Social Choice Theory

Felix Munoz-Garcia School of Economic Sciences
Washington State University

EconS 503 - Advanced Microeconomics II

Social choice theory

- MWG, Chapter 21.
- JR, Chapter 6.2-6.5.

- **Additional materials:**
- Gaertner (2009) *A Primer in Social Choice Theory*, Oxford University Press.
- It is a short book, intended for upper undergraduate students, but with all the content we need for this section. You can have a look at chapters 2-6, which I posted on Angel.

Social choice theory

- MWG, section 21.B.
- Consider a group of $I \geq 2$ individuals must choose an alternative from a set X.
  - We will first consider that set X is binary $X = \{x, y\}$
  - These two alternatives could represent the set of candidates competing for office, the policies to be implemented, etc.
- **Preferences:**
  - Every individual i's preference over x and y can be defined as a number:

    $$\alpha_i = \{1, 0, -1\}$$

    indicating that he prefers x to y, is indifferent between them, or he prefers alternative y to x, respectively.

Social choice theory

- We now seek to aggregate individual preferences with the use of a social welfare functional (or social welfare aggregator).
- **Social welfare functional:**
  - A social welfare functional (swf) is a rule

$$F(\alpha_1, \alpha_2, ..., \alpha_I) \in \{1, 0, -1\}$$

    which, for every profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I) \in \{1, 0, -1\}^I$, assigns a social preference $F(\alpha_1, \alpha_2, ..., \alpha_I) \in \{1, 0, -1\}$.
- *Example*:
  - For individual preferences $(\alpha_1, \alpha_2, \alpha_3) = (1, 0, 1)$, the swf $F(1, 0, 1) = 1$, thus prefering alternative x over y.

Social choice theory

- **Properties of swf:**
  - A swf is Paretian if it respects unanimity of strict preference;
  - That is, if it strictly prefers alternative x when all individuals strictly prefer x, i.e., $F(1, 1, ..., 1) = 1$,
  - but strictly prefers alternative y when all individuals strictly prefer y, i.e., $F(-1, -1, ..., -1) = -1$,
- *Note:*
  - This property is satisfied by many swf.
  - Weighted voting and Dictatorship are two examples (let's show that).

Social choice theory

- **Weighted voting swf:**
  - We first add individual preferences, assigning a weight $\beta_i \geq 0$ to every individual, where $(\beta_1, \beta_2, ..., \beta_I) \neq 0$, as follows $\sum_i \beta_i \alpha_i \in \mathbb{R}$.
  - We then apply the sign operator, which yields 1 when $\sum_i \beta_i \alpha_i > 0$, 0 when $\sum_i \beta_i \alpha_i = 0$, and -1 when $\sum_i \beta_i \alpha_i < 0$.
  - Hence,

$$F(\alpha_1, \alpha_2, ..., \alpha_I) = \text{sign} \sum_i \beta_i \alpha_i$$

  - In order to check if this swf is Paretian, we only need to confirm that

$$\begin{aligned} F(1, 1, ..., 1) &= 1, \text{ since } \sum_i \beta_i \alpha_i = \sum_i \beta_i > 0; \text{ and} \\ F(-1, -1, ..., -1) &= -1 \text{ since } \sum_i \beta_i \alpha_i = -\sum_i \beta_i < 0. \end{aligned}$$

- **Weighted voting swf:**
  - Needless to say, *simple majority* is a special case of weighted majority, whereby the weights satisfy $\beta_i = 1$ for every individual i.
  - The vote of every individual receives the same weight.
  - Intuitively, if the number of individuals who prefer alternative x to y is larger than the number of individuals prefering y to x, then $F(\alpha_1, \alpha_2, ..., \alpha_I) = 1$.
  - This swf is Paretian, given that

$$\begin{aligned} F(1, 1, ..., 1) &= 1, \text{ since } \sum_i \beta_i \alpha_i = N > 0; \text{ and} \\ F(-1, -1, ..., -1) &= -1 \text{ since } \sum_i \beta_i \alpha_i = -N < 0. \end{aligned}$$

Social choice theory

- **Dictatorial swf:**
  - The property of Paretian in swf is so lax that even Dictatorial swf satisfy it.
  - Let's first define a **dictatorial swf**:
  - We say that a swf is dictatorial if there exists an agent h, called the dictator, such that, for any profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$,

$$\begin{aligned} \alpha_h &= 1 \text{ implies } F(\alpha_1, \alpha_2, ..., \alpha_I) = 1, \text{ and} \\ \alpha_h &= -1 \text{ implies } F(\alpha_1, \alpha_2, ..., \alpha_I) = -1, \end{aligned}$$

    That is, the strict preference of the dictator prevails as the social preference.
  - We can understand the dictatorial swf as a extreme case of weighted voting...
    - where $\beta_h > 0$ for the dictator and $\beta_i = 0$ for all other individuals in the society $i \neq h$.

- **Dictatorial swf:**
  - Since weighted voting swf is Paretian, then the dictatorial swf (as a special case of weighted voting) must also be Paretian.
  - Extra confirmation:

$$\begin{aligned} F(1,1,...,1) &= 1, \text{ since } \sum_i \beta_i \alpha_i = \beta_h > 0; \text{ and} \\ F(-1,-1,...,-1) &= -1 \text{ since } \sum_i \beta_i \alpha_i = -\beta_h < 0. \end{aligned}$$

- **More properties of swf:**
  - *Symmetry among agents (or anonymity):*
  - The swf $F(\alpha_1, \alpha_2, ..., \alpha_I)$ is symmetric among agents (or anonymous) if the names of the agents do not matter.
  - That is, if a permutation of preferences across agents does not alter the social preference. Precisely, let

$$\pi : \{1, 2, ..., I\} \rightarrow \{1, 2, ..., I\}$$

    be an onto function (i.e., a function that, for every indvidual i, there is a j such that $\pi(j) = i$). Then, for every profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$, we have

$$F(\alpha_1, \alpha_2, ..., \alpha_I) = F\left(\alpha_{\pi(1)}, \alpha_{\pi(2)}, ..., \alpha_{\pi(I)}\right)$$

- *Example*: majority voting satisfies anonymity.

Social choice theory

Anonymity holds in simple majority

$(\alpha_1, \alpha_2, \alpha_3) = (1, -1, 1)$

$\pi(1) = 3$, red arrow

$\pi(2) = 1$, green arrow

$\pi(3) = 2$, purple arrow

$$\sum_i \alpha_i = 1 + (-1) + 1 = 1$$

$$\sum_i \alpha_i = (-1) + 1 + 1 = 1$$

Social preference coincides despite changing individual identities

Weighted voting does not necessarily satisfy anonymity

Using the same $(\alpha_1, \alpha_2, \alpha_3)$ and $\pi(i)$ function as above

$$\sum_i \beta_i \alpha_i = \beta_1 \cdot 1 + \beta_2 \cdot (-1) + \beta_3 \cdot 1 = \beta_1 + \beta_3 - \beta_2$$

$$\sum_i \beta_i \alpha_i = \beta_1 \cdot 1 + \beta_2 \cdot (-1) + \beta_3 \cdot 1 = \beta_1 + \beta_3 - \beta_2$$

Social preference does not necessarily coincide

Social choice theory

- **More properties of swf:**
  - *Neutrality between alternatives*
  - The swf $F(\alpha_1, \alpha_2, ..., \alpha_I)$ is neutral between alternatives if, for every profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$,

$$F(\alpha_1, \alpha_2, ..., \alpha_I) = -F(-\alpha_1, -\alpha_2, ..., -\alpha_I)$$

  - That is, if the social preference is reversed when we reverse the preferences of all agents.
  - This property is often understood as that the swf doesn't a priori distinguish either of the two alternatives.
- *Example*: majority voting satisfies neutrality between alternatives (see MWG pp. 792).

- **More properties of swf:**
  - *Positive responsiveness*
  - Consider a profile of individual preferences $(\alpha'_1, \alpha'_2, ..., \alpha'_I)$ where alternative x is socially preferred or indifferent to y, i.e., $F(\alpha'_1, \alpha'_2, ..., \alpha'_I) \geq 0$.
  - Take now a new profile $(\alpha_1, \alpha_2, ..., \alpha_I)$ in which some agents raise their consideration for x, i.e., $(\alpha_1, \alpha_2, ..., \alpha_I) \geq (\alpha'_1, \alpha'_2, ..., \alpha'_I)$ where $(\alpha_1, \alpha_2, ..., \alpha_I) \neq (\alpha'_1, \alpha'_2, ..., \alpha'_I)$.
  - We say that a swf is positively responsive if the new profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$ makes alternative x socially preferred, i.e., $F(\alpha_1, \alpha_2, ..., \alpha_I) = 1$.
- *Example*: majority voting satisfies neutrality between alternatives (see MWG pp. 792).

Social choice theory

Positive responsiveness

$(\alpha_1, \alpha_2, \alpha_3) = (1, 0, -1)$ and $F(1, 0, -1) = 0$ or 1

If $(\alpha_1', \alpha_2', \alpha_3') = (1, 1, -1)$ then $F(\alpha_1', \alpha_2', \alpha_3') = 1$

$(1, 0, 0)$

$(1, 0, 1)$

$(1, 1, 1)$

Any of these $(\alpha_1', \alpha_2', \alpha_3')$ satisfy $(\alpha_1', \alpha_2', \alpha_3') \geq (\alpha_1, \alpha_2, \alpha_3)$

Social choice theory

- Let us now extend our analysis to non-binary sets of alternatives X, e.g., $X = \{a, b, c, ...\}$
  - The use of majority voting swf, or weighted voting swf can be subject to non-transitivities in the resulting social preference.
  - That is, the order in which pairs of alternatives are voted can lead to cyclicalities, as shown in Condorcet's paradox (we already talked about it in the first weeks of 501, otherwise see page 270 in JR).
- An interesting question is:
  - Can we design voting systems (i.e., swf that aggregate individual preferences) that are not prone to the Condorcet's paradox and satisfy a minimal set of "desirable" properties?
  - That was the question Arrow asked himself (for his Ph.D. thesis!) obtaining a rather grim result, but a great thesis!

Arrow's impossibility theorem

- First, let us define which minimal requirements we would like to impose on swf's.
- The four minimal properties that Arrow impossed on any swf are:
  - **Unrestricted domain (U).** The domain of the swf, $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$, must include all possible combinations of individual preference relations on X.
    - In other words, we allow any sort of individual preferences over alternatives.

Arrow's impossibility theorem

- The four minimal properties that Arrow impossed on any swf are:
  - **Weak Pareto Principle (WP).** For any pair of alternatives x and y in X, if $x \succ^i y$ for every individual i, then the social preference is $x \succ y$.
    - That is, if every single member of society strictly prefers x to y, society should also prefer x to y.
    - The adjective "weak" is because WP doesn't require society to prefer x to y if, say, all but one strictly prefer x to y, yet one person is indifferent between x and y.
    - (In this case, the social preference doesn't need to prefer x to y.)

Arrow's impossibility theorem

- The four minimal properties that Arrow impossed on any swf are (cont'd):
  - **Independence of irrelevant alternatives (IIA).** Let $\succsim$ be social preferences arising from the list of individual preferences $\left(\succsim^1, \succsim^2, ... \succsim^I\right)$, and $\succsim'$ that arising when individual preferences are $\left(\succsim'^1, \succsim'^2, ... \succsim'^I\right)$. In addition, let x and y be any two alternatives in X. If each individual ranks x versus y under $\succsim^i$the same way he does under $\succsim'^1$, then the social ranking of x versus y is the same under $\succsim$ than under $\succsim'$.

- **More on IIA:**
- *Satisfaction/Violation of IIA:*
  - Note that IIA does not entail that $x \succsim^i y$ for every individual i.
  - Instead, it requires that, if $x \succsim^i y$ then $x \succsim'^i y$. But the preferences of individual j could be different, i.e., $y \succsim^j x$ and $y \succsim'^j x$.
  - In addition, the preferences of every individual i must rank x and y in the same way under $\succsim^i$than under $\succsim'^i$.
  - However, their preferences can differ in their ranking of other alternatives, i.e., $a \succsim^i b$ and $b \succsim'^i a$.

Arrow's impossibility theorem

- **More on IIA** (Interpretation)**:**
  - *Morning:* Suppose that in the morning some individuals prefer x, $x \succsim^i y$, while others prefer y, $y \succsim^i x$. However, they all rank alternative z below both x and y.
    - In addition, suppose that the swf yields a social preference of x over y, i.e., $x \succ^i y$.
  - *Afternoon:* Now alternative z is ranked above both x and y for all individuals.
    - However, the ranking that every individual had between x and y has not changed, i.e., if $x \succsim^i y$ then $x \succsim'^i y$, and if $y \succsim^i x$ then $y \succsim'^i x$.
  - The IIA says that society should still rank x over y in the afternoon, i.e., $x \succ'^i y$.

Arrow's impossibility theorem

IIA property (simple example)

Morning		Social Preference	Afternoon		Social Preference
$\succsim_1$	$\succsim_2$	$\succsim$	$\succsim'_1$	$\succsim'_2$	$\succsim'$
a	b	c	c	a	b
⋮	⋮	⋮	⋮	⋮	⋮
x	y	x	x	y	x
y	x	y	y	x	y
⋮	⋮	⋮	⋮	⋮	⋮
$x \succsim_1 y$	$y \succsim_2 x$	$x \succsim y$	$x \succsim'_1 y$	$y \succsim'_2 x$	$x \succsim' y$

Arrow's impossibility theorem

IIA property (richer example)

Morning		**Social Preference**	**Afternoon**		**Social Preference**
$\succsim_1$	$\succsim_2$	$\succsim$	$\succsim'_1$	$\succsim'_2$	$\succsim'$
a	b	b	b	a	a
b	c	a	c	c	c
⋮	a	c	⋮	⋮	d
x	⋮	⋮	x	y	⋮
c	y	x	a	b	x
y	d	d	y	d	b
⋮	x	y	⋮	x	y
⋮	⋮	⋮	⋮	⋮	⋮
$x \succsim_1 y$	$y \succsim_1 x$	$x \succsim y$	$x \succsim'_1 y$	$y \succsim'_2 x$	$x \succsim' y$

Arrow's impossibility theorem

IIA property - Violation

Morning		Social Preference	Afternoon		Social Preference
$\succsim_1$	$\succsim_2$	$\succsim$	$\succsim'_1$	$\succsim'_2$	$\succsim'$
a	b	c	c	a	b
⋮	⋮	⋮	⋮	⋮	⋮
x	y	x	x	y	y
y	x	y	y	x	x
⋮	⋮	⋮	⋮	⋮	⋮
Hence, $x \succsim_1 y$	$y \succsim_2 x$	$x \succsim y$	Hence, $x \succsim'_1 y$	$y \succsim'_2 x$	$x \succsim' y$

While individual 1's and individual 2's preference over alternatives x and y is constant along time, the social preference over these alternatives changes from $x \succsim y$ in the morning to $y \succsim' x$ in the afternoon.

Arrow's impossibility theorem

IIA property – Not a violation (premise of IIA does not hold)

Morning		Social Preference	Afternoon		Social Preference
$\succsim_1$	$\succsim_2$	$\succsim$	$\succsim'_1$	$\succsim'_2$	$\succsim'$
a	b	c	c	a	b
⋮	⋮	⋮	⋮	⋮	⋮
x	y	x	x	x	y
y	x	y	y	y	x
⋮	⋮	⋮	⋮	⋮	⋮
Hence, $x \succsim_1 y$	$y \succsim_2 x$	$x \succsim y$	Hence, $x \succsim'_1 y$	$x \succsim'_2 y$	$y \succsim' x$

- While social preference for x and y have changed from $x \succsim y$ to $y \succsim' x$, the preference of individual 2 changed as well.
- For us to check IIA holds, we need to have that every individual ranks alternatives x and y in the same way in the morning and afternoon, but here this premise does not hold. We can not claim that IIA is violated.

Arrow's impossibility theorem

- The four minimal properties that Arrow impossed on any swf are (cont'd):
  - **Non-dictatorship (D).** There is no individual h such that for all pairs of alternatives $(x, y) \in X$, $x \succ^h y$ for him implies a social preference of $x \succ y$ regardless of the preferences of all other individuals $j \neq i$, $\left(\succsim^1, \succsim^2, ..., \succsim^{h-1}, \succsim^{h+1}, ..., \succsim^I\right)$.
    - Note that this is a very mild assumption:
    - A "virtual" dictatorship in which an individual h imposes his preference on the social preferences for all pairs of alternatives, i.e., $(x, y) \in X$, but one pair of alternatives, would be considered to satisfy the non-dictatorship property.

Arrow's impossibility theorem

- Most of these assumptions are often accepted as the minimal assumptions that we should impose on any swf that aggregates individual preferences into a social preference.
- Then, Arrow's impossibility theorem comes as a surprising, even disturbing, result:
    - **Arrow's impossibility theorem.**
    - If there are at least three alternatives in the set of alternatives X, then there is no swf that simultaneously satisfies U, WP, IIA, and D.

Arrow's impossibility theorem

- **Proof:**
- We will assume that U, WP and IIA hold, and show that all swf simultaneously satisfying these three properties are dictatorial.
  - (U is used along the proof when we alter the profile of individual preferences, since all profiles are admissible.)
- *Step 1:*
  - Consider that an alternative c is placed at the bottom of the ranking of every individual i.
  - Then, by WP, alternative c must be placed at the bottom of the ranking as well.
  - (See figure)

Arrow's impossibility theorem

- **Proof (cont'd):**

$$
\begin{array}{ccccc}
\succsim^1 & \succsim^2 & \ldots & \succsim^I & \text{Soc.Pref. } \succsim \\
x & x' & \ldots & x'' & x''' \\
y & y' & \ldots & y'' & y''' \\
. & . & & & . \\
. & . & & & . \\
. & . & & & . \\
c & c & \ldots & c & c
\end{array}
$$

- **Proof (cont'd):**
  - *Step 2:*
  - Imagine now moving alternative c to the top of individual 1's ranking, leaving the ranking of all other alternatives unaffected.
  - Next, do the same move for individual 2, individual 3...
  - Let individual n be the first such that raising c to the top of his ranking causes the social ranking of alternative c to increase.
  - Figure.

Arrow's impossibility theorem

- **Proof (cont'd):**

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	x''	c
x	x'	...		...	y''	.
y	y'					.
.	.					.
.	.					.
.	.					.
w	w'	...		...	c	w'''

  - The social ranking of alternative c not only increases but actually jumps to the *top* of the social ranking.
  - (We need to show that!)

Arrow's impossibility theorem

- **Proof (cont'd):**
  - Let's prove this result by contradiction:
  - The social ranking of c increases but *not to the top*, i.e., $\alpha \succsim c$ for some states but $c \succsim \beta$ for other states, $\alpha, \beta \neq c$.

Arrow's impossibility theorem

- **Proof (cont'd):**

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	x''	α
x	x'	...		...	y''	.
y	y'					c
.	.					β
.	.					.
.	.					.
w	w'	...		...	c	w'''

  - The social ranking states that $\alpha \succsim c$ for some states but $c \succsim \beta$ for other states, $\alpha, \beta \neq c$.

Arrow's impossibility theorem

- **Proof (cont'd):**

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	β	α
α	β	...	α	...	z	.
β	α					c
.	.				α	β
.	.					.
.	.		β			.
w	w'	...		...	c	w'''

  - Because alternative c is either at the top or the bottom of every individual's ranking, we can change each individual i's preferences so that $\beta \succ^i \alpha$, while leaving the position of c unchanged for that individual.

Arrow's impossibility theorem

- **Proof (cont'd):**

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	β	α
β	β	...	β	...	z	.
α	α					c
.	.				α	β
.	.					.
.	.		α			.
w	w'	...		...	c	w'''

  - We have now changed each individual i's preferences so that $\beta \succ^i \alpha$, while leaving the position of c unchanged for that individual.

- **Proof (cont'd):**
  - This produces our desired contradiction:
    - 1) On one hand, $\beta \succ^{i} \alpha$ for every individual implies that a social preference of $\beta \succ \alpha$ since the swf satisfies WP.
    - 2) On the other hand, the rankings of α relative to c and of β relative to c have not changed for any individual, which by IIA implies that the social ranking of α relative to c and of β relative to c must remain unchanged. Hence, the social ranking still is $\alpha \succsim c$ and $c \succsim \beta$.
    - But transitivity implies that if $\alpha \succsim c$ and $c \succsim \beta$, then $\alpha \succsim \beta$ (as a conclusion of point 2) contradicting $\beta \succ \alpha$ (from point 1).
    - Hence, alternative c must have moved to the top of the social ranking.

Arrow's impossibility theorem

- **Proof (cont'd):**
  - *Step 3:*
  - Consider now any two distinct alternatives a and b, each different from c.
  - In the table on individual and social preferences, change the preferences of individual n as follows:

$$a \succ^n c \succ^n b$$

  - For every other individual $i \neq n$, rank alternatives a and b in any way that keeps the rank of c unchanged.
  - Example in the next figure.

Arrow's impossibility theorem

- **Proof (cont'd):**

$\succsim^1$	$\succsim^2$	$\dots$	$\succsim^n$	$\dots$	$\succsim^I$	Soc.Pref. $\succsim$
c	c	$\dots$	a	$\dots$	x''	c
x	x'	$\dots$	c	$\dots$	y''	.
y	y'		b		.	.
.	.				.	a
.	.				.	b
a	b	$\dots$		$\dots$	a	.
b	a	$\dots$		$\dots$	b	
.	.					.
w	w'	$\dots$		$\dots$	c	w'''

Arrow's impossibility theorem

- **Proof (cont'd):**
  - *Step 3:*
  - In the new profile of individual preferences, the ranking of alternatives a and c is the same for every individual as it was just *before* raising c to the top of individual n's ranking in Step 2.
  - Therefore, by IIA, the social ranking of alternatives a and c must be the same as it was at that moment (just *before* raising c to the top of individual n's ranking in Step 2).
    - That is, $a \succ c$, since at that moment alternative c was still at the bottom of the social ranking.

- **Proof (cont'd):**
  - Similarly, in the new profile of individual preferences, the ranking of alternatives c and b is the same for every individual as it was just after raising c to the top of individual n's ranking in Step 2.
  - Therefore, by IIA, the social ranking of alternatives c and b must be the same as it was at that moment.
    - That is, $c \succ b$, since at that moment alternative c had just risen to the top of the social ranking.

Arrow's impossibility theorem

- **Proof (cont'd):**
  - Since $a \succ c$ and $c \succ b$, we may conclude that, by transitivity, $a \succ b$.
    - Then, no matter how individuals different from n rank alternatives a and b, the social ranking agrees with individual n's ranking.
    - That is, while $a \succ^i b$ for some individuals and $b \succ^j a$ for other individuals, the fact that $a \succ^n b$ for individual n implies that $a \succ b$ for the social ranking, which is true for any two alternatives $a, b \neq c$.

$$a \succ^n b \text{ for individual } n \Rightarrow a \succ b \text{ for the social ranking}$$

  - This result shows that individual n is a dictator in all alternatives $a, b \neq c$, which completes the proof!

Arrow's impossibility theorem

- **Summary:**
- Hence, we started with properties U, WP, and IIA for a swf...
- and showed that the social preference must coincide with that of one individual,
- thus violating the non-dictatorship property (D).

Arrow's impossibility theorem - II

- **Alternative proof using diagrams**
- See section 6.2.1 in JR.
- Otherwise, see section 2.4 in Gaertner's book (posted on Angel).

Arrow's impossibility theorem - II

- Consider that the set of alternatives X is not finite, but contains infinitely many alternatives.
  - In particular, assume that X is a convex subset of $\mathbb{R}^K$, where $K \geq 1$.
- Individual preferences $\succsim^i$ on X can be represented with a continuous utility function $u_i : X \to \mathbb{R}$.
- Consider now a social welfare function $f(u^1(\cdot), ..., u^I(\cdot))$ that maps continuous individual utility functions into a continuous utility function for society.
  - For each continuous $\mathbf{u}(\cdot) = (u^1(\cdot), ..., u^I(\cdot))$, let $f_{\mathbf{u}}$ denote the social utility function, and $f_{\mathbf{u}}(x)$ represent the utility assigned to $x \in X$.

Arrow's impossibility theorem - II

- We seek to guarantee that the social ranking of alternatives is determined only by the individual preference relations $\succsim^i$.
- Hence, $f_{\mathbf{u}}$ would have to be unaffected if any individual utility function $u^i(\cdot)$ were replaced by a utility function that represents the same preference relation of this individual, $\succsim^i$,
  - i.e., if we apply a monotonic transformation on $u^i(\cdot)$.
- More formally, if we apply a strictly increasing and continuous function $\psi^i : \mathbb{R} \to \mathbb{R}$ to individual i's utility function $u^i(\cdot)$, then

$$f_{\mathbf{u}}(x) \geq f_{\mathbf{u}}(y) \iff f_{\psi \circ \mathbf{u}}(x) \geq f_{\psi \circ \mathbf{u}}(y)$$

  where $\psi \circ \mathbf{u}(\cdot) = \left(\psi^1\left(u^1(\cdot)\right), ..., \psi^I\left(u^I(\cdot)\right)\right)$.

Arrow's impossibility theorem - II

- Let's define the premises of our theorem before starting with the proof!
- Condition U in this setup means that the domain of the social utility function $f(u^1(\cdot), ..., u^I(\cdot))$ is the entire set of continuous individual utility functions, $u^1(\cdot), ..., u^I(\cdot)$.
- Condition IIA means that $f_{\mathbf{u}}(x)$ being greater, less or equal to $f_{\mathbf{u}}(y)$ can depend only on vectors

$$u^1(x), ..., u^I(x) \quad \text{and} \quad u^1(y), ..., u^I(y)$$

  indicating the utility each individual assigns to alternative x and y, and not on any other values taken by the utility function $\mathbf{u}(\cdot) = (u^1(\cdot), ..., u^I(\cdot))$.

Arrow's impossibility theorem - II

- The meaning of conditions WP and D remain as before:
  - **WP:** If $u^i(x) > u^i(y)$ for every individual i, then the social ranking satisfies $f_{\mathbf{u}}(x) > f_{\mathbf{u}}(y)$.
  - **D:** There is no individual h such that $u^h(x) > u^h(y)$ implies $f_{\mathbf{u}}(x) > f_{\mathbf{u}}(y)$ for all pairs of alternatives $(x, y) \in X$.

Arrow's impossibility theorem - II

- Consider now the following condition on f.
- **Pareto Indifference Principle (PI).**
  - If every individual $i \in I$ obtains the same utility from two distinct alternatives $x, y \in X$, i.e., $u^i(x) = u^i(y)$, then the social utility function also assigns the same utility to both options, $f_{\mathbf{u}}(x) = f_{\mathbf{u}}(y)$.

Arrow's impossibility theorem - II

- Sen (1970) showed that if f satisfies U, IIA, WP and PI, then there is a strictly increasing continuous function $W : \mathbb{R}^I \rightarrow \mathbb{R}$, such that
  - for all alternatives $x, y \in X$, and all profiles of continuous individual utility functions $\mathbf{u}(\cdot) = (u^1(\cdot), ..., u^I(\cdot))$,

$$f_{\mathbf{u}}(x) \geq f_{\mathbf{u}}(y) \iff W(u^1(x), ..., u^I(x)) \geq W(u^1(y), ..., u^I(y))$$

- Hence, the social welfare function $W(\cdot)$ only considers:
  - The utility value that each individual assigns to alternative x and y,
  - but ignores all non-utility information with respect to the alternatives.
  - For this reason, this approach is referred to as 'welfarism'

Arrow's impossibility theorem - II

- Let's now start with the proof!
- We need to show that :
  - If the social welfare function $W(\cdot)$ satisfies the above condition

    $$f_{\mathbf{u}}(x) \geq f_{\mathbf{u}}(y) \iff W(u^1(x), ..., u^I(x)) \geq W(u^1(y), ..., u^I(y))$$

    (so conditions U, IIA, WP and PI hold), then a dictator exists.

Arrow's impossibility theorem - II

- Consider two alternatives x and y, and the profiles of individual utility vectors:
  - $u^1(x), ..., u^I(x) = u^1, ..., u^I$, and
  - $u^1(y), ..., u^I(y) = \tilde{u}^1, ..., \tilde{u}^I$.
- If W ranks x as socially better than y, i.e.,

$$W(u^1, ..., u^I) \geq W(\tilde{u}^1, ..., \tilde{u}^I)$$

then we must have that

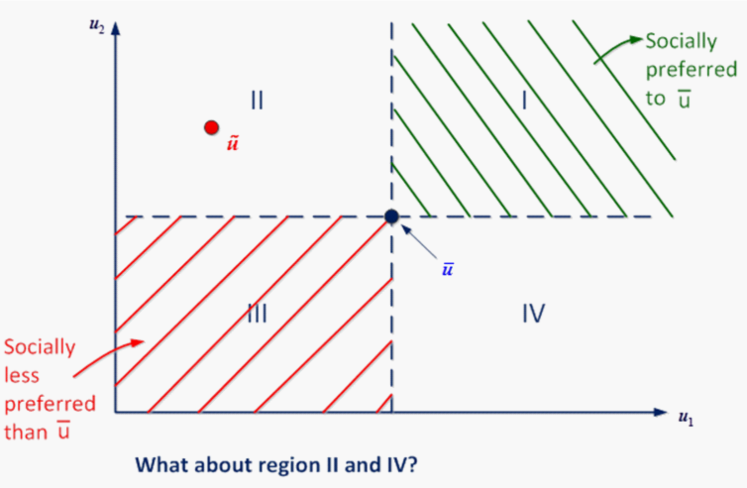
$$W(\psi^1\left(u^1\right), ..., \psi^I\left(u^I\right)) \geq W(\psi^1\left(\tilde{u}^1\right), ..., \psi^I\left(\tilde{u}^I\right))$$

for any continuous and strictly increasing transformation $\psi : \mathbb{R} \rightarrow \mathbb{R}$, since f was invariant to strictly increasing transformations on individual utility functions.

Arrow's impossibility theorem - II

- Let us consider an arbitrary utility pair $\bar{\mathbf{u}} = (\bar{u}^1, \bar{u}^2)$ in the following (u_1, u_2) quadrant.
- Region I includes utility pairs that, by WP, must be socially preferred to $\bar{\mathbf{u}}$.
- Utility pair $\bar{\mathbf{u}}$ is, by WP, socially preferred to those in Region III.
- What about the ranking between utility pair $\bar{\mathbf{u}}$ and those in Region II and IV?

Arrow's impossibility theorem - II



Arrow's impossibility theorem - II

- Consider an arbitrary point $\widetilde{\mathbf{u}}$ in Region II.
  - Is this point yielding a lower, higher or equal social welfare than $\overline{\mathbf{u}}$?

$$
\begin{aligned}
W(\overline{\mathbf{u}}) &> W(\widetilde{\mathbf{u}}),\\
W(\overline{\mathbf{u}}) &< W(\widetilde{\mathbf{u}}),\\
W(\overline{\mathbf{u}}) &= W(\widetilde{\mathbf{u}})
\end{aligned}
$$

Arrow's impossibility theorem - II

- Suppose that $W(\overline{\mathbf{u}}) < W(\widetilde{\mathbf{u}})$.
  - Recall that W's ordering is invariant to monotonic transformations of utility functions.
  - Then, consider two strictly increasing functions ψ^1 and ψ^2.
  - Assume that applying them to the coordinates of point $\overline{\mathbf{u}}$, produce

$$\psi^1(\overline{u}^1) = \overline{u}^1 \quad \text{and} \quad \psi^2(\overline{u}^2) = \overline{u}^2$$

Arrow's impossibility theorem - II

- Suppose that $W(\overline{\mathbf{u}}) < W(\widetilde{\mathbf{u}})$.
  - However, when applying them to the coordinates of point $\widetilde{\mathbf{u}}$, they produce

$$\begin{aligned} v^1 &\equiv \psi^1(\widetilde{u}^1) < \psi^1(\overline{u}^1) = \overline{u}^1 \text{ and} \\ v^2 &\equiv \psi^2(\widetilde{u}^2) > \psi^2(\overline{u}^2) = \overline{u}^2 \end{aligned}$$

  since they are both strictly increasing and $\widetilde{u}^1 < \overline{u}^1$ and $\widetilde{u}^2 > \overline{u}^2$ (i.e., $\widetilde{\mathbf{u}}$ belongs to region II).

Arrow's impossibility theorem - II

- *Example of monotonic transformation:*

$$\psi^i(t) \equiv \underbrace{\left[\frac{\overline{u}^i - u^i}{\overline{u}^i - \widetilde{u}^i}\right]}_{\alpha^i} t + \underbrace{\left[\frac{u^i - \widetilde{u}^i}{\overline{u}^i - \widetilde{u}^i}\right] \overline{u}^i}_{\beta^i}$$

which is of the form $\psi^i(t) = \alpha^i t + \beta^i$.

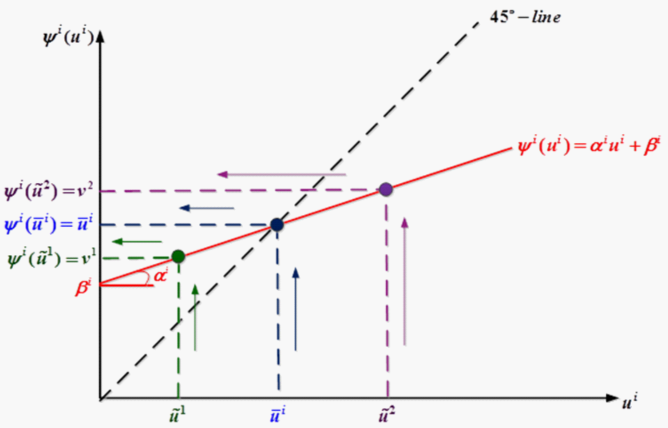
  - Note that, when $t = \overline{u}^i$, we obtain

$$\begin{aligned} \psi^i(\overline{u}^i) &\equiv \left[\frac{\overline{u}^i - u^i}{\overline{u}^i - \widetilde{u}^i}\right] \overline{u}^i + \left[\frac{u^i - \widetilde{u}^i}{\overline{u}^i - \widetilde{u}^i}\right] \overline{u}^i \\ &= \frac{\overline{u}^i - u^i + u^i - \widetilde{u}^i}{\overline{u}^i - \widetilde{u}^i} \overline{u}^i = \overline{u}^i \end{aligned}$$

as required.

Arrow's impossibility theorem - II

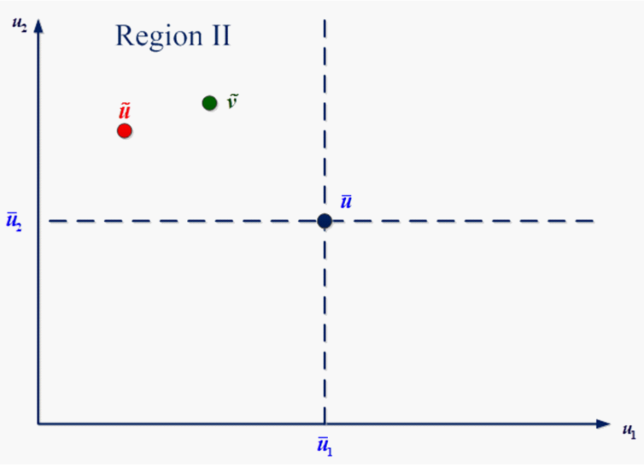
Any increasing function like the one below would do the job:



Arrow's impossibility theorem - II

- Hence, point $\widetilde{\mathbf{v}} \equiv \left(v^1, v^2\right)$ must be somewhere in region **II**, since we found that, relative to $\overline{\mathbf{u}}$, $v^1 < \overline{u}^1$ and $v^2 > \overline{u}^2$.
- However, since we have complete flexibility of the monotonic transformation functions ψ^1 and ψ^2, point $\widetilde{\mathbf{v}} \equiv \left(v^1, v^2\right)$ can be anywhere in region **II**.
- But then every point in region **II**, $\widetilde{\mathbf{v}} \equiv \left(v^1, v^2\right)$, must be ranked the same way relative to $\overline{\mathbf{u}}$.
- Then, if we started assuming that $W(\overline{\mathbf{u}}) < W(\widetilde{\mathbf{u}})$, the same argument applies to all points in region **II**, i.e., $W(\overline{\mathbf{u}}) < W(II)$.
    - The same argument applies if we started with $W(\overline{\mathbf{u}}) > W(\widetilde{\mathbf{u}})$, or with $W(\overline{\mathbf{u}}) = W(\widetilde{\mathbf{u}})$.

Arrow's impossibility theorem - II



Arrow's impossibility theorem - II

- Case $W(\overline{\mathbf{u}}) = W(\widetilde{\mathbf{u}})$ can, however, be discarded:
  - Since W is strictly increasing, then the social welfare at point $\widetilde{\mathbf{v}}$ must be higher than at point $\widetilde{\mathbf{u}}$, since $\widetilde{\mathbf{v}} >> \widetilde{\mathbf{u}}$ (in every component).
  - Hence, either $W(\overline{\mathbf{u}}) < W(II)$ or $W(\overline{\mathbf{u}}) > W(II)$.
- A similar argument applies to the points in region IV, where
  - either $W(\overline{\mathbf{u}}) < W(IV)$ or $W(\overline{\mathbf{u}}) > W(IV)$.

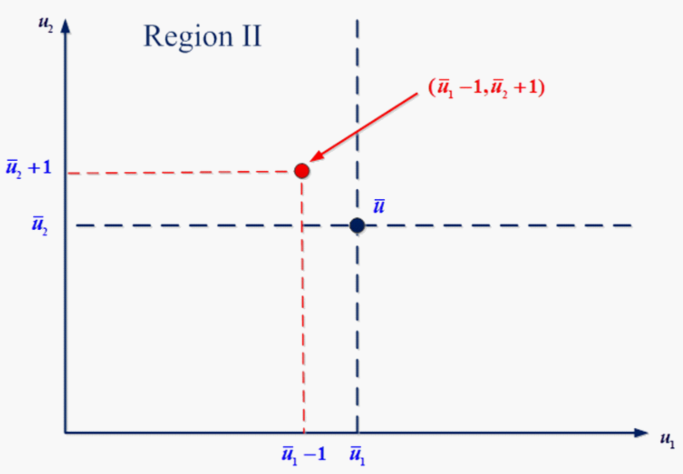
Arrow's impossibility theorem - II

- Now suppose that $W(\overline{\mathbf{u}}) < W(II)$.
  - Then, $W(\overline{\mathbf{u}}) < W(\overline{u}^1 - 1, \overline{u}^2 + 1)$, as depicted in the next figure.
  - Consider now the pair of strictly increasing functions

$$\psi^1(u^1) = u^1 + 1 \text{ and } \psi^2(u^2) = u^2 - 1$$

  - Applying these functions to point $\overline{\mathbf{u}}$ yields $\left(\overline{u}^1 + 1, \overline{u}^2 - 1\right)$
  - Applying these functions to point $(\overline{u}^1 - 1, \overline{u}^2 + 1)$, yields $\left(\overline{u}^1, \overline{u}^2\right) \equiv \overline{\mathbf{u}}$, respectively.

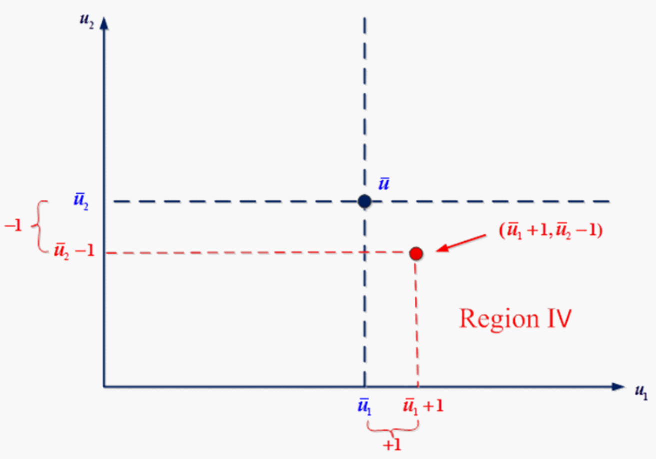
Arrow's impossibility theorem - II



Arrow's impossibility theorem - II

- Since W must be order-invariant to transformations ψ^1 and ψ^2, then
  - If $W(\overline{\mathbf{u}}) < W(\overline{u}^1 - 1, \overline{u}^2 + 1)$, then $W\left(\overline{u}^1 + 1, \overline{u}^2 - 1\right) < W(\overline{\mathbf{u}})$.
  - Since point $\left(\overline{u}^1 + 1, \overline{u}^2 - 1\right)$ lies in region IV (see next figure), we have just obtained that point $\overline{\mathbf{u}}$ is strictly socially preferred to $\left(\overline{u}^1 + 1, \overline{u}^2 - 1\right)$.
  - A similar argument applies to every point in region IV (by altering the +1 and -1 in the transformations for any other values, not necessarily symmetric in every dimension.
  - That is,
$$W(\overline{\mathbf{u}}) > W(IV)$$

Arrow's impossibility theorem - II



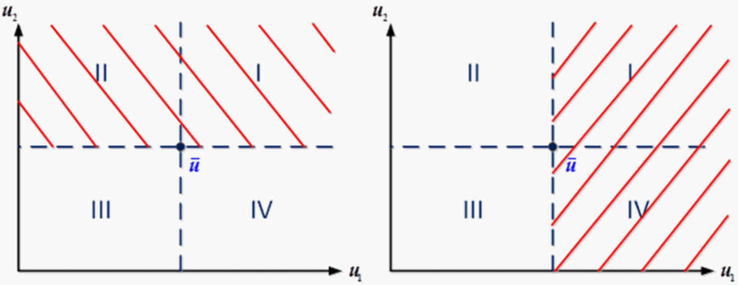
Arrow's impossibility theorem - II

- Let's start summarizing!
  - We have shown that, if $W(\bar{\mathbf{u}}) < W(II)$, then $W(\bar{\mathbf{u}}) > W(IV)$.
  - A similar argument applies if we assume that $W(\bar{\mathbf{u}}) > W(II)$, where $W(\bar{\mathbf{u}}) < W(IV)$ arises.
  - Combining these results, we have that

$$\begin{aligned} \text{either} \quad W(IV) &< W(\bar{\mathbf{u}}) < W(II), \\ \text{or} \quad W(II) &< W(\bar{\mathbf{u}}) < W(IV) \end{aligned}$$

Arrow's impossibility theorem - II

Case 1: $w(IV) < w(\bar{u}) < w(II)$ **Case 2:** $w(II) < w(\bar{u}) < w(IV)$



- Shaded areas represent (u_1, u_2)-pairs a higher social welfare than $\bar{u}$.
- Unshaded areas represent (u_1, u_2)-pairs a higher social welfare than $\bar{u}$.

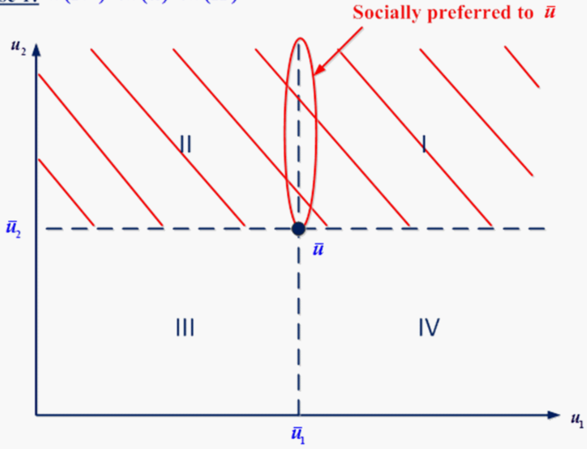
Arrow's impossibility theorem - II

- What about the dashed line depicting the frontier between regions?
  - If two adjacent regions are ranked the same way relative to $\bar{\mathbf{u}}$, then the dashed line separating the two regions must be ranked the same way relative to $\bar{\mathbf{u}}$.
  - For example, if regions I and II are ranked as socially preferred to $\bar{\mathbf{u}}$, then by the WP property, points in the dashed line above $\bar{\mathbf{u}}$ must be ranked as socially preferred to $\bar{\mathbf{u}}$.

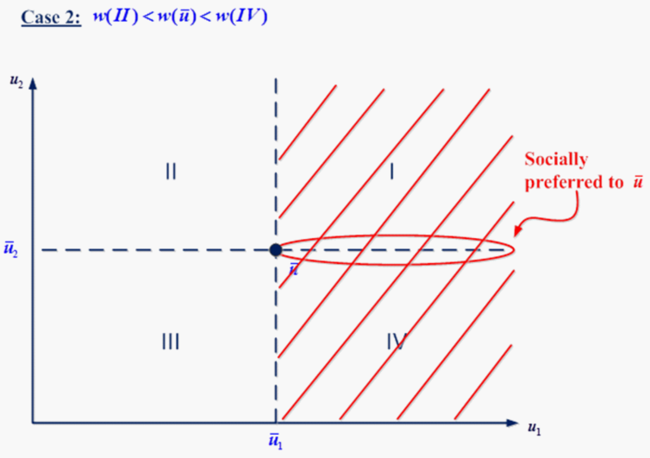
Arrow's impossibility theorem - II

Consider that regions I and II are socially preferred to $\bar{u}$.

Case 1: $w(IV) < w(\bar{u}) < w(II)$



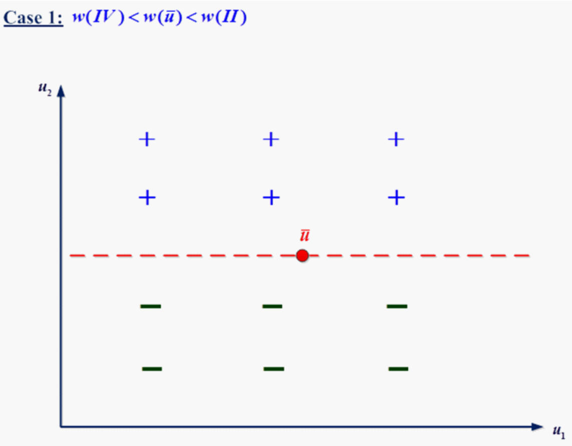
Arrow's impossibility theorem - II



Arrow's impossibility theorem - II

- As a consequence, since region I is ranked above $\overline{\mathbf{u}}$ while region III is ranked below, if $W(IV) < W(\overline{\mathbf{u}}) < W(II)$ holds,
    - the social ranking must be given as in depicted in the next figure,
    - *Legend:* $+$ signs represent utility vectors $\mathbf{u} = (u^1, u^2)$ with a social welfare $W(\mathbf{u})$ greater than $W(\overline{\mathbf{u}})$.
    - - signs represent utility vectors $\mathbf{u} = (u^1, u^2)$ with a social welfare $W(\mathbf{u})$ smaller than $W(\overline{\mathbf{u}})$.
- By the continuity of W, we can then conclude that the indifference curve through point $\overline{\mathbf{u}}$ is a straight line.

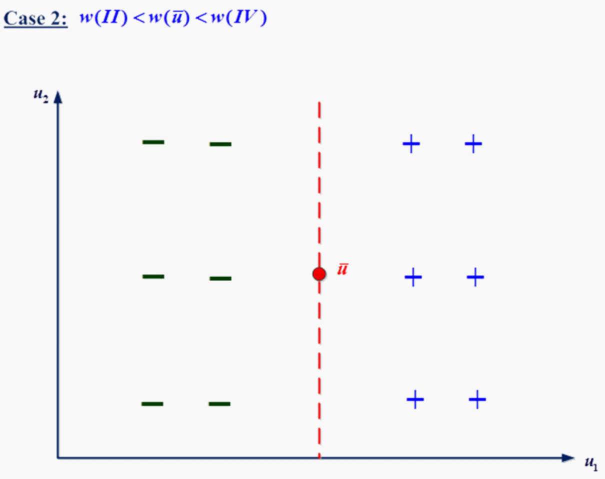
Arrow's impossibility theorem - II



Arrow's impossibility theorem - II

- A similar analysis applies to the case in which $W(II) < W(\bar{\mathbf{u}}) < W(IV)$ holds,
  - The social ranking must be given as in depicted in the next figure.
- By the continuity of W, we can then conclude that the indifference curve through point $\bar{\mathbf{u}}$ is a vertical line.

Arrow's impossibility theorem - II



Arrow's impossibility theorem - II

- Since point $\bar{\mathbf{u}}$ was arbitrary, then either all indifference curves are horizontal lines, or all indifference curves are vertical lines.
  - In the case that they are all horizontal lines, then individual 2 is a dictator.
  - In the case that they are all vertical lines, then individual 1 is a dictator.
- In either case, we have established the existence of a dictator, thus completing the proof.

Arrow's impossibility theorem - II