Explainability in Social Choice: Day 1

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

Tutorial at *Formal Models of Democracy*, Rotterdam, April 2022
http://www.illc.uva.nl/~ulle/teaching/fmd-2022/

What is this about?

A group needs to take a decision (that is: choose an alternative).
Every group member has their own preferences over those alternatives.

Given some decision, how can you explain why it is the right one?

The classic approach in *social choice theory*:

- argue why a given social choice function F is the right one
- demonstrate that that F would choose the alternative in question

Can we do better? Can we offer a direct explanation instead?

Outline

- Day 1: The Axiomatic Method in Social Choice Theory
  - Model: Social Choice Functions
  - Examples for Rules and Axioms
  - Examples for Characterisation Results
- Day 2: Justifying and Explaining Collective Decisions
  - Criticism: Characterisation Results as Explanations?
  - Approach: Normatively Grounded Explanations
  - Automation: Computing Explanations

Social Choice Theory

Social choice theory is about methods for *collective decision making*, such as *political* decision making by groups of *economic* agents.

Its methodology ranges from the *philosophical* to the *mathematical*. It is traditionally studied in *Economics* and *Political Science*, but nowadays also in *Computer Science* and *Artificial Intelligence*.

It deals with *decision-making scenarios* such as these:

- How to choose a single alternative given people's preferences?
- How to choose a set of such alternatives (say, a parliament)?
- How to aggregate the views of different judges in a court case?
- How to extract information from noisy crowdsourced data?
- How to fairly allocate resources to the members of a society?
- How to fairly divide computing time between several users?
- How to match school children to high schools?

We shall focus on the first scenario (but the ideas are more general).

Three Voting Rules

Suppose n *voters* choose from a set of m *alternatives* by stating their preferences in the form of *linear orders* over the alternatives.

Here are three *voting rules:*

- *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: Choosing a Beverage for Lunch

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

2 Germans:	Beer $\succ$ Wine $\succ$ Milk
3 French people:	Wine $\succ$ Beer $\succ$ Milk
4 Dutch people:	Milk $\succ$ Beer $\succ$ Wine

Exercise: *Which beverage wins the election for*

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

Further Voting Rules

The Borda rule is just one of many so-called *positional scoring rules*. Besides *plurality*, examples include *veto* and k*-approval*.

Even more rules:

- *Single Transferable Vote*
- *Copeland*
- *Slater*
- *Kemeny*
- *Cup Rules*
- ...

S.J. Brams and P.C. Fishburn. Voting Procedures. In K.J. Arrow et al. (eds.), *Handbook of Social Choice and Welfare*. Elsevier, 2002.

Picking a Voting Rule

So: Lots of rules. *How do you pick one?* Criteria we might use:

- *normative* requirements
- *epistemic* requirements
- *computational* requirements
- *informational* requirements

We shall focus on the first family of requirements only.

The Model

Fix a finite set $A = \{a, b, c, \ldots\}$ of *alternatives*, with $|A| = m \geqslant 2$.

Let $\mathcal{L}(A)$ denote the set of all strict linear orders R on A. We use elements of $\mathcal{L}(A)$ to model (true) *preferences* and (declared) *ballots*.

Each member i of a finite set $N = \{1, \ldots, n\}$ of *voters* supplies us with a ballot R_i, giving rise to a *profile* $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathcal{L}(A)^n$.

A *voting rule* (or *social choice function*) for N and A selects (ideally) one or (in case of a tie) more winners for every such profile:

$$F : \mathcal{L}(A)^n \to 2^A \setminus \{\emptyset\}$$

If $|F(\boldsymbol{R})| = 1$ for all profiles $\boldsymbol{R}$, then F is called *resolute*.

Most natural voting rules are *irresolute* and have to be paired with a *tie-breaking rule* to always select a unique election winner.

Examples: random tie-breaking, lexicographic tie-breaking

Axioms = Normative Requirements

We formulate normative requirements in the form of so-called *axioms*.

Some particularly convincing examples:

- *Participation Principle:* It should be in the best interest of voters to participate; voting truthfully should be no worse than abstaining.
- *Condorcet Principle:* If there is an alternative that is preferred to every other alternative by a majority of voters, then it should win.
- *Pareto Principle:* There should be no alternative that every voter strictly prefers to the alternative selected by the voting rule.

But: surprisingly hard to satisfy! ($\hookrightarrow$)

Plurality with Runoff fails the Participation Principle

No-Show Paradox: Under plurality with runoff, it may be better to abstain than to participate and vote for your favourite alternative!

25 voters: $a \succ b \succ c$
46 voters: $c \succ a \succ b$
24 voters: $b \succ c \succ a$

So b gets eliminated, and then c beats a 70:25 in the runoff.

Now suppose two voters from the first group *abstain:*

23 voters: $a \succ b \succ c$
46 voters: $c \succ a \succ b$
24 voters: $b \succ c \succ a$

Now a gets eliminated, and b beats c 47:46 in the runoff.

P.C. Fishburn and S.J Brams. Paradoxes of Preferential Voting. *Mathematics Magazine*, 1983.

Borda fails the Condorcet Principle

Consider this profile with 11 voters:

4 voters: $c \succ b \succ a$
3 voters: $b \succ a \succ c$
2 voters: $b \succ c \succ a$
2 voters: $a \succ c \succ b$

Borda elects b, but c is majority-preferred to both a and b.

In fact: *Every positional scoring rule* fails the Condorcet Principle.

Cup Rules fail the Pareto Principle

Rule given by *binary tree*, with the alternatives labelling the leaves.
To progress an alternative needs to *majority*-beat its sibling.

Such *cup rules* may fail the Pareto Principle:

```
          o
         / \
        o   d
       / \
      o   a
     / \
    b   c
```

Consider this profile with three voters:

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)

Exercise: *Do you see how I did this?*

More Axioms: Anonymity and Neutrality

Two basic fairness requirements for a voting rule F:

- F is *anonymous* if $F(R_1, \ldots, R_n) = F(R_{\pi(1)}, \ldots, R_{\pi(n)})$ for any profile $(R_1, \ldots, R_n)$ and any permutation $\pi : N \to N$.
- F is *neutral* if $F(\pi(\boldsymbol{R})) = \pi(F(\boldsymbol{R}))$ for any profile $\boldsymbol{R}$ and any permutation $\pi : A \to A$ (with π extended to profiles and sets of alternatives in the natural manner).

In other words:

- anonymity $=$ symmetry w.r.t. voters
- neutrality $=$ symmetry w.r.t. alternatives

Consequences of Axioms

For this slide only, let us restrict attention to voting rules for scenarios with just *two voters* ($n = 2$) and *two alternatives* ($m = 2$).

Exercise: *Show that there exists no resolute voting rule that is 'fair' in the sense of being both anonymous and neutral.*

Exercise: *But there still are a couple of irresolute voting rules that are both anonymous and neutral. Give some examples!*

Yet Another Axiom: Positive Responsiveness

Notation: Write $N^{\boldsymbol{R}}_{x \succ y} = \{i \in N \mid (x,y) \in R_i\}$ for the set of voters who rank alternative x above alternative y in profile $\boldsymbol{R}$.

A (not necessarily resolute) voting rule satisfies *positive responsiveness* if, whenever some voter raises a (possibly tied) winner $x^\star$ in her ballot, then $x^\star$ will become the *unique* winner. Formally:

> F is *positively responsive* if $x^\star \in F(\boldsymbol{R})$ implies $\{x^\star\} = F(\boldsymbol{R'})$ for any alternative $x^\star$ and any two *distinct* profiles $\boldsymbol{R}$ and $\boldsymbol{R'}$ s.t. $N^{\boldsymbol{R}}_{x^\star \succ y} \subseteq N^{\boldsymbol{R'}}_{x^\star \succ y}$ and $N^{\boldsymbol{R}}_{y \succ z} = N^{\boldsymbol{R'}}_{y \succ z}$ for all $y, z \in A \setminus \{x^\star\}$.

Thus, this is basically a *monotonicity* requirement.

May's Theorem

When there are only *two alternatives*, then all the voting rules we have seen coincide with the *simple majority rule*. Good news:

Theorem 1 (May, 1952) *A voting rule for two alternatives satisfies anonymity, neutrality, and positive responsiveness if and only if that rule is the simple majority rule.*

This provides a good justification for using this rule (arguing in favour of 'majority' directly is harder than arguing for anonymity etc.).

K.O. May. A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions. *Econometrica*, 1952.

Proof Sketch

Clearly, the simple majority rule satisfies all three properties. ✓

Now for the other direction:

Assume the number of voters is *odd* $\rightsquigarrow$ no ties. (other case: similar)

There are two possible ballots: $a \succ b$ and $b \succ a$.
Anonymity $\rightsquigarrow$ only *number of ballots* of each type matters.

Consider all possible profiles $\boldsymbol{R}$. Distinguish two cases:

- Whenever $|N^{\boldsymbol{R}}_{a\succ b}| = |N^{\boldsymbol{R}}_{b\succ a}| + 1$, then only a wins.
  By *PR*, a wins whenever $|N^{\boldsymbol{R}}_{a\succ b}| > |N^{\boldsymbol{R}}_{b\succ a}|$. By *neutrality*, b wins otherwise. But this is just what the simple majority rule does. ✓
- There exist a profile $\boldsymbol{R}$ with $|N^{\boldsymbol{R}}_{a\succ b}| = |N^{\boldsymbol{R}}_{b\succ a}| + 1$, yet b wins.
  Suppose one a-voter switches to b, yielding $\boldsymbol{R'}$. By *PR*, now only b wins. But now $|N^{\boldsymbol{R'}}_{b\succ a}| = |N^{\boldsymbol{R'}}_{a\succ b}| + 1$, which is symmetric to the earlier situation, so by *neutrality* a should win. Contradiction. ✓

Young's Theorem

Young's Theorem is a characterisation of the *positional scoring rules:*

A PSR is defined by a *scoring vector* $\boldsymbol{s} = (s_1, \ldots, s_m) \in \mathbb{R}^m$ with $s_1 \geqslant s_2 \geqslant \cdots \geqslant s_m$ and $s_1 > s_m$. An alternative gets s_i points for every voter putting it at the ith position.

Young's Theorem involves an axiom we have not yet seen:

F satisfies *reinforcement* if, whenever we split the electorate into two groups and some alternative wins for both groups, then that alternative also wins for the full electorate:

$$F(\boldsymbol{R}) \cap F(\boldsymbol{R'}) \neq \emptyset \;\;\Rightarrow\;\; F(\boldsymbol{R} \oplus \boldsymbol{R'}) = F(\boldsymbol{R}) \cap F(\boldsymbol{R'})$$

Young showed that a rule F is a *positional scoring rule* (with a scoring vector that need not be decreasing) <u>iff</u> it satisfies *anonymity*, *neutrality*, *reinforcement*, and a technical condition known as *continuity*.

H.P. Young. Social Choice Scoring Functions. *SIAM Journal Appl. Math.*, 1975.

Impossibility Results

Maybe the most famous results in SCT are impossibility results, such as:

Theorem 2 (Arrow, 1951) *Any resolute SCF for ⩾3 alternatives that is Paretian and independent is dictatorial.*

Theorem 3 (Gibbard-Satterthwaite, 1973/75) *Any resolute SCF for $\geqslant 3$ alternatives that is surjective and strategyproof is dictatorial.*

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

A. Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 1973.

M.A. Satterthwaite. Strategy-proofness and Arrow's Conditions. *Journal of Economic Theory*, 1975.

Discussion

Can the axiomatic method provide explainability for decision making?

Some thoughts:

- Characterisation results: provide *attractive justifications for rules*, but *lack explanatory power* for most people (too complicated!).
- Impossibility theorems: Attractive combinations of axioms too demanding. So: *no perfect rule* (yet: *good outcomes* possible).
- Are we trying to solve a harder problem than we need to? If you can justify the use of a rule, you can justify the right outcome for *every profile* (but we need to do so for just *one profile* at a time).

We shall take this up tomorrow ...

Further Reading

For a general introduction to voting theory, consult Zwicker (2016).

For more on the axiomatic method in voting, with a focus on (proving) impossibility results, consult my expository article cited below.

Finally, *Whale* (`whale.imag.fr`) is a great tool to collect ballots, compute election outcomes for several rules, and visualise your data.

W.S. Zwicker. Introduction to the Theory of Voting. In F. Brandt et al. (eds.), *Handbook of Computational Social Choice*. CUP, 2016.

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

S. Bouveret. Social Choice on the Web. In U. Endriss (ed.), *Trends in Computational Social Choice*. AI Access, 2017.