Chapter 5

Mechanism Design

In the preceding sections, we assumed that the players participate in the game in a very uncoordinated way. In this section, we will investigate how the conflicted interests of agents can be aggregated into a collective decision.

Example 5.1 (Bach, Stravinsky, or Mahler). Assume there are three people that want to visit a concert together. Player 1 prefers Bach over Mahler over Stravinsky; player 2 prefers Bach over Stravinsky over Mahler, and player 3 prefers Mahler over Stravinsky over Bach. Previously, we assumed that the players engage in an uncoordinated coordination game where each player chooses a concert independently of the other players and the players receive zero utility in case they did not meet.

In this section, we want to design mechanisms that govern the coordination of the players. We assume that there is a set of alternatives A the players want to coordinate on and each player ha preferences on the

5.1 Social Choice Functions

Let $N = \{1, \dots, n\}$ be a set of agents and $A = \{1, \dots, m\}$ be a set of alternatives. We assume that each player i has a strict preference order $\prec_i$ over A.

> **Definition 5.2** (Linear Order). A linear order on a set A is a binary relation $L \subseteq A \times A$ which is transitive, antisymmetric and total. For $a, b \in A$, we write $\preceq$ when $(a, b) \in L$. Then,

1	2	3
Bach	Bach	Mahler
$\succ$	$\succ$	$\succ$
Mahler	Stravinsky	Stravinsky
$\succ$	$\succ$	$\succ$
Stravinsky	Mahler	Bach

Figure 5.1: Preference profiles of three players.

$\{1,2,3,4\}$	$\{5,6,7\}$	$\{8,9\}$
a	c	b
$\curlyvee$	$\curlyvee$	$\curlyvee$
b	b	c
$\curlyvee$	$\curlyvee$	$\curlyvee$
c	a	a

Figure 5.2: Preference profiles of nine players.

Anti-symmetry If $a \preceq b$ and $b \preceq a$, then $a = b$.
Transitivity If $a \preceq b$ and $b \preceq c$ then $a \preceq c$.
Totality $a \preceq b$ or $b \preceq a$.

For a linear order, we sometimes write $a \prec b$ when $a \preceq b$ and $a \neq b$. For a set A, let $L(A)$ denote the set of strict linear orders.

Definition 5.3 (Social Choice Function). A social choice function is a function $f : L(A)^n \to A$.

Prominent social choice functions are, e.g., plurality, which selects an alternative ranked first by the largest number of players, and single transferable vote. In the latter, alternatives that are ranked first by the fewest number of players are eliminated successively until only one alternative remains.

Example 5.4. There are 9 players and 3 alternatives a, b and c. Four players have the preference $a \succ b \succ c$, three players have the preference $c \succ b \succ a$, and two players have the preferences $b \succ c \succ a$, see Figure 5.2.
The plurality rule would select alternative a because it is ranked first by four voters while the other two alternatives are ranked first by three and two players only.
The single transferable vote rule first eliminates alternative b because it is ranked first by two players only. Restricing the attention to the remaining alternatives, a is ranked first by four players while c is rankes first by 5 players. Thus, alternative a is eliminated next leaving c as the selected alternative.

Recall that we concluded in Example 5.4 that the plurality rule selects alternative a since it is ranked first by more players than any other alternative. However, players 8 and 9 have an incentive to misreport their preferences and claim that they prefer c over b. In fact, if we assume that ties are broken in favor of alternative c, only a single player, say player 8, needs to change their reported preference relation from $b \succ c \succ a$ to $c \succ b \succ a$ in order to change the outcome in a way that is prefered by them. Along the same lines, the single transferable vote rule is

prone to manipulation by players from the first group. When they pretend that their most prefered alternative is b, this alternative would be selected.
In the following, we are interested in social choice functions that are manipulable like this.

Definition 5.5 (Strategyproofness). A social choice function f is called manipulable if there are $i \in N$, $\succ \in L(A)^n$ and $\succ_i' \in L(A)$ such that $f(\succ_i', \succ_{-i}) \succ_i f(\succ)$ where $(\succ_i', \succ_{-i}) = (\succ_1, \ldots, \succ_{i-1}, \succ_i', \succ_{i+1}, \ldots, \succ_n)$ is the preference profile obtained by replacing the preference of player i with $\succ_i'$. The function is called strategyproof if it is not manipulable.

There are two natural ways to achieve strategyproofness. First, we may ignore all but two alternatives and use a simple majority rule to choose between the two. Second, we may choose the alternative considering only the preference profile of a single player. The first rule corresponds to non-surjective social choice functions since some of the alternatives are never chosen. The second kind of social choice functions is called a dictatorship.

Definition 5.6 (Dictatorship). A social choice function f is called dictatorial if there is $i \in N$ such that for all $\succ \in L(A)^n$ and $a \in A \setminus \{f(\succ)\}$ we have $f(\succ) \succ_i a$.

The following seminal theorem shows that the surprising result that only surjective social choice functions are dictatorships.

Theorem 5.7 (Gibbard, 1973; Satterthwaite, 1975). Let $f : L(A)^n \to A$ be a social choice function where $|A| \geq 3$. If f is surjective and strategyproof, then it is dictatorial.

To prove this theorem, we first need two lemmas. The first lemma establishes a certain monotonicity for strategyproof social choice functions.

Lemma 5.8. Let f be a strategyproof social choice function, $\succ, \succ' \in L(A)^n$ be such that $f(\succ) = a$ and $a \succ_i' b$ for all $b \in A \setminus \{a\}$ with $a \succ_i b$. Then, $f(\succ') = a$.

Proof. We start from the original profile $\succ$ and change the preferences of a single voter at a time until we reached $\succ'$ showing that the chosen alternative remains the same in every step. To this end, let $b = f(\succ_1', \succ_{-1})$. By strategyproofness $a \succeq_1 b$ and thus $a \succeq' b$ by our assumption. Also by strategyproofness, we have $b \succeq' a$ and, in conclusion, $a = b$. Repeating this argument for each player, the claim follows. □

The next lemma states that the alternative selected by a surjective and strategyproof social choice function must be Pareto optimal in the following sense.

Lemma 5.9. Let f be a surjective and strategyproof social choice function and let $a, b \in A$ and $\succ \in L(A)^n$ be such that $a \succ_i b$ for all $i \in N$. Then $f(\succ) \neq b$.

Proof. Assume for a contradiction that $f(\succ) = b$. By surjectivity, there is $\succ' \in L(A)^n$ such that $f(\succ') = a$. Let $\succ'' \in L(A)^n$ be a preference profile such that

$$a \succ''_i b \succ''_i x \qquad \text{for all } x \in A \setminus \{a, b\} \text{ and } i \in N.$$

We observe that $b \succ''_i x$ for all $i \in N$ and $x \in A \setminus$ with $b \succ''_i x$. By Lemma 5.8 $f(\succ'') = f(\succ) = b$. Similarly, $a \succ''_i x$ for all $i \in N$ and $x \in A \setminus \{a\}$ with $a \succ_i x$ and hence $f(\succ'') = f(\succ') = a$, a contradiction. □

We are now in position to prove the Gibbard-Satterthwaite-Theorem.

Proof of Theorem 5.7. We first show the result for $n = 2$ and perform an induction on n afterwards. Consider two alternatives $a, b \in A$ with $a \neq b$ and a preference profile $\succ \in L(A)^2$ such that

$$a \succ_1 b \succ_1 x \qquad \text{and} \qquad b \succ_2 a \succ_2 x$$

for all $x \in A \setminus \{a, b\}$. Then, by Lemma 5.9, $f(\succ) \in \{a, b\}$. Suppose that $f(\succ) = a$, and let $\succ' \in L(A)^2$ be such that

$$a \succ'_1 b \succ'_1 x \qquad \text{and} \qquad b \succ'_2 x \succ'_2 a$$

for all $x \in A \setminus \{a, b\}$. Then $f(\succ') = a$ since $f(\succ') \in \{a, b\}$ by Lemma 5.9 and $f(\succ') \neq b$ by strategyproofness. Lemma 5.8 then implies that f selects alternative a for any preference profile in which player 1 ranks it first.

By repeating the above argumentation for every pair of distinct alternatives in A, we obtain two sets $A_1, A_2 \in 2^A$ such that A_i is the set of alternatives that are selected for every preference profile in which player i ranks them first. Let $A_3 = A \setminus (A_1 \cup A_2)$. Note that $|A_3| \leq 1$ since otherwise, we would have applied the above argumentation for tow distinct elements in A_3 placing one of them in one of the sets A_1 or A_2.

As $|A| \geq 3$, we derive that $|A_1 \cup A_2| \geq 2$. Moreover, for $x, y \in A$ with $x \neq y$, it is impossible that $x \in A_1$ and $y \in A_2$ because this would give a contradiction when player 1 ranks x first and player 2 ranks y first. Since $a \in A_1$ this implies $A_2 = \emptyset$. Finally, we also derive that $A_3 = \emptyset$ since otherwise we obtain for $c \in A_3$ and $\succ'' \in L(A)^2$ with

$$c \succ''_1 a \succ''_1 x \qquad \text{and} \qquad a \succ''_2 c \succ''_2 x$$

for all $x \in A \setminus \{a, c\}$ and conclude that $c \in A_1$, or $a \in A_2$, a contradiction. We have proven that $A_1 = A$, so player 1 is a dictator.

To finish the proof of the theorem, assume that the statement holds for n players and consider a surjective and strategyproof social choice function $f : L(A)^{n+1} \to A$. We define a social choice function $g : L(A)^2 \to A$ by

$$g(\succ_1, \succ_2) = f(\succ_1, \succ_2, \dots, \succ_2)$$

for all $\succ_1, \succ_2 \in L(A)$. Using that f is surjective and strategyproof and Lemma 5.9, we derive that g is surjective as well. Assume for a contradiction that g is not strategyproof. Since f is strategyproof, there is no successful manipulation of player 1, so there must exist $\succ_1, \succ_2, \succ_2' \in L(A)$ and $a, b \in A$ such that

$$g(\succ_1, \succ_2) = a, \qquad g(\succ_1, \succ_2') = b \qquad \text{and} \qquad b \succ_2 a.$$

For $k \in \{0, \dots, n\}$, let $\boldsymbol{\succ}^k = (\succ_1, \succ_2', \dots, \succ_2', \succ_2, \dots, \succ_2) \in L(A)^{n+1}$ be the preference profile where k players have preference order $\succ_2'$ and $n-k$ players have preference order $\succ_2$, and let $a^k = f(\boldsymbol{\succ}^k)$. Since $a^n = b \succ_2 a = a^0$, there is $k \in \{0, \dots, n-1\}$ with $a^{k+1} \succ_2 a^k$ which contradicts the strategyproofness of f. We conclude that g is strategyproof and, thus, dictatorial.

If the dictator of g is player 1, then by Lemma 5.8, player 1 is also the dictator for f. If on the other hand, player 2 is the dictator for g, consider $h : L(A)^n \to A$ defined by

$$h(\succ_2, \dots, \succ_{n+1}) = f(\succ_1^*, \succ_2, \dots, \succ_{n+1})$$

for an arbitrary $\succ_1^* \in L(A)$. Then, h is strategyproof by the strategyproofness of f, and surjective because voter 2 is a dictator for g. Therefore, by the induction hypothesis, h is dictatorial. Without loss of generality, let player 2 be the dictator for h, and let $e : L(A)^2 \to A$ be given by

$$e(\succ_1, \succ_2) = f(\succ_1, \succ_2, \succ_2^*, \dots, \succ_{n+1}^*)$$

for arbitrary $\succ_3^*, \dots, \succ_{n+1}^* \in L(A)$. Then, e is strategyproof and surjective, and hence dictatorial. The dictator for e must be player 2, because player 1 is not a dictator for g a. Using that $\succ_i^*$ for $i = \{1, 3, 4, \dots, n+1\}$ were chosen arbitrarily, we conclude that player 2 is a dictator for f. □

5.2 General Mechanism Design Problems

In the last section, the preferences of a player were defined by a linear order $\succ \in L(A)$ over a set of alternatives. In this section, we generalize this notion to arbitrary preferences.

To this end, let N be a finite set of players and A be a set of alternatives. Each player has a set Θ_i of possible types and a utility function $u_i : A \times \Theta_i \to \mathbb{R}$. We set $\boldsymbol{\Theta} = (\Theta_i)_{i \in N}$.

In this more general setting, a social choice function is a function $f : \boldsymbol{\Theta} \to A$, i.e., a function that maps type vectors to alternatives.

Definition 5.10 (Mechanism). A mechanism is a tuple $(\mathbf{\Sigma}, g)$, where
1. $\mathbf{\Sigma} = (\Sigma_i)_{i \in N}$ and Σ_i is the message space for player i, and
2. $g : \mathbf{\Sigma} \to A$ is an outcome function.

A mechanism is said to implement a social choice function f is dominant strategies if the for all type vectors θ the outcome $f(\theta)$ appears as the dominant strategy of the mechanism.

Definition 5.11 (Implementable Social Choice Function). A mechanism $(\mathbf{\Sigma}, g)$ is said to implement the social choice function f in weakly dominant strategies if there exist functions $s_i : \Theta_i \to \Sigma_i$ for all players i such that for every $\theta \in \mathbf{\Theta}$ we have $g(s_1(\theta_1), \ldots, s_n(\theta_n)) = f(\theta)$, and in addition

$$u_i\big(g(s_i(\theta_i), \sigma_{-i}), \theta_i\big) \geq u_i\big(g(\sigma), \theta_i\big)$$

for all $i \in N$, $\theta_i \in \Theta_i$ and $\sigma \in \mathbf{\Sigma}$.

An interesting special class of mechanisms are direct mechanisms. Here, the the players directly report their type to the mechanism, i.e., $\mathbf{\Sigma} = \mathbf{\Theta}$. We will associate direct mechanisms with their outcome function $f : \mathbf{\Theta} \to A$.

Definition 5.12. A mechanism is called direct if $\mathbf{\Sigma} = \mathbf{\Theta}$.

A direct mechanism $(\mathbf{\Theta}, g)$ is called dominant strategy incentive compatible, or strategyproof if truth telling is a dominant strategy of the players.

Definition 5.13 (Strategyproof Mechanism). A direct mechanism is called dominant strategy incentive compatible or strategypoof if

$$u_i(g(\theta), \theta_i) \geq u_i(g(\theta_i', \theta_{-i}), \theta_i)$$

for all $i \in N$, $\theta \in \mathbf{\Theta}_i$ and $\theta_i' \in \Theta_i$.

The profile θ of true types is the called the truthful equilibrium of a strategyproof direct mechanism.
One may wonder whether arbitrarily complicated mechanisms are needed to implement a certain social choice function. The following results shows thatone can restrict to strategyproof direct mechanisms.

Theorem 5.14 (Revelation Principle). A social choice function is implementable if and only if it is implemented in the truthful equilibrium of a strategyproof direct mechanism.

Proof. The proof uses the observation that the direct mechanism can simulate the equilibrium strategies of the players. Formally, let f be an implementable social choice function, i.e., there is a mechanism $(\boldsymbol{\Sigma}, g)$ and functions $s_i : \Sigma_i \to \Theta_i$ for all players i such that for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we have $g(s_1(\theta_1), \ldots, s_n(\theta_n)) = f(\boldsymbol{\theta})$ and for every $i \in N$, $\theta_i \in \Theta_i$ and $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}$, we have

$$u_i\big(g(s_i(\theta_i), \boldsymbol{\sigma}_{-i}), \theta_i\big) \geq u_i\big(g(\boldsymbol{\sigma}), \theta_i\big).$$

Let $h : \boldsymbol{\Theta} \to A$ be defined as

$$h(\boldsymbol{\theta}) = g(s_1(\theta_1), \ldots, s_n(\theta_n))$$

for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then, we have $h(\boldsymbol{\theta}) = f(\boldsymbol{\theta})$ for all $\theta \in \Theta$. Furthermore, we check that for all $i \in N$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\theta_i' \in \Theta_i$ we have that

$$\begin{aligned} u_i(h(\boldsymbol{\theta}), \theta_i) &= u_i(g(s_1(\theta_1), \ldots, s_n(\theta_n)), \theta_i) \\ &\geq u_i(g(s_1(\theta_1), \ldots, s_{i-1}(\theta_{i-1}), s_i(\theta_i'), s_{i+1}(\theta_{i+1}), \ldots, s_n(\theta_n)), \theta_i) \\ &= u_i(h(\theta_i', \boldsymbol{\theta}_{-i}), \theta_i). \end{aligned}$$

We conclude that $(\boldsymbol{\Theta}, h)$ is a strategyproof direct mechanism that implements f. □

Theorems 5.7 and Theorem 5.14 together imply that only dictatorial social choice functions are implementable when there are more than two alternatives and utility functions can be arbitrary. In the next section, we look at special cases where this general impossibility result can be circumvented.

5.3 Mechanisms with Payments

The Gibbard-Statterthwaite-Theorem assumes that the players have arbitrary preferences over the set of alternatives. One way to circumvent this negative result is to include monetary payments in the outcome of the mechanism. Formally, for a set of alternatives A a direct mechanism with payments is a mechanism (Θ, f') where $f' : \boldsymbol{\Theta} \to (A \times \mathbb{R}^n)$, where the second component of the mechanism is interpreted as a vector of payments. It is useful to separate this vector of payments from the actual choice function and to write $f' = f \times \mathbf{p}$ for some social choice function $f : \boldsymbol{\Theta} \to A$ and a payment function $\mathbf{p} : \boldsymbol{\Theta} \to \mathbb{R}$.

Definition 5.15 (Direct Mechanism with Payments). A direct mechanism with payments is a tuple $(f, \mathbf{p})$ where $f : \boldsymbol{\Theta} \to A$ is a social choice function and $\mathbf{p} : \boldsymbol{\Theta} \to \mathbb{R}^n$ is a payment function.

For mechanisms with payments, we assume that the players' utilities are quasi-linear.

Definition 5.16 (Quasi-Linear Utilities). The utility of player i is called quasi-linear if there is a valuation function $v_i : A \times \Theta_i \to \mathbb{R}$ such that $u_i\big((f(\theta), \mathbf{p}(\theta)), \theta_i\big) = v_i(f(\theta), \theta_i) - \mathbf{p}_i(\theta)$.

The main positive result of this section is that there is are strategyproof direct mechanisms that maximize the social welfare defined as $V(a) = \sum_{i \in N} v_i(a, \theta_i)$, i.e., the sum of the players' valuations for the alternative. These mechanisms are called Vickrey-Clark-Groves mechanisms, or VCG mechanisms, for short.

Definition 5.17 (VCG mechanisms). A mechanism $(f, \mathbf{p})$ is called Vickrey-Clark-Groves mechanism or VCG mechanism if

$$f(\theta) \in \arg\max_{a \in A} \sum_{i \in N} v_i(a, \theta_i) \qquad \text{and}$$
$$p_i(\theta) = h_i(\theta_{-i}) - \sum_{j \in N \setminus \{i\}} v_j(f(\theta), \theta_j) \qquad \text{for all } i \in N,$$

where $h_i : \mathbf{\Theta}_{-i} \to \mathbb{R}$ is some function that depends on the types of all players except i.

The crucial component is the term $\sum_{j \in N \setminus \{i\}} v_j(f(\theta), \theta_j)$ which is equal to the social welfare of all players but i. In that way, every player is interested in maximizing the social welfare minus the term $h_i(\theta_i)$ which does not depend on θ_i. Therefore, no player has an incentive to misreport their true type.

Theorem 5.18. VCG mechanisms are strategyproof.

Proof. Let $i \in N$, $\theta \in \mathbf{\Theta}$ and $\theta_i' \in \Theta_i$ be arbitrary. We calculate

$$\begin{aligned} u_i(\theta, \theta_i) &= v_i(f(\theta), \theta_i) - p_i(\theta) \\ &= \sum_{j \in N} v_j(f(\theta), \theta_j) - h_i(\theta_{-i}) \\ &\geq \sum_{j \in N} v_j(f(\theta_i', \theta_{-i}), \theta_j) - h_i(\theta_{-i}) \\ &= u_i((\theta_i', \theta_{-i}), \theta_{-i}), \end{aligned}$$

where the inequality uses that $f(\theta)$ maximizes social welfare. □

So far, we did no impose any restrictions on the payments $p_i(\theta)$. In many cases it is natural to assume that the players are rather charged than paid and that players do not pay more than what they gain from participating in the mechanism. These two properties are reflected in the following two definitions.

Definition 5.19 (No Positive Transfers). A mechanism with payments $(f, \mathbf{p})$ makes no positive transfers if $p_i(\theta) \geq 0$ for all $i \in N$ and $\theta \in \boldsymbol{\Theta}$.

Definition 5.20 (Ex-Post Individual Rationality). A mechanism with payments $(f, \mathbf{p})$ is individually ex-post individually rational if $v_i(f(\theta), \theta_i) - p_i(\theta) \geq 0$ for all $i \in N$ and $\theta \in \boldsymbol{\Theta}$.

It turns out that we can choose the functions $h_i : \boldsymbol{\Theta}_{-i} \to \mathbb{R}$ such that the resulting VCG mechanism makes no positive transfer and is ex-post individually rational. This is achieved by the so-called Clark pivot rule.

Definition 5.21 (Clark Pivot Rule). A VCG mechanism with Clark pivot rule sets $h_i(\theta_{-i}) = \max_{a \in A} \sum_{j \in N \setminus \{i\}} v_j(a, \theta_j)$ for all $i \in N$ and $\theta_{-i} \in \boldsymbol{\Theta}_{-i}$.

With the Clark pivot rule, the payment of player i becomes

$$p_i(\theta) = \max_{a \in A} \sum_{j \in N \setminus \{i\}} v_j(a, \theta_j) - \sum_{j \in N \setminus \{i\}} v_j(f(\theta), \theta_j).$$

Intuitively, this is the amount by which the welfare of the other players drops due to player i's participation in the mechanism. The payment with the Clark pivot rule makes player i internalize this externality.

Theorem 5.22. A VCG mechanism with Clark pivot rule makes no positive transfers. If $v_i(a, \theta_i) \geq 0$ for all $i \in N$, $\theta_i \in \Theta_i$ and $a \in A$, it is also ex-post individually rational.

Proof. Let $\theta \in \boldsymbol{\Theta}$ and $i \in N$ be arbitrary and let

$$a = f(\theta) \qquad \text{and} \qquad b \in \arg\max_{a' \in A} \sum_{j \in N \setminus \{i\}} v_j(a', \theta_j).$$

Then, we obtain

$$p_i(\theta) = \sum_{j \in N \setminus \{i\}} v_j(b, \theta_j) - \sum_{j \in N \setminus [i]} v_j(a, \theta_j) \geq 0,$$

so the mechanism makes no positive transfers. In addition,

$$\begin{aligned} u_i(\theta, \theta_i) &= v_i(a, \theta_i) + \sum_{j \in N \setminus [i]} v_j(a, \theta_j) - \sum_{j \in N \setminus \{i\}} v_j(b, \theta_j) \\ &\geq \sum_{j \in N} v_j(a, \theta_j) - \sum_{j \in N} v_j(b, \theta_j) \\ &\geq 0, \end{aligned}$$

where the first inequality uses that $v_i(b, \theta_i) \geq 0$ and the second inequality uses that a maximizes the social welfare. □

Example 5.23 (Auction of a Single Good). Suppose we want to auction off a single item. In this case A corresponds to the set of possible winners of the auction, i.e., $A = N$, and for each player i the set of types Θ_i corresponds to the set of possible (monetary) valuations for receiving the item, i.e., $\Theta_i = \mathbb{R}_{\geq 0}$. Then, the valuation function $v_i : A \to \mathbb{R}$ is of the form

$$v_i(a, \theta_i) = \begin{cases} \theta_i, & \text{if } a = i \\ 0, & \text{else,} \end{cases}$$

In an auction setting, the valuations reported to by the players to the mechanism are usually called bids.

Since only a single player may receive the item, we have $\max_{a \in A} \sum_{i \in N} v_i(a) = \max_{i \in N} v_i$ and thus the VCG mechanism assigns the good to the player i^* with the highest bid. Moreover,

$$\begin{aligned} p_{i^*}(\theta) &= \max_{a \in A} \sum_{j \in N \setminus \{i^*\}} \theta_j - \sum_{j \in N \setminus \{i^*\}} v_j(f(\theta), \theta_j) \\ &= \max_{j \in N \setminus \{i^*\}} \theta_j, \end{aligned}$$

i.e., the winner pays the second highest bid. It is easy to check that the payments of all other players are zero. Thus, we obtain the well-known second-price auction.

Example 5.24 (Multiunit Auction). Consider an auction to sell k identical items to n players where each player is interested in a single item only. We assume $k < n$. In this case $A = \{S : S \subset N, |S| = k\}$ and valuations are of the form

$$v_i(a, \theta_i) = \begin{cases} \theta_i, & \text{if } i \in a \\ 0, & \text{else,} \end{cases}$$

with $\Theta_i = \mathbb{R}_{\geq 0}$. As VCG mechanisms maximize the social welfare, the items are allocated to the k players with the largest bids. One can check that each player that receives an item pays a price equal to the $k + 1$ highest bid, and all other players pay nothing.

5.3.1 Characterizations of Strategyproof Mechanisms

To obtain a characterization of strategyproof mechanism, we first observe that a mechanism is strategyproof if and only if the follwing two conditions are met. First, the payment of each player i does not depend on the reported type of player i, but only on the selected alternative and the reported types of the other players. Second, the mechanism maximizes for each player.

Lemma 5.25. A mechanism $(f, \mathbf{p})$ is strategyproof if and only if for every $i \in N$ there is a price function $t_i : \boldsymbol{\Theta}_{-i} \times A \to \mathbb{R}$ such that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the following hold

1. $p_i(\boldsymbol{\theta}) = t_i(\boldsymbol{\theta}_{-i}, f(\boldsymbol{\theta}))$
2. $f(\boldsymbol{\theta}) \in \arg\max_{a \in A(\boldsymbol{\theta}_{-i})}\{v_i(a, \theta_i) - t_i(\boldsymbol{\theta}_{-i}, a)\}$,

where $A(\boldsymbol{\theta}_{-i}) = \{f(\theta_i, \boldsymbol{\theta}_{-i}) : \theta_i \in \Theta_i\}$ is the range of f given that the reported types of all agents except i are fixed to $\boldsymbol{\theta}_{-i}$.

Proof. "$\Leftarrow$": Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $i \in N$ and $\theta_i' \in \Theta_i$ be arbitrary. We obtain

$$\begin{aligned}
u_i\big((f(\boldsymbol{\theta}), \mathbf{p}(\boldsymbol{\theta})), \theta_i\big) &= v_i(f(\boldsymbol{\theta}), \theta_i) - p_i(\boldsymbol{\theta}) \\
&= v_i(f(\boldsymbol{\theta}), \theta_i) - t_i(\boldsymbol{\theta}_{-i}, f(\boldsymbol{\theta})) \\
&= \max_{a \in A(\boldsymbol{\theta}_{-i})} \{v_i(a, \theta_i) - t_i(\boldsymbol{\theta}_{-i}, a)\} \\
&\geq v_i(f(\theta_i', \boldsymbol{\theta}_{-i}), \theta_i) - t_i(\boldsymbol{\theta}_{-i}, f(\theta_i', \boldsymbol{\theta}_{-i})) \\
&= u_i\big((f(\theta_i', \boldsymbol{\theta}_{-i}), \mathbf{p}(\theta_i', \boldsymbol{\theta}_{-i})), \theta_i\big),
\end{aligned}$$

so it is not favorable to player i to misreport their type.
"$\Rightarrow$": To see that condition 1. is necessary, assume that there are $i \in N$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\theta_i' \in \Theta_i$ such that $f(\boldsymbol{\theta}) = f(\theta_i', \boldsymbol{\theta}_{-i})$ but $p_i(\boldsymbol{\theta}) > p_i(\theta_i', \boldsymbol{\theta}_{-i})$. Then a player with type θ_i will increase their utility by declaring θ_i'.
To see that also condition 2. is necessary, assume there are $i \in N$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ such that $f(\boldsymbol{\theta}) \notin \arg\max_{a \in A(\boldsymbol{\theta}_{-i})}\{v_i(a, \theta_i) - t_i(\boldsymbol{\theta}_{-i}, a)\}$. Then, there is $\theta' \in \Theta_i$ with $f(\theta_i', \boldsymbol{\theta}_{-i}) \in \arg\max_{a \in A(\boldsymbol{\theta}_{-i})}\{v_i(a, \theta_i) - t_i(\boldsymbol{\theta}_{-i}, a)\}$ and a player with type θ_i gains from declaring type θ_i'. □

We first establish a weak form of monotonicity as a necessary condition on implementability.

Lemma 5.26 (Weak Monotonicity Lemma). Let $f : \boldsymbol{\Theta} \to A$ be implementable and let $a, b \in A$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $i \in N$ and $\theta_i' \in \Theta_i$ be such that $f(\boldsymbol{\theta}) = a \neq b = f(\theta_i', \boldsymbol{\theta}_{-i})$. Then,

$$v_i(b, \theta_i') - v_i(a, \theta_i') \geq v_i(b, \theta_i) - v_i(a, \theta_i).$$

Proof. Let $(f, \mathbf{p})$ be strategyproof and let $i \in N$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\theta_i' \in \Theta_i$ with $f(\boldsymbol{\theta}) = a \neq b = f(\theta_i', \boldsymbol{\theta}_{-i})$ be arbitrary. By Lemma 5.25, there is a price function $t_i : \boldsymbol{\Theta}_{-i} \times A \to \mathbb{R}$ such that $p_i(\boldsymbol{\theta}) = t_i(\boldsymbol{\theta}_{-i}, f(\boldsymbol{\theta}))$. By strategyproofness,

$$v_i(a, \theta_i) - t_i(\boldsymbol{\theta}_{-i}, a) \geq v_i(b, \theta_i) - p_i(\boldsymbol{\theta}_{-i}, b). \tag{5.1}$$

With the same arguments

$$v_i(a, \theta_i') - t_i(\boldsymbol{\theta}_{-i}, a) \leq v_i(b, \theta_i') - p_i(\boldsymbol{\theta}_{-i}, b). \tag{5.2}$$

Subtracting (5.2) from (5.1), we obtain $v_i(a, \theta) - v_i(a, \theta') \geq v_i(b, \theta) - v_i(b, \theta')$, as claimed. □

We are now ready to prove a necessary condition on implementability that is similar in spirit to the monotonicity condition in the setting of the Gibbard-Satterthwaite-Theorem.

Lemma 5.27 (Monotonicity Lemma). Let f be implementable and $\theta \in \boldsymbol{\Theta}$ be such that $f(\theta) = a$ and let $\theta' \in \boldsymbol{\Theta}$ be such that

$$v_i(a, \theta_i') - v_i(b, \theta_i') > v_i(a, \theta) - v_i(b, \theta_i)$$

for all $b \in A \setminus \{a\}$ and all $i \in N$. Then, $f(\theta') = a$.

Proof. Sei $v^i = (v_1', \ldots, v_i', v_{i+1}, \ldots, v_n)$ das Werteprofil, in dem die Spieler 1 bis i den durch v' gegebenen Wert und alle anderen Spieler den durch v gegebenen Wert haben. Es gilt also insbesondere, dass $v^0 = v$, $v^n = v'$ und damit auch $f(v^0) = x$. Angenommen, es gilt $f(v^{i-1}) = x$ für ein $i \in \{1, \ldots, n\}$. Für jede alternative Allokation $y \neq x$ gilt nun nach Voraussetzung, dass $v_i^i(y) - v_i^{i-1}(y) < v_i^i(x) - v_i^{i-1}(x)$. Außerdem gilt nach Konstruktion $v_{-i}^{i-1} = v_{-i}^i$. Mit der schwachen Monotonie von f folgt $f(v^i) = x$. Durch Induktion erhalten wir $f(v^n) = x$ wie behauptet. □

5.3.2 Unrestricted Domains

In the following, we write $\boldsymbol{v}(a, \theta) = (v_i(a, \theta_i)_{i \in N}$.

Definition 5.28 (Neutral Social Choice Function). A social choice function $f : \boldsymbol{\Theta} \to A$ is called neutral if $f(\theta) = a$ for all $\theta \in \boldsymbol{\Theta}$ and $a \in A$ with $\boldsymbol{v}(a, \theta) > \boldsymbol{v}(b, \theta)$ for all $b \in N$.

Theorem 5.29 (Roberts, 1979). Let $|A| \geq 0$ and $\{v_i(a, \theta_i)_{a \in A} : \theta \in \boldsymbol{\Theta}\} = \mathbb{R}^{|A|}$ for every $i \in N$. A neutral surjective social choice $f : \boldsymbol{\Theta} \to A$ is implementable by a strategyproof mechanism $(f, \boldsymbol{p})$ if and only if f is an affine maximizer of social welfare, i.e., $f(\theta) \in \arg\max_{a \in A}\{\sum_{i \in N} \alpha_i v_i(a, \theta_i)\}$ for some constants $\alpha_i \in \mathbb{R}$, $i \in N$ and all $\theta \in \boldsymbol{\Theta}$.

Lemma 5.30 (Pareto Lemma). Let $f : \boldsymbol{\Theta} \to A$ be implementable and surjective with $\{v_i(a, \theta_i)_{a \in A} : \theta_i \in \Theta_i\} = \mathbb{R}^{|A|}$ for all $i \in N$ and $\theta, \theta' \in \boldsymbol{\Theta}$, $a, b \in A$ such that $f(\theta) = a$ and $\boldsymbol{v}(a, \theta') - \boldsymbol{v}(b, \theta') > \boldsymbol{v}(a, \theta) - \boldsymbol{v}(b, \theta)$. Then, $f(\theta') \neq b$

Proof. Let $\theta, \theta' \in \boldsymbol{\Theta}$ with $f(\theta) = a$ as in the statement of the lemma. For a contradiction, assume that $f(\theta') = b$. Consider a vector $\boldsymbol{\Delta} \in \mathbb{R}^n$ such that

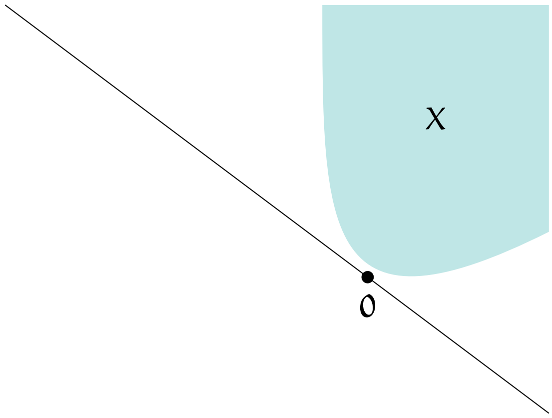

Figure 5.3: Illustration of the Separation Theorem for Convex Sets (Theorem ??).

$0 < \boldsymbol{\Delta} < \big(\boldsymbol{v}(a,\theta') - \boldsymbol{v}(b,\theta')\big) - \big(\boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta)\big)$. Consider θ'' such that

$$v_i(c,\theta_i'') = \begin{cases} \min\{v_i(c,\theta_i), v_i(c,\theta_i') + v_i(a,\theta_i) - v_i(a,\theta_i')\} - \Delta_i, & \text{if } c \in A \setminus \{a,b\} \\ v_i(a,\theta_i) - \frac{\Delta_i}{2}, & \text{if } c = a \\ v_i(b,\theta_i), & \text{if } c = b \end{cases}$$

for all $i \in N$.
We obtain $f(\theta) = a$ und $v_i(a,\theta_i'') - v_i(c,\theta_i'') > v_i(a,\theta_i) - v_i(c,\theta_i)$ for all $c \in A \setminus \{a\}$. The monotonicity property shown in Lemma 5.27 implies $f(\theta'') = a$. Similarly, $f(\theta') = b$ and for all $i \in N$ and $c \in A \setminus \{a,b\}$ we obtain

$$\begin{aligned} v_i(b,\theta_i'') - v_i(c,\theta_i'') &\geq v_i(b,\theta_i) - v_i(c,\theta_i') - v_i(a,\theta_i) + v_i(a,\theta_i') - \Delta_i \\ &> v_i(b,\theta_i') - v_i(c,\theta_i). \end{aligned}$$

Moreover, we obtain for $i \in N$ that

$$v_i(b,\theta_i'') - v_i(a,\theta_i'') > v_i(b,\theta_i) - v_i(a,\theta_i)$$

Lemma 5.27 implies $f(\theta'') = a$, a contradiction. □

We are now ready to prove Robert's Theorem. We will use the following separation theorm for convex sets, see Figure 5.3 for an illustration.

Theorem 5.31 (Separation Theorem for Convex Sets). Let $X \subset R^n$ be a non-empty, convex set with $0 \notin X$. Then there exists a hyperplane H that separates X and 0, i.e., there are $\boldsymbol{\alpha} \in \mathbb{R}^n$, $a \in \mathbb{R}$ such that $\boldsymbol{\alpha} \cdot \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in X$.

Proof of Theorem 5.29.
For two alternatives $a, b \in A$, define

$$P(a,b) = \{\boldsymbol{\lambda} \in \mathbb{R}^n : \text{ there is } \theta \in \boldsymbol{\Theta} \text{ with } f(\theta) = a \text{ and } \boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta) = \boldsymbol{\lambda}\}. \tag{5.3}$$

Moreover, let $\mathring{P}(a,b)$ denote the interior of $P(a,b)$.
It is sufficient to prove that there is a vector $\boldsymbol{\alpha} \in R$ such that $\boldsymbol{\alpha} \cdot \boldsymbol{v}(a,\theta) \geq v(b,\theta)$ for all $b \in A$ whenever $f(\theta) = a$. To this end, we will show that $\mathring{P}(a,b)$ is independent of a and b and convex.
First, note that because f is surjective, $P(a,b) \neq \emptyset$ for all $a, b \in A$. Moreover, monotonicity gives that $\boldsymbol{\lambda} \in P(a,b)$ implies $\boldsymbol{\lambda}' \in P(a,b)$ for all $\boldsymbol{\lambda}' > \boldsymbol{\lambda}$.
We proceed to prove several claims on $P(a,b)$. For the following let $\boldsymbol{\lambda}, \boldsymbol{\kappa} \in \mathbb{R}^n$ and $\boldsymbol{\epsilon}, \boldsymbol{\delta} \in R^n_{>0}$.

First claim: $\boldsymbol{\lambda} - \boldsymbol{\epsilon} \in P(a,b)$ implies $-\boldsymbol{\lambda} \notin P(b,a)$.
For a contradiction, assume that $\boldsymbol{\lambda} - \boldsymbol{\epsilon} \in P(a,b)$ and $-\boldsymbol{\lambda} \in P(b,a)$. Then, there are $\theta, \theta \in \boldsymbol{\Theta}$ with $f(\theta) = a$ and $f(\theta') = b$ and

$$\boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta) = \boldsymbol{\lambda} - \boldsymbol{\epsilon} \qquad \text{and} \qquad \boldsymbol{v}(b,\theta') - \boldsymbol{v}(a,\theta') = -\boldsymbol{\lambda}.$$

This implies $\boldsymbol{v}(a,\theta') - \boldsymbol{v}(b,\theta') > \boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta)$ for all $i \in N$ contradicting the Pareto-condition of f (Lemma 5.30).

Second claim: $\boldsymbol{\lambda} \notin P(a,b)$ implies $-\boldsymbol{\lambda} \in P(b,a)$. Let $\boldsymbol{\epsilon} \in R^n_{>0}$ be arbitrary and consider $\theta \in \boldsymbol{\Theta}$ such that

$$\boldsymbol{v}(a,\theta) - \boldsymbol{v}(c,\theta) = \begin{cases} \boldsymbol{\lambda} & \text{if } c = b, \\ \boldsymbol{\kappa}^c + \boldsymbol{\epsilon} & \text{else,} \end{cases}$$

where $\boldsymbol{\kappa}^c \in P(a,c)$ is arbitrary. By construction, for all $c \in A \setminus \{a,c\}$ we have $\boldsymbol{v}(c,\theta) - \boldsymbol{v}(a,\theta) = -\boldsymbol{\kappa}^c - \boldsymbol{\epsilon}$. With the first claim this implies $\boldsymbol{\kappa} \notin P(c,a)$ and hence $f(\theta) \in \{a,b\}$. Moreover, $\boldsymbol{\lambda} \notin P(a,b)$ implies $f(\theta) \neq a$. We conclude that $f(\theta) = b$ and hence $-\boldsymbol{\lambda} \in P(b,a)$.

Third claim: $\boldsymbol{\lambda} - \boldsymbol{\epsilon} \in P(a,b)$ and $\boldsymbol{\kappa} - \boldsymbol{\delta} \in P(b,c)$ implies $\boldsymbol{\lambda} + \boldsymbol{\kappa} - \frac{\boldsymbol{\epsilon}+\boldsymbol{\delta}}{2} \in P(a,c)$.
For all $d \in A \setminus \{a,b,c\}$ choose $\boldsymbol{\delta}^d \in P(x,w)$. Let $\theta \in \boldsymbol{\Theta}$ be such that

$$\begin{aligned} \boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta) \qquad\qquad & = \boldsymbol{\lambda} - \frac{\boldsymbol{\epsilon}}{2} \\ \boldsymbol{v}(b,\theta) - \boldsymbol{v}(c,\theta) \qquad & = \boldsymbol{\kappa} - \frac{\boldsymbol{\delta}}{2} \\ \boldsymbol{v}(a,\theta) \qquad\qquad - \boldsymbol{v}(d,\theta) & = \boldsymbol{\delta}^d + \boldsymbol{\epsilon}', \end{aligned}$$

for all $d \in A \setminus \{a,b,c\}$ where $\boldsymbol{\epsilon}' \in \mathbb{R}^n_{>0}$ is arbitrary. Using $\lambda - \boldsymbol{\epsilon} \in P(a,b)$ there is $\theta^a \in \boldsymbol{\Theta}$ with $f(\theta^a) = a$ and $\boldsymbol{v}(a,\theta^a) - \boldsymbol{v}(b,\theta^a) = \boldsymbol{\lambda} - \boldsymbol{\epsilon}$. We obtain

$$\boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta) = \boldsymbol{\lambda} - \frac{\boldsymbol{\epsilon}}{2} > \boldsymbol{\lambda} - \boldsymbol{\epsilon} = \boldsymbol{v}(a,\theta^a) - \boldsymbol{v}(a,\theta^a).$$

Lemma 5.30 then implies that $f(\theta) \neq b$. Similarly, one can show that $f(\theta) \neq d$ for all $d \in A \setminus \{a,b\}$ which then implies $f(\theta) = a$. Using that $\boldsymbol{v}(a,\theta) - \boldsymbol{v}(c,\theta) = \boldsymbol{\lambda} + \boldsymbol{\kappa} - \frac{\boldsymbol{\epsilon}+\boldsymbol{\delta}}{2}$ the claim follows.

Fourth claim: $\boldsymbol{\lambda} \in \mathring{P}(a,b)$ implies $\boldsymbol{\lambda} \in \mathring{P}(a,c)$ for all $A \setminus \{a\}$, where for a set A, $\mathring{A}$ denotes the interior of A.
If $\boldsymbol{\lambda} \in \mathring{P}(a,b)$, then $\boldsymbol{\lambda} - \boldsymbol{\epsilon} \in P(a,b)$ for some $\boldsymbol{\epsilon} > 0$. Using that f is neutral, we have $\frac{\boldsymbol{\epsilon}}{4} - \frac{\boldsymbol{\epsilon}}{8} = \frac{\boldsymbol{\epsilon}}{8} \in P(b,c)$. By the third claim, $\boldsymbol{\lambda} + \frac{\boldsymbol{\epsilon}}{4} - \frac{\boldsymbol{\epsilon}+\boldsymbol{\epsilon}/8}{2} = \boldsymbol{\lambda} - \frac{\boldsymbol{\epsilon}}{4} + \frac{\boldsymbol{\epsilon}}{16} \in P(a,c)$. We conclude $\boldsymbol{\lambda} \in \mathring{P}(a,c)$.

Fifth claim: $\boldsymbol{\lambda} \in \mathring{P}(a,c)$ implies $\boldsymbol{\lambda} \in \mathring{P}(b,c)$ for all $b \in A \setminus \{a\}$. The proof is similar to that of the forth claim and thus omitted.

The fourth and fifth claim together imply that $P = \mathring{P}(a,b) = \mathring{P}(c,d)$ for all $a, b, c, d \in A$.

Sixth claim: P is convex.
For the proof, we shall show that for arbitrary $\boldsymbol{\lambda}, \boldsymbol{\kappa} \in P$ we have $\frac{\boldsymbol{\lambda}+\boldsymbol{\kappa}}{2} \in P$. For a contradiction, assume $\frac{\boldsymbol{\lambda}+\boldsymbol{\kappa}}{2} \notin P$. The second claim then implies $-\frac{\boldsymbol{\lambda}+\boldsymbol{\kappa}}{2} \in \bar{P}$ and thus $-\frac{\boldsymbol{\lambda}+\boldsymbol{\kappa}}{2} + \boldsymbol{\epsilon} \in P$ for all $\epsilon \in \mathbb{R}^n_{>0}$. In addition, since $\boldsymbol{\lambda}, \boldsymbol{\kappa} \in P$ and P is open, there is $\boldsymbol{\epsilon} \in \mathbb{R}^n_{>0}$ such that $\boldsymbol{\lambda} - 3\boldsymbol{\epsilon}, \boldsymbol{\kappa} - 3\boldsymbol{\epsilon} \in P$. By the third claim $\boldsymbol{\lambda} + \boldsymbol{\kappa} - 3\boldsymbol{\epsilon} \in P$. Again by the the third claim then $\frac{\boldsymbol{\lambda}+\boldsymbol{\kappa}}{2} - \boldsymbol{\epsilon} \in \bar{P}$, a contradiction.

We are now ready to prove the main result. As f is neutral, $\mathbf{0} \in \bar{P} \setminus P$. Applying Theorem 5.31 to P, we derive the existence of a vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\boldsymbol{\alpha} \cdot \boldsymbol{\lambda} \geq 0$ for all $\boldsymbol{\lambda} \in \bar{P}$. Consider a type profile $\theta \in \boldsymbol{\Theta}$ with $f(\theta) = a$ and let $b \in A \setminus \{a\}$ be arbitrary. We obtain $\boldsymbol{v}(a,\theta) - \boldsymbol{v}(b,\theta) \in P(a,b)$ which implies $\boldsymbol{\alpha} \cdot (v(a,\theta) - v(b,\theta)) \geq 0$, i.e., $\boldsymbol{\alpha} \cdot v(a,\theta) \geq \boldsymbol{\alpha} \cdot v(b,\theta)$. □