Arrow's Theorem, Mechanism Design

CPSC 532A Lecture 16

October 26, 2006

Lecture Overview

Course stuff

- Assignment 2: solution posted right after class; graded assignments back Monday from Dave
- Midterm: Tuesday; 2:00 to 3:30
- Project Proposals: Tues Nov 14
- Final exam: December 10, 4:00 PM until December 12, 4:00 PM (take-home exam; paper or electronic submission)
- Projects due: December 19, 11:59:59 PM (electronic submission)
- Project reviews due: January 8, 5 PM

Lecture Overview

Notation

- N is the set of agents
- O is a finite set of outcomes with $|O| \geq 3$
- L the set of all possible preference orderings over O.
- $\succ$ 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 $\succ', \succ'' \in L^n$,
$\forall i \left(o_1 \succ_i' o_2 \leftrightarrow o_1 \succ_i'' o_2\right)$ implies that
$o_1 \succ_{W(\succ')} o_2 \Leftrightarrow o_1 \succ_{W(\succ'')} o_2$.

- 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. The argument proceeds in four steps.

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.

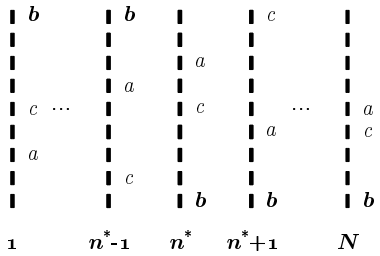
Step 2

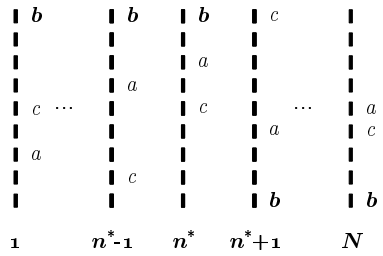
There is some voter n^* who is extremely pivotal in the sense that by changing his vote at some profile, he can move a given outcome b from the bottom of the social ranking to the top.

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.

Step 2

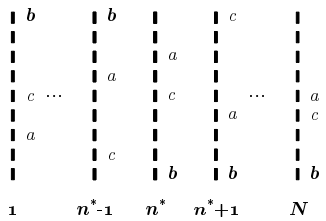
There is some voter n^* who is extremely pivotal in the sense that by changing his vote at some profile, he can move a given outcome b from the bottom of the social ranking to the top.

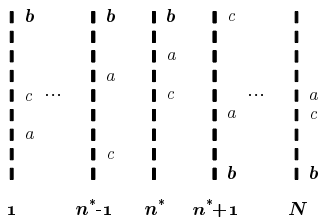


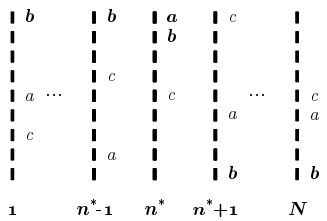


Denote by $\succ^1$ the set of preferences just before n^* moves b, and denote by $\succ^2$ the set of preferences just after n^* has moved b to the top of his ranking. In $\succ^1$, b is at the bottom in $\succ_W$. In $\succ^2$, b has changed its position in $\succ_W$, and every voter ranks b at either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

Step 3



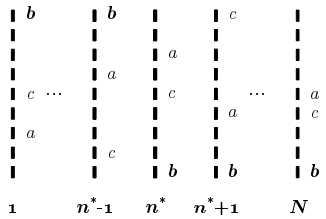


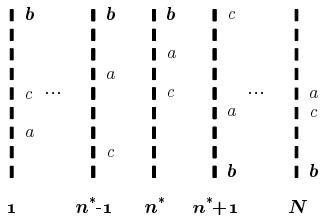


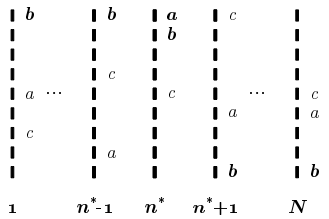
n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.

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.

Step 3



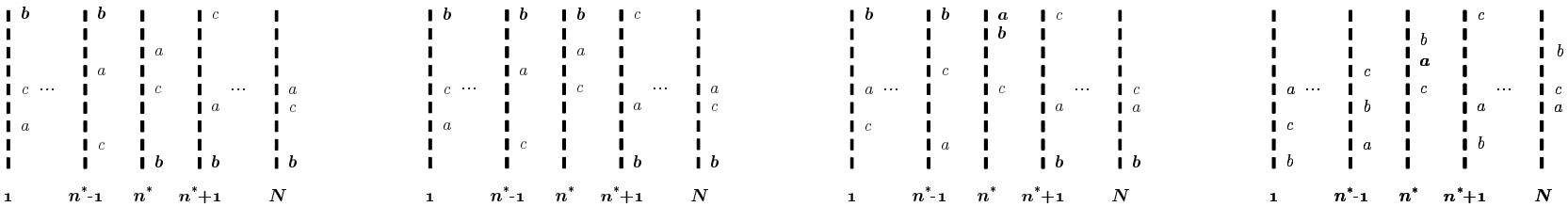




n^* (the agent who is extremely pivotal on outcome b) is a dictator over any pair ac not involving b.

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 $\succ^3$.

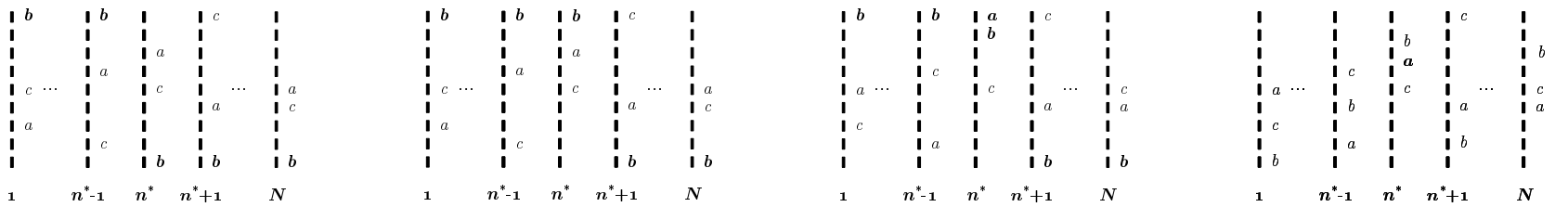
Step 3



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 ordering, $\succ^4$, 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 $a \succ_{n^*} c$.

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 $a \succ_{n^*} c$.

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.

Lecture Overview

Social Choice Functions

- Maybe Arrow's theorem held because we required a whole preference ordering.
- Idea: social choice functions might be easier to find
- We'll need to redefine our criteria for the social choice function setting; PE and IIA discussed the ordering

Weak Pareto Efficiency

Definition (Weak Pareto Efficiency)

A social choice function C is weakly Pareto efficient if, for any preference profile $\succ = (\succ_1, \ldots, \succ_n)$ (where $\succ_i \in L$), if there exist a pair of outcomes o_1 and o_2 such that $\forall i \in N$, $o_1 \succ_i o_2$, then $C(\succ) \neq o_2$.

- A dominated outcome can't be chosen.

Monotonicity

Definition (Monotonicity)

C is monotonic if, for any $o \in O$ and any preference profile $\succ = (\succ_1, \ldots, \succ_n)$ with $C(\succ) = o$, then for any other preference profile $\succ'$ with the property that $\forall i \in N, \forall o' \in O$, $o \succ_i o'$ if $o \succ_i o'$, it must be that $C(\succ') = o$.

- an outcome o must remain the winner whenever the support for it is increased relative to a preference profile under which o was already winning

Non-dictatorship

Definition (Non-dictatorship)

C is non-dictatorial if there does not exist an agent j such that C always selects the top choice in j's preference ordering.

The bad news

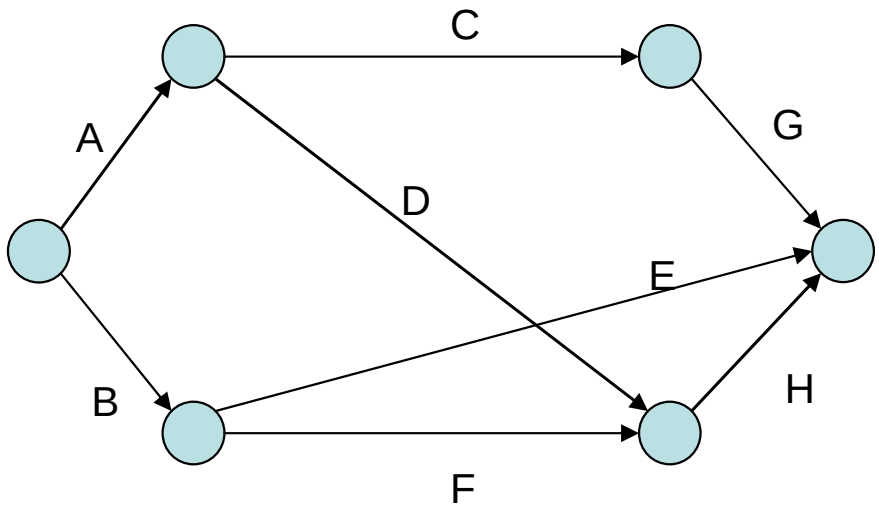
Theorem (Muller-Satterthwaite, 1977)

Any social choice function that is weakly Pareto efficient and monotonic is dictatorial.

- perhaps contrary to intuition, social choice functions are no simpler than social welfare functions after all.
- The proof repeatedly 'probes' a social choice function to determine the relative social ordering between given pairs of outcomes.
- Because the function must be defined for all inputs, we can use this technique to construct a full social welfare ordering.

Lecture Overview

Selfish Routing



- 8 people play as agents A – H; the others act as mediators.
- Agents' utility functions: $u_i =$ payment cost if your edge is chosen; 0 otherwise.
- Mediators: find me a path from source to sink, at the lowest cost you can.
- Agents: agree to be paid whatever you like; claim whatever you like; however, you can't show your paper to anyone.

Lecture Overview

Mechanism Design

- Extend the social choice setting to a new setting where agents can't be relied upon to disclose their preferences honestly.

Definition (Mechanism)

A mechanism (over a set of agents N and a set of outcomes O) is a pair (A, M), where

- $A = A_1 \times \cdots \times A_n$, where A_i is the set of actions available to agent $i \in N$, and
- $M : A \to \Pi(O)$ maps each action profile to a distribution over outcomes.

Thus, the designer gets to specify

- the action sets for the agents (though they may be constrained by the environment)
- the mapping to outcomes, over which agents have utility
- can't change agents' preferences for outcomes or type spaces

What we're up to

- The problem is to pick a mechanism that will cause rational agents to behave in a particular way, in order to maximize the mechanism designer's own "utility" or objective function
  - each agent holds private information, in the Bayesian game sense
  - often, we're interested in settings where agents' action space is identical to their type space, and an action can be interpreted as a declaration of the agent's type
- Various equivalent ways of looking at this setting
  - perform an optimization problem, given that the values of (some of) the inputs are unknown
  - choose the Bayesian game out of a set of possible Bayesian games that maximizes some performance measure
  - design a game that *implements* a particular social choice function in equilibrium, given that the designer no longer knows agents' preferences and the agents might lie

Implementation in Dominant Strategies

Definition (Implementation in dominant strategies)

A mechanism (A, M) (over N and O) is an implementation in dominant strategies of a social choice function C over (N and O) if for any vector of utility functions u, the game (N, A, O, M, u) has an equilibrium in dominant strategies, and in any such equilibrium a^* we have $M(a^*) = C(u)$.

Implementation in Bayes-Nash equilibrium

Definition (Bayes-Nash implementation)

We begin with a mechanism (A, M) over N and O. Let $\Theta = \Theta_1 \times \cdots \times \Theta_n$ denote the set of all possible type vectors $\theta = (\theta_1, \ldots, \theta_n)$, and denote agent i's utility as $u_i : O \times \Theta \to \mathbb{R}$. Let p be a (common prior) probability distribution on Θ (and hence on u). Then (A, M) is a Bayes-Nash implementation of a social choice function C, with respect to Θ and p, if there exists a Bayes-Nash equilibrium of the game of incomplete information (N, A, Θ, p, u) such that for every $\theta \in \Theta$ and every action profile $a \in A$ that can arise given type profile θ in this equilibrium, we have that $M(a) = C(u(\cdot, \theta))$.

Bayes-Nash Implementation Comments

Bayes-Nash Equilibrium Problems:

- there could be more than one equilibrium
  - which one should I expect agents to play?
- agents could miscoordinate and play none of the equilibria
- asymmetric equilibria are implausible

Refinements:

- Symmetric Bayes-Nash implementation
- *Ex-post* Bayes-Nash implementation

Implementation Comments

We can require that the desired outcome arises

- in the only equilibrium
- in every equilibrium
- in at least one equilibrium

Forms of implementation

- Direct Implementation: agents each simultaneously send a single message to the center
- Indirect Implementation: agents may send a sequence of messages; in between, information may be (partially) revealed about the messages that were sent previously like extensive form