Econ/Math C103

Social Choice Theory

Haluk Ergin

1 Social Welfare Functions

We have a fixed set of n individuals $N = \{1, 2, \ldots, n\}$ and a set of social outcomes X which we will assume to be finite unless otherwise indicated.

Definition 1 Let $\mathcal{R}$ denote the set of all linear orders (i.e. complete, transitive, and antisymmetric relations) over X.

We will assume that each individual $i \in N$ has a preference relation $R_i \in \mathcal{R}$ over the set of social alternatives.[1]

Let us define the strict preference $\succ$ and the indifference $\sim$ relations over X from $\succsim$ (analogous to our definitions of P and I from R) by:

$$x \succ y \;\Leftrightarrow\; x \succsim y \,\&\, \text{ not } y \succsim x$$

$$x \sim y \;\Leftrightarrow\; x \succsim y \,\&\, y \succsim x.$$

We will be analyzing systematic ways of constructing a social preference $\succsim$ for every possible profile of individual preferences $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$. Unlike individual preferences, we will make no a priori restrictions on the social preference any further than requiring it is a binary relation on X:

Definition 2 Let $\mathcal{B}$ denote the set of all binary relations over X.

Such systematic procedures for constructing a social preference $\succsim \in \mathcal{B}$ for each preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$ are called social welfare functions. More formally:

Definition 3 A **social welfare function (SWF)** is a function $f : \mathcal{R}^n \to \mathcal{B}$.

[1]Restricting individual preferences to be antisymmetric is not a substantial assumption. I make it here mainly because of expositional purposes. Arrow's Impossibility Theorem and the Gibbard-Satthertwaite Theorem can be extended without significant additional effort if one allows for indifferences in individuals' preferences.

Examples of SWF's:

- **Pairwise majority:** Given any preference profile $R = (R_1, \dots, R_n) \in \mathcal{R}^n$, pairwise majority selects the social preference $\succsim = f(R)$ given by:

$$x \succsim y \;\; \Leftrightarrow \;\; |\{i \in N | xR_iy\}| \geq |\{i \in N | yR_ix\}|$$

  for any $x, y \in X$. That is, x is socially weakly preferred to y if the number of individuals who weakly prefer x to y is greater than or equal to the number of individuals who weakly prefer y to x.

- **Borda count:** Given any preference profile $R = (R_1, \dots, R_n) \in \mathcal{R}^n$ and alternative $x \in X$ define the Borda count of the alternative X with respect to profile R as:

$$BC(x, R) = \sum_{i \in N} |\{z \in X | xR_iz\}|$$

  In words, $BC(x, R)$ is calculated as follows. First for each individual i, we find the rank of x from the bottom with respect to i's preference. Than we sum these numbers across the individuals. The Borda count associates to every preference profile $R = (R_1, \dots, R_n) \in \mathcal{R}^n$ the social preference $\succsim = f(R)$ represented by the Borda counts with respect to R:

$$x \succsim y \;\; \Leftrightarrow \;\; BC(x, R) \geq BC(y, R).$$

- **Dictatorship:** Fix an individual $i^* \in N$ as a dictator. For every preference profile $R = (R_1, \dots, R_n) \in \mathcal{R}^n$, set the social preference to be the preference of the dictator, that is $\succsim = f(R) = R_{i^*}$.

- **Anti-dictatorship:** Fix an individual $i^* \in N$ as an anti-dictator. For every preference profile $R = (R_1, \dots, R_n) \in \mathcal{R}^n$, set the social preference $\succsim = f(R)$ to be the preference of the anti-dictator turned upside down, that is:

$$\forall x, y \in X : \quad x \succsim y \;\; \Leftrightarrow \;\; yR_{i^*}x.$$

- **Constant SWF:** Fix any preference $\succsim^* \in \mathcal{B}$. For any preference profile $R = (R_1, \dots, R_n) \in \mathcal{R}^n$, set the social preference to be $f(R) = \succsim^*$ irrespective of the individuals' preferences.

- **Weighted-majority:** Fix a function $w : N \to \mathbb{R}$. We will interpret $w(i)$ as the "weight" of individual i in the society, although we allow $w(i)$ to be zero or negative. Given any preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$, weighted majority (with the weight function w) selects the social preference $\succsim = f(R)$ given by:

$$x \succsim y \;\; \Leftrightarrow \;\; \sum_{i \in N | x R_i y} w(i) \geq \sum_{i \in N | y R_i x} w(i)$$

for any $x, y \in X$. That is x is socially weakly preferred to y if the aggregate weight of individuals who weakly prefer x to y is greater than or equal to the aggregate weight of individuals who weakly prefer y to x.

- **Tournaments:** Given any preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$, let's say that an alternative x beats another alternative y if the number of individuals who weakly prefer x to y is greater than or equal to the number of individuals who weakly prefer y to x, that is if:

  $$|\{i \in N | x R_i y\}| \geq |\{i \in N | y R_i x\}|.$$

  Let $T(x, R)$ denote the number of alternatives that x beats under the preference profile R. The tournament SWF associates to every preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$ the social preference $\succsim = f(R)$ represented by $T(., R)$, that is:

  $$x \succsim y \;\; \Leftrightarrow \;\; T(x, R) \geq T(y, R).$$

2 Properties of Social Welfare Functions

Note that we have not yet required a social preference to be complete and transitive. We will explicitly require this as a rationality condition on the SWF when we want to restrict attention to complete and transitive social preferences.

Definition 4 A SWF f is **rational** if for any preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$, the social preference $\succsim = f(R)$ is complete and transitive.

Note that our rationality condition does not require the social preference to be strict even though we required preferences of individuals to be strict. The reason is that there are instances where a "natural" social preference can be weak although each individual has a strict preference over outcomes. For example, consider the majority rule with two alternatives when there is an equal number of individuals that prefer each alternative over the other.

Definition 5 A SWF f is **unanimous** if for any preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$ and alternatives $x, y \in X$:

$$[\forall i \in N : xP_iy] \ \Rightarrow \ x \succ y$$

where $\succsim = f(R)$.

Unanimity says that if every individual in the society unanimously ranks an alternative x strictly above y, then so should the social preference.

Definition 6 A SWF f is satisfies **independence of irrelevant alternatives (IIA)** if for any two alternatives $x, y \in X$ and preference profiles $R = (R_1, \ldots, R_n), R' = (R'_1, \ldots, R'_n) \in \mathcal{R}^n$ such that:

$$\forall i \in N : \begin{array}{ccc} xR_iy & \Leftrightarrow & xR'_iy \\ yR_ix & \Leftrightarrow & yR'_ix \end{array}$$

we have:

$$\begin{array}{ccc} x \succsim y & \Leftrightarrow & x \succsim' y \\ y \succsim x & \Leftrightarrow & y \succsim' x \end{array}$$

where $\succsim = f(R)$ and $\succsim' = f(R')$.

This one is a mouthful. It says that, if for each individual i the ranking of x versus y is the same in R_i and R'_i, then the social ranking of x versus y should be the same for the profiles R and R'. That is, the social ranking of x versus y should only depend on the individual rankings of x versus y. Any third alternative z is deemed to be "irrelevant" for the social comparison between x and y. How an individual ranks, say y versus z, should not play a role on how the social ranking between x and y is determined, as long as the individual comparisons between x and y remain the same.

The Condorcet Paradox: This example demonstrates that when there are three alternatives, the pairwise majority outcome may lead to intransitivities in the social preference. Let $N = \{1, 2, 3\}$, $X = \{x, y, z\}$, and

R_1	R_2	R_3
x	y	z
y	z	x
z	x	y

Pairwise majority outcome for $R = (R_1, R_2, R_3)$ is $x \succ y \succ z \succ x$, which is intransitive.

Here is the cornerstone result of social choice theory.

Theorem 1 *(Arrow's Impossibility Theorem) Suppose that* $|X| \geq 3$. *If a SWF* $f : \mathcal{R}^n \to \mathcal{B}$ *satisfies Rationality, Unanimity, and IIA, then* f *must be dictatorial, i.e. there exists* $i^* \in N$ *such that* $f(R) = R_{i^*}$ *for any* $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$.

Proof: See Reny (2001) which contains the proof we covered in class. $\square$

3 Strategic Considerations

In a variety of applications, we do not need to determine a full social preference over all possible alternatives, but are just interested in selecting one of the social alternatives in X as a function of the preference profile. Such procedures are called social choice functions.

Definition 7 A **social choice function (SCF)** is a function $f \colon \mathcal{R}^n \to X$.

Typically, we can not directly observe the individuals' preferences over X. In order to determine the social choice out of X, we need to ask each individual to tell us her preferences. A natural question is: "Does the SCF induce them to always tell us their true preferences, or does it in certain instances, give them incentives to manipulate the social outcome by submitting an untruthful preference ranking?" Consider the following example that illustrates this issue.

Example: Let $N = \{1, 2\}$, $X = \{x, y, z\}$, and let f be the social choice function that selects the alternative with the highest Borda count. If there are multiple such alternatives suppose that the ties are broken by favoring x, then y, and then z. Consider the true preference profile $R = (R_1, R_2)$:

R_1	R_2
x	y
y	z
z	x

Since $BC(x, R) = 4$, $BC(y, R) = 5$, and $BC(z, R) = 3$, $y = f(R)$. On the other hand if individual 1 had misrepresented her preference and instead submitted

R_1'
x
z
y

then $BC(x, R_1', R_2) = BC(y, R_1', R_2) = BC(z, R_1', R_2) = 4$, so the tie breaking rule would be applied and $x = f(R_1', R_2)$ would be selected. Since under her true preferences R_1, individual 1 strictly prefers x to y, she is strictly better-off after manipulating her submitted ranking to R_1'. Hence the SCF f gives individual 1 incentives to misrepresent her preferences when the true profile is $R = (R_1, R_2)$.

As the above example demonstrates, when the incentives to tell the truth are not there, we may never be able to recover the true preference profile by asking individuals what their preferences are. Hence a very desirable property of a SCF is that it never gives anybody an incentive to submit untruthful preferences.

Definition 8 A SCF f is **strategy-proof (or non-manipulable)** if for any $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$, $i \in N$, and $R_i' \in \mathcal{R}$:

$$f(R_1, \ldots, R_{i-1}, R_i, R_{i+1}, \ldots, R_n) \ R_i \ f(R_1, \ldots, R_{i-1}, R_i', R_{i+1}, \ldots, R_n).$$

Definition 9 Given a SCF f, the **image of** f is defined to be the subset of social alternatives that f chooses at some preference profile, that is:

$$\mathrm{Im}(f) = \{f(R) | R = (R_1, \ldots, R_n) \in \mathcal{R}^n\}.$$

Theorem 2 *(Gibbard-Sattherwaite Theorem) Suppose that f is a SCF with at least three elements in its image. If f is strategy-proof then it is dictatorial, i.e. there exists $i^* \in N$ such that for any $R = (R_1, \ldots, R_n) \in \mathcal{R}^n$, $f(R)$ is the top element in $\mathrm{Im}(f)$ with respect to R_{i^*}.*

Proof: Omitted. See again Reny (2001) for a proof. □

4 Possibility results

The impossibility theorems we have stated so far relied on at least three alternatives: Arrow's Impossibility Theorem required the existence of at least three elements, and the Gibbard- Satthertwaite Theorem required that the image of the SCF has at least three elements. One can check that these impossibility results are reversed when there are two alternatives. In the special case of $|X| = 2$, the majority rule (which coincides with Borda and pairwise majority) satisfies all the desirable properties so far introduced and is not dictatorial.

We next study an important class of social choice problems where there exist nice SWF's and SCF's. In this class of problems, the social alternatives are naturally ordered and the individuals' preferences satisfy a consistency property with respect to that order.

4.1 Single-Peaked Preferences

Let $X = [0, 1]$ represent the set of social alternatives. You can think of the interval [0,1] as the set of potential tax rates, as the set of political party positions (0 being extreme leftist, 1 being extreme rightist, and so on), or as the set of addresses in which supermarket will be built, in a linear city whose map looks like the interval [0,1].

In certain instances, including the examples given above, it is natural to think that each individual has an ideal point in the interval [0,1], and she becomes worse-off as the social outcome is further away from each side of his ideal point. Such preferences over the [0,1] are called single-peaked:

Definition 10 A preference R_i over $X = [0, 1]$ is **single-peaked** if it is rational and if there exist $x_i \in [0, 1]$ such that:

$$x_i \geq y > z \;\Rightarrow\; yP_iz.$$

$$z > y \geq x_i \;\Rightarrow\; yP_iz.$$

for any $y, z \in [0, 1]$. In this case x_i is called the **peak/ideal point** of R_i.

Definition 11 Let $\mathcal{R}^*$ denote the set of single-peaked preferences over $X = [0, 1]$.

Note that if R_i is single-peaked then it has a unique peak. Single-peakedness allows for indifference between an outcome to the left of the ideal point and another outcome to right of the ideal point.

We will assume for expositional simplicity of this section that the number of individuals n in the society is *odd*. The concepts of SWF and SCF can naturally be adapted to this framework when the preferences of the individuals are single-peaked.

Definition 12 Suppose that individuals have single-peaked preferences. A **SWF** is a function $f \colon \mathcal{R}^{*n} \to \mathcal{B}$. A **SCF** is a function $f \colon \mathcal{R}^{*n} \to X$.

Pairwise majority (PM) continues to be a well-defined SWF in this framework where for any preference profile $R = (R_1, \ldots, R_n) \in \mathcal{R}^{*n}$, the social preference $\succsim$ is given by:

$$x \succsim y \;\Leftrightarrow\; |\{i \in N | xR_iy\}| \geq |\{i \in N | yR_ix\}|$$

for any $x, y \in [0, 1]$.

PM always satisfies IIA, Unanimity, and completeness. We had seen an example with three alternatives (the Condorcet paradox) that PM may lead to intransitivities in the social ordering. Our first main result of this section tells that the PM is transitive if the individual preferences are single-peaked. But first a useful lemma.

Lemma 1 *Let* $m := \frac{n+1}{2}$, $R = (R_1, \dots, R_n) \in \mathcal{R}^{*n}$, *and let* $\succsim$ *denote the PM outcome with respect to the profile of single-peaked preferences* $R = (R_1, \dots, R_n)$*, Then for any* $x, y \in [0,1]$*:*

1. $x \succsim y \Rightarrow |\{i \in N | xR_iy\}| \geq m$,
2. $|\{i \in N | xP_iy\}| \geq m \Rightarrow x \succ y$.

Proof: Covered in class. □

Proposition 1 *When individuals have single-peaked preferences, PM is transitive.*

Proof: Covered in class. □

In contrast with Arrow's Impossibility Theorem, the above proposition yields a positive result saying that, if the domain of individual preferences does not include all possible orderings but only the single-peaked ones, then it is possible to find a non-dictatorial SWF that satisfies IIA, Unanimity, and Rationality, namely PM.

Definition 13 Given $\{x_1, \dots, x_n\} \subset [0,1]$, a median of $\{x_1, \dots, x_n\}$ is an $x^* \in \{x_1, \dots, x_n\}$ such that:

$$|\{i \in N | x_i \leq x^*\}| \geq n/2 \quad \& \quad |\{i \in N | x_i \geq x^*\}| \geq n/2.$$

Proposition 2 *For any* $\{x_1, \dots, x_n\} \subset [0,1]$*, a median exists and (since* n *is odd) it is unique. Furthermore if* x^* *is the median of* $\{x_1, \dots, x_n\}$*,* y^* *is the median of* $\{y_1, \dots, y_n\}$*, and* $x_i \leq y_i$ *for all* $i \in N$*, then* $x^* \leq y^*$*.*

Proof: I omitted the proofs of these mathematical facts. They are not very difficult, you should try to prove them if you wish. □

Definition 14 **The Median Voter Scheme** is the SCF that associates to each profile of single-peaked preferences $R = (R_1, \dots, R_n) \in \mathcal{R}^{*n}$, the median of the ideal points $\{x_1, \dots, x_n\}$, where x_i denotes the ideal point of R_i for each $i \in N$.

By Proposition 2, the Median Voter Scheme is a well-defined SCF. The next Proposition formally establishes the relationship between Pairwise Majority and the Median Voter Scheme. Namely, for each profile of single-peaked preferences, the outcome of the Median Voter Scheme is at the top of the social ranking determined by Pairwise Majority.

Proposition 3 *Let $R = (R_1, \dots, R_n) \in \mathcal{R}^{*n}$, let x_i denote the peak of R_i for each $i \in N$, let $\succsim$ be the social preference determined by pairwise majority, and let x^* be the median of the ideal points $\{x_1, \dots, x_n\}$. Then*

$$\forall y \in [0,1] \setminus \{x^*\} : \quad x^* \succ y.$$

Proof: Covered in class. $\square$

Finally the following is a positive result which shows that the Median Voter Scheme is strategy-proof when the domain of individual preferences is $\mathcal{R}^*$. It is in contrast with the Gibbard-Sattherwaite Theorem much in the same vein Proposition 1 is in contrast with Arrow's Impossibility Theorem.

Proposition 4 *When individuals have single-peaked preferences, the Median Voter Scheme is strategy-proof.*

Proof: Covered in class. $\square$

5 Optional Additional Reading

Lecture 10 of **Ru**, Chapter 21 of **MWG**: