Computational Social Choice: Autumn 2012

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

Plan for Today

Today's lecture will be devoted to classical *impossibility theorems* in social choice theory. We already proved *Arrow's Theorem* using the *"decisive coalition" technique*. Today we'll first review this result and:

- give references to alternative proofs
- discuss the challenge of *automatically* proving Arrow's Theorem

Then we'll see two further classical impossibility theorems:

- *Sen's Theorem* on the Impossibility of a Paretian Liberal (1970)
- the *Muller-Satterthwaite Theorem* (1977)

The former is easy to prove; for the latter we will again use the "decisive coalition" technique.

Arrow's Theorem

Recall terminology and axioms:

- SWF: $F : \mathcal{L}(\mathcal{X})^{\mathcal{N}} \to \mathcal{L}(\mathcal{X})$
- Pareto: $N^{\boldsymbol{R}}_{x \succ y} = \mathcal{N}$ implies $(x, y) \in F(\boldsymbol{R})$
- IIA: $N^{\boldsymbol{R}}_{x \succ y} = N^{\boldsymbol{R}'}_{x \succ y}$ implies $(x, y) \in F(\boldsymbol{R}) \Leftrightarrow (x, y) \in F(\boldsymbol{R}')$
- Dictatorship: $\exists i \in \mathcal{N}$ s.t. $\forall (R_1, \ldots, R_n)$: $F(R_1, \ldots, R_n) = R_i$

Here is again the theorem:

Theorem 1 (Arrow, 1951) *Any SWF for $\geqslant 3$ alternatives that satisfies the Pareto condition and IIA must be a dictatorship.*

K.J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963. First edition published in 1951.

Alternative Proofs

Arrow's book is an inspiring and interesting read, but his proof is very verbose and hard to follow (and the original version of 1951 famously has a small mistake in the theorem). Some alternative proofs:

- Geanakoplos (2005) gives three short proofs. The first one is particularly helpful. It uses the "pivotal voter" technique and is based on earlier work by Barberà (1980).
- Another proof involves showing that the family of decisive coalitions is an ultrafilter for $\mathcal{N}$ (Kirman and Sondermann, 1972).

J. Geanakoplos. Three Brief Proofs of Arrow's Impossibility Theorem. *Economic Theory*, 26(1):211–215, 2005.

S. Barberà (1980). Pivotal Voters: A New Proof of Arrow's Theorem. *Economics Letters*, 6(1):13–16, 1980.

A.P. Kirman and D. Sondermann. Arrow's Theorem, Many Agents, and Invisible Dictators. *Journal of Economic Theory*, 5(3):267–277, 1972.

Automated Reasoning for Social Choice Theory

There've also been attempts to automatise proving Arrow's Theorem:

- Nipkow (2009) has encoded one of Geanakoplos' proofs in the language of the higher-order logic *proof assistant* ISABELLE, resulting in an automatic verification of that proof.
- Tang and Lin (2009) have translated Arrow's Theorem for 2 individuals and 3 alternatives into a set of clauses in propositional logic, which allows for verification by means of a *SAT-solver*.

T. Nipkow. Social Choice Theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automated Reasoning*, 43(3):289–304, 2009.

P. Tang and F. Lin. Computer-aided Proofs of Arrow's and other Impossibility Theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

Logics for Social Choice Theory

More generally, it is interesting to explore the use of logics to model problems studied in SCT. This is still an under-developed strand of research. Below are some references (the last one is a survey).

R. Parikh. The Logic of Games and its Applications. *Annals of Discrete Mathematics*, 24:111–140, 1985.

M. Pauly. On the Role of Language in Social Choice Theory. *Synthese*, 163(2):227–243, 2008.

T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the Logic of Preference and Judgment Aggregation. *Auton. Agents and Multiagent Sys.*, 22(1):4–30, 2011.

U. Grandi and U. Endriss. First-Order Logic Formalisation of Impossibility Theorems in Preference Aggregation. *Journal of Philosophical Logic*. In press (2012).

U. Endriss. Logic and Social Choice Theory. In A. Gupta and J. van Benthem (eds.), *Logic and Philosophy Today*, College Publications, 2011.

Social Choice Functions

From now on we consider aggregators that take a profile of preferences and return one or several "winners" (rather than a full social ranking). This is called a *social choice function* (SCF):

$$F : \mathcal{L}(\mathcal{X})^{\mathcal{N}} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$$

A SCF is called *resolute* if $|F(\boldsymbol{R})| = 1$ for any given profile $\boldsymbol{R}$, i.e., if it always selects a unique winner.

Remark: We can think of a SCF as a *voting rule*, particularly if it tends to select "small" sets of winners (we won't make this precise). Voting rules are often required to be resolute ($\leadsto$ *tie-breaking rule*).

Examples: The plurality and the Borda rule are both (irresolute) SCFs. Approval voting is *not* a SCF (inputs are sets, not linear orders).

Alternative Definition

In the literature you will sometimes find the term SCF being used for functions $F : \mathcal{L}(X)^{\mathcal{N}} \times 2^{\mathcal{X}} \setminus \{\emptyset\} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$. Two readings:

- The input of F is a profile of preferences (as before) + a set of *feasible alternatives*. The output should be a subset of the feasible alternatives, selected in view of the preference profile.
- The input of F is just a profile of preferences (as before). The output is a *choice function* $C : 2^{\mathcal{X}} \setminus \{\emptyset\} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$ that will select a set of winners from any given set of alternatives.

This refinement is not relevant for the results we want to discuss here, so we shall take a SCF to be a function $F : \mathcal{L}(\mathcal{X})^{\mathcal{N}} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$.

The Pareto Condition for Social Choice Functions

A SCF F satisfies the *Pareto condition* if, whenever all individuals rank x above y, then y cannot win:

$$N^{\boldsymbol{R}}_{x \succ y} = \mathcal{N} \text{ implies } y \notin F(\boldsymbol{R})$$

Liberalism

Think of $\mathcal{X}$ as the set of all possible "social states". Certain aspects of such a state will be some individual's private business. Example:

> If x and y are identical states, except that in x I paint my bedroom white, while in y I paint it pink, then I should be able to dictate the relative social ranking of x and y.

Sen (1970) proposed the following axiom:

A SCF F satisfies the axiom of *liberalism* if, for every individual $i \in \mathcal{N}$, there exist two distinct alternatives $x, y \in \mathcal{X}$ such that i is *two-way decisive* on x and y:

$$i \in N^{\boldsymbol{R}}_{x \succ y} \text{ implies } y \notin F(\boldsymbol{R}) \text{ and } i \in N^{\boldsymbol{R}}_{y \succ x} \text{ implies } x \notin F(\boldsymbol{R})$$

A.K. Sen. The Impossibility of a Paretian Liberal. *Journal of Political Economics*, 78(1):152–157, 1970.

The Impossibility of a Paretian Liberal

Sen (1970) showed that liberalism and the Pareto condition are incompatible (recall that we required $|\mathcal{N}| \geqslant 2$, which matters here):

Theorem 2 (Sen, 1970) *No SCF satisfies both liberalism and the Pareto condition.*

As we shall see, the theorem holds even when liberalism is enforced for only two individuals. The number of alternatives does not matter.

Again, a surprising result (but easier to prove than Arrow's Theorem).

A.K. Sen. The Impossibility of a Paretian Liberal. *Journal of Political Economics*, 78(1):152–157, 1970.

Proof

Let F be a SCF satisfying Pareto and liberalism. Get a contradiction:

Take two distinguished individuals i_1 and i_2, with:

- i_1 is two-way decisive on x_1 and y_1
- i_2 is two-way decisive on x_2 and y_2

Assume x_1, y_1, x_2, y_2 are pairwise distinct (other cases: easy).

Consider a profile with these properties:

(1) Individual i_1 ranks $x_1 \succ y_1$.
(2) Individual i_2 ranks $x_2 \succ y_2$.
(3) All individuals rank $y_1 \succ x_2$ and $y_2 \succ x_1$.
(4) All individuals rank x_1, x_2, y_1, y_2 above all other alternatives.

From liberalism: (1) rules out y_1 and (2) rules out y_2 as winner.
From Pareto: (3) rules out x_1 and x_2 and (4) rules out all others.

Thus, there are no winners. Contradiction. ✓

Monotonicity

Next we want to formalise the idea that when a winner receives increased support, she should not become a loser.

We focus on *resolute* SCFs. Write $x^\star = F(\boldsymbol{R})$ for $\{x^\star\} = F(\boldsymbol{R})$.

- *Weak monotonicity:* F is weakly monotonic if $x^\star = F(\boldsymbol{R})$ implies $x^\star = F(\boldsymbol{R}')$ for any alternative $x^\star$ and any two profiles $\boldsymbol{R}$ and $\boldsymbol{R}'$ with $N^{\boldsymbol{R}}_{x^\star \succ y} \subseteq N^{\boldsymbol{R}'}_{x^\star \succ y}$ and $N^{\boldsymbol{R}}_{y \succ z} = N^{\boldsymbol{R}'}_{y \succ z}$ for all $y, z \in \mathcal{X} \setminus \{x^\star\}$.
- *Strong monotonicity:* F is strongly monotonic if $x^\star = F(\boldsymbol{R})$ implies $x^\star = F(\boldsymbol{R}')$ for any alternative $x^\star$ and any two profiles $\boldsymbol{R}$ and $\boldsymbol{R}'$ with $N^{\boldsymbol{R}}_{x^\star \succ y} \subseteq N^{\boldsymbol{R}'}_{x^\star \succ y}$ for all $y \in \mathcal{X} \setminus \{x^\star\}$.

The latter property is also known as *Maskin monotonicity* or *strong positive association*.

Example

Even *weak monotonicity* is not satisfied by some common voting rules.

Under *plurality with runoff* the two alternatives with the highest plurality score enter a second round and the majority winner of that round is the winner (used to elect the French president). Example:

27 voters:	$A \succ B \succ C$
42 voters:	$C \succ A \succ B$
24 voters:	$B \succ C \succ A$

B is eliminated in the first round and C beats A 66:27 in the runoff. But if 4 of the voters in the first group *raise C to the top* (i.e., join the second group), then B will win.

But many other rules (e.g., *plurality*) do satisfy weak monotonicity. How about *strong monotonicity*?

The Muller-Satterthwaite Theorem

Strong monotonicity turns out to be (desirable but) too demanding:

Theorem 3 (Muller and Satterthwaite, 1977) *Any resolute SCF for $\geqslant 3$ alternatives that is surjective and strongly monotonic must be a dictatorship.*

Here, a resolute SCF F is called *surjective* (or *nonimposed*) if for every alternative $x \in \mathcal{X}$ there exists a profile $\boldsymbol{R}$ such that $F(\boldsymbol{R}) = x$.

And: a SCF F is a *dictatorship* if there exists an $i \in \mathcal{N}$ such that $F(R_1, \ldots, R_n) = \mathrm{top}(R_i)$ for every profile $(R_1, \ldots, R_n)$.

Remark: Above theorem, which is what is nowadays usually referred to as the Muller-Satterthwaite Theorem, is in fact a corollary of their main theorem and the better known Gibbard-Satterthwaite Theorem.

E. Muller and M.A. Satterthwaite. The Equivalence of Strong Positive Association and Strategy-Proofness. *Journal of Economic Theory*, 14(2):412–418, 1977.

Proof

We use again the "decisive coalition" technique. Full details are available in the review paper cited below.

Claim: *Any resolute SCF for $\geqslant 3$ alternatives that is surjective and strongly monotonic must be a dictatorship.*

Let F be a SCF for $\geqslant 3$ alt. that is surjective and strongly monotonic.

Proof Plan:

- Show that F must be *independent* (to be defined).
- Show that F must be *Pareto* efficient.
- Prove a version of Arrow's Theorem for SCFs.

U. Endriss. Logic and Social Choice Theory. In A. Gupta and J. van Benthem (eds.), *Logic and Philosophy Today*, College Publications, 2011.

Independence

Call a SCF F *independent* if it is the case that $x \neq y$, $F(\boldsymbol{R}) = x$, and $N^{\boldsymbol{R}}_{x \succ y} = N^{\boldsymbol{R}'}_{x \succ y}$ together imply $F(\boldsymbol{R}') \neq y$.

That is, if y lost to x under profile $\boldsymbol{R}$, and the relative rankings of x vs. y do not change, then y will still lose (possibly to a different winner).

Claim: F is independent.

Proof: Suppose $x \neq y$, $F(\boldsymbol{R}) = x$, and $N^{\boldsymbol{R}}_{x \succ y} = N^{\boldsymbol{R}'}_{x \succ y}$.

Construct a third profile $\boldsymbol{R}''$:

- All individuals rank x and y in the top-two positions.
- The relative rankings of x vs. y are as in $\boldsymbol{R}$, i.e., $N^{\boldsymbol{R}''}_{x \succ y} = N^{\boldsymbol{R}}_{x \succ y}$.
- Rest: whatever

By strong monotonicity, $F(\boldsymbol{R}) = x$ implies $F(\boldsymbol{R}'') = x$.
By strong monotonicity, $F(\boldsymbol{R}') = y$ *would* imply $F(\boldsymbol{R}'') = y$.
Thus, we must have $F(\boldsymbol{R}') \neq y$. ✓

Pareto Condition

Recall the Pareto condition: if everyone ranks $x \succ y$, then y won't win.

Claim: F satisfies the Pareto condition.

Proof: Take any two alternatives x and y.

From surjectivity: x will win for *some* profile $\boldsymbol{R}$.

Starting in $\boldsymbol{R}$, have everyone move x above y (if not above already).
From strong monotonicity: x still wins.

From independence: y does not win for *any* profile where all individuals continue to rank $x \succ y$. ✓

Plan for the Rest of the Proof

We now know that F must be a SCF for $\geqslant 3$ alternatives that is *independent* and *Pareto* efficient. We want to infer that F must be a *dictatorship*.

Call a coalition $G \subseteq \mathcal{N}$ *decisive* on (x, y) iff $G \subseteq N^{\boldsymbol{R}}_{x \succ y} \;\Rightarrow\; y \neq F(\boldsymbol{R})$.

Proof plan:

- Pareto condition $=$ $\mathcal{N}$ is decisive for all pairs of alternatives
- Lemma: G with $|G| \geqslant 2$ *decisive* for all pairs $\Rightarrow$ some $G' \subset G$ as well
- Thus (by induction), there's a decisive coalition of size 1 (a *dictator*).

About Decisiveness

Recall: $G \subseteq \mathcal{N}$ *decisive* on (x, y) iff $G \subseteq N^{\boldsymbol{R}}_{x \succ y} \;\Rightarrow\; y \neq F(\boldsymbol{R})$

Call $G \subseteq \mathcal{N}$ *weakly decisive* on (x, y) iff $G = N^{\boldsymbol{R}}_{x \succ y} \;\Rightarrow\; y \neq F(\boldsymbol{R})$.

Claim: G weakly decisive on (x, y) $\Rightarrow$ G decisive on *any* pair (x', y')

Proof: Suppose x, y, x', y' are all distinct (other cases: similar).

Consider a profile where individuals express these preferences:

- Members of G: $x' \succ x \succ y \succ y'$
- Others: $x' \succ x$, $y \succ y'$, and $y \succ x$ (note that x'-vs.-y' is not specified)
- All rank x, y, x', y' above all other alternatives.

From G being weakly decisive for $(x, y) \Rightarrow y$ must lose
From Pareto $\Rightarrow x$ must lose (to x') and y' must lose (to y)

Thus, x' must win (and y' must lose). By independence, y' will still lose when everyone changes their non-x'-vs.-y' rankings.

Thus, for *any* profile $\boldsymbol{R}$ with $G \subseteq N^{\boldsymbol{R}}_{x' \succ y'}$ we get $y' \neq F(\boldsymbol{R})$. ✓

Contraction Lemma

Claim: If $G \subseteq \mathcal{N}$ with $|G| \geqslant 2$ is a coalition that is decisive on all pairs of alternatives, then so is some nonempty coalition $G' \subset G$.

Proof: Take any nonempty G_1, G_2 with $G = G_1 \cup G_2$ and $G_1 \cap G_2 = \emptyset$.

Recall that there are $\geqslant 3$ alternatives. Consider this profile:

- Members of G_1: $x \succ y \succ z \succ$ *rest*
- Members of G_2: $y \succ z \succ x \succ$ *rest*
- Others: $z \succ x \succ y \succ$ *rest*

As $G = G_1 \cup G_2$ is decisive, z cannot win (it loses to y). Two cases:

(1) The winner is x: Exactly G_1 ranks $x \succ z \Rightarrow$ By independence, in any profile where exactly G_1 ranks $x \succ z$, z will lose (to x) $\Rightarrow G_1$ is weakly decisive on (x, z). Hence (previous slide): G_1 is decisive on all pairs.

(2) The winner is y, i.e., x loses (to y). Exactly G_2 ranks $y \succ x \Rightarrow \cdots \Rightarrow$ G_2 is decisive on all pairs.

Hence, one of G_1 and G_2 will always be decisive. ✓

Summary

We have by now see three important impossibility theorems, establishing the incompatibility of certain desirable properties:

- *Arrow:* Pareto, IIA, nondictatoriality
- *Sen:* Pareto, liberalism
- *Muller-Satterthwaite:* surjectivity, strong monotonicity, nondictat.

We have discussed these results in two formal frameworks (none of the results heavily depend on the choice of framework):

- social welfare functions (*SWF*)
- (resolute) social choice functions (*SCF*)

This has also been an introduction to the *axiomatic method:*

- formulate desirable properties of aggregators as axioms
- explore the consequences of imposing several such axioms

What next?

As discussed, the impossibility theorems we have seen can also be interpreted as axiomatic characterisations of the class of dictatorships.

Soon we will see *characterisations* of more attractive (classes of) voting rules:

- using (again) the axiomatic method; and
- using different methods.

But first we will see more examples for *practical voting rules* and discuss their properties, including also their *algorithmic properties* (how hard is it to compute the winners?).