CS294-1 Algorithmic Aspects of Game Theory	**Spring 2001**

Lecture 4: February 6, 2001

Lecturer: Christos Papadimitriou *Scribes: Eric Kuo, Tracy Riggs, Alice Zheng*

4.1 Overview

In this lecture, we discuss social choice theory and introduce Arrow's Impossibility Theorem. We then move on to the primary focus of this course, mechanism design.

4.2 Social Choice Theory

Social choice theory deals with the aggregation of individual preferences into a collective prefence, as in voting. Essentially, we attempt to answer the question: given a set of individual preferences, how do we make the best choice globally?

Definition 4.1 (Preference Relation). *Suppose there are n agents and a set of social choices $C = \{c_1, c_2, \ldots, c_m\}$. The preference relation $\succeq_i$ over C is defined as the ordering of set C according to the preference of agent i.*

The preference relation may be either partially or totally ordered. We shall often assume it is a total order (i.e., no ties), in which case we shall denote it $\succ_i$, because this assumption avoids uninteresting complications in proofs.

Definition 4.2 (Social Welfare Functional). *A social welfare functional is a function $\boldsymbol{\succeq} (\succeq_1, \ldots, \succeq_n)$ that takes as input the individual preferences, and outputs a "global preference".*

The following are examples of social welfare functionals.

- Dictatorship: $\boldsymbol{\succeq} (\succeq_1, \ldots, \succeq_n) = \succeq_1$
- Democracy: When $|C| = 2$, $c_1 \boldsymbol{\succeq} (\succeq_1, \ldots, \succeq_n)\, c_2$ iff $|\{i : c_1 \succeq_i c_2\}| \geq |\{i : c_2 \succeq_i c_1\}|$.

  Theorem 4.3. *When $|C| = 2$, majority voting is the only social welfare functional that is monotone (an agent's preference change from c_1 to c_2 does not cause the global preference to change from c_2 to c_1) and symmetric with respect to agents (permuting the agents does not change the result) and outcomes (reversing the role of the two choices in the input results in the same reversal in the output).*

  When $|C| \geq 3$, however, the situation becomes more difficult. Some common ideas are these:

- *Borda's rule*: Order the global preference by increasing $\sum_{i=1}^{n} rank_i(c_j)$. The problem with this functional is that whether $x \succeq y$ depends on z, otherwise known as a violation of *independence from irrelevant alternatives.*
- *Kemeny's rule*: Choose $\boldsymbol{\succeq}$ that minimizes $\sum_{i=1}^{n} dist(\succeq_i, \boldsymbol{\succeq})$, where $dist(\succeq, \succeq')$ is the "bubble sort distance" or how many changes are required to get from $\succeq$ to $\succeq'$. Kemeny's rule also violates independence from irrelevant alternatives, and furthermore is NP-hard to compute. (See discussion in the paper by Dwork et al. in the course readings.)

- Majority voting on pairs: Another way to derive a global preference is by examining the majority vote between each pair of social choices. However, this may lead to the *Condercet paradox*, which occurs in the following situation: agent 1 chooses $x \succeq y \succeq z$, agent 2 chooses $y \succeq z \succeq x$, and agent 3 chooses $z \succeq x \succeq y$. Majority voting over pairs gives a cyclical preference: x is preferred over y, y over z, and z over x (each by two votes to one).

4.3 Arrow's Impossibility Theorem

In fact, this difficulty is inherent:

Theorem 4.4 (Arrow's Impossibility Theorem). *Let $|C| \geq 3$. Then there is no social welfare functional $\succ$ that satisfies all of the following properties:*

1. *$\succ$ is independent from irrelevant alternatives.*
2. *$\succ$ respects unanimity.*
3. *$\succ$ is non-dictatorial.*

Proof. Notice that in this theorem, for simplicity, we assume all orders are total. For the proof we shall assume that $\succ$ is independent from irrelevant alternatives and that it respects unanimity. We show that $\succ$ must be dictatorial.

We say that a set $S \subseteq \{1, 2, \ldots, n\}$ is *decisive for X over Y* if, whenever $X \succ_i Y$ for each $i \in S$ and $X \prec_i Y$ for each $i \notin S$, then $X \succ Y$. In other words, if S is decisive for X over Y, each agent in S prefers X over Y, and each agent outside S prefers Y over X, then the functional must choose X over Y.

In order to prove that S is decisive for X over Y, first we create one situation in which $X \succ_i Y$ for each $i \in S$, $X \prec_i Y$ for each $i \notin S$, and $X \succ Y$. Then from the property of independence from irrelevant alternatives, we can conclude that S is decisive for X over Y since varying the agents' preferences for all other remaining candidates will not change the fact that $X \succ Y$.

Lemma 4.5. *If S is decisive for X over Y, then for any other candidate Z, S is decisive for X over Z, and S is decisive for Z over Y.*

Proof. Let S be decisive for X over Y. Suppose that $X \succ_i Y \succ_i Z$ for each agent $i \in S$, and $Y \succ_i Z \succ_i X$ for each agent $i \notin S$. Since S is decisive for X over Y, we have $X \succ Y$. Because $Y \succ_i Z$ for every agent i (in or not in S), it follows that $Y \succ Z$ since $\succ$ respects unanimity. Then by transitivity, $X \succ Z$. But we also have $X \succ_i Z$ for all $i \in S$, and $X \prec_i Z$ for all $i \notin S$. Then since $\succ$ is independent from irrelevant alternatives (in this case Y), we conclude that S is decisive for X over Z.

To show that S is decisive for Z over Y, we suppose that $Z \succ_i X \succ_i Y$ if $i \in S$, and $Y \succ_i Z \succ_i X$ otherwise. Then $X \succ Y$ since S is decisive for X over Y; $Z \succ X$ from respecting unanimity; and $Z \succ Y$ from transitivity. Thus S is decisive for Z over Y. $\square$

Given that S is decisive for X over Y, we deduce that S is decisive for X over Z and Z over Y. But then we can reapply the lemma with decision for Z over Y as the hypothesis and conclude that S is decisive for Z over X, which in turn implies that S is decisive for Y over X, which finally implies that S is decisive for Y over Z. Thus we can state our next result:

Lemma 4.6. *If S is decisive for X over Y, then for each pair of candidates U and V, S is decisive for U over V.*

Thus, if a set S is decisive for one choice over another, then we say that S is *decisive*. Let $\mathcal{D}$ be the set of decisive subsets of $\{1, 2, \ldots, n\}$. We prove that for every set S, either S or its complement (but not both) must be decisive. That is,

$$|\{S, \bar{S}\} \cap \mathcal{D}| = 1.$$

Suppose that for each $i \in S$, $X \succ_i Y$, and for each $i \in \bar{S}$, $X \prec_i Y$. We ask whether $X \succ Y$. If so, then clearly S is decisive. If not, $\bar{S}$ is decisive for Y over X and is thus decisive.

Next we show that decisiveness is closed under intersection.

Lemma 4.7. *If $S, T \in \mathcal{D}$, then $S \cap T \in \mathcal{D}$.*

Proof. Let $S = \{i : Z \succ_i Y \succ_i X\} \cup \{i : X \succ_i Z \succ_i Y\}$ and $T = \{i : Y \succ_i X \succ_i Z\} \cup \{i : X \succ_i Z \succ_i Y\}$. And for the remaining $i \notin S \cup T$, $Y \succ_i Z \succ_i X$. Since S is decisive, we can say $Z \succ Y$. Since T is decisive, $X \succ Z$. Then by transitivity, $X \succ Y$. But note that if $X \succ_i Y$, then $i \in S \cap T$. Hence $S \cap T$ is decisive; $S \cap T \in \mathcal{D}$. □

Finally, we show that if $S = S_1 \cup S_2 \in \mathcal{D}$ where S_1 and S_2 are disjoint and nonempty, then either S_1 or $S_2 \in \mathcal{D}$. For suppose that S_1 and $S_2 \notin \mathcal{D}$. Then both complements $\bar{S_1}$ and $\bar{S_2}$ are decisive. Since decisiveness is closed under intersection, $\bar{S_1} \cap \bar{S_2} = \bar{S} \in \mathcal{D}$. But we assumed that S is decisive. Thus we have a contradiction.

Clearly, the whole set $\{1, 2, \ldots, n\}$ of agents is decisive. We showed that we can split any decisive set into two subsets, at least one of which is decisive. Thus by repeatedly splitting a decisive subset, we can find a smaller decisive subset until only one agent remains. That single agent becomes dictator. □

Arrow's Impossibility Theorem can also be expressed in an alternative form:

Theorem 4.8 (Arrow's Impossibility Theorem, Second Form). *Let $F(\succ_1, \succ_2, \ldots, \succ_n) \in C$ be a social choice function that simply chooses a winner based on the agents' preferences. Then if F is monotonic, respects unanimity, and nonbinary, then F must also be dictatorial.*

The proof is by a reduction to the previous form: Assume that such a function exists, and extend it to a function*al*, which chooses not just a winner, but a global order. This is done by constructing $\succ$ from F as follows: To determine whether $a \succ (\succ_1, \ldots, \succ_n) b$, modify the $\succ_i$'s so that a and b are at the top (with their order, and also all that of all other elements, unchanged) and then see which of the two is a winner under F. It turns out that the $\succ$ constructed this way is a total order, and it satisfies the conditions of the first form of Arrow's Impossibility Theorem.

4.4 Mechanism Design

Mechanism design is the real "meat" of this course. In a way, it is the most real-world oriented topic in game theory. The problem involves designing a game where selfish agents will ultimately play according to the desire of the game designer. In this section, we formally introduce the problem, the next two lectures will concentrate on solutions to the problem.

4.4.1 Definition

Suppose we are given an **environment** (n, C, U), where $n \in \mathbb{Z}$ denotes the number of agents/players, C denotes the set of possible *outcomes*, and U is a set of tuples of the form $(u_1, \ldots u_n), u_i : C \to \mathbb{R}^+$, the set of all possible combinations of agent utilities. Suppose there is also a **choice rule** $f : U \to 2^C$, which maps agent utilities to sets of outcomes (this captures the mechanism designer's "desire" alluded to above). be the set of strategies for each agent, The goal of **mechanism design** is to design a game with sets of strategies $X_1, \ldots X_n$ and an *outcome function* $g : X_1 \times \cdots \times X_n \to C$, such that the utility of player I in the game is $u_i(x_1, \ldots x_n) = u_i(g(x_1, \ldots x_n))$, and such that the resulting game G satisfies the following two conditions:

1. For each player, there is a *weak dominating strategy* $\mathrm{Dom} \neq \emptyset$.
2. If $(x_1, \ldots x_n) \in \mathrm{Dom}(G)$, then $g(x_1, \ldots x_n) \in f(u_1, \ldots u_n)$.