PS 12 NOTES: WEEK 5

TA: WADE HANN-CARUTHERS

1. Social preferences

1.1. **Preferences.** Given a set A of alternatives, a *(strict) preference* $\succ$ is a linear order on A. When A is finite, we just think of a preference as a ranking over the alternatives. For example, if $A = \{banana, orange, chocolate\}$, then we would write

$$chocolate \succ banana \succ orange$$

to express the preference of an individual who likes chocolate more than bananas and bananas more than oranges. Notice that since $\succ$ is a linear order, this also implies that the individual prefers chocolate to oranges: $chocolate \succ orange$.

1.2. **Social welfare functions.** A *social welfare function* is a function that maps each n-tuple of preferences $(\succ_1, \ldots, \succ_n)$ to a preference $\succ$, called the *social preference*. Social welfare functions, also called *preference aggregation rules*, are intended to provide a societal ranking over alternatives on the basis of the individuals' rankings.

It turns out that useful social welfare functions do not exist in general. A social welfare function satisfies *Pareto efficiency* if for any two alternatives $x, y \in A$, $x \succ y$ whenever $x \succ_i y$ for all $i = 1, \ldots, n$. In other words, a social welfare function satisfies Pareto efficiency if it ranks one alternative higher than another whenever all individuals do–so if everyone likes chocolate more than oranges, the social preference should also rank chocolate above oranges. A social welfare function satisfies *Independence of Irrelevant Alternatives (IIA)* if for any two alternatives $x, y \in A$, the relative ranking of x to y in the social preference depends only on the individuals' relative rankings of x to y. For example, suppose there are three individuals $1, 2, 3$, with preferences

$$banana \succ_1 orange \succ_1 chocolate$$
$$orange \succ_2 banana \succ_2 chocolate$$
$$chocolate \succ_3 banana \succ_3 orange$$

and suppose the social welfare function assigns this preference profile a preference $\succ$ such that $banana \succ orange$. Then under a preference profile where all agents feel the same about orange and banana, like this one

$$chocolate \succ_1 banana \succ_1 orange$$
$$orange \succ_2 chocolate \succ_2 banana$$
$$banana \succ_3 orange \succ_3 chocolate$$

the social welfare function must also assign this preference profile a preference $\succ$ such that $banana \succ orange$. Finally, a social welfare function is *dictatorial* if there is some individual i such that the social preference is always equal to i's preference.

Theorem 1 (Arrow's impossibility theorem)**.** *If there are at least three alternatives, then any social welfare function that satisfies Pareto efficiency and Independence of Irrelevant Alternatives is dictatorial.*

1.3. **Condorcet paradox.** There is (what would seem to be) a natural candidate for a social welfare function. We could say that for any two alternatives $x, y \in A$, $x \succ_{maj} y$ if and only if $x \succ_i y$ for a majority of individuals i. This is just an extension of how we normally decide between two alternatives—so-called majority rule. So what's wrong with this rule? Suppose there are three individuals $1, 2, 3$, three alternatives $A = \{a, b, c\}$, and the preferences are

$$a \succ_1 b \succ_1 c$$
$$b \succ_1 c \succ_1 a$$
$$c \succ_1 a \succ_1 b$$

Then what would the resulting social preference be? Since 1 and 3 prefer a to b, $a \succ_{maj} b$. Since 1 and 2 prefer b to c, $b \succ_{maj} c$. So our social preference should be $a \succ_{maj} b \succ_{maj} c$. But since 2 and 3 prefer c to a, $c \succ_{maj} a$! The whole point of the social preference is that it is intended to provide a ranking over the alternatives, but if we use this rule, it seems like c should be ranked below *and* above a, which is nonsense. The fact that it is possible for the majority rule relation $\succ_{maj}$ to have cycles, like $a \succ_{maj} b \succ_{maj} c \succ_{maj} a$, is called the *Condorcet paradox*.

1.4. **Condorcet winner.** In practice, the Condorcet paradox may not always be an issue. If we are only concerned with choosing a best alternative, we don't need to be so concerned if there are weird issues with aggregating preferences over all of the alternatives. An alternative is called a *Condorcet winner* if it is preferred to any other alternative by a majority of individuals–that is, x is a Condorcet winner if for every other alternative y, $x \succ_{maj} y$ is true and $y \succ_{maj} x$ is not true.

2. Agenda voting

A *voting agenda* is an ordering on the alternatives, which we interpret as follows. The first alternative is taken as the tentative outcome. At each step, all individuals vote between the current tentative outcome and the next alternative on the agenda, and the winner (the alternative that gets a majority of the votes) becomes the tentative outcome. Whichever alternative is the tentative outcome after the final vote is then the final outcome.

For example, suppose the set of alternatives is $A = \{1, 2, 3, 4\}$ and the voting agenda is $(4, 2, 1, 3)$. First, there is a vote between 4 and 2. Imagine 2 wins this vote. Then next, there is a vote between 2 and 1. Imagine 1 wins this vote. Then finally, there is a vote between 1 and 3. Imagine 3 wins this vote. Then the final outcome is 3.

One reason agenda voting is important is because it cleanly illustrates the sensitivity of the final outcome to the details of the voting procedure. For example, suppose the individuals, alternatives, and preferences are as in the Condorcet paradox example, and assume that individuals vote for their preferred alternative whenever faced with a vote. If the voting agenda is (a, b, c), then $a \succ_{maj} b$, so a wins the vote between a and b, and $c \succ_{maj} a$, so c will win the subsequent vote between a and c; hence, the final outcome will be c. However, if the voting agenda is (a, c, b), $c \succ_{maj} a$, so c will win the vote between a and c, and $b \succ_{maj} c$, so b will win the subsequent vote between c and b; hence, the final outcome will be b.

2.1. Sophisticated voting.

Another reason agenda voting is important is because it illustrates the sensitivity of the final outcome to how "sophisticated" the voting behavior is. When individuals always simply vote for their preferred alternative when faced with a vote, we call their behavior *naive voting*. Now, let's reconsider the (a, b, c) agenda, but this time assume that the voters all know each other's preferences and are forward-looking. Then voters recognize that if a wins the first vote, the result is effectively the voting agenda (a, c), and if b wins the first vote, the result is effectively the voting agenda (b, c). Now, c would be the final outcome in the first case and b would be the final outcome in the second case. Thus, a vote for a in the first round is effectively a vote for c, and a vote for b in the first round is effectively a vote for b. Since a majority of voters prefer b to c, we should expect a majority of voters to vote for b in the first round, and for the final outcome to be b! When individuals use their knowledge of the other individuals' preferences and are forward-looking in this way, we call their behavior *sophisticated voting*.

Understanding what happens under sophisticated voting for longer agendas is slightly more involved, but the ideas are the same. Consider the following procedure on the agenda $(x_1, \dots, x_k)$.

- Circle x_k.
- For j starting from $k - 1$ and going down to 1,
  - if x_j is majority-preferred to all alternatives that have already been circled, circle x_j
  - otherwise, do not circle x_j

Proposition 2. *The last alternative circled under the above procedure will be the final outcome under sophisticated voting.*

The intuition is straightforward and uses the same observation we made when analyzing the (a, b, c) agenda. The first vote is really a vote between two shorter agendas, and hence is really a vote between the outcomes under those two shorter agendas. If we assume this procedure works for determining the outcomes under those two agendas, a little bit of casework shows that it must also work for determining the outcome under the original agenda, and so the procedure works for determining the outcome for any agenda by induction.

2.2. **Condorcet winners and agenda voting.** Given an agenda, it is clear that if there is an alternative x on the agenda that is a Condorcet winner among the alternatives on the agenda, it will win under naive voting: there will eventually be a vote involving x, x will win and become the tentative outcome, and x will remain the tentative outcome since it will win all of the subsequent votes, ultimately becoming the final outcome. What about under sophisticated voting?

Proposition 3. *If x is a Condorcet winner among alternatives on the agenda, then x will be the final outcome under sophisticated voting.*

To see why, consider what happens when we use the procedure described above. When we get to x, it will be preferred by a majority of individuals to any of the alternatives that have already been circled, so it will be circled. For each remaining alternative y, it is impossible for y to be circled, since it is not preferred by a majority of individuals to x. So x will be the last circled alternative!

2.3. **Covering.** Our procedure for finding the final outcome under sophisticated voting also provides another insight that will be useful later on. For $x, y \in A$, we say that x *covers* y if $x \succ_{maj} y$ and for any $z \in A$ such that $y \succ_{maj} z$, we also have $x \succ_{maj} z$. When this occurs, we also say that y is *covered* (by x). Let's consider what happens when we use the procedure described above. If y is on the agenda before x, then either x will be circled, in which case y will not be circled, or x will not be circled because z is already circled for some $z \succ_{maj} x$, in which case y will also not be circled since $z \succ_{maj} y$. If y is on the agenda after x, then either y will not be circled, or y will be circled and x will be circled, or y will be circled and something between y and x will be circled. In any of these cases, y will not be the last alternative circled.

Proposition 4. *If y is covered (within the alternatives on the agenda), then y is not the final outcome of sophisticated voting.*

California Institute of Technology