Computational Social Choice: Autumn 2011

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

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"?

$$\triangle \succ_1 \bigcirc \succ_1 \square$$

$$\square \succ_2 \triangle \succ_2 \bigcirc$$

$$\bigcirc \succ_3 \square \succ_3 \triangle$$

$$\overline{}$$

$$?$$

SCT is traditionally studied in Economics and Political Science, but now also by "us": *Computational Social Choice*.

Introduction

The course will cover issues at the interface of *computer science* and *mathematical economics*, including in particular:

- (computational) logic
- multiagent systems
- artificial intelligence
- social choice theory
- game theory
- decision theory

There has been a recent *trend* towards research of this sort. The broad philosophy is generally the same, but people use different names to identify various flavours of this kind of work, e.g.:

- Algorithmic Game Theory
- Social Software
- *and:* Computational Social Choice

Very few specific *prerequisites* are required to follow the course. Nevertheless, we will frequently touch upon *current research* issues.

Organisational Matters

- **Lecturer:** Ulle Endriss (`u.endriss@uva.nl`), Room C3.140
- **TA:** Umberto Grandi (`u.grandi@uva.nl`), Room C3.119
- **Timetable:** Tuesdays 11–13 in Room F2.21 (G3.05 in block 2)
- **Examination:** There will be several *homework* assignments on the material covered in the course. And every student will have to study a *recent paper* from the literature, write a short essay on the topic, and present their findings in a talk.
- **Website:** Lecture slides, homework assignments, and other important information will be posted on the course website:

  `http://www.illc.uva.nl/~ulle/teaching/comsoc/2011/`
- **Seminars:** There are occasional talks at the ILLC that are directly relevant to the course and that you are welcome to attend (e.g., at the Computational Social Choice Seminar).

Topics

The main topic for 2011 will be *logic and social choice*, which we will investigate from all sorts of angles. Some keywords:

- axiomatic method: impossibility theorems, characterisation results
- logics for modelling social choice scenarios
- social choice in combinatorial domains
- judgment aggregation

Part of this material is covered in a recent review article, cited below.

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

Prerequisites

There are no formal prerequisites. But: you should be comfortable with *formal* material and you will be asked to *prove* stuff.

To appreciate the bigger picture it is useful (but not necessary) to be aware of basic concepts in *game theory* (e.g., strategy, equilibrium).

For a couple of lectures (and homework questions) you will need basic knowledge in *complexity theory* (NP-completeness, polynomial reductions). It's possible to learn this along the way, if needed.

Related Courses

- Strategic Games [not offered next year]
  Krzysztof Apt

- Cooperative Games
  Stéphane Airiau

- Games and Complexity
  Peter van Emde Boas

- Autonomous Agents and Multiagent Systems (MSc AI)
  Shimon Whiteson

- Topics in Dynamic Epistemic Logic
  Alexandru Baltag

Plan for Today

Today's lecture has two parts:

- Part I. Informal introduction to some of the topics of the course
- Part II. A classical result: Arrow's Theorem

Part I: Examples, Problems, Ideas

Three Voting Rules

Voting is the prototypical form of collective decision making.

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

- *Plurality:* elect the candidate ranked first most often (i.e., each voter assigns one point to a candidate of her choice, and the candidate receiving the most votes wins)
- *Borda:* each voter gives $m-1$ points to the candidate she ranks first, $m-2$ to the candidate she ranks second, etc., and the candidate with the most points wins
- *Approval:* voters can approve of as many candidates as they wish, and the candidate with the most approvals wins

Example

Suppose there are three *candidates* (A, B, C) and 11 *voters* with the following *preferences* (where boldface indicates *acceptability*, for AV):

$$\text{5 voters think:} \quad \mathbf{A} \succ B \succ C$$

$$\text{4 voters think:} \quad \mathbf{C} \succ B \succ A$$

$$\text{2 voters think:} \quad \mathbf{B} \succ \mathbf{C} \succ A$$

Assuming the voters vote *sincerely*, who *wins* the election for

- the plurality rule?
- the Borda rule?
- approval voting?

Conclusion: We need to be very clear about which *properties* we are looking for in a voting rule . . .

The Axiomatic Method: May's Theorem

Three attractive properties ("axioms") of voting rules:

- *Anonymity:* voters should be treated symmetrically
- *Neutrality:* candidates should be treated symmetrically
- *Positive Responsiveness:* if a (sole or tied) winner receives increased support, then she should become the sole winner

One of the classical results in voting theory:

Theorem 1 (May, 1952) *A voting rule for two candidates satisfies anonymity, neutrality and pos. responsiveness iff it is the plurality rule.*

K.O. May. A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions. *Econometrica*, 20(4):680–684, 1952.

Example

Suppose the *plurality rule* is used to decide an election: the candidate receiving the highest number of votes wins.

Assume the preferences of the people in, say, Florida are as follows:

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

So even if nobody is cheating, Bush will win this election. But:

- It would have been in the interest of the Nader supporters to *manipulate*, i.e., to misrepresent their preferences.

Is there a better voting rule that avoids this problem?

The Gibbard-Satterthwaite Theorem

More properties of voting rules:

- A voting rule is *manipulable* if it may give a voter an incentive to misrepresent her preferences.
- A voting rule is *dictatorial* if the winner is always the top candidate of a particular voter (the dictator).

Another classical result (not stated 100% precisely here):

Theorem 2 (Gibbard-Satterthwaite) *For more than two candidates, every voting rule is either dictatorial or manipulable.*

What to do? One approach in COMSOC has been to look for voting rules that make manipulation (possible but) *computationally hard* ...

A. Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4):587–601, 1973.

M.A. Satterthwaite. Strategy-proofness and Arrow's Conditions. *Journal of Economic Theory*, 10:187–217, 1975.

Social Choice in Combinatorial Domains

Suppose 13 voters are asked to each vote *yes* or *no* on three issues; and we use the plurality rule for each issue independently to select a winning combination:

- 3 voters each vote for YNN, NYN, NNY.
- 1 voter each votes for YYY, YYN, YNY, NYY.
- No voter votes for NNN.

But then NNN wins: 7 out of 13 vote *no* on each issue.

This is an instance of the *paradox of multiple elections:* the winning combination received the fewest number of (actually: *no*) votes.

What to do instead? The number of combinatorial alternatives is *exponential* in the number of issues (e.g., $2^3 = 8$), so even just representing voter preferences is a challenge ...

S.J. Brams, D.M. Kilgour, and W.S. Zwicker. The Paradox of Multiple Elections. *Social Choice and Welfare*, 15(2):211–236, 1998.

Judgment Aggregation

Preferences are not the only structures we may wish to aggregate.

In JA we aggregate people's judgments regarding complex propositions:

	p	$p \rightarrow q$	q
Judge 1:	yes	yes	yes
Judge 2:	no	yes	no
Judge 3:	yes	no	no
Majority:	yes	yes	no

Problem: While each individual set of judgments is logically *consistent*, the *collective* judgement produced by the *majority rule* is not.

C. List. The Theory of Judgment Aggregation: An Introductory Review. *Synthese*. In press (2011).

Computational Social Choice

Research can be broadly classified along two dimensions —

The kind of *social choice problem* studied, e.g.:

- electing a winner given individual preferences over candidates
- aggregating individual judgements into a collective verdict
- fairly dividing a cake given individual tastes
- finding a stable matching of students to schools

The kind of *computational technique* employed, e.g.:

- algorithm design to implement complex mechanisms
- complexity theory to understand limitations
- logical modelling to fully formalise intuitions
- knowledge representation techniques to compactly model problems
- deployment in a multiagent system

Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. A Short Introduction to Computational Social Choice. Proc. SOFSEM-2007.

Part II: Arrow's Theorem

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.

What we will see next:

- formal framework: *social welfare functions*
- the *axiomatic method* in SCT, and some axioms
- the theorem, its interpretation, and a proof

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

Formal Framework

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* $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathcal{L}(\mathcal{X})^{\mathcal{N}}$ 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 $\boldsymbol{R}$.

For today 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):

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

The Axiomatic Method

Many important classical results in social choice theory are *axiomatic*. They formalise desirable properties as "*axioms*" and then establish:

- *Characterisation Theorems*, showing that a particular (class of) mechanism(s) is the only one satisfying a given set of axioms
- *Impossibility Theorems*, showing that there exists *no* voting mechanism satisfying a given set of axioms

Anonymity and Neutrality

Two very basic axioms (that we won't actually need for the theorem):

- 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:

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

This is a standard condition going back to the work of the Italian economist Vilfredo Pareto (1848–1923).

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.

IIA has been proposed by Arrow.

Universal Domain

This "axiom" is not really an axiom ...

Sometimes the fact that any SWF must be defined over *all* profiles is stated explicitly as a *universal domain* axiom.

Instead, I prefer to think of this as an integral part of the definition of the framework (for now) or as a *domain condition* (later on).

Arrow's Theorem

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 3 (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 the Pareto condition and IIA.
- Arrow's Theorem is often read as an *impossibility theorem:*

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

- The importance of Arrow's Theorem is due to the result itself ("there is no good way to aggregate preferences!"), but also to the *method:* for the first time (a) the desiderata had been rigorously specified and (b) an argument was given that showed that there can be *no* good procedure (rather than just pointing out flaws in concrete existing procedures).

Caveat

A common misinterpretation of Arrow's Theorem is that it just says that the outcome always happens to coincide with one of the individual preferences (which sounds ok).

No, it's much stronger: to satisfy Pareto and IIA, we must *first* fix the dictator i; *then* the outcome will always be R_i.

Proof

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

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 J. van Benthem and A. Gupta (eds.), *Logic and Philosophy Today*, College Publications. In press (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: homework).

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.

What next?

In the next lecture we will see an alternative proof of Arrow's Theorem.

And we will see two further classical impossibility theorems:

- Sen's Theorem on the Impossibility of a Paretian Liberal
- The Muller-Satterthwaite Theorem