Modeling Meta-Agreement through Deliberation: An Adaptation of the DeGroot Model

Amir Sahrani
Computational Science MSc
University of Amsterdam

24 July 2025

Choosing a contractor

Alice thinks i3 delivers the best quality, and has no monetary restrictions.

Bob does not care as much for the quality, and thinks dwm is cheapest.

Charlie thinks qtile delivers the best quality, and is the cheapest contractor.

Alice: i3 $\succ$ qtile $\succ$ dwm

Bob: dwm $\succ$ i3 $\succ$ qtile

Charlie: qtile $\succ$ dwm $\succ$ i3

Instead of directly trying to pick a winner, we ask them to talk a bit before.

Choosing a contractor

They realize that Bob and Charlie hold mutually exclusive beliefs, namely dwm and qtile cannot both be the cheapest options. It turns out that qtile is running a deal!

Alice: i3 $\succ$ qtile $\succ$ dwm
Bob: qtile $\succ$ dwm $\succ$ i3
Charlie: qtile $\succ$ dwm $\succ$ i3

These preferences are now single-peaked. This now allows us to pick qtile as the winner, using the rule of picking the median alternative.

Motivation

The example introduces the two main objects of interests of this work. Namely, deliberation and strategyproofness.

Following work by List (2002) proposing deliberation as a mechanism of enforcing shared ‘issue-dimensions’.

We set out to find a formal description of deliberation, as well as a mechanistic computational model.

Using this, we hope to be able to understand deliberation and the process by which it increases shared issue-dimensions

Classical Result

Theorem

Gibbard-Satterthwaite theorem (1973, 1975). There exists no resolute social choice function for elections with 3 or more candidates that is surjective, strategyproof and non-dictatorial.

Solution: Single-peaked preferences allow for the median-voter rule to satisfy all axioms. (Black 1948)

Deliberation, a political science perspective

Tenets of deliberation (Cohen, 2002): Free, Equal, Reasoned, Consensus.

List (2002): Meta-agreement, unanimous consensus too strong on substantive agreement. Meta-agreement requires three hypotheses to be satisfied.

Honesty

Can deliberation really be considered strategyproof?

Using definitions of Rad and Roy (2021), we show that this is in fact not the case.

Trivially: people be artificially more stubborn

Fixed bias: People can misreport preferences, and minimize outcomes under different metrics.

New tenet of deliberation: Honesty

America in one Room (2020)

A large-scale deliberative experiment measuring the impact of structured political discussion on voter attitudes.

- Large scale deliberative experiment
- Measured pre- and post-intervention opinions of participants
- Deliberation caused participants to be more likely to vote, have more favorable opinions of political rivals, and be more likely to support Joe Biden

We use the data from this experiment to validate our own model.

Our model: The Adapted DeGroot model

DeGroot model reduces the group dynamics of opinion change to a network of trust.

Shown to be more accurate at modeling human belief updating than Bayesian updating

Meta-agreement $\rightarrow$ arguing over positions of candidates.

A deliberation step can be modeled as a matrix multiplication

$$P^{(1)} = TP^{(0)}$$

Our model: Computational Complexity

Problem (δ-DBVM(S))

Given: $A, B \in S^{n \times n}$, $k \in \mathbb{R}_{\geq 0}$

Decision: Does there exist a bijection $f : [n] \to [n]$, such that $\delta(A, f(B)) \leq k$?

Theorem

δ-DBVM(S) is NP-complete, for $\delta \in \{\ell_1, \ell_2\}$ and $S \in \{0, 1\}$

Sketch: We reduce to the 0-1 MAX-QAP

Experimental Setup

Using data from the America in one room experiment to inform each voter's opinion.

We randomly generate candidates by averaging over 1 or 10 random voters, voters inaccurately judge candidate positions.

Sample n random voters for a deliberation group, forming a dense network.

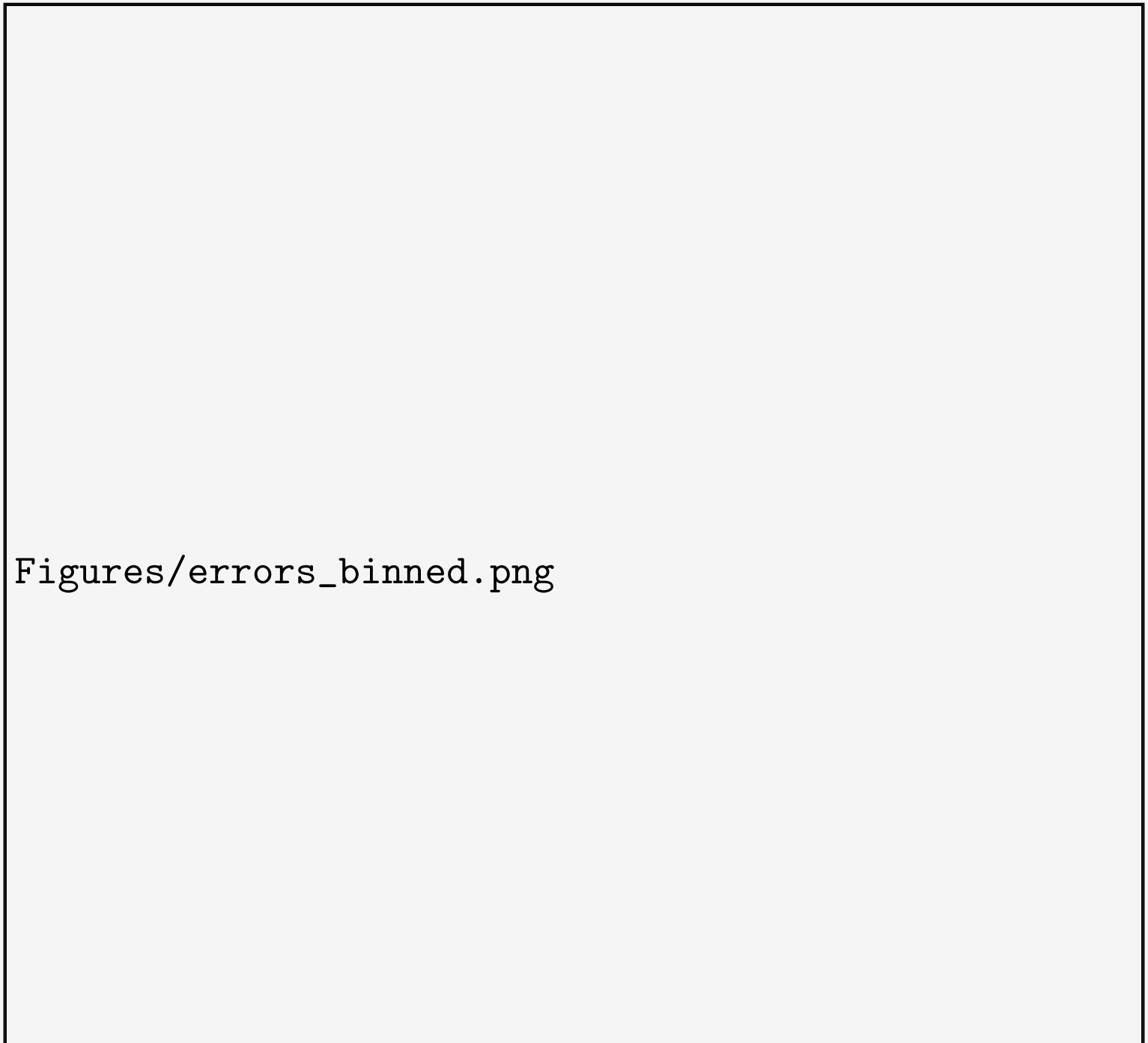
Trust matrices generated using:

- Knowledge
- Ego
- Similarity
- Bias

Results - PBS

Results - Breaking down PBS

Results - Errors



Results - Sensitivity

Results - Proximity to Single-Peakedness

Limitations

The model has poor individual predictive power.

Trust matrices are likely not realistic.

Voters' inaccuracy in judging candidates is likely not normally distributed.

Future work

Richer computational model

- Dynamic trust
- Non-linear interactions
- Negative influence
- Agent behavior based on social science literature

Proper data

Conclusion

Deliberation:

- Increases (proximity to) single-peakedness
- Does not ensure strategyproofness in a broad sense

Adapted DeGroot model:

- Poor predictor of individual opinion change
- Limited application with separate network data
- Requires more nuanced interactions