Social Choice Theory and Machine Learning
Lecture 5

Eric Pacuit, University of Maryland

August 9, 2024

Course Plan

- ✓ introduction to mathematical analysis of voting methods, voting paradoxes;
- ▶ probabilistic voting methods (time permitting);
- ✓ quantitative analysis of voting methods (e.g., Condorcet efficiency);
- ✓ learning voting rules (PAC-learning, MLPs, other approaches);
- ▶ using modern deep learning techniques to generate synthetic election data;
- ⇒ strategic voting, learning to successfully manipulate voting rules based on limited information about how the other voters will vote using neural networks (multi-layer perceptrons);
- ⇒ RLHF (reinforcement learning with human feedback) and social choice;
- ▶ using large-language models to improve group decision-making; and
- ▶ liquid democracy (time permitting).

Learning to Manipulate under Limited Information

We use machine learning to gauge how resistant a preferential voting method is to manipulation under limited information about how other voters will vote.

Wesley Holliday, Alexander Kristoffersen, Eric Pacuit. *Learning to Manipulate under Limited Information*. arxiv.org/abs/2401.16412, 1st Workshop on Social Choice and Learning Algorithms (SCaLA 2024).

How to manipulate

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$$\downarrow$$

$$Borda(\mathbf{P}) = \{b\}$$

Winners

How to manipulate

Rankings

$R \in \{a\ b\ c\ d,\ \dots,\ d\ c\ b\ a\}$

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$$Borda(\mathbf{P}) = \{b\}$$

Winners

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
a	b	b	b	b
d	d	d	a	a
b	a	a	c	c

$$Borda(R, \mathbf{P}_{-v_1}) = \{c, d\}$$

Winners

How to manipulate

Rankings

$R \in \{a\ b\ c\ d,\ \ldots,\ d\ c\ b\ a\}$

v_1	v_2	v_3	v_4	v_5
c	c	c		
b	b	b		
a	d	d		
d	a	a	c	c

$Borda(\mathbf{P}) = \{b\}$

Winners

Which ranking R *should* v_1 submit?

v_1	v_2	v_3	v_4	v_5
		c	d	d
		b	b	b
		d	a	a
b	a	a	c	c

$Borda(R, \mathbf{P}_{-v_1}) = \{c, d\}$

Winners

Profitable manipulations

Given a profile of utilities for each voters, we can define the profile of rankings submitted by each voter, where alternative a is ranked above alternative b when the utility of a is greater than the utility of b:

Voters	a	b	c	d
v_1	0.1	0.65	0.9	0.08
v_2	0.7	0.9	1.0	0.8
v_3	0.01	0.03	0.5	0.02
v_4	0.1	0.5	0	0.9
v_5	0.1	0.2	0.05	1.0

U

$\Rightarrow$

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

P

Profitable manipulations

Given a profile of utilities for each voters, we can define the profile of rankings submitted by each voter, where alternative a is ranked above alternative b when the utility of a is greater than the utility of b:

Voters	a	b	c	d
v_1	0.1	0.65	0.9	0.08
v_2	0.7	0.9	1.0	0.8
v_3	0.01	0.03	0.5	0.02
v_4	0.1	0.5	0	0.9
v_5	0.1	0.2	0.05	1.0

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

P

Profitable manipulations

A ranking R is a *profitable manipulation* for voter i in preference profile $\mathbf{P}$ generated from a utility profile $\mathbf{U}$ for voting method F provided that

$$\mathbf{EU}_i(F_\ell\,(\,R, \mathbf{P}_{-i}\,)) > \mathbf{EU}_i(F_\ell\,(\,\mathbf{P}\,))$$

Profitable manipulations

A ranking R is a *profitable manipulation* for voter i in preference profile $\mathbf{P}$ generated from a utility profile $\mathbf{U}$ for voting method F provided that

$$\mathbf{EU}_i(F_\ell(R, \mathbf{P}_{-i})) > \mathbf{EU}_i(F_\ell(\boxed{\mathbf{P}}))$$

The profile where all voters submit their "true" ranking.

Profitable manipulations

A ranking R is a *profitable manipulation* for voter i in preference profile $\mathbf{P}$ generated from a utility profile $\mathbf{U}$ for voting method F provided that

$$\mathbf{EU}_i(F_\ell\left(R, \mathbf{P}_{-i}\right)) > \mathbf{EU}_i(F_\ell\left(\mathbf{P}\right))$$

The profile where all voters except i submit their “true” ranking and i submits R.

The profile where all voters submit their “true” ranking.

Profitable manipulations

A ranking R is a *profitable manipulation* for voter i in preference profile $\mathbf{P}$ generated from a utility profile $\mathbf{U}$ for voting method F provided that

Expected utility for voter i of the winners according to the voting method F, using even-chance tiebreaking if needed

$$\mathbf{EU}_i(F_\ell(R, \mathbf{P}_{-i})) > \mathbf{EU}_i(F_\ell(\mathbf{P}))$$

The profile where all voters except i submit their "true" ranking and i submits R.

The profile where all voters submit their "true" ranking.

Profitable manipulations

The *profitability* of voter i's submitting ranking R given utility profile $\mathbf{U}$ that induces preference profile $\mathbf{P}$ is given by

$$\frac{\mathbf{EU}_i(F_\ell(R, \mathbf{P}_{-i})) - \mathbf{EU}_i(F_\ell(\mathbf{P}))}{\max(\{\mathbf{U}_i(x) \mid x \in X\}) - \min(\{\mathbf{U}_i(x) \mid x \in X\})},$$

adopting the normalization of Relative Utilitarianism (Dhillon and Mertons 1999).

Limited information

U

Ideal Voter

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$
Winners

v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$
Winners

Choose an R that maximizes profitability

Limited information

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$\longrightarrow$

Plurality Scores

a	b	c	d
0	0	3	2

$\longrightarrow$

v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(\mathbf{P})$
Winners

$F(R, \mathbf{P}_{-1})$
Winners

Limited information

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$

Winners

Plurality Scores

a	b	c	d
0	0	3	2

v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$

Winners

v_1	v_2	v_3	v_4	v_5
c	d	c	c	d
b	a	b	b	a
a	b	d	a	b
d	c	a	d	c

,

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	a	d	a	a
a	b	b	b	b
d	d	a	c	c

, $\cdots$

Limited information

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$

Winners

Plurality Ranking

$$c > d > (a\ b)$$

v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$

Winners

v_1	v_2	v_3	v_4	v_5
c	d	c	c	d
b	a	b	b	a
a	b	d	a	b
d	c	a	d	c

,

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	a	d	a	a
a	b	b	b	b
d	d	a	c	c

, $\cdots$

Limited information

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$
Winners

Sincere Winners

a	b	c	d
0	1	0	0

v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$
Winners

v_1	v_2	v_3	v_4	v_5
c	d	b	b	a
b	b	a	d	d
a	a	d	a	c
d	c	c	c	b

,

v_1	v_2	v_3	v_4	v_5
c	b	b	d	b
b	a	c	a	d
a	c	a	c	c
d	d	d	b	a

, $\cdots$

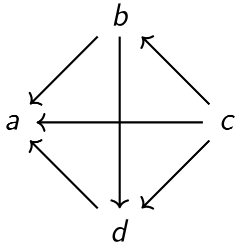
Limited information

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$
Winners

Majority Matrix



v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$
Winners

v_1	v_2	v_3	v_4	v_5
c	b	c	a	b
b	c	b	c	d
a	d	d	b	c
d	a	a	d	a

,

v_1	v_2	v_3	v_4	v_5
c	d	c	b	c
b	a	b	d	a
a	b	d	a	b
d	c	a	c	d

$, \cdots$

Limited information

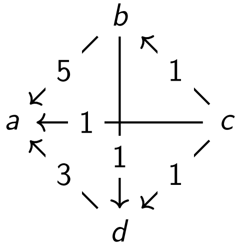
U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$

Winners

Margin Matrix



v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$

Winners

v_1	v_2	v_3	v_4	v_5
c	d	c	c	d
b	b	b	b	b
a	a	d	d	a
d	c	a	a	c

,

v_1	v_2	v_3	v_4	v_5
c	c	d	c	b
b	d	b	b	d
a	b	a	d	a
d	a	c	a	c

, $\cdots$

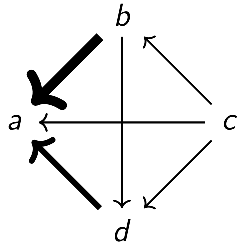
Limited information

U

v_1	v_2	v_3	v_4	v_5
c	c	c	d	d
b	b	b	b	b
a	d	d	a	a
d	a	a	c	c

$F(\mathbf{P})$
Winners

Qualitative Margin Matrix



v_1	v_2	v_3	v_4	v_5
?	c	c	d	d
?	b	b	b	b
?	d	d	a	a
?	a	a	c	c

$F(R, \mathbf{P}_{-1})$
Winners

v_1	v_2	v_3	v_4	v_5
c	d	c	c	d
b	b	b	b	b
a	a	d	d	a
d	c	a	a	c

,

v_1	v_2	v_3	v_4	v_5
c	c	d	c	b
b	d	b	b	d
a	b	a	d	a
d	a	c	a	c

, $\cdots$

Learning to manipulate under limited information

Learning to manipulate under limited information

- We trained $\approx$ 100,000 multi-layered perceptrons (MLP) of 26 sizes to manipulate against 8 different voting methods, under 6 types of limited information, in profiles with 5-21 voters and 3-6 alternatives.

Learning to manipulate under limited information

- We trained $\approx$ 100,000 multi-layered perceptrons (MLP) of 26 sizes to manipulate against 8 different voting methods, under 6 types of limited information, in profiles with 5-21 voters and 3-6 alternatives.
- These networks act as function approximators for profitable manipulation policies for a given voting method and type of limited information.

Learning to manipulate under limited information

- We trained $\approx$ 100,000 multi-layered perceptrons (MLP) of 26 sizes to manipulate against 8 different voting methods, under 6 types of limited information, in profiles with 5-21 voters and 3-6 alternatives.
- These networks act as function approximators for profitable manipulation policies for a given voting method and type of limited information.
- We evaluate the manipulation resistance of a voting method by the size and complexity of the network required to learn a profitable manipulation policy, as well as the average profitability of learned policies.

Setup

1. **Generate Utility Profiles**: We generate utility profiles for the voters for training, validation, and evaluation according to some probability model:

Setup

1. **Generate Utility Profiles**: We generate utility profiles for the voters for training, validation, and evaluation according to some probability model:
    - **Random Utility Model**: for each voter, the utility of each alternative is drawn independently from the uniform distribution on the $[0, 1]$ interval.

Setup

1. **Generate Utility Profiles**: We generate utility profiles for the voters for training, validation, and evaluation according to some probability model:
   - **Random Utility Model**: for each voter, the utility of each alternative is drawn independently from the uniform distribution on the $[0, 1]$ interval.
   - **2D Spatial Model**: each alternative and each voter is independently placed in $\mathbb{R}^2$ according to the multivariate normal distribution with no correlation between the two dimensions; then the utility of a alternative for a voter is the square of the Euclidean distance between the alternative and the voter

Setup

1. **Generate Utility Profiles**: We generate utility profiles for the voters for training, validation, and evaluation according to some probability model:
   - **Random Utility Model**: for each voter, the utility of each alternative is drawn independently from the uniform distribution on the $[0, 1]$ interval.
   - **2D Spatial Model**: each alternative and each voter is independently placed in $\mathbb{R}^2$ according to the multivariate normal distribution with no correlation between the two dimensions; then the utility of a alternative for a voter is the square of the Euclidean distance between the alternative and the voter
   - **Mallows Model**: generate a linear profile with the Mallows model ($\phi = 0.8$); then for each ranking generate a random utility that represents the ranking

Setup

1. **Generate Utility Profiles**: We generate utility profiles for the voters for training, validation, and evaluation according to some probability model:
   - **Random Utility Model**: for each voter, the utility of each alternative is drawn independently from the uniform distribution on the $[0, 1]$ interval.
   - **2D Spatial Model**: each alternative and each voter is independently placed in $\mathbb{R}^2$ according to the multivariate normal distribution with no correlation between the two dimensions; then the utility of a alternative for a voter is the square of the Euclidean distance between the alternative and the voter
   - **Mallows Model**: generate a linear profile with the Mallows model ($\phi = 0.8$); then for each ranking generate a random utility that represents the ranking
2. **Labeling**: For a given training profile and voting method, compute the optimal rankings that the manipulator could possibly submit.

Setup

3. **Training**:
   - The input to an MLP is (1) the manipulator's own utility function plus (2) some limited information about the profile.

Setup

3. **Training**:
   - The input to an MLP is (1) the manipulator's own utility function plus (2) some limited information about the profile.
   - Applying a softmax to the output yields a probability distribution π over all rankings that the manipulator could submit, which we reduce to the probability of choosing an optimal ranking or not.

Setup

3. **Training**:
   - The input to an MLP is (1) the manipulator's own utility function plus (2) some limited information about the profile.
   - Applying a softmax to the output yields a probability distribution π over all rankings that the manipulator could submit, which we reduce to the probability of choosing an optimal ranking or not.
   - We compute the final loss as the mean-squared error between the reduced distribution and the distribution assigning probability 1 to choosing an optimal-labeled ranking and 0 to choosing a non-optimal-labeled ranking.

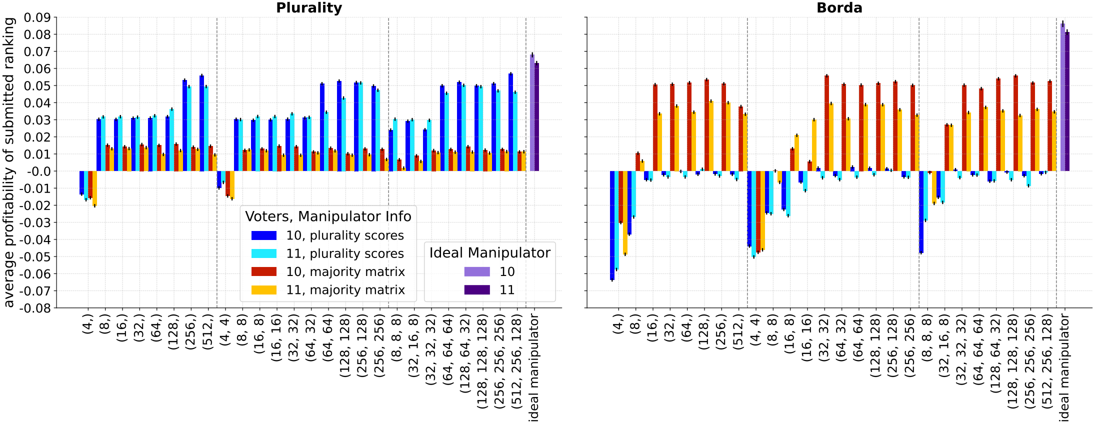
Setup

3. **Training**:
   - The input to an MLP is (1) the manipulator's own utility function plus (2) some limited information about the profile.
   - Applying a softmax to the output yields a probability distribution π over all rankings that the manipulator could submit, which we reduce to the probability of choosing an optimal ranking or not.
   - We compute the final loss as the mean-squared error between the reduced distribution and the distribution assigning probability 1 to choosing an optimal-labeled ranking and 0 to choosing a non-optimal-labeled ranking.

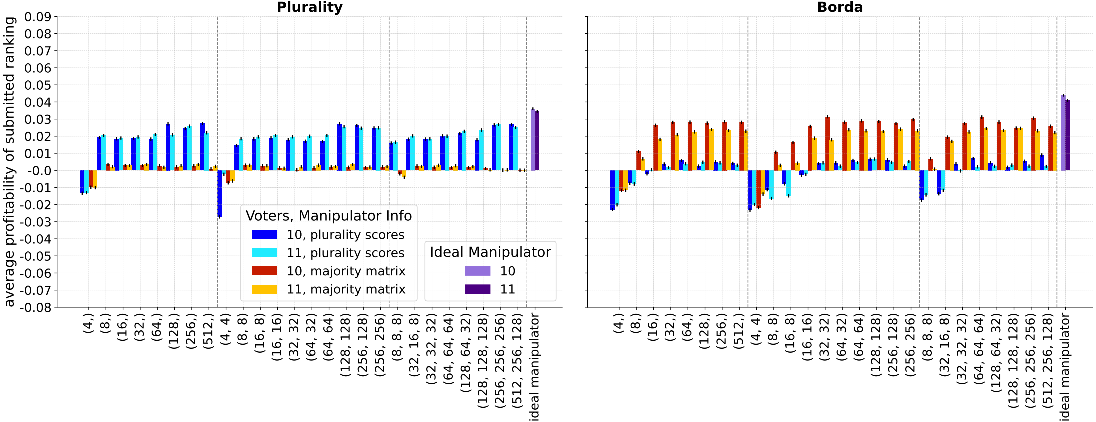
4. **Evaluation**: When evaluating the MLP, we take the most probable ranking R according to π to be submitted, and we compute the profitability of R.

Demonstration

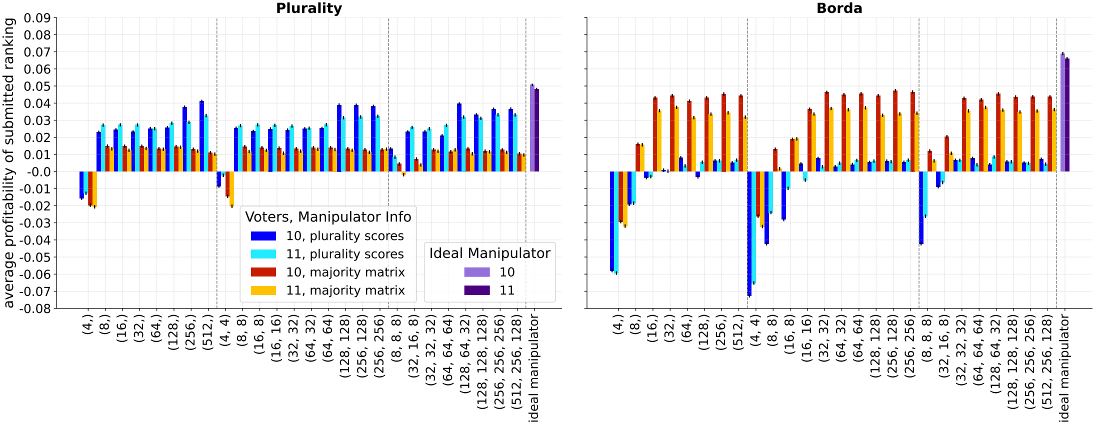
Results: Random Utility Model, 6 alternatives



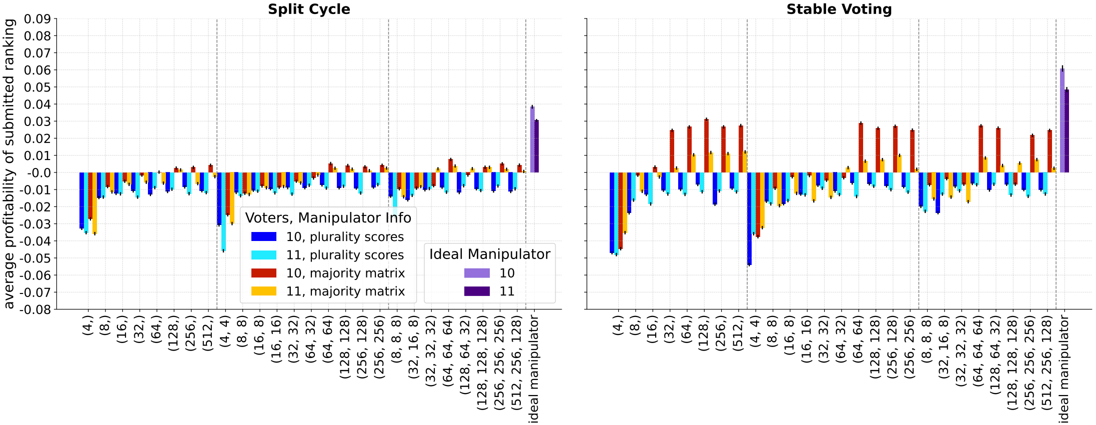
Results: 2D Spatial Model, 6 alternatives



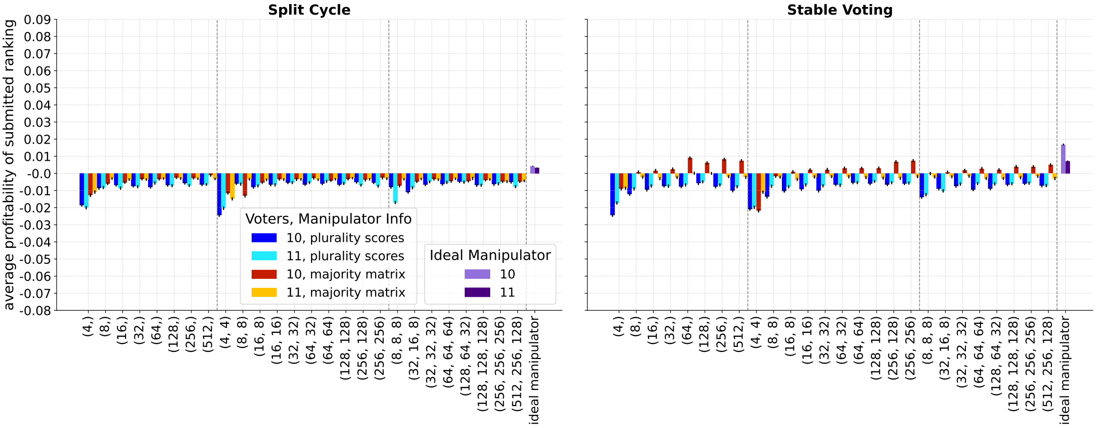
Results: Mallows Model, 6 alternatives



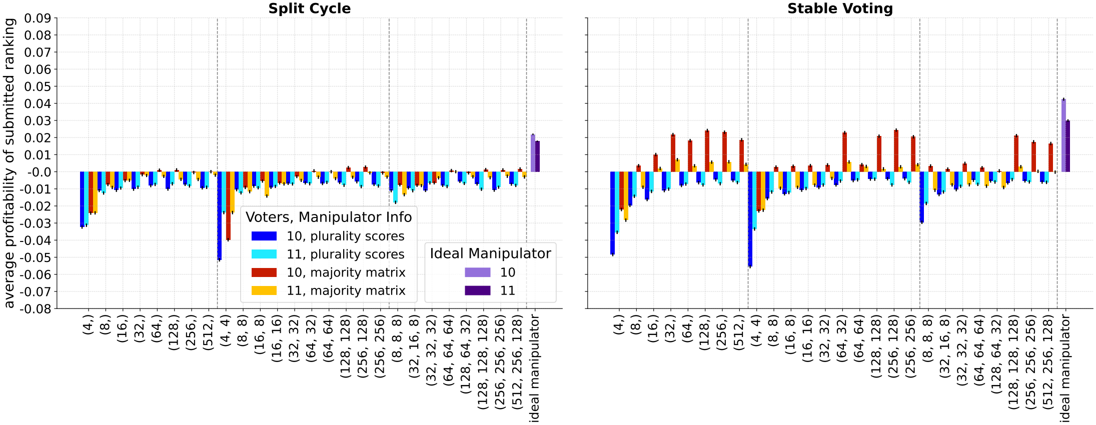
Results: Random Utility Model, 6 alternatives



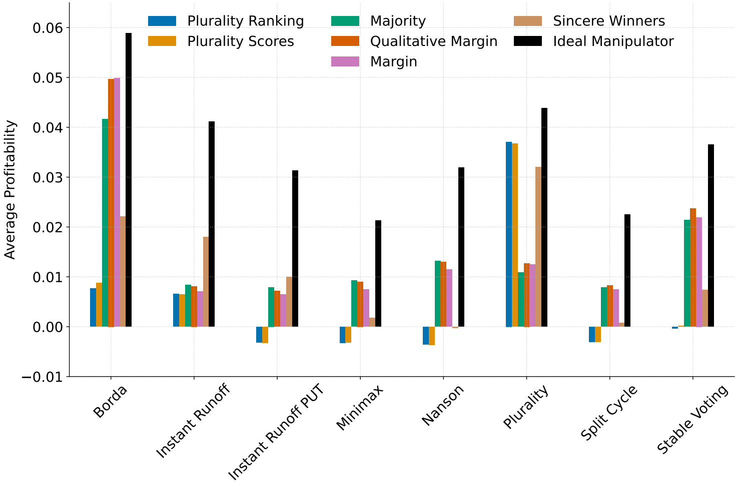
Results: 2D Spatial Model, 6 alternatives



Results: Mallows Model, 6 alternatives

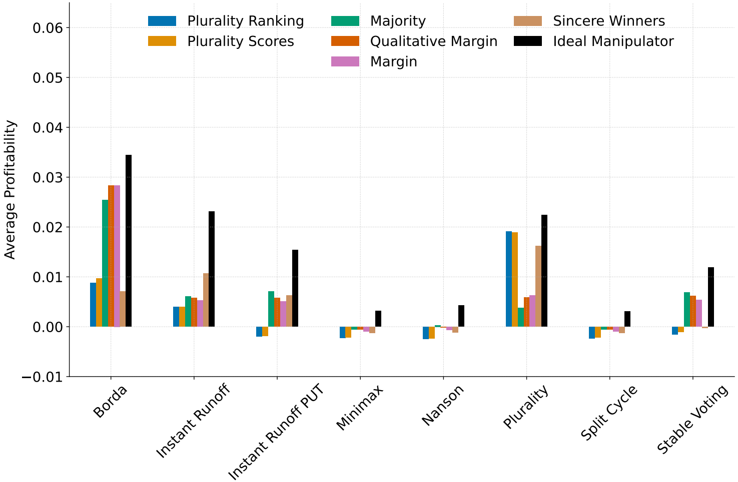


Results: Random Utility Model, 3-6 alternatives



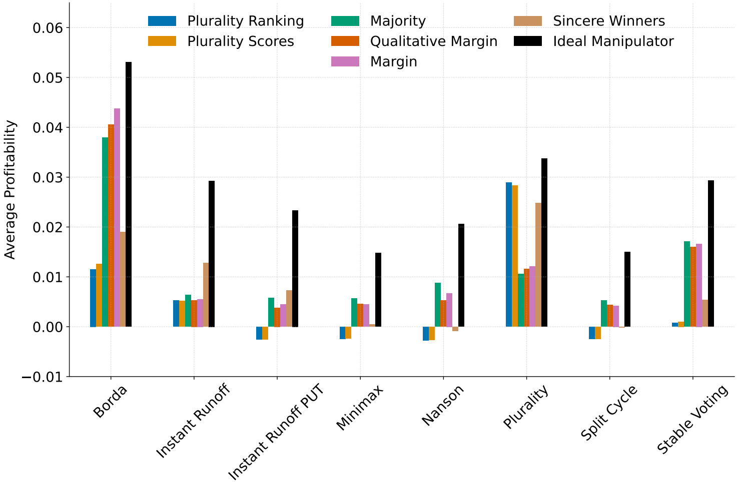
Average profitability of the best performing MLP with any hidden layer configuration for a given voting method and information type.

Results: 2D Spatial Model, 3-6 alternatives



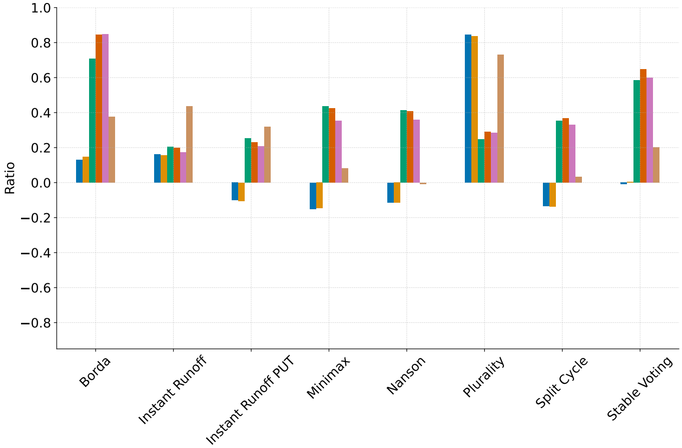
Average profitability of the best performing MLP with any hidden layer configuration for a given voting method and information type.

Results: Mallows Model, 3-6 alternatives



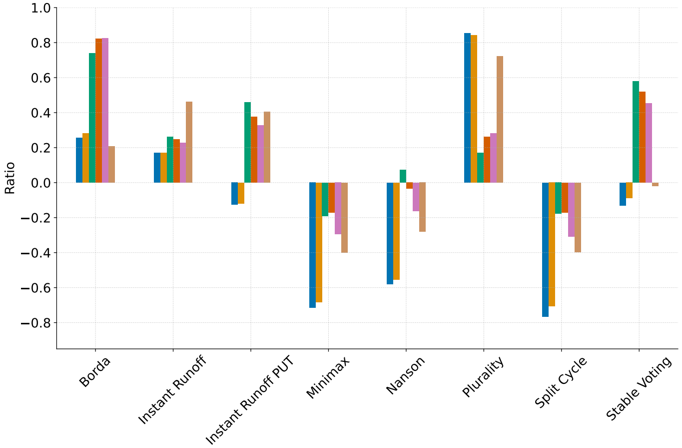
Average profitability of the best performing MLP with any hidden layer configuration for a given voting method and information type.

Results: Random Utility Model, 3-6 alternatives



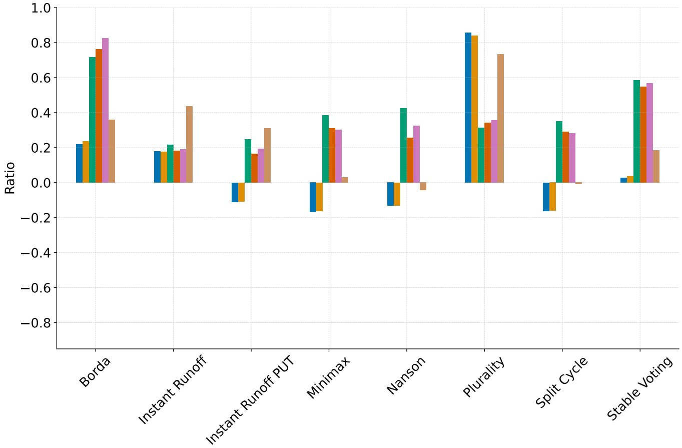
The ratio of the average profitability of the MLP's submitted ranking to that of the ideal manipulator's submitted ranking.

Results: 2D Spatial Model, 3-6 alternatives



The ratio of the average profitability of the MLP's submitted ranking to that of the ideal manipulator's submitted ranking.

Results: Mallows Model, 3-6 alternatives



The ratio of the average profitability of the MLP's submitted ranking to that of the ideal manipulator's submitted ranking.

Conclusion

It is possible for MLPs to learn to manipulate preferential voting methods on the basis of limited information, though the profitability of doing so varies significantly between different voting methods and types of information.

Conclusion

It is possible for MLPs to learn to manipulate preferential voting methods on the basis of limited information, though the profitability of doing so varies significantly between different voting methods and types of information.

Roughly three types of methods:

- **Highly manipulable even under limited info**: e.g., Borda;
- **Significantly manipulable under full info but not under limited**: e.g., Instant Runoff (though somewhat manipulable with sincere winners info);
- **Highly resistant to manipulation, especially under limited info**: e.g., Minimax.

Conclusion

Additional research questions:

- What about manipulation by a *coalition* of voters?

Conclusion

Additional research questions:

- What about manipulation by a *coalition* of voters?
- What if all voters simultaneously strategize?

Conclusion

Additional research questions:

- What about manipulation by a *coalition* of voters?
- What if all voters simultaneously strategize?
- What is the social cost or benefit of the learned manipulations?

Conclusion

Additional research questions:

- What about manipulation by a *coalition* of voters?
- What if all voters simultaneously strategize?
- What is the social cost or benefit of the learned manipulations?

Cf. K. Dowding and M. van Hees (2008), "In Praise of Manipulation," *British Journal of Political Science*, 38(1), pp. 1-15.

Conclusion

Based on considerations of manipulability, William H. Riker's (1988) wrote:

> I conclude that the meaning of social choices is quite obscure. They may consist of the amalgamation of the true tastes of the majority. . . or they may consist simply of the tastes of some people (whether a majority or not) who are skillful or lucky manipulators. If we assume social choices are often the latter, they may consist of what the manipulators truly want, or they may be an accidental amalgamation of what the manipulators (perhaps unintentionally) happened to produce. Furthermore, since we can by observation know only expressed values (never true values), we can never be sure, for any particular choice, which of these possible interpretations are correct. (p. 167)

Conclusion

Based on considerations of manipulability, William H. Riker's (1988) wrote:

> I conclude that the meaning of social choices is quite obscure. They may consist of the amalgamation of the true tastes of the majority. . . or they may consist simply of the tastes of some people (whether a majority or not) who are skillful or lucky manipulators. If we assume social choices are often the latter, they may consist of what the manipulators truly want, or they may be an accidental amalgamation of what the manipulators (perhaps unintentionally) happened to produce. Furthermore, since we can by observation know only expressed values (never true values), we can never be sure, for any particular choice, which of these possible interpretations are correct. (p. 167)

Can we mitigate these worries to some extent by the use of more manipulation-resistant preferential voting methods?

Wesley Holliday, Alexander Kristoffersen, Eric Pacuit. *Learning to Manipulate under Limited Information*. arxiv.org/abs/2401.16412, 1st Workshop on Social Choice and Learning Algorithms (SCaLA 2024).

https://github.com/epacuit/ltm

Models of voters behavior: IC (Impartial culture), IAC (Impartial anonymous culture), IANC (Impartial anonymous and neutral culture), Mallows models, Spatial models, Structured Preferences (e.g., Single Peaked models)

- https://comsoc-community.github.io/prefsampling/
- https://pref-voting.readthedocs.io/en/latest/generate_spatial_profiles.html
- http://preflib.org

N. Boehmer, P. Faliszewski, L. Janeczko, A. Kaczmarczyk, G. Lisowski, G. Pierczyński, S. Rey, D. Stolicki, S. Szufa, and T. Was (2024). *Guide to Numerical Experiments on Elections in Computational Social Choice*. arXiv preprint arXiv:2402.11765.

Probability of Condorcet Cycles

Is there *empirical evidence* that Condorcet cycles have shown up in real elections?

W. Riker. *Liberalism against Populism*. Waveland Press, 1982.

G. Mackie. *Democracy Defended*. Cambridge University Press, 2003.

Against the IC model

"...changing the distribution in *any fashion* (whether we call it 'realistic' or not) away from an impartial culture over linear orders will automatically have the effect of reducing the probability of majority cycles in infinite samples..." (pg., 28, 29)

This means that assuming an impartial culture is a *worst case analysis*.

M. Regenwetter, B. Gromfan, A. Marley, and I. Tsetlin. *Behavioral Social Choice*. Cambridge University Press, 2006.

See, also,

W. Gehrlein. *Condorcet's Paradox*. Springer, 2006.

F. Plassmann and T. N. Tideman. *How frequently do different voting rules encounter voting paradoxes in three-candidate elections?*. Social Choice and Welfare, 42, pp. 31 - 75, 2014.

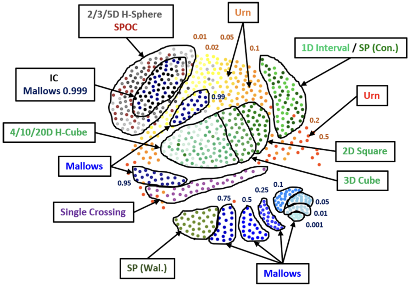


Figure 2: Visual representation of the election testbed. Each election is a dot whose colors give the statistical culture from which it was generated (the color of the statistical culture's label matches the color of its elections). For Urn and Mallows elections we also provide the value of their ϕ and α parameter.

S. Szufa, P. Faliszewski, P. Skowron, A. Slinko, and N. Talmod (2020). *Drawing a Map of Elections in the Space of Statistical Cultures*. AAMAS 2020, May 9 - 13, Auckland, New Zealand.

Using ML Techniques to Generate Election Data

Jui Chien Lin, Farhad Mohsin, Sahith Bhamidipati, and Lirong Xia (2023). *Generating Election Data Using Deep Generative Models*. AI4SG workshop at AAAI-23.

Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback

Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, Emanuel Tewolde, and William S. Zwicker (2024). *Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback*. Proceedings of ICML.

Many collective decision procedures

4	4	9	4	2
A	A	B	C	C
B	C	C	A	B
C	B	A	B	A

Borda Count: CBA
Instant Runoff: ABC
Ranked Pairs: BCA

Figure: Individual rankings on the left (4 voters say ABC, 4 say ACB, etc.) lead to different aggregations on the right, depending on the aggregation rule.

Many collective decision procedures

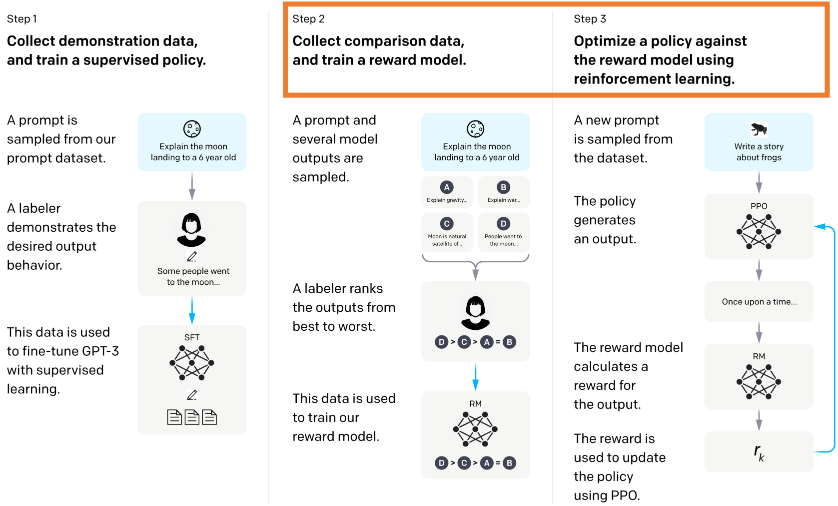
4	4	9	4	2
A	A	B	C	C
B	C	C	A	B
C	B	A	B	A

Borda Count: CBA
Instant Runoff: ABC
Ranked Pairs: BCA

Figure: Individual rankings on the left (4 voters say ABC, 4 say ACB, etc.) lead to different aggregations on the right, depending on the aggregation rule.

Reinforcement learning from human feedback (RLHF)

From Open AI (https://openai.com/index/instruction-following/):



Constitutional AI and RLAIF

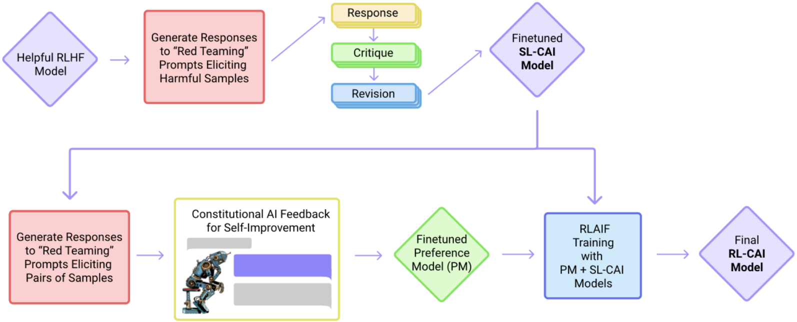


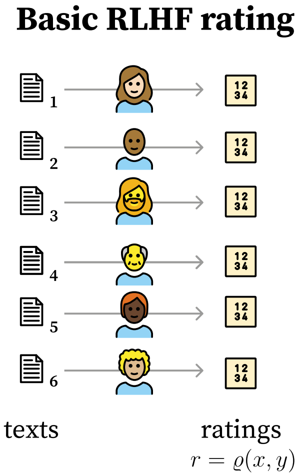
Figure 1 We show the basic steps of our Constitutional AI (CAI) process, which consists of both a supervised learning (SL) stage, consisting of the steps at the top, and a Reinforcement Learning (RL) stage, shown as the sequence of steps at the bottom of the figure. Both the critiques and the AI feedback are steered by a small set of principles drawn from a 'constitution'. The supervised stage significantly improves the initial model, and gives some control over the initial behavior at the start of the RL phase, addressing potential exploration problems. The RL stage significantly improves performance and reliability.

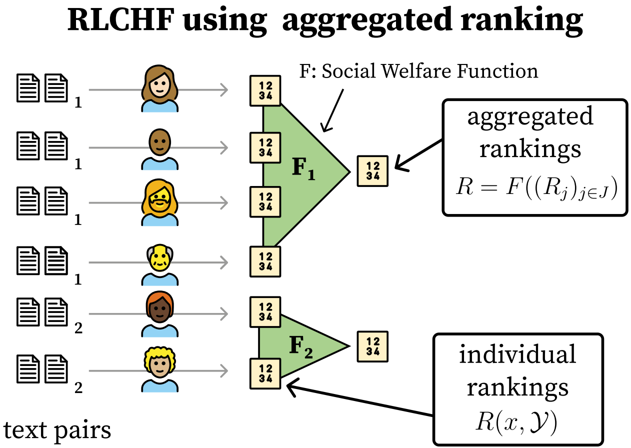
Bai et al., Constitutional AI: Harmlessness from AI Feedback.
https://arxiv.org/abs/2212.08073

Questions raised in the paper

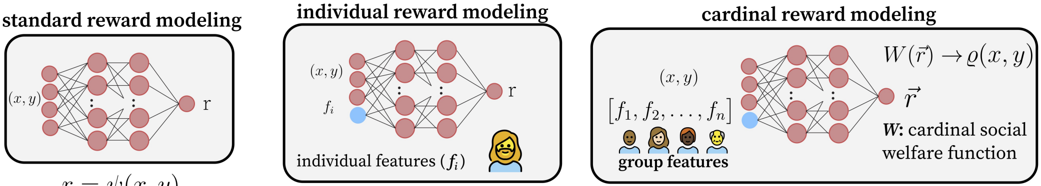
- How should we think about what the space of alternatives is?
- What type(s) of feedback should humans give?
- Who gets to give feedback, and how is it weighed?
  - How is a representative pool of stakeholders selected to give feedback?
- What about behavioral aspects / how should human cognitive structures be taken into account?
- What traditional social choice concepts are most relevant for AI alignment?
- When should we have multiple AI systems, and how do we avoid conflict between them?
- What are the limitations to dealing with diverging feedback?
- and more . . .

RLCHF using aggregated rankings





RLCHF using evaluator features and aggregated ratings



$$r = \psi(x, y)$$

$$r = \psi(x, f_i, y)$$

$$\varrho(x, y) = W(\psi(x, f_1, y), \ldots, \psi(x, f_N, y))$$

Relevant concepts from social choice

There are many concepts from social choice theory that are relevant to AI alignment:

- Axioms—but some are more relevant than others
- Strategic voting
- Anonymity vs. weighted votes
- Principles as voters
- and more...

Independence of clones

In social choice problems, sometimes multiple alternatives, say A and B, compare very similarly against every other alternative X, according to the preferences of individuals.

Independence of clones

In social choice problems, sometimes multiple alternatives, say A and B, compare very similarly against every other alternative X, according to the preferences of individuals.

Such alternatives are referred to as **clones**.

- According to a strict notion of clones (Tideman 1987), A and B are clones if, for every individual, if that individual prefers A to some other alternative X, then they also prefer B to X, and if they instead prefer X to A, then they also prefer X to B.

Independence of clones

In social choice problems, sometimes multiple alternatives, say A and B, compare very similarly against every other alternative X, according to the preferences of individuals.

Such alternatives are referred to as **clones**.

- According to a strict notion of clones (Tideman 1987), A and B are clones if, for every individual, if that individual prefers A to some other alternative X, then they also prefer B to X, and if they instead prefer X to A, then they also prefer X to B.
- According to a more liberal notion (Laffond et al. 1996), A and B are clones if, whenever a majority of individuals prefer A to some other alternative X, then a majority prefers B to X as well, and whenever a majority prefers some X to A, then a majority prefers X to B as well.

Example

Suppose a group of people are voting over where to go for dinner, and the only two alternatives are a **Chinese restaurant** and an **Indian restaurant**.

Example

Suppose a group of people are voting over where to go for dinner, and the only two alternatives are a **Chinese restaurant** and an **Indian restaurant**.

- 52% of the voters prefer the Chinese restaurant; 48% prefer Indian.

Example

Suppose a group of people are voting over where to go for dinner, and the only two alternatives are a **Chinese restaurant** and an **Indian restaurant**.

- 52% of the voters prefer the Chinese restaurant; 48% prefer Indian.

But then, someone points out that the Chinese restaurant has two floors and argues that the two floors should be considered separate options. So now the alternatives are C_1, C_2, and I.

Example

Suppose a group of people are voting over where to go for dinner, and the only two alternatives are a **Chinese restaurant** and an **Indian restaurant**.

- 52% of the voters prefer the Chinese restaurant; 48% prefer Indian.

But then, someone points out that the Chinese restaurant has two floors and argues that the two floors should be considered separate options. So now the alternatives are C_1, C_2, and I. Nobody really cares all that much about the floor, but suppose that

- 26% of the voters prefer $C_1 \succ C_2 \succ I$, and
- 26% of the voters prefer $C_2 \succ C_1 \succ I$ (adding up to the original 52%).

Example

Suppose a group of people are voting over where to go for dinner, and the only two alternatives are a **Chinese restaurant** and an **Indian restaurant**.

- 52% of the voters prefer the Chinese restaurant; 48% prefer Indian.

But then, someone points out that the Chinese restaurant has two floors and argues that the two floors should be considered separate options. So now the alternatives are C_1, C_2, and I. Nobody really cares all that much about the floor, but suppose that

- 26% of the voters prefer $C_1 \succ C_2 \succ I$, and
- 26% of the voters prefer $C_2 \succ C_1 \succ I$ (adding up to the original 52%).

Further suppose that the voting rule used is Plurality, in which the alternative that appears at the very top of voters' rankings the most often wins.

Example

Suppose a group of people are voting over where to go for dinner, and the only two alternatives are a **Chinese restaurant** and an **Indian restaurant**.

- 52% of the voters prefer the Chinese restaurant; 48% prefer Indian.

But then, someone points out that the Chinese restaurant has two floors and argues that the two floors should be considered separate options. So now the alternatives are C_1, C_2, and I. Nobody really cares all that much about the floor, but suppose that

- 26% of the voters prefer $C_1 \succ C_2 \succ I$, and
- 26% of the voters prefer $C_2 \succ C_1 \succ I$ (adding up to the original 52%).

Further suppose that the voting rule used is Plurality, in which the alternative that appears at the very top of voters' rankings the most often wins.
Then the Indian restaurant ends up winning now with 48% of the vote.

Independence of clones

Independence of cones, as an axiom for voting rules, states that introducing a clone should not affect whether a non-clone (e.g., the Indian restaurant in our example) is selected or which non-clone is selected.[1]

[1] But it may affect which clone, if any, is selected. For instance, a clone-independent rule could select C_1 over C_2 in our example, if among the 48% of people who prefer I, a strict majority of them prefer C_1 to C_2.

Independence of clones

Independence of cones, as an axiom for voting rules, states that introducing a clone should not affect whether a non-clone (e.g., the Indian restaurant in our example) is selected or which non-clone is selected.[1]

This axiom can also be generalized to rules that output a ranking.

[1] But it may affect which clone, if any, is selected. For instance, a clone-independent rule could select C_1 over C_2 in our example, if among the 48% of people who prefer I, a strict majority of them prefer C_1 to C_2.

Fact. Instant Runoff Voting satisfies independence of clones (even when generalized to allow voters to submit ties).

T. Delemazure and D. Peters (2024). *Generalizing Instant Runoff Voting to Allow Indifferences*. In Proceedings of EC '24.

Fact. Beat Path, Ranked Pairs, and Split Cycle all satisfy Independence of Clones

Independence of clones, Borda, and RLHF

The **Borda count** seriously violates independence of clones.

Independence of clones, Borda, and RLHF

The **Borda count** seriously violates independence of clones. For example:

Fact

Given a preference profile $\mathbf{P}$ in which candidate A is not Pareto dominated by any other candidate, one can add clones of A below A in each voter's ranking so that in the resulting profile $\mathbf{P}'$, A is the unique Borda winner.

Independence of clones, Borda, and RLHF

The **Borda count** seriously violates independence of clones. For example:

Fact

Given a preference profile $\mathbf{P}$ in which candidate A is not Pareto dominated by any other candidate, one can add clones of A below A in each voter's ranking so that in the resulting profile $\mathbf{P}'$, A is the unique Borda winner.

This is noteworthy in light of the following connection between Borda and RLHF.

Theorem (Siththaranjan et al. 2023)

Given voters' pairwise preferences on a set X of alternatives, let u be the utility function on X inferred using MLE with the **Bradley-Terry** model. Then the ordinal ranking of X derived from u is exactly the **Borda ranking**.

Conclusion

Our position on alignment: **methods from social choice should be applied to address questions such as which humans should provide input, what type of feedback should be collected, and how it should be aggregated and used**.

Conclusion

Our position on alignment: **methods from social choice should be applied to address questions such as which humans should provide input, what type of feedback should be collected, and how it should be aggregated and used**.

There have been several other recent papers at this intersection, including:

- "Axioms for AI Alignment from Human Feedback," Luise Ge, Daniel Halpern, Evi Micha, Ariel D. Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu, arXiv:2405.14758.
- "Mapping Social Choice Theory to RLHF," Jessica Dai and Eve Fleisig, arXiv:2404.13038
- "Distributional Preference Learning: Understanding and Accounting for Hidden Context in RLHF," Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell, arXiv:2312.08358.

Conclusion

Our position on alignment: **methods from social choice should be applied to address questions such as which humans should provide input, what type of feedback should be collected, and how it should be aggregated and used**.

There have been several other recent papers at this intersection, including:

- "Axioms for AI Alignment from Human Feedback," Luise Ge, Daniel Halpern, Evi Micha, Ariel D. Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu, arXiv:2405.14758.
- "Mapping Social Choice Theory to RLHF," Jessica Dai and Eve Fleisig, arXiv:2404.13038
- "Distributional Preference Learning: Understanding and Accounting for Hidden Context in RLHF," Anand Siththaranjan, Cassidy Laidlaw, and Dylan Hadfield-Menell, arXiv:2312.08358.

We encourage researchers in both AI and social choice to join this effort!

Linear Social Choice

Luise Ge, Daniel Halpern, Evi Micha, Ariel D. Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu (2024). *Axioms for AI Alignment from Human Feedback*. arXiv:2405.14758.

Linear Social Choice

In social choice theory, axioms are typically defined for rules that map rankings over candidates to a single winner (social choice functions) or a ranking of the candidates (social welfare functions).

By contrast, we are interested in rules that assign a *reward to each candidate*. This gap is easy to bridge, though: we simply consider a ranking of the candidates by decreasing reward.

Linear Social Choice

A much more significant gap is that in classical social choice, all relevant candidates appear in the input preferences, whereas in our setting (where candidates correspond, e.g., to prompts and their responses), we are only given preferences over a relatively small set of candidates identified by their (known) features, and we need to generalize from this information.

Linear Social Choice

A much more significant gap is that in classical social choice, all relevant candidates appear in the input preferences, whereas in our setting (where candidates correspond, e.g., to prompts and their responses), we are only given preferences over a relatively small set of candidates identified by their (known) features, and we need to generalize from this information.

In practice, this entails using a restricted—commonly, parametric—class of reward models which map candidate features to real-valued rewards, and which we fit to existing data.

Linear Social Choice

Specifically, we assume that a *linear reward function* defined by a parameter vector determines the reward of each candidate by computing the inner product of the parameter vector and the feature vector of the candidate...

Each human participant (henceforth referred to as a voter) is associated with a parameter vector, which is unknown to us and is used to specify ordinal preferences over the candidates.

Linear Social Choice

Specifically, we assume that a *linear reward function* defined by a parameter vector determines the reward of each candidate by computing the inner product of the parameter vector and the feature vector of the candidate...

Each human participant (henceforth referred to as a voter) is associated with a parameter vector, which is unknown to us and is used to specify ordinal preferences over the candidates.

Our task is to design **linear rank aggregation rules**, which aggregate rankings induced by these individual linear functions into a collective ranking that is also induced by a linear function

Linear Social Choice

A ranking method satisfies **Pareto Optimality** provided that for all candidates a and b is every voter ranks a above b, then a must be ranked above b.

A ranking method satisfies **Pairwise Majority Consistency** provided that for all candidates a and b, a is ranked above b if, and only if, a majority of voters rank a above b.

Linear Social Choice

A ranking method satisfies **Pareto Optimality** provided that for all candidates a and b is every voter ranks a above b, then a must be ranked above b.

A ranking method satisfies **Pairwise Majority Consistency** provided that for all candidates a and b, a is ranked above b if, and only if, a majority of voters rank a above b.

Theorem

If a linear rank aggregation rule f optimizes a loss function ℓ that satisfies $\inf_x \ell(x) < \ell(0)$ and is either nondecreasing and weakly convex, or strictly convex (and possibly nonmonotone), then it fails Pairwise Majority Consistency and Pareto Optimality.

Generative Social Choice

S. Fish, P. Gölz, David Parkes, Ariel Procaccia, Gili Rusak, Itai Shapira, and Manuel Wüthrich (2024). *Generative social choice*. Proceedings of EC 2024.

Generative Social Choice

In our view, there are two fundamental obstacles to using classical social choice to answer open-ended questions, both of which can be circumvented by LLMs.

- *Unforeseen Alternatives*. In classical social choice, the set of alternatives is explicitly specified and static.

Generative Social Choice

In our view, there are two fundamental obstacles to using classical social choice to answer open-ended questions, both of which can be circumvented by LLMs.

- *Unforeseen Alternatives*. In classical social choice, the set of alternatives is explicitly specified and static.

    - By contrast, LLMs have the capability of generating alternatives that were not initially anticipated but find common ground between participants.

    - In principle, the possible outcomes of an LLM-augmented democratic process may span the universe of all relevant outcomes for the problem at hand, e.g., all possible bills or statements.

Generative Social Choice

In our view, there are two fundamental obstacles to using classical social choice to answer open-ended questions, both of which can be circumvented by LLMs.

- *Extrapolating Preferences*. In classical social choice theory, agents specify their preferences in a rigid format.

Generative Social Choice

In our view, there are two fundamental obstacles to using classical social choice to answer open-ended questions, both of which can be circumvented by LLMs.

- *Extrapolating Preferences*. In classical social choice theory, agents specify their preferences in a rigid format.
  - This approach clearly does not suffice if a democratic process may produce alternatives that were not previously anticipated, and therefore not elicited: to even know which alternatives would be promising to generate, the process must be able to extrapolate participants' preferences.
  - LLMs can address this problem by allowing participants to implicitly specify their preferences by expressing their opinions, values, or criteria in natural language.
  - The LLM can act as a proxy for the participant, predicting their preferences over any alternative, whether foreseen or newly generated.

Course Plan

- ✓ introduction to mathematical analysis of voting methods, voting paradoxes;
- ~~probabilistic voting methods (time permitting)~~;
- ✓ quantitative analysis of voting methods (e.g., Condorcet efficiency);
- ✓ learning voting rules (PAC-learning, MLPs, other approaches);
- ✓ using modern deep learning techniques to generate synthetic election data;
- ✓ strategic voting, learning to successfully manipulate voting rules based on limited information about how the other voters will vote using neural networks (multi-layer perceptrons);
- ✓ RLHF (reinforcement learning with human feedback) and social choice;
- ✓ using large-language models to improve group decision-making; and
- ~~liquid democracy (time permitting)~~.

Thank you!!

https://pacuit.org/esslli2024/social-choice-machine-learning/

https://pref-voting.readthedocs.io/

https://stablevoting.org