Social Welfare Functions

Shivansh Nair, Maths and Physics Club

July 2025

§1 Introduction

In the last few weeks, we have explored many actual games with strategies and rewards, and tried to find out the optimal strategies. Now this week let us instead try to model real life democracy, and figure out what should be the outcome based on the votes of people. By the end we will obtain a cool result about the impossibility of democracy.

§2 Voting Examples

First, let us look at various examples where we have to make a decision based on society's preferences, and try to rationalize some outcomes.

Example 22.1: The Committee Choice Problem

A committee composed of 21 people needs to select one individual from among three candidates: A, B, and C. The preference profile of the committee members is summarized below:

No. of Members	First Choice	Second Choice	Third Choice
1	A	B	C
7	A	C	B
7	B	C	A
6	C	B	A

There are multiple ways of selecting the winning candidate.
Since majority voting is a common method, it is natural to ask whether one candidate defeats every other candidate in a head-to-head majority contest. Such a candidate is called a **Condorcet winner**.

Definition 2.1. A candidate is a **Condorcet winner** if they defeat every other candidate in a pairwise majority vote.

In this case:

- If only A and C exist, then A has 8 votes while C has 13 votes.
- If only B and C exist, then B has 8 votes while C has 13 votes.
- If only A and B exist, then A has 8 votes while B has 13 votes.

Since C would win against both A and B individually, they are the Condorcet winner.

However, what if we simply took everyone's vote to be their first preference? In this case A gets 8 votes, B gets 7 votes, and C only gets 6 votes. We see that just by changing how we interpret the votes we get a completely reverse order than our earlier method, even though everyone's vote is the same. So choosing this method is very important as it can easily affect the results. It is also not necessary for a Condorcet winner to always exist.

Now instead of just choosing 1 winner, let us look at the more general case, where given every voter's preferences between a set of candidates, we order all of the candidates in a reasonable manner.

§3 Social Welfare Functions

Let A be a finite set of alternatives, and $N = \{1, 2, \ldots, n\}$ be a set of individuals (voters or decision makers). Each individual has a preference relation over the elements of A.

§3.1 Preference Relations

Definition 3.1. A **preference relation** $\succsim_i$ of player i over a set A is a relation satisfying:

- **Completeness:** For every $a, b \in A$, either $a \succsim_i b$ or $b \succsim_i a$.
- **Reflexivity:** $a \succsim_i a$ for all $a \in A$.
- **Transitivity:** If $a \succsim_i b$ and $b \succsim_i c$, then $a \succsim_i c$.

Definition 3.2. A **strict preference relation** $\succ_i$ of player i over A satisfies:

- For every distinct $a, b \in A$, either $a \succ_i b$ or $b \succ_i a$.
- **Irreflexivity:** $a \not\succ_i a$ for all $a \in A$.
- **Transitivity:** If $a \succ_i b$ and $b \succ_i c$, then $a \succ_i c$.

> **Remark 3.3.** If $a \succsim_i b$ and $b \succsim_i a$, then a and b are **equivalent** under $\succsim_i$, denoted $a \sim_i b$. Note that $a \sim_i b$ does not imply $a = b$.

§3.2 Examples of Preference Relations

- Let $A = \{-m, -m+1, \ldots, m\} \subset \mathbb{Z}$. Then the usual numerical ordering $\leq$ defines a preference relation, and $<$ defines a strict preference relation.
- Define $\succsim_i$ by $a \succsim_i b \iff |a| \leq |b|$. Then $\succsim_i$ is a preference relation where, for example, $k \sim_i -k$.
- Define a **lexicographic preference relation** on a set $A = \{(n, m) \mid 1 \leq n \leq N,\ 1 \leq m \leq M\}$ by:

$$(n, m) \succsim_L (n', m') \iff (n > n') \text{ or } (n = n' \text{ and } m \geq m').$$

§3.3 Preference Profiles and Social Welfare Functions

Denote by $\mathcal{P}^*(A)$ the set of all preference relations over A, and by $\mathcal{P}(A)$ the set of all strict preference relations over A.

Definition 3.4. A **strict preference profile** is a list $P^N = (P_i)_{i\in N}$ of strict preferences, one for each individual. The collection of all strict preference profiles is the Cartesian product:

$$\mathcal{P}(A)^N = \mathcal{P}(A) \times \mathcal{P}(A) \times \cdots \times \mathcal{P}(A).$$

A strict preference profile describes how each individual in society ranks all the alternatives. The problem before us is how to *aggregate* all the preferences in a strict preference profile into one preference relation, called the **social preference relation**.

Definition 3.5. A **social welfare function** is a function F that maps each strict preference profile $P^N = (P_i)_{i\in N} \in \mathcal{P}(A)^N$ to a preference relation in $\mathcal{P}^*(A)$. This function is denoted by $F(P^N)$.

In other words, a social welfare function summarizes the opinions of everyone in society: given the strict preference relations $P^N = (P_i)_{i\in N}$ of all individuals, society as a collective ranks the alternatives in A via the preference relation $F(P^N)$.

> **Remark 3.6.** If society ranks a above b, that is, $a \succsim_{F(P^N)} b$, we say that society (weakly) prefers a to b.

In the input of the social welfare function we are assuming that it is a strict preference profile, that is no one equally values or is indifferent between 2 candidates. However our output preference relation is not strict, 2 candidates can be valued equally. Let us look at an example to see why:

If there are 2 candidates, and each voter strictly chooses one over the other, what happens if both of them get equal votes? If we stick to giving a strict preference relation in the output, we would have to randomly choose between the candidates. Hence to avoid this, we relax our output to be a not necessarily strict preference relation, even though all the inputs are strict.

§4 Requirements from social welfare function's

Now that we have defined social welfare functions, lets look at some reasonable properties we would like them to have.

§4.1 Dictatorship

Firstly, since we want a democratic outcome, we would not want there to be a dictator present, someone who can only by himself decide the final outcome based on his votes. Formally written as:

Definition 4.1 (Dictatorship)**.** A social welfare function F is called **dictatorial** if there exists an individual $i \in N$ such that

$$F(P^N) = P_i \quad \text{for every profile of strict preferences } P^N.$$

In other words, for every pair of alternatives $a, b \in A$, and every strict preference profile P^N,

$$a \succ_{P_i} b \;\Rightarrow\; a \succ_{F(P^N)} b.$$

In this case, individual i is called a **dictator**.

We would like our social welfare functions to be **non dictatorial**.

§4.2 Unanimity

Another reasonable need from our function would be unanimity. If every voter prefers a to b in their preference profiles, society also prefers a to b. Formally written as:

Definition 4.2 (Unanimity)**.** A social welfare function F satisfies the property of **unanimity** if the following holds:

For every pair of alternatives $a, b \in A$, and every strict preference profile $P^N = (P_i)_{i \in N}$, if

$$a \succ_{P_i} b \quad \text{for every } i \in N,$$

then

$$a \succ_{F(P^N)} b.$$

In words, if every individual strictly prefers a over b, then society must also strictly prefer a over b.

§4.3 Independence of irrelevant alternatives

We would also like our function to determine the preference between a and b solely between the voter's preferences of a and b, and want it to remain unchanged if unrelated preferences are changed. Formally written as:

Definition 4.3 (Independence of Irrelevant Alternatives (IIA))**.** A social welfare function F satisfies the property of **independence of irrelevant alternatives (IIA)** if:

For every pair of alternatives $a, b \in A$, and every pair of strict preference profiles P^N and Q^N, if

$$a \succ_{P_i} b \iff a \succ_{Q_i} b \quad \text{for all } i \in N,$$

then

$$a \succsim_{F(P^N)} b \iff a \succsim_{F(Q^N)} b.$$

In other words, if every individual has the same strict preference between a and b in both profiles, then the social ranking between a and b must also be the same in both social rankings.

If a social welfare is dictatorial, it can easily be proven that it satisfies unanimity and IIA. We can also see that for 2 candidates, a function choosing the majority vote clearly satisfies all 3 of our desired properties. But can we find similar functions for more than 3 candidates? Suprisingly, it is not possible.

§5 Arrow's impossibility theorem

Theorem 5.1 (Arrow's Impossibility Theorem)

If $|A| \geq 3$, then there does not exist a social welfare function that satisfies all three of the following properties:

- **Unanimity**
- **Independence of Irrelevant Alternatives (IIA)**
- **Nondictatorship**

To prove this theorem, we need to define and explore some new terms.

Definition 5.2. A **coalition** is a subset of individuals $S \subseteq N$.

Definition 5.3. Let F be a social welfare function, and let $a, b \in A$ be two distinct alternatives. A coalition $S \subseteq N$ is called **decisive for** a **over** b (relative to F) if for every strict preference profile $P^N \in (P(A))^N$ satisfying:

1. $a \succ_{P_i} b$ for every $i \in S$,
2. $b \succ_{P_j} a$ for every $j \notin S$,

it holds that $a \succ_{F(P^N)} b$.

The coalition S is called **decisive** (relative to F) if there exists a pair of alternatives $a, b \in A$ such that S is decisive for a over b.

In words, a set of individuals S is decisive for a over b if when every member of S prefers a to b, and all the other individuals prefer b to a, society prefers a to b.

If we apply unanimity onto coalitions, then we know that if every pair chooses a over b society chooses a over b. Hence we get the following theorem.

Theorem 5.4

Let F be a social welfare function satisfying the unanimity property. Then for every $a, b \in A$:

- The coalition N is **decisive** for a over b.
- The empty coalition $\varnothing$ is **not decisive** for a over b.

Now if we apply IIA onto coalitions, the ranking between a and b only depends on everyone's preference between a and b. Hence the *for all* condition in Definition 5.3 can be changed to *there exists*, since the ranking of a and b is unchanged in every such preference profile. Hence we get the folllowing theorem.

Theorem 5.5

Let F be a social welfare function satisfying the independence of irrelevant alternatives (IIA) property, and let $a, b \in A$ be two alternatives. A coalition $S \subseteq N$ is decisive for a over b if and only if there exists a strict preference profile P^N satisfying:

(a1) $a \succ_{P_i} b$ for all $i \in S$,

(a2) $b \succ_{P_j} a$ for all $j \notin S$,

(a3) $a \succ_{F(P^N)} b$.

It follows that if a coalition S is not decisive for a over b, and if a strict preference profile P^N satisfies:

(b1) $a \succ_{P_i} b$ for all $i \in S$,

(b2) $b \succ_{P_j} a$ for all $j \notin S$,

then $b \succsim_{F(P^N)} a$.

Now we will prove the following theorem, which states that if a coalition is decisive for any single pair, it is decisive for every single pair.

Theorem 5.6

Suppose that $|A| \geq 3$ and that F satisfies the unanimity and independence of irrelevant alternatives (IIA) properties. If a coalition V is decisive for some pair a^* over b^*, then V is decisive for every pair of alternatives in A.

Proof. Let a and b be a pair of alternatives.

Part 1: If V is decisive for a over b, then V is decisive for a over c, for any alternative $c \in A \setminus \{a\}$.

If $c = b$, the claim follows by assumption. Otherwise, let $c \in A \setminus \{a, b\}$. Consider the following strict preference profile P^N:

$$\begin{cases} a \succ_{P_i} b \succ_{P_i} c & \text{if } i \in V, \\ b \succ_{P_i} c \succ_{P_i} a & \text{if } i \notin V. \end{cases}$$

All the other alternatives in A are ordered arbitrarily by each individual.

Since V is decisive for a over b, it follows that $a \succ_{F(P^N)} b$. Since F satisfies the unanimity property, we also have $b \succ_{F(P^N)} c$. By transitivity of $F(P^N)$, it follows that $a \succ_{F(P^N)} c$. By Theorem 5.5, V is decisive for a over c.

Part 2: If V is decisive for a over b, then V is decisive for b over c, for any $c \in A \setminus \{a, b\}$.

Let $c \in A \setminus \{a, b\}$. Consider the following strict preference profile P^N:

$$\begin{cases} b \succ_{P_i} a \succ_{P_i} c & \text{if } i \in V, \\ c \succ_{P_i} b \succ_{P_i} a & \text{if } i \notin V. \end{cases}$$

All the other alternatives in A are ordered arbitrarily by each individual.

From Part 1 and the fact that V is decisive for a over b, it follows that V is decisive for a over c, hence $a \succ_{F(P^N)} c$. Since F satisfies unanimity, $b \succ_{F(P^N)} a$. Then by transitivity, $b \succ_{F(P^N)} c$, and so by Theorem 5.5, V is decisive for b over c.

Part 3: The first two parts are sufficient to prove the theorem.

Let $a \neq b$ be any pair of alternatives in A. We aim to show that V is decisive for a over b. Recall that V is decisive for some fixed pair a^* over b^*. Consider three cases:

- If $a = a^*$, then from Part 1 and the fact that V is decisive for a^* over b^*, we deduce that V is decisive for a over b.
- If $a \neq a^*$ and $b \neq a^*$, then from Part 1 and the fact that V is decisive for a^* over b^*, we have that V is decisive for a^* over a. From Part 2, it then follows that V is decisive for a over b.
- If $a \neq a^*$ and $b = a^*$, then choose an alternative $c \in A \setminus \{a, b\}$ (such a c exists since $|A| \geq 3$). From Part 1 and the fact that V is decisive for a^* over b^*, we have that V is decisive for b over c. Then from Part 2, V is decisive for c over a, and again from Part 2, V is decisive for a over b.

Hence, in all cases, V is decisive for a over b. This completes the proof. □

Now we can finally prove arrow's impossibility theorem. To do this, we will first prove that there exists a decisive coalition containing a single individual. Then we will show, that individual is the dictator required by our proof of Arrow's Impossibility theorem.

Claim 5.7 — There exists a decisive coalition V containing a single individual.

Proof. Let V be a nonempty decisive coalition containing a minimal number of individuals. Since by Theorem 5.4 the coalition N is decisive and the empty coalition $\emptyset$ is not decisive, there must be such a coalition. If V has only 1 individual, our proof is done. Now we will show that if V has more than 1 individual, it leads to a contradiction.

Let $j \in V$, and define $U = V \setminus \{j\}$ and $W = N \setminus V$. Since V contains at least two individuals, the coalition U is nonempty. By minimality of V, both U and $\{j\}$ are non-decisive coalitions, since their size is smaller than V.

Since $|A| \geq 3$, we can choose three distinct alternatives $a, b, c \in A$. Consider the strict preference profile $P^N = (P_i)_{i \in N}$ defined as follows:

$$\begin{cases} a \succ_{P_i} b \succ_{P_i} c & \text{if } i = j, \\ c \succ_{P_i} a \succ_{P_i} b & \text{if } i \in U, \\ b \succ_{P_i} c \succ_{P_i} a & \text{if } i \in W. \end{cases}$$

The rest of the alternatives in A are ordered arbitrarily by each individual.

Since V is decisive, by Theorem 5.6, it is decisive for every pair of alternatives. In particular, it is decisive for a over b. Since $a \succ_i b$ for every $i \in V = U \cup \{j\}$, and $b \succ_i a$ for every $i \in N \setminus V = W$, we have $a \succ_{F(P^N)} b$.

Now, U is not decisive for c over b, but $c \succ_i b$ for every $i \in U$, and $b \succ_i c$ for every $i \in N \setminus U = W \cup \{j\}$. Therefore, by Theorem 5.5, we must have $b \succ_{F(P^N)} c$, as otherwise U would be decisive.

Since $F(P^N)$ is transitive, and we have $a \succ_{F(P^N)} b$ and $b \succ_{F(P^N)} c$, it follows that $a \succ_{F(P^N)} c$.

Now observe that $a \succ_j c$, but $c \succ_i a$ for all $i \neq j$. Hence, by Theorem 5.5, $\{j\}$ must be a decisive coalition for a over c, contradicting the assumption that $\{j\}$ is not decisive. Hence proved by contradiction, V can't have more than 1 element.

Therefore, the coalition V must contain exactly one individual. Let $V = \{j\}$. We next prove that this individual j is a dictator. □

Claim 5.8 — Individual j is a dictator, i.e., $F(P^N) = P_j$ for every $P^N = (P_i)_{i \in N} \in (\mathcal{P}(A))^N$.

It might seem obvious that if j is decisive, it is a dictator. Since even when j chooses a over b, and every single other person chooses b over a, the social function still chooses a. However this alone does not prove the cases where someone other than j also chooses a over b. Even though it might seem logical that if even more people are choosing a over b the choice shouldn't change, none of our theorems directly state that. Hence we need a proper proof.

Proof. Let P^N be a strict preference profile, and let $a, b \in A$ be two different alternatives such that $a \succ_j b$. We wish to show that $a \succ_{F(P^N)} b$.
Since A contains at least three alternatives, there exists an alternative $c \in A \setminus \{a, b\}$. Consider the following strict preference profile Q^N:

$$Q_i = \begin{cases} a \succ_{Q_i} c \succ_{Q_i} b & \text{if } i = j, \\ c \succ_{Q_i} a \succ_{Q_i} b & \text{if } i \neq j \text{ and } a \succ_{P_i} b, \\ c \succ_{Q_i} b \succ_{Q_i} a & \text{if } i \neq j \text{ and } b \succ_{P_i} a. \end{cases}$$

The other alternatives in A are ordered arbitrarily by each individual.
Since $\{j\}$ is decisive for any pair of alternatives, it is in particular decisive for a over c, and therefore $a \succ_{F(Q^N)} c$. Since F satisfies the unanimity property, it follows that $c \succ_{F(Q^N)} b$. By transitivity of $F(Q^N)$, we deduce that $a \succ_{F(Q^N)} b$.

Now, individual j prefers a over b in both P_j and Q_j, and every individual $i \neq j$ prefers a over b in P_i if and only if they do in Q_i. Since F satisfies the independence of irrelevant alternatives (IIA), and $a \succ_{F(Q^N)} b$, we conclude $a \succ_{F(P^N)} b$, as required. □

Now we have proved that there is a decisive coalition containing a single individual who is a dictator, hence our proof of Arrow's Impossibility theorem is complete.

§6 Further Reading

In this handout, we have covered Chapter 22.1 of *Game Theory* by Maschler, Solan, and Zamir, which deals with inputting preference orders and outputting a preference order.

However in real life, we often only care about a single winner and not an entire preference order. 22.2 and 22.3 of the same book deal with *social choice functions*, which are similar to social welfare functions except for only outputting a single winner. Interested people can read these sections.

Some of what you will read will include:

- A theorem similar to Arrow's impossibility theorem for social choice functions
- Manipulability of a social choice function, where a person can lie about his true preference order in his votes to reach a better outcome for himself.
- Gibbard Satterthwaite theorem, which states that every social choice function with more than 3 candidates that can not be manipulated in such a way, is dictatorial.