Logic and Social Choice Theory

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

[http://www.illc.uva.nl/~ulle/teaching/esslli-2013/]

Social Choice Theory

SCT studies collective decision making: how should we aggregate the preferences of the members of a group to obtain a "social preference"?

Agent 1: $\triangle \succ \bigcirc \succ \square$

Agent 2: $\bigcirc \succ \square \succ \triangle$

Agent 3: $\square \succ \triangle \succ \bigcirc$

Agent 4: $\square \succ \triangle \succ \bigcirc$

Agent 5: $\bigcirc \succ \square \succ \triangle$

?

Plan for Today

The purpose of this tutorial is to give an introduction to social choice theory and to highlight the role of logic in the field.

This is the plan for today:

- Examples to introduce some of the concerns of SCT
- Preference aggregation and the axiomatic method
- A classical result: Arrow's Theorem

Example: Electing a President

Remember Florida 2000 (simplified):

49%:	Bush $\succ$ Gore $\succ$ Nader
20%:	Gore $\succ$ Nader $\succ$ Bush
20%:	Gore $\succ$ Bush $\succ$ Nader
11%:	Nader $\succ$ Gore $\succ$ Bush

Questions:

- Who wins?
- Is that a fair outcome?
- What would your advice to the Nader-supporters have been?

Three Voting Rules

How should n *voters* choose from a set of m *alternatives*?

Here are three *voting rules* (there are many more):

- *Plurality:* elect the alternative ranked first most often (i.e., each voter assigns 1 point to an alternative of her choice, and the alternative receiving the most points wins)
- *Plurality with runoff:* run a plurality election and retain the two front-runners; then run a majority contest between them
- *Borda:* each voter gives $m-1$ points to the alternative she ranks first, $m-2$ to the alternative she ranks second, etc.; and the alternative with the most points wins

Example

Consider this election with nine *voters* having to choose from three *alternatives* (namely what drink to order for a common lunch):

4 Dutchmen:	Milk	$\succ$	Beer	$\succ$	Wine
2 Germans:	Beer	$\succ$	Wine	$\succ$	Milk
3 Frenchmen:	Wine	$\succ$	Beer	$\succ$	Milk

Which beverage *wins* the election for

- the plurality rule?
- plurality with runoff?
- the Borda rule?

Positional Scoring Rules

We can generalise the idea underlying the Borda rule as follows:

A *positional scoring rule* is given by a *scoring vector* $s = \langle s_1, \ldots, s_m \rangle$ with $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_m$ and $s_1 > s_m$.

Each voter submits a ranking of the m alternatives. Each alternative receives s_i points for every voter putting it at the ith position.

The alternative(s) with the highest score (sum of points) win(s).

Examples:

- *Borda rule* = PSR with scoring vector $\langle m{-}1, m{-}2, \ldots, 0 \rangle$
- *Plurality rule* = PSR with scoring vector $\langle 1, 0, \ldots, 0 \rangle$
- *Antiplurality rule* = PSR with scoring vector $\langle 1, \ldots, 1, 0 \rangle$
- For any $k \leqslant m$, *k-approval* = PSR with $\langle \underbrace{1, \ldots, 1}_{k}, 0, \ldots, 0 \rangle$

The Condorcet Principle

The Marquis de Condorcet was a public intellectual living in France during the second half of the 18th century.

An alternative that beats every other alternative in pairwise majority contests is called a *Condorcet winner*.

There may be no Condorcet winner; witness the *Condorcet paradox:*

Ann:	$A \succ B \succ C$
Bob:	$B \succ C \succ A$
Cindy:	$C \succ A \succ B$

Whenever a Condorcet winner exists, then it must be *unique*.

A voting rule satisfies the *Condorcet principle* if it elects (only) the Condorcet winner whenever one exists.

M. le Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilté des décisions rendues a la pluralité des voix*. Paris, 1785.

All PSR's Violate the Condorcet Principle (!)

Consider the following example:

3 voters:	$A \succ B \succ C$
2 voters:	$B \succ C \succ A$
1 voter:	$B \succ A \succ C$
1 voter:	$C \succ A \succ B$

A is the *Condorcet winner*; she beats both B and C $4:3$. But any *positional scoring rule* makes B win (because $s_1 \geqslant s_2 \geqslant s_3$):

$$A: \quad 3 \cdot s_1 + 2 \cdot s_2 + 2 \cdot s_3$$
$$B: \quad 3 \cdot s_1 + 3 \cdot s_2 + 1 \cdot s_3$$
$$C: \quad 1 \cdot s_1 + 2 \cdot s_2 + 4 \cdot s_3$$

Thus, *no positional scoring rule* for three (or more) alternatives will satisfy the *Condorcet principle*.

Another Example: Sequential Majority Voting

Yet another rule: *sequential majority voting* means running a series of pairwise majority contests, with the winner always getting promoted to the next stage. This is guaranteed to meet the Condorcet principle.

But there is another problem. Consider this example:

```
        o
       / \
      o   D
     / \
    o   A
   / \
  B   C
```

Take this profile with three agents:

Ann: $A \succ B \succ C \succ D$

Bob: $B \succ C \succ D \succ A$

Cindy: $C \succ D \succ A \succ B$

D *wins!* (despite being dominated by C)

This is a violation of the (weak) *Pareto principle:* if you can make a change that impoves everyone's welfare, then do make that change.

Vilfredo Pareto was an Italian economist active around 1900.

Judgment Aggregation

Preferences are not the only structures we may wish to aggregate.
In JA we aggregate people's judgments regarding complex propositions:

Suppose a court with three judges is considering a case in contract law. Legal doctrine stipulates that the defendant is *liable* *iff* the contract was *valid* (p) and it has been *breached* (q). So we care about $p \wedge q$.

	p	q	$p \wedge q$
Judge 1:	Yes	Yes	Yes
Judge 2:	No	Yes	No
Judge 3:	Yes	No	No

What will/should be the collective decision regarding $p \wedge q$?

Insights so far

Our examples have demonstrated:

- There are different *frameworks* in which we need to aggregate the views of several individuals. They include preference aggregation, voting, and judgment aggregation.
- There are different *methods* of aggregation (especially in voting). We need clear *citeria* for choosing one.
- There are all sorts of *paradoxes* (counterintuitive outcomes). We need to clearly specify *desiderata* for methods of aggregtion to have a chance of understanding these problems.

Today we explore the framework of *preference aggregation* (later on also voting and judgment aggregation, but not, e.g., fair division or matching).

We will focus on the *axiomatic method* when specifying desiderata for assessing aggregators and exploring their consequences.

Formal Framework: Preference Aggregation

Basic terminology and notation:

- finite set of *individuals* $\mathcal{N} = \{1, \ldots, n\}$, with $n \geqslant 2$
- (usually finite) set of *alternatives* $\mathcal{X} = \{x_1, x_2, x_3, \ldots\}$
- Denote the set of *linear orders* on $\mathcal{X}$ by $\mathcal{L}(\mathcal{X})$.
  Preferences (or *ballots*) are taken to be elements of $\mathcal{L}(\mathcal{X})$.
- A *profile* $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathcal{L}(\mathcal{X})^n$ is a vector of preferences.
- We shall write $N^{\boldsymbol{R}}_{x \succ y}$ for the set of individuals that rank alternative x above alternative y under profile $\boldsymbol{R}$.

We are interested in preference aggregation mechanisms that map any profile of preferences to a single collective preference.

The proper technical term is *social welfare function* (SWF):

$$F : \mathcal{L}(\mathcal{X})^n \to \mathcal{L}(\mathcal{X})$$

Anonymity and Neutrality

Two examples for *axioms* (= formally specified desirable properties):

- A SWF F is *anonymous* if *individuals* are treated symmetrically:

$$F(R_1, \ldots, R_n) = F(R_{\pi(1)}, \ldots, R_{\pi(n)})$$

for any profile $\boldsymbol{R}$ and any permutation $\pi : \mathcal{N} \to \mathcal{N}$

- A SWF F is *neutral* if *alternatives* are treated symmetrically:

$$F(\pi(\boldsymbol{R})) = \pi(F(\boldsymbol{R}))$$

for any profile $\boldsymbol{R}$ and any permutation $\pi : \mathcal{X} \to \mathcal{X}$

(with π extended to preferences and profiles in the natural manner)

Keep in mind:

- not every SWF will satisfy every axiom we state here
- axioms are meant to be *desirable* properties (always arguable)

The Pareto Condition

A SWF F satisfies the *Pareto condition* if, whenever all individuals rank x above y, then so does society:

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

Independence of Irrelevant Alternatives (IIA)

A SWF F satisfies *IIA* if the relative social ranking of two alternatives only depends on their relative individual rankings:

$$N^{\boldsymbol{R}}_{x \succ y} = N^{\boldsymbol{R}'}_{x \succ y} \text{ implies } (x, y) \in F(\boldsymbol{R}) \Leftrightarrow (x, y) \in F(\boldsymbol{R}')$$

In other words: if x is socially preferred to y, then this should not change when an individual changes her ranking of z.

Arrow's Theorem

This is probably the most famous theorem in social choice theory.
It was first proved by Kenneth J. Arrow in his 1951 PhD thesis.
He later received the Nobel Prize in Economic Sciences in 1972.

A SWF F is a *dictatorship* if there exists a "dictator" $i \in \mathcal{N}$ such that $F(\boldsymbol{R}) = R_i$ for any profile $\boldsymbol{R}$, i.e., if the outcome is always identical to the preference supplied by the dictator.

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

Next: some remarks, then a proof

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

Remarks

- Note that this is a *surprising* result!
- Note that the theorem does *not* hold for *two* alternatives.
- Note that the *opposite direction* clearly holds: any dictatorship satisfies both Pareto and IIA (so they *characterise* dictatorships).
- Common misunderstanding: the SWF being *dictatorial* does not just mean that the outcome coincides with the preferences of some individual (rather: it's *the same* dictator for any profile).
- Arrow's Theorem is often read as an *impossibility theorem:*

  There exists no SWF for $\geqslant 3$ *alternatives that is Paretian, independent, and nondictatorial.*

- Significance of the result: (a) the result itself; (b) *general* theorem rather than just another observation of a flaw about a specific procedure; (c) *methodology* (precise statement of "axioms").

Proof

We'll sketch a proof adapted from Sen (1986), using the "decisive coalition" technique. Full details are in my review paper cited below.

Claim: *Any SWF for $\geqslant 3$ alternatives that satisfies the Pareto condition and IIA must be a dictatorship.*

So let F be a SWF for $\geqslant 3$ alternatives that satisfies Pareto and IIA.

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

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*).

A.K. Sen. *Social Choice Theory*. In K.J. Arrow and M.D. Intriligator (eds.), *Handbook of Mathematical Economics*, Volume 3, North-Holland, 1986.

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

About Decisiveness

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

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

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: $x' \succ x$ and $y \succ y'$ and $y \succ x$ (rest still undetermined)

From G being weakly decisive for (x, y): society ranks $x \succ y$
From Pareto: society ranks $x' \succ x$ and $y \succ y'$
Thus, from transitivity: society ranks $x' \succ y'$

Note that this works for any ranking of x' vs. y' by non-G individuals.
By IIA, it still works if individuals change their non-x'-vs.-y' rankings.

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

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, society ranks $y \succ z$. Two cases:

(1) Society ranks $x \succ z$: Exactly G_1 ranks $x \succ z \Rightarrow$ By IIA, in any profile where exactly G_1 ranks $x \succ z$, society will rank $x \succ z \Rightarrow G_1$ is weakly decisive on (x, z). Hence (previous slide): G_1 is decisive on all pairs.

(2) Society ranks $z \succ x$, i.e., $y \succ x$: 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. ✓
This concludes the proof of Arrow's Theorem.

Aside

If you are familiar with *ultrafilters*, then you may have noticed that the family of decisive coalitions forms an ultrafilter on $\mathcal{N}$.

We will exploit this correspondence more explicitly later in the course.

Summary

SCT is about *collective decision making*, instances of which are preference aggregation, voting, fair division, matching, and judgment aggregation.

In the first part, we have seen examples for some of the problems that can arise in some of these frameworks of aggregation:

- *Condorcet paradox:* majority preference may be cyclic
- Voters may benefit from *strategic lies* about their true preferences
- Intuitively appealing voting rules all give *different results*
- Intuitively appealing voting rules may fail to elect the *Condorcet winner*
- Intuitively appealing voting rules may violated the *Pareto principle*
- *Discursive dilemma:* majority judgments may be inconsistent

This calls for a formal treatment!

In the second part, we have provided one for *preference aggregation:*

- formal framework of *social welfare functions*, several axioms
- *Arrow's Theorem:* Pareto + IIA $\Rightarrow$ dictatorship (for $\geqslant 3$ alternatives)

This has been a first example for the *axiomatic method* in SCT.

Plan for the Rest of the Week

High-level overview of what I plan to cover:

- Day 1: introduction, preference aggregation, Arrow's Theorem
- Day 2: voting, strategic behaviour, Gibbard-Satterthwaite Thm
- Day 3: logical modelling and automated reasoning for social choice
- Day 4: basic judgment aggregation, List-Pettit Theorem
- Day 5: advanced judgment aggregation, wrap-up

Most of the material is also covered in my review paper cited below. Slides will get posted on the course website every day:

`http://www.illc.uva.nl/~ulle/teaching/esslli-2013/`

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