Exploring the Intersection between Voting Theory and AI

Zoi Terzopoulou

GATE, St-Etienne School of Economics
University of Lyon/St-Etienne

Plan for Today

Social choice theory studies collective decision making.

The main problem is to design a voting rule with “good“ properties.

- Voting Theory (originating in Economics and Philosophy)
- Machine Voting (explored in Computer Science)

All voting preliminaries can be found in the following review chapter:

Zwicker. *Introduction to the Theory of Voting*. Handbook of Computational Social Choice, 2016.

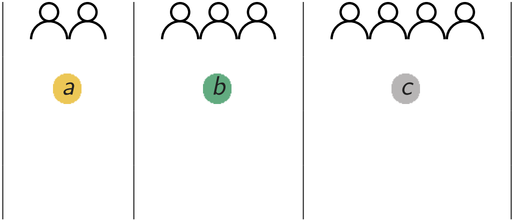
Voting

Example: Different Rules, Different Outcomes

👤👤	👤👤👤	👤👤👤👤
a	b	c
b	a	a
c	c	b

What should be the voting outcome?

Example: Plurality



Winner with the most first positions: *c*

*Used in the House of Commons in the United Kingdom.

Example: Borda

	👤👤	👤👤👤	👤👤👤👤
2	*a*	*b*	*c*
1	*b*	*a*	*a*
0	*c*	*c*	*b*

Winner with the most accumulated linear scores: *a*

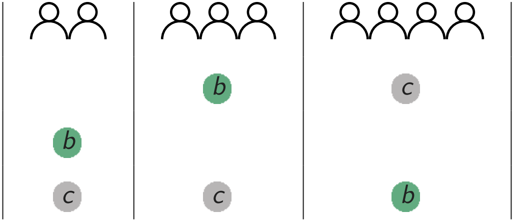
*Used to elect the Italian-speaking ethnic minority members in the National Assembly of Slovenia.

Example: Plurality with Runoff—Round 1

👤👤	👤👤👤	👤👤👤👤
a	*b*	*c*
b	*a*	*a*
c	*c*	*b*

Two alternatives with the most first positions are promoted: *b*, *c*

Example: Plurality with Runoff—Round 2



The majority alternative wins: *b*

*Used in France for presidential elections.

Formal Framework

- A finite set of voters $N = \{1, ..., n\}$.
- Voters need to choose from a finite set of m alternatives A.
- Voters have preferences and cast ballots $\succ$, which are strict linear orders over the set of alternatives $\mathcal{L}(A)$.
- All ballots of the voters together provide us with an election:

$$\boldsymbol{P} = (\succ_1, \ldots, \succ_n) \in \mathcal{L}(A)^n$$

- A voting rule selects the winner(s) for each such election:

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

Other Types of Ballots

Approval Voting

- You can approve of any subset of the alternatives.
  The alternative with the most approvals wins.

Majority judgment

- You award a grade to each alternative (“excellent”, “good”, etc.).
  The alternative with the highest median grade wins.

Normative Principles and Voting Rules

Recall Plurality, Borda, and Plurality with runoff.

Which of them satisfy the following axioms?

- Anonymity: The names of the voters don't matter.
- Neutrality: The names of the alternatives don't matter.
- Monotonicity: If a winning alternative receives additional support (it is ranked higher by some voter), then it should still win the election.
- Reinforcement: If alternative x wins in two disjoint electorates, then x should also win when we join those two electorates into one.

Condorcet Principle

The Condorcet winner is an alternative x such that for every other alternative y, a majority of voters ranks x higher than y.

Condorcet principle: If there exists a Condorcet winner, then it should win the election. A rule satisfying this principle is a Condorcet extension.

The Borda rule infamously fails the Condorcet principle.

Condorcet Extensions

Under the Copeland rule, an alternative gets +1 point for every pairwise majority contest won and -1 point for every such contest lost. The alternatives with the most points win.

Other proposals exist, often based on the majority graph of an election: A directed graph with nodes the alternatives in A, and with an edge from x to y whenever x beats y in a pairwise majority contest.

Brandt, Brill & Harrenstein. *Tournament Solutions*. Handbook of Computational Social Choice, 2016.

Axiom: The Pareto Principle

A voting rule satisfies Pareto if, whenever all voters rank one alternative higher than another, then the latter does not win.

👤👤👤	👤👤	👤	👤
a	*b*	*b*	*a*
b	*c*	*a*	*b*
c	*a*	*c*	*c*

Does the Borda rule satisfy the Pareto principle?

Axiom: Independence of Irrelevant Alternatives

If a wins and b loses, then a is socially preferred to b.

Whether an alternative is socially preferred to another should depend only on their relative rankings in the election, and not on how other, irrelevant, alternatives are ranked.

a	a	c
c	c	b
b	b	a

vs.

c	a	b
a	b	c
b	c	a

Does the Borda rule satisfy independence?

Arrow's Impossibility Theorem

A voting rule is a dictatorship if there is a voter i (the dictator) such that **for all** elections $\boldsymbol{P}$, the winner is the first alternative of $\succ_i$.

Theorem. *Any (resolute) voting rule for* $m \geqslant 3$ *that satisfies Pareto and independence must be a dictatorship.*

- Impossibility: independence + Pareto + nondictatoriality
- Characterisation: dictatorship = independence + Pareto

Arrow. *Social Choice and Individual Values*, 1951.

Strategyproofness: An Example

Consider the plurality rule.

👤	👤👤👤	👤👤👤
c	a	b
b	b	a
a	c	c

What would you do?
Is there a better voting rule to avoid this problem?

Definitions

We need to distinguish:

- The preference a voter has over the alternatives.
- The ballot the voter reports.

They coincide when the voter is truthful.

F is strategyproof (or immune to manipulation) if for no voter $i \in N$ there exist an election $\boldsymbol{P} = (\succ_1, \ldots, \succ_i, \ldots, \succ_n)$ and a ballot $\succ'$ such that:

$$F(\succ_1, \ldots, \succ_i', \ldots, \succ_n) \succ_i F(\succ_1, \ldots, \succ_i, \ldots, \succ_n)$$

Voter i prefers the outcome obtained by reporting the untruthful ballot $\succ_i'$ to the outcome obtained by reporting the truthful preference $\succ_i$.

Why Strategyproofness?

- Voters should not have to waste resources pondering over what other voters will do and trying to figure out how to respond. Most often, they will not be able to figure this out perfectly.

- If voters strategise, then the final outcome of the election will be based on a skewed election, and the winner may not be a good compromise overall.

The Gibbard-Satterthwaite Theorem

A voting rule is surjective if for every alternative can win some election.

Gibbard and Satterthwaite independently proved that:

Theorem. *For $m \geqslant 3$, any voting rule that is surjective and strategyproof is a dictatorship.*

However, the existence of one election where a voter has an incentive to lie does not speak about the frequency of such instances or the difficulty of recognizing the possibility.

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

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

The Condition of Complete Information

The classical work on strategic manipulation in voting assumes that the manipulation has full information about the ballots of other voters.
This makes sense sometimes:

- In small committees (e.g., students electing the classroom representative), this is more realistic.
- Many elections are preceeded by polls, which allow voters to have fairly accurate information about the votes of others.
- We may need to use a voting rule that is safe against strategic manipulation in the worst case.

Incomplete Information

We say that an election $\boldsymbol{P'}$ is compatible with the information obtained by an election $\boldsymbol{P}$ when a voter cannot distinguish between the two.

F is strategyproof under incomplete information if for no voter $i \in N$ there exist an election $\boldsymbol{P}$ and a ballot $\succ'_i$ such that:

- $F(\boldsymbol{P'}_{-i}, \succ'_i) \succ_i F(\boldsymbol{P'}_{-i}, \succ_i)$ for some $\boldsymbol{P'}$ compatible with i's info
- $F(\boldsymbol{P''}_{-i}, \succ_i) \succ_i F(\boldsymbol{P''}_{-i}, \succ'_i)$ for no $\boldsymbol{P''}$ compatible with i's info

Strategyproofness under Incomplete Information

Recall the plurality rule. What would you do now?

👤	👤👤👤	👤👤👤
c	?	?
b	?	?
a	?	?

Theorem: *When $n > 3$, any Condorcet extension is immune to zero-manipulation.*

Reijngoud & Endriss. *Voter Response to Iterated Poll Information*. AAMAS, 2012.

Machine Voting

AI involved with Voting

Pros

- Efficiency: Speeds up vote counting and result computation.
- Optimization: Designs new fair voting systems.

Cons

- Transparency: Does not always ensure reliability and accountability.
- Prediction: May foresee and influence outcomes and trends.

Kubacka, Slavkovik & Rückmann. *Predicting the Winners of Borda, Kemeny, and Dodgson Elections with Supervised Machine Learning*. In the 17th Eur. Work. on MultiAgent Systems, 2020.

Types of Neural Networks

- Multilayer Perceptrons (MLPs) learn patterns to make predictions, such as classifying emails as spam or predicting house prices.
- Convolutional Neural Networks (CNNs) are great at recognizing and analyzing images, for example at identifying faces in photos or detecting objects in videos.
- Transformers are advanced models used to understand and generate human language, and perform well in making translations and summaries, or answering questions.

Which one best captures the voting domain?

Modelling Challenges

- Data Representation: Effectively representing votes in a format that neural networks can process is non-trivial.
- Data Availability: Real election data is often not accessible, and generating realistic artificial data is an open question.
- Scalability: Handling elections with many voters and alternatives requires efficient algorithms and architectures.
- Parameter Setting: The network size, the data volume, the random seeds, the number of elections per gradient step, and the epochs can significantly impact the risk of overfitting or underfitting.

Learning Existing Voting rules (1)

Goal: Mimic the predefined outcomes via supervised learning.

- Burka et al. use an MLP and generate artificial elections of up to 11 voters and 5 alternatives (training on about 1000 of those).

  The MLP mimics more closely Borda, no matter on which rule it is trained: e.g., for 3 alternatives and 7 voters, trained on Plurality, the MLP mimics Borda with 95% and Plurality with 86% accuracy.

  The training data size impacts the results (e.g., when trained on Condorcet winner, the MLP mimics more closely Borda in sample-size up to 1000, and Copeland in larger samples).

  Increasing the size of the MLP (adding layers) is not important.

Burka, Puppe, Szepesváry, & Tasnádi. *Voting: A Machine Learning Approach*. European J. of Operational Research, 2022.

Learning Existing Voting rules (2)

Goal: Mimic the predefined outcomes via supervised learning.

- Anil and Bao use a kind of transformer and generate elections of up to 99 voters and 29 alternatives (testing on about 16,000 of those). They use 320,000 gradient steps.

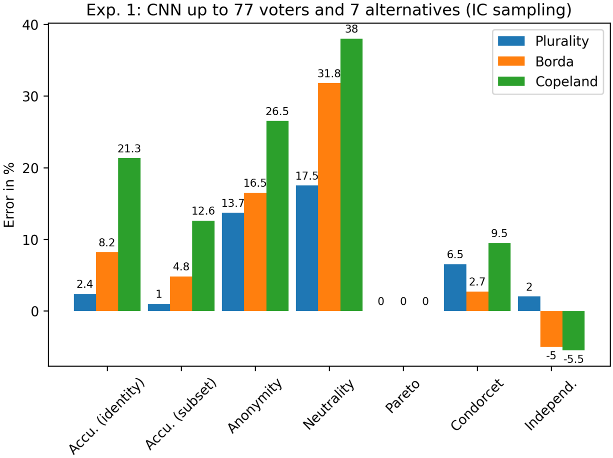
  The network can mimic Plurality and Borda with accuracy almost 100%, and Copeland with around 80%. It also generalizes to unseen real datasets and elections with an unseen number of voters.

Is learning connected to the "conceptual complexity" of a rule?

Anil. and Bao. *Learning to Elect*. In Advances in Neural Information Processing Systems, 2021.

Accuracy vs. Interpretability

- The architecture can have basic requirements built in; e.g., transformers are permutation-invariant (in voting: anonymous).
- Otherwise, we should test if the network learns for the right reasons.



Designing New Voting Rules (1)

- Anil and Bao generate synthetic data based on underlying utility destributions and define an optimal oracle for these utilities, according to some notion of social welfare. The network is trained to mimic that oracle with access only to the votes, not to the utilities.

  The transformer discovers a voting rule that beats classical ones in voting theory, and is very close to the optimal oracle.

- "Good" rules are also those satisfying some principles. Armstrong and Larson define a penalty for each alternative given an election; e.g., the number of axioms that are violated when this alternative is elected. They train a regression network to predict the penalty of the alternatives and elect the least penalized one.

  They use real data from Canadian federal elections and discover a rule that is highly Condorcet-consistent.

Armstrong and & Larson. *Machine Learning to Strengthen Democracy*. In NeurIPS Joint Work. on AI for Social Good, 2019.

Designing New Voting Rules (2)

Recall Arrow: No voting rule can simultaneously satisfy several desirable axioms. How does a rule coming closest to this ideal look like?

- We phrase this as an optimization problem: To employ gradient descent we formulate differentiable versions of the axioms via loss functions and obtain an unsupervised learning task.

  We observe that excluding certain axioms from the optimization task may actually increase their satisfaction.

  A (by construction anonymous) transformer is much better than an MLP and a CNN. It superceeds most existing rules for Pareto, Condorcet, and independence, and comes close to a rule similar to plurality with runoff (the french rule).

What does this tell us for the theory of voting?

Hornischer & Terzopoulou. Work in Progress.

LLMs as Voters

- Yang et al. use a human voting experiment with 180 participants to establish a baseline for human preferences and conduct a corresponding experiment with LLM (e.g., GPT-4) agents. They create persona descriptions from participant responses in the survey.

  The voting behavior of the LLMs is affected by the presentation order of the alternatives, and the numerical voting IDs of the LLMs.

  Different voting rules (like Borda) show that LLMs may lead to less diverse collective outcomes, although they are generally consistent with human choices. Adding randomness through temperature increases diversity but deviates from human choices. GPT-4 seems to over-rely on default/stereotypical demographics of the personas: e.g., university students always vote for better public transport.

Yang, Dailisan, Korecki, Hausladen, & Helbing. *LLM Voting: Human Choices and AI Collective Decision-Making*. arXiv, 2024

Manipulating Elections

- Holliday et al. train over 70,000 MLPs of 26 sizes to manipulate against 8 different voting rules, with 6 types of limited information, in about 4,000 elections with 5–21 voters and 3–6 alternatives.

  Generating preferences uniformly at random, sufficiently large MLPs learned to profitably manipulate all examined voting rules only with information about the pairwise majority victories between alternatives. But some Condorcet-extensions (e.g., Split Cycle) seemed more resistant than other rules (e.g., Plurality and Borda).

  Assumption: required model size indicates difficulty of manipulation.

  Generating preferences under a more realistic spatial model (2-D Euclidean), MLPs never learned to manipulate some rules like Split Cycle (perhaps because of rare and case-specific manipulations).

Holliday, Kritoffersen & Pacuit. *Learning to Manipulate under Limited Information*. arXiv, 2024

Social Choice for AI alignment

- Noothigattu et al. use data from the Moral Machine Experiment to build a model of aggregated moral preferences.
- Mishra proves via Arrow that there isn't AI treating all users and reinforcers (i.e., representative human supervisors) equally.
- Conitzer et al. connect social choice and AI alignment. E.g., the alternatives could be: all possible parameterizations of a network, or all its possible (prob. distributions over) answers.

Noothigattu et al. *A Voting-Based System for Ethical Decision Making*. AAAI, 2018

Mishra. *AI Alignment and Social Choice: Fundamental Limitations and Policy Implications*. arXiv, 2024

Conitzer et al. *Position: Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback*. ICML, 2024

Summary

We discussed the voting framework of Social Choice Theory from Economics, relevant for AI via Computational Social Choice.

Relevant topics:

- Learning existing voting rules.
- Accuracy vs. interpretability.
- Designing new voting rules.
- LLMs as voters.
- Manipulating elections.
- AI alignment.
- ...