Social Choice Theory

Nick Netzer

University of Zurich

Spring 2022

Conflicting Interests

The fundamental problem of collective decision-making is that the involved individuals will often have conflicting interests:

- Some citizens support buying the F-35 jet, others not.
- Attitudes towards taxes and subsidies depend on the own income.
- Farmers like agricultural subsidies, taxpayers might be against them.
- Producers like high product prices, consumers like low product prices.
- Boxing referees sometimes disagree about the winner.
- University selection committees might not agree on the best candidate.
- Different countries have different preferences at the Eurovision Song Contest.
- ...

Social Choice Theory

Social choice theory investigates how such conflicts can be resolved.

- Normative aspect:

  How should conflicts be resolved? Which decision procedure is fair?

  In this sense, social choice theory is about fairness and justice.

- Positive aspect:

  What outcomes do procedures such as plurality voting yield? How can certain voting paradoxes be explained?

  In this sense, social choice theory is a theory of voting and voting systems.

Part 1: Preliminaries

Organizational Remarks

Information

- Target group: master students economics (major and minor)
- Prerequisites: solid microeconomics, mathematical methods
- 4 hours/week
  Mon 10:15-12:00, Thu 10:15-12:00, break?
- 6 ECTS points, final exam
- Lecture slides and other materials on OLAT
- Integrated tutorial (time share about 1/4)

Literature

Mandatory textbook:

[GA] W. Gaertner: A Primer in Social Choice Theory, LSE Perspectives in Economic Analysis, Revised Edition, 2009, Oxford University Press.

Other books:

[BK] F. Breyer, M. Kolmar: Grundlagen der Wirtschaftspolitik, 2014, Mohr-Siebeck.

[KR] D.M. Kreps: Notes on the Theory of Choice, Underground Classics in Economics, 1988, Westview Press.

[MC] A. Mas-Colell, M.D. Whinston, J.R. Green: Microeconomic Theory, 1995, Oxford University Press.

Articles:

[AS] A. Sen (1970): The Impossibility of a Paretian Liberal, *Journal of Political Economy*, 78, 152-157.

[SE] A. Sen (1977): Social Choice Theory: A Re-Examination, *Econometrica*, 45, 53-88.

Contents I

Contents II

Binary Relations

This part relies on [GA] and [SE].

Binary Relations

Let X be an arbitrary set (e.g. consumption bundles, candidates, allocations).
Example: $X = \{x, y, z\}$.

The Cartesian product $X \times X$ is the set of all ordered pairs of the elements of X.
For the previous example we obtain:

$$X \times X = \{(x,x),(x,y),(x,z),(y,x),(y,y),(y,z),(z,x),(z,y),(z,z)\}.$$

A binary relation R on the set X is a subset of $X \times X$, i.e. $R \subseteq X \times X$.
Example: $R = \{(x,x),(x,y),(y,y),(y,z),(z,z)\}$.

We use the following notational conventions:

- $(x, y) \in R$ is also written as xRy.
- $(x, y) \notin R$ is also written as $\neg xRy$.
- $[(x, y) \notin R \wedge (y, x) \notin R]$ is also written as $x\, nc\, y$.

Example: Magnitudes

The relation “at least as large” for numbers is a binary relation:

- $X = \mathbb{N} = \{1, 2, 3, \ldots\}$.
- $X \times X = \{(1,1), (1,2), (1,3), \ldots, (2,1), (2,2), (2,3), \ldots\}$.
- $\geq = \{(1,1), (2,1), (2,2), (3,1), (3,2), (3,3), \ldots\}$.
- $(2,1) \in \geq$ is also written as $2 \geq 1$.
- $2 \geq 1$ means that 2 is at least as large as 1.

Example: Preferences

Preferences of economic agents are binary relations:

- $X = \{a(pple), o(range), p(ear)\}$.
- $X \times X = \{(a, a), (a, o), (a, p), (o, a), (o, o), (o, p), (p, a), (p, o), (p, p)\}$.
- $\succeq = \{(a, a), (a, p), (o, a), (o, o), (o, p), (p, a), (p, p)\}$.
- $(a, p) \in \succeq$ is also written as $a \succeq p$.
- $a \succeq p$ means that the agent weakly prefers the apple over the pear.

Symmetric and Asymmetric Part

Any binary relation R on X can be partitioned into a symmetric part I and an asymmetric part P, both of which are themselves binary relations.

Formally, for all $x, y \in X$ let

$$xIy \;\Leftrightarrow\; [xRy \wedge yRx],$$

$$xPy \;\Leftrightarrow\; [xRy \wedge \neg yRx].$$

Consider the previous two examples:

- $\geq = \{(1,1),(2,1),(2,2),(3,1),(3,2),(3,3),\ldots\}$, at least as large.
  $> = \{(2,1),(3,1),(3,2),\ldots\}$, strictly larger.
  $= = \{(1,1),(2,2),(3,3),\ldots\}$, equally large.
- $\succeq = \{(a,a),(a,p),(o,a),(o,o),(o,p),(p,a),(p,p)\}$, preference.
  $\succ = \{(o,a),(o,p)\}$, strict statements.
  $\sim = \{(a,a),(a,p),(o,o),(p,a),(p,p)\}$, indifference statements.

Properties of Binary Relations

Definition

A binary relation R on a set X is

- reflexive if $xRx, \ \forall x \in X$.
- irreflexive if $\neg xRx, \ \forall x \in X$.
- complete if $xRy \vee yRx, \ \forall x, y \in X$ with $x \neq y$.
- transitive if $[xRy \wedge yRz] \Rightarrow xRz, \ \forall x, y, z \in X$.
- quasi-transitive if its induced asymmetric part P is transitive.
- acyclical if there is no finite sequence $(x_1, x_2, \ldots, x_k)$ of elements from X such that $x_1 P x_2$, $x_2 P x_3$, …, $x_{k-1} P x_k$ and $x_k P x_1$.
- symmetric if $xRy \Rightarrow yRx, \ \forall x, y \in X$.
- asymmetric if $xRy \Rightarrow \neg yRx, \ \forall x, y \in X$.
- antisymmetric if $xRy \Rightarrow \neg yRx, \ \forall x, y \in X$ with $x \neq y$.

Applications

Consider the previous two examples:

- $\geq$ is reflexive, complete, transitive, and antisymmetric.

  $>$ is irreflexive, complete, transitive, and asymmetric.

  $=$ is reflexive, transitive, and both symmetric and antisymmetric.

- $\succeq$ is reflexive, complete, and transitive.

  $\succ$ is irreflexive, transitive, and asymmetric.

  $\sim$ is reflexive, transitive, and symmetric.

Remarks

Remark

When a binary relation R on X satisfies any of these properties, then it also satisfies them when restricted to any subset $S \subseteq X$.

Remark

R is asymmetric if and only if R is antisymmetric and irreflexive.

Transitivity I

Proposition

(i) If R is transitive, then R is quasi-transitive.
(ii) If R is quasi-transitive, then R is acyclical.

Proof:

(i) Suppose R is transitive. Let P be its induced asymmetric part. Suppose further that xPy and yPz for some $x, y, z \in X$. We need to show that xPz.

By definition

$$xPy \;\Leftrightarrow\; [xRy \;\wedge \neg yRx],$$

$$yPz \;\Leftrightarrow\; [yRz \;\wedge \neg zRy].$$

But xRy and yRz implies xRz by transitivity. It remains to be shown that $\neg zRx$.

Assume to the contrary that zRx was true. Then, since yRz holds, we would also obtain yRx by transitivity, a contradiction. Hence $\neg zRx$ and thus xPz, which establishes that P is transitive, and R is therefore quasi-transitive. $\diamond$

Transitivity II

(ii) Suppose R is quasi-transitive, i.e. the induced P is transitive.
Suppose further that there exists a finite sequence $(x_1, x_2, \ldots, x_k)$ of elements from X such that $x_1 P x_2$, $x_2 P x_3$, …, $x_{k-1} P x_k$ and $x_k P x_1$.

From $x_1 P x_2$ and $x_2 P x_3$ we obtain $x_1 P x_3$ by transitivity of P.
From $x_1 P x_3$ and $x_3 P x_4$ we obtain $x_1 P x_4$ by transitivity of P.
…
Iterating, we obtain $x_1 P x_k$ after a finite number of steps, which contradicts $x_k P x_1$. Hence finite cycles cannot exist, so R is acyclical. $\diamond$ $\square$

Important Classes of Binary Relations

Definition

A binary relation R on a set X is called

- preference (or preference ordering, complete pre-ordering) if it is reflexive, complete, and transitive.
- equivalence (or equivalence relation) if it is reflexive, transitive, and symmetric.
- partial order if it is reflexive, transitive, and antisymmetric.
- linear order (or complete order) if it is reflexive, complete, transitive, and antisymmetric.

Preferences

For a given set X, we will denote by

- $\mathscr{R}$ the set of all possible preferences on X.
- $\mathscr{P}$ the set of all antisymmetric preferences (linear orders) on X.

A preference $R \in \mathscr{P} \subset \mathscr{R}$ is also called a strict preference.

If R is a preference (strict or not) then the induced

- symmetric part I is reflexive, transitive, and symmetric (hence an equivalence relation).
- asymmetric part P is irreflexive, transitive, and asymmetric.

Note that neither I nor P are themselves preferences.

Maximal and Greatest Elements I

Let R be a binary relation on X and let $S \subseteq X$.

Definition

Element $x \in S$ is called a maximal element of S with respect to R if there does not exist $y \in S$ with yPx. The set of all maximal elements is denoted by $M(S, R)$.

$M(S, R)$ can be empty, if R allows for cycles or if S is infinite.
$M(S, R)$ can contain more than one element, if R has ties or is incomplete.

Definition

Element $x \in S$ is called a greatest element of S with respect to R if xRy, $\forall y \in S$. The set of all greatest elements is denoted by $G(S, R)$.

$G(S, R)$ can be empty, if R allows for cycles or is incomplete, or if S is infinite.
$G(S, R)$ can contain more than one element, if R has ties.

Maximal and Greatest Elements II

Proposition

(i) $G(S,R) \subseteq M(S,R)$. If R is reflexive and complete, then $G(S,R) = M(S,R)$.
(ii) It holds that $M(S,R) \neq \varnothing$ for all finite and non-empty $S \subseteq X$ if and only if R is acyclical.
(iii) It holds that $G(S,R) \neq \varnothing$ for all finite and non-empty $S \subseteq X$ if and only if R is reflexive, complete and acyclical.

Proof:

(i) Let $x \in G(S,R)$, so $xRy\ \forall y \in S$. Since yPx requires $\neg xRy$, there is no $y \in S$ with yPx. Hence $x \in M(S,R)$. This implies $G(S,R) \subseteq M(S,R)$.

Now suppose R is reflexive and complete. We need to show $M(S,R) \subseteq G(S,R)$. Let $x \in M(S,R)$, so there is no $y \in S$ with yPx. By completeness and reflexivity of R, we have xRy or yRx, $\forall y \in S$. Since $\neg xRy$ would then imply yRx and thus yPx, we must have $xRy\ \forall y \in S$. Hence $x \in G(S,R)$. $\diamond$

Maximal and Greatest Elements III

(ii)

"$\Rightarrow$" Suppose $M(S, R) \neq \varnothing$ for all finite and non-empty $S \subseteq X$. By contradiction: If R was not acyclical, $\exists x_1, x_2, \dots x_k \in X$ s.t. $x_1 P x_2$, $x_2 P x_3$, ... and $x_k P x_1$, so that $M(\{x_1, x_2, \dots, x_k\}, R) = \varnothing$.

"$\Leftarrow$" Suppose R is acyclical.

By contradiction, suppose $M(S, R) = \varnothing$ for some finite and non-empty $S \subseteq X$, so for each $x \in S$ there exists $y \in S$ such that yPx. Hence we can find a strict (and finite) cycle in S, contradicting that R is acyclical. $\diamond$

(iii)

"$\Rightarrow$" Suppose $G(S, R) \neq \varnothing$ for all finite and non-empty $S \subseteq X$. By contradiction: If R was not reflexive, $\exists x \in X$ s.t. $x\,nc\,x$ and $G(\{x\}, R) = \varnothing$.

If R was not complete, $\exists x, y \in X$, $x \neq y$, s.t. $x\,nc\,y$ and $G(\{x, y\}, R) = \varnothing$.

If R was not acyclical, $\exists x_1, x_2, \dots x_k \in X$ s.t. $x_1 P x_2$, $x_2 P x_3$, ... and $x_k P x_1$, so that $G(\{x_1, x_2, \dots, x_k\}, R) = \varnothing$.

"$\Leftarrow$" Follows from (i) and (ii). $\diamond$ $\square$

Basics of Choice Theory

This part relies on [GA], [SE] and [KR].

Choice Functions

Let X be finite (from now on). Let K be the set of all non-empty subsets of X.
Example: $X = \{x, y, z\}$, $K = \{\{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\}$.

Definition

A choice function $C : K \to K$ assigns a non-empty $C(S) \subseteq S$ to each $S \in K$.

Interpretation:

- A choice function can be seen as a complete data set: it describes how an individual chooses in every conceivable choice situation S.
- If S are the available alternatives, then the individual chooses the alternatives $C(S) \subseteq S$ among them. If $C(S)$ has more than one element, this means that the individual chooses differently from time to time.
- Choice functions are the primitive of choice theory, because they are observable (as opposed to "preferences" or "utility").

Example: $C(\{x\}) = \{x\}$, $C(\{y\}) = \{y\}$, $C(\{z\}) = \{z\}$, $C(\{x, y\}) = \{x, y\}$, $C(\{x, z\}) = \{z\}$, $C(\{y, z\}) = \{z\}$, $C(\{x, y, z\}) = \{z\}$.

From Binary Relations to Choice Functions

A binary relation R might generate a choice function C, by

$$C(S) = G(S, R) \text{ for all } S \in K.$$

Proposition

A binary relation R generates a choice function C if and only if it is reflexive, complete and acyclical.

The result has been proven before, because reflexivity, completeness and acyclicity are necessary and sufficient for $G(S, R)$ to be non-empty for all $S \in K$.

We obtain as a corollary that a preference generates a choice function.

From Choice Functions to Binary Relations

A choice function C is rationalizable if there exists a binary relation R that generates it, as defined on the previous slide.

How can we find such a relation?

Given a choice function C, we can construct its base relation R_C by letting xR_Cy if and only if $x \in C(\{x, y\})$, for all $x, y \in X$. Clearly, R_C is reflexive and complete.

Proposition

C is rationalizable if and only if it is generated by its base relation R_C.

Proof:

"$\Rightarrow$" Assume R' generates C, so R' is reflexive and complete. If $R' \neq R_C$, then $\exists x, y \in X$ s.t. $G(\{x, y\}, R') \neq G(\{x, y\}, R_C) = C(\{x, y\})$, a contradiction.

"$\Leftarrow$" Trivial. $\square$

Properties of Choice Functions

Under which conditions is C rationalizable, i.e. under which conditions is it generated by R_C? And how do properties of C translate into properties of R_C?

Definition

A choice function C might have the following properties:

α: If $x \in S \subseteq T$ and $x \in C(T)$, then $x \in C(S)$, $\forall x, S, T$.

β: If $x, y \in C(S)$ and $S \subseteq T$, then $x \in C(T)$ if and only if $y \in C(T)$, $\forall x, y, S, T$.

γ: If $x \in C(S)$ for all $S \in M \subseteq K$, then $x \in C(\cup M)$, $\forall x, M$.

WARP: If $x, y \in S$ and $x, y \in T$, and if $x \in C(S)$ and $y \in C(T)$, then $x \in C(T)$, $\forall x, y, S, T$.

Each property captures aspects of consistency and rationality of choice behavior.

α is also called contraction consistency.
β is also called expansion consistency.
WARP is an abbreviation of weak axiom of revealed preference.

Results (Without Proofs)

Does R_C generate a choice function at all?

Proposition

If C satisfies α, then R_C is acyclical.

Does R_C generate the original C, i.e. is C rationalizable?

Proposition

C is rationalizable if and only if C satisfies α and γ.

Is the generating R_C actually a preference?

Proposition

C is rationalizable and R_C is a preference if and only if C satisfies α and β.

Why is WARP interesting?

Proposition

C satisfies α and β if and only if it satisfies WARP.

Part 2: The Problem of Social Choice

Motivation and Concepts

This part relies (partially) on [GA].

Example, Plurality Voting

Let us consider an example (found on Wikipedia).

Assume a group of 21 individuals must decide among alternatives $X = \{x, y, z\}$. Individual preferences are given in the following table (shorthand notation):

#	preferences
6	$x \, P \, y \, P \, z$
5	$z \, P \, x \, P \, y$
5	$y \, P \, x \, P \, z$
3	$z \, P \, y \, P \, x$
2	$y \, P \, z \, P \, x$

Let us solve the problem by plurality voting, i.e. each individual casts a vote for the preferred alternative. We obtain 6 votes for x, 7 votes for y, and 8 votes for z.

The social ranking is therefore $zPyPx$, and alternative z wins.

Example, Plurality Voting with Runoff

Alternative z wins with a relative majority (8 votes), but it fails to achieve the absolute majority (11 votes).

Before selecting z as a winner, we might require a runoff vote between z and y. We obtain 13 votes for y and 8 votes for z, so y wins.

Consider the Ständeratswahl Ticino 2019, where two candidates were elected:

- In the first round, no candidate achieved an absolute majority (52'882 votes):
  Filippo Lombardi (CVP, 34'318), Marco Chiesa (SVP, 32'576),
  Giovanni Merlini (FDP, 30'371), Marina Carobbio Guscetti (SP, 30'263),
  Greta Gysin (GPS, 22'012), Battista Ghiggia (Lega, 20'546).
- In the second round, only four candidates remained:
  Marco Chiesa (SVP, 42'552), Marina Carobbio Guscetti (SP, 36'469)
  Filippo Lombardi (CVP, 36'424), Giovanni Merlini (FDP, 33'278).

Example, Pairwise Majority Voting

Why not try a completely different procedure? With pairwise majority voting, we say that one alternative is better than another when it wins the pairwise vote.

x against y yields 11 against 10 votes, so x is better than y.

x against z yields 11 against 10 votes, so x is better than z.

y against z yields 13 against 8 votes, so y is better than z.

The social ranking is therefore $xPyPz$, and alternative x wins.
This is exactly the opposite of plurality voting!

Alternative x is the Condorcet winner (Marquis de Condorcet, 1743-1794), as it wins in every pairwise comparison. Alternative z is the Condorcet loser.

A method satisfies the Condorcet criterion (or is a Condorcet method) if it selects or uniquely ranks on top the Condorcet winner, whenever one exists.

Example, Borda Count

Maybe it is helpful to try a final method. With Borda count, each individual assigns 2 points to the most preferred alternative, 1 point to the second, and 0 points to the worst alternative.

#	preferences	x	y	z
6	$x\ P\ y\ P\ z$	2	1	0
5	$z\ P\ x\ P\ y$	1	0	2
5	$y\ P\ x\ P\ z$	1	2	0
3	$z\ P\ y\ P\ x$	0	1	2
2	$y\ P\ z\ P\ x$	0	2	1
		22	23	18

The social ranking is therefore $yPxPz$, which is again a different result! Alternative y wins.

Example, Conclusions

The example illustrates several interesting points:

- There is more than one way of making democratic decisions.
- In the given example, every alternative wins for some voting method. Therefore, the voting method has a profound impact on the result.
- It is necessary (and strategically helpful) to examine the properties of voting methods in greater detail.
- Common methods such as plurality voting can have severe drawbacks. In the example, plurality voting selects the Condorcet loser, i.e. the winner would be defeated by any of the losing alternatives in a pairwise vote.

The Formal Problem I

- X is a set of alternatives (initiatives, candidates, tax rates, boxers...)
  X is often finite, and m then denotes the number of alternatives.
- $N = \{1, \ldots, n\}$, $n \geq 2$, is a set of voters (citizens, referees...)
- $\mathscr{R}$ is the set of all possible preferences on X.
- $\mathscr{P}$ is the set of all possible strict preferences on X.
- $\mathscr{C}$ is the set of all reflexive, complete and acyclical binary relations on X.
- Hence we have $\mathscr{P} \subset \mathscr{R} \subset \mathscr{C}$.
- We denote by $R_i \in \mathscr{R}$ the preference of voter i,
  and by P_i and I_i its induced asymmetric and symmetric parts.
- We denote by $\mathbf{R} = (R_1, \ldots, R_n) \in \mathscr{R}^n$ a profile of preferences of all voters.

The Formal Problem II

Let $\mathscr{A} \subseteq \mathscr{R}^n$ denote some set of admissible preference profiles.

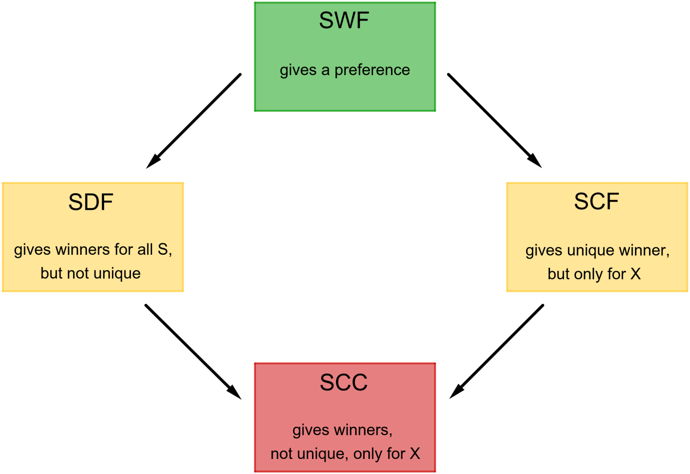
Definition

- A social welfare function (SWF) is a mapping $f : \mathscr{A} \to \mathscr{R}$.
- A social decision function (SDF) is a mapping $d : \mathscr{A} \to \mathscr{C}$.
- A social choice function (SCF) is a mapping $c : \mathscr{A} \to X$.
- A social choice correspondence (SCC) is a mapping $\gamma : \mathscr{A} \to 2^X$.

Interpretation

- An SWF is a rule or voting method that delivers a preference of society $f(\mathbf{R}) \in \mathscr{R}$ for every possible profile of preferences of the voters $\mathbf{R} \in \mathscr{A}$.
- An SDF is a rule or voting method that only delivers a reflexive, complete and acyclical binary relation $d(\mathbf{R}) \in \mathscr{C}$.
  The relation $d(\mathbf{R})$ is not necessarily transitive and hence not a preference, but it still generates non-empty social choices (see Part 1 of the lecture).
- An SCF is a rule or voting method that delivers a winning alternative $c(\mathbf{R}) \in X$, instead of a social ranking.
- An SCC is a rule or voting method that only delivers a set of winners $\gamma(\mathbf{R}) \in 2^X$ (or, equivalently, $\gamma(\mathbf{R}) \subseteq X$), but not necessarily a unique winner.

Connection Between Concepts



Back to the Example

In the previous example...

- ...plurality voting delivered a social preference (even a strict one).
  This is true in general: plurality voting is an SWF on any domain $\mathscr{A}$.
- ...plurality voting with a runoff was designed to select a single winner.
  Hence it is an SCF, on any domain $\mathscr{A}$ (tie-breaking might be necessary).
- ...pairwise majority voting delivered a social preference (even a strict one).
  However, this is not true in general (see Exercise 2.1).
- ...Borda count delivered a social preference (even a strict one).
  This is true in general: Borda count is an SWF on any domain $\mathscr{A}$.

Voting Methods

This part relies (partially) on [GA].

Even More Methods

We will now look at the following methods in greater detail:

- Plurality voting (PV)
- Instant-runoff voting (IR)
- Pairwise majority voting (PM)
- Copeland method (CO)
- Borda count (BC)
- Pareto efficiency method (EM)
- Pareto extension rule (PE)

There are many more rules:
Black's method, Dodgson method, Kemeny-Young method, Blackballing...

Plurality Voting (PV)

With plurality voting, each voter casts a vote for the most preferred alternative. We use the name majority voting when there are only two alternatives ($m = 2$).

If a voter has $a \geq 2$ best alternatives, we could either assume a tie-breaking rule (e.g. alphabetical order) or split the vote into a votes of size $1/a$ each.

Formal definition:
Assume $\mathbf{R} \in \mathscr{P}^n$ for simplicity. Then $N(x, \mathbf{R}) = |\{i \in N | \, xP_iz \; \forall z \neq x\}|$ is the number of votes for $x \in X$. A social preference $f^{PV}(\mathbf{R})$ can then be defined by

$$xf^{PV}(\mathbf{R})y \;\; \leftrightarrow \;\; N(x, \mathbf{R}) \geq N(y, \mathbf{R}),$$

for all $x, y \in X$. This is clearly reflexive, complete and transitive, so that PV is in fact an SWF on $\mathscr{A} = \mathscr{P}^n$ (and also on $\mathscr{A} = \mathscr{R}^n$).

Discussion:

PV violates the Condorcet criterion (think of Bush, Gore and Nader). The winner might be defeated in a pairwise vote .

Instant-Runoff Voting (IR)

With instant-runoff voting (a version of single transferable vote), voters submit their complete preference rankings. The outcome of plurality voting is then determined first. If an alternative obtains the absolute majority, it wins. If not, the alternative with the smallest number of votes is eliminated, all preferences are updated, and plurality voting takes place again. This procedure is repeated until an alternative gets the absolute majority.

In a conventional runoff vote, candidates without the absolute majority have to be voted upon again, after worse candidates have been eliminated (Ständerat, French presidential election). IR implies that runoffs take place automatically.

Discussion:

Since IR selects a winner, it is an SCF.

IR also violates the Condorcet criterion (see Exercise 2.4).

Pairwise Majority Voting (PM)

With pairwise majority voting, each pair of alternatives is voted on once. An alternative is socially better than those which it defeats in these pairwise votes.

Formal definition:
Let $N(x, y, \mathbf{R}) = |\{i \in N | \, xP_i y\}|$ and define

$$xf^{PM}(\mathbf{R})y \;\; \leftrightarrow \;\; N(x, y, \mathbf{R}) \geq N(y, x, \mathbf{R}),$$

for all $x, y \in X$.

Discussion:

If a Condorcet winner exists, it will be ranked above any other candidate, so PM satisfies the Condorcet criterion.

However, as seen earlier, $f^{PM}(\mathbf{R})$ might not be transitive, and not even acyclical (paradox of voting). Hence PM is in general not an SWF (SDF, SCF, SCC), so using the notation f is slightly abusive.

Copeland Method (CO)

For the Copeland method (Arthur Copeland) we run all pairwise votes, as for PM. An alternative gets $+1$ point for every pairwise vote that it wins, and -1 point for every pairwise vote that it loses. Alternatives are then ranked according to the sum of these points.

Formal definition:

$$CO(x, \mathbf{R}) = |\{y \in X | N(x, y, \mathbf{R}) > N(y, x, \mathbf{R})\}| - |\{y \in X | N(x, y, \mathbf{R}) < N(y, x, \mathbf{R})\}|$$

and

$$x f^{CO}(\mathbf{R}) y \quad \leftrightarrow \quad CO(x, \mathbf{R}) \geq CO(y, \mathbf{R}),$$

for all $x, y \in X$. This is clearly reflexive, complete and transitive, so that CO is in fact an SWF on $\mathscr{A} = \mathscr{R}^n$.

Discussion:

The final ranking is based on scores $CO(., \mathbf{R})$, so transitivity problems are avoided.

If x is the Condorcet winner, then $CO(x, \mathbf{R}) = m - 1$ while $CO(y, \mathbf{R}) < m - 2$ for all $y \neq x$. Hence x is ranked top, and CO satisfies the Condorcet criterion.

Sports leagues are often ranked based on similar methods.

Borda Count (BC)

With Borda count (Jean-Charles de Borda, 1733-1799) each voter casts $m - k$ votes for the kth-best alternative ($m - 1$ for the best ... and 0 for the worst).

If a alternatives tie for ranks k to $k + a - 1$, each gets $(\sum_{j=k}^{k+a-1}(m - j))/a$ votes.

Formal definition:

Assume $\mathbf{R} \in \mathscr{P}^n$ for simplicity. Let $r(x, R_i)$ denote the rank of x in R_i, i.e. $r(x, R_i) = 1$ if x is the best in R_i ... and $r(x, R_i) = m$ if x is the worst in R_i.

Then alternative x gets $BC(x, \mathbf{R}) = \sum_{i=1}^{n}(m - r(x, R_i))$ votes and we have

$$xf^{BC}(\mathbf{R})y \quad \leftrightarrow \quad BC(x, \mathbf{R}) \geq BC(y, \mathbf{R}),$$

for all $x, y \in X$. This is clearly reflexive, complete and transitive, so that BC is in fact an SWF on $\mathscr{A} = \mathscr{P}^n$ (and also on $\mathscr{A} = \mathscr{R}^n$).

Discussion:

BC has the advantage that the complete preference profiles are taken into account. However, it violates the Condorcet criterion.

Pareto Efficiency Method (EM)

For the Pareto efficiency method, all voters submit their preferences and we divide alternatives into two classes, those which are Pareto efficient and those which are Pareto inefficient. Society is indifferent between alternatives within the same class, but strictly prefers efficient over inefficient alternatives.

Formal definition:

Let $X^I(\mathbf{R}) = \{x \in X | \exists y \in X \text{ with } yR_i x \ \forall i \in N \text{ and } yP_j x \text{ for at least one } j \in N\}$ and $X^E(\mathbf{R}) = X \backslash X^I(\mathbf{R})$. Define, for all $x, y \in X$,

$$xf^{EM}(\mathbf{R})y \ \text{ except if } \ y \in X^E(\mathbf{R}) \text{ and } x \in X^I(\mathbf{R}).$$

This is reflexive, complete and transitive, so that EM is an SWF on $\mathscr{A} = \mathscr{R}^n$.

Discussion:

EM violates the Condorcet criterion (see Exercise 2.4).

It is also unsatisfactory because it does not solve conflicts of interest.

Pareto Extension Rule (PE)

The Pareto extension rule compares alternatives pairwise. Society is indifferent between two alternatives except if one Pareto dominates the other, in which case the Pareto better alternative is strictly preferred. In terms of voting, unanimity is required in the pairwise vote, for one alternative to be ranked above another.

Formal definition:

For all $x, y \in X$, we define

$$x d^{PE}(\mathbf{R}) y \quad \text{except if} \quad yR_i x \ \forall i \in N \text{ and } yP_j x \text{ for at least one } j \in N.$$

Discussion:

PE violates the Condorcet criterion (see Exercise 2.4).

Similar to EM, PE does not solve conflicts of interest.

As the notation d^{PE} already indicates, PE is an SDF on $\mathscr{A} = \mathscr{R}^n$, not an SWF. Relation $d^{PE}(\mathbf{R})$ is reflexive, complete and acyclical, but not necessarily transitive.

PE: Non-Transitivity

It is easy to see that $d^{PE}(\mathbf{R})$ is reflexive and complete.

To see that $d^{PE}(\mathbf{R})$ is not necessarily transitive, consider the example

#	preferences
1	$y \; P \; z \; P \; x$
1	$z \; P \; x \; P \; y$

We obtain $xd^{PE}(\mathbf{R})y$ and $yd^{PE}(\mathbf{R})z$ but $\neg xd^{PE}(\mathbf{R})z$.

PE: Acyclicity

Proposition

For any $\mathbf{R} \in \mathscr{R}^n$, the binary relation $d^{PE}(\mathbf{R})$ is quasi-transitive, hence acyclical.

Proof:

Suppose $x d_P^{PE}(\mathbf{R}) y$ and $y d_P^{PE}(\mathbf{R}) z$, where $d_P^{PE}(\mathbf{R})$ is the strict part of $d^{PE}(\mathbf{R})$.
By definition of d^{PE} this requires that

$$xR_iy \; \forall i \in N \text{ and } xP_jy \text{ for at least one } j \in N,$$

as well as

$$yR_iz \; \forall i \in N \text{ and } yP_jz \text{ for at least one } j \in N.$$

By transitivity of all R_i (i.e. $\mathbf{R} \in \mathscr{R}^n$), we therefore have

$$xR_iz \; \forall i \in N \text{ and } xP_jz \text{ for at least one } j \in N,$$

which implies $x d_P^{PE}(\mathbf{R}) z$ and thus quasi-transitivity of d^{PE}. $\square$

Scoring Methods and Pairwise Comparison Methods

This part relies (partially) on [GA].

Scoring Methods

Let $\mathbf{R} \in \mathscr{P}^n$ for simplicity. An SWF f is a scoring method if it works as follows:

- There is a vector $s = (s_1, s_2, \ldots, s_m)$ of scores, with $s_1 \geq s_2 \geq \ldots \geq s_m$.
- Voter i assigns s_k votes to her kth-best alternative.
  Formally, alternative x gets $s_{r(x,R_i)}$ votes from voter i.
- Votes are added across voters, so x obtains an overall score of

$$SC(x, \mathbf{R}) = \sum_{i=1}^{n} s_{r(x,R_i)}.$$

- Alternatives are ranked according to these scores, i.e.

$$xf(\mathbf{R})y \quad \leftrightarrow \quad SC(x, \mathbf{R}) \geq SC(y, \mathbf{R}).$$

- If individual preferences are not strict, we need a rule how to split votes.

Examples of Scoring Methods

The following methods are examples of scoring methods:

- Borda count: $s = (m-1, m-2, \ldots, 0)$
- Plurality voting: $s = (1, 0, \ldots, 0)$
- Anti-plurality voting: $s = (1, \ldots, 1, 0)$
- Rejection voting: $s = (0, \ldots, 0, -1)$
- Nameless example I: $s = (2, 1, \ldots, 1, 0)$
- Nameless example II: $s = (1, 0, \ldots, 0, -1)$

Quiz: Do you know where $s = (12, 10, 8, 7, 6, 5, 4, 3, 2, 1, 0, \ldots, 0)$ is used?

Discussion of Scoring Methods I

Scoring methods are well-behaved when voters are added.

For instance, if x is ranked top by a given scoring method in two separate groups of voters, then the same holds for the joint group (because scores are added).

A generalized version of this property is called consistency.

In contrast, consider PV with a runoff in the following example:

#	group 1
6	$x\ P\ y\ P\ z$
4	$y\ P\ z\ P\ x$
3	$z\ P\ y\ P\ x$

#	group 2
3	$x\ P\ y\ P\ z$
4	$y\ P\ z\ P\ x$
6	$z\ P\ y\ P\ x$

#	joint
9	$x\ P\ y\ P\ z$
8	$y\ P\ z\ P\ x$
9	$z\ P\ y\ P\ x$

In group 1, x and y enter the runoff and y wins.
In group 2, z and y enter the runoff and y wins.
In the joint group, x and z enter the runoff and z wins.

Discussion of Scoring Methods II

Scoring methods are less well-behaved when alternatives are added.

For instance, BC is sensitive to manipulation by cloning of candidates. Consider

#	preferences	x	y	z
4	$x\ P\ y\ P\ z$	2	1	0
3	$z\ P\ y\ P\ x$	0	1	2
		8	7	6

where BC yields the social preference $xPyPz$.
Now suppose the 3 voters who prefer z nominate an additional clone $\hat{z}$ of z.

#	preferences	x	y	z	$\hat{z}$
4	$x\ P\ y\ P\ z\ P\ \hat{z}$	3	2	1	0
3	$z\ P\ \hat{z}\ P\ y\ P\ x$	0	1	3	2
		12	11	13	6

BC now yields $zPxPyP\hat{z}$. Introducing the clone moves z on top.

Scoring methods also tend to violate "independence of irrelevant alternatives".

Discussion of Scoring Methods III

Consider a first scoring method with $s^1 = (s^1_1, s^1_2, \ldots, s^1_m)$.

Construct a second method with $s^2 = (s^2_1, s^2_2, \ldots, s^2_m)$ by $s^2_k = a \cdot s^1_k + b$, for arbitrary scalars $a > 0$ and b and all $k = 1, \ldots, m$.

Then, for any x and $\mathbf{R}$ we have

$$SC^2(x, \mathbf{R}) = \sum_{i=1}^{n} s^2_{r(x,R_i)} = \sum_{i=1}^{n} \left(a \cdot s^1_{r(x,R_i)} + b \right) = a \cdot SC^1(x, \mathbf{R}) + n \cdot b.$$

Hence the scores coincide up to a positive affine transformation. The two methods induce the same social ranking over alternatives, and are therefore equivalent.

Discussion of Scoring Methods IV

BC (and equivalent methods) plays a central role in the class of scoring methods.

It has the property that voters can express their full preference ranking in the vote.

This is not true for PV, for instance, where preferences over non-top-ranked alternatives cannot be expressed.

Pairwise Comparison Methods

With a pairwise comparison method, the social ranking between any two alternatives x and y depends only on how x and y are ranked in $\mathbf{R}$, not on the relative ranking to other alternatives.

Among the previous methods, PM and PE are pairwise comparison methods.

Advantages:

- simplicity of pairwise comparisons
- satisfy "independence of irrelevant alternatives"

Disadvantage:

- transitivity problems

Note 1: The literature also knows "pairwise methods", which are different.
Note 2: Some methods are neither scoring nor pairwise comparison (IR, CO, EM).

Part 3: Arrow's Theorem

Impossibility for SWFs

This part relies on [MC].

Introduction

Where we stand:

- We know what social choice theory is about.
- We have become acquainted with several voting methods.
- We have discussed some of their properties and problems.
- We have developed a formal framework for studying preference aggregation.

What we do next:

- We take a more systematic approach to the comparison of SWFs.
- We postulate plausible properties and try to find SWFs that satisfy them.

Reminder

- X is the set of alternatives, with $m = |X|$.
- $N = \{1, \ldots, n\}$, $n \geq 2$ is the set of voters.
- $\mathscr{R}$ is the set of possible preferences on X.
- Voter i's preference is denoted by R_i.

  Its asymmetric part is P_i, its symmetric part is I_i.
- The profile of preferences of all voters is denoted by $\mathbf{R} = (R_1, \ldots, R_n)$.
- Let $\mathscr{A} \subseteq \mathscr{R}^n$ be a set of admissible preference profiles. A social welfare function is a mapping $f : \mathscr{A} \to \mathscr{R}$, which assigns to every admissible profile of individual preferences $\mathbf{R} \in \mathscr{A}$ a social preference $f(\mathbf{R}) \in \mathscr{R}$.
- The asymmetric part of $f(\mathbf{R})$ is $f_P(\mathbf{R})$, its symmetric part is $f_I(\mathbf{R})$.

Arrow's Axioms, Intuitively

What are plausible properties that an SWF f should satisfy?

Kenneth Arrow has proposed the following:

[U] Universality, also called Universal Domain:

The SWF should work for all conceivable preference profiles.

Besides *practical reasons*, this incorporates a notion of *tolerance*.

[I] Independence of Irrelevant Alternatives:

The social ranking of two alternatives should depend only on how the individuals rank these two alternatives (pairwise comparison method).

Besides *practical reasons*, this alleviates problems of *strategic nomination*.

[P] Weak Pareto Principle:

If all individuals strictly prefer x over y, this should also hold for society.

This means that the SWF should *respect agreement* by all individuals.

[D] Non-Dictatorship:

No individual should be able to always impose the own preference on society.

This is a *minimal democratic requirement*.

Arrow's Axioms, Formally

Definition

[U] $\mathscr{A} = \mathscr{R}^n$

[I] For any pair of alternatives $x, y \in X$, if two profiles $\mathbf{R}, \mathbf{R}' \in \mathscr{A}$ satisfy

$$xR_iy \;\leftrightarrow\; xR'_iy \text{ and } yR_ix \;\leftrightarrow\; yR'_ix \text{ for all } i \in N,$$

then

$$xf(\mathbf{R})y \;\leftrightarrow\; xf(\mathbf{R}')y \;\text{ and }\; yf(\mathbf{R})x \;\leftrightarrow\; yf(\mathbf{R}')x$$

must be true.

[P] For any pair of alternatives $x, y \in X$, if a profile $\mathbf{R} \in \mathscr{A}$ satisfies

$$xP_iy \text{ for all } i \in N,$$

then $xf_P(\mathbf{R})y$ must be true.

[D] There is no individual $i \in N$ such that, for all $x, y \in X$ and $\mathbf{R} \in \mathscr{A}$, xP_iy implies $xf_P(\mathbf{R})y$.

Arrow's Axioms, Remarks

- We could replace axiom [U] by the weaker requirement $\mathscr{A} = \mathscr{P}^n$, without getting different results in the following.
- Axiom [P] is called *weak* Pareto principle because it imposes a requirement on f only in case all individuals have a strict preference between x and y.

  The *strong* Pareto principle would require $xf_P(\mathbf{R})y$ already when xR_iy for all $i \in N$ and xP_jy for at least one $j \in N$. This is stronger than [P] because it imposes a requirement in more cases.
- According to axiom [D], a dictator is someone who can impose the own *strict* preference on society (someone else might decide when the dictator is indifferent).
- The axioms are normative requirements. You don't have to agree with them.

Arrow's Impossibility Theorem

Theorem

If $m \geq 3$, there is no SWF f that satisfies [U], [I], [P] and [D].

- Arrow's theorem is the most important result in social choice theory.
- It shows that preference aggregation faces fundamental logical problems.
- If we require [U], [I] and [P], then we end up with a dictatorship.

Arrow's Theorem: Structure of Proof I

We will now prove Arrow's impossibility theorem. We fix an arbitrary SWF f and assume it satisfies [U], [I] and [P]. We will then show that it must violate [D].

We will need the following definitions:

Definition

Given an SWF f, a subset of voters $S \subseteq N$ is

- decisive for x over y, for some $x, y \in X$, if

$$xP_iy \ \ \forall i \in S \ \text{ and } \ yP_ix \ \ \forall i \in N \backslash S \ \text{ implies } \ xf_P(\mathbf{R})y.$$

- completely decisive for x over y, for some $x, y \in X$, if

$$xP_iy \ \ \forall i \in S \ \text{ implies } \ xf_P(\mathbf{R})y.$$

- decisive if S is decisive for x over y for all $x, y \in X$.
- completely decisive if S is completely decisive for x over y for all $x, y \in X$.

Observe that i is a dictator if and only if $S = \{i\}$ is completely decisive.

Arrow's Theorem: Structure of Proof II

The proof has 7 steps. Given an SWF that satisfies [U], [I] and [P], we show:

1. If some S is decisive for x over y for some $x, y \in X$, then S is decisive.
2. If both S and T are decisive, then the intersection $S \cap T$ is decisive as well.
3. For any S, either S or $N \backslash S$ is decisive.
4. If S is decisive and $S \subseteq T$, then T is also decisive.

Based on these insights we can then show:

5. If S with $|S| \geq 2$ is decisive, there exists an $S' \subset S$ that is also decisive.
6. There exists an $h \in N$ such that $\{h\}$ is decisive.
7. Set $\{h\}$ from step 6 is completely decisive, hence h is a dictator.

Throughout, we will indicate in red where our presumptions enter.

Arrow's Theorem: Proof I

Step 1

If some S is decisive for x over y for some $x, y \in X$, then S is decisive.

Proof:

Claim A: If S is decisive for x over y, then S is decisive for x over z, $\forall z \neq x$.
Proof of Claim A:
Fix any $z \in X$ with $z \neq x$ and $z \neq y$ (using $m \geq 3$).
Consider $\mathbf{R} \in \mathscr{R}^n$ (using [U]) such that

$$xP_iy, yP_iz, xP_iz \;\; \forall i \in S \;\; \text{and} \;\; yP_iz, zP_ix, yP_ix \;\; \forall i \in N \backslash S.$$

Decisiveness of S for x over y implies $xf_P(\mathbf{R})y$.
[P] implies $yf_P(\mathbf{R})z$.
We then have $xf_P(\mathbf{R})z$ by transitivity of $f(\mathbf{R})$ (using $f(\mathbf{R}) \in \mathscr{R}$).
By [I], $xf_P(\mathbf{R}')z$ must then also hold in any other $\mathbf{R}'$ where

$$xP'_iz \;\; \forall i \in S \;\; \text{and} \;\; zP'_ix \;\; \forall i \in N \backslash S.$$

Hence S is also decisive for x over z. $\diamond$

Arrow's Theorem: Proof II

Claim B: If S is decisive for x over y, then S is decisive for z over y, $\forall z \neq y$.
Proof of Claim B:
Fix any $z \in X$ with $z \neq x$ and $z \neq y$ (using $m \geq 3$).
Consider $\mathbf{R} \in \mathscr{R}^n$ (using [U]) such that

$$zP_ix, xP_iy, zP_iy \;\; \forall i \in S \;\; \text{and} \;\; yP_iz, zP_ix, yP_ix \;\; \forall i \in N \backslash S.$$

Decisiveness of S for x over y implies $xf_P(\mathbf{R})y$.
[P] implies $zf_P(\mathbf{R})x$.
We then have $zf_P(\mathbf{R})y$ by transitivity of $f(\mathbf{R})$ (using $f(\mathbf{R}) \in \mathscr{R}$).
By [I], $zf_P(\mathbf{R}')y$ must then also hold in any other $\mathbf{R}'$ where

$$zP_i'y \;\; \forall i \in S \;\; \text{and} \;\; yP_i'z \;\; \forall i \in N \backslash S.$$

Hence S is also decisive for z over y. $\diamond$

To complete the proof of Step 1, we still have to show that S is decisive for arbitrary $v \in X$ over arbitrary $w \in X$.

Arrow's Theorem: Proof III

Suppose S is decisive for x over y. Let z, with $z \neq x$ and $z \neq y$, be a third alternative ($m \geq 3$). Let $v, w \in X$, $v \neq w$, be two arbitrary alternatives.

If $v = z$, then S is decisive for v over y by Claim B, and decisive for v over w by a repeated application of Claim A.

If $w = z$, then S is decisive for x over w by Claim A, and decisive for v over w by a repeated application of Claim B.

If $v \neq z$ and $w \neq z$, then S is decisive for z over y by Claim B, hence decisive for z over w by a repeated application of Claim A, hence decisive for v over w by a repeated application of Claim B.

We have exhausted all cases, so S is decisive for v over w. Hence S is decisive. $\square$

Arrow's Theorem: Proof IV

Step 2

If both S and T are decisive, then the intersection $S \cap T$ is decisive as well.

Proof:

Let $x, y, z \in X$ be three distinct alternatives ($m \geq 3$).
Consider $\mathbf{R} \in \mathscr{R}^n$ (using [U]) such that

$$\begin{array}{ll} xP_iz, zP_iy, xP_iy & \forall i \in S \cap T, \\ zP_iy, yP_ix, zP_ix & \forall i \in S \backslash (S \cap T), \\ yP_ix, xP_iz, yP_iz & \forall i \in T \backslash (S \cap T), \\ yP_iz, zP_ix, yP_ix & \forall i \in N \backslash (S \cup T). \end{array}$$

Since $[S \cap T] \cup [S \backslash (S \cap T)] = S$ is decisive, we have $zf_P(\mathbf{R})y$.
Since $[S \cap T] \cup [T \backslash (S \cap T)] = T$ is decisive, we have $xf_P(\mathbf{R})z$.
We then also have $xf_P(\mathbf{R})y$, by transitivity of $f(\mathbf{R})$ (using $f(\mathbf{R}) \in \mathscr{R}$).
By [I], $xf_P(\mathbf{R}')y$ must then also hold in any other $\mathbf{R}'$ where

$$xP'_iy \;\; \forall i \in S \cap T \;\; \text{and} \;\; yP'_ix \;\; \forall i \in N \backslash (S \cap T).$$

Hence $S \cap T$ is decisive for x over y, hence decisive by Step 1. $\square$

Arrow's Theorem: Proof V

Step 3

For any S, either S or $N \backslash S$ is decisive.

Proof:

Let $x, y, z \in X$ be three distinct alternatives ($m \geq 3$).
Consider $\mathbf{R} \in \mathscr{R}^n$ (using [U]) such that

$$\begin{array}{ll} xP_iz, zP_iy, xP_iy & \forall i \in S, \\ yP_ix, xP_iz, yP_iz & \forall i \in N \backslash S. \end{array}$$

We must have $xf_P(\mathbf{R})y$ or $yf(\mathbf{R})x$, by completeness of $f(\mathbf{R})$ (using $f(\mathbf{R}) \in \mathscr{R}$).

If $xf_P(\mathbf{R})y$, then S is decisive for x over y by [I], hence decisive by Step 1.

If $yf(\mathbf{R})x$, then we must have $yf_P(\mathbf{R})z$, because $xf_P(\mathbf{R})z$ by [P] and transitivity of $f(\mathbf{R})$ (using $f(\mathbf{R}) \in \mathscr{R}$). But then $N \backslash S$ is decisive for y over z by [I], hence decisive by Step 1. □

Arrow's Theorem: Proof VI

Step 4

If S is decisive and $S \subseteq T$, then T is also decisive.

Proof:

The empty set $\varnothing$ is not decisive by [P].

Now suppose $N \setminus T$ was decisive. Then $(N \setminus T) \cap S = \varnothing$ must be decisive by Step 2, a contradiction. Hence $N \setminus T$ is not decisive.

But then $N \setminus (N \setminus T) = T$ is decisive by Step 3. $\square$

Arrow's Theorem: Proof VII

Step 5

If S with $|S| \geq 2$ is decisive, there exists an $S' \subset S$ that is also decisive.

Proof:

Take any $h \in S$. If $S' = S \backslash \{h\}$ is decisive, we are done.

Hence assume $S \backslash \{h\}$ is not decisive. Then $N \backslash (S \backslash \{h\}) = (N \backslash S) \cup \{h\}$ is decisive by Step 3. Then $S' = S \cap ((N \backslash S) \cup \{h\}) = \{h\}$ is decisive by Step 2. $\square$

Step 6

There exists an $h \in N$ such that $\{h\}$ is decisive.

Proof:

The complete set N is decisive by [P]. Since N is finite, we can iterate Step 5 until arriving at a singleton decisive set $\{h\}$. $\square$

Arrow's Theorem: Proof VIII

Step 7

Set $\{h\}$ from step 6 is completely decisive, hence h is a dictator.

Proof:

Let $x, y, z \in X$ be three distinct alternatives ($m \geq 3$) and let $T \subseteq N \backslash \{h\}$.
Consider $\mathbf{R} \in \mathscr{R}^n$ (using [U]) such that

$$\begin{array}{ll} xP_h z, zP_h y, xP_h y & \\ xR_i y, yP_i z, xP_i z & \forall i \in T, \\ yP_i z, zP_i x, yP_i x & \forall i \in N \backslash (T \cup \{h\}). \end{array}$$

$\{h\}$ is decisive, so $zf_P(\mathbf{R})y$. $\{h\} \cup T$ is decisive by Step 4, so $xf_P(\mathbf{R})z$.
Then $xf_P(\mathbf{R})y$ by transitivity of $f(\mathbf{R})$ (using $f(\mathbf{R}) \in \mathscr{R}$).

Since T was arbitrary and by [I], $\{h\}$ is completely decisive for x over y.
Since x and y were arbitrary, $\{h\}$ is completely decisive, so h is a dictator. □

This completes the proof of Arrow's impossibility theorem for SWFs.

Impossibility for SCFs

This part relies on [MC].

A Solution?

According to Arrow, any SWF must violate plausible axioms when $m \geq 3$.

In many cases we might not need a complete social preference $f(\mathbf{R}) \in \mathscr{R}$.
If we are satisfied with a winner $c(\mathbf{R}) \in X$, maybe there is a solution?

In other words, can we get a possibility result for social choice functions (SCFs)?

Reminder:
Let $\mathscr{A} \subseteq \mathscr{R}^n$ be a set of admissible preference profiles. A social choice function is a mapping $c : \mathscr{A} \to X$, which assigns to every admissible profile of individual preferences $\mathbf{R} \in \mathscr{A}$ a social winner $c(\mathbf{R}) \in X$.

Arrow's Axioms for SCFs, Intuitively

Since we consider a different type of rule, we need to adjust Arrow's axioms.

[$\bar{\text{U}}$] Universality:

As before, the SCF should work for all conceivable preference profiles.

[$\bar{\text{M}}$] Monotonicity:

Monotonicity replaces Independence of Irrelevant Alternatives. The winner should be unaffected if preferences change for irrelevant alternatives.

If individual preferences change, but the previously winning alternative does not drop in anyone's preference, then it should remain the winner.

[$\bar{\text{P}}$] Weak Pareto Principle:

If all individuals strictly prefer x over y, then y should not be the winner.

[$\bar{\text{D}}$] Non-Dictatorship:

There should be no individual such that the winner is always one of the top-ranked alternatives of that individual.

Arrow's Axioms for SCFs, Formally

Definition

An alternative $x \in X$ maintains its position from $\mathbf{R}$ to $\mathbf{R}'$ if xR_iy implies xR'_iy and xP_iy implies xP'_iy, for all $i \in N$ and $y \in X$.

Definition

- [$\bar{\text{U}}$] $\mathscr{A} = \mathscr{R}^n$
- [$\bar{\text{M}}$] For all $x \in X$ and $\mathbf{R}, \mathbf{R}' \in \mathscr{A}$, if $x = c(\mathbf{R})$ and x maintains its position from $\mathbf{R}$ to $\mathbf{R}'$, then $x = c(\mathbf{R}')$ must hold.
- [$\bar{\text{P}}$] For any pair of alternatives $x, y \in X$, if a profile $\mathbf{R} \in \mathscr{A}$ satisfies xP_iy for all $i \in N$, then $y \neq c(\mathbf{R})$ must be true.
- [$\bar{\text{D}}$] There is no $i \in N$ such that, for all $\mathbf{R} \in \mathscr{A}$, $c(\mathbf{R})R_iy$ holds for all $y \in X$.

Arrow's Axioms for SCFs, Remarks

- We could again replace [$\bar{\text{U}}$] by the weaker requirement $\mathscr{A} = \mathscr{P}^n$, without getting different results.
- You will find a slightly different definition of [$\bar{\text{M}}$] in [MC], because the underlying definition of "maintaining position" is slightly different. With this definition, a dictatorship would never be monotonic.
- Axiom [$\bar{\text{P}}$] is again a *weak* Pareto principle, because it still allows the winner to be weakly Pareto dominated by some other alternative.

  The *strong* Pareto principle would require $y \neq c(\mathbf{R})$ already when xR_iy for all $i \in N$ and xP_jy for at least one $j \in N$.
- According to [$\bar{\text{D}}$], a dictator can impose *one of* his most preferred alternatives on society (someone else might decide in case of indifference).

Arrow's Impossibility Theorem for SCFs

Theorem

If $m \geq 3$, there is no SCF c that satisfies [$\bar{\text{U}}$], [$\bar{\text{M}}$], [$\bar{\text{P}}$] and [$\bar{\text{D}}$].

Moving from SWFs to the weaker concept of SCFs does not solve the problem.

The proof will illustrate that the concept of an SCF is in fact *not* much weaker than the concept of an SWF.

Arrow's Theorem for SCFs: Structure of Proof I

We will again prove the theorem. We fix an arbitrary SCF c and assume it satisfies [$\bar{\text{U}}$], [$\bar{\text{M}}$] and [$\bar{\text{P}}$]. We will then show that it must violate [$\bar{\text{D}}$].

We will need the following definition:

Definition

Let $X' \subseteq X$ and $\mathbf{R} \in \mathscr{R}^n$. The profile $\mathbf{R}' \in \mathscr{R}^n$ takes X' to the top from $\mathbf{R}$ if

$$xR_i'y \leftrightarrow xR_iy \quad \text{for all } x, y \in X'$$

but

$$xP_i'y \quad \text{for all } x \in X', y \in X \backslash X'.$$

Remarks:

- Given X' and $\mathbf{R}$, there are generally different $\mathbf{R}'$ that all take X' to the top from $\mathbf{R}$, because we are free to change preferences within $X \backslash X'$.
- Note that, if $\mathbf{R}'$ takes X' to the top from $\mathbf{R}$, then every $x \in X'$ maintains its position from $\mathbf{R}$ to $\mathbf{R}'$.

Arrow's Theorem for SCFs: Structure of Proof II

The proof has 4 steps. Given an SCF c that satisfies [$\bar{\text{U}}$], [$\bar{\text{M}}$] and [$\bar{\text{P}}$], we show:

1. If $\mathbf{R}'$ and $\mathbf{R}''$ both take $X' \neq \varnothing$ to the top from $\mathbf{R}$, then $c(\mathbf{R}') = c(\mathbf{R}'') \in X'$.

Using this insight:

2. We construct an SWF f from the SCF c.
   We do so by taking elements pairwise to the top and check which one wins.
3. The constructed SWF f violates [D].
4. The dictator for f is also a dictator for c, so c violates [$\bar{\text{D}}$].

Throughout, we will indicate in red where our presumptions enter.

Arrow's Theorem for SCFs: Proof I

Step 1

If $\mathbf{R}'$ and $\mathbf{R}''$ both take $X' \neq \varnothing$ to the top from $\mathbf{R}$, then $c(\mathbf{R}') = c(\mathbf{R}'') \in X'$.

Proof:

By [$\bar{\text{P}}$] we must have $c(\mathbf{R}') \in X'$, because every alternative from $X \backslash X'$ is strictly Pareto dominated by every alternative from X'.

Any $x \in X'$, so also $c(\mathbf{R}')$, maintains its position from $\mathbf{R}'$ to $\mathbf{R}''$.
Therefore [$\bar{\text{M}}$] implies $c(\mathbf{R}'') = c(\mathbf{R}')$. $\square$

Arrow's Theorem for SCFs: Proof II

Step 2

The following construction yields an SWF $f : \mathscr{R}^n \to \mathscr{R}$ from the SCF c:
For any $\mathbf{R} \in \mathscr{R}^n$ and $x, y \in X$, let $xf(\mathbf{R})y$ if and only if
(i) $x = y$, or
(ii) $x = c(\mathbf{R}')$ for some $\mathbf{R}'$ that takes $\{x, y\}$ to the top from $\mathbf{R}$, when $x \neq y$.

Proof:

By Step 1, it does not matter which $\mathbf{R}'$ is used for (ii), and there exists an admissible one by [$\bar{\text{U}}$], so the construction yields a unique result.

It remains to be shown that $f(\mathbf{R}) \in \mathscr{R}$ for all $\mathbf{R} \in \mathscr{R}^n$.
Fix some $\mathbf{R} \in \mathscr{R}^n$. Clearly, $f(\mathbf{R})$ is reflexive by (i) and complete by (ii).
By (ii) it is also antisymmetric.

Transitivity of $f(\mathbf{R})$ remains to be shown. Suppose $xf(\mathbf{R})y$ and $yf(\mathbf{R})z$ for three distinct alternatives $x, y, z \in X$. We will show that $xf(\mathbf{R})z$ must then also hold.

Arrow's Theorem for SCFs: Proof III

Let $\mathbf{R}'$ take $\{x, y, z\}$ to the top from $\mathbf{R}$, so that $c(\mathbf{R}') \in \{x, y, z\}$ by [$\bar{\text{P}}$].

Suppose $y = c(\mathbf{R}')$. Let $\mathbf{R}''$ take $\{x, y\}$ to the top from $\mathbf{R}'$. Then [$\bar{\text{M}}$] implies $y = c(\mathbf{R}'')$. But $\mathbf{R}''$ also takes $\{x, y\}$ to the top from $\mathbf{R}$, so $yf(\mathbf{R})x$ by definition of f. Since $f(\mathbf{R})$ is antisymmetric, this contradicts $xf(\mathbf{R})y$. Hence $y \neq c(\mathbf{R}')$.

Suppose $z = c(\mathbf{R}')$. Let $\mathbf{R}''$ take $\{y, z\}$ to the top from $\mathbf{R}'$. Then [$\bar{\text{M}}$] implies $z = c(\mathbf{R}'')$. But $\mathbf{R}''$ also takes $\{y, z\}$ to the top from $\mathbf{R}$, so $zf(\mathbf{R})y$ by definition of f. Since $f(\mathbf{R})$ is antisymmetric, this contradicts $yf(\mathbf{R})z$. Hence $z \neq c(\mathbf{R}')$.

We conclude that $x = c(\mathbf{R}')$. Let $\mathbf{R}''$ take $\{x, z\}$ to the top from $\mathbf{R}'$. Then [$\bar{\text{M}}$] implies $x = c(\mathbf{R}'')$. Since $\mathbf{R}''$ also takes $\{x, z\}$ to the top from $\mathbf{R}$, it follows that $xf(\mathbf{R})z$ by definition of f. This establishes transitivity. $\square$

Arrow's Theorem for SCFs: Proof IV

Step 3

The constructed SWF f violates [D].

Proof:

We will show that f satisfies the original Arrow axioms [U], [I] and [P].

[U] This has already been shown in Step 2.

[I] Suppose $\mathbf{R}$ and $\mathbf{R}'$ coincide with respect to x and y, for all $i \in N$. Let $\mathbf{R}''$ take $\{x, y\}$ to the top from $\mathbf{R}$. Then $\mathbf{R}''$ also takes $\{x, y\}$ to the top from $\mathbf{R}'$. By definition of f we then have $xf(\mathbf{R})y \leftrightarrow xf(\mathbf{R}')y$ and $yf(\mathbf{R})x \leftrightarrow yf(\mathbf{R}')x$.

[P] Let $\mathbf{R}$ be such that xP_iy for all $i \in N$, for some $x, y \in X$. Let $\mathbf{R}'$ take $\{x, y\}$ to the top from $\mathbf{R}$. Then $x = c(\mathbf{R}')$ by [$\bar{\text{P}}$], so $xf_P(\mathbf{R})y$ by construction of f.

It now follows from Arrow' impossibility theorem for SWFs that f violates [D], i.e. there exists $h \in N$ such that $xf_P(\mathbf{R})y$ whenever xP_hy. □

Arrow's Theorem for SCFs: Proof V

Step 4

The dictator for f is also a dictator for c.

Proof:

Consider any $\mathbf{R} \in \mathscr{R}^n$ and the winner $c(\mathbf{R})$.
For an arbitrary $y \in X$, $y \neq c(\mathbf{R})$, let $\mathbf{R}'$ take $\{c(\mathbf{R}), y\}$ to the top from $\mathbf{R}$.
Since $c(\mathbf{R})$ maintains position from $\mathbf{R}$ to $\mathbf{R}'$, we have $c(\mathbf{R}') = c(\mathbf{R})$ by [$\bar{\text{M}}$].
Hence $c(\mathbf{R}) f(\mathbf{R}) y$ by definition of f, for all $y \in X$.

We have just shown that $c(\mathbf{R})$ is a greatest element of X according to $f(\mathbf{R})$.
Hence it is a greatest element according to the preference R_h of the dictator h of f. Therefore h is also a dictator for c, so that c violates [$\bar{\text{D}}$]. $\square$

This completes the proof of Arrow's impossibility theorem for SCFs.

Possibility for SDFs

This part relies on [GA].

A Solution!?

We have now established impossibility results for both SWFs and SCFs.

Maybe we are satisfied with a social binary relation $d(\mathbf{R}) \in \mathscr{C}$, which generates a non-empty set of social winners from every non-empty and finite subset $S \subseteq X$.

In other words, can we get a possibility result for social decision functions (SDFs)?

Reminder:
Let $\mathscr{A} \subseteq \mathscr{R}^n$ be be a set of admissible preference profiles. A social decision function is a mapping $d : \mathscr{A} \to \mathscr{C}$, which assigns to every admissible profile of preferences $\mathbf{R} \in \mathscr{A}$ a reflexive, complete and acyclical binary relation $d(\mathbf{R}) \in \mathscr{C}$.

Even though we consider a different type of rule, Arrow's axioms for SWFs can be applied without modification, since $d(\mathbf{R})$ is still a binary relation.

Arrow's Axioms for SDFs

Definition

[U] $\mathscr{A} = \mathscr{R}^n$

[I] For any pair of alternatives $x, y \in X$, if two profiles $\mathbf{R}, \mathbf{R}' \in \mathscr{A}$ satisfy

$$xR_iy \;\leftrightarrow\; xR_i'y \text{ and } yR_ix \;\leftrightarrow\; yR_i'x \text{ for all } i \in N,$$

then

$$xd(\mathbf{R})y \;\leftrightarrow\; xd(\mathbf{R}')y \;\text{ and }\; yd(\mathbf{R})x \;\leftrightarrow\; yd(\mathbf{R}')x$$

must be true.

[P] For any pair of alternatives $x, y \in X$, if a profile $\mathbf{R} \in \mathscr{A}$ satisfies

$$xP_iy \text{ for all } i \in N,$$

then $xd_P(\mathbf{R})y$ must be true.

[D] There is no individual $i \in N$ such that, for all $x, y \in X$ and $\mathbf{R} \in \mathscr{A}$, xP_iy implies $xd_P(\mathbf{R})y$.

Sen's Possibility Result

Theorem

There exists an SDF that satisfies [U], [I], [P] and [D] for any m.

We prove the theorem by giving an example of an SDF that satisfies the axioms. In fact, consider the Pareto extension rule (which is even quasi-transitive):

For all $x, y \in X$, we define

$$xd^{PE}(\mathbf{R})y \quad \text{except if} \quad yR_i x \; \forall i \in N \text{ and } yP_j x \text{ for at least one } j \in N.$$

Proof:

[U]: We have shown earlier that PE is an SDF on the universal domain $\mathscr{A} = \mathscr{R}^n$.

[I]: Immediate, since PE is a pairwise comparison method.

[P]: If $xP_i y$ for all $i \in N$, then clearly $xd^{PE}(\mathbf{R})y$ but $\neg yd^{PE}(\mathbf{R})x$, so $xd_P^{PE}(\mathbf{R})y$.

[D]: Immediate, a single pair i, j with $xP_i y$ and $yP_j x$ suffices for $xd_I^{PE}(\mathbf{R})y$. □

Discussion

Note that a possibility result for SDFs implies a corresponding result for SCCs.

Is PE a reasonable solution to the problem of social choice?

- PE does not solve conflicts of interest. Whenever two alternatives x and y cannot be Pareto compared, it "solves" the problem by $xd_I^{PE}(\mathbf{R})y$.
- The induced socially optimal set $G(S, d^{PE}(\mathbf{R}))$ will generally be very large. It contains all elements that are Pareto efficient within S:
  - If $x \in S$ is not Pareto dominated by any $y \in S$, then $xd^{PE}(\mathbf{R})y$ for all $y \in S$, so $x \in G(S, d^{PE}(\mathbf{R}))$.
  - If $x \in S$ is Pareto dominated by some $y \in S$, then $\neg xd^{PE}(\mathbf{R})y$, so $x \notin G(S, d^{PE}(\mathbf{R}))$.
- While no individual is a dictator in the sense of [D], every individual is a weak dictator: for every $i \in N$ and all $x, y \in X$, xP_iy implies $xd^{PE}(\mathbf{R})y$. Every individual has veto power even if all others strictly agree.
- Unfortunately, there is essentially no other SDF that satisfies all the axioms.

  Unfortunately, we get an impossibility result if we exclude weak dictators.

Possibility with 2 Alternatives

This part relies on [GA] and [MC].

Binary Choice

Arrow's impossibility result for SWFs applies when $m \geq 3$.

Many real-world problems concern social decisions among only two alternatives:

- F-35 yes or no?
- Make a job offer to an applicant or not?
- Tyson Fury or Dillian Whyte?
- ...

Can we get a possibility result for SWFs when $m = 2$?

Note that the difference between SWFs and SCFs becomes minor when $m = 2$. We do not consider SCFs separately.

Some Methods

Let us check the following methods (recall Exercise 3.1):

- Borda count. It satisfies all Arrow axioms except [I].
  But [I] is satisfied by any method when $m = 2$, because we have $\mathbf{R} = \mathbf{R}'$ when every voter's preference between x and y does not change.
  Hence BC satisfies all Arrow axioms when $m = 2$!
- Copeland method. It satisfies all Arrow axioms except [I].
  By the same argument, CO also satisfies all axioms when $m = 2$!
- Pairwise majority voting. It satisfies all Arrow axioms except [U].
  PM might generate a non-transitive outcome. However, non-transitivity is impossible with $m = 2$, so [U] is satisfied.
  Hence PM also satisfies all Arrow axioms when $m = 2$!
- Plurality voting. It satisfies all Arrow axioms except [I] and [P].
  [I] is not a problem when $m = 2$. Furthermore, if xP_iy for all $i \in N$, then clearly $xf_P^{PV}(\mathbf{R})y$ when $m = 2$, so [P] is also not a problem.
  PV also satisfies all Arrow axioms when $m = 2$!

Majority Voting

Clearly, all these methods are the same and coincide with majority voting (MV) when there are only two alternatives (recall Exercise 2.2).

Let us also check the (different) Pareto efficiency method (recall Exercise 3.1):

- It satisfies all Arrow axioms except [I] and [P].

  [I] is not a problem when $m = 2$. Furthermore, [P] is not a problem since $X^I(\mathbf{R})$ can contain at most one alternative.

  Hence EM also satisfies all Arrow axioms when $m = 2$!

We will now investigate MV more closely. We will try to find a characterization of MV. We will look for a list of properties that are equivalent to MV. Not only does MV satisfy these properties, but MV is the only method that satisfies them.

Formal Framework

- Let $X = \{x, y\}$, so $m = 2$.
- There are now exactly three possible preferences on X:
  - $R^0 = \{(x,x),(y,y),(x,y),(y,x)\}$, shortcut notation xIy.
  - $R^{+1} = \{(x,x),(y,y),(x,y)\}$, shortcut notation xPy.
  - $R^{-1} = \{(x,x),(y,y),(y,x)\}$, shortcut notation yPx.

  Hence we have $\mathscr{R} = \{R^{-1}, R^0, R^{+1}\}$.
- A profile $\mathbf{R} = (R_1, \ldots, R_n) \in \mathscr{R}^n$ can therefore also be written as $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \{-1, 0, +1\}^n$, where

$$\alpha_i = \begin{cases} +1 & & xP_iy \\ 0 & \text{represents preference} & xI_iy & \text{of voter } i. \\ -1 & & yP_ix \end{cases}$$

- Let $\mathscr{A} \subseteq \{-1, 0, +1\}^n$ be a set of admissible preference profiles. An SWF is a mapping $f : \mathscr{A} \to \{-1, 0, +1\}$, which assigns to every admissible profile of preferences $\boldsymbol{\alpha} \in \mathscr{A}$ a social preference $f(\boldsymbol{\alpha}) \in \{-1, 0, +1\}$.

Examples

- Majority voting: $f^{MV}(\boldsymbol{\alpha}) = \text{sign} \sum_{i=1}^{n} \alpha_i$, where

$$\text{sign } z = \begin{cases} +1 & \text{if} \quad z > 0, \\ 0 & \text{if} \quad z = 0, \\ -1 & \text{if} \quad z < 0. \end{cases}$$

- Generalized class: $f(\boldsymbol{\alpha}) = \text{sign} \sum_{i=1}^{n} \beta_i \alpha_i$ for given weights $\beta_1, \ldots, \beta_n \geq 0$.
  - Majority voting: $\beta_1 = \ldots = \beta_n > 0$.
  - Dictatorship of h: $\beta_h > 0$ and $\beta_i = 0$ for all $i \neq h$.
  - At a shareholders' meeting, β_i could correspond to the number of shares that voter i holds in the company.

- Pareto efficiency method:

$$f^{EM}(\boldsymbol{\alpha}) = \begin{cases} +1 & \text{if } \alpha_i \geq 0 \ \forall i \in N \text{ and } \alpha_j > 0 \text{ for at least one } j \in N, \\ -1 & \text{if } \alpha_i \leq 0 \ \forall i \in N \text{ and } \alpha_j < 0 \text{ for at least one } j \in N, \\ 0 & \text{otherwise.} \end{cases}$$

May's Axioms, Intuitively

The following properties will characterize majority voting:

[U] Universality, as before.

[N] Neutrality:

The method should treat both alternatives equally, there should be no bias.

[PR] Positive Responsiveness:

The method should be (minimally) sensitive to preference changes. If x is socially at least as good as y, and if individual preferences change further in favor of x, then x should become socially strictly preferred over y.

As we will see, this can be thought of as a strengthening of [P].

[A] Anonymity (also called symmetry):

The method should treat all voters equally, their names should not matter.

As we will see, this can be thought of as a strengthening of [D].

May's Axioms, Formally

Definition

[U] $\mathscr{A} = \{-1, 0, +1\}^n$

[N] For any $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathscr{A}$ it must hold that

$$f(\alpha_1, \ldots, \alpha_n) = -f(-\alpha_1, \ldots, -\alpha_n),$$

provided that $-\boldsymbol{\alpha} = (-\alpha_1, \ldots, -\alpha_n) \in \mathscr{A}$.

[PR] Suppose $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathscr{A}$ satisfy $\alpha'_i \geq \alpha_i$ for all $i \in N$ and $\alpha'_j > \alpha_j$ for at least one $j \in N$. Then $f(\boldsymbol{\alpha}) \geq 0$ implies $f(\boldsymbol{\alpha}') = +1$.

[A] For all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathscr{A}$, if $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ are permutations of each other, then $f(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}')$ must be true.

Note that $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ are permutations of each other if there exists a bijective function $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $\alpha_i = \alpha'_{\pi(i)}$ for all $i \in N$.

Example: $\boldsymbol{\alpha} = (+1, -1, 0, 0)$ and $\boldsymbol{\alpha}' = (0, -1, 0, +1)$, based on $\pi(1) = 4$, $\pi(2) = 2$, $\pi(3) = 1$ and $\pi(4) = 3$.

May's Axioms, Remarks I

- [PR] is defined only in one direction, i.e. for the case that x gains support. However, if combined with [N], the analogous property is ensured also for y.
- [PR] can be seen as a strengthening of [P].

  With two alternatives, [P] reduces to requiring $f(+1, \dots, +1) = +1$ and $f(-1, \dots, -1) = -1$, whenever these preference profiles are admissible.

  We show on the following slide that [PR] implies [P], whenever [U] and [N] are also satisfied. Hence [PR] alone is not stronger than [P].
- [A] can be seen as a strengthening of [D].

  With two alternatives, [D] reduces to the requirement that there is no $i \in N$ such that, for all $\boldsymbol{\alpha} \in \mathscr{A}$, $\alpha_i = +1$ implies $f(\boldsymbol{\alpha}) = +1$ and $\alpha_i = -1$ implies $f(\boldsymbol{\alpha}) = -1$.

  We show on the next slide that [A] implies [D], whenever [U] is also satisfied. Hence [A] alone is not stronger than [D].

May's Axioms, Remarks II

Proposition

Assume $m = 2$. If f satisfies [U], [N] and [PR], then it also satisfies [P].

Proof:

Let f satisfy [U], [N] and [PR]. By [U] and [N], $f(+1, \ldots, +1) = +1$ if and only if $f(-1, \ldots, -1) = -1$, so we only need to examine one of these cases.

Assume to the contrary that $f(+1, \ldots, +1) \leq 0$.

If there exists $\boldsymbol{\alpha} \neq (+1, \ldots, +1)$ for which $f(\boldsymbol{\alpha}) \geq 0$, then $f(+1, \ldots, +1) = +1$ would have to hold by [PR], a contradiction. Hence $f(\boldsymbol{\alpha}) = -1$ must hold for all $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$, $\boldsymbol{\alpha} \neq (+1, \ldots, +1)$. In particular, $f(-1, \ldots, -1) = -1$. But this again implies $f(+1, \ldots, +1) = +1$ by [N], a contradiction. $\square$

Remark 1: To see why we require [U], consider $\mathscr{A} = \{(+1, \ldots, +1)\}$ and $f(+1, \ldots, +1) = -1$, which violates [P] but satisfies [PR].

Remark 2: To see why we require [N], consider f with $f(\boldsymbol{\alpha}) = -1$ for all $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$, which violates [P] but satisfies [PR].

May's Axioms, Remarks III

Proposition

Assume $m = 2$. If f satisfies [U] and [A], then it also satisfies [D].

Proof:

Let f satisfy [U] and [A]. Assume to the contrary that h is a dictator.

Consider any $\boldsymbol{\alpha} \in \mathscr{A}$ where $\alpha_h = +1$ and $\alpha_j = -1$ for some $j \neq h$, which exists by [U]. Then $f(\boldsymbol{\alpha}) = +1$ must hold.

Let $\boldsymbol{\alpha}'$ be obtained from $\boldsymbol{\alpha}$ by permuting the preferences of h and j, i.e. $\alpha'_h = -1$ and $\alpha'_j = +1$, and leaving all other preferences unchanged. Then $f(\boldsymbol{\alpha}') = +1$ must hold by [A], which contradicts that h is a dictator. $\square$

Remark 1: To see why we require [U], consider $\mathscr{A} = \{(+1, \ldots, +1)\}$ and $f(+1, \ldots, +1) = +1$, which satisfies [A]. However, every voter is a dictator on this (extremely) restricted domain.

May's Theorem

Theorem

Assume $m = 2$. An SWF f satisfies [U], [N], [PR] and [A] if and only if it is MV.

- When we apply MV, we know that the axioms are satisfied.
- When we want the axioms to be satisfied, we must use MV.
  Whenever we apply a different method, then at least one axiom is violated.

May's Theorem: Proof I

Proof:

"$\Leftarrow$" We have to show that $f^{MV}(\boldsymbol{\alpha}) = \text{sign} \sum_{i=1}^{n} \alpha_i$ satisfies the four axioms.

[U] Clearly $f^{MV}(\boldsymbol{\alpha}) \in \{-1, 0, +1\}$ for all $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$.

[N] We have $-f^{MV}(-\boldsymbol{\alpha}) = -\text{sign} \sum_{i=1}^{n} -\alpha_i = \text{sign} \sum_{i=1}^{n} \alpha_i = f^{MV}(\boldsymbol{\alpha})$.

[PR] If $\boldsymbol{\alpha}, \boldsymbol{\alpha}'$ are as required for [PR] to apply, then $\sum_{i=1}^{n} \alpha'_i > \sum_{i=1}^{n} \alpha_i$. Hence $\sum_{i=1}^{n} \alpha'_i > 0$ whenever $\sum_{i=1}^{n} \alpha_i \geq 0$, so $f^{MV}(\boldsymbol{\alpha}') = +1$ whenever $f^{MV}(\boldsymbol{\alpha}) \geq 0$.

[A] Fix any bijective function $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$. Then clearly $\sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \alpha_{\pi(i)}$. Hence $f^{MV}(\boldsymbol{\alpha}) = f^{MV}(\boldsymbol{\alpha}')$ whenever $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ are permutations of each other. $\diamond$

"$\Rightarrow$" Let f be an SWF that satisfies [U], [N], [PR] and [A].

For any $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$, define

$$n^+(\boldsymbol{\alpha}) = |\{i \in N | \alpha_i = +1\}| \text{ and } n^-(\boldsymbol{\alpha}) = |\{i \in N | \alpha_i = -1\}|.$$

By [A] only the *number* of voters with different preferences matters, so we can write $f(\boldsymbol{\alpha}) = g(n^+(\boldsymbol{\alpha}), n^-(\boldsymbol{\alpha}))$ for some function g.

May's Theorem: Proof II

- Case 1: Consider any $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$ where $n^+(\boldsymbol{\alpha}) = n^-(\boldsymbol{\alpha})$.

  Then $n^-(-\boldsymbol{\alpha}) = n^+(\boldsymbol{\alpha}) = n^-(\boldsymbol{\alpha}) = n^+(-\boldsymbol{\alpha})$. But then $f(\boldsymbol{\alpha}) = g(n^+(\boldsymbol{\alpha}), n^-(\boldsymbol{\alpha})) = g(n^+(-\boldsymbol{\alpha}), n^-(-\boldsymbol{\alpha})) = f(-\boldsymbol{\alpha}) = -f(\boldsymbol{\alpha})$, the last step by [N].

  Since $f(\boldsymbol{\alpha}) = -f(\boldsymbol{\alpha})$ can only be true when $f(\boldsymbol{\alpha}) = 0$, we have shown that $f(\boldsymbol{\alpha}) = 0$ must hold for all $\boldsymbol{\alpha}$ with $n^+(\boldsymbol{\alpha}) = n^-(\boldsymbol{\alpha})$.

May's Theorem: Proof III

- Case 2: Consider any $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$ where $n^+(\boldsymbol{\alpha}) > n^-(\boldsymbol{\alpha})$.

  Let $H = n^+(\boldsymbol{\alpha})$ and $J = n^-(\boldsymbol{\alpha})$ and assume, without loss of generality due to [A], that $\alpha_i = +1$ for all $i \leq H$ (and hence $\alpha_i \leq 0$ for all $i > H$).

  Consider the profile $\boldsymbol{\alpha}'$ where

$$\begin{array}{rll} \alpha'_i = +1 & \text{for all} & i \leq J, \\ \alpha'_i = 0 & \text{for all} & J < i \leq H, \\ \alpha'_i = \alpha_i & \text{for all} & i > H, \end{array}$$

  so that $n^+(\boldsymbol{\alpha}') = J = n^-(\boldsymbol{\alpha}')$ and hence $f(\boldsymbol{\alpha}') = 0$ by case 1.
  Since x is favored by $\boldsymbol{\alpha}$ compared to $\boldsymbol{\alpha}'$ in the sense of [PR],
  we must have $f(\boldsymbol{\alpha}) = +1$.

  Hence we have shown that $f(\boldsymbol{\alpha}) = +1$ for all $\boldsymbol{\alpha}$ with $n^+(\boldsymbol{\alpha}) > n^-(\boldsymbol{\alpha})$.

May's Theorem: Proof IV

- Case 3: Consider any $\boldsymbol{\alpha} \in \{-1, 0, +1\}^n$ where $n^+(\boldsymbol{\alpha}) < n^-(\boldsymbol{\alpha})$.

  Then $n^+(-\boldsymbol{\alpha}) > n^-(-\boldsymbol{\alpha})$, so $f(-\boldsymbol{\alpha}) = +1$ by case 2, and therefore $f(\boldsymbol{\alpha}) = -1$ by [N].

  Hence we have shown that $f(\boldsymbol{\alpha}) = -1$ for all $\boldsymbol{\alpha}$ with $n^+(\boldsymbol{\alpha}) < n^-(\boldsymbol{\alpha})$.

In summary, we must have
$f(\boldsymbol{\alpha}) = \text{sign}\,[n^+(\boldsymbol{\alpha}) - n^-(\boldsymbol{\alpha})] = \text{sign} \sum_{i=1}^n \alpha_i = f^{MV}(\boldsymbol{\alpha})$. $\diamond\ \Box$

Possibility with Restricted Domains

This part relies on [GA] and [MC].

Weakening Universality

If we want an SWF for $m \geq 3$ alternatives, we need to relax at least one of the axioms [U], [I], [P] or [D]. If we are willing to relax [I], then both Borda count and the Copeland method are a solution (recall Exercise 3.1).

For SCFs, we need to relax at least one of the axioms [$\bar{\text{U}}$], [$\bar{\text{M}}$], [$\bar{\text{P}}$] or [$\bar{\text{D}}$]. If we are willing to relax [$\bar{\text{M}}$], instant-runoff voting is a solution (recall Exercise 3.2).

We now investigate whether relaxing [U] or [$\bar{\text{U}}$] can also help.

Pairwise Majority Voting

Consider pairwise majority voting (PM). We can ask the following questions:

- Since PM satisfies axioms [I], [P] and [D] (recall Exercise 3.1), can we find a reasonable restricted domain $\mathscr{A} \subset \mathscr{R}^n$ so that PM becomes an SWF $f^{PM} : \mathscr{A} \to \mathscr{R}$ (recall Exercise 2.1)?
- We can also define an SCF $c^{PM} : \mathscr{A} \to X$ based on PM, by selecting as a winner the alternative which PM ranks above all other alternatives, i.e. the Condorcet winner. Can we find a domain $\mathscr{A} \subset \mathscr{R}^n$ so that a Condorcet winner always exist? The axioms [$\bar{\text{M}}$], [$\bar{\text{P}}$] and [$\bar{\text{D}}$] are then clearly satisfied.

Preferences Regularities

In many applications, individual preferences will show certain regularities:

- If $X = \{\text{Left}, \text{Center}, \text{Right}\}$ is a set of political parties, then preference

$$\text{Right } P_i \text{ Left } P_i \text{ Center}$$

  would be rather unusual.

- If $X = \{7\%, 8\%, 9\%, 10\%\}$ is a set of value-added tax rates, then preference

$$7\% \; P_i \; 10\% \; P_i \; 8\% \; P_i \; 9\%$$

  would be rather unusual.

- If $X = \{\text{Bern}, \text{Winterthur}, \text{Zurich}\}$ are possible workplaces for someone living in Zurich, then preference

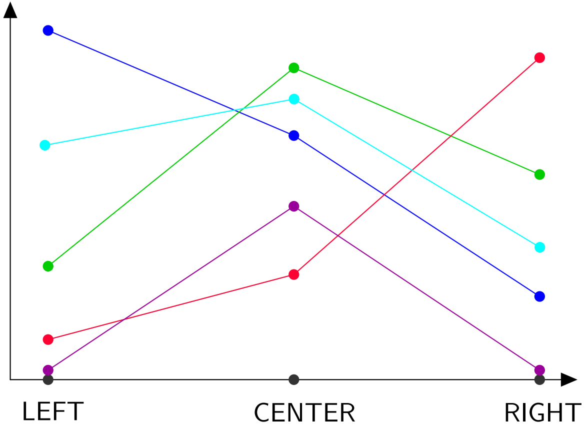
$$\text{Bern } P_i \text{ Winterthur } P_i \text{ Zurich}$$

  would be rather unusual.

The examples have in common that the alternatives can be ordered (political spectrum, size, location) and preferences often somehow adhere to this order.

Single-Peaked Preferences, Illustration I

Suppose three political parties can be ordered from left to right. The following graph depicts all five different single-peaked preferences:



Single-Peaked Preferences, Definition I

We first have to define how alternatives are ordered. Let $\geq$ by a binary relation on X that is reflexive, complete, transitive and antisymmetric, hence a linear order. The induced strict part of $\geq$ is denoted by $>$.

We do not interpret $\geq$ as a preference here, but as some objective reference order:

- Ordering of parties in the political spectrum:

$$\text{Right} \;>\; \text{Center} \;>\; \text{Left}.$$

- Ordering of tax rates according to size:

$$10\% \;>\; 9\% \;>\; 8\% \;>\; 7\%.$$

- Ordering of workplaces according to location along railway:

$$\text{Bern} \;>\; \text{Zurich} \;>\; \text{Winterthur}.$$

Single-Peaked Preferences, Definition II

Definition

A preference $R \in \mathscr{R}$ is single-peaked with respect to the linear order $\geq$ if there exists an alternative $x \in X$ such that

$$x \geq z > y \quad \text{implies} \quad z \; P \; y$$

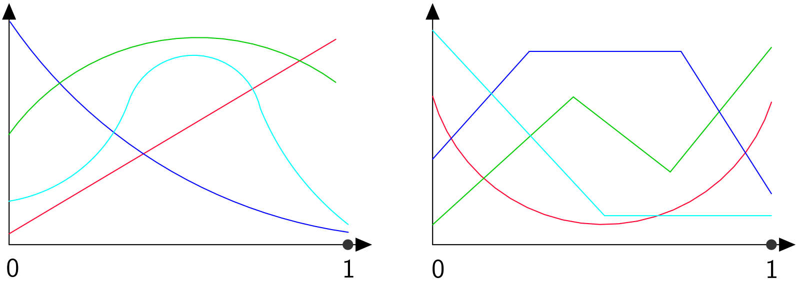
and

$$y > z \geq x \quad \text{implies} \quad z \; P \; y.$$

- Alternative x is the peak of R, i.e. the most preferred alternative.
- As we move away from x in either direction according to the linear order $\geq$, the alternatives become strictly worse according to the preference R.
- Note that we cannot make general statements about the preference between two alternatives on different sides of the peak.

Single-Peaked Preferences, Illustration II

Suppose $X = [0, 1]$ are tax rates, ordered according to size. The following graph depicts single-peaked (left) and non-single-peaked (right) preferences:



Single-Peaked Preferences, Discussion

- A preference $R \in \mathscr{R}$ can be single-peaked with respect to some order $\geq$ but not with respect to a different order $\geq'$.

  Example: $xPyPz$ is single-peaked w.r.t. $y > x > z$ but not $x >' z >' y$.

- A preference $R \in \mathscr{R}$ can be single-peaked with respect to different orders.

  Example: $xPyPz$ is single-peaked w.r.t. both $y > x > z$ and $x >' y >' z$.

- Some preferences $R \in \mathscr{R}$ are not single-peaked with respect to any order.

  Example: $wPxIyIz$.

- Any strict preference $R \in \mathscr{P}$ is single-peaked for some order, e.g. $\geq = R$.

  Example: $xPyPz$ and $x > y > z$ as above.

- Preferences $R \in \mathscr{R} \backslash \mathscr{P}$ can be single-peaked with respect to some order $\geq$.

  Example: $xPyIz$ is single-peaked w.r.t. $y > x > z$.

Single-Peaked Preference Profiles

For a given linear order $\geq$, let $\mathscr{R}_{\geq} \subset \mathscr{R}$ denote the set of all preferences that are single-peaked with respect to $\geq$.

Example: For $X = \{\text{Left}, \text{Center}, \text{Right}\}$ and Right $>$ Center $>$ Left, $\mathscr{R}_{\geq}$ contains:
Right P Center P Left, Left P Center P Right, Center P Right P Left,
Center P Left P Right, Center P Left I Right.

Then $\mathscr{R}^n_{\geq}$ is the set of all preference *profiles* $\mathbf{R} = (R_1, \ldots, R_n)$ in which every voter's preference R_i is single-peaked with respect to the *same* linear order $\geq$.

In the example, every voter's preference must be one of the five possibilities.

Median Voter

Consider a profile $\mathbf{R} \in \mathscr{R}_{\geq}^{n}$ of preferences that are single-peaked with respect to $\geq$.

For every $i \in N$, let $x_i \in X$ be the peak (the most preferred alternative) of voter i.

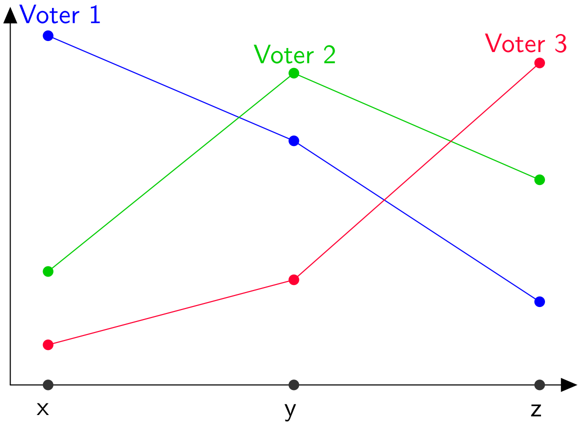
Definition

Given profile $\mathbf{R} \in \mathscr{R}_{\geq}^{n}$, voter $h \in N$ is a median voter if

$$|\{i \in N | x_i \geq x_h\}| \geq n/2 \quad \text{and} \quad |\{i \in N | x_h \geq x_i\}| \geq n/2.$$

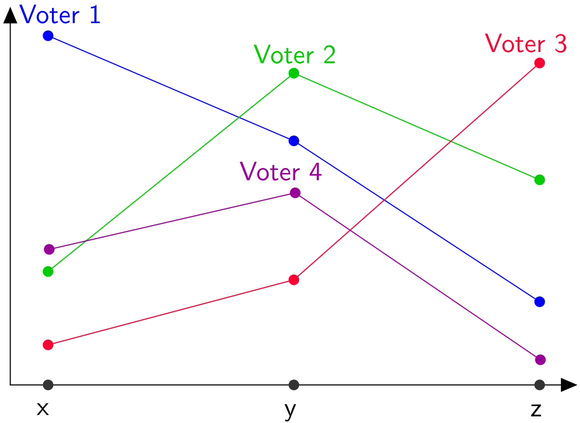
The peak of the median voter is a median in the sense that (at least) half of the population has a (weakly) higher and (at least) half has a (weakly) lower peak.

Median Voter, Illustration I



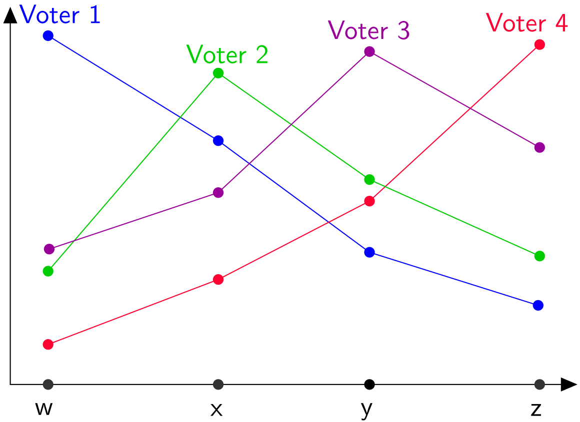
We have $X = \{x, y, z\}$, $z > y > x$, $n = 3$.
Peaks are $x_1 = x$, $x_2 = y$, $x_3 = z$.
Voter 2 with peak y is the median voter.

Median Voter, Illustration II



We have $X = \{x, y, z\}$, $z > y > x$, $n = 4$.
Peaks are $x_1 = x$, $x_2 = y$, $x_3 = z$, $x_4 = y$.
Voters 2 and 4 are both median voters, with identical peak y.

Median Voter, Illustration III



We have $X = \{w, x, y, z\}$, $z > y > x > w$, $n = 4$.
Peaks are $x_1 = w$, $x_2 = x$, $x_3 = y$, $x_4 = z$.
Voters 2 and 3 are both median voters, with different peaks x and y.

Median Voter Theorem for c^{PM}

Lemma

Suppose $\mathbf{R} \in \mathscr{R}^n_{\geq}$ for some linear order $\geq$. Let $h \in N$ be a median voter. Then $x_h \, f^{PM}(\mathbf{R}) \, y$ for all $y \in X$.

- The lemma follows immediately from the observation that, in any pairwise vote, at least half of the population votes for a median voter's peak.
- Hence a median voter's peak cannot be defeated in any pairwise vote. In this sense, it is a weak Condorcet winner.

A true Condorcet winner must actually win every pairwise vote. It can be shown that there is always a unique median peak if n is odd, which is then a Condorcet winner in this sense. Hence we obtain:

Theorem

If n is odd, c^{PM} is an SCF on the domain $\mathscr{A} = \mathscr{R}^n_{\geq}$, for any linear order $\geq$.

Median Voter Theorem for f^{PM}

If n is odd and all preferences are single-peaked (with respect to the same linear order), then a Condorcet winner always exists and the SCF c^{PM} is well-defined.

It is still possible that pairwise majority voting f^{PM} is not transitive. However, the additional assumption of strictness of all individual preferences ensures that problems with transitivity cannot occur:

Theorem

If n is odd, f^{PM} is an SWF on the domain $\mathscr{A} = \mathscr{P}^n_{\geq}$, for any linear order $\geq$.

Remark: $\mathscr{A} = \mathscr{P}^n_{\geq}$ can be further weakened to "value-restricted" preferences.

Application: Basic Income I

Consider a (too) simple model of voting on a non-means-tested basic income:

- $N = \{1, 2, \ldots, n\}$ is the set of voters, where n is odd.
  Voter i has an exogenous income y_i, where $0 \leq y_1 < y_2 < \ldots < y_n$.
  $y = \sum_{i=1}^{n} y_i$ is total income and $\bar{y} = y/n$ is average income.
- $X = [0, \bar{y}]$ is the set of possible basic income levels b.
  Basic income is financed by a constant marginal tax rate t.
  Budget-balance $nb = \sum_{i=1}^{n} ty_i$ can be rewritten as $t = b/\bar{y}$.
- Ex-post income $\hat{y}_i$ of i as a function of b is

$$\hat{y}_i(b) = b + (1 - t)y_i = y_i + \left(1 - \frac{y_i}{\bar{y}}\right) b.$$

  Preference R_i is given by $\hat{y}_i(b)$: $b' R_i b''$ if and only if $\hat{y}_i(b') \geq \hat{y}_i(b'')$.

The model is equivalent to voting on redistributive taxation (e.g. [MC], p. 814).

Application: Basic Income II

Suppose $y_i \neq \bar{y}$, $\forall i \in N$, so each $\hat{y}_i(b)$ is strictly increasing or strictly decreasing in b. Hence preferences are strict and single-peaked (w.r.t. the usual order):

- If $y_i < \bar{y}$, then $\hat{y}_i(b)$ is strictly increasing in b, with peak $b_i = \bar{y}$.
- If $y_i > \bar{y}$, then $\hat{y}_i(b)$ is strictly decreasing in b, with peak $b_i = 0$.

Since the peaks b_i are (weakly) decreasing in i, voter $m = (n+1)/2$ with the median income y_m is also a median voter (in fact, every voter with the same peak as m is also a median voter):

- If $y_m < \bar{y}$, then $b_m = \bar{y}$ is the Condorcet winner.
- If $y_m > \bar{y}$, then $b_m = 0$ is the Condorcet winner.

SCF c^{PM} selects the Condorcet winner b_m, SWF f^{PM} selects preference R_m.

Application: Basic Income III

The outcome of pairwise voting depends on the relation between average and median income, and hence on the shape of the income distribution:

	y_1	y_2	y_3	y_m	$\bar{y}$	b_m
Example 1	60	90	150	90	100	100
Example 2	60	110	130	110	100	0

A more realistic model might have to include:

- endogenous income, incentive effects of basic income and taxation
- unemployment, income risk
- non-linear taxation, VAT
- ...

Summary of Part 3

Main insights:

- The details of the decision process matter.
- Arrow's impossibility results for SWFs and SCFs show that preference aggregation faces fundamental logical problems.
- Group decisions must be "less rational" than individual decisions.

Solutions to the dilemma:

- Consider SWFs for $m \geq 3$ but
  - give up [I]: Borda count, Copeland method
  - give up [U]: f^{PM} for single-peaked preferences
- Consider SCFs for $m \geq 3$ but
  - give up [$\bar{\text{M}}$]: instant-runoff voting
  - give up [$\bar{\text{U}}$]: c^{PM} for single-peaked preferences
- Consider SDFs: Pareto extension rule (clearly unsatisfactory)
- $m = 2$: majority voting

Part 4: Individual Rights

The Liberal Paradox

This part relies on [GA] and [AS].

Collective Decisions and Individual Rights

We have now investigated in great detail the possibility or impossibility to resolve conflicts of interest by using different voting methods.

One can, however, also adopt the position that not all decisions should be done collectively by voting:

- Should we vote on a religion that all citizens have to belong to?
- [AS] used the example of whether or not people should be allowed to read the book "Lady Chatterley's Lover".
- Similar issues might arise in the context of drugs, LGBTQ+ rights,...

It is a central idea of liberalism that everybody should have a private sphere of decisions that nobody else is allowed to control (e.g. Mill, von Hayek). In this part, we will try to incorporate such concerns into our social choice framework.

Example I

Consider the following example of occupational choice:

- $N = \{1, 2\}$, person 1 can work as either a butcher or a baker, person 2 can work as either a banker or a professor.
- Alternatives $X = \{w, x, y, z\}$ are the possible job allocations:

	Banker	Professor
Butcher	w	x
Baker	y	z

- Both individuals have a preference $R_i \in \mathscr{R}$ over these alternatives. For instance, person 1 might have one of the following preferences:
  - $w\, I_1\, x\, P_1\, y\, I_1\, z$ (prefers being a butcher, other's job irrelevant)
  - $w\, P_1\, x\, P_1\, y\, P_1\, z$ (prefers being a butcher, but having a bank is useful)
  - $x\, P_1\, z\, P_1\, w\, P_1\, y$ (strongly dislikes bankers, but prefers to be a butcher)
  - ...
- Given individual preferences, we can now apply some voting method.

Example II

- Assume, for instance, that preferences are given by

$$w \; P_1 \; y \; P_1 \; x \; P_1 \; z \quad \text{and} \quad y \; P_2 \; z \; P_2 \; w \; P_2 \; x.$$

  Person 1 prefers being a butcher, but having a bank is very important. Person 2 prefers being a banker, but strongly cares for animal rights.

  If we apply Borda count, we obtain the social ranking

$$y \; P \; w \; P \; z \; P \; x,$$

  and we would implement job allocation y.

- Person 1 is forced to become a baker, even though he would like to switch to being a butcher, holding person 2's job fixed.

  From a liberal's perspective, occupational choice might belong to the protected private sphere. Here, voting does not respect this sphere.

Example III

- To formalize private sphere, we could require that the social preference...

  ...between w and y is determined by person 1 only,
  ...between x and z is determined by person 1 only,
  ...between w and x is determined by person 2 only,
  ...between y and z is determined by person 2 only.

  Together with the observation that y strictly Pareto dominates x, this leads us to the social preference $w \, P \, y \, P \, x \, P \, z$.

  If we select w, then nobody must be forced not to switch jobs.

Rights Systems I

Let $N = \{1, \ldots, n\}$ be the set of voters and X be an arbitrary set of alternatives.

Definition

A rights system is a profile $\mathbf{D} = (D_1, \ldots, D_n)$ where, for each $i \in N$, D_i is a binary relation on X which is irreflexive and symmetric.

- D_i represents the private sphere of voter i.
- If we want to respect $\mathbf{D}$, then the social ranking of x and y should not contradict i's preferences when $(x, y) \in D_i$ (and $(y, x) \in D_i$ by symmetry).
- In the previous example, we have $\mathbf{D} = (D_1, D_2)$ given by $D_1 = \{(w, y), (y, w), (x, z), (z, x)\}$, $D_2 = \{(w, x), (x, w), (y, z), (z, y)\}$.

Rights Systems II

We consider social decision functions in the following.

Definition

An SDF $d : \mathscr{A} \to \mathscr{C}$ respects the rights systems $\mathbf{D}$ if, $\forall\, \mathbf{R} \in \mathscr{A}$, $i \in N$, $x, y \in X$, $(x, y) \in D_i$ and xP_iy implies $xd_P(\mathbf{R})y$.

- If $(x, y) \in D_i$, and hence $(y, x) \in D_i$ by symmetry, then a strict preference of voter i between x and y (in either direction) must carry over to society.
- Phrased differently, if $(x, y) \in D_i$ then $\{i\}$ is completely decisive for x over y and for y over x.
- In our example, Borda count did not respect the rights system $\mathbf{D} = (D_1, D_2)$.

Sen's Axioms, Intuitively

We will again ask if there are SDFs that satisfy some plausible and weak axioms:

[U*] Universality:

The SDF should work for all conceivable preference profiles.

[P*] Weak Pareto Principle:

If all individuals strictly prefer x over y, this should also hold for society.

[L*] Minimal Liberalism:

At least two individuals should each have at least two alternatives over which they are completely decisive in the above sense. Hence there should be some (minimal) rights system that is respected by the SDF.

The axiom captures an absolutely minimal notion of liberalism.

Sen's Axioms, Formally

Definition

[U*] $\mathscr{A} = \mathscr{R}^n$

[P*] For any pair of alternatives $x, y \in X$, if a profile $\mathbf{R} \in \mathscr{A}$ satisfies

$$xP_iy \text{ for all } i \in N,$$

then $xd_P(\mathbf{R})y$ must be true.

[L*] There exists a rights system $\mathbf{D} = (D_1, \ldots, D_n)$, with $D_i \neq \varnothing$ for at least two different voters $i \in N$, that is respected by d.

Sen's Impossibility Result

Theorem

There is no SDF d that satisfies [U*], [P*] and [L*].

- This is also called impossibility of a Paretian liberal or liberal paradox.
- If you are a liberal, you have to give up Pareto efficiency.

  If you insist on efficiency, you cannot grant rights to more than one person.
- The result requires no assumptions about the number of alternatives m.
- Clearly, the impossibility result for SDFs implies the corresponding result for SWFs. An analogous version can also be proven for SCFs.

Example IV

Consider our example again:

- Assume that preferences are given by

$$x \; P_1 \; z \; P_1 \; w \; P_1 \; y \quad \text{and} \quad y \; P_2 \; z \; P_2 \; w \; P_2 \; x.$$

  Person 1 prefers being a butcher, but finds science extremely important. Person 2's preferences are as before.

- Let individual rights be given by $D_1 = \{(w,y),(y,w),(x,z),(z,x)\}$ and $D_2 = \{(w,x),(x,w),(y,z),(z,y)\}$.
- Suppose we respect these rights. We then obtain $wd_P(\mathbf{R})y$ and $yd_P(\mathbf{R})z$. If we respect [P*], then we also obtain $zd_P(\mathbf{R})w$. But this is a strict cycle!
- The above preferences essentially turn our occupational choice problem into a prisoners' dilemma. If everybody exercises the own rights, the outcome is not Pareto efficient. But such preferences are always admissible by [U*].

Proof of Sen's Impossibility Result

Proof:

By contradiction, suppose d is an SDF that satisfies [U*], [P*] and [L*].

By [L*], there exists a rights system $\mathbf{D} = (D_1, \ldots, D_n)$ that is respected by d, with $D_i \neq \varnothing$ and $D_j \neq \varnothing$ for some $i, j \in N$, $i \neq j$. Denote by $(x, y) \in D_i$ and $(z, w) \in D_j$ two pairs of alternatives over which i and j, respectively, are decisive.

Suppose first that w, x, y, z are all distinct alternatives. Consider a preference profile $\mathbf{R}$ where xP_iy, zP_jw, and wP_kx, yP_kz for all $k \in N$, which is admissible by [U*]. Now [L*] implies $xd_P(\mathbf{R})y$ and $zd_P(\mathbf{R})w$, and [P*] implies $wd_P(\mathbf{R})x$ and $yd_P(\mathbf{R})z$. Hence we have a strict cycle, contradicting that d is an SDF.

Now suppose (x, y) and (z, w) have one alternative in common, say $z = x$ without loss of generality. Consider a preference profile $\mathbf{R}$ where xP_iy, wP_jx, and yP_kw for all $k \in N$, which is admissible by [U*]. Now [L*] implies $xd_P(\mathbf{R})y$ and $wd_P(\mathbf{R})x$, and [P*] implies $yd_P(\mathbf{R})w$. Hence we again have a contradiction.

If (x, y) and (z, w) are the same pair, the contradiction is immediate. $\square$

Discussion

- As with Arrow's impossibility result, we can think about relaxing the axioms. For instance, we could think about a restricted domain $\mathscr{A}$ where all voters are "mutually disinterested", or a different notion of liberalism...
- Does Sen's result contradict well-known market efficiency results?

  A competitive market assigns an equilibrium allocation to each individual preference profile, hence you can think of it as an SCF. This Walrasian SCF satisfies both [P*] and [L*], as defined for SCFs.

  But this holds only for specific preference profiles! The first fundamental theorem of welfare economics, for instance, assumes that consumers care only about their own consumption. With interdependent preferences, markets can suffer from externalities, and the outcome can be inefficient. Hence [U*] is not satisfied.
- With individual rights, situations related to the prisoners' dilemma can arise. But such dilemmas can be solved by writing contracts (Coase Theorem). Here, this would require that individuals trade or give up their rights D_i, some of which might be human rights that cannot/should not be traded.

Part 5: Manipulability

The Gibbard-Satterthwaite Theorem

This part relies on [GA] and [MC].

Unobservability

Up to this point, we have studied the problem of aggregating preferences, implicitly assuming that we know these preferences. This requires that either

- preferences are directly observable, or
- voters voluntarily reveal their preferences truthfully.

Since direct observability is usually not given in practice, we will now study the incentives of voters to reveal or not reveal their true preferences in elections.

Example: Instant-Runoff Voting I

Consider the following profile of true preferences:

#	true preferences
3	$x\ P\ y\ P\ z$
3	$z\ P\ y\ P\ x$
2	$y\ P\ z\ P\ x$

If we apply IR to this profile, we eliminate y first, and then select z.

This outcome is really bad for the first three voters. Suppose one of them pretends to have preference $yPzPx$ instead. Then the ballot looks as follows:

#	ballot
2	$x\ P\ y\ P\ z$
3	$z\ P\ y\ P\ x$
3	$y\ P\ z\ P\ x$

If we apply IR to this profile, we eliminate x first, and then select y.

Example: Instant-Runoff Voting II

By misrepresenting the own preference, the voter can manipulate the outcome from z to y. Observe that he strictly prefers y to z according to the *true* preference. Hence, provided that everybody else votes truthfully, this voter has an incentive to vote strategically and not reveal the own preference truthfully.

We still don't know what would actually happen in this election, but we know that truthful voting is not a stable situation with IR in this example.

Example: Borda Count I

Consider the following profile of true preferences and apply Borda count:

#	true preferences	w	x	y	z
1	$w\ P\ x\ P\ y\ P\ z$	3	2	1	0
1	$x\ P\ y\ P\ w\ P\ z$	1	3	2	0
1	$z\ P\ y\ P\ x\ P\ w$	0	1	2	3
		4	6	5	3

We obtain the social preference $xPyPwPz$ and select x.

Suppose the third voter pretends to have preference $yPzPwPx$ instead:

#	ballot	w	x	y	z
1	$w\ P\ x\ P\ y\ P\ z$	3	2	1	0
1	$x\ P\ y\ P\ w\ P\ z$	1	3	2	0
1	$y\ P\ z\ P\ w\ P\ x$	1	0	3	2
		5	5	6	2

We obtain the social preference $yPwIxPz$ and select y.

Example: Borda Count II

The voter has again successfully manipulated the election, because she strictly prefers y over x according to her true preference. Truthful voting is not a stable situation with BC in this example.

These examples illustrate that the assumption of truthful voting is not innocuous!

The Problem of Manipulability

Why is manipulability of a voting method a problem?

- When we allow for strategic voting, a distinction arises between the voting method (the game) and its outcome (the induced SCF).

  For instance, if we apply the voting method IR, it is not guaranteed that the outcome is the SCF c^{IR} as applied to the true preferences.

  The actual outcome might satisfy/violate completely different axioms.

- Some voters might better understand than others how to manipulate. This could be considered as unfair.

Formal Framework

- Finite set X of alternatives, with $m = |X|$.
- Set of voters/players $N = \{1, \ldots, n\}$, with preferences $R_i \in \mathscr{R}$.
- We consider SCFs $c : \mathscr{A} \to X$, where $\mathscr{A} \subseteq \mathscr{R}^n$.

  An SCF is surjective (or onto) if for each $x \in X$ there exists some $\mathbf{R} \in \mathscr{A}$ such that $c(\mathbf{R}) = x$. This is also written as $c(\mathscr{A}) = X$.

Reminder: When $m \geq 3$, there is no SCF c that satisfies [$\bar{\text{U}}$], [$\bar{\text{M}}$], [$\bar{\text{P}}$] and [$\bar{\text{D}}$].

In the following, we skip the axioms [$\bar{\text{M}}$] and [$\bar{\text{P}}$], so there are viable SCFs.
We will, however, introduce a new axiom that captures non-manipulability.

Gibbard-Satterthwaite's Axioms, Intuitively

- [$\bar{\text{U}}$] Universality:

  The SCF should work for all conceivable preference profiles.

- [$\bar{\text{D}}$] Non-Dictatorship:

  There should be no individual such that the winner is always one of the top-ranked alternatives of that individual.

- [$\bar{\text{S}}$] Strategy-proofness:

  It should not be possible for any individual to manipulate the outcome through insincere voting, for no possible profile of true preferences.

Gibbard-Satterthwaite's Axioms, Formally

Definition

An SCF $c : \mathscr{A} \to X$ is manipulable by voter i at profile $\mathbf{R} \in \mathscr{A}$ if there exists a profile $\mathbf{R}' \in \mathscr{A}$ with $R'_j = R_j$, $\forall j \neq i$, such that $c(\mathbf{R}')P_i c(\mathbf{R})$.

Definition

- [$\bar{\text{U}}$] $\mathscr{A} = \mathscr{R}^n$ (or $\mathscr{A} = \mathscr{P}^n$)
- [$\bar{\text{D}}$] There is no $i \in N$ such that, for all $\mathbf{R} \in \mathscr{A}$, $c(\mathbf{R})R_i y$ holds for all $y \in X$.
- [$\bar{\text{S}}$] There is no $i \in N$ and $\mathbf{R} \in \mathscr{A}$ such that c is manipulable by i at $\mathbf{R}$.

Gibbard-Satterthwaite's Theorem

Theorem

If $m \geq 3$, there is no surjective SCF c that satisfies [$\bar{\mathrm{U}}$], [$\bar{\mathrm{D}}$], and [$\bar{\mathrm{S}}$].

- This is [GA]'s version of the Gibbard-Satterthwaite theorem.
- If $m \geq 3$ and we consider surjective SCFs on the universal domain, then...
  - ...strategy-proofness implies that the SCF must be dictatorial.
  - ...non-dictatorship implies that the SCF is manipulable.
- Without surjectiveness, an SCF could select among only two alternatives, turning any problem into a problem with $m = 2$.

  We did not need this requirement in part 3 of the lecture, because such a "solution" would have violated [$\bar{\mathrm{P}}$], which is not imposed here.
- Our proof will follow [MC], with small notational adjustments.

  We will prove the theorem only for the case where [$\bar{\mathrm{U}}$] requires $\mathscr{A} = \mathscr{P}^n$. It can also be proven for $\mathscr{A} = \mathscr{R}^n$.

Gibbard-Satterthwaite's Theorem, Proof I

Proof:

Assume $m \geq 3$ and let c be a surjective SCF that satisfies [$\bar{\text{U}}$] and [$\bar{\text{S}}$]. We will show that it must violate [$\bar{\text{D}}$]. We do this by showing that c must satisfy [$\bar{\text{M}}$] and [$\bar{\text{P}}$]. The conclusion then follows from Arrow's impossibility theorem for SCFs.

[$\bar{\text{M}}$]: Consider any $\mathbf{R} \in \mathscr{P}^n$ and denote by $x = c(\mathbf{R})$ the winner. Suppose x maintains its position from $\mathbf{R}$ to $\mathbf{R}' \in \mathscr{P}^n$. We need to show that $x = c(\mathbf{R}')$.

Consider profile $\mathbf{R}'' = (R_1', R_2, \ldots, R_n)$. By [$\bar{\text{S}}$] we must have $xR_1c(\mathbf{R}'')$, since otherwise voter 1 could manipulate at $\mathbf{R}$. Since x maintains its position from $\mathbf{R}$ to $\mathbf{R}'$, we then also have $xR_1'c(\mathbf{R}'')$. By [$\bar{\text{S}}$] we must have $c(\mathbf{R}'')R_1'x$, since otherwise voter 1 could manipulate at $\mathbf{R}''$. Hence $c(\mathbf{R}'')I_1'x$, which implies $x = c(\mathbf{R}'')$ by strictness of preferences.

The argument can now be iterated, switching one voter's preference from R_i to R_i' at a time, until arriving at the desired conclusion $x = c(\mathbf{R}')$. $\diamond$

Gibbard-Satterthwaite's Theorem, Proof II

[$\bar{\text{P}}$]: By contradiction, suppose [$\bar{\text{P}}$] is violated, so there exists $\mathbf{R} \in \mathscr{P}^n$ and $x, y \in X$, $x \neq y$, such that xP_iy, $\forall i \in N$, but $y = c(\mathbf{R})$.

Now consider any $\mathbf{R}' \in \mathscr{P}^n$ where, for all $i \in N$ and all $z \neq x, y$, $x\underline{P}'_i yP'_i z$ holds. Since y maintains its position from $\mathbf{R}$ to $\mathbf{R}'$ and c satisfies [$\bar{\text{M}}$] as shown before, we must have $y = c(\mathbf{R}')$.

Since c is surjective, there exists $\mathbf{R}'' \in \mathscr{P}^n$ for which $x = c(\mathbf{R}'')$. But x maintains is position from $\mathbf{R}''$ to $\mathbf{R}'$, where it is uniquely ranked top for every $i \in N$. Hence [$\bar{\text{M}}$] implies $x = c(\mathbf{R}')$, a contradiction (because $x \neq y$). $\diamond$ $\square$

A Game-Theoretic Approach

The axiom of strategy-proofness can be expressed in game-theoretic terms:

- Suppose $\mathscr{A} = \mathscr{R}^n$ and fix true preferences $\mathbf{R} = (R_1, \dots, R_n) \in \mathscr{R}^n$. Assume for now that preferences are common knowledge among voters.
- In an election, each voter can submit an arbitrary preference $\hat{R}_i \in \mathscr{R}$, so insincere voting $\hat{R}_i \neq R_i$ is possible.
- When $\hat{\mathbf{R}} = (\hat{R}_1, \dots, \hat{R}_n) \in \mathscr{R}^n$ has been submitted, the outcome is $c(\hat{\mathbf{R}})$.
- This is a normal-form game $\Gamma(\mathbf{R})$, where the n voters are the players. Each voter's strategy set is $\mathscr{R}$. Outcomes are evaluated by the true preferences $\mathbf{R}$.
- Truthful voting $\hat{\mathbf{R}} = \mathbf{R}$ is a Nash equilibrium of game $\Gamma(\mathbf{R})$ if and only if c is not manipulable by any voter at $\mathbf{R}$.

Dominant Strategies

Strategy-proofness [$\bar{\text{S}}$] requires that truthful voting is a Nash equilibrium in *every* game $\Gamma(\mathbf{R})$, for all $\mathbf{R} \in \mathscr{R}^n$ (also called ex-post Nash equilibrium).

In our context, this implies that truthful voting is even a (weakly) dominant strategy for every voter in every game $\Gamma(\mathbf{R})$, $\mathbf{R} \in \mathscr{R}^n$.

Proof: Suppose voting truthfully is not weakly dominant for some $i \in N$ in $\Gamma(\mathbf{R})$ for some $\mathbf{R} = (R_i, R_{-i})$, i.e. there exists $\hat{R}_{-i} \neq R_{-i}$ and $R_i' \neq R_i$ such that $c(R_i', \hat{R}_{-i}) P_i c(R_i, \hat{R}_{-i})$. Then truthful voting is not a Nash equilibrium in $\Gamma(\mathbf{R}'')$ for $\mathbf{R}'' = (R_i, \hat{R}_{-i})$, because i prefers to announce $R_i' \neq R_i$, a contradiction. $\square$

Hence [$\bar{\text{S}}$] implies that it is not necessary for voters to know the preferences and strategies of the other voters to make truthful voting optimal, which is a great advantage. On the other hand, [$\bar{\text{S}}$] is clearly a very strong requirement.

Solutions

There are several ways out of the Gibbard-Satterthwaite impossibility theorem:

- For $m = 2$, majority voting clearly satisfies [$\bar{\text{U}}$], [$\bar{\text{D}}$] and [$\bar{\text{S}}$].
- [$\bar{\text{U}}$] can be relaxed. If n is odd and (true and announced) preferences are single-peaked, then c^{PM} is surjective and satisfies [$\bar{\text{D}}$] and [$\bar{\text{S}}$].
- [$\bar{\text{S}}$] can be relaxed. For instance, we could require only Bayes Nash equilibrium instead of dominant strategies.
- We could compare different non-strategy-proof SCFs according to how often or easily they can be manipulated.
  Manipulation might also become less of a problem as $n \to \infty$.

Such issues belong to the field of mechanism design theory, which studies in detail the problem of social choice without observability of individual preferences.

Part 6: Distributive Justice

Social Evaluation Functions

This part relies on [GA] and [BK].

Preference Aggregation or Utility Aggregation

We have now studied many different aspects of the aggregation of preferences. Preferences capture pairwise comparisons and as such are an ordinal concept.

In economics, we often work with utility functions instead of preferences:

- A utility function may simply represent a preference.
  Working with functions is easier than working with binary relations.
  Utility is interpreted only ordinally in this case.
- A utility function might be more substantive than that.
  Maybe utility can be interpreted as happiness?
  In that case, utility aggregation can rely on more than ordinal information.

Whether or not utility is more than ordinal is an unsettled problem (in philosophy and, more recently, also in neuroscience).

Preferences and Utility

Fix a set of alternatives X and consider a single decision-maker.

- A preference R is a reflexive, complete and transitive binary relation on X. The set of all possible preferences is denoted $\mathscr{R}$.
- A utility function U is a mapping $U : X \to \mathbb{R}$. The set of all possible utility functions is denoted $\mathscr{U}$.

Definition

Utility function $U \in \mathscr{U}$ represents preference $R \in \mathscr{R}$ if, for any pair $x, y \in X$,

$$U(x) \geq U(y) \quad \leftrightarrow \quad xRy.$$

Proposition

Suppose $U \in \mathscr{U}$ represents $R \in \mathscr{R}$. Then $\arg\max_{x \in S} U(x) = G(S, R), \ \forall S \subseteq X$.

Proof:
For any $S \subseteq X$, we have $\arg\max_{x \in S} U(x) = \{x \in S | U(x) \geq U(y), \forall y \in S\} = \{x \in S | xRy, \forall y \in S\} = G(S, R)$. □

Utility Transformation I

Definition

A function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is called a

- strictly increasing transformation if $a > b$ implies $\varphi(a) > \varphi(b)$, $\forall a, b \in \mathbb{R}$.
- positive affine transformation if

$$\varphi(a) = \alpha + \beta a$$

  for some $\alpha, \beta \in \mathbb{R}$ with $\beta > 0$, $\forall a \in \mathbb{R}$.

- positive linear transformation if

$$\varphi(a) = \beta a$$

  for some $\beta \in \mathbb{R}$ with $\beta > 0$, $\forall a \in \mathbb{R}$.

Utility Transformation II

- Any positive linear transformation is a positive affine transformation. Any positive affine transformation is a strictly increasing transformation.
- Given a utility function $U \in \mathscr{U}$ and a function $\varphi : \mathbb{R} \to \mathbb{R}$, we can define another utility function $V \in \mathscr{U}$ by $V(x) = \varphi(U(x))$ for all $x \in X$. We also write $V = \varphi(U)$ in that case.

Proposition

Suppose $U \in \mathscr{U}$ represents $R \in \mathscr{R}$. Then any $V \in \mathscr{U}$ given by $V = \varphi(U)$ for some strictly increasing transformation φ also represents R.

Proof:
For any pair $x, y \in X$, we have

$$V(x) \geq V(y) \;\leftrightarrow\; \varphi(U(x)) \geq \varphi(U(y)) \;\leftrightarrow\; U(x) \geq U(y) \;\leftrightarrow\; xRy,$$

so that V also represents R. $\square$

Formal Framework

- X is a set of alternatives.
- $N = \{1, \dots, n\}$, $n \geq 2$, is a set of agents.
- We denote by $R_i \in \mathscr{R}$ a preference of agent i.
  Profiles of preferences are denoted $\mathbf{R} = (R_1, \dots, R_n) \in \mathscr{R}^n$.
- We denote by $U_i \in \mathscr{U}$ a utility function of agent i.
  Profiles of utility functions are denoted $\mathbf{U} = (U_1, \dots, U_n) \in \mathscr{U}^n$.
- We say that $\mathbf{U}$ represents $\mathbf{R}$ if U_i represents R_i, for each $i \in N$.

Social Evaluation Functions

Definition

- A social welfare function (SWF) is a mapping $f : \mathscr{R}^n \to \mathscr{R}$.
- A social evaluation function (SEF) is a mapping $e : \mathscr{U}^n \to \mathscr{R}$.

Interpretation:

- We implicitly make the universal domain assumption here.
- An SEF is a rule that delivers a preference of society $e(\mathbf{U}) \in \mathscr{R}$ for every possible profile of utility functions of the agents $\mathbf{U} = (U_1, \ldots, U_n) \in \mathscr{U}^n$.
- We denote by $e_P(\mathbf{U})$ and $e_I(\mathbf{U})$ the asymmetric and symmetric parts.
- SEFs are sometimes also called social evaluation functionals, because their arguments are (profiles of) functions.

From SWFs to SEFs

- For any given SWF f we can construct a corresponding SEF e by

$$e(\mathbf{U}) = f(\mathbf{R})$$

  where $\mathbf{R}$ is the profile of preferences represented by $\mathbf{U}$, for any $\mathbf{U} \in \mathscr{U}^n$.

- The resulting SEF uses only the ordinal information in the utility functions: If $\mathbf{U}$ and $\mathbf{U}'$ represent the same $\mathbf{R}$, then $e(\mathbf{U}) = e(\mathbf{U}')$.

- In general, however, SEFs can use more information than that: Even if $\mathbf{U}$ and $\mathbf{U}'$ represent the same $\mathbf{R}$, we could have $e(\mathbf{U}) \neq e(\mathbf{U}')$.

- Hence we have more degrees of freedom with SEFs than with SWFs. We will now make precise how much information a given SEF actually uses.

Information Structures

Definition

An SEF e is consistent with the utility functions being

- ordinally measurable and non-comparable (OM-NC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi_i(U_i)$ for some (i-specific) strictly increasing transformation φ_i, $\forall i \in N$.
- ordinally measurable and level-comparable (OM-LC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi(U_i)$ for some (common) strictly increasing transformation φ, $\forall i \in N$.
- cardinally measurable and non-comparable (CM-NC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi_i(U_i)$ for some (i-specific) positive affine transformation φ_i, $\forall i \in N$.
- cardinally measurable and unit-comparable (CM-UC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi_i(U_i)$ for some positive affine transformation $\varphi_i(a) = \alpha_i + \beta a$, $\forall i \in N$.
- cardinally measurable and level-comparable (CM-LC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi(U_i)$ for some (common) positive affine transformation φ, $\forall i \in N$.
- ratio measurable and non-comparable (RM-NC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi_i(U_i)$ for some (i-specific) positive linear transformation φ_i, $\forall i \in N$.
- ratio measurable and level-comparable (RM-LC), if $e(\mathbf{U}) = e(\mathbf{U}')$ whenever $U_i' = \varphi(U_i)$ for some (common) positive linear transformation φ, $\forall i \in N$.

Overview of Information Structures

	no comparability	unit comparability	level comparability
ordinal measurability	OM-NC		OM-LC
cardinal measurability	CM-NC	CM-UC	CM-LC
ratio measurability	RM-NC		RM-LC

Measurability and Comparability

Consider the measurability dimension first, holding agent i fixed:

- Ordinal measurability allows only ranking statements (e.g. sports leagues):
  $U_i(x)$ is larger than $U_i(y)$.
- Cardinal measurability allows also difference statements (e.g. temperature):
  $U_i(w) - U_i(x)$ is twice as large as $U_i(y) - U_i(z)$.
- Ratio measurability allows also ratio statements (e.g. money):
  $U_i(x)$ is twice as large as $U_i(y)$.

Comparability captures whether such statements are also possible across agents:

- OM-LC: *$U_i(x)$ is larger than $U_j(x)$.*
- CM-UC: *$U_i(w) - U_i(x)$ is twice as large as $U_j(w) - U_j(x)$.*
- CM-LC: *$U_i(w) - U_j(w)$ is twice as large as $U_i(x) - U_j(x)$.*
- RM-LC: *$U_i(x)$ is twice as large as $U_j(x)$.*

Minimal Information Requirements

For any given SEF e, we can identify the information structures with which it is consistent. However, we do not have to check all seven of them separately:

- The set of admissible transformations shrinks as we strengthen measurability (OM $\rightarrow$ CM $\rightarrow$ RM) and/or comparability (NC $\rightarrow$ UC $\rightarrow$ LC) requirements.
- If SEF e is consistent with some structure, it is also consistent with all unambiguously stronger structures. For example, consistency with CM-UC implies consistency with CM-LC and RM-LC.
- If SEF e is not consistent with some structure, it is also not consistent with all unambiguously weaker structures. For example, inconsistency with CM-UC implies inconsistency with CM-NC and OM-NC.
- Some information structures cannot be compared unambiguously and hence have to be checked separately, for example CM-UC and OM-LC.

Borda Count I

Suppose X is finite, with $m = |X|$. For any $x \in X$ and $U_i \in \mathscr{U}$, define $r(x, U_i) = r(x, R_i)$ for the $R_i \in \mathscr{R}$ that is represented by U_i, and let

$$BC(x, \mathbf{U}) = \sum_{i=1}^{n} (m - r(x, U_i)).$$

Then $e^{BC}(\mathbf{U})$ is the preference represented by $BC(., \mathbf{U})$, that is

$$x e^{BC}(\mathbf{U}) y \quad \leftrightarrow \quad BC(x, \mathbf{U}) \geq BC(y, \mathbf{U}),$$

for all $x, y \in X$.

Borda Count II

Proposition

e^{BC} is consistent with OM-NC.

Proof:

Consider any pair $\mathbf{U}, \mathbf{U}' \in \mathscr{U}^n$ where $U_i' = \varphi_i(U_i)$ for some strictly increasing transformation φ_i, $\forall i \in N$. For any $x \in X$ we then have

$$BC(x, \mathbf{U}') = \sum_{i=1}^{n}(m - r(x, U_i')) = \sum_{i=1}^{n}(m - r(x, U_i)) = BC(x, \mathbf{U}),$$

so that $e^{BC}(\mathbf{U}) = e^{BC}(\mathbf{U}')$. $\square$

Utilitarian Welfare I

For any $x \in X$ and $\mathbf{U} \in \mathscr{U}^n$, let

$$UT(x, \mathbf{U}) = \sum_{i=1}^{n} U_i(x).$$

Then $e^{UT}(\mathbf{U})$ is the preference represented by $UT(., \mathbf{U})$, that is

$$xe^{UT}(\mathbf{U})y \quad \leftrightarrow \quad UT(x, \mathbf{U}) \geq UT(y, \mathbf{U}),$$

for all $x, y \in X$.

Discussion: Utilitarianism dates back to at least the 18th century, e.g. Jeremy Bentham. A modern proponent was John Harsanyi.

Utilitarian Welfare II

Proposition

e^{UT} is consistent with CM-UC, but not with RM-NC or OM-LC.

Proof:

- CM-UC: Consider any pair $\mathbf{U}, \mathbf{U}' \in \mathscr{U}^n$ where $U_i' = \varphi_i(U_i)$ for some positive affine transformation $\varphi_i(a) = \alpha_i + \beta a$, $\forall i \in N$. For any $x \in X$ we then have

$$UT(x, \mathbf{U}') = \sum_{i=1}^{n} U_i'(x) = \sum_{i=1}^{n} (\alpha_i + \beta U_i(x)) = \sum_{i=1}^{n} \alpha_i + \beta \sum_{i=1}^{n} U_i(x) = \alpha + \beta UT(x, \mathbf{U}),$$

so that $e^{UT}(\mathbf{U}) = e^{UT}(\mathbf{U}')$.

Utilitarian Welfare III

- RM-NC: $n = m = 2$, positive linear transformations $\beta_1 = 3, \beta_2 = 1$.

U	x	y
U_1	1	2
U_2	5	3
UT	6	5

U$'$	x	y
U_1'	3	6
U_2'	5	3
UT	8	9

- OM-LC: $n = m = 2$, strictly increasing transformation $\varphi(a) = a^3$.

U	x	y
U_1	2	4
U_2	6	5
UT	8	9

U$'$	x	y
U_1'	8	64
U_2'	216	125
UT	224	189

□

Maximin Welfare I

For any $x \in X$ and $\mathbf{U} \in \mathscr{U}^n$, let

$$MM(x, \mathbf{U}) = \min\{U_1(x), \ldots, U_n(x)\}.$$

Then $e^{MM}(\mathbf{U})$ is the preference represented by $MM(., \mathbf{U})$, that is

$$x e^{MM}(\mathbf{U}) y \quad \leftrightarrow \quad MM(x, \mathbf{U}) \geq MM(y, \mathbf{U}),$$

for all $x, y \in X$.

Discussion: Maximin is often used as a formalization of John Rawls' difference principle (this amounts to a severe shortening of Rawls' ideas).

Maximin Welfare II

Proposition

e^{MM} is consistent with OM-LC, but not with CM-UC or RM-NC.

Proof:

- OM-LC: Consider any pair $\mathbf{U}, \mathbf{U}' \in \mathscr{U}^n$ where $U_i' = \varphi(U_i)$ for some strictly increasing transformation φ, $\forall i \in N$. For any $x \in X$ we then have

$$\begin{aligned} MM(x, \mathbf{U}') &= \min\{U_1'(x), \ldots, U_n'(x)\} \\ &= \min\{\varphi(U_1(x)), \ldots, \varphi(U_n(x))\} \\ &= \varphi(\min\{U_1(x), \ldots, U_n(x)\}) \\ &= \varphi(MM(x, \mathbf{U})), \end{aligned}$$

so that $e^{MM}(\mathbf{U}) = e^{MM}(\mathbf{U}')$.

Maximin Welfare III

- CM-UC: $n = m = 2$, positive affine transformations $\alpha_1 = 3, \alpha_2 = 0, \beta = 1$.

U	x	y
U_1	1	2
U_2	5	3
MM	1	2

U$'$	x	y
U_1'	4	5
U_2'	5	3
MM	4	3

- RM-NC: $n = m = 2$, positive linear transformations $\beta_1 = 4, \beta_2 = 1$

U	x	y
U_1	1	2
U_2	5	3
MM	1	2

U$'$	x	y
U_1'	4	8
U_2'	5	3
MM	4	3

□

Leximin Welfare I

Consider e^{MM} in the following example:

$\mathbf{U}$	x	y
U_1	1	1
U_2	3	2
MM	1	1

Society is indifferent between x and y, even though x Pareto dominates y.

Leximin is a variant of maximin that alleviates this problem. For any $x \in X$ and $\mathbf{U} \in \mathscr{U}^n$, let $i_k(x, \mathbf{U}) \in N$ be the k'th best-off person in x under $\mathbf{U}$, i.e.

$$U_{i_1(x,\mathbf{U})}(x) > U_{i_2(x,\mathbf{U})}(x) > \ldots > U_{i_n(x,\mathbf{U})}(x),$$

where we abstract from ties.

In the above example: $i_1(x, \mathbf{U}) = 2$, $i_2(x, \mathbf{U}) = 1$, $i_1(y, \mathbf{U}) = 2$, $i_2(y, \mathbf{U}) = 1$.

Leximin Welfare II

The previous definition of e^{MM} can be rewritten as

$$xe^{MM}(\mathbf{U})y \quad \leftrightarrow \quad U_{i_n(x,\mathbf{U})}(x) \geq U_{i_n(y,\mathbf{U})}(y),$$

for all $x, y \in X$.

The SEF e^{LM} is defined by $xe_P^{LM}(\mathbf{U})y$ if and only if there exists $h \in N$ such that

$$U_{i_h(x,\mathbf{U})}(x) > U_{i_h(y,\mathbf{U})}(y)$$

while

$$U_{i_k(x,\mathbf{U})}(x) = U_{i_k(y,\mathbf{U})}(y)$$

for all $k > h$. We move up the ranks until a difference between x and y appears.

Leximin Welfare III

Proposition

e^{LM} is consistent with OM-LC, but not with CM-UC or RM-NC.

Proof:

- OM-LC: Consider any pair $\mathbf{U}, \mathbf{U}' \in \mathscr{U}^n$ where $U_i' = \varphi(U_i)$ for some strictly increasing transformation φ, $\forall i \in N$. For any $x \in X$ and any $k \in N$ we then have $i_k(x, \mathbf{U}') = i_k(x, \mathbf{U})$ and therefore

$$U'_{i_k(x,\mathbf{U}')}(x) = \varphi(U_{i_k(x,\mathbf{U}')}(x)) = \varphi(U_{i_k(x,\mathbf{U})}(x)).$$

  Hence the argument about the n'th rank for e^{MM} generalizes to any rank, and we obtain $e^{LM}(\mathbf{U}) = e^{LM}(\mathbf{U}')$.

- CM-UC: See the counterexample for e^{MM}.
- RM-NC: See the counterexample for e^{MM}. □

Axioms, Intuitively

As for SWFs, we can postulate axioms for SEFs:

- [$\tilde{\text{U}}$] Universality (has implicitly been assumed already)
- [$\tilde{\text{I}}$] Independence of Irrelevant Alternatives in Utility Terms:

  The social ranking of two alternatives should not change if no agent's utility from any of the two changes.
- [$\tilde{\text{P}}$] Strong Pareto Principle:

  If all agents weakly prefer x over y, so should society. If, in addition, at least one agent strictly prefers x, society should also strictly prefer x.
- [$\tilde{\text{A}}$] Anonymity:

  The SEF should treat all agents equally, their names should not matter.

Axioms, Formally

Definition

[Ũ] $\mathscr{A} = \mathscr{U}^n$

[Ĩ] For any pair of alternatives $x, y \in X$, if two profiles $\mathbf{U}, \mathbf{U}' \in \mathscr{A}$ satisfy

$$U_i(x) = U_i'(x) \text{ and } U_i(y) = U_i'(y) \text{ for all } i \in N,$$

then

$$xe(\mathbf{U})y \;\leftrightarrow\; xe(\mathbf{U}')y \;\text{ and }\; ye(\mathbf{U})x \;\leftrightarrow\; ye(\mathbf{U}')x.$$

[P̃] For any pair of alternatives $x, y \in X$, if a profile $\mathbf{U} \in \mathscr{A}$ satisfies

$$U_i(x) \geq U_i(y) \text{ for all } i \in N,$$

then $xe(\mathbf{U})y$. If, in addition, $U_j(x) > U_j(y)$ for some $j \in N$, then $xe_P(\mathbf{U})y$.

[Ã] For all $\mathbf{U}, \mathbf{U}' \in \mathscr{A}$, if $\mathbf{U}$ and $\mathbf{U}'$ are permutations of each other, then $e(\mathbf{U}) = e(\mathbf{U}')$ must be true.

Axioms, Remarks

- Consider an SEF e that is consistent with OM-NC.

  In that case, the axioms [$\tilde{\text{U}}$], [$\tilde{\text{I}}$], [$\tilde{\text{P}}$] and [$\tilde{\text{A}}$] are stronger than the original Arrow axioms [U], [I], [P] and [D] (showing this for [$\tilde{\text{I}}$] is not trivial).

  Hence there is no such SEF that satisfies [$\tilde{\text{U}}$], [$\tilde{\text{I}}$], [$\tilde{\text{P}}$] and [$\tilde{\text{A}}$] when $m \geq 3$.

- Axioms [$\tilde{\text{U}}$], [$\tilde{\text{I}}$], [$\tilde{\text{P}}$] and [$\tilde{\text{A}}$] can be applied to any SEF.

  In this sense, the following possibility results will show that stronger measurability and comparability requirements are another way out of Arrow's impossibility problem.

Another Axiom

[Ẽ] Equity:

Suppose that all but two agents are indifferent between x and y. The two agents have opposite preferences concerning x and y, but one of the two agents is better off than the other agent in both x and y. Then society should follow the worse off agent's preference.

Definition

[Ẽ] For any pair of alternatives $x, y \in X$, if a profile $\mathbf{U} \in \mathscr{A}$ satisfies

$$U_k(x) = U_k(y) \text{ for all } k \in N \backslash \{i, j\}$$

and

$$U_i(y) < U_i(x) < U_j(x) < U_j(y),$$

for some $i, j \in N$, $i \neq j$, then $x e_P(\mathbf{U}) y$.

- Axiom [Ẽ] implicitly assumes OM-LC.

Characterization Results

Theorem

An SEF is consistent with CM-UC and satisfies [Ũ], [Ĩ], [P̃] and [Ã] if and only if it is utilitarian welfare e^{UT}.

- Utilitarian welfare is a way out of Arrow's impossibility result under CM-UC.
- If you accept CM-UC and [Ũ], [Ĩ], [P̃] and [Ã], you are a utilitarian.

Theorem

An SEF is consistent with OM-LC and satisfies [Ũ], [Ĩ], [P̃], [Ã] and [Ẽ] if and only if it is leximin welfare e^{LM}.

- Leximin welfare is a way out of Arrow's impossibility result under OM-LC.
- If you accept OM-LC and [Ũ], [Ĩ], [P̃], [Ã] and [Ẽ], you believe in leximin.
- Removing [Ẽ] allows for some more methods, such as leximax.

Discussion I

- The characterization results reveal all normative judgements implicit in utilitarian and leximin welfare.
- Utilitarianism and leximin seem to be quite opposing views of justice.

  However, their main difference is not in the axioms (except for [$\tilde{\text{E}}$]) but in the assumption about measurability and comparability of utility.

  Hence utilitarianism and leximin capture very similar ideas of justice for different information structures (CM-UC versus OM-LC).
- The information structure puts strong restrictions on what operators are available for utility aggregation.

  CM-UC: Summation works, but not rank comparisons or multiplication.

  OM-LC: Rank comparisons work, but not summation or multiplication.

Discussion II

Aside from axiomatic foundations, contractualistic justifications can be given for utilitarianism and maximin/leximin, using veil of ignorance arguments:

- Suppose you do not yet know which role you will take in society.
- Inequality between people then becomes a lottery.
- Which society/lottery do you prefer?
- Shouldn't we all agree on the same social contract?

This thought experiment turns the question of redistribution into a question of insurance. Our model of lottery choice will determine our conclusion:

- Harsanyi: Expected utility maximization yields utilitarianism.
- Rawls: Extreme risk or uncertainty aversion yields maximin/leximin.