $A.2$. Then F can be implemented in subgame perfect Nash equilibrium.*

Proof: Let $a \in F(\theta) - F(\phi)$. From $A.2$, it follows that there exists $k \in A$ and $x, y \in A$ such that $xR_k(\theta)y$, $yP^(\phi)x$ and x, y are not $R_i(\psi)$ maximal for any $i \in I$ and $\psi \in \Theta$. Henceforth, we refer to these outcomes and agents as $x(a, \theta, \phi)$, $y(a, \theta, \phi)$ and $k(a, \theta, \phi)$ respectively. Since F is efficient, there exists an agent $j(a, \theta, \phi)$ such that $aR_{j(a,\theta,\phi)}x(a, \theta, \phi)$.

The mechanism Γ has two stages.

STAGE 0

Let $M_i^0 = \{(a_i, \theta_i, n_i^0, c_i^0) \in A \times \Theta \times \mathbb{N} \times A | a_i \in F(\theta_i)\}, \;\; i \in I$.

(i) if $m_i^0 = (a, \theta, ., .)$ for all $i \in I$, then the outcome is a. STOP

(ii) if $m_i^0 = (a, \theta, ., .)$ for all $i \in I - \{j\}$ and $m_j^0 = (a_j, \phi, ., .)$ and

(iia) $j = j(a, \theta, \phi)$, then go to Stage 1.

(iib) $j \neq j(a, \theta, \phi)$, then the outcome is a. STOP

(iii) if (i) and (ii) do not hold, then the outcome is c_k^0 where k is the lowest index in the set of agents who who announce the highest integer , i.e $k = \arg\min\{i \in I | n_i^0 \geq n_j^0 \text{ for all } j \neq i\}$. STOP

STAGE 1

Let $M_i^1 = \{(B_i, n_i^1, c_i^1) \in \{0,1\} \times \mathbb{N} \times A\}, \ \ i \in I$.

(i) if $\#\{i | m_i^1 = (0, ., .)\} \geq N - 1$ then the outcome is c_j^1 where $j = j(a, \theta, \phi)$. STOP

(ii) if $\#\{i | m_i^1 = (1, ., .)\} \geq N - 1$ then

(iia) the outcome is $x(a, \theta, \phi)$ if $m_k^1 = (1, ., .)$ where $k = k(a, \theta, \phi)$. STOP

(iib) the outcome is $y(a, \theta, \phi)$ if $m_k^1 = (0, ., .)$ where $k = k(a, \theta, \phi)$. STOP

(iii) if (i) and (ii) do not hold, then the outcome is c_k^1 where k is the lowest index in the set of agents who who announce the highest integer , i.e $k = \arg\min\{i \in I | n_i^1 \geq n_j^1 \text{ for all } j \neq i\}$. STOP

Let θ be the true state.

Claim **3** *$F(\theta) \subset SPE(\Gamma, \theta)$.*

Consider the following strategy-profile:

- $m_i^0 = (a, \theta, ., ., .)$ for all $i \in I$
- $m_i^1 = (1, ., .)$ for all $i \in I$ and for all Stage 0 histories.

The outcome is a. We first need to check that these strategies induce Nash equilibrium in Stage 1. Suppose Stage 1 has been reached because $j(a, \theta, \phi)$ deviated in Stage 0 and announced $m_j^0 = (b, \phi, ., ., .)$. On the path specified by these strategies, the outcome is $x(a, \theta, \phi)$. The only deviation which can change the outcome is by $k(a, \theta, \phi)$ who can obtain $y(a, \theta, \phi)$. But $x(a, \theta, \phi) R_{k(a,\theta,\phi)} y(a, \theta, \phi)$ by assumption, so that this agent will not deviate. Now, in Stage 0, the only agent who deviation "matters" is agent $j(a, \theta, \phi)$. By deviating, this agent will obtain $x(a, \theta, \phi)$. Since $a R_{j(a,\theta,\phi)} x(a, \theta, \phi)$ so that this agent has no incentive to deviate. This establishes the Claim.

Claim 4 $SPE(\Gamma, \theta) \subset F(\theta)$.

Observe first that as a consequence of $A.1$ and $A.2$ all candidate equilibria must be of the following form:

1. $m_i^0 = (a, \phi, ., ., .)$ for all $i \in I$. in Stage 0. In Stage 1 following an arbitrary history either A or B below must hold.

A. $m_i^1 = (1, ., .)$ for all $i \in I$

B. $m_i^1 = (0, ., .)$ for all $i \in I$.

Note that if messages in any stage are non-unanimous, at least two agents can trigger the integer game. Since this game has no equilibrium because of our assumptions, such message profiles cannot be part of an equilibrium.

In the candidate equilibrium, the outcome is a. If $a \in F(\theta)$, there is nothing to prove; assume therefore that $a \notin F(\theta)$, i.e. $a \in F(\phi) - F(\theta)$. Consider a deviation by agent $j = j(a, \phi, \theta)$ who announce $m_j^0 = (b, \theta, ., .)$ where $b \in F(\theta)$ and sends the game to Stage 1. Suppose that A applies in the continuation game. The outcome is $x(a, \phi, \theta)$. If agent $k = k(a, \phi, \theta)$ deviates by announcing $m_k^1 = (0, ., .)$, the outcome is $y(a, \phi, \theta)$. Since $y(a, \phi, \theta) P_{k(a,\phi,\theta)}(\theta) x(a, \phi, \theta)$ by assumption, k will indeed deviate. Suppose then that B applies in Stage 2. The outcome is then c_j^1. Since $a \in F(\phi)$ and F is interior, there exists c_j^1 such that $c_j^1 P_j(\theta) a$. Clearly j can obtain c_j^1 by his Stage 0 deviation. Therefore, if the candidate equilibrium strategies are indeed an equilibrium, it must be the case that $a \in F(\theta)$. This proves the Claim. ∎

Observation **2** The assumption of efficiency in the result above is not essential (see Moore and Repullo (1988)). It only ensures that two-stage mechanisms suffice. Abreu and Sen (1990) prove a more general result which also applies to voting environments. In fact, they establish the counterpart of MM for subgame perfect implementation. This condition is significantly weaker than MM but there are still important sccs that fail to satisfy it.

8.9 Discussion

There is a large body of literature which establishes that the scope for implementation increases very dramatically if either (i) the solution concept is "refined" from Nash to subgame-perfect, iterated elimination of dominated strategies, elimination of weakly dominated strategies, trembling-hand perfect equilibrium and so on and (ii) randomization is allowed in the mechanism and the notion of implementation is weakened to concepts like "virtual implementation" etc.

9 Incomplete Information

In the incomplete information model, each agent receives private information but cannot deduce the state of the world from the information she receives. In other words, an agent does not know the information received by other agents. The mechanism designer does not observe the information received by any agent. Consider the following well-known examples.

9.1 Examples

9.1.1 Voting

Assume that there are N voters and assume for convenience that N is odd. Voters have to collectively select one of the two proposals a or b. Each voter i either believes that a is better than b or b is better than a. Importantly, these preference ordering is known *only* to i. Voters therefore need to reveal their preferences by voting.

Consider the majority voting rule: all voters vote either a or b and the proposal which gets the highest aggregate number of votes is selected. Voters realize that they are playing a game. They can vote either a or b (their strategy sets) and the outcome and payoff depends not only on how they vote but also on how *everyone else* votes. How will they vote? Note that voting according to their true preferences is a weakly dominant strategy. Their vote does not matter unless the other voters are exactly divided in their opinion on a and b. In this case a voter gets to choose the proposal she wants. She will clearly hurt herself by misrepresenting her preferences.

What if there are three proposals or candidates a, b and c? Consider a generalization of the rule proposed above. Each voter votes for her best proposal. Select the proposal which is best for the largest number number of voters. If no such proposal exists, select a (which can be thought of as a *status quo* proposal).

What behaviour does this rule induce? Is truth-telling a dominant strategy once again? No, as the following example for three players demonstrates.

1	2	3
c	b	a
b	a	b
a	c	c

Table 9: Voter Preferences

Now suppose voter 1's true preference is c better than b than a while she believes that voters 2 and 3 are going to vote for b and a respectively. Then voting truthfully will yield a while lying and voting for b will get b which is better than a according to her *true* preferences.

Are there voting rules which will induce voters to reveal their true preferences? Note that if voters do not vote truthfully, the actual outcome could be very far from the desired one.

9.1.2 Bilateral Trading

There are two agents, a seller S and a buyer B. The seller has a single object which the buyer is potentially interested in buying. The seller and buyer have valuations v_s and v_b which are known only to themselves. Assume that they are independently and identically distributed random variables. Assume further that they are uniformly distributed on $[0, 1]$.

Consider the following trading rule proposed by Chatterjee and Samuelson . Seller and buyer announce "bids" x_s and x_b. Trade takes place only if $x_b > x_s$. If trade occurs, it does so at price $\frac{x_b+x_s}{2}$. Agents have quasi-linear utility, i.e. if no trade occurs both agents get 0; if it occurs, then payoffs for the buyer and seller are $v_b - \frac{x_b+x_s}{2}$ and $\frac{x_b+x_s}{2} - v_s$ respectively.

This is a game of incomplete information. A linear Bayes-Nash equilibrium of the game exists where $x_b = \frac{2}{3}v_b + \frac{1}{12}$ and $x_s = \frac{2}{3}v_s + \frac{1}{4}$. Therefore trade takes place only if $v_b - v_s > \frac{1}{4}$. However efficiency would require trade to take place whenever $v_b > v_s$. There are realizations of v_b, v_s where there is no trade in equilibrium where it would be efficient to have it.

Are there other trading rules where agents participate voluntarily and equilibrium outcomes are always efficient?

9.2 A General Model

As before, the set of agents is $I = \{1, \ldots, N\}$, the set of feasible alternatives or outcomes or allocations is A. Each agent i has some private information $\theta_i \in \Theta_i$. The parameter θ_i is often referred to as agent i's *type*. Agent i has a payoff function $v_i : \Theta_i \times A \to \Re$. Thus every realization of θ_i determines a payoff function for i. [2] A profile $\theta \equiv (\theta_1, \ldots, \theta_N)$ is an N tuple which describes the state of the world. The notation $(\theta_i^{'}, \theta_{-i})$ will refer to the profile where the i^{th} component of the profile θ is replaced by $\theta_i^{'}$.

Definition **34** *A Social Choice Function (scf) is a mapping* $f : \Theta_1 \times \Theta_2 \times ... \times \Theta_N \to A$.

As before, a scf represents the collective goals of the agents and the objectives of a Principal/Designer.

Definition **35** *A SCF f is strategy-proof if*

[2] A more general model is one where the payoff function for agent i is $v_i : \Theta_1 \times \ldots \times \Theta_N \times A \to \Re$, i.e an agent's payoff depends on the types of all agents. This is the model of common or interdependent valuations and has many interesting applications.

$$v_i(f(\theta), \theta_i) \geq v_i(f(\theta_i', \theta_{-i}), \theta_i)$$

holds for all θ_i, θ_i', θ_{-i} and $i \in I$.

If a scf is strategy-proof, then truth-telling is a dominant strategy for each agent. Strategy-proofness is dominant-strategy incentive-compatibility.

An alternative (and weaker) notion of incentive compatibility is Bayes-Nash incentive-compatibility. Here truth-telling gives a higher expected utility than lying for each agent when these expectations are computed with respect to beliefs regarding the types of other agents and assuming that other agents are telling the truth.

Assume that $\mu_i : \Theta_1 \times ... \times \Theta_N \to [0, 1]$ denotes the beliefs of agent i over the possible types of other agents, i.e $\mu(\theta) \geq 0$ and $\int_\theta d\mu_i(\theta) = 1$. Let $\mu_i(.|\theta_i)$ denote agent i's beliefs over the types of other agents conditional on her type being θ_i.

Definition **36** *A scf f is Bayesian incentive-compatible (BIC) if*

$$\int_{\theta_{-i}} v_i(f(\theta), \theta_i) d\mu_i(\theta_{-i}|\theta_i) \geq \int_{\theta_{-i}} v_i(f(\theta_i', \theta_{-i}), \theta_i) d\mu_i(\theta_{-i}|\theta_i)$$

for all θ_i, $i \in I$.

A scf which is strategy-proof is BIC with respect to *all* priors. One goal of the theory is to identify scfs which are strategy-proof or BIC. Another one is to identify the "best" or optimal scf within the class of incentive-compatible scfs. For instance, we might wish to design an auction which maximizes expected revenue to the seller and so on.

Two fundamental issues are:

- The choice of a solution concept, i.e. strategy-proofness vs BIC. As we have remarked earlier, the former is a more robust notion (we can be more confident that agents will play weakly dominant strategies when they exist). However the difficulty is that it the class of scfs which satisfy it are severely restricted and is smaller than the class of scfs that satisfy BIC.

- The domain of preferences. Specifically, what is the structure of the set A, the sets Θ_i and the nature of the function v_i? As subsequent examples will show, these are determined by the particular choice of model. For instance, a critical choice is whether or not monetary compensation is allowed (voting vs. exchange).

Observation **3** A question which arises naturally is the following: why are we interested in truth-telling? Why don't we consider a general mechanism where people send messages from some artificially constructed set? Perhaps the mechanism could be constructed so that the equilibrium (either dominant strategies or Bayes-Nash) outcomes at any profile are exactly the "optimal" ones for that profile? There is no sensible notion of truth-telling in such mechanisms. The answer is that there is no loss of generality in restricting attention to direct mechanisms (where agents directly reveal their types and the scf itself is the mechanism) and requiring truth-telling to be an equilibrium. This simple fact/observation is known as *The Revelation Principle.*

9.3 The Complete Domain

In this section voting models will be considered. These are models where monetary compensation is not permitted. The goal will be to present a well-known result which characterizes the class of strategy-proof scfs.

The set $A = \{a, b, c, ...\}$ is a set of m proposals/candidates/alternatives and $I = \{1, 2, \ldots, N\}$ is a set of voters. Voter i's type, θ_i is his ranking of the elements of the set A. This ranking will be more conveniently written as P_i. For convenience, we assume that P_i is a linear order i.e. it is complete, reflexive, transitive and anti-symmetric. Hence, for all $a, b \in A$, aP_ib is interpreted as "a is strictly preferred to b under P_i".

Let $\mathbb{P}$ be the set of all linear orderings over A (there are $m!$ such orders). A preference profile $P = (P_1, ..., P_N) \in \mathbb{P}^N$ is an n -list of orderings, one for each voter. A scf or a voting rule f is a mapping $f : \mathbb{P}^N \to A$.

The strategy-proofness property introduced in the previous section is restated below for this environment.

Definition **37** *The scf f is manipulable if there exists a voter i, a profile $P \in \mathbb{P}^N$ and an ordering P_i' such that*

$$f(P_i', P_{-i})P_if(P_i, P_{-i})$$

Definition **38** *The SCF f is strategy-proof if it is not manipulable.*

One class of voting rule which is always strategy-proof is the constant SCF which selects the same alternative at all profiles. In order to rule out this possibility, it will be assumed that SCFs under consideration satisfy the property of *unanimity.*

For all voters i and $P_i \in \mathbb{P}$, let $\tau(P_i)$ denote the maximal element in A according to P_i.

Definition **39** *The SCF f satisfies unanimity if $f(P) = a$ whenever $\tau(P_i) = a$ for all $i \in I$.*

Definition **40** *The voting rule f is dictatorial if there exists $i \in I$ such that for all $P \in \mathbb{P}^N$, $f(P) = \tau(P_i)$.*

THEOREM **7 (Gibbard (1973), Satterthwaite (1975))** *Assume $m \geq 3$. If f satisfies unanimity then it is strategy-proof if and only if it is dictatorial.*

Proof: Sufficiency is obvious and we only prove necessity.

STEP 1: We prove the result in the case of $N = 2$. Let $I = \{1, 2\}$ and assume $f : \mathbb{P}^2 \to A$ satisfies unanimity and strategy-proofness.

CLAIM **5** *Let $P = (P_1, P_2)$ be such that $\tau(P_1) \neq \tau(P_2)$. Then $f(P_1, P_2) \in \{\tau(P_1), \tau(P_2)\}$.*

Proof: Suppose not i.e. suppose that there exists P_i, P_2 and a, b, c such that $\tau(P_1) = a \neq b = \tau(P_2)$ and $f(P_1, P_2) = c \neq a, b$. Let $P_1' = \begin{pmatrix} a \\ b \\ \vdots \end{pmatrix}$ and $P_2' = \begin{pmatrix} b \\ a \\ \vdots \end{pmatrix}$.

If $f(P_1', P_2) = a$ then voter 1 manipulates at (P_1, P_2) by voting P_1' because $a = f(P_1', P_2) P_1 f(P_1, P_2) = c$. If on the other hand, $f(P_1', P_2) = x$ where $bP_i'x$, then voter 1 manipulates at (P_1', P_2) by voting $\bar{P}_1$ where $\tau(\bar{P}_1) = b$. Then $f(\bar{P}_1, P_2) = b$ (by unanimity) and $b = f(\bar{P}_1, P_2) P_1' f(P_1', P_2) = x$. Hence $f(P_1', P_2) = b$.

Now suppose $f(P_1', P_2') = x \neq b$, i.e. $bP_2'x$. Then 2 will manipulate by voting P_2 because $b = f(P_1', P_2) P_2' f(P_1', P_2') = x$. Hence $f(P_1', P_2') = b$.

By a symmetric argument, $f(P_1, P_2') = f(P_1', P_2') = a$. However this contradicts our earlier conclusion that $f(P_1', P_2') = b$. ■

CLAIM **6** *Let $P, \bar{P} \in \mathbb{P}^2$ be such that $\tau(P_1) = a \neq b = \tau(P_2)$ and $\tau(\bar{P}_1) = c \neq d = \tau(\bar{P}_2)$. Then $[f(P) = \tau(P_1)] \to [f(\bar{P}) = \tau(\bar{P}_1)]$ and $[f(P) = \tau(P_2)[\to [f(\bar{P}) = \tau(\bar{P}_2)]$.*

Proof: Let $P_1 = \begin{pmatrix} a \\ \vdots \end{pmatrix}$ and $P_2 = \begin{pmatrix} b \\ \vdots \end{pmatrix}$. Assume, without loss of generality that $f(P_1, P_2) = a = \tau(P_1)$. Note that for all P_1' such that $\tau(P_1') = a$, we must have $f(P_1', P_2) = a$. Hence we can assume that c is the second ranked outcome at P_1, i.e. we can assume that $P_1 = \begin{pmatrix} a \\ c \\ \vdots \end{pmatrix}$.

Let $\bar{P}_1 = \begin{pmatrix} c \\ a \\ \vdots \end{pmatrix}$. By Claim 1, $f(\bar{P}_1, P_2) \in \{b, c\}$. Suppose $f(\bar{P}_1, P_2) = b$. Then 1 manipulates at $(\bar{P}_1, P_2)$ by voting P_1 because $a = f(P_1, P_2) \bar{P}_1 f(\bar{P}_1, P_2) = b$. Hence $f(\bar{P}_1, P_2) = c$.

Observe that for any $P_2' \in \mathbb{P}$ such that $\tau(P_2') = b$ we must have $f(\bar{P}_1, P_2') = c$. If this is not true, then Claim 1 would imply $f(\bar{P}_1, P_2') = b$ and 2 would manipulate at $(\bar{P}_1, P_2)$

by voting P'_2. We can therefore assume without loss of generality that the second ranked alternative in P_2 is d. Now Claim 1 implies that $f(\bar{P}_1, \bar{P}_2) \in \{c, d\}$. If $f(\bar{P}_1, \bar{P}_2) = d$, then 2 would manipulate at $(\bar{P}_1, P_2)$ via $\bar{P}_2$. Hence $f(\bar{P}_1, \bar{P}_2) = c$. ■

Claims **??** and **??** establish that f is dictatorial.

STEP 2: We now show that the Theorem holds for general N. In particular, we show that the following two statements are equivalent.

(a) $f : \mathbb{P}^2 \to A$ is strategy-proof and satisfies unanimity $\Rightarrow f$ is dictatorial

(b) $f : \mathbb{P}^N \to A$ is strategy-proof and satisfies unanimity $\Rightarrow f$ is dictatorial, $N \geq 2$.

(b) $\Rightarrow$ (a) is trivial. We now show that $(a) \Rightarrow$(b). Let $f : \mathbb{P}^N \to A$ be a non-manipulable scf satisfying unanimity. Pick $i, j \in I$ and construct a scf $g : \mathbb{P}^2 \to A$ as follows: for all $P_i, P_j \in \mathbb{P}$, $g(P_i, P_j) = f(P_i, P_j, P_j \dots, P_j)$.

Since f satisfies unanimity, it follows immediately that g satisfies this property. We claim that g is strategy-proof. If i can manipulate g at (P_i, P_j), then i can manipulate f at $(P_i, P_j, ..., P_j)$ which contradicts the assumption that f is strategy-proof. Suppose j can manipulate g, i.e. there exists $P_i, P_j, \bar{P}_j \in \mathbb{P}$ such that $b = g(P_i, \bar{P}_j) P_j g(P_i, P_j) = a$. Now consider the sequence of outcomes obtained when individuals other than i progressively switch preferences from P_j to $\bar{P}_j$. Let $f(P_i, \bar{P}_j, P_j, ..., P_j) = a_1$. If a and a_1 are distinct, then $a P_j a_1$ since f is non-manipulable. Let $f(P_i, \bar{P}_j, \bar{P}_j, P_j, ..., P_j) = a_2$. Again, since f is non-manipulable, $a_1 P_j a_2$ whenever a_1 and a_2 are distinct . Since P_j is transitive, $a P_j a_2$. Continuing in this manner to the end of the sequence, we obtain $a P_j b$ which contradicts our initial assumption.

Since g is strategy-proof and satisfies unanimity, statement (a) applies, so that either i or j is a dictator. Let $O_{-i}(P_i) = \{a \in A | a = f(P_i, P_{-i}) \text{ for some } P_{-i} \in \mathbb{P}^{N-1}\}$. We claim that $O_{-i}(P_i)$ is either a singleton or the set A. Suppose i is the dictator in the scf g. Let $P_i \in \mathbb{P}$ with $r_1(P_i) = a$. Since g satisfies unanimity, it follows that $g(P_i, P_j) = a$ where $r_1(P_j) = a$. Therefore $a \in O_{-i}(P_i)$. Suppose there exists $b \neq a$ such that $b \in O_{-i}(P_i)$, i.e. there exists $P_{-i} \in \mathbb{P}^{N-1}$ such that $f(P_i, P_{-i}) = b$. Let $\bar{P}_j \in \mathbb{P}$ be such that $r_1(\bar{P}_j) = b$. Observe that $f(P_i, \bar{P}_j, \dots, \bar{P}_j) = b$ (progressively switch preferences of individuals j other than i from P_j to $\bar{P}_j$ and note that the outcome at each stage must remain b; otherwise an individual who can shift the outcome from b will manipulate). Therefore, $g(P_i, \bar{P}_j) = b$. This contradicts the assumption that i is the dictator. Therefore, $O_{-i}(P_i)$ is a singleton. Suppose j is the dictator. Then $A = \{a \in A | g(P_i, P_j) = a \text{ for some } P_j \in \mathbb{P}\} \subseteq O_{-i}(P_i)$, so that $O_{-i}(P_i) = A$.

We now complete the proof by induction on N. Observe that statements (a) and (b) are identical when $N = 2$. Suppose it is true for all societies of size less than or equal to $N - 1$. Consider the case where there are N individuals. Pick $i \in I$. From the earlier argument, either $O_{-i}(P_i)$ is a singleton or the set A. Suppose the latter case holds. Fix

$P_i \in \mathbb{P}$ and define a scf $g : \mathbb{P}^{N-1} \to A$ as follows : $g(P_{-i}) = f(P_i, P_{-i})$ for all $P_{-i} \in \mathbb{P}^{N-1}$. Since $O_{-i}(P_i) = A$, g satisfies unanimity because it is strategy-proof and its range is A. Applying the induction hypothesis, it follows that there exists an individual $j \neq i$ who is a dictator. We need to show that the identity of this dictator does not depend on P_i. Suppose that there exists $P_i, \bar{P}_i \in \mathbb{P}$ such that the associated dictators are j and k respectively. Pick $a, b \in A$ such that aP_ib and $a \neq b$. Pick $P_j, P_k \in \mathbb{P}$ such that $r_1(P_j) = b$ and $r_1(P_k) = a$. Let P_{-i} be the $N-1$ profile where j has the ordering P_j and k has the ordering P_k. Then $f(P_i, P_{-i}) = b$ and $f(\bar{P}_i, P_{-i}) = a$ and i manipulates at (P_i, P_{-i}). Therefore f is dictatorial. Suppose then that $O_{-i}(P_i)$ is a singleton. We claim that $O_{-i}(P_i)$ must be a singleton for all $P_i \in \mathbb{P}$. Suppose not, i.e. there exists $\bar{P}_i \in \mathbb{P}$ such that $O_{-i}(\bar{P}_i) = A$. From our earlier argument, there exists an individual $j \neq i$ who is a dictator. But this would imply that $O_{-i}(\bar{P}_i)$ is a singleton. Therefore, it must be the case that $O_{-i}(P_i)$ is a singleton for all $P_i \in \mathbb{P}$. But this implies that individual i is a dictator. ■

Observation **4** *There is a large class of scfs, called committee rules which are strategy-proof in the case where $|A| = 2$.*

9.4 Maskin Monotonicity and Strategy-Proofness

There is a close connection between scfs satisfying Maskin Monotonicity and strategy-proof scfs as the Proposition below shows.

Proposition **10 (Muller and Satterthwaite (1977))** *Let $\mathbb{D} \subset \mathbb{P}$. If a scf $f : \mathbb{D}^N \to A$ is strategy-proof, it satisfies MM. If a scf $f : \mathbb{P}^N \to A$ satisfies MM, it is strategy-proof.*

Proof: Let $f : \mathbb{D}^N \to A$ be a strategy-proof scf. Let $P, \bar{P} \in \mathbb{P}^N$ and $a \in A$ be such that $f(P) = a$ and $aP_ib \to a\bar{P}_ib$ for all $b \neq a$ and $i \in I$. Let $f(\bar{P}_1, P_{-1}) = c$. Assume $c \neq a$. Since P_1 is a strict order, either cP_1a or aP_1c must hold. Suppose the former is true. Then agent 1 manipulates f at P via $\bar{P}_1$. Suppose aP_1c. Then $a\bar{P}_1c$ by hypothesis. In this case, agent 1 manipulates f at $(\bar{P}_1, P_{-1})$ via P_1. Hence $f(\bar{P}_1, P_{-1}) = a$. Now progressively switch the preferences of agents 2 through N from $P_2 \dots P_N$ to $\bar{P}_2 \dots \bar{P}_N$. At each stage, the argument above can be applied to show that the outcome remains fixed at a. Hence $f(\bar{P}) = a$ and f satisfies MM.

Suppose that $f : \mathbb{P}^N \to A$ satisfies MM but is not strategy-proof. Thus, there exists $i \in I$, $P \in \mathbb{P}^N$ and $\bar{P}_i \in \mathbb{P}$ such that $f(\bar{P}_i, P_{-i})P_if(P)$. Let $f(P) = a$ and $f(\bar{P}_i, P_{-i}) = b$. Let $P_i' \in \mathbb{P}$ be an ordering where b is ranked first and a is ranked second. Note that $b\bar{P}_ix \to bP_i'x$ for all $x \in A$. Since $f(\bar{P}_i, P_{-i}) = b$ and f satisfies MM, we must have $f(P_i', P_{-i}) = b$. Also observe that $aP_ix \to aP_i'x$ for all $x \in A$ (since the only alternative ranked above a in P_i' is

b which was ranked above a in P_i). Since $f(P) = a$, MM implies that $f(P_i', P_{-i}) = a$. This contradicts our earlier conclusion. ■

Corollary **3** *Assume $|A| \geq 3$. If a scf $f : \mathbb{P}^N \rightarrow A$ satisfies MM and unanimity, it must be dictatorial.*

Proof: By Proposition **??**, f must be strategy-proof. The conclusion now follows from Theorem **??**. ■

9.5 Restricted Domains: Single-Peaked Domains

A natural way of evading the negative conclusions of the Gibbard-Satterthwaite Theorem is to assume that admissible preferences are subject to certain restrictions. One of the most natural domain restrictions is that of single-peaked preferences. These domains were introduced by Black (1948) and Inada (1964) and form the cornerstone of the modern theory of political economy.

We assume that there is an exogenous strict (or linear) order $<$ on the set A. If $a < b$, we say that a is to the left of b or equivalently, b is to the right of a. Suppose $a < b$. We let $[a, b] = \{x : a < x < b\} \cup \{a, b\}$ i.e. $[a, b]$ denotes all the alternatives which lie "between" a and b including a and b.

Definition **41** *The ordering P_i is single-peaked if*

1. *For all $a, b \in A$, $[b < a < \tau(P_i)] \Rightarrow [aP_ib]$.*
2. *For all $a, b \in A$, $[\tau(P_i) < a < b] \Rightarrow [aP_ib]$.*

We let $\mathbb{D}^{SP}$ denote the set of all single-peaked preferences. Clearly $\mathbb{D}^{SP} \subset \mathbb{P}$. It is worth emphasizing that the order $<$ is fixed for the domain $\mathbb{D}^{SP}$.

For expositional convenience, we have considered the case where A is finite and single-peaked preferences are linear orders. There are no conceptual or technical difficulties in extending these ideas to the case where, for instance $A = [0, 1]$ and single-peaked preferences admit indifference. The set A could be the proportion of the national budget to the spent on primary education. If a voter's peak is 0.25, then she would prefer an expenditure of 0.20 to 0.16; she would also prefer 0.40 to 0.75. Note however that no restrictions are placed on the comparison between 0.40 and 0.20.

Example **4** *Let $A = \{a, b, c\}$ and assume $a < b < c$. Table* **??** *shows all single-peaked preferences in this case.*

a	b	b	c
b	a	c	b
c	c	a	a

Table 10: Single-Peaked Preferences

Observation **5** In general, $\#\mathbb{D}^{SP} = 2^{m-1}$. Recall that $|A| = m$.

Let $B \subset A$ and assume $|B| = 2k+1$ for some positive integer k. We say that $b \in B$ is the *median* of B if (i) $|\{x \in B : x \leq b\}| \geq k+1$ and (ii) $|\{x \in B : b \leq x\}| \geq k+1$. In other words, there are at least $k+1$ alternatives including b which lie to the left of b and $k+1$ alternatives including b which lie to the right of b.

We denote the median of B by $\text{med}(B)$.

A scf is *anonymous* if its outcome at any profile is unchanged if the names of the agents are permuted.

The following is a characterization of strategy-proof, anonymous and efficient scfs defined over the domain of single-peaked preferences.

Theorem **8 (Moulin (1980))** *The following two statements are equivalent.*

1. *The scf* $f : [\mathbb{D}^{SP}]^N \rightarrow A$ *is strategy-proof, efficient and anonymous.*
2. *There exists a set* $B \subset A$ *with* $|B| = N-1$ *such that for all* $P \in [\mathbb{D}^{SP}]^N$, $f(P) = \textit{med}\{\{\tau(P_1), \ldots, \tau(P_N)\} \cup B\}$.

Proof: We will prove the result only for the case $N = 2$. We start by showing $1 \Rightarrow 2$.

We begin with a preliminary result which states that the outcome of a strategy-proof scf can only depend on the peaks of agent 1 and 2's preferences.

Claim **7** *Let* f *be a strategy-proof scf satisfying unanimity. Let* P, $\bar{P}$ *be profiles such that* $\tau(P_1) = \tau(\bar{P}_1)$ *and* $\tau(P_2) = \tau(\bar{P}_2)$. *Then* $f(P) = f(\bar{P})$.

Proof: Let P be a profile and $\bar{P}_1$ be a single-peaked preference such that $\tau(P_1) = \tau(\bar{P}_1) = a$. Suppose $f(P) = x \neq y = f(\bar{P}_1, P_2)$. Let $\tau(P_2) = b$ and assume without loss of generality that $a < b$. Suppose that x and y both lie to the left of a. Assume without loss of generality that $x < y$. Since $x < y < a$, voter 1 will manipulate at P via $\bar{P}_1$. By a similar argument, a cannot lie to the left of both x and y. Therefore x and y must lie on different sides of a. Assume without loss of generality $x < a$. Since $a < b$, voter 2 will manipulate at P via an ordering which has a as its peak. By unanimity, this will yield a which he prefers to x. Hence $f(P) = f(\bar{P}_1, P_2)$. An identical argument for a change in voter 2's preferences yields $f(\bar{P}_1, P_2) = f(\bar{P})$. ■

Now assume that f is strategy-proof, anonymous and efficient. Let a and b be the left-most and right-most alternative in A respectively. Let P' be the profile where $\tau(P_1') = a$ and $\tau(P_2') = b$. Let $f(P') = x$. We will show that for all profiles P, $f(P) = \text{ med}\{\tau(P_1), \tau(P_2), x\}$, i.e. the set B in the statement of the Theorem is $\{x\}$. There are three cases to consider.

Case 1: x is distinct from both a and b. Let P be a profile where $\tau(P_1) \in [a, x]$ and $\tau(P_2) \in [x, b]$. We claim that $f(P) = x$. Suppose first that $f(P_1, P_2') = y \neq x$. If $x < y$, then $\tau(P_1) < x < y$ implies that voter 1 manipulates at P via P_1'. If $y < x$, then $a, y < x$ implies that voter 1 manipulates at (P_1', P_2) via P_1. Therefore $f(P_1, P_2') = x$. By replicating these arguments for voter 2, we can conclude that $f(P) = x$.

Let P be a profile where $\tau(P_1) < \tau(P_2) < x$. We claim that $f(P) = \tau(P_2)$. By efficiency $f(P) \in [\tau(P_1), \tau(P_2)]$. Suppose $f(P) = y < \tau(P_2)$. Applying Claim **??**, we can assume without loss of generality that P_2 is an ordering where all alternatives to the right of $\tau(P_2)$ are preferred to all alternatives to the left of $\tau(P_2)$. Therefore voter 2 will manipulate via an ordering whose peak is to the right of x. By the arguments in the previous paragraph, the outcome of such a profile is x which is better than y according to P_2.

Finally suppose that P is a profile where $x < \tau(P_1) < \tau(P_2)$. The arguments in the previous paragraph can be adapted in a straightforward manner to yield the conclusion that $f(P) = \tau(P_1)$.

Applying anonymity, it follows that whenever voters have peaks on either side of x, the outcome is x; whenever both voters have peaks to the left of x, the outcome is the right-most peak of the two peaks and whenever both voters have peaks to the right of x, the outcome is the left-most of the two peaks. Clearly, in all cases the outcome is the median between the two peaks and x.

Case 2: We have $x = a$. Pick profile P where $\tau(P_1) < \tau(P_2)$. Observe that $f(P_1', P_2) = x$; otherwise voter 2 manipulates at P' via P_2. We claim that $f(P) = \tau(P_1)$. Suppose that this is not the case. Efficiency implies that $f(P)$ must lie to the right of $\tau(P_1)$. Applying Claim **??**, we can assume without loss of generality that all alternatives to the left of $\tau(P_1)$ are preferred to all alternatives to its right according to P_1. But then voter 1 will manipulate at P via P_1'. Therefore $f(P) = \tau(P_1)$.

Applying anonymity, it follows that when voters have peaks to the right of x, the outcome is the left-most of the two peaks. Once again, the outcome is the median of the two peaks and x.

Case 3: We have $x = b$. Using the symmetric analogue of the arguments used in Case 2, we can conclude that when both voters have peaks to the left of x, the outcome is the right-most of the two peaks, i.e. the median of the two peaks and x.

Cases 1, 2 and 3 exhaust all possibilities. Hence statement 2 holds.

We now show that $2 \Rightarrow 1$.

Note that the scf is efficient (because $f(P) \in [\tau(P_1, \tau(P_2)]$ for all profiles P) and anonymous. We only show that that the scf is strategy-proof. Let a and b be the left-most and right-most alternatives in A. Suppose $B = \{x\}$ and $x \neq a, b$. Consider a profile where the voters have peaks on either side of x. The outcome is then x. In this situation, a voter can only change the outcome to y where x lies in the interval between y and her peak. This is clearly worse than x because of single-peakedness. Similar arguments apply when both peaks are on the "same side" as x and when x is one of the "extreme alternatives" a and b. ■

Observation **6** The set B is known as the set of "phantom voters", fictitious voters whose peaks are fixed and independent of the profile. Note that the median voter rule with an arbitrary number of phantom voters, is strategy-proof. However adding more than $N-1$ phantoms makes the rule inefficient. There are other ways to characterize strategy-proof scfs in single-peaked domains and to extend the notion of single-peakedness to more than one dimension (see Barberà, Gul and Stachetti (1994)).

9.6 Restricted Domains: Random Social Choice Functions

Randomization has been used as a method of resolving conflicts of interest since antiquity. From the point of view of mechanism design theory, allowing for randomization expands the set of incentive-compatible social choice functions because domain restrictions are inherent in the preference ranking of lotteries that satisfy the expected utility hypothesis.

Let $\mathcal{L}(A)$ denote the set of lotteries over the elements of the set A. If $\lambda \in \mathcal{L}(A)$, then λ_a will denote the probability that λ puts on $a \in A$. Clearly $\lambda_a \geq 0$ and $\sum_{a \in A} \lambda_a = 1$.

Definition **42** *Let $\mathbb{D} \subset \mathbb{P}$. A Random Social Choice Function (RSCF) is a map $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$.*

In models where the outcome of voting is a probability distribution over outcomes, there are several ways to define strategy-proofness. Here we follow the approach of Gibbard (1977).

Definition **43** *A utility function $u : A \to \Re$ represents the ordering P_i over A if for all $a, b \in A$,*

$$[aP_ib] \Leftrightarrow [u(a) > u(b)]$$

Definition **44** *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ is strategy-proof if, for all $i \in I$, for all $P \in \mathbb{D}^N$, for all $\bar{P}_i \in \mathbb{D}$ and all utility functions u representing P_i, we have*

$$\sum_{a \in A} u(a)\varphi_a(P_i, P_{-i}) \geq \sum_{a \in A} u(a)\varphi_a(\bar{P}_i, P_{-i}).$$

A RSCF is strategy-proof if at every profile no voter can obtain a higher expected utility by deviating from her true preference ordering than she would if she announced her true preference ordering. Here, expected utility is computed with respect an arbitrary utility representation of her true preferences. It is well-known that this is equivalent to requiring that the probability distribution from truth-telling stochastically dominates the probability distribution from misrepresentation in terms of a voter's true preferences. This is stated formally below.

For any $i \in I$, $P_i \in \mathbb{D}$ and $a \in A$, we let $B(a, P_i) = \{b \in A : bP_ia\} \cup \{a\}$, i.e. $B(a, P_i)$ denotes the set of alternatives that are weakly preferred to a according to the ordering P_i.

Definition **45** *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ is strategy-proof if for all $i \in I$, for all $P \in \mathbb{D}^N$, for all $\bar{P}_i \in \mathbb{D}$ and all $a \in A$, we have*

$$\textstyle\sum_{b \in B(a,P_i)} \varphi_b(P_i, P_{-i}) \geq \sum_{b \in B(a,P_i)} \varphi_b(\bar{P}_i, P_{-i}).$$

We also introduce the a notion of unanimity for RSCFs. This requires an alternative which is first-ranked by all voters in any profile to be selected with probability one in that profile.

Definition **46** *A RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ satisfies unanimity if for all $P \in \mathbb{D}^N$ and $a \in A$,*

$$[a = \tau(P_i, A) \text{ for all } i \in I] \Rightarrow [\varphi_a(P) = 1].$$

Definition **47** *The RSCF $\varphi : \mathbb{D}^N \to \mathcal{L}(A)$ is a random dictatorship if there exist non-negative real numbers β_i, $i \in I$ with $\sum_{i \in I} \beta_i = 1$ such that for all $P \in \mathbb{D}$ and $a \in A$,*

$$\varphi_a(P) = \textstyle\sum_{\{i : \tau(P_i) = a\}} \beta_i$$

In a random dictatorship, each voter i gets weight β_i where the sum of these β_i's is one. At any profile, the probability assigned to an alternative a is simply the sum of the weights of the voters whose maximal element is a. A random dictatorship is clearly strategy-proof for any domain; by manipulation, a voter can only transfer weight from her most-preferred to a less-preferred alternative. A fundamental result in Gibbard (1977) states that the converse is also true for the complete domain $\mathbb{P}$. [3]

Theorem **9** *Assume $|A| \geq 3$. A RSCF $\varphi : \mathbb{P}^N \to \mathcal{L}(A)$ is strategy-proof and satisfies unanimity if and only if it is a random dictatorship.*

[3] Gibbard's result is actually more general than Theorem **??** below because it does not assume unanimity. However since unanimity will be a maintained hypothesis throughout the paper, we state only the version of the result with unanimity.

Proof: The proof of sufficiency is straightforward. We prove necessity in the case of $I = \{1,2\}$. The arguments are an extension of those used to prove the Gibbard-Satterthwaite Theorem in the $N = 2$ case.

In what follows assume $\varphi : \mathbb{P}^2 \to \mathcal{L}(A)$ satisfies unanimity and strategy-proofness.

CLAIM **8** *Let* $P = (P_1, P_2)$ *be such that* $\tau(P_1) \neq \tau(P_2)$. *Then* $[\varphi_a(P_1, P_2) > 0] \Rightarrow [a \in \{\tau(P_1), \tau(P_2)\}]$.

Proof: Suppose not i.e. suppose that there exists P_1, P_2 and $a, b \in A$ such that $\tau(P_1) = a \neq b = \tau(P_2)$ and $\varphi_a(P_1, P_2) + \varphi_b(P_1, P_2) < 1$. Let $\alpha = \varphi_a(P_1, P_2)$ and $\beta = \varphi_b(P_1, P_2)$. Let $P_1' = \begin{pmatrix} a \\ b \\ \vdots \end{pmatrix}$ and $P_2' = \begin{pmatrix} b \\ a \\ \vdots \end{pmatrix}$.

Strategy-proofness implies $\varphi_a(P_1', P_2) = \alpha$. Also $\varphi_a(P_1', P_2) + \varphi_b(P_1', P_2) = 1$; otherwise voter 1 will manipulate via P_2, thereby obtaining probability one on b by unanimity. Hence $\varphi_b(P_1', P_2) = 1 - \alpha$. Strategy-proofness also implies $\varphi_b(P_1', P_2') = \varphi_b(P_1', P_2) = 1 - \alpha$ and $\varphi_a(P_1', P_2') = \alpha$.

By a symmetric argument, $\varphi_b(P_1', P_2') = \varphi_b(P_1, P_2') = \beta$ and $\varphi_a(P_1', P_2') = 1 - \beta$. Comparing the probabilities on a and b given by φ at the profile (P_1', P_2'), we conclude that $\alpha + \beta = 1$ contradicting our earlier conclusion. ■

CLAIM **9** *Let* $P, \bar{P} \in \mathbb{P}^2$ *be such that* $\tau(P_1) = a \neq b = \tau(P_2)$ *and* $\tau(\bar{P}_1) = c \neq d = \tau(\bar{P}_2)$. *Then* $[\varphi_a(P) = \varphi_c(\bar{P})]$ *and* $[\varphi_b(P) = \varphi_d(\bar{P})]$.

Proof: Let $P_1 = \begin{pmatrix} a \\ \vdots \end{pmatrix}$ and $P_2 = \begin{pmatrix} b \\ \vdots \end{pmatrix}$.

Let $\hat{P}$ be an arbitrary profile where $\tau(\hat{P}_1) = a$ and $\tau(\hat{P}_2) = b$. Strategy-proofness implies that $\varphi_a(\hat{P}_1, P_2) = \varphi_a(P_1, P_2)$. Claim **??** implies $\varphi_b(\hat{P}_1, P_2) = \varphi_b(P_1, P_2)$. Now changing voter 2's ordering from P_2 to $\hat{P}_2$ and applying the same arguments, it follows that $\varphi_a(\hat{P}_1, \hat{P}_2) = \varphi_a(P_1, P_2)$ and $\varphi_b(\hat{P}_1, \hat{P}_2) = \varphi_b(P_1, P_2)$.

The argument in the previous paragraph implies that we can assume without loss without loss of generality that c is the second ranked outcome at P_1, i.e. we can assume that $P_1 = \begin{pmatrix} a \\ c \\ \vdots \end{pmatrix}$. Let $\bar{P}_1 = \begin{pmatrix} c \\ a \\ \vdots \end{pmatrix}$. Strategy-proofness implies $\varphi_a(\bar{P}_1, P_2) + \varphi_c(\bar{P}_1, P_2) = \varphi_a(P_1, P_2) + \varphi_c(P_1, P_2) = 1$. By Claim **??**, $\varphi_c(P_1, P_2) = \varphi_a(\bar{P}_1, P_2) = 0$. Hence $\varphi_a(P_1, P_2) = \varphi_c(\bar{P}_1, P_2)$ while $\varphi_b(P_1, P_2) = \varphi_b(\bar{P}_1, P_2)$. Now switching voter 2's preferences from P_2 to

$\bar{P}_2$ and applying the same argument as above, we conclude $\varphi_c(\bar{P}_1, P_2) = \varphi_c(\bar{P}_1, \bar{P}_2)$ while $\varphi_b(\bar{P}_1, P_2) = \varphi_b(\bar{P}_1, \bar{P}_2)$. The claim follows immediately. ■

The Claims above establish that φ is a random dictatorship. ■

Observation **7** The general result can be established for general N as in the proof of the Gibbard-Satterthwaite Theorem.

9.7 Restricted Domains: Quasi-Linear Domains

These are models where monetary compensation is feasible. Moreover money enters the utility function in an additively separable way.

Once again assume that A is the set of alternatives. Agent i's type is $\theta_i \in \Theta$ determines her valuation for every $a \in A$ according to the utility function $u_i : \Theta \times A \to \Re$, i.e. $u_i(a, \theta_i)$ is the valuation of alternative a when her type is θ_i. The agent may also receives a monetary payment $x_i \in \Re$. The overall utility of the agent is given by $v_i(a, x_i; \theta_i) = u_i(a, \theta_i) + x_i$. We re-define the earlier notions in this environment.

Definition **48** *A scf is a mapping* $f : \Theta^N \to A$.

Definition **49** *A transfer scheme is a collection of mappings* $x \equiv (x_1, \ldots, x_N)$ *where* $x_i : \Theta^N \to \Re$ *for all* $i \in I$.

Definition **50** *A pair* (f, x) *where* f *is a scf and* x *is a transfer scheme, is strategy-proof if*

$$u(f(\theta_i, \theta_{-i}), \theta_i) + x_i(\theta_i, \theta_{-i}) \geq u(f(\theta_i^{'}, \theta_{-i}), \theta_i) + x_i(\theta_i^{'}, \theta_{-i}).$$

for all $\theta_i, \theta_i^{'} \in \Theta$, *for all* $\theta_{-i} \in \Theta^{N-1}$ *and for all* $i \in I$.

Let f be a scf. We say that f is *implementable* if there exists a transfer scheme x such that the pair (p, x) is strategy-proof. This notion of implementability should not be confused with the same term defined earlier in the context of complete information.

QUESTION: What are the scfs which are implementable?

Below we provide an example of an important implementable scf.

Example **5** The following is the *efficient* scf f^e. For all $\theta \in \Theta^N$

$$f^e(\theta) = \arg\max_{a \in A} \textstyle\sum_{i \in I} u_i(a, \theta_i).$$

We claim that f^e is implementable. Let $x_i(\theta) = \sum_{j \neq i} u_j(f^e(\theta), \theta_j) + h_i(\theta_{-i})$ where h_i is an arbitrary function $h_i : \Theta^{N-1} \to \Re$. We show that (f^e, x) is strategy-proof.

Observe that

$$\begin{aligned} u_i(f^e(\theta_i, \theta_{-i}), \theta_i) + x_i(\theta_i, \theta_{-i}) &= u_i(f^e(\theta_i, \theta_{-i}), \theta_i) + \sum_{j \neq i} u_j(f^e(\theta_i, \theta_{-i}), \theta_j) + h_i(\theta_{-i}) \\ &= \sum_{i \in I} u_i(f^e(\theta_i, \theta_{-i}), \theta_i) + h_i(\theta_{-i}) \\ &\geq \sum_{i \in I} u_i(f^e(\theta_i^{'}, \theta_{-i}), \theta_i) + h_i(\theta_{-i}) \\ &= u_i(f^e(\theta_i^{'}, \theta_{-i}), \theta_i) + \sum_{j \neq i} u_j(f^e(\theta_i^{'}, \theta_{-i}), \theta_j) + h_i(\theta_{-i}) \\ &= u_i(f^e(\theta_i^{'}, \theta_{-i}), \theta_i) + x_i(\theta_i^{'}, \theta_{-i}) \end{aligned}$$

Therefore (f^e, x) is strategy-proof. The transfer scheme is known as the Vickrey-Clarke-Groves (VCG) scheme. If Θ is "rich enough", this scheme is the unique scheme with the property that (f^e, x) is strategy-proof. This is a special case of a class of results called *Revenue Equivalence Theorems*.

Very general characterizations of implementability in general domains exist in terms of "monotonicity properties". Below are explicit characterizations in special domains.

9.7.1 The Complete Domain

The domain Θ is *unrestricted* if, for all $\alpha \in \Re$, $a \in A$, and $i \in I$, there exists $\theta_i \in \Theta$ such that $u_i(a, \theta_i) = \alpha$.

THEOREM **10 (Roberts (1979))** *Assume $|A| \geq 3$. Let Θ be an unrestricted domain. The scf $f : \Theta^N \to A$ is implementable if and only if there exist non-negative real numbers $k_1, \ldots, k_N$ and real numbers $\gamma(a)$ for all $a \in A$ such that for all $\theta \in \Theta$,*

$$f(\theta) = \arg\max_{a \in A} \textstyle\sum_{i \in I} \{k_i u_i(a, \theta_i) + \gamma(a)\}$$

Moreover the associated transfers are of the form $x_i(\theta) = \frac{1}{k_i} \sum_{j \neq i} \{k_j u_j(a, \theta_j) + \gamma(a)\} + h_i(\theta_{-i})$.

9.7.2 An Auction Model

In this model there is a single object with one seller and n buyers or bidders. The set of alternatives $A = \{e_0, ..., e_n\}$ where e_i is the allocation where the object is given to bidder i, $i = 1, \ldots, n$ and e_0 is the allocation where the object is unsold and remains with the seller. We let x_i denote the payment by bidder i.

Bidder i's valuation for the object (her type) is θ_i which is a non-negative real number. We assume for convenience that $\theta_i \in [0,1]$. The payoff of bidder i, of type θ_i for allocation e_i and payment x_i is

$$v_i(e_j, x_i, \theta_i) = \begin{cases} \theta_i - x_i & \text{if } e_j = e_i \\ -x_i & \text{o.w.} \end{cases}$$

An auction is a pair (p, x) where $p : [0,1]^n \to \Delta^n$ is a probability distribution over $\{e_1, ...e_n\}$ and $x : [0,1]^n \to \Re^n$ is the vector of payments by bidders. If $\theta \equiv (\theta_1, ..., \theta_n)$ is the vector of announced valuations, then $p_i(\theta)$, $i = \{1, .., n\}$ is the probability that agent i gets the object. Clearly $p_i(\theta) \geq 0$ and $\sum_i p(\theta) = 1$. Furthermore, $x_i(\theta)$ is the payment made by i. The p component of an auction will be called an allocation rule and the x component, a transfer scheme/rule.

Fix an auction (p, x). The utility of bidder i (whose true valuation is θ_i, who bids θ_i' given that others bids θ_{-i}) is given by

$$v_i((\theta_i', \theta_{-i}), \theta_i) = p_i(\theta_i', \theta_{-i})\theta_i - x_i(\theta_i', \theta_{-i}).$$

In accordance with our earlier definition, an auction (p, x) is strategy-proof if $v_i((\theta), \theta_i) \geq v_i((\theta_i'), \theta_{-i}), \theta_i)$ holds for all θ_i, θ_i', θ_{-i} and i.

The following result characterizes strategy-proof auctions

THEOREM **11 (Myerson (1981))** *1. If $(p, x)$is strategy-proof, then $p_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_i for all θ_{-i}.*

2. Suppose $p_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_i for all θ_{-i}. Then there exists a transfer rule x such that (p, x) is strategy-proof. Moreover x must be as follows:

$$x_i(\theta_i, \theta_{-i}) = p_i(\theta_i, \theta_{-i})\theta_i - \int_0^{\theta_i} p_i(s_i, \theta_{-i}) ds_i + h_i(\theta_{-i})$$

where h_i is an arbitrary function of θ_{-i}.

Proof: We first establish statement 1. Fix an arbitrary θ_i. Since (p, x) is strategy-proof, the following inequalities must hold

1. $p_i(\theta_i, \theta_{-i})\theta_i - x_i(\theta_i, \theta_{-i}) \geq p_i(\theta_i', \theta_{-i})\theta_i - x_i(\theta_i', \theta_{-i})$

2. $p_i(\theta_i', \theta_{-i})\theta_i' - x_i(\theta_i', t_{-i}) \geq p_i(\theta_i, \theta_{-i})\theta_i' - x_i(\theta_i, \theta_{-i})$

Adding the two inequalities we obtain:

$$[p_i(\theta_i, \theta_{-i}) - p_i(\theta_i', \theta_{-i})][\theta_i - \theta_i'] \geq 0$$

This implies that if $\theta_i > \theta_i'$ then $p_i(\theta_i, \theta_{-i}) \geq p_i(\theta_i', \theta_{-i})$ i.e. $p_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_i.

We now establish the second part of the Theorem. Let $p_i(\theta_i, \theta_{-i})$ be weakly increasing in θ_i. Let x be the following transfer function:

$$x_i(\theta_i, \theta_{-i}) = p_i(\theta_i, \theta_{-i})\theta_i - \int_0^{\theta_i} p_i(s_i, \theta_{-i})ds_i + h_i(\theta_{-i}).$$

We claim that (p, x) is strategy-proof. Note that

$$\begin{aligned} v_i((\theta_i, \theta_{-i}), \theta_i) &= p_i(\theta_i, \theta_{-i})\theta_i - x_i(\theta_i, \theta_{-i}) \\ &= \int_0^{\theta_i} p_i(s_i, \theta_{-i})ds_i + h_i(\theta_{-i}) \end{aligned}$$

Also,

$$\begin{aligned} v_i((\theta_i', \theta_{-i}), \theta_i) &= p_i(\theta_i', \theta_{-i})\theta_i - x_i(\theta_i', \theta_{-i}) \\ &= p_i(\theta_i', \theta_{-i})\theta_i - p_i(\theta_i', \theta_{-i})\theta_i' + \int_0^{\theta_i'} p_i(s_i, \theta_{-i})ds_i + h_i(\theta_{-i}) \\ &= p_i(\theta_i', \theta_{-i})(\theta_i - \theta_i') + \int_0^{\theta_i'} p_i(s_i, \theta_{-i})ds_i + h_i(\theta_{-i}) \end{aligned}$$

Let $\Delta = v_i((\theta_i, \theta_{-i}), \theta_i) - v_i((\theta_i', \theta_{-i}), \theta_i)$. There are two cases to consider.

Case 1: $\theta_i > \theta_i'$. Then,

$$\Delta = \int_{\theta_i'}^{\theta_i} p_i(s_i, \theta_{-i}) ds_i - p_i(\theta_i', \theta_{-i})(\theta_i - \theta_i')$$
$$\geq 0$$

where the inequality follows from the fact that $p_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_i.

Case 2: $\theta_i' > \theta_i$. Then

$$\Delta = p_i(\theta_i', \theta_{-i})(\theta_i' - \theta_i) - \int_{\theta_i}^{\theta_i'} p_i(s_i, \theta_{-i}) ds_i$$
$$\geq 0$$

where the last inequality again follows from the fact that $p_i(\theta_i, \theta_{-i})$ is weakly increasing in θ_i.

Hence (p, x) is strategy-proof.

Finally, we show that if (p, x) is strategy-proof and p is increasing, then x must be of the form described in Part 2 of the statement of the Theorem.

$$\begin{aligned} v_i((\theta_i, \theta_{-i}), \theta_i) &= p_i(\theta_i, \theta_{-i})\theta_i - x_i(\theta_i, \theta_{-i}) \\ &\geq p_i(\theta_i', \theta_{-i})\theta_i - x_i(\theta_i', \theta_{-i}) \\ &= p_i(\theta_i', \theta_{-i})\theta_i' - x_i(\theta_i', \theta_{-i}) + (\theta_i - \theta_i')p_i(\theta_i', \theta_{-i}) \\ &= v_i((\theta_i', \theta_{-i}), \theta_i') + (\theta_i - \theta_i')p_i(\theta_i', \theta_{-i}) \end{aligned}$$

Let $\overline{v}_i(\theta_i) = v_i((\theta_i, \theta_{-i}), \theta_i)$ and $\overline{v}_i(\theta_i') = v_i((\theta_i', \theta_{-i}), \theta_i')$. We suppress the dependence of $\overline{v_i}$ on θ_{-i} for notational convenience. In other words, $\overline{v}_i(\theta_i)$ is bidder i's truth-telling utility when she is of type θ_i (when others announce θ_{-i}.) The last inequality reduces to:

$$\overline{v}_i(\theta_i) - \overline{v}_i(\theta_i') \geq (\theta_i - \theta_i')p_i(\theta_i', \theta_{-i})$$

In the case where $\theta_i > \theta_i'$, we have

$$\frac{\overline{v}_i(\theta_i) - \overline{v}_i(\theta_i')}{\theta_i - \theta_i'} \geq p_i(\theta_i', \theta_{-i})$$

By considering the symmetric counterpart of the case considered above where θ_i' does not gain by bidding θ_i, we have:

$$\frac{\overline{v}_i(\theta_i)-\overline{v}_i(\theta_i')}{\theta_i-\theta_i'} \leq p_i(\theta_i, \theta_{-i})$$

Hence,

$$p_i(\theta_i, \theta_{-i}) \geq \frac{\overline{v}_i(\theta_i)-\overline{v}_i(\theta_i')}{\theta_i-\theta_i'} \geq p_i(\theta_i', \theta_{-i})$$

Since $p_i(\theta_i, \theta_{-i})$ is (weakly) increasing in θ_i (for any given θ_{-i}), it is continuous almost everywhere. Considering sequences $\theta_i \to \theta_i'$, we have $p_i(\theta_i, \theta_{-i}) \to p_i(\theta_i', \theta_{-i})$ almost everywhere. Observe that at points of continuity θ_i', $\lim_{\theta_i \to \theta_i'} \frac{\overline{v}_i(\theta_i')-\overline{v}_i(\theta_i)}{\theta_i'-\theta_i}$ exists and equals $p_i(\theta_i', \theta_{-i})$, i.e. $\frac{\partial \overline{v_i}(\theta_i)}{\partial \theta_i} = p_i(\theta_i, \theta_{-i})$ almost everywhere.

Since $p_i(\theta_i, \theta_{-i})$ is increasing in θ_i, it is Riemann integrable. Applying the Fundamental Theorem of Calculus, we have,

$$v_i((\theta_i, \theta_{-i}), \theta_i) = v_i((0, \theta_{-i}), 0) + \int_0^{\theta_i} p_i(s_i, \theta_{-i}) ds_i$$

Thus $x_i(\theta_i, \theta_{-i}) = p_i(\theta_i, \theta_{-i})\theta_i - \int_0^{\theta_i} p_i(s_i, \theta_{-i}) ds_i + h_i(\theta_{-i})$ where $h_i(\theta_{-i}) \equiv v_i((0, \theta_{-i}), 0)$.
■

OBSERVATION **8** Myerson actually proved the result for BIC auctions with the assumption that valuations are distributed independently. However the result as well as its proof can be straightforwardly adapted from the statement and proof of Theorem 2 (More accurately, Theorem 2 is an adaptation of Myerson's result). The main change required is that the θ_{-i}'s have to be "integrated out". For instance, instead of requiring $p_i(\theta_i, \theta_{-i})$ to be increasing in θ_i for all θ_{-i}, BIC requires $\bar{p}_i(\theta_i)$ to be increasing where $\bar{p}_i(\theta_i) = \int_{\theta_{-i}} p_i(\theta_i, \theta_{-i}) dF_{-i}(\theta_{-i})$ and F_{-i} is the joint distribution function for θ_{-i}.

Mechanism design in the environment considered above has an important *decomposition* property. In order to check whether an auction is strategy-proof, it suffices to check whether the allocation rule satisfies a increasingness or monotonicity property. The transfer rule is then *uniquely* determined by the allocation rule for every type of bidder i (given θ_{-i}) upto a constant which depends only θ_{-i}. This constant can be interpreted as the truth-telling utility obtained by bidder i whose valuation is 0, again given that the other bidders have valuation θ_{-i}. This unique determination of the transfer rule once the allocation rule is fixed, is known widely as *The Revenue Equivalence Principle.* This principle and the decomposition property holds quite generally in environments where agents have quasi-linear utility, i.e. models where agents have utility functions of the form $v_i(d, x_i, \theta_i) = u_i(d, \theta_i) + x_i$ where $d \in D$ is a decision chosen from some arbitrary set D and x_i is the transfer received by agent i.

Observation **9** A special case of the auction (p^*, x^*) above is the one where $h_i(\theta_{-i}) = 0$ for all θ_{-i} and i. This auction is *individually rational*, i.e. bidders who do not get the object have a payoff of 0. This is the class of *Vickrey* auctions.