**Computational Aspects of Game Theory** **Bertinoro Spring School 2011**

# Lecture 9: Social Choice Theory

*Lecturer: Bruno Codenotti*

The theory of social choice aims at aggregating the preferences of the individuals towards a single joint decision. We introduce some of the basic impossibility theorems, which point out the main difficulties of the field, and highlight some positive results obtained by making suitable restrictions to the general model.

## 9.1 Introduction

We consider the problem of individual decision making in the following abstract setting. The individual must *choose* from a set $X$ of mutually exclusive alternatives. The primitive characteristic of the individual is expressed by her *preference relation* $\succeq$ over the set $X$. For $x, y \in X$, we read $x \succeq y$ as "x is at least as good as y". From $\succeq$ we can derive two other natural relations on $X$, the *strict* preference relation $\succ$ and the *indifference* relation $\sim$.

The strict preference relation $\succ$ is defined from $\succeq$ by: $x \succ y$ if and only if $x \succeq y$ and not $y \succeq x$.

The indifference relation $\sim$ is defined from $\succeq$ by: $x \succ y$ if and only if $x \succeq y$ and $y \succeq x$.

**Definition 9.1 (Rational Preference Relation)** *The preference relation $\succeq$ is rational if the following two properties hold:*

1. *For all $x, y \in X$ either $x \succeq y$ or $y \succeq x$, or both. (Completeness, or Universal Domain)*
2. *For all $x, y, z \in X$ if $x \succeq y$ and $y \succeq z$, then $x \succeq z$. (Transitivity)*

Consider now a *society* of $N$ individuals, and a set $X$ of mutually exclusive alternatives. Every individual $i$ has a rational preference relation $\succeq_i$ on $X$. We want to study the extent to which these individual preferences can be aggregated into a *social preference* or into a social decision in a "reasonable" way.

**Definition 9.2 (Social Welfare Function)** *A social welfare function (SWF) is a function that associates with every $N$-tuple of rational preferences (called* preference profile*) $(\succeq_1, \succeq_2, \ldots, \succeq_N)$ a rational preference $\succeq$, which is called the social preference.*

## 9.2 Impossibility Theorems

We now impose some conditions on social welfare functions.

1. A SWF respects **unanimity** if society strictly prefers $x \in X$ to $y \in X$ whenever all the individuals strictly prefer $x \in X$ to $y \in X$. Formally, if $x \succ_i y$ for all $i$, then $x \succ y$.
2. A SWF satisfies **independence of irrelevant alternatives** if the social relative ranking of two alternatives $x, y \in X$ depends only on their relative rankings by each individual, and not on the ranking of the other choices.

3. A SWF is a **dictatorship** by individual $i$ if for all $x, y \in X$ we have that $x \succ_i y$ implies $x \succ y$.

**Theorem 9.3 (Arrow)** *Let $N \geq 2$ and $|X| \geq 3$. Any SWF that satisfies rationality, unanimity, and independence of irrelevant alternatives is a dictatorship.*

**Proof:** The proof is the composition of three claims.

Let $y \in X$ be chosen arbitrarily. Consider a preference profile where every individual puts $y$ either at the top or at the bottom of her rankings. We claim that the social preference must also put $y$ either at the very top or at the very bottom. Assume by contradiction that for such a profile there are $x \neq y \neq z \in X$ such that $x \succeq y$ and $y \succeq z$. Independence of irrelevant alternatives implies that this will continue to hold even if all the individuals moved $z$ above $x$, i.e., they all strictly prefer $z$ to $x$. In fact these changes in the individual rankings do not affect the relative rankings of $x$ vs $y$ or $y$ vs $z$, since $y$ occupies an extreme position. But then we reach a contradiction, because $x \succeq z$ (by transitivity) and $z \succ x$ (by unanimity).

We claim that there is an *extremely pivotal* individual $i(y)$, who can move $y$ from the bottom to the top of the social rankings, by changing her vote. Assume each individual puts $y$ at the bottom of her ranking. By unanimity, the same must happen for the social ranking. Now let the individuals from 1 to $N$ successively move $y$ from the bottom to the top (without changing the other relative rankings), and let $i(y)$ be the first individual whose change causes the social ranking of $y$ to change. It must change because at the end of the process, by the first claim, $y$ must be at the top.

Consider two preference profiles: profile A, given by the individual rankings *just before* $i(y)$ moves $y$ in the social rankings, and profile B, given by the individual rankings *just after* $i(y)$ moves $y$ in the social rankings. By our previous claim, we have that the social preference corresponding to profile B must put $y$ at the top.

Now we claim that $i(y)$ is a dictator over any pair $x, z \in X$ such that $x \neq y$ and $z \neq y$. We choose an element from the pair $x, z$, say $x$. From profile B, we construct profile C by moving $x$ above $y$ in $i(y)$'s rankings, and by letting all the other individuals rearrange arbitrarily their relative rankings of $x$ and $z$ while leaving $y$ in the extreme position. By independence of irrelevant alternatives, the social preference for profile C must necessarily put $x$ above $y$ (since all individual relative rankings of $x$ and $y$ are as in profile A where $i(y)$ put $y$ at the bottom) and $y$ above $z$ (since all individual relative rankings of $y$ and $z$ are as in profile B where $i(y)$ put $y$ at the top). By transitivity, we must have $x \succ z$. By independence of irrelevant alternatives, we see that the social preference over $x, z \in X$ must agree with $i(y)$'s preference.

We conclude the proof by showing that $i(y)$ must be a dictator over every pair $x, y$. Take $z$ different from both $x$ and $y$, and put $z$ at the bottom in the construction of our second claim. By our third claim there is an individual $i(z)$ who is a dictator on each pair not involving $z$, in particular over the pair $x, y$. But individual $i(y)$ can affect the social ranking of $x$ vs $y$, e.g., at profiles A and B, and thus this dictator $i(z)$ must be $i(y)$. ■

For simplicity, in the rest of this section, we assume that the individual preferences are asymmetric binary relations, i.e., we do not allow for indifferences. We represent preferences by utility functions $u \in U$, so that $x \succeq y$ if and only if $u(x) \geq u(y)$. Preference profiles are elements of the set $U^N$. A preference profile $u = (u_1, u_2, \ldots, u_N)$ can also be denoted by $(u_i, u_{-i})$, $i \in \{1, 2, \ldots, N\}$, where $u_{-i} = (u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_N)$.

**Definition 9.4 (Social Choice Function)** *A social choice function (SCF) is a function that associates with every preference profile $(u_1, u_2, \ldots, u_N)$ an alternative $x \in X$, which is called the social choice. A social choice function is sometimes called a* voting rule.

**Definition 9.5 (Incentive Compatibility)** *An SCF $f$ can be* strategically manipulated *if there are an* *individual $i$, preferences $u_i' \in U$, and a preference profile $u \in U^N$ such that $u_i(f(u_i', u_{-i})) > u_i(f(u))$. An SCF is incentive compatible (or* strategy-proof*) if it cannot be strategically manipulated.*

We say that an SCF $f$ is *dictatorial* if there exists $i$ such that $u_i(f(u)) \geq u_i(x)$, for all $x \in X$ and for all $u \in U^N$.

We say that an SCF $f$ is *onto* if for all $x \in X$ there exists $u \in U^N$ such that $f(u) = x$.

**Theorem 9.6 (Gibbard-Satterthwaite)** *Let $|X| \geq 3$. An incentive compatible SCF function that is onto must be dictatorial.*

## 9.3 Possibility Results

In order to cope with the difficulties expressed by the Impossibility Theorems, one has to either restrict the domain of the individual preferences or radically change the model. We now show a scenario where the domain of the preferences is restricted, and a non-dictatorial social welfare function can be defined.

**Definition 9.7 (Linear Order)** *A binary relation $\geq$ on $X$ is a linear order on $X$ if it is reflexive, transitive and total.*

For example, for $X \subset R$, the ordinary $\geq$ relation (greater than or equal to) is a linear order on $X$.

**Definition 9.8 (Single-Peaked Preferences)** *A preference relation $\succeq_i$ is single-peaked with respect to the linear order $\geq$ on $X$ if there exists $x \in X$ such that $\succeq_i$ is increasing with respect to $\geq$ on $\{y \in X | x \geq y\}$ and decreasing on $\{y \in X | y \geq x\}$. This means that if $x \geq z > y$ then $z \succ_i y$, and if $y > z \geq x$ then $z \succ_i y$.*

**Definition 9.9 (Majority Decision Rule)** *A SWF implements the majority decision rule if the social relative ranking of any two alternatives $x, y \in X$ puts $x$ above $y$ if and only if the majority of individuals rank $x$ above $y$.*

**Theorem 9.10 (Black)** *Assume the number of individuals is odd. If individual preferences are single-peaked, then the majority decision rule is a non-dictatorial SWF that satisfies independence of irrelevant alternatives.*

## Bibliographic notes

A general treatment of Social Choice Theory can be found in Chapter 21 of [5].

The Arrow's Impossibility Theorem has been proved in [1] (see also the book by Arrow [2]). We have followed the first of the three simple proofs presented by Geanakoplos in [7].

The Gibbard-Satterthwaite's Theorem has been proved in [8, 9]. A simple proof can be found in [10]. A popular presentation of the manipulability of voting systems is [11].

Black's possibility result for single peaked preferences has been shown in [3].

## References

[1] K. Arrow, A Difficulty in the Concept of Social Welfare, Journal of Political Economy 58(4) (August, 1950), pp. 328346.

[2] K. Arrow, Social Choice and Individual Values. New York: John Wiley and Sons (1951).

[3] D. Black, On the Rationale of Group Decision-making, Journal of Political Economy, 56:2334, 1948.

[4] D. Black. Theory of Committees and Elections. Cambridge University Press, Cambridge, 1958.

[5] A. Mas-Colell, M.D. Whinston, J.R. Green, Microeconomic Theory, Oxford University Press (1995).

[6] N. de Condorcet, Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Paris (1785).

[7] J. Geanakoplos, Three Brief Proofs of Arrow's Impossibility Theorem. Cowles Foundation discussion paper No.1123RRR. Cowles Foundation for Research in Economics, Yale University, New Haven, Connecticut (2001).

[8] A. Gibbard, Manipulation of voting schemes: a general result, Econometrica, Vol. 41, No. 4 (1973), pp. 587601.

[9] M.A. Satterthwaite, Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions, Journal of Economic Theory 10 (April 1975), 187217.

[10] L.G. Svensson, The Proof of the Gibbard-Satterthwaite Theorem revisited, Manuscript (1999).

[11] A. Taylor, The manipulability of voting systems, The American Mathematical Monthly, April 2002.