Computational Social Choice: Autumn 2010

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

Plan for Today

We have seen already that we need to be precise about the properties we would like to see in a voting procedure and that it can be hard to satisfy all the desiderata we might have. Using the *axiomatic method*, today we will see two *impossibility theorems*:

- *Arrow's Theorem* [1951]
- the *Muller-Satterthwaite Theorem* [1977]

This is (very) classical social choice theory, but we will also briefly touch upon some modern COMSOC concerns:

- Can we go beyond the mathematical *rigour* of SCT and achieve a *formalisation* in the sense of symbolic logic?
- Can we *automate* the proving of theorems in SCT?
- What changes if we alter the notion of *ballot*, which classically is assumed to be a (usually strict) ranking of the alternatives?

Formal Framework

Basic terminology and notation:

- finite set of *voters* $\mathcal{N} = \{1, \ldots, n\}$, the *electorate*
- (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* are assumed to be elements of $\mathcal{L}(\mathcal{X})$. *Ballots* are elements of $\mathcal{L}(\mathcal{X})$.

A *voting procedure* is a function $F : \mathcal{L}(\mathcal{X})^{\mathcal{N}} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$, mapping profiles of ballots to nonempty sets of alternatives.

Remark 1: Approval Voting, Majority Judgment, Cumulative and Range Voting don't fit this framework; everything else we've seen does.

Remark 2: If we wanted to be a bit more general, we could introduce a *ballot language* $\mathcal{B}(\mathcal{X})$ and work with functions $F : \mathcal{B}(\mathcal{X})^{\mathcal{N}} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$.

Remark 3: A voting procedure *parametrised by $\mathcal{N}$ and $\mathcal{X}$* (e.g., Borda) is a family of functions $F^{\mathcal{N},\mathcal{X}} : \mathcal{L}(\mathcal{X})^{\mathcal{N}} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$.

Resoluteness and Tie-Breaking

$F : \mathcal{L}(\mathcal{X})^{\mathcal{N}} \to 2^{\mathcal{X}} \setminus \{\emptyset\}$ is called *resolute* if $|F(\underline{b})| = 1$ for any ballot profile $\underline{b} \in \mathcal{L}(\mathcal{X})^{\mathcal{N}}$, i.e., if F always produces a unique winner.

Terminology:	*voting rule*	*vs.*	*voting correspondence*
	(resolute)		(irresolute)

We can turn an irresolute procedure F into a resolute procedure $F \circ T$ by pairing F with a (deterministic) *tie-breaking rule* $T : 2^{\mathcal{X}} \setminus \{\emptyset\} \to \mathcal{X}$ with $T(X) \in X$ for any $X \in 2^{\mathcal{X}} \setminus \{\emptyset\}$. Examples:

- select the lexicographically first alternative
- select the preferred alternative of some chair person

We will (mostly) just analyse either irresolute or resolute procedures, without worrying about tie-breaking in particular.

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

Today, we will see two examples for the latter.

We first discuss some of these axioms, starting with very basic ones.

Universal Domain

The first axiom is not really an axiom ...

Sometimes the fact that voting procedures F are defined over *all* ballot 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).

Anonymity and Neutrality

A voting rule is *anonymous* if the *voters* are treated symmetrically: if two voters switch ballots, then the winners don't change. Formally:

> F is anonymous if $F(b_1, \ldots, b_n) = F(b_{\pi(1)}, \ldots, b_{\pi(n)})$ for any ballot profile $(b_1, \ldots, b_n)$ and any permutation $\pi : \mathcal{N} \to \mathcal{N}$.

A voting procedure is *neutral* if the *alternatives* are treated symmetrically. Formally:

> F is neutral if $F(\pi(\underline{b})) = \pi(F(\underline{b}))$ for any ballot profile $\underline{b}$ and any permutation $\pi : \mathcal{X} \to \mathcal{X}$ (with π extended to ballot profiles and sets of alternatives in the natural manner).

Nonimposition

A voting procedure satisfies *nonimposition* if each alternative is the unique winner under at least one ballot profile. Formally:

> F satisfies nonimposition if for any alternative $x \in \mathcal{X}$ there exists a ballot profile $\underline{b} \in \mathcal{L}(\mathcal{X})^{\mathcal{N}}$ such that $F(\underline{b}) = \{x\}$.

Remark 1: Any *surjective* (*onto*) voting procedure satisfies nonimposition. For resolute procedures, the two properties coincide.

Remark 2: Any *neutral* resolute voting procedure satisfies nonimposition.

Dictatorships

A voting procedure is *dictatorial* if there exists a voter (the dictator) such that the unique winner will always be her top-ranked alternative.

A voting procedure is *nondictatorial* if it is not dictatorial. Formally:

> F is nondictatorial if there exists *no* voter $i \in \mathcal{N}$ such that $F(\underline{b}) = \{x\}$ whenever $i \in \underline{b}(x \succ y)$ for all $y \in \mathcal{X} \setminus \{x\}$.

Remark: Any *anonymous* voting procedure is nondictatorial.

Notation: $\underline{b}(x \succ y)$ is the set of voters ranking x above y in profile $\underline{b}$.

Unanimity and the Pareto Condition

A voting procedure is *unanimous* if it elects (only) x whenever all voters say that x is the best alternative. Formally:

> F is unanimous if $\underline{b}(x \succ y) = \mathcal{N}$ for all $y \in \mathcal{N} \setminus \{x\}$ implies $F(\underline{b}) = \{x\}$.

The *weak Pareto condition* holds if an alternative y that is dominated by some other alternative x in all ballots cannot win. Formally:

> F is weakly Pareto if $\underline{b}(x \succ y) = \mathcal{N}$ implies $y \notin F(\underline{b})$.

Remark: The weak Pareto condition entails unanimity, but the converse is not true.

Independence of Irrelevant Alternatives (IIA)

A voting procedure is *independent of irrelevant alternatives* (IIA) if, whenever y loses to some winner x and the relative ranking of x and y does not change in the ballots, then y cannot win (independently of any possible changes wrt. other, irrelevant, alternatives). Formally:

> F satisfies IIA if $x \in F(\underline{b})$ and $y \notin F(\underline{b})$ together with $\underline{b}(x \succ y) = \underline{b}'(x \succ y)$ imply $y \notin F(\underline{b}')$ for any profiles $\underline{b}$ and $\underline{b}'$.

Remark: IIA was introduced by Arrow (1951/1963), originally for *social welfare functions* (SWFs), where the outcome is a preference ordering. Above variant of IIA (for voting) is due to Taylor (2005).

K.J. Arrow. *Social Choice and Individual Values.* 2nd edition. Cowles Foundation, Yale University Press, 1963.

A.D. Taylor. *Social Choice and the Mathematics of Manipulation.* Cambridge University Press, 2005.

Arrow's Theorem for Voting Procedures

This is widely regarded as *the* seminal result in social choice theory. Kenneth J. Arrow received the Nobel Prize in Economics in 1972.

Theorem 1 (Arrow, 1951) *No voting procedure for $\geqslant 3$ alternatives can be weakly Pareto, IIA, and nondictatorial.*

This particular version of the theorem is due to Taylor (2005).

Maybe the most accessible proof (of the standard formulation of the theorem) is the first proof in the paper by Geanakoplos (2005).

K.J. Arrow. *Social Choice and Individual Values.* 2nd edition. Cowles Foundation, Yale University Press, 1963.

A.D. Taylor. *Social Choice and the Mathematics of Manipulation.* Cambridge University Press, 2005.

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

Remarks

- Note that this is a *surprising* result!
- Note that the theorem does *not* hold for *two* alternatives.
- We can interpret the theorem as a *characterisation* result:

  A voting procedure for $\geqslant 3$ alternatives satisfies the weak Pareto condition and IIA if and only if it is a dictatorship.

- *IIA* is the most debatable of the three axioms featuring in the theorem. Indeed, it is quite hard to satisfy.
- 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).

Proof of Arrow's Theorem

We'll sketch a proof adapted from Sen (1986), who proves the standard formulation of Arrow's Theorem using the "decisive coalition" technique.

Definitions:

$G \subseteq \mathcal{N}$ is *decisive* for $(x,y) \in \mathcal{X}^2$ if $\underline{b}(x \succ y) \supseteq G$ implies $y \notin F(\underline{b})$

$G \subseteq \mathcal{N}$ is *almost dec.* for $(x,y) \in \mathcal{X}^2$ if $\underline{b}(x \succ y) = G$ implies $y \notin F(\underline{b})$

Proof Plan:

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

The proof of the lemma relies on another lemma:

- Lemma: G *almost decisive* for *some* (x,y) $\Rightarrow$ G *decisive* for *all* (a,b)

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

Proof of the Decisiveness Lemma

Suppose F is a voting procedure that is weakly Pareto and IIA.

Lemma 1 (Decisiveness) *If $G \subseteq \mathcal{N}$ is an almost decisive coalition for some pair $(x,y) \in \mathcal{X}^2$, then G is decisive for all pairs $(a,b) \in \mathcal{X}^2$.*

Proof: Suppose x, y, a, b are all distinct (other cases work similarly).

Consider this (class of) ballot profile(s):

G: $a \succ x \succ y \succ b$

Rest: $a \succ x$ and $y \succ b$ and $y \succ x$ (rest unspecified)

Note that "the rest" could have *any* ranking for a and b.

From G being almost decisive on (x,y) $\Rightarrow$ y loses (to x)
From the weak Pareto condition $\Rightarrow$ x loses (to a) and b loses (to y)

Thus, b loses and (only) a wins in a situation where G has $a \succ b$, independently of how the rest rank a and b. So, by IIA, b will lose to a for *any* profile in which G has $a \succ b$, i.e., G is decisive for (a,b). ✓

Proof of the Contraction Lemma

Lemma 2 (Contraction) *If a coalition $G \subseteq \mathcal{N}$ with $|G| > 1$ is decisive for all pairs, then so is some smaller coalition $G' \subset G$.*

Proof: Let $G = G_1 \uplus G_2$, both nonempty. Consider this ballot profile:

G_1:	$x \succ y \succ z$	$\Rightarrow$ as G decisive, z loses (against y)
G_2:	$y \succ z \succ x$	thus, two possibilities:
Rest:	$z \succ x \succ y$	(1) x wins *or* (2) only y wins

Case (1): x wins, z loses, and only G_1 has $x \succ z$

- by IIA, z will lose for *any* profile in which only G_1 has $x \succ z$
- in other words, G_1 is almost decisive for (x,z)
- by Lemma 1, G_1 is thus decisive for all pairs

Case (2): y wins, x loses, and only G_2 has $y \succ x$

- by the same argument, G_2 is decisive for all pairs

Hence, either G_1 or G_2 will be decisive under all circumstances. ✓

Logic and Automated Reasoning

Logic has long been used to *formally specify* computer systems, facilitating formal or even *automatic verification* of various properties.

Can we apply this methodology also to *social choice* mechanisms?

- What logic fits best?
- Which automated reasoning methods are useful?

Computer-aided Proof of Arrow's Theorem

Tang and Lin (2009) prove two inductive lemmas:

- If there exists an Arrovian aggregator for n voters and $m{+}1$ alternatives, then there exists one for n and m (if $n \geqslant 2$, $m \geqslant 3$).
- If there exists an Arrovian aggregator for $n{+}1$ voters and m alternatives, then there exists one for n and m (if $n \geqslant 2$, $m \geqslant 3$).

Tang and Lin then show that the *"base case"* of Arrow's Theorem with 2 voters and 3 alternatives can be fully modelled in *propositional logic*.

A SAT solver can *verify* $\textsc{Arrow}(2,3)$ to be correct in < 1 second — that's $(3!)^{3! \times 3!} \approx 10^{28}$ aggregators [SWFs] to check.

Discussion: Opens up opportunities for quick sanity checks of hypotheses regarding new impossibility theorems.

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

Related Work

- Ågotnes et al. (2010) propose a modal logic to model preferences and their aggregation that can express Arrow's Theorem.
- Arrow's Theorem holds *iff* the set $T_{\textsc{Arrow}}$ of FOL formulas (defined in the paper) has no finite models (Grandi and E., 2009).
- Nipkow (2009) formalises and verifies a known *proof* of Arrow's Theorem in the HOL proof assistant Isabelle.

T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the Logic of Preference and Judgment Aggregation. *J. Auton. Agents and Multiagent Sys.* In press (2010).

U. Grandi and U. Endriss. *First-order Logic Formalisation of Arrow's Theorem*. Proc. 2nd Internat. Workshop on Logic, Rationality and Interaction (LORI-2009).

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

Monotonicity

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

We restrict attention to *resolute* voting procedures (unique winner).

- *Weak monotonicity*: F is weakly monotonic if $F(\underline{b}) = \{x\}$ implies $F(\underline{b}') = \{x\}$ for any alternative x and any two ballot profiles $\underline{b}$ and $\underline{b}'$ with $\underline{b}(x \succ y) \subseteq \underline{b}'(x \succ y)$ and $\underline{b}(y \succ z) = \underline{b}'(y \succ z)$ for all alternatives y and z different from x.
- *Strong monotonicity*: F is strongly monotonic if $F(\underline{b}) = \{x\}$ implies $F(\underline{b}') = \{x\}$ for any alternative x and any two ballot profiles $\underline{b}$ and $\underline{b}'$ with $\underline{b}(x \succ y) \subseteq \underline{b}'(x \succ y)$ for all alternatives y different from x.

Strong monotonicity is also known as *Maskin monotonicity* or *strong positive association*.

Example

Even *weak monotonicity* is not satisfied by some voting procedures. Consider *Plurality with Run-off* (with some tie-breaking rule).

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 run-off. 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 other procedures (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 2 (Muller and Satterthwaite, 1977) *No resolute voting procedure for $\geqslant 3$ alternatives can be surjective, strongly monotonic, and nondictatorial.*

Proof omitted. The "decisive coalition" technique used to prove Arrow's Theorem is applicable here as well (see e.g. Myerson, 1996).

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 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.

R.B. Myerson. Fundamentals of Social Choice Theory. CMSEMS Discussion Paper 1162. Northwestern University, 1996.

Partially Ordered Preferences

In the classical framework of voting theory, preferences and ballots are assumed to be *linear orders* (or possibly weak orders).

In AI in particular, this may not always be appropriate:

- An agent may be unable to rank certain alternatives, e.g., for lack of information or lack of computational resources (bounded rationality).
- Transmitting a full ranking may be unnecessarily verbose (elicitation costs, communication costs, on-demand computation).

Therefore, some work in COMSOC has investigated what happens when we move to different types of structures to model preferences and ballots.

For example, Pini et al. (2009) have investigated what happens to impossibility theorems such as those discussed today when we work with *partial orders* (in short: in terms of results, not much changes).

M.S. Pini, F. Rossi, K.B. Venable, and T. Walsh. Aggregating Partially Ordered Preferences. *Journal of Logic and Computation*, 19(3):475–502, 2009.

Summary

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

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

We have seen two classical *impossibility theorems* that apply as soon as we have *three or more* alternatives:

- *Arrow's Theorem:* only dictatorial rules are weakly Pareto and IIA
- *Muller-Satterthwaite Theorem:* only dictatorial rules are resolute, surjective, and strongly monotonic

Regarding modern COMSOC concerns we have discussed:

- using *logic* to fully formalise social choice problems and theorems, and *automated reasoning* to support analysis
- exploring *ballot languages* other than linear orders

Remark: Note that so far we have not made any (formal) use of the notion of "preference"; we have only talked about "ballots". This will not change until we start discussing strategic issues.

What next?

As discussed, the impossibility theorems we have seen today could be interpreted as different axiomatic characterisations of the class of dictatorial voting procedures.

Next, we will see characterisations of more attractive (classes of) voting procedures:

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