Arrow's Impossibility Theorem

Lecture 12

Lecture Overview

Social Choice

Definition (Social choice function)

Assume a set of agents $N = \{1, 2, \ldots, n\}$, and a set of outcomes (or alternatives, or candidates) O. Let L_{-} be the set of non-strict total orders on O. A social choice function (over N and O) is a function $C : L_{-}^{n} \mapsto O$.

Definition (Social welfare function)

Let N, O, L_{-} be as above. A social welfare function (over N and O) is a function $W : L_{-}^{n} \mapsto L_{-}$.

Some Voting Schemes

- Plurality
  - pick the outcome which is preferred by the most people
- Plurality with elimination ("instant runoff")
  - everyone selects their favorite outcome
  - the outcome with the fewest votes is eliminated
  - repeat until one outcome remains
- Borda
  - assign each outcome a number.
  - The most preferred outcome gets a score of $n-1$, the next most preferred gets $n-2$, down to the n^{th} outcome which gets 0.
  - Then sum the numbers for each outcome, and choose the one that has the highest score
- Pairwise elimination
  - in advance, decide a schedule for the order in which pairs will be compared.
  - given two outcomes, have everyone determine the one that they prefer

Condorcet Condition

- If there is a candidate who is preferred to every other candidate in pairwise runoffs, that candidate should be the winner
- While the Condorcet condition is considered an important property for a voting system to satisfy, there is not always a Condorcet winner
- sometimes, there's a cycle where A defeats B, B defeats C, and C defeats A in their pairwise runoffs

Notation

- N is the set of agents
- O is a finite set of outcomes with $|O| \geq 3$
- L is the set of all possible strict preference orderings over O.
  - for ease of exposition we switch to strict orderings
  - we will end up showing that desirable SWFs cannot be found *even if* preferences are restricted to strict orderings
- $[\succ]$ is an element of the set L^n (a preference ordering for every agent; the input to our social welfare function)
- $\succ_W$ is the preference ordering selected by the social welfare function W.
  - When the input to W is ambiguous we write it in the subscript; thus, the social order selected by W given the input $[\succ']$ is denoted as $\succ_{W([\succ'])}$.

Pareto Efficiency

Definition (Pareto Efficiency (PE))

W is Pareto efficient if for any $o_1, o_2 \in O$, $\forall i\, o_1 \succ_i o_2$ implies that $o_1 \succ_W o_2$.

- when all agents agree on the ordering of two outcomes, the social welfare function must select that ordering.

Independence of Irrelevant Alternatives

Definition (Independence of Irrelevant Alternatives (IIA))

W is independent of irrelevant alternatives if, for any $o_1, o_2 \in O$ and any two preference profiles $[\succ'], [\succ''] \in L^n$, $\forall i\,(o_1 \succ'_i o_2$ if and only if $o_1 \succ''_i o_2)$ implies that $(o_1 \succ_{W([\succ'])} o_2$ if and only if $o_1 \succ_{W([\succ''])} o_2)$.

- the selected ordering between two outcomes should depend only on the relative orderings they are given by the agents.

Nondictatorship

Definition (Non-dictatorship)

W does not have a dictator if $\neg\exists i\,\forall o_1, o_2 (o_1 \succ_i o_2 \Rightarrow o_1 \succ_W o_2)$.

- there does not exist a single agent whose preferences always determine the social ordering.
- We say that W is dictatorial if it fails to satisfy this property.

Lecture Overview

Arrow's Theorem

Theorem (Arrow, 1951)

Any social welfare function W that is Pareto efficient and independent of irrelevant alternatives is dictatorial.

We will assume that W is both PE and IIA, and show that W must be dictatorial. Our assumption that $|O| \geq 3$ is necessary for this proof. The argument proceeds in four steps.

Arrow's Theorem, Step 1

Step 1: If every voter puts an outcome b at either the very top or the very bottom of his preference list, b must be at either the very top or very bottom of $\succ_W$ as well.

Consider an arbitrary preference profile $[\succ]$ in which every voter ranks some $b \in O$ at either the very bottom or very top, and assume for contradiction that the above claim is not true. Then, there must exist some pair of distinct outcomes $a, c \in O$ for which $a \succ_W b$ and $b \succ_W c$.

Arrow's Theorem, Step 1

Step 1: If every voter puts an outcome b at either the very top or the very bottom of his preference list, b must be at either the very top or very bottom of $\succ_W$ as well.

Now let's modify $[\succ]$ so that every voter moves c just above a in his preference ranking, and otherwise leaves the ranking unchanged; let's call this new preference profile $[\succ']$. We know from IIA that for $a \succ_W b$ or $b \succ_W c$ to change, the pairwise relationship between a and b and/or the pairwise relationship between b and c would have to change. However, since b occupies an extremal position for all voters, c can be moved above a without changing either of these pairwise relationships. Thus in profile $[\succ']$ it is also the case that $a \succ_W b$ and $b \succ_W c$. From this fact and from transitivity, we have that $a \succ_W c$. However, in $[\succ']$ every voter ranks c above a and so PE requires that $c \succ_W a$. We have a contradiction.

Arrow's Theorem, Step 2

Step 2: There is some voter n^* who is extremely pivotal in the sense that by changing his vote at some profile, he can move a given outcome b from the bottom of the social ranking to the top.

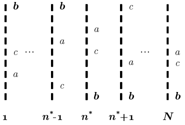
Consider a preference profile $[\succ]$ in which every voter ranks b last, and in which preferences are otherwise arbitrary. By PE, W must also rank b last. Now let voters from 1 to n successively modify $[\succ]$ by moving b from the bottom of their rankings to the top, preserving all other relative rankings. Denote as n^* the first voter whose change causes the social ranking of b to change. There clearly must be some such voter: when the voter n moves b to the top of his ranking, PE will require that b be ranked at the top of the social ranking.

Arrow's Theorem, Step 2

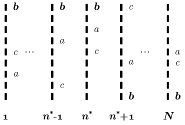
Step 2: There is some voter n^* who is extremely pivotal in the sense that by changing his vote at some profile, he can move a given outcome b from the bottom of the social ranking to the top.

Denote by $[\succ^1]$ the preference profile just before n^* moves b, and denote by $[\succ^2]$ the preference profile just after n^* has moved b to the top of his ranking. In $[\succ^1]$, b is at the bottom in $\succ_W$. In $[\succ^2]$, b has changed its position in $\succ_W$, and every voter ranks b at either the top or the bottom. By the argument from Step 1, in $[\succ^2]$ b must be ranked at the top of $\succ_W$.

Profile $[\succ^1]$:



Profile $[\succ^2]$:

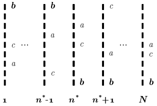


Arrow's Theorem, Step 3

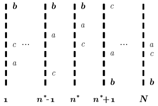
Step 3: n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.

We begin by choosing one element from the pair ac; without loss of generality, let's choose a. We'll construct a new preference profile $[\succ^3]$ from $[\succ^2]$ by making two changes. First, we move a to the top of n^*'s preference ordering, leaving it otherwise unchanged; thus $a \succ_{n^*} b \succ_{n^*} c$. Second, we arbitrarily rearrange the relative rankings of a and c for all voters other than n^*, while leaving b in its extremal position.

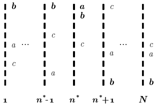
Profile $[\succ^1]$:



Profile $[\succ^2]$:



Profile $[\succ^3]$:

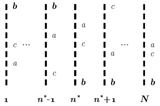


Arrow's Theorem, Step 3

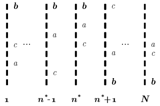
Step 3: n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.

In $[\succ^1]$ we had $a \succ_W b$, as b was at the very bottom of $\succ_W$. When we compare $[\succ^1]$ to $[\succ^3]$, relative rankings between a and b are the same for all voters. Thus, by IIA, we must have $a \succ_W b$ in $[\succ^3]$ as well. In $[\succ^2]$ we had $b \succ_W c$, as b was at the very top of $\succ_W$. Relative rankings between b and c are the same in $[\succ^2]$ and $[\succ^3]$. Thus in $[\succ^3]$, $b \succ_W c$. Using the two above facts about $[\succ^3]$ and transitivity, we can conclude that $a \succ_W c$ in $[\succ^3]$.

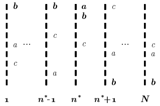
Profile $[\succ^1]$:



Profile $[\succ^2]$:



Profile $[\succ^3]$:

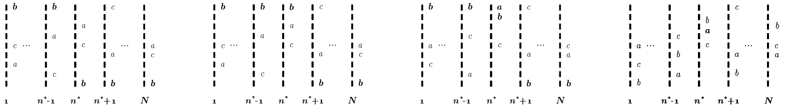


Arrow's Theorem, Step 3

Step 3: n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.

Now construct one more preference profile, $[\succ^4]$, by changing $[\succ^3]$ in two ways. First, arbitrarily change the position of b in each voter's ordering while keeping all other relative preferences the same. Second, move a to an arbitrary position in n^*'s preference ordering, with the constraint that a remains ranked higher than c. Observe that all voters other than n^* have entirely arbitrary preferences in $[\succ^4]$, while n^*'s preferences are arbitrary except that $a \succ_{n^*} c$.

Profile $[\succ^1]$: Profile $[\succ^2]$: Profile $[\succ^3]$: Profile $[\succ^4]$:

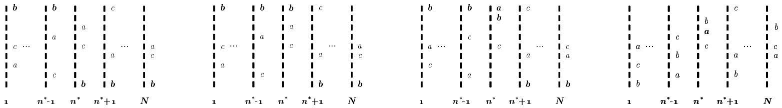


Arrow's Theorem, Step 3

Step 3: n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.

In $[\succ^3]$ and $[\succ^4]$ all agents have the same relative preferences between a and c; thus, since $a \succ_W c$ in $[\succ^3]$ and by IIA, $a \succ_W c$ in $[\succ^4]$. Thus we have determined the social preference between a and c without assuming anything except that $a \succ_{n^*} c$.

Profile $[\succ^1]$: Profile $[\succ^2]$: Profile $[\succ^3]$: Profile $[\succ^4]$:



Arrow's Theorem, Step 4

Step 4: n^* is a dictator over all pairs ab.

Consider some third outcome c. By the argument in Step 2, there is a voter n^{**} who is extremely pivotal for c. By the argument in Step 3, n^{**} is a dictator over any pair $\alpha\beta$ not involving c. Of course, ab is such a pair $\alpha\beta$. We have already observed that n^* is able to affect W's ab ranking—for example, when n^* was able to change $a \succ_W b$ in profile $[\succ^1]$ into $b \succ_W a$ in profile $[\succ^2]$. Hence, n^{**} and n^* must be the same agent.