Anchor-proofness in Voting

Zoi Terzopoulou

CNRS, GATE, Lyon/Saint-Etienne

joint work with Federico Fioravanti

Anchoring Bias and Approval Voting

Framing effects guide individual and thus collective decisions.

Can the order of alternatives affect the outcome of an election?

- Voters have ordinal preferences and acceptability thresholds.
- A planner sets the presentation order of alternatives.
- Early alternatives serve as anchors, affecting which later alternatives receive approval.

This raises two questions:

- **Anchor-proofness:** When is the outcome independent of the presentation order?
- **Strategic ordering:** When can a planner manipulate the order of presentation to obtain a desirable outcome?

Example: Individual Preferences

A group of Greek friends is deciding how to spend a summer day.

- Everyone likes both going to the sea and going to the tavern.
- Some prefer the sea, others prefer the tavern.
- The alternatives are assessed one at a time.

Consider a friend who prefers to .

- Assesses the sea first: approves only the sea.
- Assesses the tavern first: approves both options.

The order of presentation affects approval ballots.

Example: Collective Outcome

Personalized order

Everyone sees their favorite alternative first

V1	V2	V3	V4	V5

Winner: (3 approvals vs. 2)

Common order

Everyone sees the same alternative first

V1	V2	V3	V4	V5

Winner: (5 approvals vs. 3)

Same rankings

Different order

↓

Different winner

(A glimpse of the) Results

- **No non-constant rule is anchor-proof** — but reasonable rules can be robust for restricted preference or order structures.

- **Zero information protects against manipulation** — but more information gives room to it.

Model

- Voters $N = \{1, \ldots, n\}$;
- Alternatives X, with $|X| > 2$;
- Each voter i has a preference-approval p_i: a ranking with acceptability threshold $t_i \in \{1, \ldots, |X|\}$.

Special cases:

- Tolerant: all alternatives acceptable for all voters.
- Intolerant: only top alternative acceptable for each voter.

Alternatives presented **sequentially** to voter i via permutation σ_i.

Ballot formation (anchoring bias):

- **Step 1:** Approve σ_i^1 if acceptable.
- **Step *k*:** Approve σ_i^k if acceptable and strictly preferred to all previously approved alternatives.

At each step the anchor is the best alternative encountered so far.

Example: From Preferences to a Ballot

Preference ranking

a

b

$\underline{c}$ ✓

Anchor: c

Presentation order

$c,\ a,\ b$

Approved: $\{c\}$

Example: From Preferences to a Ballot

Preference ranking

$\underline{a}$ ✓

b

c ✓

Anchor: a

Presentation order

$c,\ a,\ b$

Approved:$\{c, a\}$

Example: From Preferences to a Ballot

Preference ranking

$\underline{a}$ ✓

b ✗

c ✓

Anchor: a

Presentation order

$c,\ a,\ b$

Approved: $\{c, a\}$

Ballots may be inconsistent:
a voter approves c and rejects a preferred b.

Voting Rules and Axioms

Approval profile $(\boldsymbol{p}, \sigma)$: the approval ballot of every i via p_i and σ_i.

Approval-based voting rules $F : (\boldsymbol{p}, \sigma) \mapsto W \subseteq X$.

- Standard Approval Voting (SAV): largest approval count.
- Nomination Rule (Nom): approved by at least one voter.

Axioms:

- Neutrality/Anonymity: alternative/voter names are irrelevant.
- Weak Unan: if there are unanimously accepted alternatives, winners are among them.
- Total Unan: all alternatives unanimously accepted, then all win.
- Unanimity: if there are unanimously accepted alternatives, winners are exactly them.

SAV: satisfies Unan $\Rightarrow$ Weak Unan. & Total Unan.
Nom: satisfies Total Unan. but not Weak Unan. $\Rightarrow$ not Unan.

Anchor-Proofness

Collective outcomes faithfully represent voters' underlying preferences, immune to the choice of presentation order.

- F is **anchor-proof for preference-approval profile $\boldsymbol{p}$** if $F(\boldsymbol{A}_{\boldsymbol{p},\sigma}) = F_a(\boldsymbol{A}_{\boldsymbol{p},\pi})$ for every two orders σ, π.
- F is **anchor-proof** if this holds for every profile.

Main Impossibility: All Profiles

Theorem

Let F be **anchor-proof**. Then F is **constant**.

This result holds if we restrict to the domain of tolerant profiles.

Existence of an Anchor-Proof Profile

Does there exist *some* profile for which a given rule is anchor-proof for any pair of orders?

- **Yes** (general domain): every intolerant profile — only the top alternative can be approved, so the order is irrelevant.
- **No** (tolerant domain, reasonable rules):

> **Theorem**
>
> Let F satisfy weak and total unanimity.
> Then F is **not anchor-proof** for any tolerant profile.

Exceptions: Non-neutral rules can be anchor-proof for certain tolerant profiles.

Order Pairs Preserving Outcomes for All Profiles

Does there exist *some* pair of distinct orders that produce the same winners for all profiles?

> **Theorem**
>
> Let F satisfy weak and total unanimity.
> Then no such pair exists (even restricted to tolerant profiles).

Exceptions:

- Nom (violates weak unan.): if each alternative is shown first once, Nom selects all alternatives for all tolerant profiles.
- Non-neutral rules: certain order pairs can be made equivalent.

What is Manipulation?

The planner has strict preferences over 2^X and some *info* about $\boldsymbol{p}$.

A presentation order σ^* is the planner's optimal strategy if:

- **It is safe:** for every possible profile according to *info*, σ^* yields an outcome at least as good as any other order.
- **Sometimes it strictly helps:** for some possible profile, σ^* yields a strictly better outcome.

F is manipulable under *info* if an optimal strategy exists for some planner's preferences and some $\boldsymbol{p}$.

More information : more opportunities to manipulate.

Example (plurality *info*): The planner prefers to win.

She knows 2 voters rank first and 3 rank first.

She shows first to everyone: Safe and strictly beneficial!

Manipulation under Partial Information

Theorem

Let F satisfy weak and total unanimity.
Then F is **not manipulable** under zero *info*.

SAV and Nom are not manipulable under zero *info*.

Theorem

Let F satisfy weak and total unanimity.
Then F is **manipulable** under plurality *info*.

Theorem

Let F satisfy weak unanimity, anonymity, neutrality.
Then F is **manipulable** under acceptability *info*.

Anchor-Proof Profiles

Full *info* manipulation is impossible for anchor-proof profiles.

Theorem

Every weakly unanimous rule is anchor-proof for p iff p is:

- **intolerant**, or
- has a **unique unanimously accepted alternative ranked first by every voter**.

The planner almost always has room to act.

Anchor-Proof Profiles for SAV and Nom

Theorem

SAV is anchor-proof for $\boldsymbol{p}$ iff for every plurality maximizer x:

$$plur_{\boldsymbol{p}}(x) \geq acc_{\boldsymbol{p}}(y) \quad \forall\, y \in X,$$

with equality only among plurality maximizers.

Plurality leaders are so dominant no alternative can overtake them via acceptability votes.

Theorem

Nom is anchor-proof for $\boldsymbol{p}$ iff the set of top alternatives is the set of acceptable alternatives.

All acceptable alternatives are already someone's top choice; reordering cannot unlock new winners.

Conclusion

So far:

- **Anchoring bias in voting:** model and basic definitions.
- **Impossibility:** no non-constant rule is anchor-proof; the planner can always exploit "natural" rules.
- **Information is key:** no manipulation under zero information; extra information restores manipulation.

Next:

- **Richer ballots:** trichotomous ballots — a preferred alternative seen after a better anchor could be neutral rather than rejected.
- **Optimization:** min. vulnerability; approx. unbiased outcome.
- **Other biases:** framing, loss aversion, social conformity.