Online Deliberation

Optimized Democracy guest lecture · Carter Blair and Jakob de Raaij · Oct 28, 2026

Agenda

1. **Introduction** — Carter
   Deliberation and online deliberation

2. **Bridging** — Jakob
   How online deliberation platforms operationalize bridging
   A new formulation of bridging

3. **Community Notes** — Carter

4. **Wrap-up** — Carter

Where Do Preferences Come From?



- So far in the course we have looked at the relationship between **underlying and reported preferences**: strategic misreporting, distortion, noisy signals of a ground truth.

Aggregative vs. Deliberative Democracy

Two competing models?

> *"**Voting-centric** views see democracy as the arena in which **fixed** preferences and interests compete via fair mechanisms of aggregation. In contrast, **deliberative** democracy focuses on the communicative processes of opinion and will-formation…"*
>
> Simone Chambers, "Deliberative Democratic Theory," *Annual Review of Political Science* (2003)

Aggregative

Voting-centric

Fixed preferences, combined by a rule

vs.

Deliberative

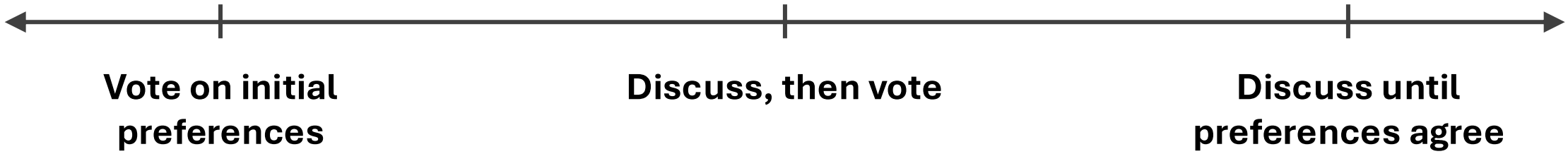
Talk-centric

Preferences formed through discussion

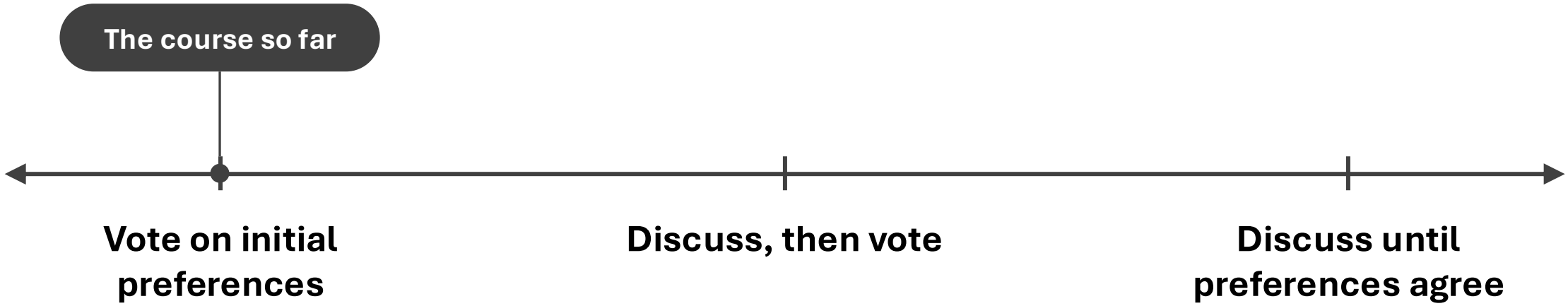
Aggregative vs. Deliberative Democracy

Better seen as a spectrum

> *"The field of deliberative democracy now generally recognizes the* ***co-dependence of deliberation and voting****."*
>
> Simone Chambers & Mark Warren, "Why Deliberation and Voting Belong Together," *Res Publica* (2023)



Aggregative vs. Deliberative Democracy



Deliberation as an Operation on the Profile

σ = preference profile: each participant's ranking over the alternatives



Discussion changes the profile: **participants pool information and exchange arguments.**

A Quick Aside

Why is deliberation useful for social choice?

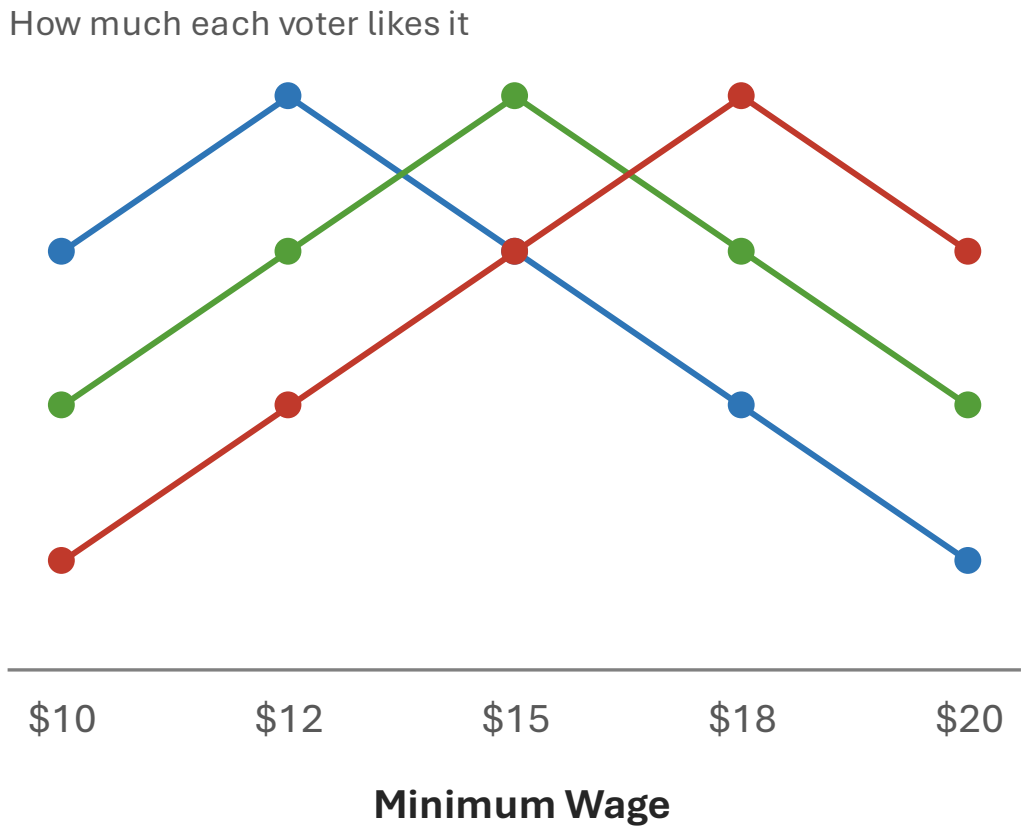
Recall: Single-Peaked Preferences Tame Majority Rule

Single-peaked

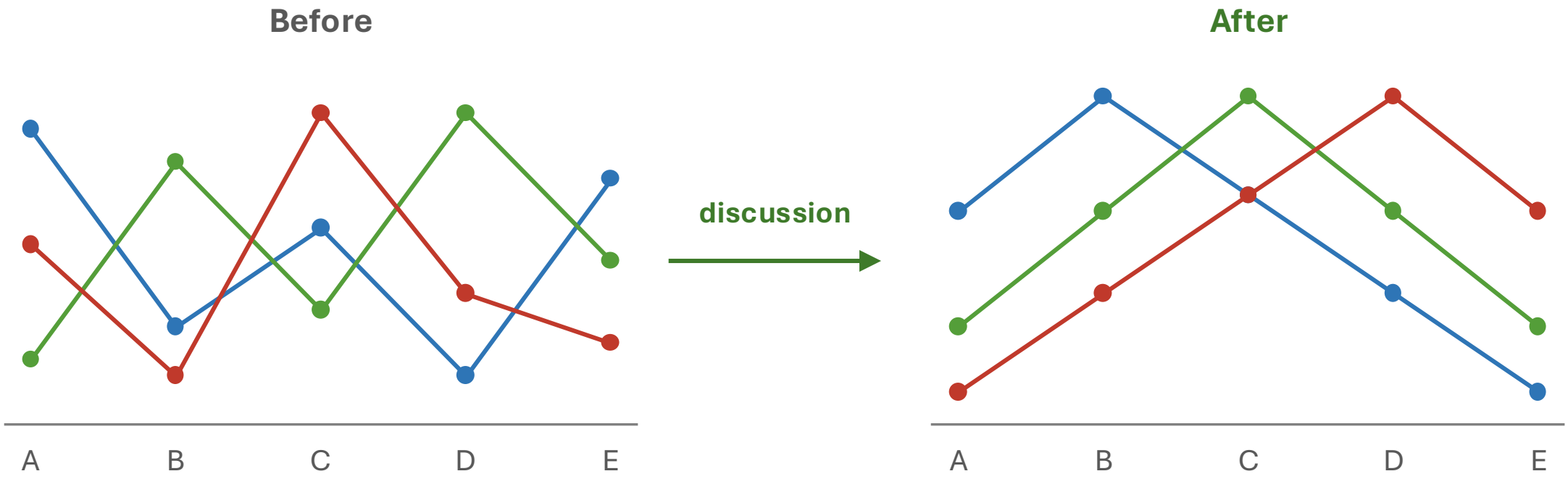
- Options lie on **one dimension**; each voter has **one favorite**

Then

- A **Condorcet winner** exists (the median's favorite)
- **No majority cycles**
- Majority rule is **strategy-proof**



Discussion Can Line Preferences Up on One Dimension



Dryzek & List (2003), *British Journal of Political Science*

Deliberation Increases Single-Peakedness

+0.10

average rise in proximity to single-peakedness (0–1 scale)

across 13 issues in Deliberative Polls (+0.13 on low- and moderate-salience issues)

- Participants ranked policy options **before and after** deliberating
- Proximity = share of participants whose rankings are **single-peaked on one common ordering**
- Not just more agreement: rankings became **more structured, not more alike**
- Not fully single-peaked, but **closer means fewer cycles**[1]

[1] Niemi (1969), "Majority Decision-Making with Partial Unidimensionality," *American Political Science Review*
Source: List, Luskin, Fishkin & McLean (2013), "Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy," *Journal of Politics*; averages from Farrar et al. (2010), *BJPS*

Deliberation in Practice

Ireland

Citizens' Assembly, 2016–18

99 randomly selected members. Referendum passed 25 May 2018 with 66.4%.

France

Citizens' Convention on Climate, 2019–20

150 members. Produced 149 climate proposals the president had promised to pass "unfiltered."

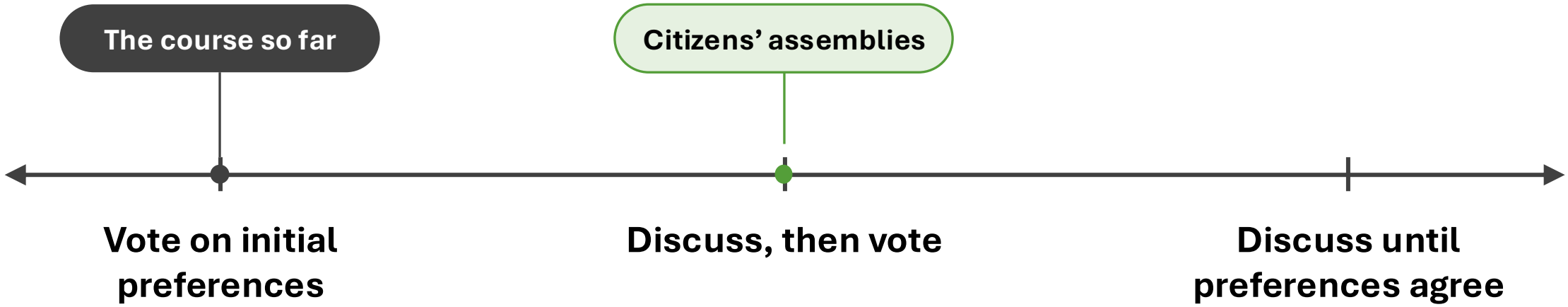
British Columbia

Citizens' Assembly on Electoral Reform, 2004

161 members. Designed a new electoral system (STV) that went to a province-wide referendum.

In-person deliberation is hard to scale beyond a few hundred people

Aggregative vs. Deliberative Democracy



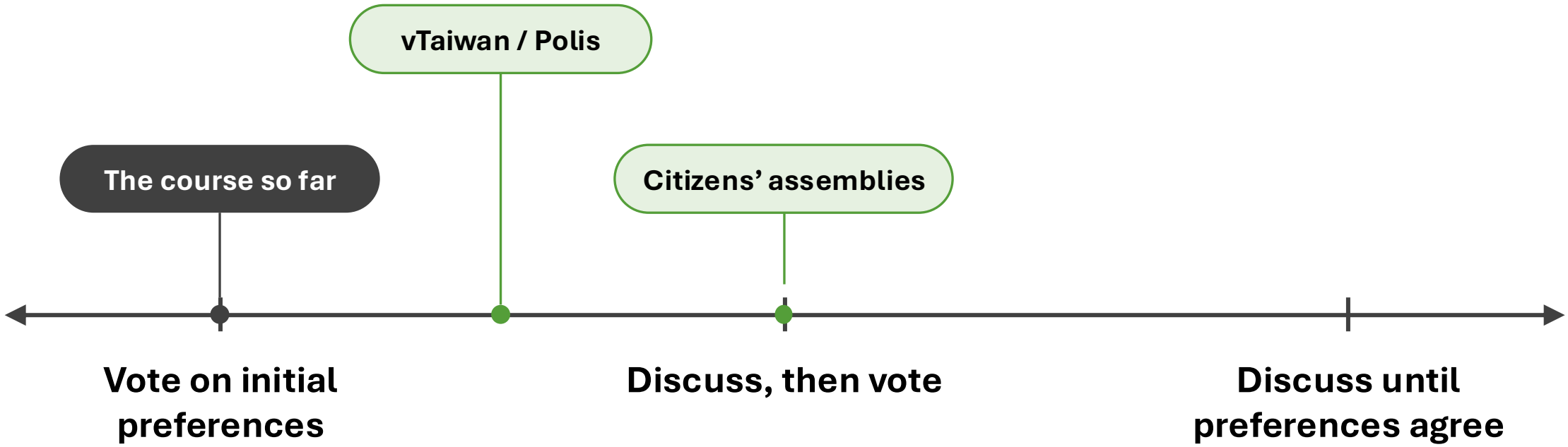
vTaiwan (2015)

Online, deliberation scales, so aggregation does more of the work

- **30× the scale:** regulating Uber, 2015, drew ~3,000 participants over four weeks on Polis, against ~100 in a citizens' assembly
- **Thinner discussion:** people vote agree / disagree / pass on others' statements; no replies
- **So the rule carries more weight:** Polis clusters participants by voting pattern and surfaces statements supported across clusters

Aggregative vs. Deliberative Democracy

At scale, the rule does more of the work. Which rule?



Bridging

What ideas/statements/policies should we look for?

An idea is **bridging** if it is approved by participants that usually hold opposing views

[Ovadya, 2022]



Find compromises

by highlighting ideas with wide and diverse support

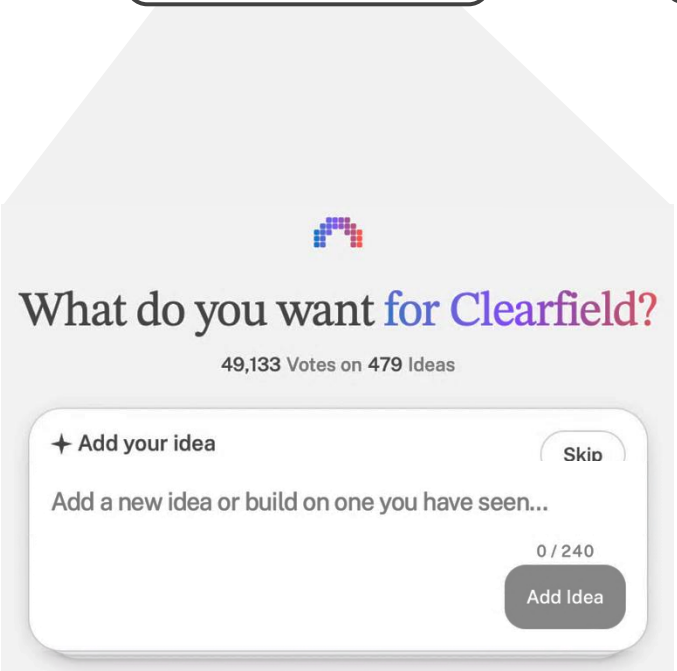


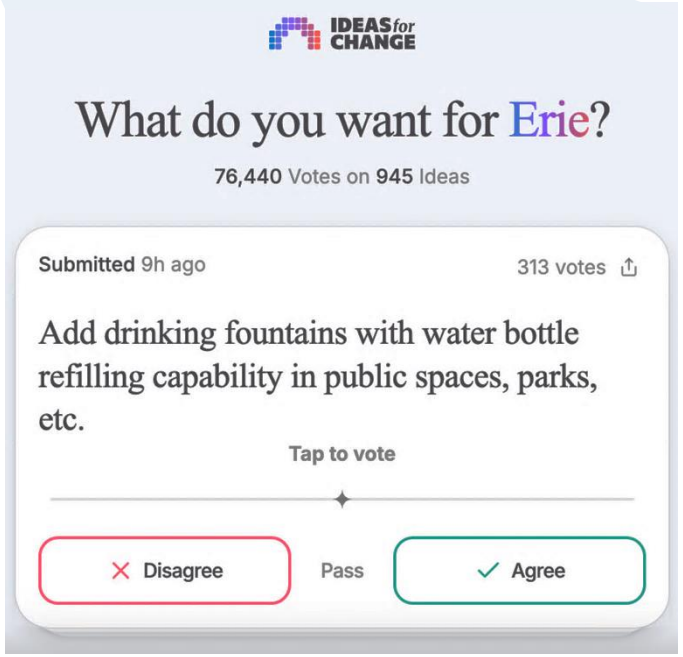
Increase mutual understanding

by emphasizing common ground across polarized groups

Approval Votes

Ideas from **participants** → **Approval votes** → **Bridging Metric** → **Bridging** comments





Approval Votes

Ideas from **participants**

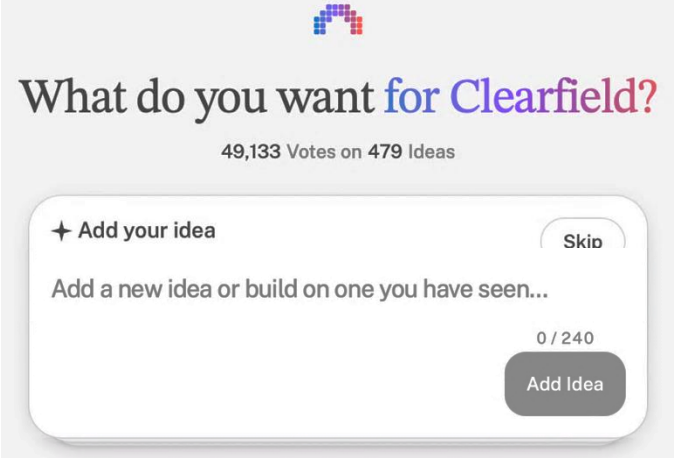
Approval votes

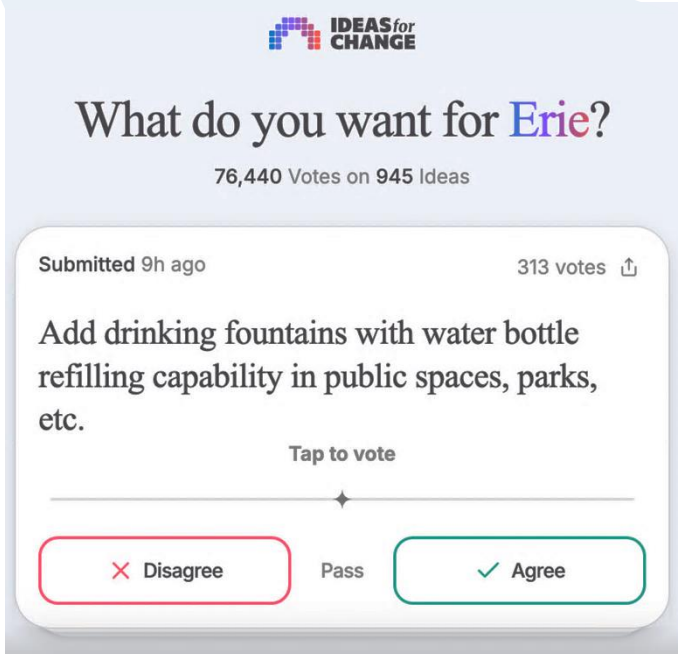
Bridging Metric

Bridging comments

LLM or small-group **deliberation**

new ideas



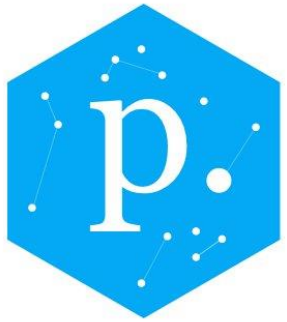


Vote Matrix

- A vote matrix records known and unknown votes

$$\begin{array}{c} \\ \text{ideas } C \end{array} \begin{array}{c} \text{voters } N \\ \begin{bmatrix} 1 & ? & ? & 0 & ? & 0 \\ 1 & 1 & ? & 0 & 0 & ? \\ ? & ? & 0 & 1 & 1 & ? \\ ? & 0 & ? & ? & ? & 1 \\ 1 & ? & 1 & ? & ? & 0 \end{bmatrix} \end{array}$$

- Can work with incomplete data, or run matrix completion
- For simplicity, assume matrix is completed (no unknown votes)

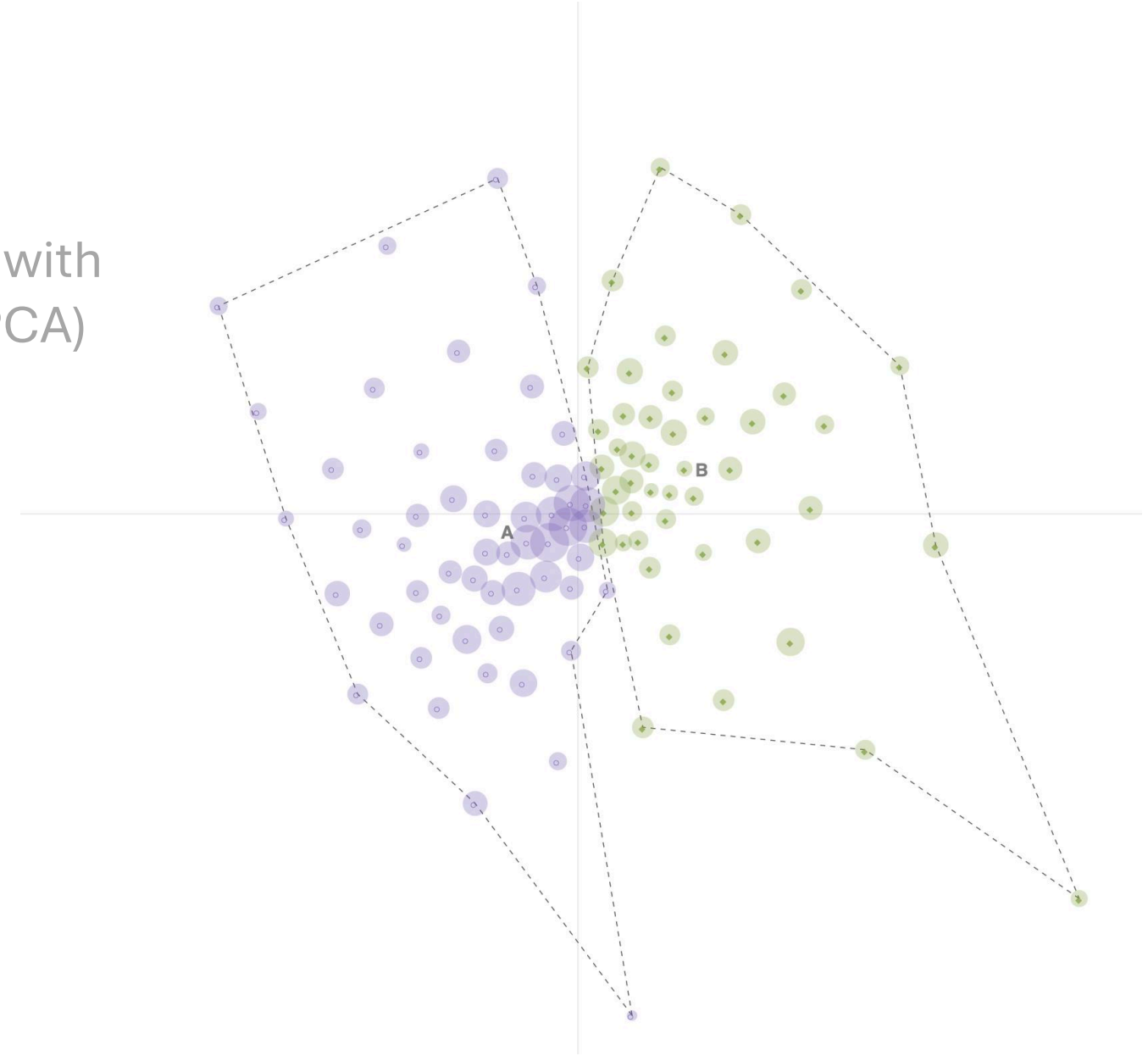


Pol.is

Used in Taiwan following the Sunflower Student Movement to allow citizens to discuss proposed laws. Led to widely approved policies on Uber, online alcohol sales, and others.

Polis - Bridging

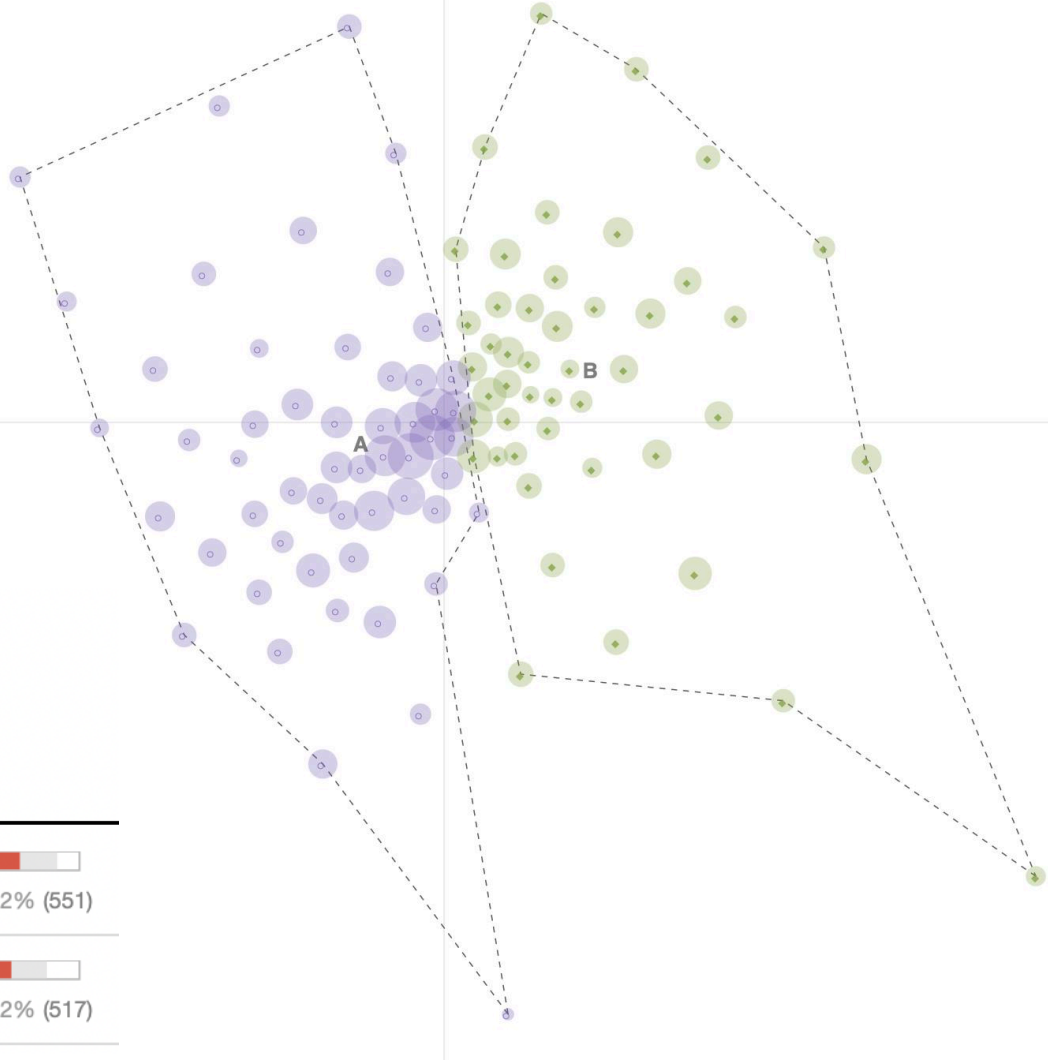
- Project voters into 2 dimensions with principle-component analysis (PCA)



https://pol.is/report/r2xcn2cdbmrzjmmuuytdk

Polis - Bridging

- Project voters into 2 dimensions with principle-component analysis (PCA)
- Run clustering algorithm to identify groups in population



Group A: 959 participants

Statements which make this group unique, by their votes:

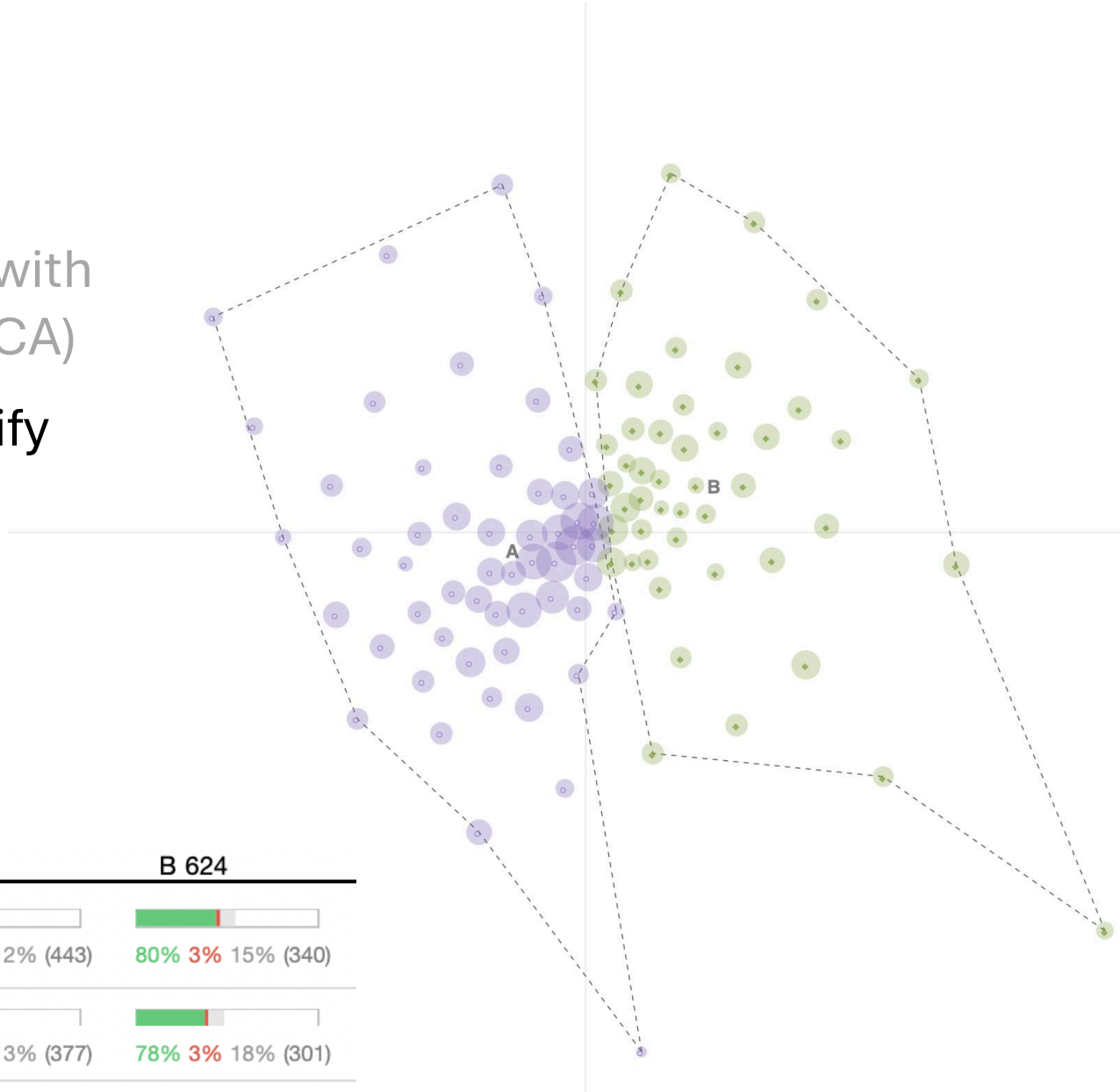
	STATEMENT	OVERALL 1583	A 959	B 624
20	It is embarrassing that our city is the largest in the state not to have a fairness ordinance.	43% 35% 21% (1287)	72% 6% 20% (736)	4% 72% 22% (551)
86	Bowling Green needs a fairness ordinance to protect the rights of its citizens.	48% 33% 18% (1208)	78% 6% 15% (691)	8% 68% 22% (517)

https://pol.is/report/r2xcn2cdbmrzjmmuuytdk

Polis - Bridging

- Project voters into 2 dimensions with principle-component analysis (PCA)
- Run clustering algorithm to identify groups in population
- The bridging score of idea c is

$$\prod_{group\ G} (\text{approval rate of } c \text{ in } G)$$



	STATEMENT	OVERALL 1583	A 959	B 624
83	Impact on traffic flow should be considered with any proposed new developments.	83% 2% 13% (783)	86% 1% 12% (443)	80% 3% 15% (340)
200	Bowling Green needs more competitive cable rates.	80% 3% 15% (678)	83% 3% 13% (377)	78% 3% 18% (301)

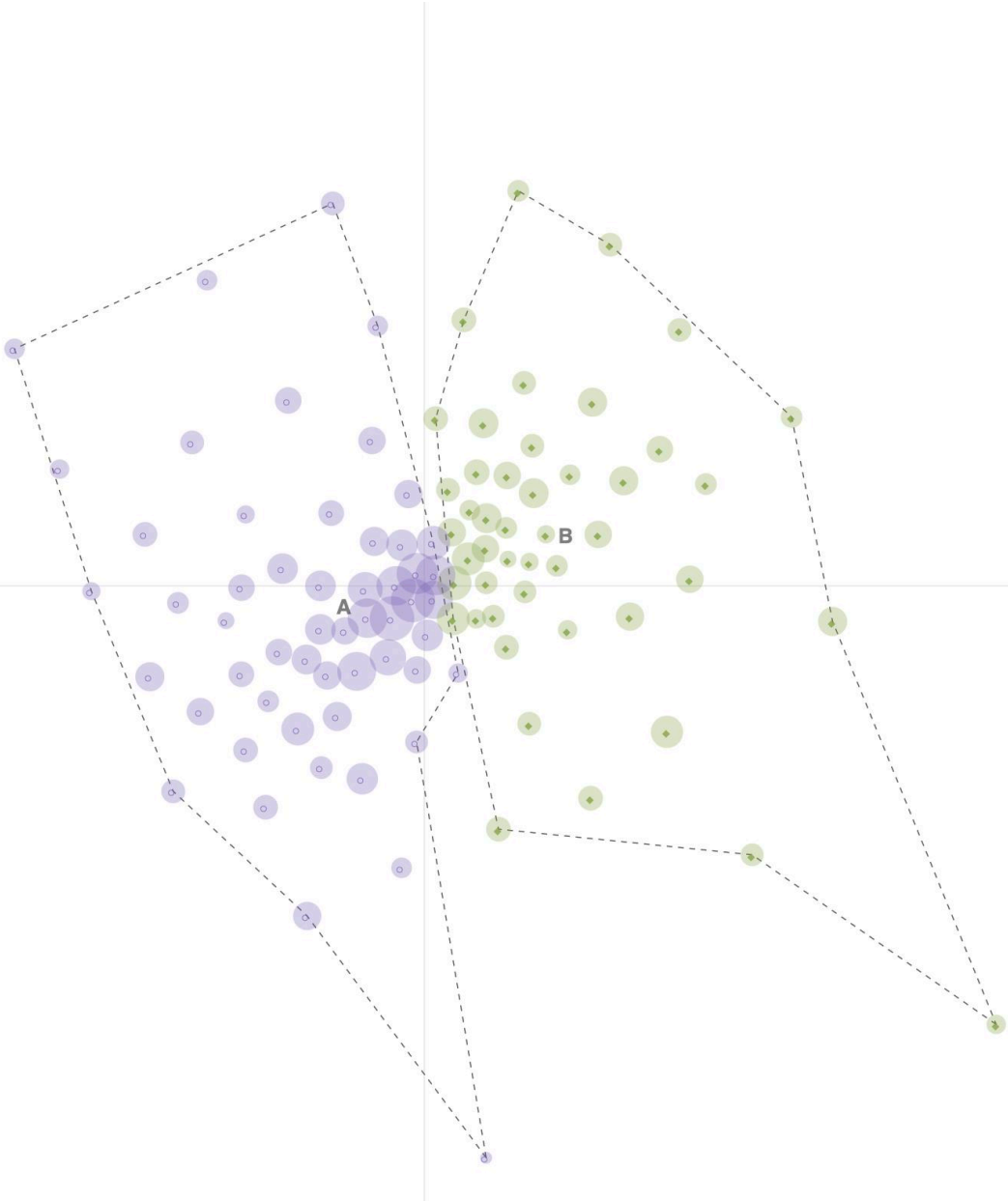
https://pol.is/report/r2xcn2cdbmrzjmmuuytdk

Polis - Bridging

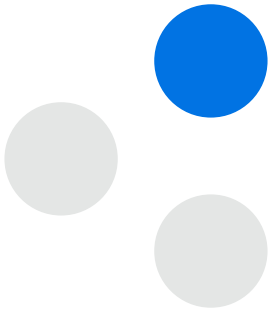
- Project voters into 2 dimensions with principle-component analysis (PCA)
- Run clustering algorithm to identify groups in population
- The bridging score of idea c is

$$\prod_{group\ G} (\text{approval rate of } c \text{ in } G)$$

- Only user information needed are votes
- Users are encouraged to build on old ideas to propose more bridging ideas



https://pol.is/report/r2xcn2cdbmrzjmmuuytdk



Remesh

Used by the United Nations in online deliberations as part of peacebuilding efforts, among other places in Yemen, Libya, and Iraq.

Remesh - Bridging

- Collect demographics or political alignment of users
- Group participants by these features, e.g.

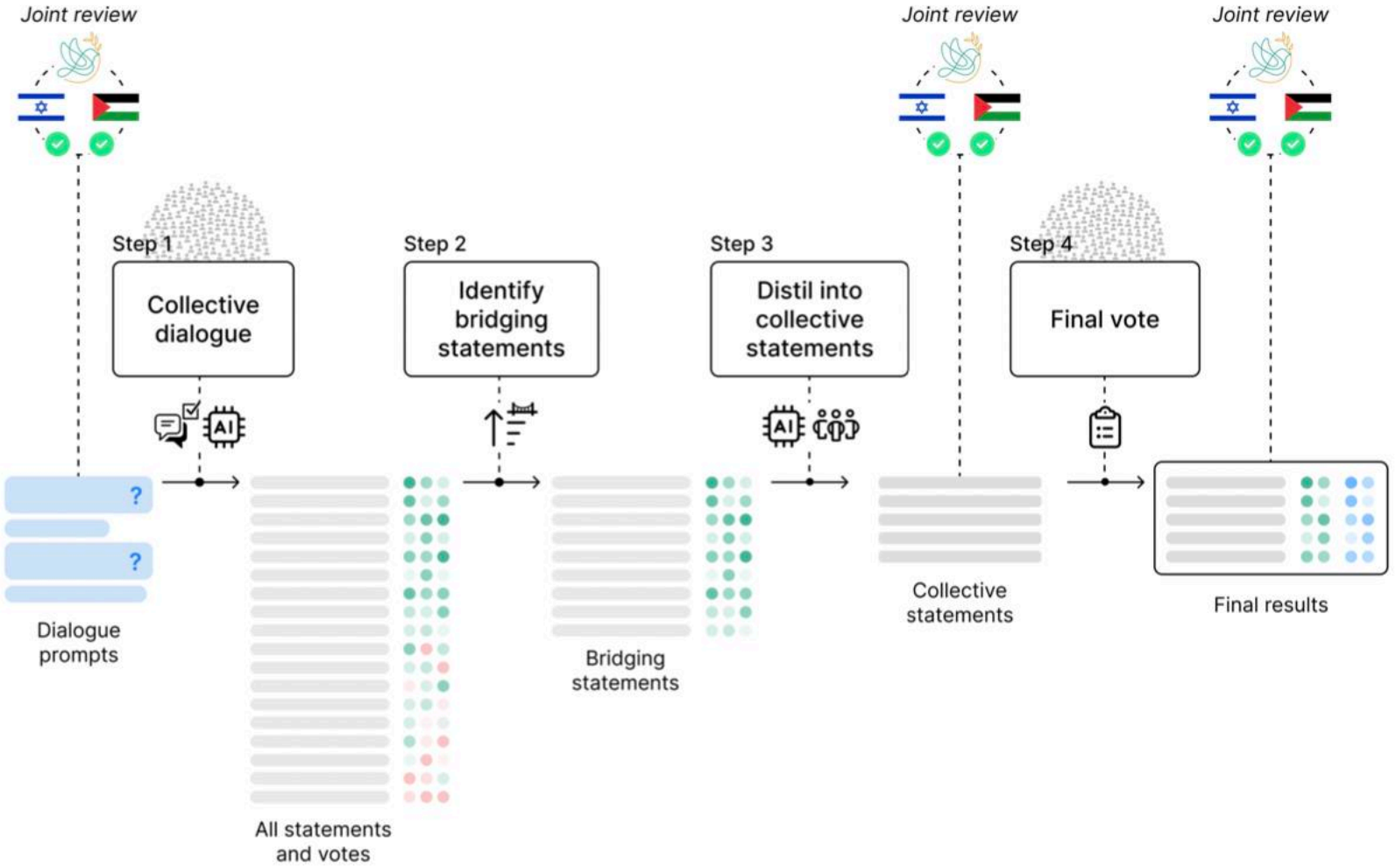
Republicans Independents Democrats

- The bridging score of idea c is

$$\min_{group\ G} (\text{approval rate of } c \text{ in } G)$$

- Use bridging comments as building blocks for collective statements

Remesh – Example Pipeline



Andrew Konya, Luke Thorburn, Wasim Almasri, Oded Adomi Leshem, Ariel D. Procaccia, Lisa Schirch, and Michiel A. Bakker. 2025. *Using Collective Dialogues and AI to Find Common Ground Between Israeli and Palestinian Peacebuilders*.

Poll

Polis: $\prod_{group\ G}(\text{approval rate in } G)$. Remesh: $\min_{group\ G}(\text{approval rate in } G)$

	Group 1: 70 people	Group 2: 30 people	Total Approval
Comment 1	100%	40%	82%
Comment 2	80%	80%	80%
Comment 3	70%	100%	79%

Poll: Rank the comments by bridging score according to Polis and Remesh.

A. Polis: 3 > 2 > 1, Remesh: 2 > 3 = 1

B. Polis: 2 > 3 = 1, Remesh: 2 > 3 > 1

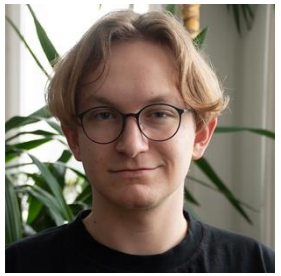
C. Polis: 2 > 3 > 1, Remesh: 1 > 2 > 3

D. Polis: 3 > 2 > 1, Remesh: 2 > 3 > 1

The Structure of Bridging



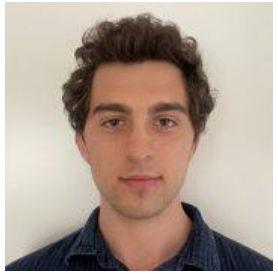
Carter
Blair
Harvard



Jakob
de Raaij
Harvard



Ariel D.
Procaccia
Harvard



Maxon
Rubin-Toles
Harvard



Nisarg
Shah
U of Toronto



Michelle
Si
Harvard



Serena
Wang
UBC

Finding Groups to Bridge

Demographic or political groups (Remesh):

- information may not be available (reliably)
- do we care about bridging demographic groups?

PCA+Cluster groups (Polis):

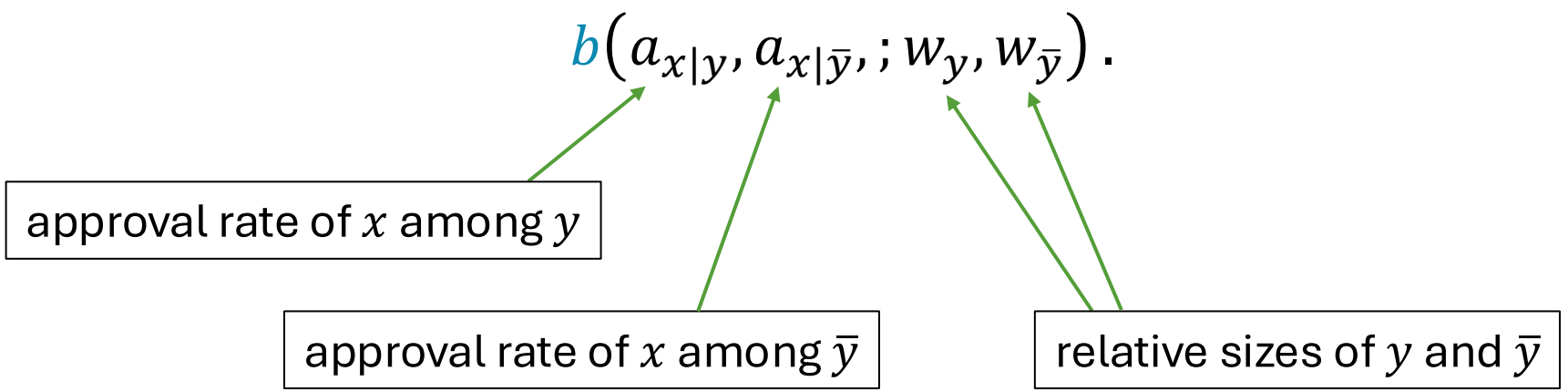
- sensitive to changes in data
- only finds one division into groups

Goal: Find a bridging function that is **robust**, considers **all opinion divides**, and **does not need auxiliary data**.

Finding Groups to Bridge

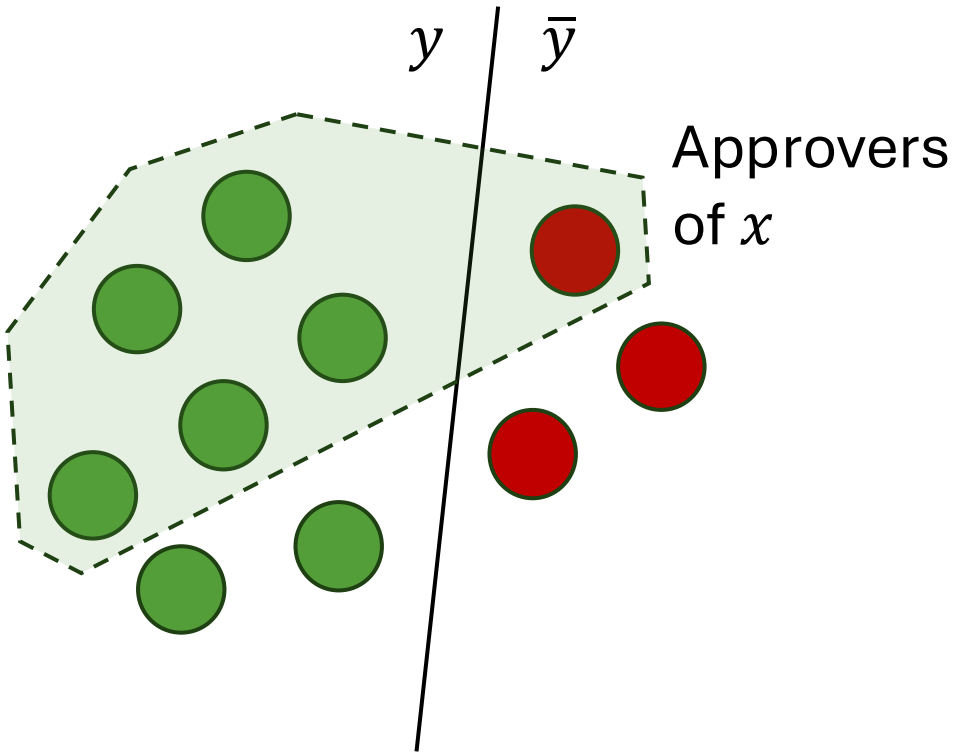
Our approach:

- each comment $y \in C$ induces two groups: Approvers y and Disapprovers $\bar{y}$
- b evaluates how well comment x bridges the groups $(y, \bar{y})$



Example

- y partitions the 10 voters into 7 approvers y and 3 disapprovers $\bar{y}$
- x has approval rate $5/7$ in group y and approval rate $1/3$ in group $\bar{y}$



$$b\left(\frac{5}{7},\frac{1}{3};\frac{7}{10},\frac{3}{10}\right)$$

approval rate of x among y

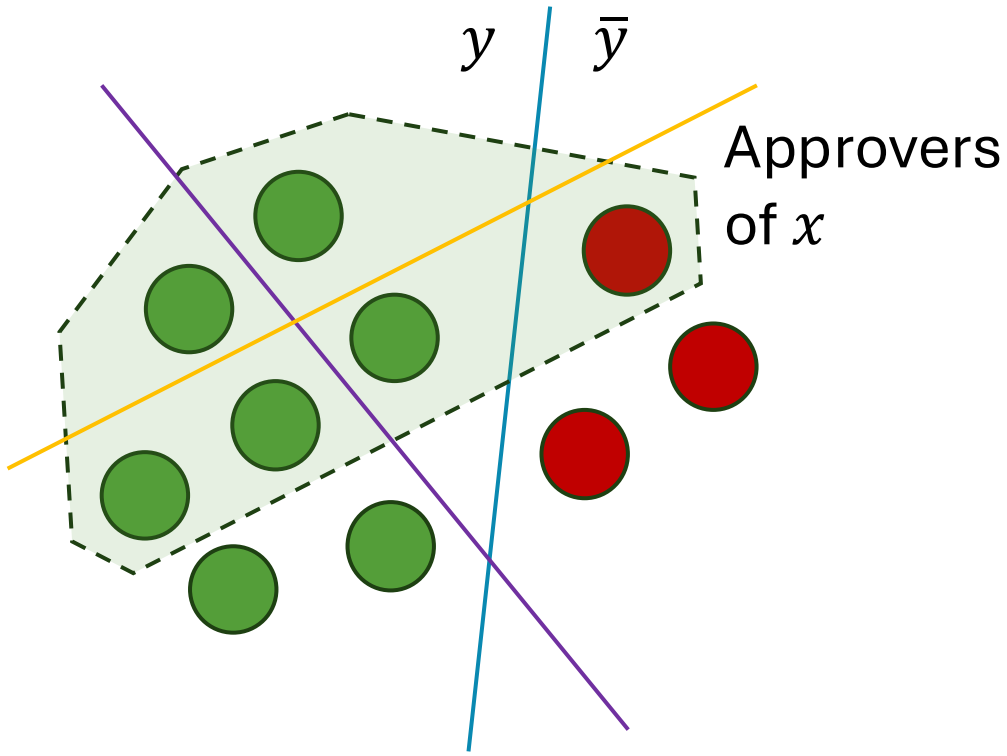
approval rate of x among $\bar{y}$

relative sizes of y and $\bar{y}$

Example

- y partitions the 10 voters into 7 approvers y and 3 disapprovers $\bar{y}$
- x has approval rate $5/7$ in group y and approval rate $1/3$ in group $\bar{y}$

$$\frac{1}{3}\left(b\left(\frac{5}{8}, 1; \frac{8}{10}, \frac{2}{10}\right) + b\left(\frac{5}{7}, \frac{1}{3}; \frac{7}{10}, \frac{3}{10}\right) + b\left(1, \frac{2}{5}; \frac{1}{2}, \frac{1}{2}\right)\right)$$



Mean-Split Bridging

Def: A *mean split bridging function* with split evaluator b is

$$\mathcal{B}(x,\mathcal{R}) = \frac{1}{|C|}\sum_{y\in C} b\left(a_{x|y}, a_{x|\bar{y}}, ; w_y, w_{\bar{y}}\right).$$

> **Theorem:** Every function $\mathcal{B}: C\times\{\mathcal{R}\} \to \mathbb{R}$ that satisfies **A1**, **A2**, **A3**, is a mean-split bridging function.

A1 Election Additivity: Let $\mathcal{R}_1 = (N, C_1), \mathcal{R}_2 = (N, C_2)$. Let $\mathcal{R} = (N, C_1 \cup C_2)$.
Then, $\mathcal{B}(x,\mathcal{R}) = \frac{|C_1|}{|C_1|+|C_2|}\cdot\mathcal{B}(x,\mathcal{R}_1) + \frac{|C_2|}{|C_1|+|C_2|}\cdot\mathcal{B}(x,\mathcal{R}_2)$

A2 Anonymity: Permuting the voters doesn't change $\mathcal{B}(x,\mathcal{R})$

A3 Homogeneity: Cloning every voter $k \in \mathbb{N}^+$ times doesn't change $\mathcal{B}(x,\mathcal{R})$

What b?

- Given two groups of size w and $1 - w$ in which an ideas has respective approval rates a_1 and a_2, how bridging is it?

Remesh

$$b^{Remesh}(a_1, a_2, w) = \min(a_1, a_2)$$

Polis

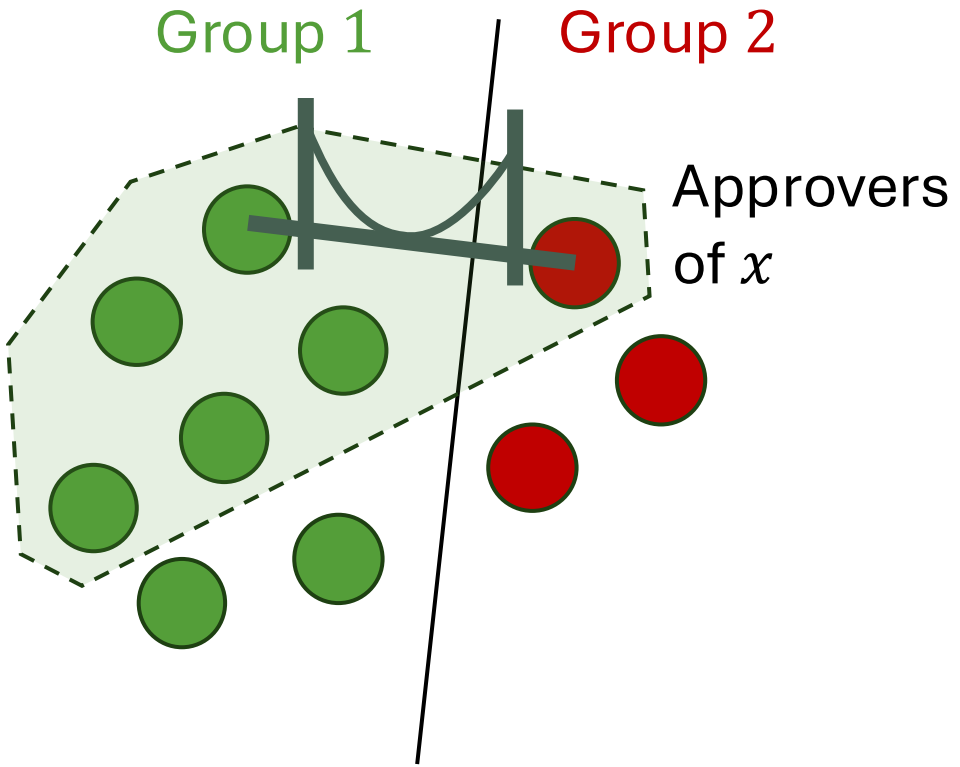
$$b^{Polis}(a_1, a_2, w) = a_1 a_2$$

What b?

- Given two groups of size w and $1-w$ in which an ideas has respective approval rates a_1 and a_2, how bridging is it?
- A **bridge** is a pair of voters (i, j) that disagree on y but both approve x

$$\#\text{bridges} = (a_1 \cdot w \cdot |N|) \cdot (a_2 \cdot (1-w) \cdot |N|)$$
$$= \left(\frac{5}{7} \cdot \frac{7}{10} \cdot 10\right) \cdot \left(\frac{1}{3} \cdot \frac{3}{10} \cdot 10\right) = 5$$

$$b^{PD}(a_1, a_2, w) = 4 \cdot a_1 \cdot w \cdot a_2 \cdot (1-w)$$
$$= \#\text{bridges} \cdot \frac{4}{|N|^2}$$

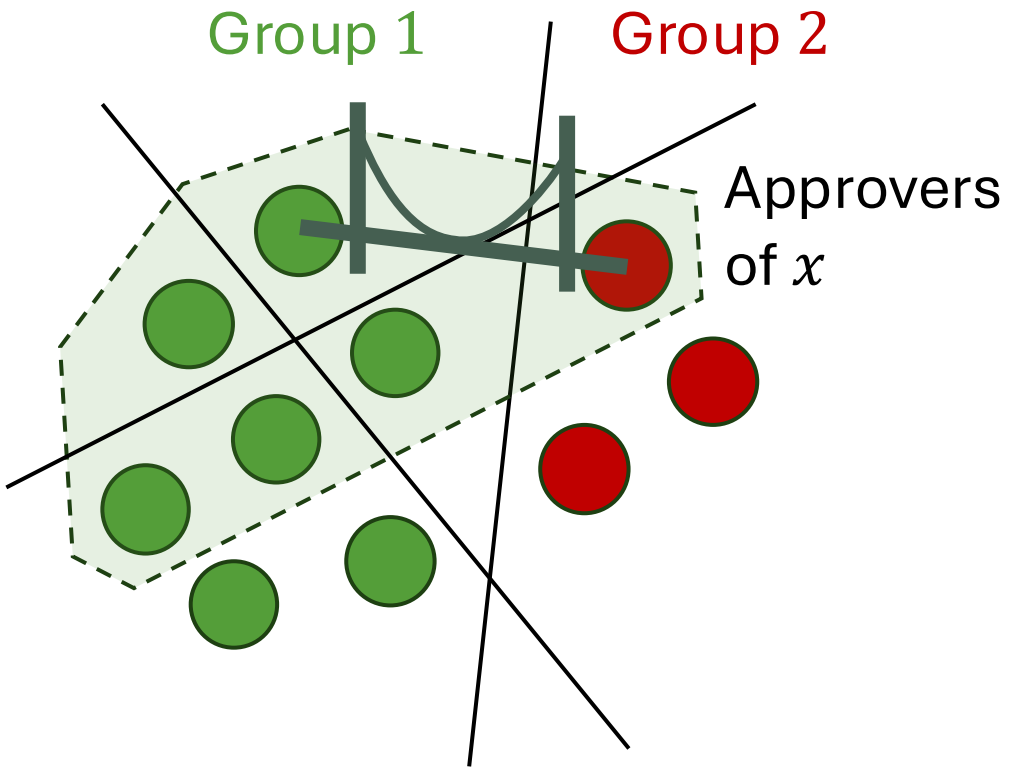


What b?

The pairwise disagreement bridging function is

$$\mathcal{B}(x,\mathcal{R}) = \frac{4}{|C|}\sum_{y\in C} a_{x|y} w_y a_{x|\bar{y}} w_{\bar{y}}$$

$$= \frac{4}{|C|}\sum_{y\in C} \#(\text{bridges across } y)$$



What b?

The pairwise disagreement bridging function is

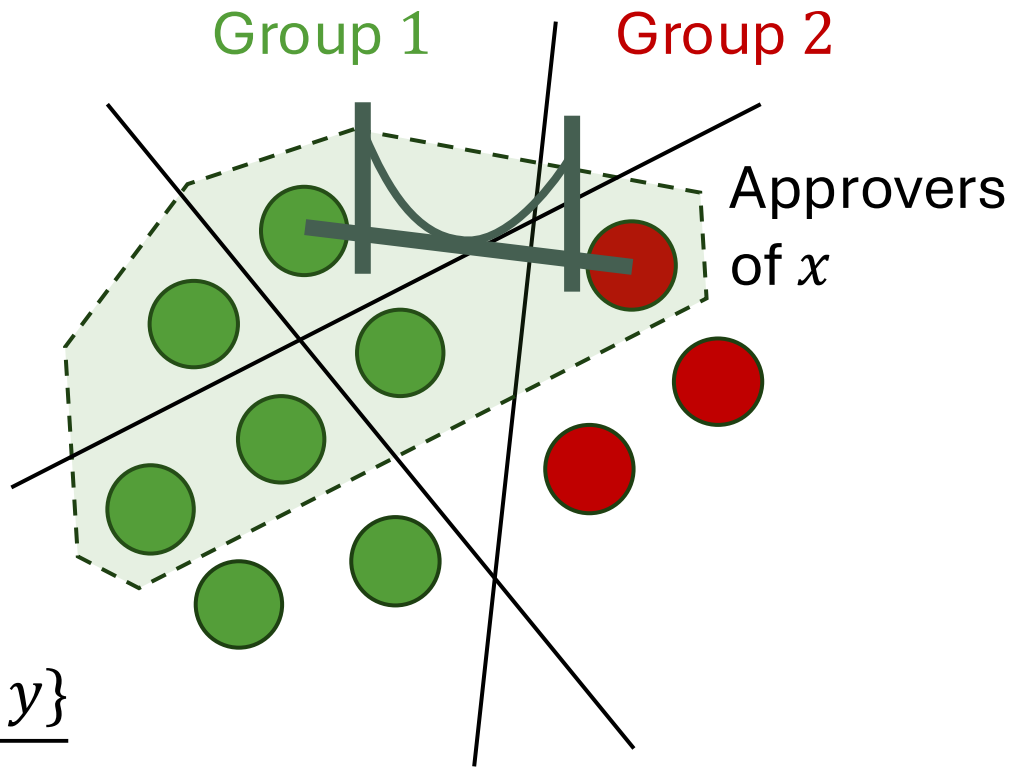
$$\mathcal{B}(x,\mathcal{R}) = \frac{4}{|C|}\sum_{y\in C} a_{x|y} w_y a_{x|\bar{y}} w_{\bar{y}}$$

$$= \frac{4}{|C|}\sum_{y\in C} \#(\text{bridges across } y)$$

$$= \frac{4}{|C|}\sum_{i,j\in N} \#(\text{bridges between } i,j)$$

$$= 4 \sum_{i,j \text{ approve } x} \frac{|\{y \in C : i,j \text{ disagree on } y\}}{|C|}$$

$$= 4 \sum_{i,j \text{ approve } x} d_{i,j}$$



Pairwise Disagreement

The pairwise disagreement bridging function is

$$\mathcal{B}(x,\mathcal{R}) = \frac{1}{|C|}\sum_{y\in C} 4 \cdot a_{x|y} w_y a_{x|\bar{y}} w_{\bar{y}} = 4 \sum_{i,j \text{ approve } x} d_{i,j}$$

- **Robust:** A single vote changing has small impact on $\mathcal{B}$
- **Does not need auxiliary data:** uses approval votes only
- **All opinion divides:** finds ideas that bridge many opinion divides at once

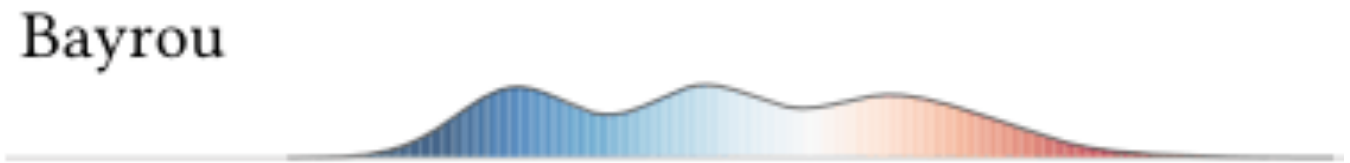
French Presidential Election 2007

- 2836 voters in Cigne,´Louvigny, Illkirch

- 12 candidates, including:
  - Sarkozy, a conservative
  - Royal, a socialist
  - Bayrou, a moderate liberal

How to Read the Ridgeline Plots

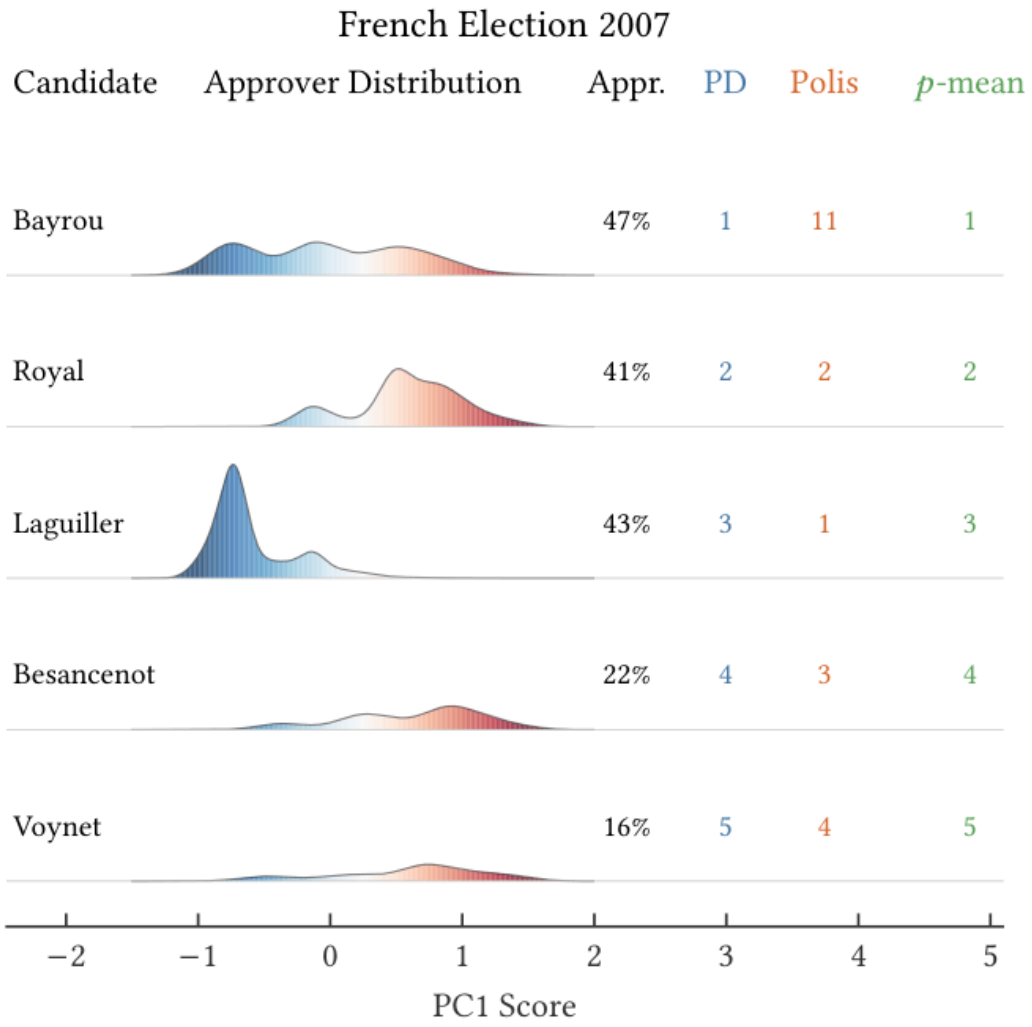
- Place every voter on one axis: **the first principal component** of the vote matrix
- Voters who approve similar candidates end up close together; the two ends are the main opposing groups (**blue** / **red**)
- For each candidate, plot **where its approvers sit** on that axis
- Narrow curve = support from one side. **Wide curve = support across the divide**



French Presidential Election 2007

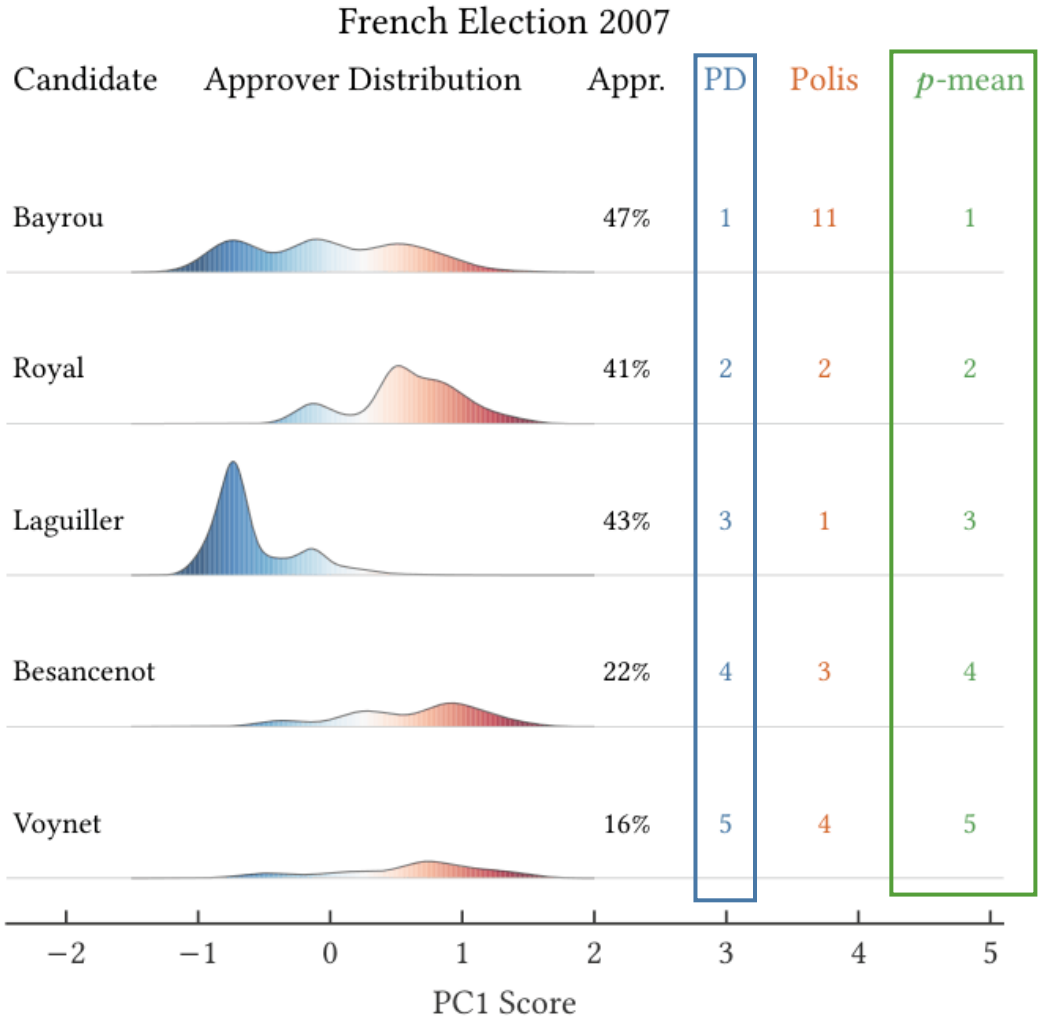
- 2836 voters in Cigne,´Louvigny, Illkirch

- 12 candidates, including:
  - Sarkozy, a conservative
  - Royal, a socialist
  - Bayrou, a moderate liberal



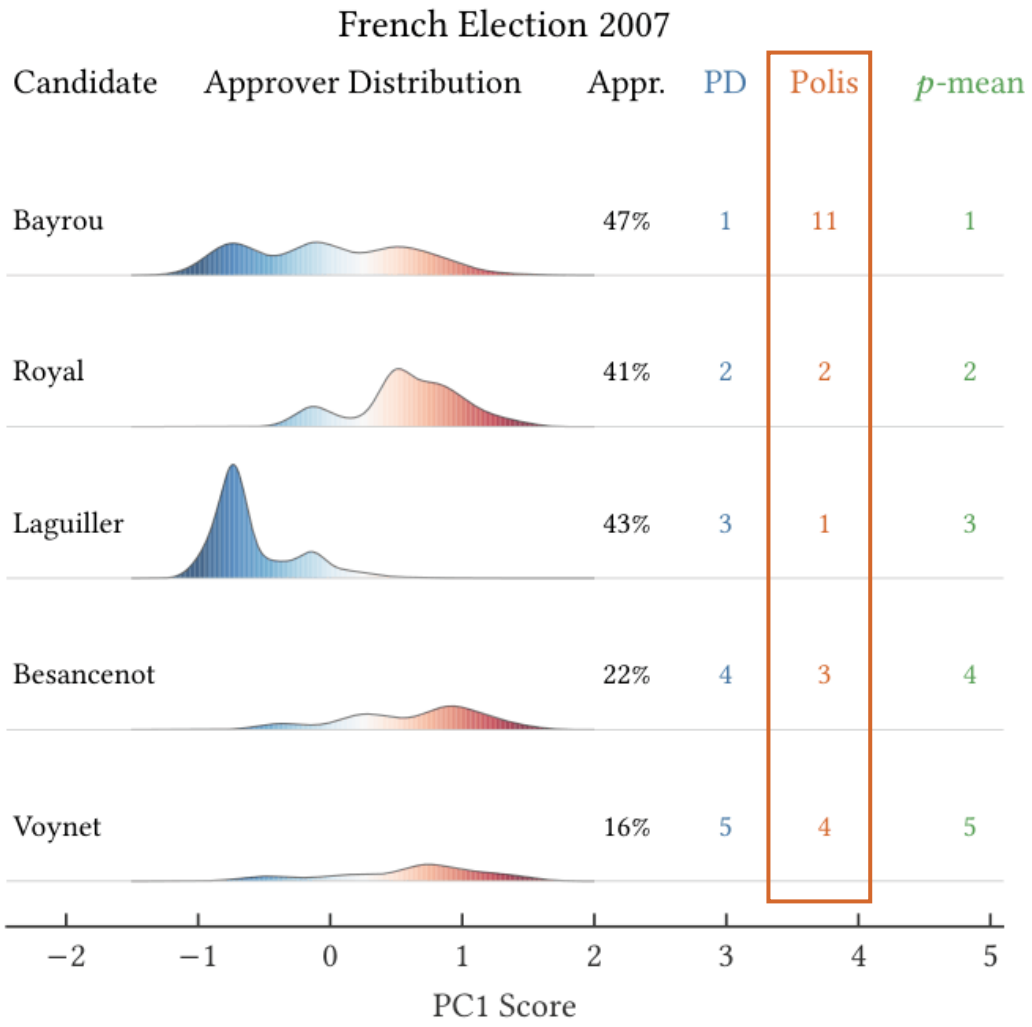
French Presidential Election 2007

- 2836 voters in Cigne,´Louvigny, Illkirch
- 12 candidates, including:
  - Sarkozy, a conservative
  - Royal, a socialist
  - Bayrou, a moderate liberal
- PD & p-mean find the centrist

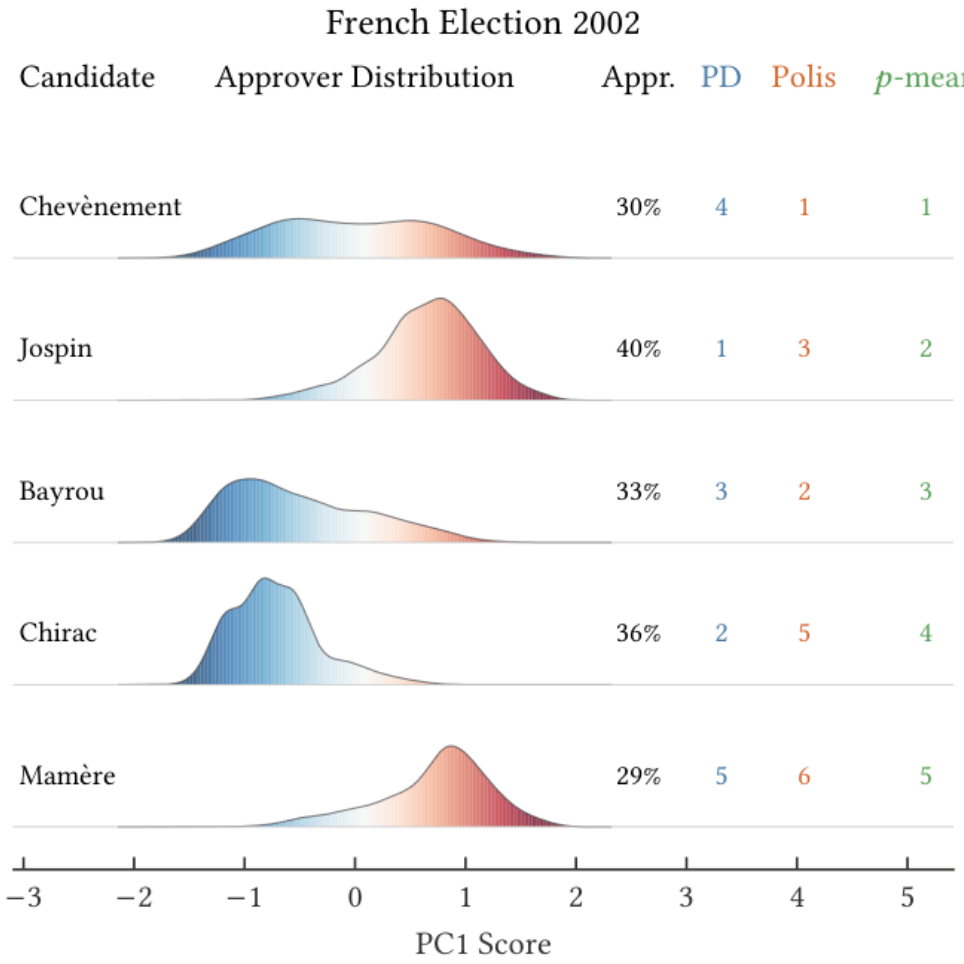


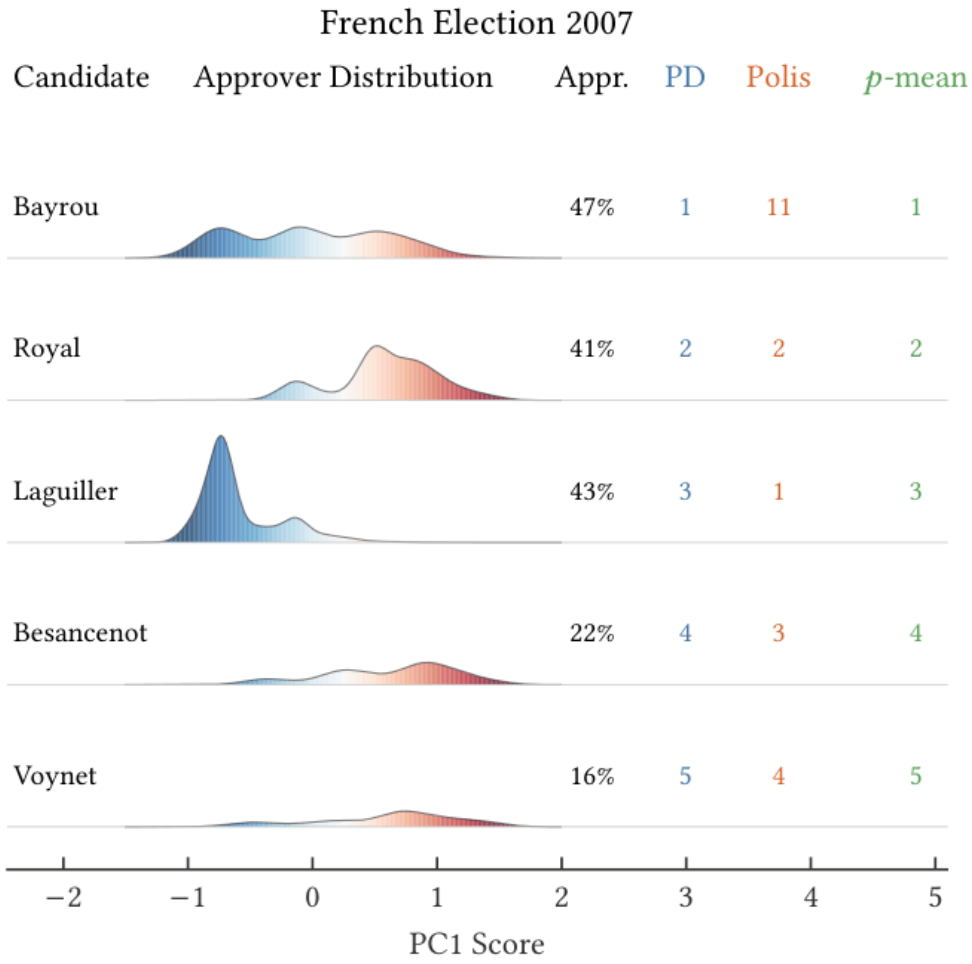
French Presidential Election 2007

- 2836 voters in Cigne,´Louvigny, Illkirch
- 12 candidates, including:
  - Sarkozy, a conservative
  - Royal, a socialist
  - Bayrou, a moderate liberal
- PD & p-mean find the centrist
- Polis fails



French Presidential Election 2007





Robustness to Incomplete Votes

- not all voters will see all alternatives → incomplete votes
- PD and p-mean can be estimated well

Robustness to Incomplete Votes

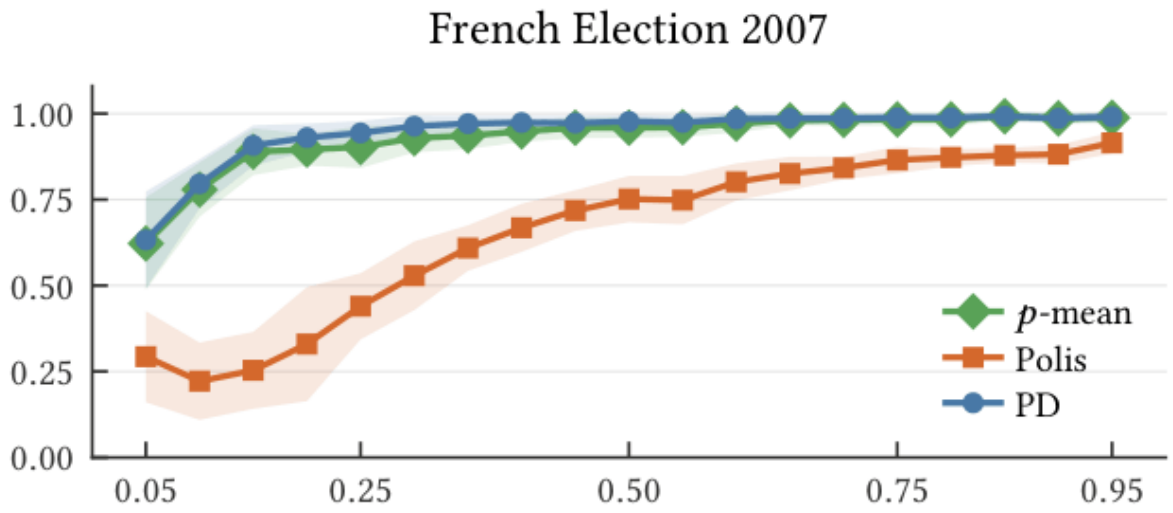
- not all voters will see all alternatives → incomplete votes
- PD and p-mean can be estimated well

- Assume random q-fraction of votes are observed:

Robustness to Incomplete Votes

- not all voters will see all alternatives → incomplete votes
- PD and p-mean can be estimated well

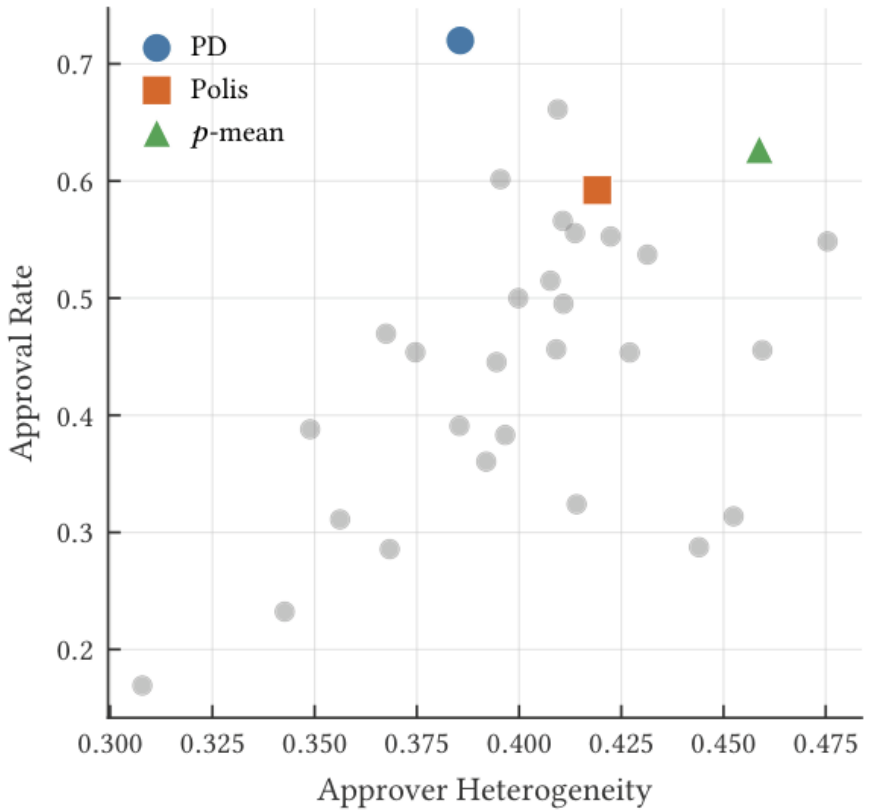
- Assume random q-fraction of votes are observed:



Online Deliberation

- Topic: Introduce $15 minimum wage in Seattle
- 339 voters, 54 comments
- Incomplete: ~16% of votes known

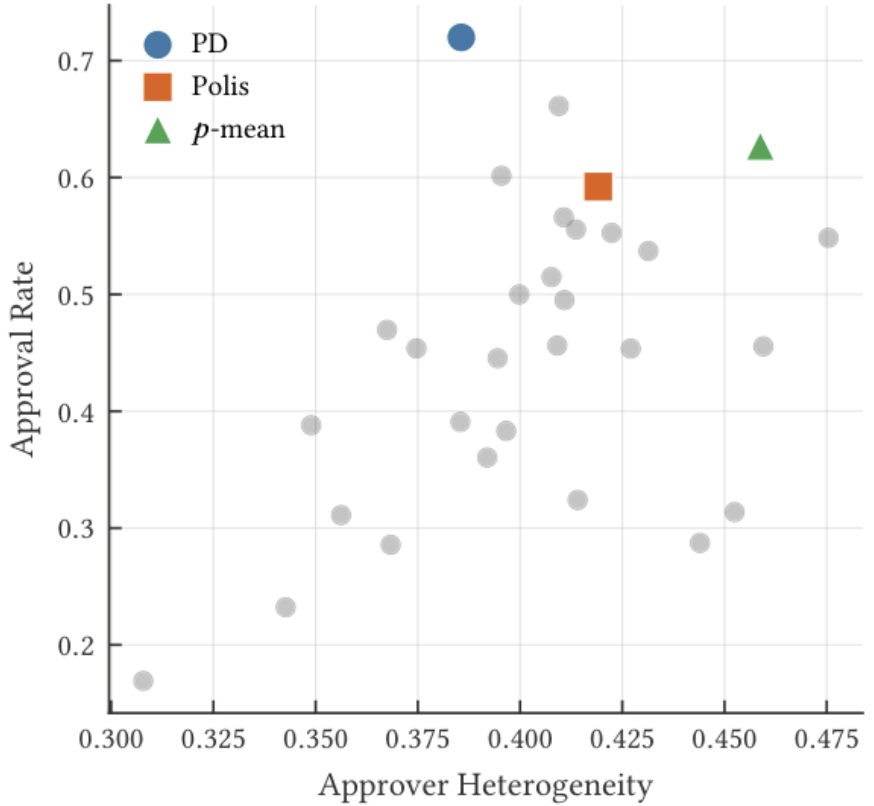
- Polis Pareto-dominated by p-mean
- PD more strongly correlated with approval rate



[…] Why should we subsidize businesses that depend on inadequate wages? If a business's product is good, then this shouldn't be a challenge.

Something needs to be done to address income inequality and this is a good if imperfect start

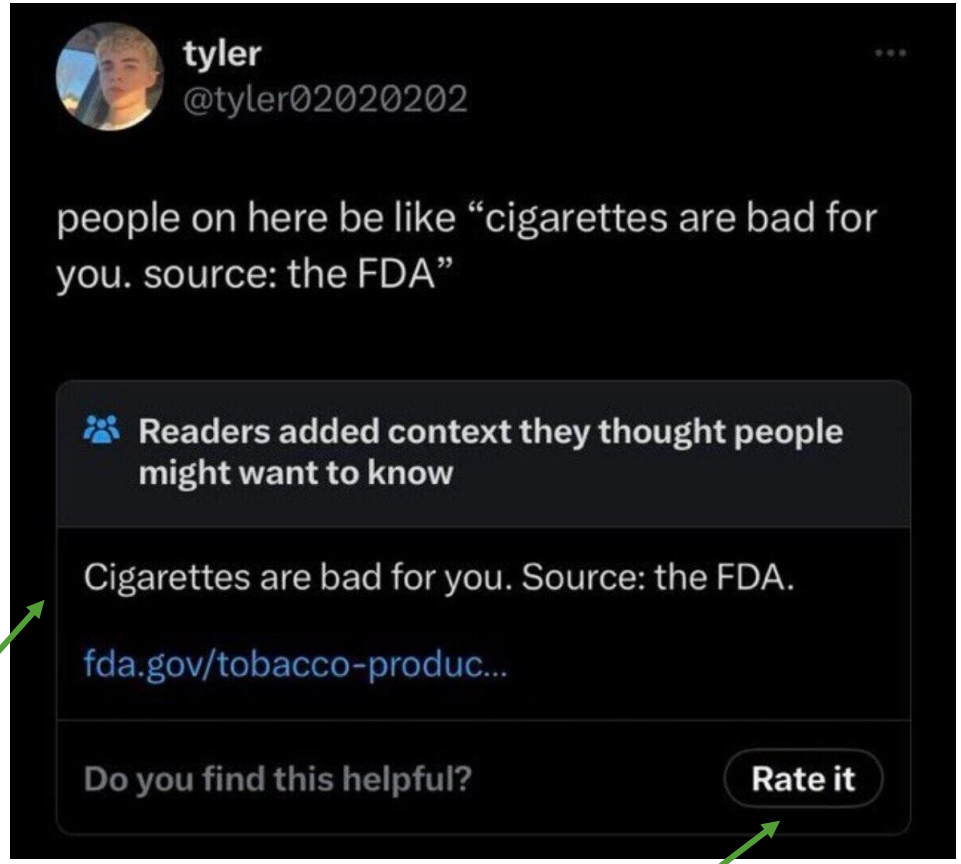
At the very least, customers across the board are going to have to pay more […] I worry that customers will start spending less […]



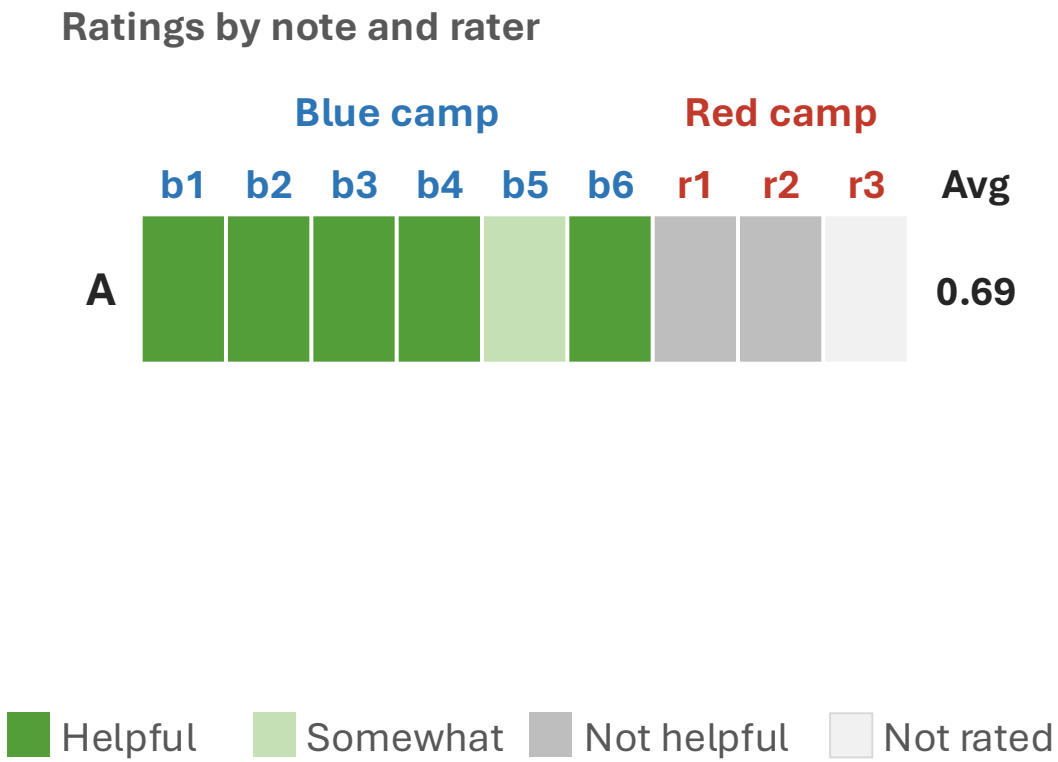
Community Notes

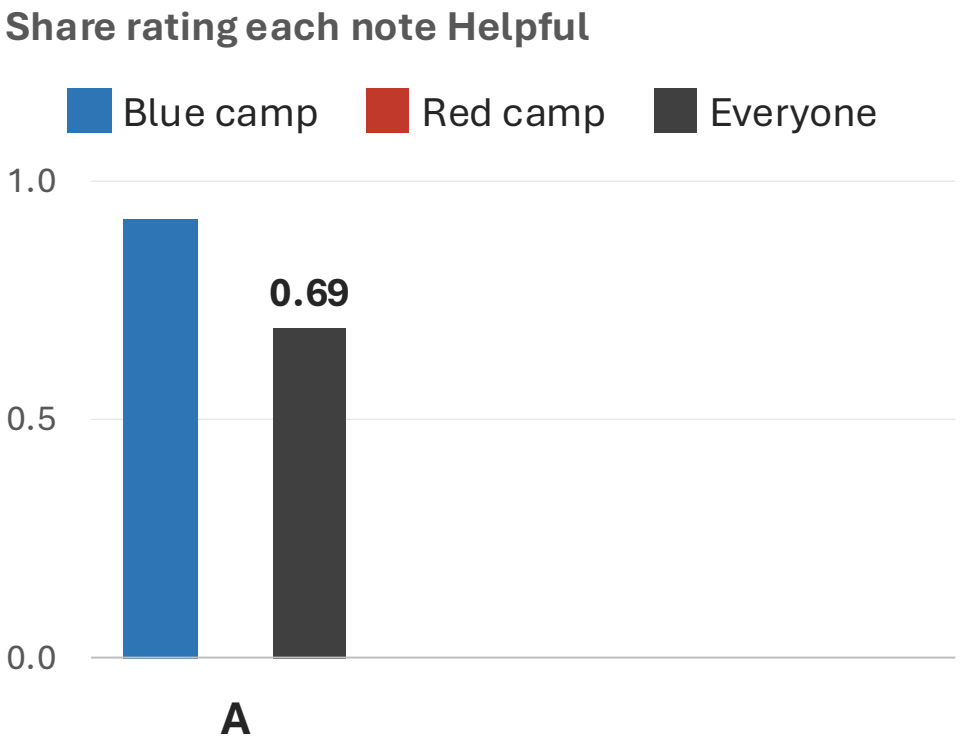
Community Notes: Crowd Fact-Checking with a Polarized Crowd

- Anyone enrolled can write a note on a post; anyone enrolled can rate notes helpful / somewhat / not helpful.
- Notes rated Helpful are shown publicly under the post.
- Raters are polarized: errors are correlated within camps.

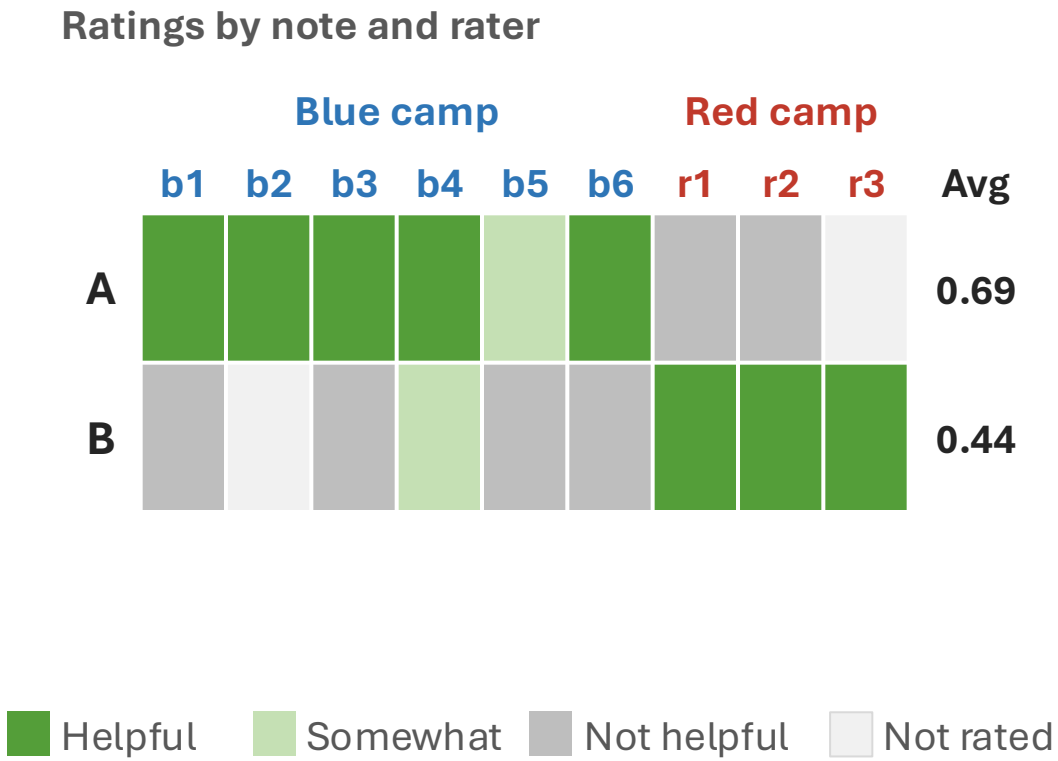


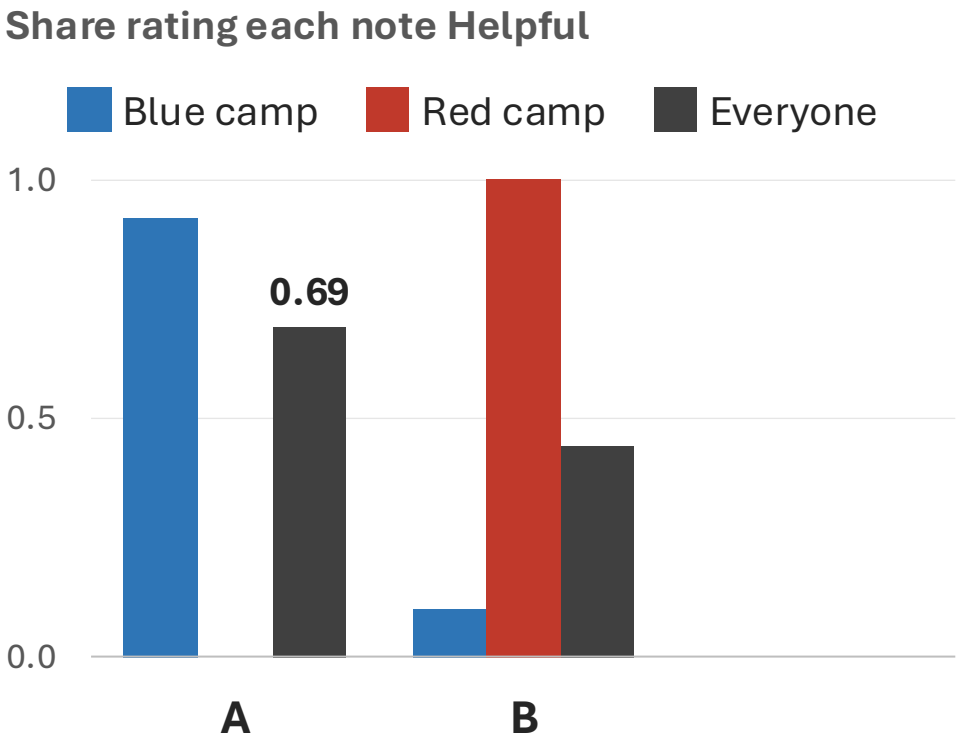
Raw Approval Rewards the Bigger Camp



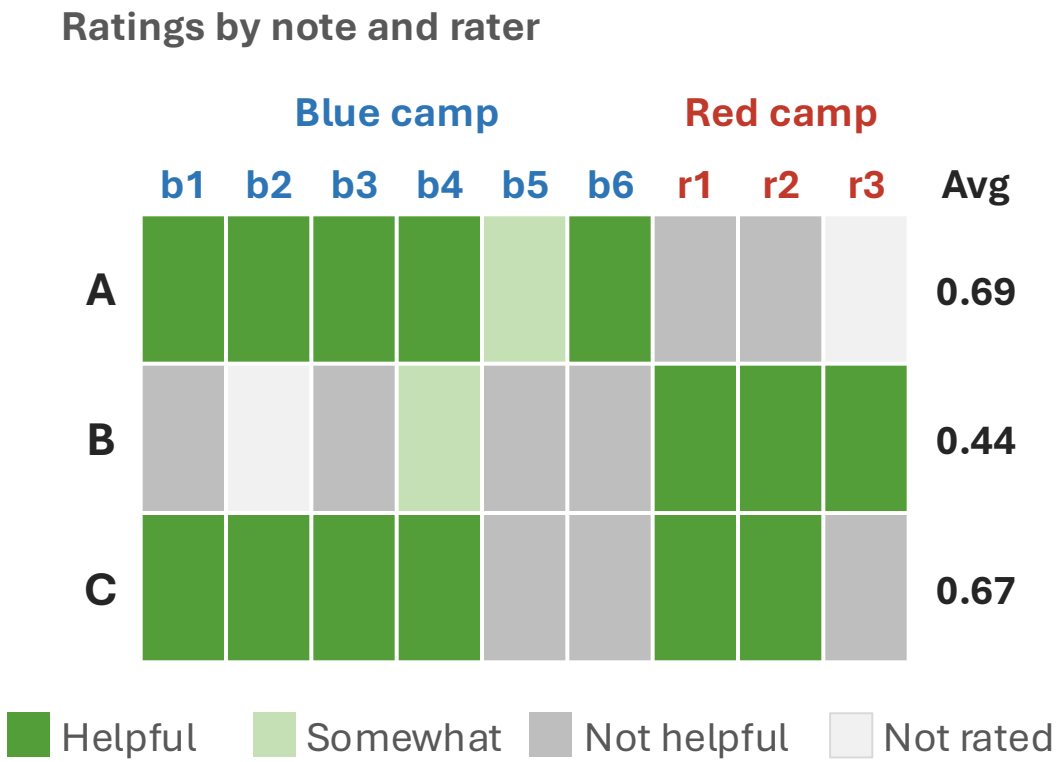


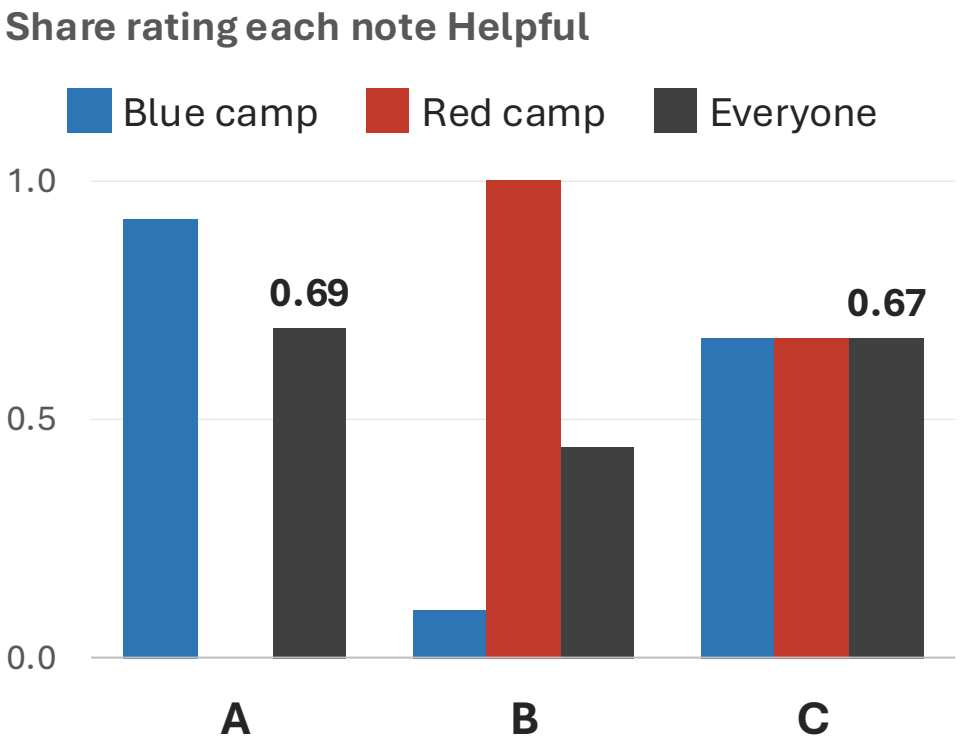
Raw Approval Rewards the Bigger Camp





Raw Approval Rewards the Bigger Camp





A outscores C on average (0.69 vs. 0.67), **but only C is liked by both camps.**

Borrow a Recommender System

$$\hat{r}_{un} = \mu + i_u + \boldsymbol{i_n} + f_u \cdot f_n$$

u , n	a rater u and a note n
r_{un} , $\hat{r}_{un}$	the rating (1, 0.5, 0) and its prediction
μ	global mean rating
i_u	rater leniency
f_u , f_n	rater and note positions on a learned one-dimensional axis
$\boldsymbol{i_n}$	**the note's intercept, its helpfulness after viewpoint is accounted for. This is the score.**

Loss: $\Sigma\,(r_{un} - \hat{r}_{un})^2 + \lambda_i\,(i_u^2 + i_n^2 + \mu^2) + \lambda_f(\|f_u\|^2 + \|f_n\|^2),\ \ \lambda_i = 0.15,\ \lambda_f = 0.03$

Simplifying the Model

- To build intuition, let's drop μ and i_u
- That leaves two terms: $\hat{r}_{un} \approx \boldsymbol{i_n} + \boldsymbol{f_u} \cdot \boldsymbol{f_n}$
  - $\boldsymbol{i_n}$: how good the note is, apart from viewpoint
  - $\boldsymbol{f_u} \cdot \boldsymbol{f_n}$: how well the rater and note line up

Placing Raters and Notes

$$\hat{r}_{un} \approx i_n + \boldsymbol{f}_u \cdot \boldsymbol{f}_n$$

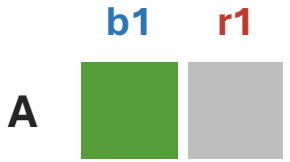
Ratings

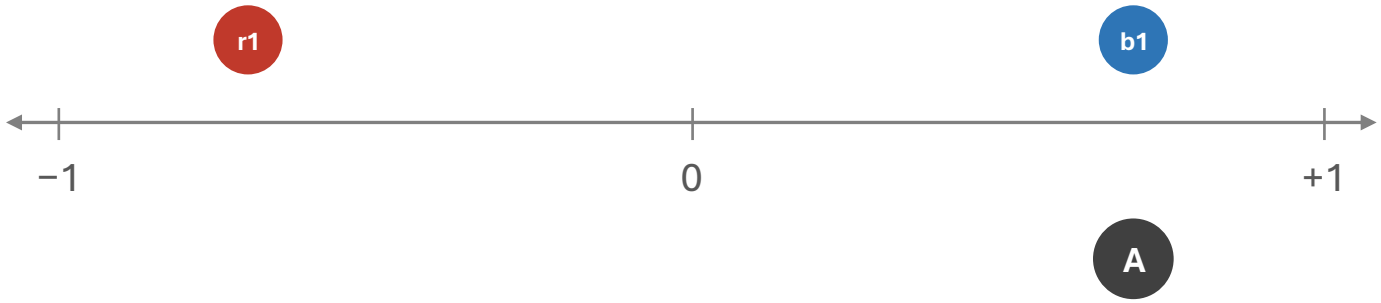


Placing Raters and Notes

$$\hat{r}_{un} \approx i_n + \boldsymbol{f}_u \cdot \boldsymbol{f}_n$$

Ratings

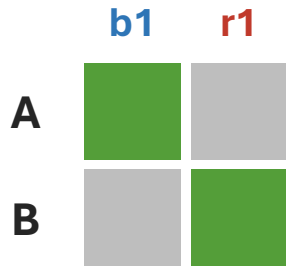


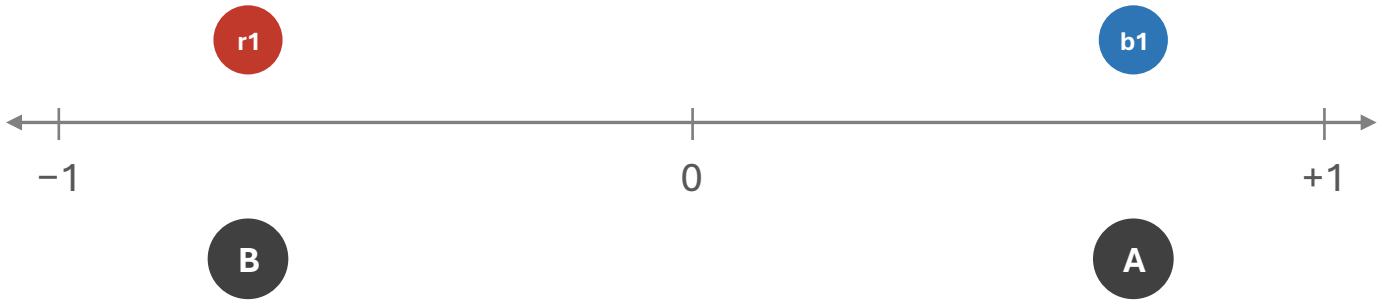


Placing Raters and Notes

$$\hat{r}_{un} \approx i_n + f_u \cdot f_n$$

Ratings

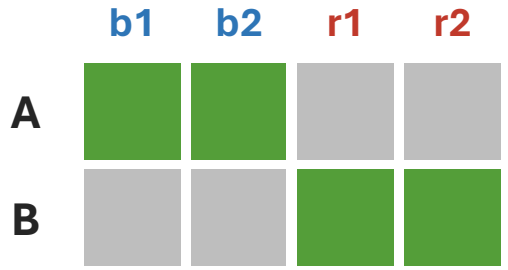


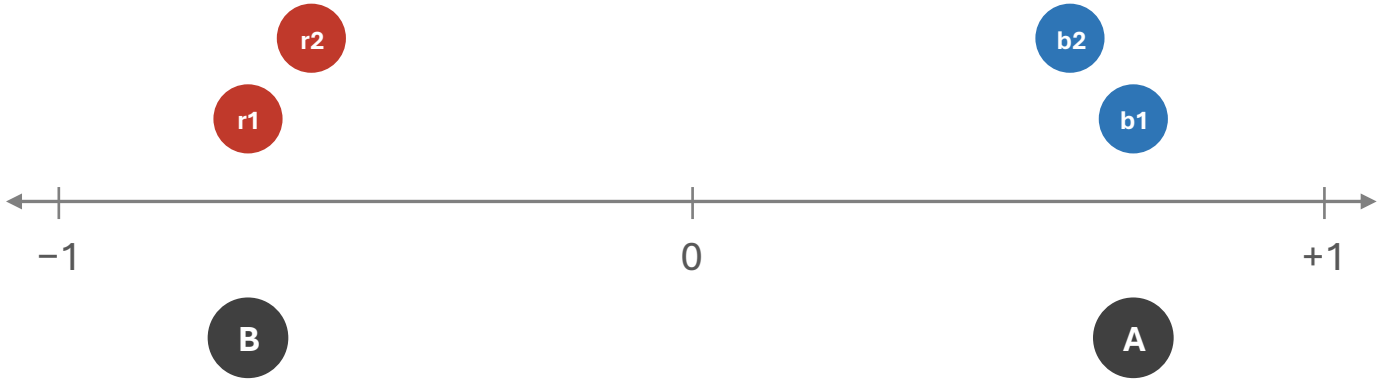


Placing Raters and Notes

$$\hat{r}_{un} \approx i_n + f_u \cdot f_n$$

Ratings

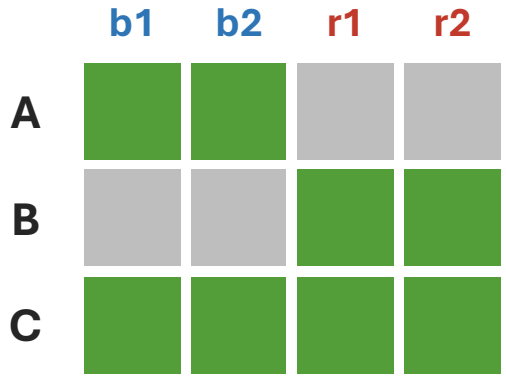


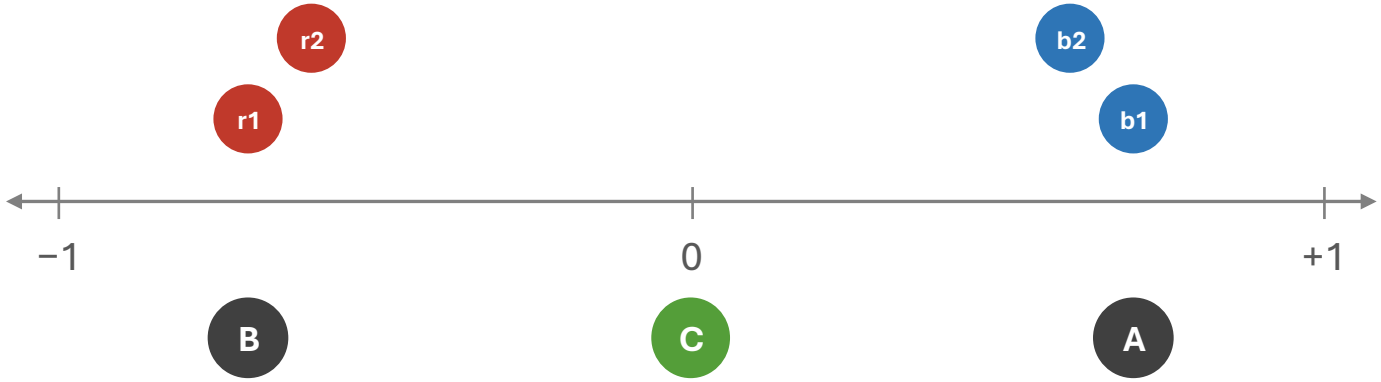


Placing Raters and Notes

$$\hat{r}_{un} \approx i_n + f_u \cdot f_n$$

Ratings

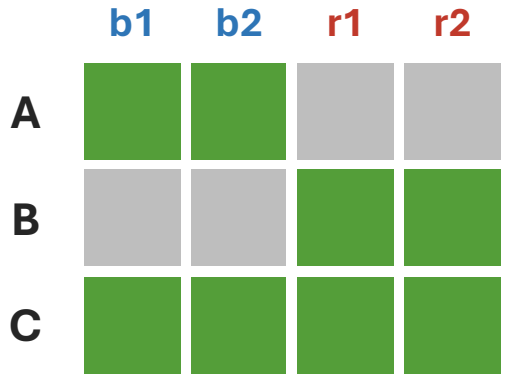


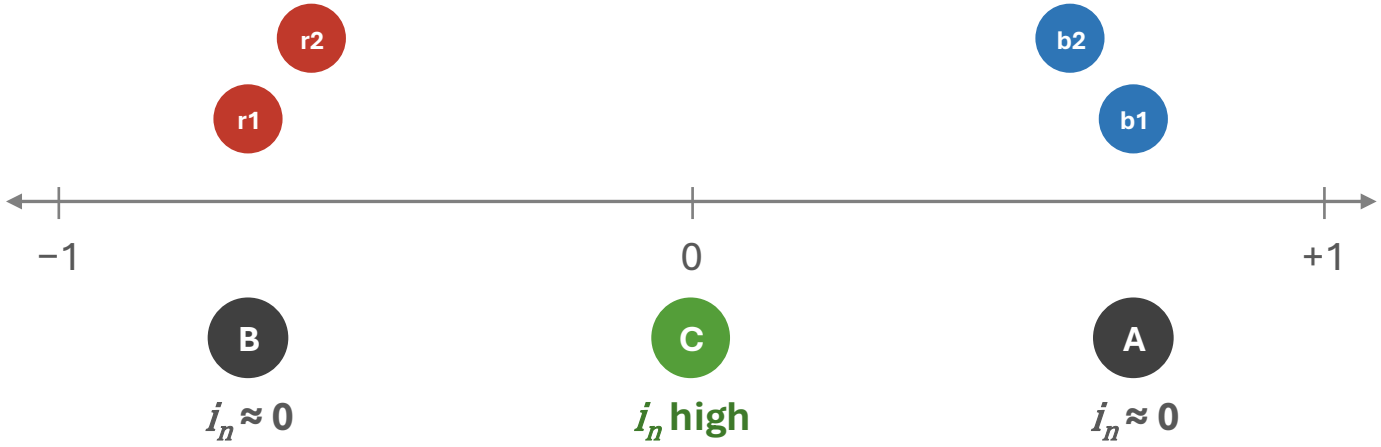


Adding the Note Intercept

$$\hat{r}_{un} \approx i_n + f_u \cdot f_n$$

Ratings



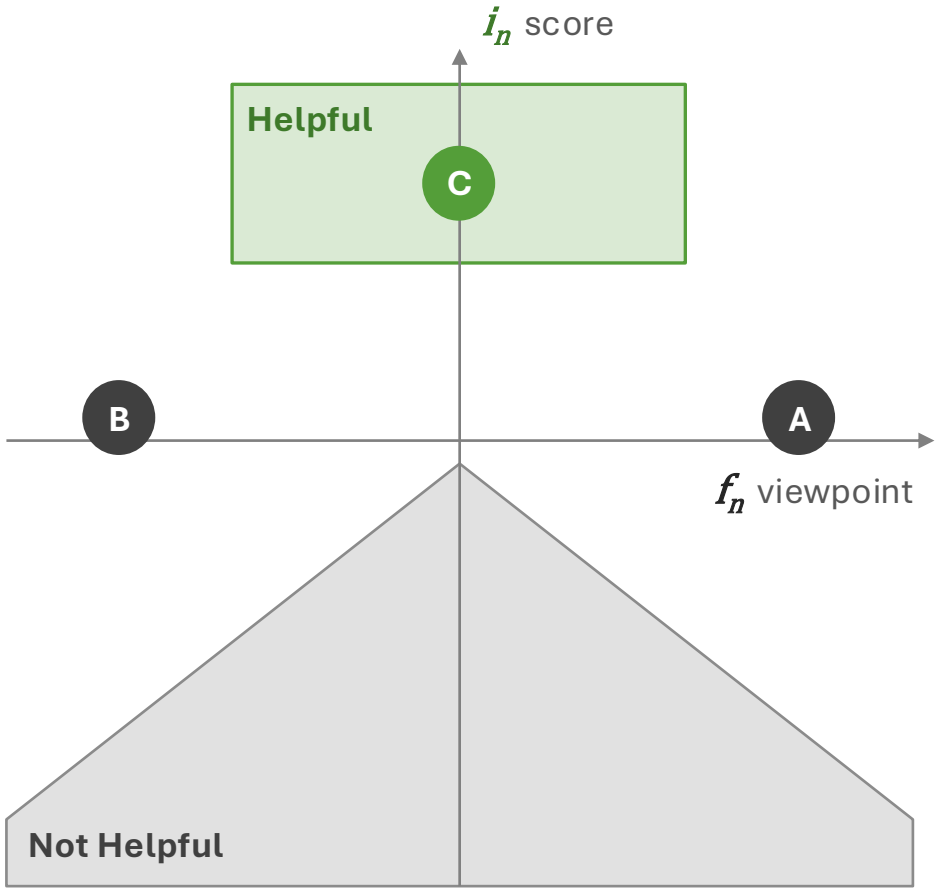


Bridging in Community Notes

- A note scores well only if **people who usually disagree** both find it helpful
  - "Usually disagree" means far apart on the learned axis
- The score i_n is the support left over once viewpoint is explained
  - Support from one camp is absorbed by $f_u \cdot f_n$
  - Only cross-camp support raises i_n

Status Depends on Where It Lands

- Each note gets a viewpoint f_n and a score i_n
- **Helpful:** high score and not lopsided
- **Not Helpful:** below the V
  - Near the center, mild dislike from both sides is enough
  - Far out, a note needs much more dislike: its Not Helpful votes come mostly from the other camp, and one camp can't bury the other's notes
- Everything else: **Needs More Ratings**

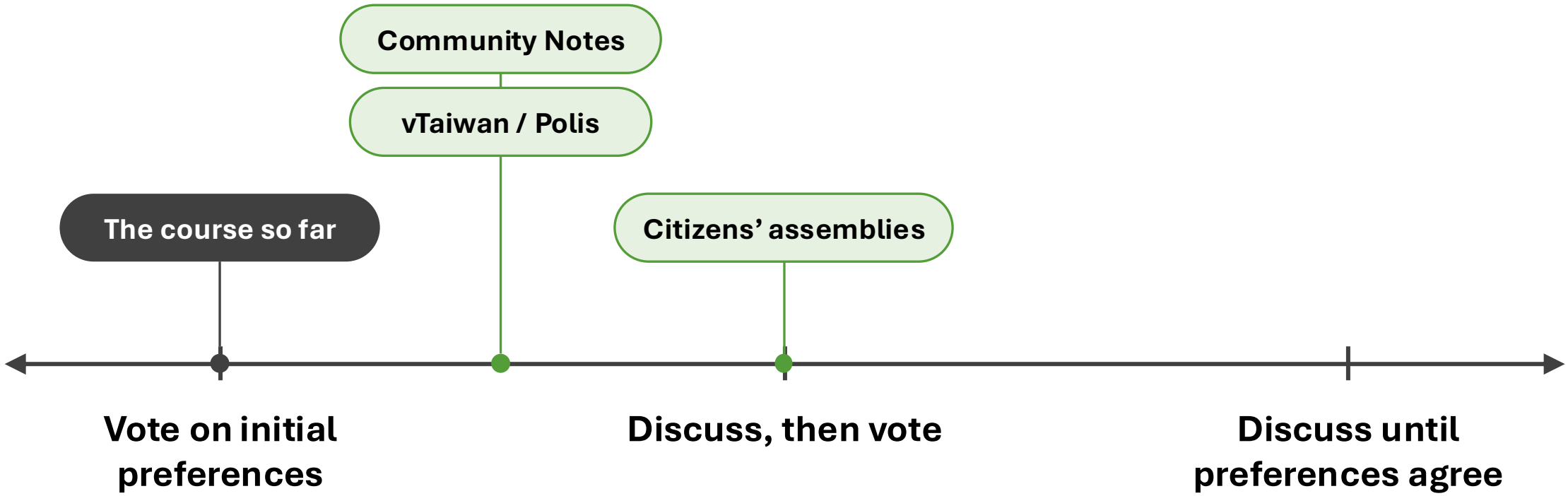


Who Counts as “Otherwise Disagreeing”?

- **Polis / Remesh:** one partition, fixed in advance (clustering, or declared groups)
- **Our paper:** average over every partition induced by another comment’s approvers and disapprovers
- **Community Notes:** a learned continuous axis (no partition at all). Bridging is the residual after projecting out the axis

Aggregative vs. Deliberative Democracy

Bridging is an aggregation rule that serves a deliberative aim



A true claim one side rejects is never surfaced. Which objective should a fact-checking system have?

Wrap-Up

- Social choice usually takes preferences as given, **deliberation thinks about how they are formed**
- Discussion can line preferences up on one dimension, which **helps majority rule**
- It doesn’t scale past ~100 people, so online platforms sacrifice some discussion and rely more on aggregation
- Bridging rules (Polis, our paper, Community Notes) reward **support across the main divide**