The Efficacy of Different Analysis Algorithms for Summarizing Online Deliberations

by

Naveen Venkat

B.S. Computer Science and Computer Engineering, MIT, 2023

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2024

Authored by: Naveen Venkat
Department of Electrical Engineering and Computer Science
September 6, 2024

Certified by: Lily L. Tsai
Ford Professor of Political Science, Thesis Supervisor

Accepted by: Katrina LaCurts
Chair, Master of Engineering Thesis Committee

The Efficacy of Different Analysis Algorithms for Summarizing Online Deliberations

by

Naveen Venkat

Submitted to the Department of Electrical Engineering and Computer Science on September 6, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

ABSTRACT

For the past decade, online deliberation platforms like *Polis* have expanded the reach of deliberative democracy, which calls for political decisions to be based on the results of fair and balanced discussions among citizens, by enabling larger deliberations. However, as these discussions often generate a large volume of comments, which is infeasible for policy-makers to thoroughly review, these platforms often include analysis algorithms that distill the conversation into a small set of comments, which policy-makers can use as the base of citizen input into decision-making. While *Polis* currently provides a clustering-analysis summary of the discussion, two newer aggregation algorithms, inspired by computational social choice theory and abstract argumentation theory, have recently been proposed. These algorithms seek to provide more representative (i.e. portraying all perspectives) and consistent (i.e. comments within a perspective do not oppose each other) summaries of the discussion, respectively. Still, though these newer algorithms may have theoretical advantages over *Polis*'s current methods, they have yet to be evaluated in a real-world application. Through a randomized controlled trial of all three approaches using a nationally representative sample, we compare their practical effectiveness, as measured by participants' subjective experiences regarding how well these summaries represent their concerns. We find that the computational social choice-inspired algorithm consistently outperforms *Polis*'s current methods in this regard, though future theoretical work is still needed to fully adapt this approach to a real-world setting.

Thesis supervisor: Lily L. Tsai
Title: Ford Professor of Political Science

Acknowledgments

First, I would like to thank my supervisor, Lily Tsai, for fostering my interest in deliberative democracy, and for allowing me to explore this project at the intersection of political science and theoretical computer science. I could not have completed this thesis without her help and support.

Next, I would like to thank my primary advisor, Lula Chen, for her assistance in developing and implementing this project. Her generous availability for long discussions and brainstorming sessions were instrumental to this work, and her many valuable insights into conducting public trials were integral to the design of this project.

I would also like to thank Manuel Wüthrich for providing the Python implementation for the ucb-α-pav algorithm as used in the paper he helped co-author.

Many thanks to Jenny Huang, Meghana Potta, Vishwak Srinivasan, and Mansi Joisher of the MIT EECS Communications Lab for helping me to revise this thesis.

Finally, I am eternally grateful for my family's unconditional love, support and encouragement. Without them, I would not be who or where I am today. I also want to especially thank my parents for helping me to label data, as well as test the deliberation platform, for this project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the largest flaws (or features) of representative democracy is that the public generally have little direct input into the political decision-making process, except by voting for politicians who align with their interests or by voting on the rare direct ballot initiative. One would expect that a government of the people, by the people, and for the people would more greatly involve *the people* in policy-making. Thus, the arrival of such a democratizing technology as the internet brought hope of transforming our society into a stronger [1], more participatory democracy [2][3]. Indeed, disciplines like deliberative democracy [4], which calls for political decisions to be based on the results of fair and balanced discussions among citizens, inspired a new wave of projects [5][6][7] for digital democracy [8]. In general, there have been numerous trials of electronic participation (e-participation) programs, which seek to promote citizen participation in policy-making through the use of information and communication technologies [9][10][11][12][13][14].

However, these programs have had decidedly mixed results [15][16][17][18]. On the one hand, digital technology promotes online public spaces where citizens can overcome physical barriers to collaborate and share information [19][20], as long as policymakers take the proper steps to ensure broad participation [21][22][23]. On the other hand, these experiments have also been criticized for a variety of reasons: involving a demanding conception of citizenship while having a negligible impact on policy-making [24][25], permitting only a limited level of citizen participation [26][27][17], and limiting use cases to information sharing and dialogue rather than the core stages of decision-making and implementation [26][28][17]. For a full characterization and analysis of e-participation tools, see Shin et al. [29].

For the remainder of this thesis, we focus on participatory and deliberative democracy platforms, some of the more broadly-used e-participation technology in practice [30][31][32][27][7][33]. In fact, some deliberative democracy platforms—including *Consul*[1], *Decidim*[2], and *Polis*[3]—emerged from popular movements, are open source, and are being used by both local governments (like the Generalitat de Catalunya[4]) and municipalities (such as Barcelona[5], Madrid[6], and Reykjavik[7]). Still, the use of these forums is not as extensive as one might hope.

Given the prevalence of these platforms, as well as interest in these tools gaining even more widespread adoption, it is crucial to address their more recently-observed shortcomings, including declining participation rates, low quality of deliberations, and limited impact of initiatives on legislation [7][34][35][33]. To do so, we investigate newly-proposed augmentations to the design of *Polis*, an existing deliberation platform.

Polis allows hosting online deliberations on a specific issue, promoting the discussion of different perspectives and/or arguments pertaining to the issue. It does this by allowing participants to submit comments on the issue, as well as to vote their approval (or disapproval) of others' comments. By then presenting the results of the discussion to policy-makers, *Polis* facilitates the integration of citizen input into political decision-making. But given the large volume of comments, it would be prohibitive to have to examine all, or even part, of this input before incorporating it into new policy proposals. Thus, *Polis* tries to provide a way

[1] https://consuldemocracy.org/
[2] https://decidim.org/
[3] https://pol.is/
[4] https://participa.gencat.cat/
[5] https://decidim.barcelona
[6] https://decide.madrid.es/
[7] http://reykjavik.is/en/participation/

to summarize the discussion by compiling a small set of comments that are indicative of the sentiments of the broader discussion.

As noted by Deseriis et al. [36] when comparing six of the most popular democratic decision-making systems, the design of each platform embeds a specific notion of democracy and political participation. Thus, since voting introduces a more "minimalist" notion of participation [37] that places a lower decisional burden on users, *Polis*'s platform design is an ideal starting point to address the aforementioned modern shortcomings of deliberative platforms. It offers the low-commitment feature of voting on comments, to promote greater rates of participation, while still offering anyone the ability to contribute more than just a vote by submitting comments themselves. Additionally, because *Polis* forgoes direct replies, which lead to a breakdown in information structure when implemented at scale[8], it also leads to higher-quality discussions for larger deliberations. Finally, by trying to provide a subset of the comments that 'summarize' the discussion, *Polis* can help facilitate the incorporation of citizen feedback into actual legislation (note that in this thesis, a summarization is a *subset of the comments* from the deliberation). However, as this selection forms the backbone of citizen feedback that policy-makers consider, it is imperative that the provided summarization is representative of all perspectives, an obligation that is not guaranteed by Polis[9].

Within the past year, two new algorithmic approaches have been published that seek to better tackle this challenge. For reference, for its summarization, Polis presents the comments with the greatest majority support, in addition to clustering its participants into 'opinion groups' according to similar voting patterns and displaying the distinguishing comments for each group. However, Polis's approach does not *guarantee* fair representation of the entire deliberation, nor does it *guarantee* consistent summaries of different viewpoints.

In choosing these newer algorithmic approaches, we consider the analysis of online deliberation platforms by Tsai et al. [38], who recognize two key objectives of deliberative

[8] https://blog.pol.is/pol-is-in-taiwan-da7570d372b5

[9] In this thesis, we distinguish the platform design, *Polis*, from its current algorithmic approach, Polis, by use of italics.

processes. First, they note that these processes must represent a diverse range of perspectives, especially those of the marginalized or underrepresented, to both foster inclusiveness as well as strengthen the legitimacy and acceptance of the discussion's results. This is why we examine a new computational social choice (ComSoC)-inspired algorithm introduced by Halpern et al. [39], which constructs a representative set of comments with guarantees of fair representation. Second, Tsai et al. [38] identify that an effective deliberation is marked by reasonable and coherent conclusions, for which there is collective understanding and acceptance. Therefore, we also investigate a new argumentation- and ComSoC-inspired algorithm presented by Bernreiter et al. [40] that produces a small but representative set of consistent and justifiable viewpoints.

Nonetheless, though these new algorithms improve on Polis's performance in objective metrics, neither was inherently designed to consider the subjective experiences of participants, which is crucial for the real-world usefulness of these approaches. Indeed, though an algorithm may have strong theoretical guarantees, it matters just as much, if not more, that people *feel* represented in the summary of the deliberation. As noted by Tsai et al. [38], high levels of participant satisfaction are important for ensuring that the results of the deliberation are viewed as legitimate.

Additionally, these algorithms include methods to adaptively present participants comments to vote, known as comment routing methods, to efficiently collect the viewpoints and perspectives of discussions for summarization. Thus, if we seek to examine the subjective experiences of participants, we also need to look into their experience with these different comment routing methods.

This work seeks to investigate potential improvements to the *Polis* deliberative platform by evaluating the relative usefulness of these new summarization algorithms as compared to Polis's current methods. Our primary contribution is to implement all three algorithms—Polis, ComSoC, and Argumentation—in a real-world setting with human participants, to compare the relative strengths and weaknesses of these three paradigms in successfully distilling the diverse viewpoints of a deliberation into a set of key comments to inform policy-making. Specifically, for each of three different topics, we lead a deliberation using each algorithm, and collect feedback on how well participants feel the discussion's summarization represents their viewpoints, and therefore, would make a good basis of concerns for policymakers to keep in mind while drafting legislation.

1.1 Related Work

Some of the most closely related work to this research is from the papers that propose the very algorithms we seek to test. Indeed, Halpern et al. [39] run their own empirical tests of their algorithm, using real-world data. However, they only use modified historical data, for which they have computationally 'inferred' missing votes, to allow their algorithm to request a historical participant's vote on a comment they did not see. Bernreiter et al. [40] also empirically test their algorithm, but they only do so using purely synthetic data, citing the difficulty of labelling attack relations on real data. It is important to note, though, that neither conducted live, real-world tests of their methodologies, and neither considers the subjective experiences of voters when designing their methods.

In a similar vein to our research into improving the *Polis* platform, López-Sánchez et al. [41], building off of their previous work [42], propose and investigate aggregation methods—including the Proposal Argument Map (PAM) [43], the Target oriented discussion framework (TODF) [42], and a hybrid of the two [41]—to add to the *Decidim* platform. Given that *Decidim* structures conversations somewhat differently from the design of *Polis*, their work is not applicable to our research, since their algorithms make use of structures unique to *Decidim*. Furthermore, they propose comment aggregation in order to stimulate further discussion within a deliberation, rather than to provide representative summarizations of the discussion.

Finally, inspired by the apparent explosion in use of Large Language Models (LLMs) in recent years, Fish et al. [44] introduce the theory of generative social choice, which seeks to combine LLMs with ComSoC to produce better democratic outcomes than ComSoC alone. Not only do they test their implementation in a real-world discussion, finding respectable results, but in fact, a similar study by Konya et al. [45] demonstrated a process for democratic policy development using collective dialogues and LLMs that has been provably carried out on a large scale. In three different 1500-participant tests of their process conducted on somewhat contentious issues, they achieved a 75% approval rate for final policy proposals, with ≥70% approval rates among demographic splits spanning age, race, gender, education, and political party.

However, there are also many different risks involved with using LLMs, including but not limited to: hallucinations (whereby the LLM 'makes up' a response that has no grounding in reality), biases against groups of people and viewpoints (which are a well-documented issue with LLMs), and lack of transparency (due to the complex and incomprehensible nature of large neural networks). As acknowledged by Small et al. [46], who investigate how LLMs could be applied to help other areas of *Polis*, as well as by Fish et al. [44], such risks are detrimental to the goal of fair and accountable democratic deliberations. Since we are wary of these risks and fear we cannot apply the care needed to mitigate them—if such is even possible—we do not include this method as a treatment in our experiments.

1.2 Thesis Organization

The remainder of this thesis is organized as follows:

- In Chapter 2, we discuss, as described in their respective papers, the theoretical framework underpinning each algorithm, going over both the implementation of and motivations behind each approach.

- In Chapter 3, we describe the steps needed to adapt each of the algorithms for real-world use in our platform, and evaluate our modifications against the original implementations that accompany their respective papers.

- In Chapter 4, we outline the experimental design of our study, including how we set up our simulated deliberations as well as the specific outcomes we measure.
- In Chapter 5, we compile and analyze the results of our study, and discuss the major takeaways from our research.
- In Chapter 6, we summarize our findings from this study and present potential areas of exploration for future work.

Chapter 2

Theoretical Background

Though all three of these algorithms—Polis, ComSoC, and Argumentation—operate within the *Polis* platform's design model, they each differ in the method they use to generate a summary of the discussion, as well as the method that they use to route comments (i.e. offer participants comments to vote on). In this chapter, we review, in detail, the theoretical framework behind each of these algorithms, as outlined in their respective papers, after briefly addressing our motivation for investigating these algorithms in our study.

We try to make these technical explanations easy to follow regardless of theoretical background, though we occasionally use mathematical notation for summation ($\sum$), product ($\prod$), and set-theoretical concepts like element of ($\in$), where ($|$), there exists ($\exists$), such that ($:$), cardinality of a set ($|\cdot|$), set difference ($\setminus$), cross ($\times$), subset / superset ($\subset$ / $\supset$), and union / intersection ($\cup$/$\cap$), for convenience. We also use the mathematical operators max (which gives the maximum value over a given set of arguments), min (which gives the minimum value over a given set of arguments), and argmax (which gives the *argument* that maximizes a value over a given set of arguments). Finally, from complexity theory, we use the notation $\mathcal{O}(f(x))$, which essentially means 'on the order of' function $f(x)$.

2.1 The Polis Algorithm

While its summarization method may have its shortcomings (namely, that it does not guarantee representativeness or consistency), the Polis algorithm has a proven track record in facilitating discussion that has impacted real-world policy, such as in Taiwan's vTaiwan process[1] and the Bowling Green Public Assembly in Bowling Green, Kentucky[2]. Thus, it serves as a useful baseline for comparison against the two newly proposed algorithmic approaches.

2.1.1 Theoretical Framework

As detailed by Small et al. [47], Polis's deliberations begin with a topic being created on a particular issue, with optional seed comments. Users are then invited to submit comments of their own, as well as vote their (dis)agreement with other people's comments (including the seed comments). During the deliberation, users are are shown comments to vote on one at a time, each selected at random from a non-uniform distribution over the comments, where each comment is weighted according to a computed priority metric. As this data is collected, it is continually recombined and reanalyzed to cluster individuals into opinion groups, identify the distinguishing comments for each group, and place groups in the political landscape, in addition to informing the priority metric.

More specifically, in real time, votes are collected in a voting matrix V, with rows indexed by participant and columns indexed by comment, such that element $v_{i,j}$ corresponds to the vote of participant i on comment j (agree (a) is encoded as $+1$, disagree (d) as -1, and pass (p) as 0). To analyze the deliberation after each update to the voting matrix, the current voting matrix is used as a basis to generate an analysis matrix. For this analysis matrix, missing entries of the voting matrix, which correspond to comments that a participant has not voted on, are imputed by taking the column-wise average of non-missing values, i.e. the

[1]https://www.centreforpublicimpact.org/case-study/building-consensus-compromise-uber-taiwan

[2]https://web.archive.org/web/20210414093745/https://civichall.org/civicist/testing-tech-consensus-purple-town/

average approval for that comment in the conversation. Meanwhile, rows corresponding to participants who have voted on fewer than seven comments are removed from the analysis matrix to prevent the "clumping up" of participants around the center of conversation when analyzing their votes.

Then, dimensionality reduction is performed on the data using principal component analysis (PCA) [48][49]. This produces a 2D representation of the data (which can be thought of as a 2D "map" of the opinion space) presented as a two-column matrix, where each row corresponds to the location in 2D space of a participant's position. Each row in this representation is further scaled by the factor of $\sqrt{\frac{C}{C_p}}$, where C is the total number of comments and C_p is the number of comments voted on by participant p, to correct for the fact that participants with lesser engagement get projected closer to the center of the conversation, since they are assumed to vote the average for any comment they have not seen.

Finally, this 2D projection is used to perform a fine-grained clustering analysis using K-means clustering [50][51], with $K = 100$, to produce a set of base clusters. These base clusters then serve as the basis for coarse-grained clustering, also using K-means, to determine opinion group clusters. Here, multiple runs of K-means are performed, for values of K from 2 to 5. The K with the greatest silhouette coefficient (a measure of within-cluster similarity vs. between-cluster dissimilarity) [52] is chosen as the number of opinion group clusters, with a smoothing function applied (i.e. for the number of clusters to change, the new value for K must be consistently observed across multiple rounds of analysis) to ensure that this number does not fluctuate too frequently. This smoothing is especially necessary for the beginning of a conversation, when the opinion landscape shifts more rapidly with each vote.

These clusters are used not only to provide a real-time visualization of opinion groups placed in the political landscape, but also to compute distinguishing comments for each cluster, which, along with the computed consensus comments, serve as the provided summarization of the deliberation.

Summarizing the Deliberation

Small et al. [47] explain that in Polis, comments are analyzed for how strongly they represent each opinion group by the representativeness metric $R_v(g, c)$, which, for group g, comment c, and vote v, estimates how much more likely participants in group g are to vote v on this comment than participants outside of g. Letting $N_v(g, c)$ be the number of participants in group g who cast vote v on comment c, and $N(g, c)$ be the total number of votes on comment c within group g, they compute

$$P_v(g, c) = \frac{1 + N_v(g, c)}{2 + N(g, c)}$$

as an estimate on the probability that a given person in group g votes v on this comment (where the 1 and 2 pseudocounts ensure that this metric defaults to $\frac{1}{2}$ in the absence of votes). Then the representativeness metric is defined as the estimated relative odds ratio,

$$R_v(g, c) = \frac{P_c(g, c)}{P_v(\overline{g}, c)} \ ,$$

where $\overline{g}$ is the complement of g, i.e. everyone in the conversation not in g.

To determine distinguishing comments for each group, the two-property Fisher exact test is also performed [53], and the corresponding Fisher Z-statistic is multiplied by $R_v(g, c)$ to reflect both the estimated effect size and the statistical confidence associated with that effect. These metrics are computed for both the agree and disagree votes for every comment, for every group, so that the top distinguishing comments for each group can be selected.

Meanwhile, the group clusters also inform the group-aware consensus metric, calculated as

$$C_v(c) = \prod_{g \in G} P_v(g, c)$$

for $v = a$ (or $v = d$), which is maximized when *all* groups tend to agree (or disagree, respectively) with a comment, helping to protect from the tyranny of the majority and allow minority dissent to be respected. Using this metric, the comments with the top group-aware consensus are also selected.

Comment Routing

Additionally, Small et al. [47] detail that in Polis, comments are chosen to send to participants by sampling randomly from a non-uniform distribution over the comments, which is formed by a given priority metric. This priority metric reflects each comment's likeliness to place participants in the opinion landscape, seeks to build consensus, and highlights comments new to the conversation. In particular, during voting, participants are sent c as the next comment to vote on with probability $\textsc{Priority}(c)$, normalized by the sum of such values for all comments.

Let $P_v(c) = P_v(G, c)$ as defined above, where G is the set of all participants; $N(c)$ be the total number of votes on comment c; and $E(c)$ be the extremity of comment c, defined as the distance from the center of the conversation of a theoretical participant who *only* voted 'agree' on comment c and voted on no other comments. Then

$$\textsc{Priority}(c) = \left[P_{v=a}(c) \cdot (1 - P_{v=p}(c)) \cdot (1 + E(c)) \cdot (1 + 2^{3-\frac{N(c)}{5}})\right]^2 .$$

The equation is constructed so that each of the terms in the product has value greater than 1 for comments that should be sent more to participants, and value between 0 and 1 for comments that should not be shown as often. The $P_{v=a}(c)$ term is meant to boost consensus by promoting comments with higher agreement, and decreases to 0 for comments with little support. The $(1 - P_{v=p}(c))$ term decreases to 0 for comments that have been mostly passed on. The $(1 + E(c))$ term helps to place participants in the conversational landscape, by promoting comments with strong opinions. Finally, the $(1 + 2^{3-\frac{N(c)}{5}})$ term helps highlight new comments by emphasizing those with fewer votes. The outer square term strengthens the effect of the bias towards comments boosted by these factors.

2.2 The Computational Social Choice (ComSoC) Algorithm

While Polis's analysis tries to offer some distinguishing comments from each opinion group, to provide disparate perspectives in its summary, this summary fails to satisfy any specific representation and satisfaction guarantees. To address this issue, Halpern et al. [39] introduce a new summarization algorithm using tools from computational social choice (ComSoC) theory, which *does* satisfy such guarantees (namely, EJR and OAS, as defined below). This allows it to capture the diversity of opinions inherent to a discussion, making it a great choice to explore in our search for a better summarization algorithm for *Polis*-like deliberations.

2.2.1 Theoretical Framework

Here, the ComSoC algorithm is incrementally developed following the constructions presented by Halpern et al. [39].

Preliminaries

In the basic approval-based committee-selection setting [54], there is a set $N = \{1, \ldots, n\}$ of n voters and a set C of m comments. Each voter $i \in N$ approves of a set of comments $A_i \subseteq C$, and the sequence $\mathbf{A} = (A_1, \ldots, A_n)$ is the voters' *approval profile*. Given these inputs, for a specified *target committee size* $k \leq m$, a *k-committee-selection algorithm* is one that chooses a committee $W \subseteq C$ of size k.

As an example, Proportional Approval Voting (PAV) is one such (broadly-studied) committee selection algorithm, which, given an approval profile $\mathbf{A}$ and a committee size k, outputs a committee W of size k maximizing the *PAV-score*,

$$\text{PAV-SC}(W) = \frac{1}{n} \sum_{i \in V} \sum_{j=1}^{|A_i \cap W|} \frac{1}{j} \ .$$

When discussing representation, a group of voters $V \subseteq N$ is *ℓ-large* if $|V| \geq \ell \cdot \frac{n}{k}$ and *ℓ-cohesive* if $|\bigcap_{i \in V} A_i| \geq \ell$ (i.e. they all agree on at least ℓ comments). Aziz et al. [54] introduced the following two notions of fairness:

Definition 2.2.1 (Justified Representation (JR)) *A committee W satisfies JR if for every 1-large, 1-cohesive group of voters V, there exists a member $i \in V$ who approves at least one comment in W, i.e. $|W \cap A_i| \geq 1$.*

Definition 2.2.2 (Extended Justified Representation (EJR)) *A committee W satisfies EJR if for every $\ell \in \{1, \ldots, k\}$, and every ℓ-large, ℓ-cohesive group of voters V, there exists a member $i \in V$ who approves at least ℓ comments in W, i.e. $|W \cap A_i| \geq \ell$.*

From this Halpern et al. [39] look at the approximate version of EJR:

Definition 2.2.3 (α-Extended Justified Representation (α-EJR)) *A committee W satisfies α-EJR if for every $\ell \in \{1, \ldots, k\}$, and every $\frac{\ell}{\alpha}$-large, ℓ-cohesive group of voters V, there exists a member $i \in V$ who approves at least ℓ comments in W, i.e. $|W \cap A_i| \geq \ell$.*

Also, Fernandez et al. [55] proposed another notion of fairness:

Definition 2.2.4 (Average Satisaction) *The average satisfaction of a group of voters V with a committee W, $avs_W(V)$, is the average overlap of the approval profile of a voter in V with the committee W, i.e. $avs_W(V) = \frac{1}{|V|} \sum_{i \in V} |A_i \cap W|$.*

From this, Halpern et al. [39] defined a-OAS, which measures how close a committee is to the maximum average satisfaction that can hold for all elections.

Definition 2.2.5 (α-Optimal Average Satisfaction (α-OAS)) *A committee W satisfies α-OAS if for every $\lambda \in \{0, \ldots, k\}$, and every $\frac{\lambda+1}{\alpha}$-large, λ-cohesive group of voters V, $avs_W(V) \geq \lambda$.*

As $\alpha = 1$ is its maximum possible setting [56][57], they refer to 1-OAS simply as OAS. Analogously, note that 1-EJR is just EJR.

The aforementioned PAV satisfies EJR and OAS [55][56]; however, it is also computationally intractable[3]. Therefore, Aziz et al. [56] introduced a local search approximation for PAV, LS-PAV, which still satisfies EJR and OAS, but can also be computed efficiently[4].

Simpler Algorithms using Exact Queries

First, Halpern et al. [39] consider the exact query setting, in which an algorithm can perform a query $Q \subseteq C$ of size t, and essentially receives in response each set $Q \cap A_i$ (i.e. the comments within Q that voter i approves of), for each $i \in N$ (though in a way that information from separate responses cannot be combined). In this setting, they introduce an algorithm based on LS-PAV [56] that satisfies EJR and OAS for a practically feasible number of queries.

For committee W and comments $c \in W$, $c' \notin W$,

$$\Delta(W, c', c) := \text{PAV-SC}(W \cup \{c'\} \setminus \{c\}) - \text{PAV-SC}(W)$$

is the difference in PAV score from replacing c with c' in W, and

$$\Delta(W, c') := \text{PAV-SC}(W \cup \{c'\}) - \text{PAV-SC}(W)$$

is the marginal increase in PAV score from adding c' to W.

LS-PAV begins with an arbitrary committee W and repeatedly replaces a comment $c \in W$ with another comment $c' \notin W$, as long as $\Delta(W, c', c) \geq \frac{1}{k^2}$. As proven by Aziz et al. [56], at most $\mathcal{O}(k^2 \log k)$ such swap pairs can be found, after which point W satisfies EJR and OAS.

Halpern et al. [39] note that LS-PAV can be implemented using exact queries. For any W, $c \in W$, and $c' \notin W$, $\Delta(W, c', c)$ can be computed from a query Q that includes both W

[3]Specifically, PAV is NP-hard to compute [58].

[4]Specifically, LS-PAV runs in polynomial time [56].

Algorithm 1 (k, t)-α-PAV

Choose $W \in \binom{C}{k}$, $c \in W$, and $c' \notin W$ arbitrarily
$\gamma \leftarrow \infty$
while $\gamma \geq \frac{1}{\alpha k}$ **do**
  $W \leftarrow W \cup \{c'\} \setminus \{c\}$
  Choose $\mathcal{Q} = (Q_1, \ldots, Q_j)$, with $|Q_i| = t$, such that $W \subseteq \bigcup_i Q_i$ and $C \subseteq \bigcap_i Q_i$ ▷ ($\mathcal{Q}$ is chosen so that every query Q_i contains the committee W, and every comment $c \in C$ is covered by some query Q_i)
  $c' \leftarrow \operatorname{argmax}_{x \notin W} \Delta(W, x)$ ▷ (using, for each x, the query Q_i that contains both W and x)
  $c \leftarrow \operatorname{argmax}_{y \in W} \Delta(W, c', y)$ ▷ (using the query Q that contains both W and c')
  $\gamma \leftarrow \Delta(W, c')$
return W

and c' (i.e. $W \cup \{c'\} \subseteq Q$). By using $j = \frac{m-k}{t-k}$ queries of size t, all $m - k$ comments not in W can be covered by one of these queries, to complete a round of the algorithm. This gives LS-PAV an overall (worst-case) query complexity of $\mathcal{O}(mk^2 \log k)$.

They then present the following version of LS-PAV, called α-PAV (Algorithm 1), which only has (worst-case) query complexity $\mathcal{O}(mk \log k)$ in order to satisfy approximate ($\alpha < 1$) α-EJR and α-OAS[39].

In addition to the approximation parameter α, Algorithm 1 differs in two key ways from LS-PAV: First, the termination condition that there is no alternate $c' \notin W$ such that $\Delta(W, c') \geq \frac{1}{k}$ (when $\alpha = 1$) is weaker than the termination condition of LS-PAV that there is no pair c, c' such that $\Delta(W, c', c) \geq \frac{1}{k^2}$, implying it may terminate earlier. Second, instead of considering all possible swaps c, c', it only considers adding to W the alternate c' with the greatest marginal increase in PAV score, $\Delta(W, c')$, which is slightly more computationally efficient (by a factor of k).

Better Algorithms using Noisy Queries

Now, Halpern et al. [39] turn to the noisy query setting, in an effort to use a more realistic model. To represent voters coming into the platform one at a time, in this setting, each time the algorithm performs a query $Q \subseteq C$ of size t, it receives in response the set $Q \cap A_i$, for a

Algorithm 2 (k, t)-noisy-α-PAV

$\ell \leftarrow \frac{1}{\epsilon^2} \log(\frac{m}{\delta})$
Choose $W \in \binom{C}{k}$, $c \in W$, and $c' \notin W$ arbitrarily
$\gamma \leftarrow \infty$
while $\gamma \geq \frac{1}{\alpha k} - \epsilon$ **do**
  $W \leftarrow W \cup \{c'\} \setminus \{c\}$
  Choose $\mathcal{Q} = (Q_1, \ldots, Q_j)$, with $|Q_i| = t$, such that $W \subseteq \bigcup_i Q_i$ and $C \subseteq \bigcap_i Q_i$
  Ask each query $Q \in \mathcal{Q}$ to ℓ new voters
  $\hat{\Delta}(W, x) \leftarrow$ estimate of $\Delta(W, x)$ using the ℓ voters who answered the query Q_i that contains both W and x ▷ For each $x \notin W$
  $c' \leftarrow \operatorname{argmax}_{x \notin W} \hat{\Delta}(W, x)$
  $\hat{\Delta}(W, c', y) \leftarrow$ estimate of $\Delta(W, c', y)$ using the ℓ voters who answered the query Q that contains both W and c' ▷ For each $y \in W$
  $c \leftarrow \operatorname{argmax}_{y \in W} \hat{\Delta}(W, c', y)$
  $\gamma \leftarrow \hat{\Delta}(W, c')$
return W

randomly chosen voter $i \in N$.

They note that an algorithm with noisy queries can approximate an exact query Q by aggregating estimates of Q from repeated noisy queries. By standard sample complexity bounds [39], with probability $1 - \delta$, using $\mathcal{O}(\frac{1}{\epsilon^2} \log(\frac{r}{\delta}))$ queries, a noisy-query algorithm could guarantee an estimate for each of the r values that Q is used to calculate within an ϵ tolerance for each. Since, in Algorithm 1, each exact query Q is used to calculate a maximum of $r = m$ values of the form Δ[5], they modify it to allow for additive ϵ error and use $\ell = \frac{1}{\epsilon^2} \log(\frac{m}{\delta})$ noisy queries to approximate each exact query. This results in Algorithm 2, called noisy-α-pav, which they prove has (worst case) query complexity $\mathcal{O}(mk^3 \log k \log m)$ to satisfy approximate ($\alpha < 1$) α-EJR and α-OAS with probability $1 - \delta$ [39].

However, in order to achieve this, the algorithm must choose the tolerance ϵ so that ℓ is large enough that if the termination condition of the loop is not met (i.e. $\hat{\Delta}(W, c') \geq \frac{1}{\alpha k} - \epsilon$), the resulting swap is guaranteed to result in a positive improvement to the PAV score. They choose $\epsilon = \frac{(1-\alpha)k+1}{\alpha k^2}$, which finally allows Algorithm 2 to satisfy the above conditions.

[5] In particular, there are a maximum of $r = m$ values of the form Δ because in line 6 of Algorithm 1, each exact query Q is used to calculate $m - k$ values $\Delta(W, x)$, and in line 7, an exact query Q is used to calculate k values $\Delta(W, c', y)$, for a total of $m - k + k = m$ values.

From here, they apply further optimizations to improve the average-case performance of the final algorithm, known as ucb-α-pav. First of all, after every swap, Algorithm 2 discards all previous information, reassessing every alternate from scratch. To speed this up, in the final algorithm, past votes are used to compute bounds on the estimated values $\hat{\Delta}(W, c', c)$ and $\hat{\Delta}(W, c')$, even though the working committee W may have since changed. Furthermore, Algorithm 2 presents every possible alternate $c' \notin W$ to the same number of voters, even though it can quickly become apparent which ones are more or less promising. To address this, promising candidates are shown to voters more often in the final algorithm. Additionally, the final algorithm performs swaps as soon as it is confident an alternate yields a marginal increase in PAV-Score above a certain threshold, rather than always querying a predetermined number of voters.

For the full implementation of this final algorithm, see Algorithm 4 in [39].

2.3 The Argumentation Algorithm

Another shortcoming of Polis's analysis is that the summaries of different opinion groups are not guaranteed to be consistent. Furthermore, its results are generally not explainable (i.e. observers cannot truly follow the decision-making process of a specific instance of the algorithm), which makes it difficult for the democratic process to be accountable. While the ComSoC algorithm may be somewhat explainable through the working history of its representative 'committee', Polis's PCA dimensionality reduction is too opaque for regular citizens to follow. To address these issues, Bernreiter et al. [40] use tools from ComSoC and abstract argumentation to propose an algorithm that provides a consistent, explainable, and representative summary of a deliberation, another great option to explore in our search for better summarization algorithms for *Polis*-like deliberations.

2.3.1 Theoretical Framework

Again, as in the case of the previous algorithm, the ideas for the Argumentation algorithm are slowly developed following the constructions presented by Bernreiter et al. [40].

Preliminaries

The problem of selecting representative comments starts, as before, with the basic approval-based committee-selection setting, where there is a set $N = \{1, \dots, n\}$ of n voters and a set C of m comments. Each voter $i \in N$ approves of a set of comments $A_i \subseteq C$, and the sequence $\mathbf{A} = (A_1, \dots, A_n)$ is the voters' *approval profile*. However, instead of selecting a committee, Bernreiter et al. [40] try to choose a set Ω of k subsets of C (i.e. $\Omega \subseteq 2^C$, $|\Omega| = k$, where 2^C denotes the power set of C). Rather than explicitly limiting the cardinality of the selected subsets in Ω, they impose consistency constraints using abstract argumentation.

Argumentation Frameworks (AFs) [59] are a widely-studied concept in artificial intelligence and related fields, through which discussions can be represented and reasoned about. Arguments (which, in our case, represent comments) in AFs are abstract entities, meaning the focus is not on their internal structure but rather on the relationships between them. Particularly, if an argument x attacks an argument y, then they are in conflict: they cannot both be accepted. Furthermore, in order to accept y, it must be defended from x's attack, i.e. either it attacks x itself or another argument z that attacks x (that *can* be accepted alongside y) must be jointly accepted.

Definition 2.3.1 (Argumentation Framework (AF)) *An AF* $F = (\mathrm{Arg}, \mathrm{Att})$ *consists of a set of arguments* Arg *and an attack relation* $\mathrm{Att} \subseteq \mathrm{Arg} \times \mathrm{Arg}$ *between arguments. For* $S \subseteq \mathrm{Arg}$*,*

- S attacks $b \in \mathrm{Arg}$ *if* $(a, b) \in \mathrm{Att}$ *for some* $a \in S$*.*
- $S_F^+ = \{b \in \mathrm{Arg} \mid \exists a \in S : (a, b) \in \mathrm{Att}\}$ *denotes the set of arguments attacked by* S*.*
- *An argument* $a \in S$ *is* defended *by* S *if, for each* b *with* $(b, a) \in \mathrm{Att}$*,* $b \in S_F^+$*.*

AF-semantics are functions σ that assign a set $\sigma(F) \subseteq 2^{\text{Arg}}$ of extensions (i.e. subsets of arguments) to an AF $F = (\text{Arg}, \text{Att})$. Conflict-free semantics ($\sigma = \text{cf}$) choose sets $S \subseteq \text{Arg}$ such that no two elements attack one another. Admissible semantics ($\sigma = \text{adm}$) choose conflict-free sets that defend themselves. Finally, preferred semantics ($\sigma = \text{prf}$) choose subset-maximal admissible sets (i.e. no further arguments can be added to any preferred extension).

Definition 2.3.2 *For AF $F = (\text{Arg}, \text{Att})$, $S \subseteq \text{Arg}$, it holds that:*

- *$S \in \text{cf}(F)$ if and only if there are no $a, b \in S$ such that $(a, b) \in \text{Att}$.*
- *$S \in \text{adm}(F)$ if and only if $S \in \text{cf}(F)$ and each $a \in S$ is defended by S.*
- *$S \in \text{prf}(F)$ if and only if $S \in \text{adm}(F)$ and there does not exist $T \in \text{adm}(F)$ such that $S \subsetneq T$*

Approval-based Social AFs

Bernreiter et al. [40] combine AFs with approval-based committee selection to create Approval-based Social AFs, which can be used to model deliberations.

Definition 2.3.3 (Approval-based Social AFs) *An ABSAF $\mathcal{S} = (F, N, \mathbf{A})$ consists of an AF $F = (\text{Arg}, \text{Att})$, a set $N = \{1, \ldots, n\}$ of n voters, and an approval profile $\mathbf{A} = (A_1, \ldots, A_n)$, where each voter $i \in N$ approves of a set of comments $A_i \subseteq \text{Arg}$.*

There are no constraints on the submitted approval ballots—not even that they be conflict-free—as ballots containing conflicts appear in real-world examples [40]. The goal in using ABSAFs is to select a small set of coherent perspectives that represent the voters.

Definition 2.3.4 *An outcome $\Omega \subseteq \sigma(F)$ of an ABSAF $\mathcal{S} = (F, N, \mathbf{A})$ is a set of σ-extensions of F. Then $\pi \in \Omega$ is called a viewpoint.*

In order to find a representation that both contains few (namely, k) viewpoints and represents as many voters as possible, they define a measure of how well a voter is represented by an outcome or viewpoint.

Definition 2.3.5 *For ABSAF $\mathcal{S} = (F, N, \mathbf{A})$, voter $i \in N$, outcome $\Omega \subseteq \sigma(F)$ (for $\sigma = \mathrm{prf}$), and viewpoint $\pi \in \Omega$:*

- $rep_i(\pi) = \frac{|\pi \cap A_i|}{|A_i|}$
- $rep_i(\Omega) = \max_{\pi \in \Omega}(rep_i(\pi))$

In theory, voters are assumed to be rational agents [60], arriving at a consistent and defendable position, called their *ideal* position, if given enough time to properly consider all viewpoints. However, since participants in a deliberation often do not have such time, such restrictions on their voting behavior cannot be assumed. Thus, $rep_i(\pi)$ can also be interpreted as a measure of how consistent a participant's voting behavior is with the premise that π represents their ideal voting behavior. Due to the rationality assumptions of ideal positions, and the desire for viewpoints to represent as many voters as possible, Bernreiter et al. [40] focus on the subset-maximal cohesive viewpoint of preferred extensions ($\sigma = \mathrm{prf}$). After all, if any admissible extension π contains a voter's approval preferences, so does the preferred extension $\pi' \supseteq \pi$. This then motivates the definition that $rep_i(\Omega) = \max_{\pi \in \Omega}(rep_i(\pi))$.

Optimizing Representation

Now that they have resolved to look at only preferred extensions, Bernreiter et al. [40] consider how to choose an outcome that optimally represents the voters in an ABSAF.

One approach common in social choice theory is the *Utilitarian* rule [61], which seeks to maximize the average representation across all voters (i.e. $\frac{1}{n}\sum_{i \in N} rep_i(\Omega)$). Another is the *Egalitarian* rule [62], which aims to maximize the representation of the least-represented voter (i.e. $\min_{i \in N} rep_i(\Omega)$).

These notions are generalized by a family of rules inspired by ordered weight averaging (OWA) vectors [63]. For an outcome Ω, and a given OWA rule, if $\vec{s} = (s_1, \ldots, s_n)$ is the vector $(rep_1(\Omega), \ldots, rep_n(\Omega))$ sorted in non-decreasing order (i.e. s_1 corresponds to the least represented voter), and $\vec{w} = (w_1, \ldots, w_n)$ is the non-increasing vector of non-negative weights of the rule, the chosen OWA rule is defined as

$$\mathrm{OWA}_{\vec{w}}(\mathcal{S}) = \underset{\Omega \subseteq \mathrm{prf}(F), |\Omega| \leq k}{\mathrm{argmax}} \; \vec{w} \cdot \vec{s}(\Omega) \ ,$$

where $\cdot$ is the dot product. For the Utilitarian and Egalitarian rules, the corresponding weight vectors are $(1, \ldots, 1)$ and $(1, 0, \ldots, 0)$, respectively. More importantly, inspired by the Proportional Approval Voting (PAV) algorithm from the previous section, the *Harmonic* rule can also be defined, with weight vector $(1, \frac{1}{2}, \ldots, \frac{1}{n})$.

Unfortunately, this $\mathrm{OWA}_{\vec{w}}(\mathcal{S})$ rule is computationally intractable because it is provably NP-hard [40]. The best-known running time of the algorithm is achieved by first enumerating all preferred extensions in $\mathcal{O}(3^{\frac{m}{3}})$ time, where $m = |\mathrm{Arg}|$, followed by enumerating all outcomes of size k in $\mathcal{O}(p^k)$ time, where $p = |\mathrm{prf}(F)|$, which not practical.

This motivates the greedy variant of $\mathrm{OWA}_{\vec{w}}$, called $\mathrm{GreedOWA}_{\vec{w}}$, where viewpoints π are sequentially added to the outcome Ω. Assuming ℓ viewpoints $\pi_1, \ldots, \pi_\ell$ have already been chosen, the $(\ell + 1)$th is chosen by

$$\pi_{\ell+1} = \underset{\pi \in \mathrm{prf}(F) \setminus \{\pi_1, \ldots, \pi_\ell\}}{\mathrm{argmax}} \; \vec{w} \cdot \vec{s}(\{\pi_1, \ldots, \pi_\ell, \pi\}) \ ,$$

stopping when k viewpoints have been chosen.

Sadly, $\mathrm{GreedOWA}_{\vec{w}}$ is also computationally intractable because it is provably NP-hard [40]. Once again, the best-known running time of the algorithm is achieved by first enumerating all preferred extensions is $\mathcal{O}(3^{\frac{m}{3}})$ time. However, the second part of the algorithm only needs to enumerate all viewpoints each time it chooses a new one in the greedy procedure, which can be achieved in the much more useful $\mathcal{O}(pk)$ time. Thus, if the preferred extensions are precomputed before the start of the deliberation, e.g. by powerful argumentation solvers [64][65], this method could potentially be practically feasible.

Chapter 3

Adaptations for Real-World Usage

The three algorithms are in various stages of readiness to be incorporated into the deliberation platform we are using for this study. In this chapter, we discuss the potential shortcomings of these algorithms that prevent their direct application to real-world use and detail the modifications that we applied to these algorithms to make them better suited for use in our platform.

While the full details of our experimental design are detailed in the next chapter, we briefly preview certain aspects of the setup here to more easily manage discussing our adaptations below. Most importantly, each discussion uses a fixed set of 50 previously-collected comments, from which each participant is offered 20 to vote on. Also, for convenience, we say that a participant *requests* a comment to vote on whenever they need to be given a comment to vote on while interacting with the platform.

3.1 The Polis Algorithm

Given that *Polis* is the prototypical platform for the deliberative platform design we are testing, as well as the fact that Polis has a proven track record in facilitating discussions that have impacted real-world policy, we need to make little change to its theoretical framework to prepare it for use on our platform. Indeed, given that it is meant to serve as a baseline, to compare the other approaches against one of the state-of-the-art in this format, we do not *want* to make many modifications to Polis for this study. However, one change we *do* end up making is to use the group-aware consensus, rather than just the regular consensus, when generating the summarization of the discussion. This option is already also computed by Polis, and is meant to be more representative of diverse perspectives, which is important for good citizen input to policy-making.

However, while Polis's implementation is open source, with the source code freely available on GitHub[1], its algorithms are implemented in Clojure. Thus, we have completely reimplemented its methods in Python to make it compatible with our platform's back end (for reference, the provided experimental implementations of the other two algorithms are already in Python).

3.2 The ComSoC Algorithm

We are interested in the ComSoc algorithm because, in theory, it improves on the Polis algorithm by guaranteeing high representation and satisfaction metrics, namely α-EJR and α-OAS. However, in order for such a guarantee to be achieved, the algorithm might, in the worst case, need to query an order of magnitude more participants than there are comments, which is not very practical for a *Polis*-like deliberation platform, where every participant can submit comments.

Luckily though, since this algorithm continuously improves a working committee, with an initial rapid rate of improvement and finer tuning as it goes on, early termination of the algorithm still achieves surprisingly good results, despite losing its strict representativeness guarantees. Indeed, in the experiments performed [39] using past *Polis* data sets (in which missing data values were inferred), each voter was only queried once to vote on 20 comments, with the algorithm terminating afterwards. This implementation not only provided a more

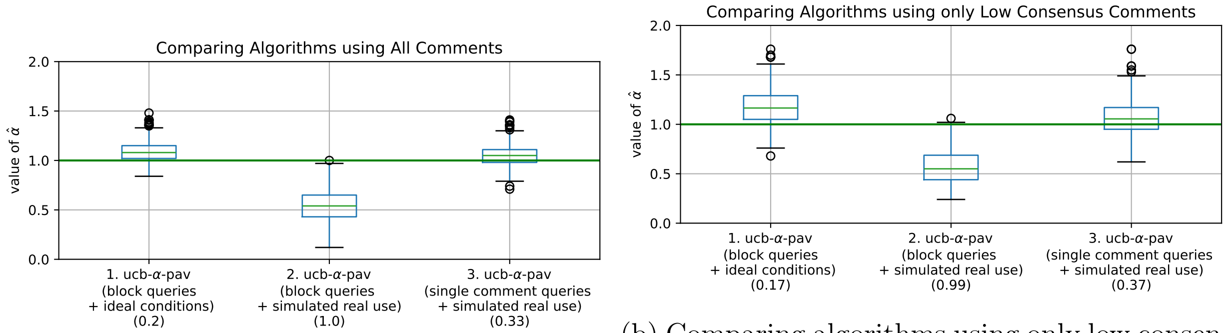
[1] https://github.com/compdemocracy/polis/

applicable construction for real-world use, but also demonstrated a respectable performance compared to an exact computation that had access to all votes.

Still, this 'practical' implementation has shortcomings. For one, the algorithm assumes that it immediately receives the responses to one query before needing to query the next voter. This is unrealistic for a discussion in which multiple participants may join before one participant has finished responding to their query on 20 comments. The naive solution to this—allowing the algorithm to send a new query to each voter as they join—is quite inefficient. Since the algorithm is deterministic, such an adjustment would lead to multiple participants being given the same query set of comments to vote on, which throws away chances for the algorithm to elicit votes on different sets of comments. This would lead to much slower convergence to a representative committee of comments, and therefore much worse performance.

Thus, we instead tackle this challenge by modifying the algorithm to send single-comment queries. In particular, whenever the original algorithm sends out a (comment-block) query Q of 20 comments to be answered by a single voter, we instead split it into its 20 component single-comment queries $q_1, \ldots, q_{20}$. Then, each of the next 20 times some participant requests a comment to vote on, they are randomly given one of the remaining unanswered q_i (for a comment that they have not previously voted on) to vote on. The algorithm is then allowed to run until it generates the next query Q', which is then used for the next 20 times some participant requests a comment to vote on. This cycle repeats as long as participants use the platform to vote on comments.

While we cannot guarantee the optimality of our modifications, we *can* empirically evaluate their performance. In their paper, Halpern et al. [39] experimentally test their algorithms using historical data of previous Polis discussions for which missing votes had been inferred. To measure how representative the final committee of votes for an algorithm is, they use this final committee W to calculate the performance metric $\hat{\alpha}$ (see [39] for the full details of how $\hat{\alpha}$ is calculated). Importantly, they prove that said algorithm satisfies α-EJR and α-OAS for



(a) Comparing algorithms using all comments

(b) Comparing algorithms using only low consensus comments

Figure 3.1: Comparison of modified ComSoC algorithms

any $\alpha < \hat{\alpha}$, so that if $\hat{\alpha} \geq 1$, the corresponding committee W satisfies EJR and OAS.

Across 13 different historical data sets, three different committee sizes, and 10 different randomness seeds, they run each algorithm, and gather the resulting $\hat{\alpha}$ values from these trials together into a boxplot for each algorithm. Comparing algorithms, they demonstrate that their ucb-α-pav algorithm has a respectable performance, by showing that in the vast majority of cases, it results in $\hat{\alpha} \geq 1$, i.e. it finds a final committee that satisfies EJR and OAS.

In a similar vein, we collect, across these same trials, the corresponding $\hat{\alpha}$ values for ucb-α-pav run in its standard setting, as well as both the single query and naive block query variants when run in a simulated concurrent setting. In this setting, voters' requests for comments to vote on are all randomly interleaved, and a voter does not record their response to a request for a random amount of time after the request. Thus, it simulates the worst-case scenario for these algorithms, as far as possible from the sequential setting that ucb-α-pav was designed for. Though real-world use of these algorithms will take place somewhere between these two extremes, it is better to have an algorithm that gracefully handles concurrency.

Specifically, we compare, in Figure 3.1, the relative performance of the original ucb-α-pav algorithm under ideal conditions (where each participant can answer a full block query before the next one joins) against that of the naive block-comment query algorithm and our single-comment query algorithm in these more realistic conditions. The number in parentheses below each algorithm's label represents the fraction of trials in which the respective algorithm did not achieve an $\hat{\alpha} \geq 1$, i.e. did not achieve EJR and OAS. When simulating real-world use in the aforementioned setting, single-comment queries fared much better than naive block-comment queries, as is apparent for both the full historical deliberations (Figure 3.1a) and the simulated contentious discussions (Figure 3.1b), where (as in the paper, to emulate a more challenging setting) comments with high (>60%) consensus are filtered out. In fact, the naive block-comment query approach was usually unable to collect enough useful data to ever update its initial committee, illustrating truly how unsuitable it is for real-world use. In contrast, the single-comment query strategy, in simulated realistic conditions, performs remarkably close to the original algorithm in ideal conditions, affirming our use of this modification for our real-world study.

Another shortcoming of this algorithm is that, unlike Polis, which considers both approval and disapproval in building the viewpoints of different opinion group, this algorithm only considers approvals. However, we have a fairly straightforward way to remedy this: for every comment $c \in C$, we also introduce the comment $\overline{c}$, and record disapproval of c as approval of $\overline{c}$ (passing is simply recorded as approval of neither).

Lastly, we note that even with these modifications, the ComSoC algorithm does not support allowing participants to respond to more than the initial queries, even if they want to contribute more to the discussion (and by the same token, participants cannot be analyzed as part of the conversation if they can only respond to less than the initial queries). Although it is not difficult to directly add such functionality with our setup, the analysis is not constructed to support variable levels of engagement. As this algorithm treats each query response as having been sampled independently and uniformly at random, such variance in engagement would cause the algorithm to give more significance to the opinions of participants who vote more often, which makes this 'solution' untenable. Indeed, the underlying algorithm itself would need to be tweaked to avoid this bias. While it could be theoretically possible to add this functionality, we do not consider this, as it is outside the scope of this project. After all, since every participant in our study votes on the same number of comments, this lack of flexibility does not affect our experiment. Still, we note that this limitation is somewhat antithetical to the goals of deliberative democracy, which seek to encourage, not limit, citizen participation.

3.3 The Argumentation Algorithm

We are interested in the Argumentation algorithm because it improves on Polis by guaranteeing conflict-free summaries of viewpoints, at least in theory. However, it faces a few major challenges to the practical applicability of doing so.

For one, it is difficult to label attack relations in practice, as they either have to be labeled by moderators, crowdsourced from participants, or mined from natural language text using Natural Language Processing (NLP)-based argumentation mining (which has shown to be challenging, as it faces many problems [66]). In fact, the empirical tests of the algorithm run by Bernreiter et al. [40] were run using purely synthetic data, because, in practice, attack relations are incredibly hard to label at scale.

Additionally, while the regular algorithm has an exponential runtime, which makes it unsuitable for larger discussions, the greedy variant of the algorithm partially improves upon this runtime. However, this variant still includes an exponential-runtime precomputation step, run on the set of comments, that makes this algorithm infeasible for *Polis*-like deliberations, where participants can add comments during the discussion.

Nevertheless, this algorithm is still worth considering, if only with a slightly modified deliberation design with separate comment-gathering and voting phases, separated by some time. Between these phases, the precomputation step can often be performed by relatively fast Answer Set Programming (ASP) solvers, since it is only provably costly in the worst case. Furthermore, as AI and NLP continue to make advances in the coming years, mining attack relations from natural language comments will become increasingly feasible. In fact, this is one of the stated goals of the FAME[2] Project [66].

In the meantime, our setup, with its fixed comment sets, mitigates some of these issues. Additionally, for each of our data sets, the attack relations between comments have been *painstakingly* labelled by hand. Since we provide this information when constructing the deliberation, we do not need to implement any natural language processing (NLP) to mine argumentations from the data. Most importantly, all of this allows us to precompute preferred extensions, which makes the exponential running time of this step less of an issue for our construction.

In this study, we use the greedy-harmonic variant of the argumentation algorithm. As discussed above, we choose the greedy variant for better real-world performance. Furthermore, we choose the harmonic variant, because, as noted by Bernreiter et al. [40], it is a good compromise between efficiency and fairness, allowing us to best represent the discussion at large, while still capturing the nuance of disparate viewpoints.

Additionally, just as with the ComSoC algorithm, this algorithm only considers explicit approval of comments, in contrast to Polis's consideration of both approvals and disapprovals. Luckily, we can use the same solution: for every comment $c \in C$, we introduce the comment $\overline{c}$, and record disapproval of c as approval of $\overline{c}$ (and again, passing is simply recorded as approval of neither). Note, it follows from this definition that c attacks $\overline{c}$ and vice versa.

Finally, this algorithm makes no attempt to minimize the number of comments each participant votes on; in fact, it assumes that every participant has voted on every comment, which is an unreasonable assumption for real-world discussions. While *Polis*'s analysis algorithm also makes this assumption, it extrapolates this data by inferring that a participant voted the mean vote on any comment they haven't yet seen. However, inferring an average position on every unseen comment for each participant would wreak havoc on the argumentation algorithm by making every participant seem to have a more inconsistent worldview

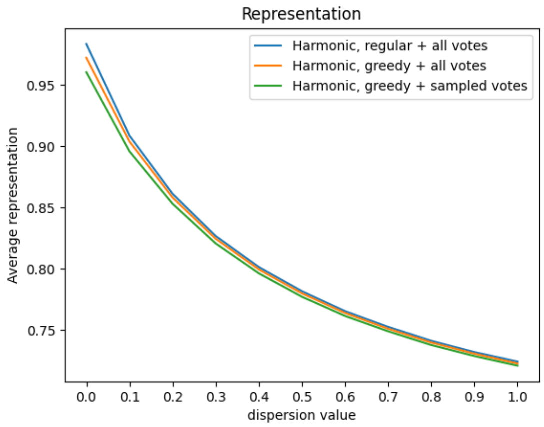
[2] A Framework for Mining and Formal Evaluation of Arguments

than they do in reality, which would reduce the effectiveness of the algorithm.

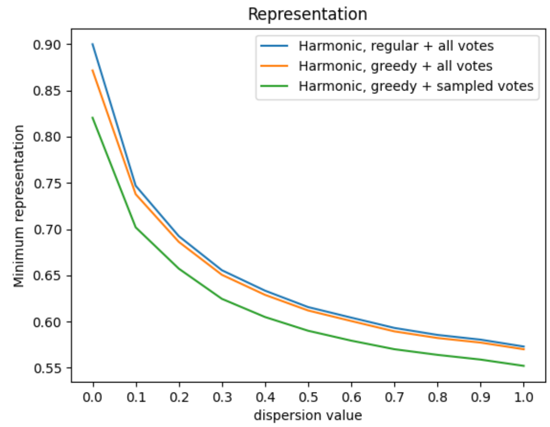
Fortunately, since the algorithm merely looks for preferred extensions that are most representative of participants' approval profiles, a smaller approval profile only makes this task easier. Therefore, for the final algorithm, we can simply assume that the participants do not approve of any comments they have not seen. Even so, we still need to make sure that we have a thoroughly accurate sample of each participant's preferences. Thus, to improve this algorithm's practical applicability, we create the following comment routing algorithm.

Inspired by Polis's comment-routing strategy, we design an algorithm that also randomly chooses comments according to a non-uniform distribution, informed by the structures unique this algorithm. At the beginning, we consider the undirected attack relation graph, where each comment c represents a node, and for any two comments c, c' such that either c attacks c' or c' attacks c, there is an edge between their corresponding nodes. We divide the comments into connected components, which are sections of this graph that are connected to each other by these undirected attack relation edges. Since only comments that disagree about a particular issue are connected by attack relations, each connected component can be thought to represent a different dimension of the topic being discussed. Also, for each participant, we keep track of the number of comments they approve of in each connected component.

Then, when a participant requests a comment to vote, we randomly choose a connected component from which to provide them a comment to vote on, among those for which they have voted the least approvals. In this way, we make sure that we equally capture the participant's positions on all dimensions of the issue. Once we have chosen a connected component, we randomly choose a comment from it to present to the participant by sampling from a non-uniform distribution on these comments, which is formed by a chosen priority metric. This priority metric promotes comments a participants is more likely to agree with, seeks to build consensus, and highlights lesser-seen comments. In particular, a comment c in the connected component is chosen with probability $\textsc{Priority}(c)$, normalized by the sum of such values for all comments in the connected component.



(a) Comparing algorithms by average representation



(b) Comparing algorithms by minimum representation

Figure 3.2: Comparison of modified Argumentation algorithms

Let $P_v(c)$ be defined as in Chapter 2.1.1, an estimate of the probability that a participant voted v on comment c; $N(c)$ be as defined in Chapter 2.1.1, the total number of votes on comment c; and $Nbr_a(c)$ be the number of neighbors of c (i.e. adjacent comments) in the undirected attack graph that this participant has approved of. Then

$$\textsc{Priority}(c) = P_{v=a}(c) \cdot (1 - P_{v=p}(c)) \cdot (1 + 2^{3-\frac{N(c)}{5}}) \cdot (2^{-\frac{Nbr_a(c)}{5}}) \ .$$

Like Polis, the equation is constructed so that each of the terms in the product has value greater than 1 for comments that should be sent more to participants, and value between 0 and 1 for comments that should not be shown as often. The $P_{v=a}(c)$ term is meant to boost consensus by promoting comments with higher agreement, and decreases to 0 for comments with little support. The $(1 - P_{v=p}(c))$ term decreases to 0 for comments that have been mostly passed on. The $(1 + 2^{3-\frac{N(c)}{5}})$ term helps highlight new comments by emphasizing those with fewer votes. Finally, the $(2^{-\frac{Nbr_a(c)}{5}})$ term deprioritizes comments that have an attack relation with comments the participant has already approved, which are comments that the participant is likely to disagree with.

Again, though we cannot guarantee the optimality of our modifications, we *can* empirically evaluate their performance. In their paper, Bernreiter et al. [40] experimentally test the representativeness of their algorithms using synthetic data spanning a range of dispersion values (which measure the proportion of the generated voters' votes that differ from a consistent viewpoint), an evaluation method that we can replicate. As discussed in Chapter 2.3.1, we can measure how represented each voter is by the algorithm's generated outcome; and two important metrics to evaluate the algorithm are the average and minimum representations of voters under this outcome. The harmonic variant of the Argumentation algorithm performs well under both of these metrics, which is why we choose to use this variant in our trials, so we focus on this variant in our evaluation.

In particular, we consider the performance of the greedy algorithm when votes are sampled using our comment routing algorithm, against that of both the greedy and non-greedy algorithms when all votes are used. As in the paper [40], we compare these algorithms according to the aforementioned metrics using synthetic data spanning a range of dispersion values. For both average (Figure 3.2a) and minimum (Figure 3.2b) representation, we find that our comment routing algorithm does not significantly decrease the representativeness of the greedy-harmonic Argumentation algorithm. This affirms our choice of this algorithm to accurately capture participants' preferences even while only sampling their votes.

However, the algorithm still has another shortcoming, for which we need to make additional modifications. Since the outcome Ω consists of multiple preferred extensions—each of which can often include more than half of all comments—we need to truncate these viewpoints to be able to display them for our final summaries. Otherwise, reading these summaries might take more effort than reading the entire comment set, which defeats the entire point of having summaries. We choose to show only the top comments that are shared amongst viewpoints, as well as those unique to each viewpoint. While truncating viewpoints in this way loses the property that each viewpoint's presented summary is defendable, each is still importantly conflict-free.

Once again, we still note that these modifications do not support allowing participants to respond to more than the initial queries, even if they want to contribute more to the discussion (and by the same token, participants cannot be assessed as part of the conversation if they can only respond to less than the initial queries). Although our comment-routing algorithm maintains no hard limit to the number of comments it can give to each participant, the analysis algorithm does not properly support variable levels of engagement. As this algorithm seeks only to maximize the proportion of each participant's approval profile included in a viewpoint (weighted according to an OWA vector), such variance in engagement would cause the algorithm to give more significance to the opinions of participants who vote more often, which makes this 'solution' untenable. Once again, the underlying algorithm itself would need to be tweaked to avoid this bias. While it may be theoretically possible to add this functionality, we do not consider this as it is outside the scope of this project. After all, since every participant in our study votes on the same number of comments, this lack of flexibility does not affect our experiment. Still, we again note that this limitation is somewhat antithetical to the goals of deliberative democracy, which seek to encourage, not limit, citizen participation.

Chapter 4

Experimental Design

In this chapter, we outline the setup of our main experiment. As a reminder, our goal in this study is to facilitate the incorporation of citizen input into the political decision-making process by improving the usefulness of *Polis*'s deliberative online platform design. To accomplish this, we focus on enhancing the subjective experiences of users with both the voting process and the algorithmically-generated summarization of the discussion. In order to test these outcomes, we have designed the following experimental methodology.

In particular, across three different deliberation topics, we study the effect of three treatments[1]—the Polis algorithm, the ComSoC algorithm, and the Argumentation algorithm—on the subjective experience of the platform user. We recruit about 900 participants into this study from a nationally representative sample of the US population, and randomly assign them to one of the treatment conditions, such that 100 participants are in each condition for each data set. Participants are recruited to this study using Prolific[2], a platform that connects researchers with high-quality research participants. In the first round, each of the participants is given comments to vote on according to their respective algorithm, after which they are asked to rate how useful the presented comments were for expressing their

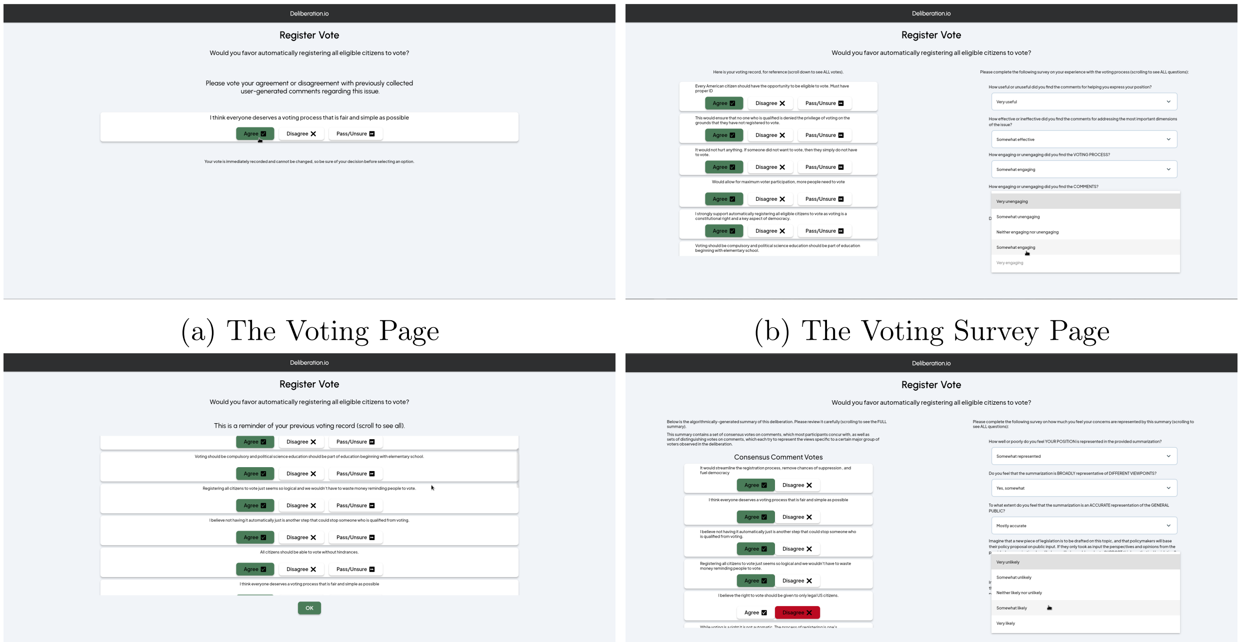
[1] The code for our back end implementations of the algorithms is available at https://github.com/jrenriquez/deliberation.io , specifically, the Python files prefixed with "adaptive_".

[2] https://prolific.com

viewpoint. When all votes are collected, the respective algorithm is run to produce the representative “summary" set of comments, and participants are then asked (in the second round) to rate how much they feel their concerns are represented by the summary, as well as how they would have felt if the summary was the main feedback taken from this discussion as input by policy-makers. (These are just a selected few of the questions that were asked; to see the full list of ten survey questions and multiple-choice answers for each, see Appendix A.1)

To make the comparison as direct as possible, we control for the different comments that could generate in each dialogue by using three pre-selected bodies of 50 comments each that each discussion begins with. In particular, these comment sets were chosen to represent the breadth of perspectives found in each of the original sets of 300+ comments that were collected from a representative sample of the U.S. population in a previously conducted survey (See Appendix A.2 for the comment sets used). The discussion topics—Insurrection Act, Register Vote, and Abortion—were chosen due to the varying levels of disagreement and consensus seen in this prior survey, to better test the algorithms in a variety of situations. Insurrection Act, which involves responses to whether the Insurrection Act should be invoked for peaceful protests during the next inauguration, was chosen for its large consensus; Abortion, which involves responses to whether the length of pregnancy should inform the legality of abortion, was chosen for its diversity of opinions; and Register Vote, which involves responses to whether voter registration should be automatic, was chosen for its being in between.

Although this is slightly less accurate for modeling a real deliberative discussion where participants also submit comments, we are only focusing on the subjective experiences of participants with these algorithms, not the entire platform. Thus, this consistency of comments is actually beneficial, as it allows us to more directly compare outcomes on the same group of data when analyzed by different algorithms. Though each trial has a slightly different distribution of participant positions as each has different participants, the treatments



(a) The Voting Page

(b) The Voting Survey Page

(c) The Voting Record Reminder Page

(d) The Summary Survey Page

Figure 4.1: The pages we designed for this study on Deliberation.io

are still comparable, as participants are randomized into each condition.

Another upside of this approach is that it allows us to pre-process the data to label attack relations between comments, which is necessary for the argumentation algorithm (See Appendix A.3 for these attack relations, which were labeled by hand).

4.1 Platform Design

This study is carried out using Deliberation.io, a new open source discussion platform for research that allows us to implement these three algorithms, as well as interject surveys to gauge participants' experience with the three paradigms.

In particular, we have designed four different key pages for participants to interact with, two for each round. For the first round, on the comment voting page (Figure 4.1a), participants are shown 20 comments, one at a time, to vote their agreement or disagreement on. On the voting survey page (Figure 4.1b), they are shown a reminder of the comments they just voted on on the left, and asked to complete survey questions about the voting process on the right. Meanwhile, when participants return to the platform for the second round, they are first reminded of their previous voting record on the voting record reminder page (Figure 4.1c), to re-familiarize themselves with the deliberation and their stated position. Finally, on the summary survey page (Figure 4.1d), they are presented with the algorithmically-generated summarization of the discussion on their left and asked to fill out a survey about this summarization on their right.

Chapter 5

Results and Discussion

In this chapter, we discuss the results from our trials. Recall that we had nine trials, one per algorithm per comment set. Furthermore, we had asked ten different survey questions to gauge participants' subjective experiences with the platform. Of these, we choose to highlight the three questions (one question regarding the comment routing of the voting phase of the discussion, as well as two questions regarding the usefulness of the algorithmically-generated summarization of the deliberation) that we feel are most pertinent to our exploratory goals. In the following, we explore the distribution of participant responses to these questions and discuss what this data can tell us about these different algorithmic approaches (To see visualizations of answers to all 10 questions, see the omitted graphs in Appendix B).

Before we analyze the responses to our study, we note our response rates to contextualize our results. In the first round of our study, where users participated in the deliberation by voting on comments, then answered questions regarding the voting process, we had a 92% response rate. This resulted in just over 100 (specifically, between 103 and 105) participants responding for each of the nine treatments. Of these, about 75% (specifically, between 74 and 82 participants) returned to participate in the second round of the study, where they responded to questions regarding the algorithmically-generated summarization of the discussion.

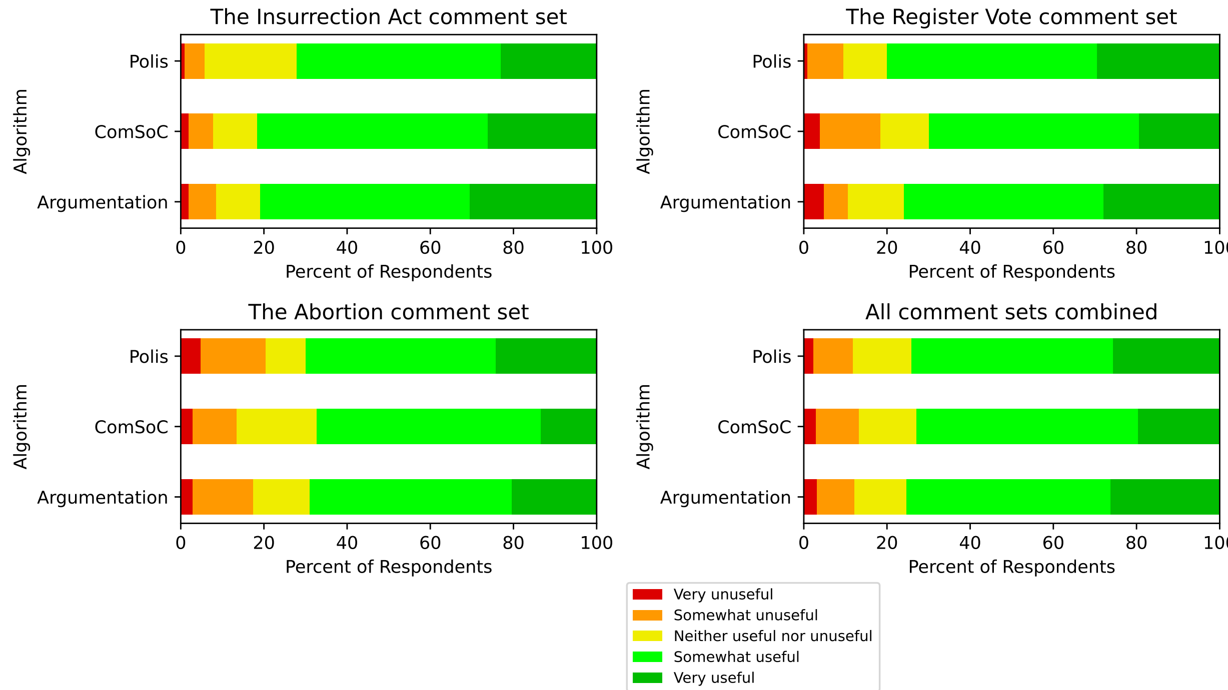


Figure 5.1: Usefulness of comment routing algorithms for expressing personal position.

5.1 Comment Routing

Usefulness for expressing personal position

First, we consider the experiences of users with the voting process. Indeed, to encourage citizen participation in political decision-making, we first need to ensure that their experience with the platform does not dissuade them from participating. To measure this, we consider, among the three algorithms studied, the perceived usefulness of each approach's method for routing comments, through responses to the following survey question: "How useful or unuseful did you find the comments for helping you express your position?"

As evidenced by Figure 5.1, all three algorithms across all three data sets were either "Somewhat useful" or "Very useful" in this regard for about 70-80% of participants. By a two-sample t-test, the only treatments with a statistically significant difference (using a 95% confidence interval) are those of Polis and ComSoC in the Register Vote data set; thus, it is fairly straightforward to conclude that the choice of comment routing algorithm does not significantly impact participants' perceived usefulness of the comments for expressing their preferences.

5.2 Summarization

Now, we arrive at our main object of study in this project: namely, how participants subjectively feel about the algorithmically-generated summaries of the discussion that are provided by the platform. We consider both the perceived representation of participants' positions in the summarization, as well as their support for legislation using the summarization as input.

Representation of participants' positions

If we truly want to foster the incorporation of citizen feedback into the policy-making process, we need to encourage participation through a platform that makes people feel that their voices are being heard. To this end, we look at responses to the following survey question: "How well or poorly do you feel YOUR POSITION is represented in the provided summarization?"

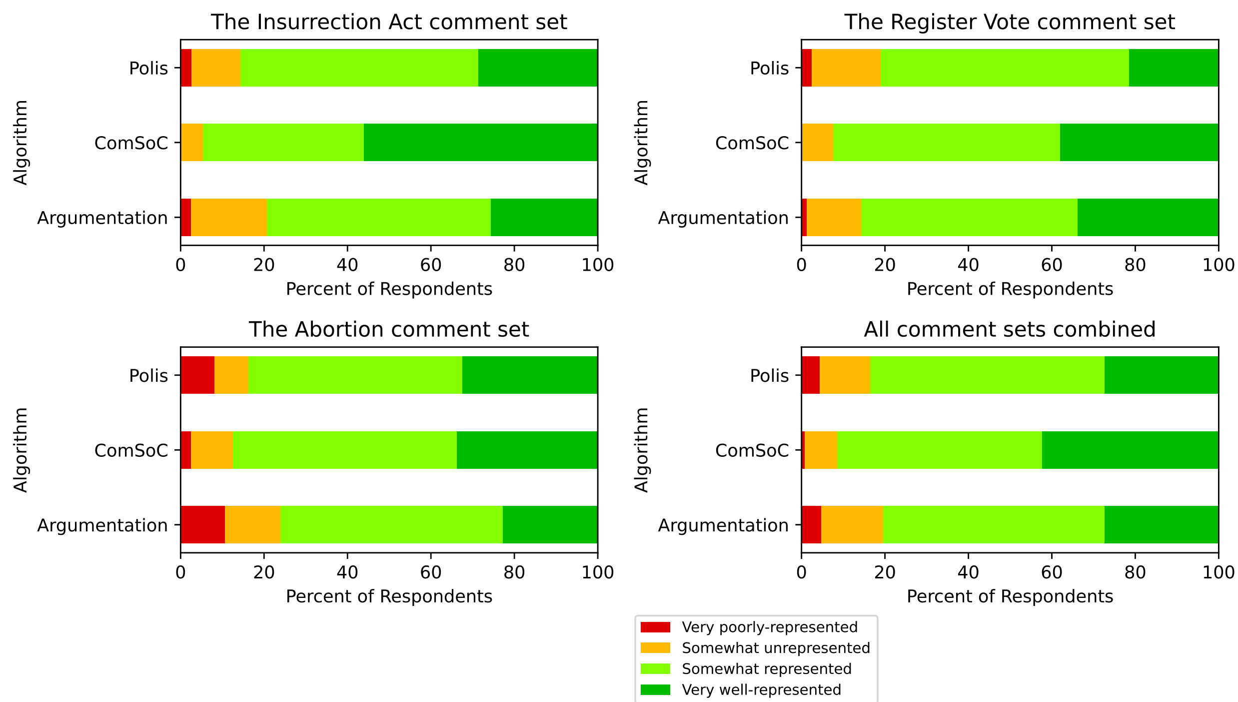


Figure 5.2: How well participants felt their positions were represented by the summarization.

Algorithm A	Algorithm B	Data Set
ComSoC	Polis	Insurrection Act
ComSoC	Argumentation	Insurrection Act
ComSoC	Polis	Register Vote
ComSoC	Argumentation	Abortion
ComSoC	Polis	Overall
ComSoC	Argumentation	Overall

Table 5.1: Statistically significant (by a 95% confidence interval) differences in representation between algorithms as determined by a two-sample t-test. For each line, we find that Algorithm A results in participants feeling more represented by its algorithmically-generated summarization than Algorithm B for the given data set (or overall), by a statistically significant margin.

Here, again (Figure 5.2), we have a largely positive result, finding that all three algorithms, across all three datasets, produced summarizations that made about 80-90% of participants feel "Somewhat represented" or "Very well-represented". However, as noted in Table 5.1, the ComSoC algorithm quite consistently results in participants feeling their positions are more represented by the summarization than either of the other algorithms, by

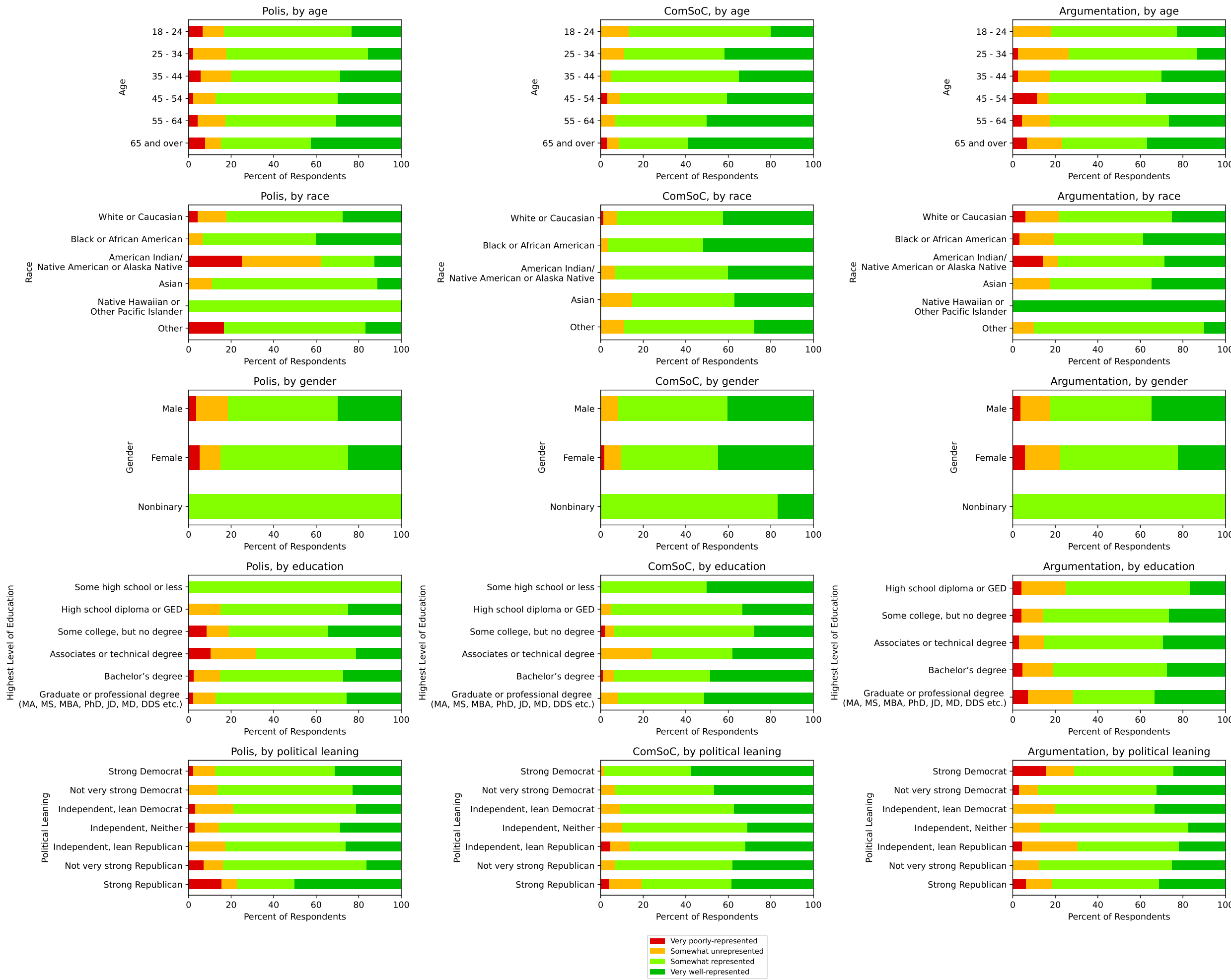


Figure 5.3: How well participants felt their positions were represented by the summarization, by demographic category.

a statistically significant margin.

Indeed, it is not surprising that the ComSoC algorithm performs well here; after all, it was specifically designed to achieve high representativeness and satisfaction guarantees. Still, it is notable that these guarantees have transferred over from the objective assurances of theory to the subjective experiences of real life.

Algorithm	Demographic Category	Group A	Group B
Polis	Race	Black or Afri...	Asian
ComSoC	Age	55 - 64	18 - 24
ComSoC	Political Leaning	Strong Democrat	Independent, Neither
ComSoC	Political Leaning	Strong Democrat	Independent, Lean Republican

Table 5.2: Statistically significant (by a 99% confidence interval, excluding outliers; since this analysis requires us to split up the data more, we require a higher level of confidence, and exclude outliers) differences in representation by these algorithms between demographic categories as determined by a two-sample t-test. For each line, we find that the given algorithm results in participants from Group A feeling more represented by its algorithmically-generated summarization than those from Group B, by a statistically significant margin.

Furthermore, as shown in Figure 5.3, we find that all three algorithms, across demographic splits spanning age, race, gender, education, and political leaning, produced summarizations that made about 70-95% of participants feel "Somewhat represented" or "Very well-represented". Unfortunately, as shown in Table 5.2, the ComSoC algorithm is somewhat more likely than the other algorithms to result in a statistically significant difference in perceived representation across demographic categories. Still, we note that even among this variance, the ComSoC algorithm usually achieves better representation than the other two.

Participants' support for legislation using this input

Another important outcome of deliberative platforms is that they provide a useful means of gathering citizen input to inform policy-making. To this end, we look at responses to the following survey question: "Imagine that a new piece of legislation is to be drafted on this topic, and that policymakers will base their policy proposal on public input. If they only took as input the perspectives and opinions from the provided summarization, how likely or unlikely would you be to SUPPORT this hypothetical legislation?"

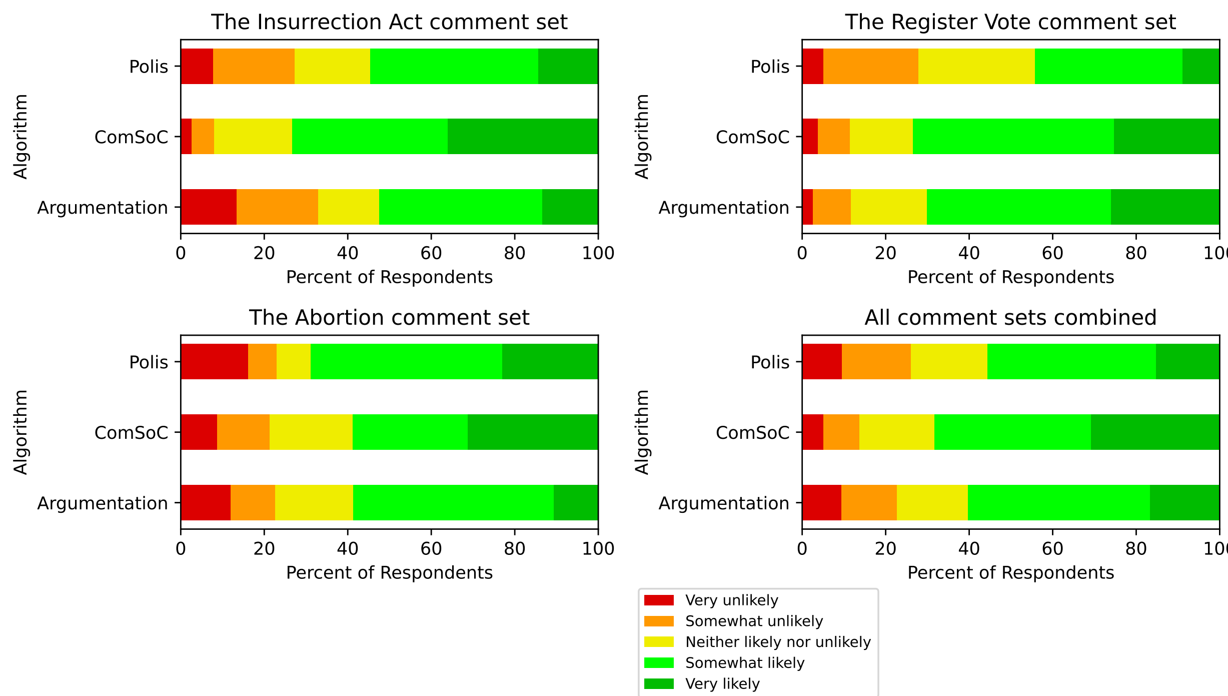


Figure 5.4: How participants would have felt if the algorithmically-generated summarization were the main feedback taken from this discussion as input by policy-makers.

Algorithm A	Algorithm B	Data Set
ComSoC	Polis	Insurrection Act
ComSoC	Argumentation	Insurrection Act
ComSoC	Polis	Register Vote
Argumentation	Polis	Register Vote
ComSoC	Polis	Overall
ComSoC	Argumentation	Overall

Table 5.3: Statistically significant (by a 95% confidence interval) differences in support for hypothetical legislation between algorithms as determined by a two-sample t-test. For each line, we find that Algorithm A results in participants having greater support for hypothetical legislation informed solely by the algorithmically-generated summarization than Algorithm B for the given data set (or overall), by a statistically significant margin.

Yet again (Figure 5.4), we have a fairly positive result, finding that all three algorithms, across all three datasets, produced summarizations that had about 55-70% of participants "Somewhat likely" or "Very likely" to support hypothetical legislation informed by these summaries. However, as noted in Table 5.3, the ComSoC algorithm quite consistently results in participants having greater support for hypothetical legislation informed solely by the summarization than either of the other algorithms, by a statistically significant margin.

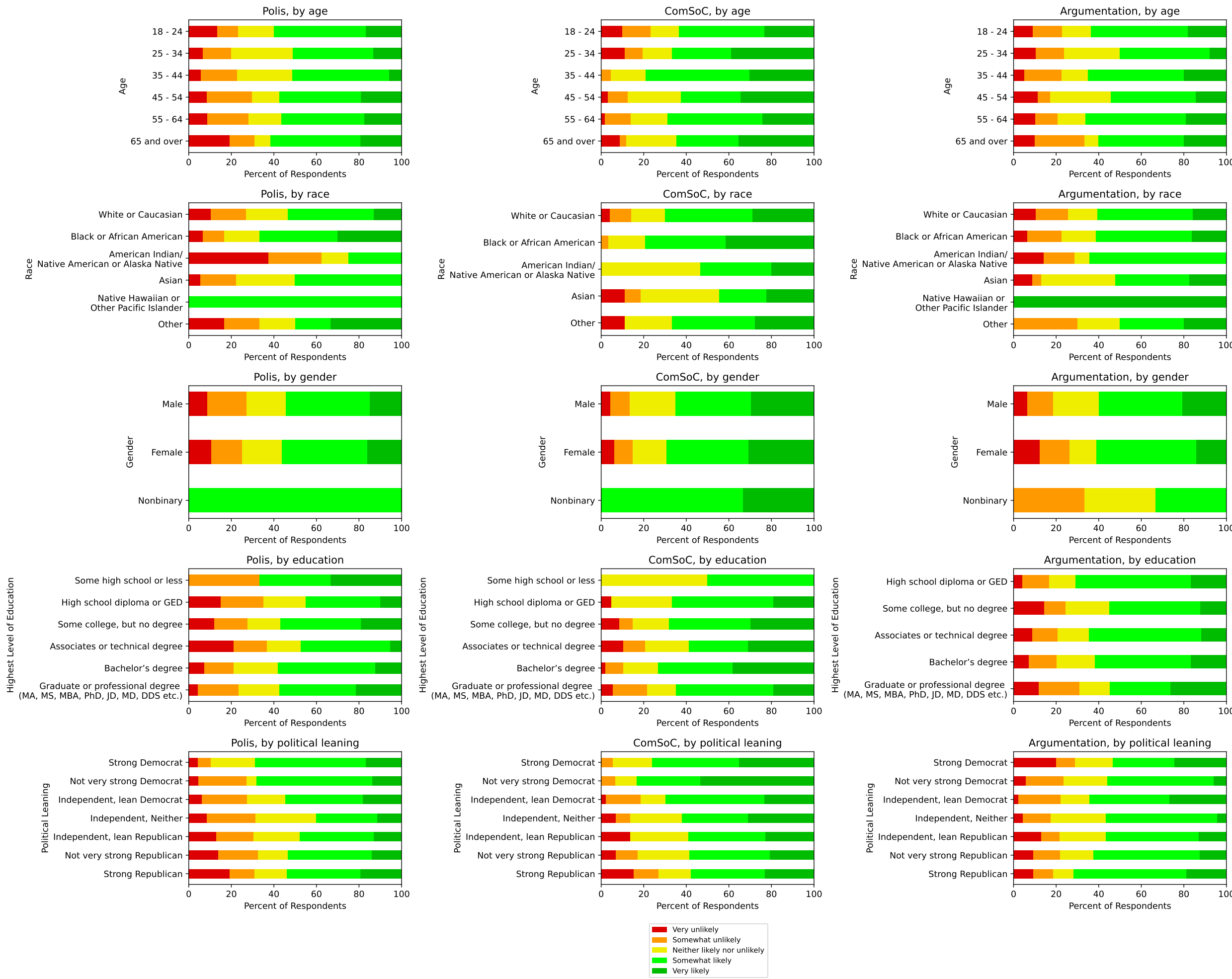


Figure 5.5: How participants would have felt if the algorithmically-generated summarization were the main feedback taken from this discussion as input by policy-makers, by demographic category.

Algorithm	Demographic Category	Group A	Group B
ComSoC	Race	Black or African American	Asian
ComSoC	Political Leaning	Not very strong Democrat	Not very strong Re...
ComSoC	Political Leaning	Not very strong Democrat	Strong Republican

Table 5.4: Statistically significant (by a 99% confidence interval, excluding outliers; since this analysis requires us to split up the data more, we use a higher level of confidence, and exclude outliers) differences in representation by these algorithms between demographic categories as determined by a two-sample t-test. For each line, we find that the given algorithm results in participants from Group A having greater support for hypothetical legislation informed solely by its algorithmically-generated summarization than those from Group B, by a statistically significant margin.

Furthermore, as shown in Figure 5.5, we find that that all three algorithms, across demographic splits spanning age, race, gender, education, and political leaning, produced summarizations that that had about 45-75% of participants "Somewhat likely" or "Very likely" to support hypothetical legislation informed by these summaries. Once again, as shown in Table 5.2, the ComSoC algorithm is somewhat more likely than the other algorithms to result in a statistically significant difference in support for hypothetical legislation across demographic categories. Again, we note that even among this variance, the ComSoC algorithm still usually achieves higher levels of support than the other two.

5.3 Key Takeaways

In general, we found that while all three approaches provided an essentially equivalent voting experience, the ComSoC algorithm consistently resulted in participants feeling more represented by, and more likely to support legislation informed by, its provided summarizations. Indeed, this seems to be exactly what we seek, namely, an improvement in the usefulness of *Polis*'s deliberative online platform design, to facilitate the incorporation of citizen input into the political decision-making process.

However, this does not immediately merit a replacement of the *Polis* platform's current algorithms. For one, we still caution that though it might have been slight, the ComSoC algorithm exhibited the greatest variance in these metrics across demographic splits among these three algorithms. More importantly, we note that since we completely reimplemented the Polis algorithm for use in this study, we acknowledge that its under-performance in this study could be due to a mistake in our implementation.

Moreover, even if we were to recommend the ComSoC algorithm's use on the *Polis* platform, there are still barriers to its real-life applicability. Namely, as it currently stands, the algorithm does not support allowing participants to vote on a variable number of comments (note that it *does*, however, support the ability for participants to submit comments during the voting phase [39]). Until this shortcoming is addressed, the ComSoC algorithm cannot be employed in a practical setting.

Meanwhile, as its performance was largely comparable to the existing Polis algorithm, there is little incentive to recommend the use of the Argumentation algorithm. In fact, due to the many challenges we faced in utilizing this algorithm, we generally discourage its real-world use. For one, labelling attack relations in order to use this algorithm is a long and tedious process to carry out by hand. Also, trying to crowdsource this task by splitting it among participants presents an undue burden on users that would depress participation. Lastly, as noted previously, there is not currently an automated way to mine arguments from natural language.

Even if attack labelling was a solved problem, we would still not recommend the Argumentation algorithm due to its excessive running time. Using the more efficient greedy variant, a conversation still requires separate comment submission and voting phases, as the algorithm needs to run a costly precomputation step between these phases. Furthermore, this step does not at all scale for larger conversations. Even for the medium-sizes deliberation that we simulated in this study, running this computation step took multiple days on a modern processor, a performance that is infeasible for practical applications.

Chapter 6

Conclusion

Summary

In this work, we have sought to promote greater citizen participation in the political decision-making process through the use of online deliberation platforms like *Polis*. In particular, the algorithmically-generated summarizations of these discussions can be used to incorporate citizen input into the policy-making process. However, since these selections form the backbone of citizen feedback that policy-makers consider, we investigated newly proposed algorithms that seek to provide more representative (ComSoC) and consistent (Argumentation) summarizations than those currently provided by Polis.

After conducting randomized controlled trials of these three algorithms across multiple different topics, we found that the ComSoC algorithm consistently resulted in participants feeling more represented by, and more likely to support legislation informed by, its provided summarizations than Polis's current approach. While this indeed represents exactly the kind of improvement we sought, to foster greater adoption of deliberation platforms for democratic processes, we note that the ComSoC algorithm still requires future theoretical work to make it fully suitable for a real-world setting.

Future Work

While our results showed that the ComSoC algorithm meaningfully improved upon the Polis algorithm in producing summaries in which participants felt represented, a major hurdle still preventing its real-world applicability is its lack of support for allowing participants to vote on a variable number of comments. Thus, a primary direction for future work would be to build on our modifications to the ComSoC algorithm to produce a theoretically sound algorithm that allows this flexibility.

Another possible topic for future study would be to investigate the performance of a hybrid approach that combines the Polis and ComSoC algorithms. Indeed, while the ComSoC algorithm provided the most representative summarizations, these could potentially still be improved by using the Polis algorithm's analyses to inform the grouping of comments by viewpoint. For instance, such an approach could use primarily the ComSoC algorithm for comment routing (with any potential ties broken according to Polis's priority metric), and show as summary the committee of comments that ComSoC produced, but group these comments according to the clusters from Polis's analysis.

Appendix A

Data

A.1 Survey Questions

We collected the following multiple-choice survey questions while participants used our platform.

A.1.1 Voting Survey

Table A.1: Voting Survey.

Begin of Table	
Survey Question	Response Options
How useful or unuseful did you find the comments for helping you express your position?	Very unuseful Somewhat unuseful Neither useful nor unuseful Somewhat useful Very useful
How effective or ineffective did you find the comments for addressing the most important dimensions of the issue?	Very ineffective Somewhat ineffective Neither effective nor ineffective Somewhat effective Very effective
How engaging or unengaging did you find the VOTING PROCESS?	Very unengaging Somewhat unengaging Neither engaging nor unengaging Somewhat engaging Very engaging
How engaging or unengaging did you find the COMMENTS?	Very unengaging Somewhat unengaging Neither engaging nor unengaging Somewhat engaging Very engaging

Continuation of Table A.1	
Survey Question	Response Options
Do you think the group of people who made these comments tend to lean left or right politically?	Lean strongly right Lean somewhat right Lean neither particularly left nor right Lean somewhat left Lean strongly left Lean both left and right

A.1.2 Summarization Survey

Table A.2: Summarization Survey.

Begin of Table	
Survey Question	Response Options
How well or poorly do you feel YOUR POSITION is represented in the provided summarization?	Very poorly-represented Somewhat unrepresented Somewhat represented Very well-represented
Do you feel that the summarization is BROADLY representative of DIFFERENT VIEWPOINTS?	No, not much at all No, not that much Yes, somewhat Yes, very much
To what extent do you feel that the summarization is an ACCURATE representation of the GENERAL PUBLIC?	Mostly inaccurate Somewhat inaccurate Neither accurate nor inaccurate Somewhat accurate Mostly accurate
Imagine that a new piece of legislation is to be drafted on this topic, and that policymakers will base their policy proposal on public input. If they only took as input the perspectives and opinions from the provided summarization, how likely or unlikely would you be to SUPPORT this hypothetical legislation?	Very unlikely Somewhat unlikely Neither likely nor unlikely Somewhat likely Very likely

Continuation of Table A.2	
Survey Question	Response Options
Imagine that a new piece of legislation is to be drafted on this topic, and that policymakers will base their policy proposal on public input. If they only took as input the perspectives and opinions from the provided summarization, how likely or unlikely would you be to ACCEPT this hypothetical legislation?	Very unlikely Somewhat unlikely Neither likely nor unlikely Somewhat likely Very likely

A.2 Comment Sets

First, we include the three comment sets used in our study. Each consists of 50 comments, chosen to represent the breadth of perspectives found in the original sets of 300+ comments that were collected from a representative sample of the U.S. population.

A.2.1 Register Vote

Table A.3: Register Vote.

Begin of Table Would you favor automatically registering all eligible citizens to vote?	
Index	Comment
0	I somewhat favor this for convenience purposes; however, people deserve to not be inundated with political mail.
1	The government is by the people, for the people. So all should make their voice heard. But some might coerce others to vote against wishes.
2	It would greatly simplify things but some people do not wish to be registered because it means they will be called up for jury duty
3	You cannot force someone to vote, but not having that extra step may be more of an incentive vs. figuring out registering.
4	With modern technology this seems like it should be feasible without compromising the integrity of elections.
5	I dont 100% oppose all people being auto registered, but I do oppose any lumped in secondary requirements or repercussions of not voting.
6	Every American citizen should have the opportunity to be eligible to vote. Must have proper ID
7	Registering citizens to vote will help with voter fraud.
8	This would ensure that no one who is qualified is denied the privilege of voting on the grounds that they have not registered to vote.
9	I think it's a great idea. Not everyone has the ability or the $ on hand to get to a DMV to buy an official Photo id.

Continuation of Table A.3	
Would you favor automatically registering all eligible citizens to vote?	
Index	Comment
10	It would not hurt anything. If someone did not want to vote, then they simply do not have to vote.
11	It would make it easier for people to vote if they were to all be registered automatically.
12	Would allow for maximum voter participation, more people need to vote
13	I strongly support automatically registering all eligible citizens to vote as voting is a constitutional right and a key aspect of democracy.
14	Voting should be compulsory and political science education should be part of education beginning with elementary school.
15	Registering all citizens to vote just seems so logical and we wouldn't have to waste money reminding people to vote.
16	There are people who don't understand the process or don't have time to get a voter registration card. Automatically registering is fair.
17	I believe not having it automatically just is another step that could stop someone who is qualified from voting.
18	I should not be required to register to exercise a right guaranteed by the Constitution.
19	All citizens should be able to vote without hindrances.
20	participating in the democratic process is not only a right, but a responsibility and civic duty for those living in a democracy
21	I think everyone deserves a voting process that is fair and simple as possible
22	It would streamline the registration process, remove chances of suppression , and fuel democracy
23	Citizens of most European countries are automatically registered to vote. This makes the process easier, more efficient and accountable.
24	This would be so efficient. It would certainly be cost effective. It would be easier for the people, too.
25	If all eligible voters are not registered, then only a small amount of, typically white, voices are being heard.
26	Automatic voter registration would boost turnout and voter participation, so why wouldn't we do it?
27	This would significantly assist historically disenfranchised groups vote
28	Getting to a polling place to vote is already a hurdle for many voters, an additional, prior step doesn't need to be necessary.
29	It's kind of intimidating process at first i think a lot more people would vote if it was auto
30	Such a thing might impel more people to vote, and think about the importance of voting. It would also remove barriers to registration.
31	I would prefer that whomever is voting is voting because they want to and not because they were automatically registered
32	While voting is a right it is not automatic. The process of registering is one's responsibility

Continuation of Table A.3	
Would you favor automatically registering all eligible citizens to vote?	
Index	Comment
33	Not everyone wants to vote. It feels like wasted resources for those individuals who choose not to vote.
34	I believe the right to vote should be given to only legal US citizens.
35	Having people that are motivated and informed enough to register to vote is a good thing.
36	That seems a bit unlawful to automatically register people.
37	I believe that this country is based on freedom, which includes the freedom to choose to vote or not.
38	It might go against someone's beliefs and culture of not voting and it must be accepted.
39	There may be challenges in maintaining accurate and up-to-date voter rolls.
40	If this was just for American citizens, I would 100% support this. But illegal immigrants shouldn't be able to vote. They broke the law.
41	Automatic registration will lead to widespread fraud. People need to show proof of identity to be registered to vote.
42	Some people who may be qualified to vote still shouldn't because they don't understand the consequences of implications of their votes
43	I don't want Alzheimer's or dementia patients voting. I don't want addicts and mentally ill people voting.
44	This violates privacy. If the person does not wish to affiliate with a party, then this information could be used to go to the Dem/Rep
45	Automatic registration would be rife with fraud. People should take initiative and register themselves.
46	People should not be forced to vote if they don't want to vote.
47	Registering to vote also requires selective service registration.
48	If someone chooses not to vote that is also their right, votes don't matter anyway because electors are the ones who actually decide election
49	Citizens who can't be bothered to register to vote are unlikely to make well informed decisions.

A.2.2 Insurrection Act

Table A.4: Insurrection Act.

Begin of Table	
The Insurrection Act of 1807 allows the president to use military force against US civilians during a rebellion. Would you support invoking the Insurrection Act if there are peaceful protests when the next president takes office?	
Index	Comment
0	I think it's good policy to have, in case something gets out of control It's a way to protect others.
1	If the protesters are violent then yes use military force if needed
2	There needs to be a police presence just in case, but if everyone behaves, no action should be taken.
3	I think that with the mental health of people nowadays, and their idea of peaceful protest are getting dangerous and they need to be stopped
4	protesters are a waste of time and should be banned
5	force has to be used to maintain control., there is no other option
6	I think it's important to use them if they are getting out of hand we can't let the transition from peace to violence happen
7	Peaceful protests are not an insurrection, however rioters should be dealt with but peaceful protestors should not be harmed.
8	Peaceful protests are fine but most do not stay that way and many are intentionally violent
9	It depends on how bad. If its peaceful but controversial such as pro palestine, then no. If people are hurting each other then yes
10	If a protest is peaceful, it's fine if it exists. When protests become violent, I would think they could step in
11	Storming the halls of Congress merited the use of military force but not a peaceful protest
12	If the protest is peaceful, then I am against it. If the protest becomes a riot, then I am partial to it
13	No I feel like you should be able to control yourself with the results and what happened
14	While I believe peaceful protest can sometimes do harm, I don't see military action needed for those protest
15	I don't support it because using the military against citizens would lead to abuse. Only in very extreme circumstances.
16	If the protests are peaceful then it's not a rebellion therefore not a valid use of the insurrection act. Police can deal with troublemakers.
17	If the protest is peaceful, then there is no reason to invoke the Insurrection Act. The act is to be used in the case of a rebellion only.
18	A Peaceful protest or even a mildly disorderly one is not an insurrection.
19	I don't think the president should have the right to stop me from speaking when I have something important to say.
20	Suppressing the people isn't going to make them not want to protest. Martial law should be used during emergencies only not for protest.

Continuation of Table A.4	
The Insurrection Act of 1807 allows the president to use military force against US civilians during a rebellion. Would you support invoking the Insurrection Act if there are peaceful protests when the next president takes office?	
Index	Comment
21	Allowing peaceful protests is a sign of a healthy democracy so using military to prevent it is a sign of an unhealthy democracy.
22	I do not believe the brute force would be required for peaceful protestors.
23	I do not trust the judgment of the government or those involved.
24	There needs to be a way to protest without the extreme consequence of use of military force.
25	Without protests there would be no rights. Peaceful protest or other.
26	I would not support invoking the Insurrection Act against peaceful protests as it undermines constitutional rights to free speech and assembly
27	Invoking the insurrection act on peaceful protestors is a one way street towards the erosion of legal privileges.
28	A peaceful protest is not one that deserves a violent response as no violence is taking place,
29	Disallowing people to express their opinions and views and demonstrate them can lead down a slippery slope
30	People should be able to protest without threats of violence from the government.
31	I am completely against using force or violence or anything in that nature against civilians regardless of their stance.
32	It is never the job of the government to use force against US citizens, if they are not destroying property/violent then leave them alone
33	If people are peacefully protesting, police violence should not be invoked.
34	We should be able to overturn government or authorities that do not serve is
35	Such an Act violates the rights of Americans. Invoking such Act could escalate a protest leading to further unrest among citizens.
36	We have the right to protest for something we feel is wrong, especially when it comes to the government.
37	excessive use of federal power to suppress peaceful dissenting citizens is inherently undemocratic
38	Peaceful protest is a fundamental of modern democracy, being able to oppose government policy is essential to a working democracy.
39	All people in a democracy have the right to a voice, whether assenting or dissenting.
40	Free speech and peaceful demonstrations are a cornerstone of the US.
41	I don't believe using (potentially) lethal military force against peaceful protesters is justified, ever.
42	People have a right to peacefully protest. Military force against our own people would be terrible.
43	You should not use military force on your civilians

Continuation of Table A.4	
The Insurrection Act of 1807 allows the president to use military force against US civilians during a rebellion. Would you support invoking the Insurrection Act if there are peaceful protests when the next president takes office?	
Index	Comment
44	Military force is not needed for a peaceful protest. If emergency services can move through the area & people are safe, then let them be.
45	I think using the military against civilians that are not an active threat is veering into dictator territory
46	The president should not have unilateral power to break up peaceful protests because it infringes on our rights.
47	Peaceful protests are a method of showing those in power that we the people care about a specific subject, in a way that can't be missed.
48	Peaceful protests are not a rebellion. We don't need overreaction in real life, in addition to online spaces.
49	Protesting is an important part of American culture and tradition and you cannot silence voices

A.2.3 Abortion

Table A.5: Abortion.

Begin of Table	
Would you favor or oppose considering how long a woman has been pregnant in determining whether it is legal or illegal to have an abortion?	
Index	Comment
0	I am not on with aborting creatures that can feel, but bundles of cells is fine to exterminate.
1	If a pregnancy has advanced to the point the fetus is viable outside the women's body then I believe it has acquired human rights
2	After a certain period of time the baby is alive and at that point I would consider it murder.
3	Late term abortions should only be legal if the mother's life is in danger or the fetus is not viable.
4	Grey area, if passed the term that is a human, however it's not in the world yet.
5	There should be a cut-off for when safe, viable pregnancies could be terminated.
6	It could be one consideration among many, and only in the very late stages.
7	There is a point in pregnancy where I say that a woman should have decided earlier on if they wanted to discontinue it
8	Late term abortions are cruel to fully formed fetuses, but early term fetuses are not developed enough to be considered humans.
9	It is more harmful if done later in pregnancy to the mother. Some women are severely harmed and likely sterilized

Continuation of Table A.5	
Would you favor or oppose considering how long a woman has been pregnant in determining whether it is legal or illegal to have an abortion?	
Index	Comment
10	There should be no restrictions when medically necessary. Late term abortions due to lack of planning should not available mainstream.
11	I would completely favor this option abortions should not be done after the woman is far along
12	It's worth having a discussion about when life begins and what constitutes being alive at what stage.
13	I think it's a good idea to factor in how far along a pregnancy is when deciding if abortion should be legal. This way, we can balance women
14	Abortion should be a right at least until viability. After that, the doctor and patient must determine next step based on health and risks
15	There is a point where abortions should not be had, but a lot of laws are overly strict about terms.
16	Abortion should be legal up until the fetus is viable.
17	If more than six weeks, no abortion should be allowed. A woman knows when she is pregnant before this time.
18	Above 3 months, I believe that there is no point to abort the child since it's now grown.
19	I oppose considering the length of pregnancy as the sole determining factor for the legality of abortion
20	Legality or being illegal is not the problem here when it comes to when a woman can have an abortion
21	we do not need more people. It is selfish and rude to others
22	Until a certain gestational age, genetic factors will not be able to be determined.
23	Although I am pro choice I think there are few circumstances where it's too late but I still believe it should be her choice
24	while I believe a woman has a right to choose, late term abortions should be guided by doctors to ensure the safety of the woman
25	There should not be a set rule as there could be a medical issue for mother or baby making abortion necessary. Let doctor & patient decide.
26	It shouldn't matter how long it's been as long as it's safe
27	Abortion should be legal if it is harming the fetus or the mother. Abortion should be legal in cases of rape. Abortion should be legal.
28	I feel strongly that what someone decides about their own pregnancy is none of my business and support choice.
29	I oppose having a limit to a legal abortion. And should ultimately be decided by the woman's right to determine.
30	Most bans on abortion that are based on length of pregnancy don't consider how long it can take to realize you're pregnant.
31	Women who choose to have late-term abortions usually have a really good reason to do so. It isn't something that is done lightly.

Continuation of Table A.5	
Would you favor or oppose considering how long a woman has been pregnant in determining whether it is legal or illegal to have an abortion?	
Index	Comment
32	I believe in autonomy, women know their own body better than me. But it also needs to be balanced with avoiding cruelty.
33	Women should be able to do whatever they can do with their bodies. No one should be able to tell them what to do with their body and future
34	Abortion fundamentally cannot be considered a legal issue.
35	Banning late term abortions could put mothers at risk, as late term abortions are usually performed when there is a serious health risk.
36	Deciding when it is ok for someone to have an abortion is nonsensical. It is a personal health decision no arbitrary date range can dictate.
37	Abortion is fully up to the woman that has to carry the baby and many women don't even know they are pregnant till after the new cut off date.
38	I think a woman should have totally control of her body no matter how far along she is.
39	There can be medical complications up until birth that require a woman to have an abortion
40	Late abortions are astonishingly rare, and there isn't a consensus on when a fetus becomes a human. The citizen should decide for herself.
41	Abortion is a choice that needs to be taken into consideration by women and their doctors not their government.
42	Life is important. While I believe in people having choices, I also feel that children have a right to live from conception.
43	I believe that a baby is human from the moment of conception but it is difficult to regulate pregnancies very early on
44	When having a baby, it should be planned. It is already a living human. If you can't commit, then abort right away. Be responsible.
45	No pregnancy should be allowed to abort unless it is from rape; or incest or if the mother's life is at risk during the pregnancy or delivery
46	It shouldn't be an afterthought. Human life begins from the moment of conception.
47	I feel that abortion shouldn't be legal at all. Regardless of how long the woman has been pregnant.
48	All abortions are murder, no matter how long gestation has been.
49	No matter how developed the baby is, it is still a baby and someone's sense of freedom does not take precedence over someone's life

A.3 Attack Relations

The attack relations between comments in each of these comment sets was painstakingly labeled by hand. These relations are provided below. Note that here, we refer to comments by their index, as noted above. Furthermore, comment j, for $j \geq 50$ simply represents the negation (disapproval) of comment $j - 50$ (in keeping with our modification to allow disapprovals to the Argumentation algorithm).

A.3.1 Register Vote

Table A.6: Register Vote.

Begin of Table	
Attacker	Comments Attacked
0	39, 41, 45, 50, 65
1	7, 10, 14, 32, 36, 42, 45, 51
2	32, 33, 36, 41, 44, 45, 52
3	1, 32, 33, 36, 38, 53, 61
4	31, 32, 36, 39, 41, 45, 46, 54
5	14, 36, 38, 41, 43, 45, 55, 74
6	36, 37, 38, 39, 41, 56, 69
7	31, 32, 36, 37, 38, 41, 44, 45, 57, 72
8	31, 32, 33, 36, 38, 39, 58, 67
9	31, 32, 36, 38, 39, 41, 44, 45, 59
10	36, 38, 60
11	39, 45, 61
12	36, 41, 62
13	36, 39, 45, 63
14	31, 33, 36, 64
15	33, 36, 65
16	32, 44, 66
17	32, 36, 58, 67
18	32, 36, 68
19	36, 41, 68, 69
20	31, 42, 43, 46, 63, 70, 71
21	36, 41, 42, 52, 71
22	36, 41, 45, 50, 72, 76
23	32, 36, 41, 45, 69, 73
24	32, 36, 38, 41, 57, 74, 76
25	32, 44, 45, 68, 75, 76
26	36, 38, 41, 45, 72, 74, 76
27	36, 41, 49, 76, 77
28	32, 35, 41, 45, 68, 77, 78
29	32, 36, 41, 45, 59, 61, 79
30	32, 36, 42, 45, 46, 65, 71, 80
31	12, 14, 19, 20, 21, 24, 71, 74, 81, 93
32	0, 13, 18, 25, 27, 52, 57, 82, 91
33	7, 8, 10, 14, 15, 24, 83, 85, 93, 96
34	14, 19, 20, 27, 82, 84, 86, 90

Continuation of Table A.6	
Attacker	Comments Attacked
35	12, 14, 15, 16, 19, 20, 81, 82, 85, 96
36	5, 9, 10, 13, 14, 15, 26, 82, 83, 84, 86, 89
37	1, 12, 13, 14, 20, 26, 81, 82, 83, 87, 96
38	12, 13, 14, 19, 20, 26, 82, 86, 88, 96
39	2, 4, 5, 7, 18, 27, 81, 84, 89
40	9, 10, 11, 20, 21, 25, 84, 86, 90, 99
41	2, 4, 7, 22, 24, 84, 86, 91
42	10, 12, 13, 14, 92, 95, 96
43	11, 13, 14, 19, 92, 93, 96
44	1, 10, 13, 19, 82, 94, 96
45	4, 7, 10, 22, 82, 84, 95
46	10, 14, 15, 30, 85, 96, 98
47	5, 10, 15, 94, 97
48	4, 14, 17, 96, 98
49	12, 14, 18, 99
50	0, 9, 57, 58
51	1, 53
52	2, 86, 89
53	3, 8, 51
54	4
55	5
56	6, 19
57	7, 91
58	3, 8, 22
59	9
60	3, 10, 26
61	8, 11, 24, 26
62	3, 12, 17, 26
63	13, 19
64	14
65	8, 9, 15
66	15, 16
67	8, 17
68	18, 19, 20
69	19, 21
70	18, 20
71	11, 19, 21
72	8, 22
73	23
74	15, 24
75	25, 26, 27
76	26, 75
77	27

Continuation of Table A.6	
Attacker	Comments Attacked
78	28
79	29
80	8, 11, 30
81	8, 19, 20, 31
82	32, 33
83	33
84	34, 40
85	35
86	36
87	31, 37
88	38
89	39, 41, 57
90	34, 40
91	6, 7, 41, 57
92	35, 42, 43, 44
93	42, 43
94	44, 46
95	32, 41, 45
96	46
97	47
98	48
99	49

A.3.2 Insurrection Act

Table A.7: Insurrection Act.

Begin of Table	
Attacker	Comments Attacked
0	14, 31, 41, 50
1	31, 41, 43, 51
2	31, 43, 52
3	14, 18, 21, 22, 26, 31, 32, 33, 41, 42, 43, 53, 60
4	18, 25, 26, 27, 28, 30, 31, 32, 33, 41, 49, 54
5	14, 19, 24, 25, 26, 30, 31, 37, 41, 42, 43, 46, 49, 55
6	31, 36, 41, 43, 56
7	25, 31, 43, 57, 83
8	25, 26, 31, 41, 43, 58
9	4, 5, 31, 41, 43, 59
10	14, 31, 41, 43, 60
11	31, 43, 61
12	14, 41, 43, 62

Continuation of Table A.7	
Attacker	Comments Attacked
13	5, 6, 63, 76, 81
14	4, 5, 6, 12, 50, 64, 76, 83
15	3, 4, 5, 31, 41, 43, 65, 76, 81
16	4, 5, 66
17	67, 72
18	4, 5, 57, 68
19	69
20	5, 31, 70
21	4, 5, 71, 88
22	3, 4, 5, 72, 91
23	0, 1, 3, 4, 5, 6, 7, 10, 12, 73
24	0, 1, 3, 4, 5, 10, 72, 74, 92
25	4, 75
26	3, 4, 5, 8, 57, 66, 72, 76, 81
27	5, 77
28	3, 5, 6, 8, 57, 78, 83
29	4, 79
30	0, 1, 2, 5, 6, 9, 10, 12, 57, 66, 76, 78, 80, 81, 82, 92
31	0, 1, 2, 5, 6, 7, 8, 9, 10, 12, 65, 67, 72, 81, 93
32	3, 5, 6, 9, 10, 12, 82, 96
33	3, 5, 6, 12, 57, 62, 67, 76, 78, 83, 91
34	0, 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 84
35	0, 1, 3, 5, 6, 9, 10, 12, 72, 81, 85, 91
36	1, 3, 4, 5, 6, 8, 10, 12, 80, 86, 96
37	3, 5, 8, 72, 83, 87
38	3, 4, 5, 80, 86, 88
39	4, 5, 80, 81, 89, 99
40	4, 64, 69, 90, 99
41	1, 2, 4, 5, 6, 64, 72, 83, 91, 93
42	4, 5, 6, 11, 72, 83, 91, 92
43	0, 1, 3, 5, 6, 9, 10, 11, 12, 68, 72, 81, 91, 93
44	1, 5, 6, 72, 74, 91, 93, 94
45	2, 5, 77, 95
46	5, 69, 96
47	3, 4, 5, 69, 81, 86, 97
48	4, 5, 98
49	1, 3, 4, 5, 99
50	0, 6, 64, 81
51	0, 1
52	2
53	3
54	4, 81, 83
55	5, 81, 91, 93

Continuation of Table A.7	
Attacker	Comments Attacked
56	6
57	7, 18
58	8
59	9
60	10
61	11
62	12
63	13, 31, 41, 43
64	14, 22, 31, 33, 41, 43
65	15, 16, 22, 26, 41
66	16, 26, 31, 35, 41
67	16, 17, 28, 31, 41
68	18
69	19, 46
70	20, 35
71	17, 21, 38
72	22, 33, 41
73	23
74	24
75	25, 54
76	26, 31, 43
77	27
78	28
79	29
80	30
81	15, 22, 31, 43
82	31, 32, 43
83	30, 31, 33
84	34
85	35
86	36
87	21, 37
88	21, 38
89	39
90	40
91	15, 22, 30, 31, 41, 43
92	22, 31, 33, 41, 42
93	31, 33, 43
94	26, 44
95	45
96	19, 46
97	47
98	48

Continuation of Table A.7	
Attacker	Comments Attacked
99	49

A.3.3 Abortion

Table A.8: Abortion.

Begin of Table	
Attacker	Comments Attacked
0	19, 23, 25, 26, 28, 29, 33, 36, 37, 41, 47, 48, 50, 97, 98
1	25, 26, 29, 30, 33, 36, 37, 38, 51
2	19, 25, 26, 33, 37, 41, 52, 55
3	19, 26, 29, 32, 36, 37, 39, 53
4	54
5	19, 26, 29, 31, 32, 36, 37, 41, 55
6	29, 33, 36, 41, 48, 56
7	19, 25, 26, 28, 33, 36, 38, 42, 43, 49, 57
8	19, 26, 37, 43, 48, 58
9	59
10	26, 30, 36, 37, 41, 47, 48, 60
11	19, 61, 95, 97, 98, 99
12	62, 93
13	19, 41, 61, 63
14	47, 48, 64, 66, 76
15	26, 33, 38, 47, 48, 65, 66
16	29, 36, 43, 47, 48, 66
17	19, 26, 29, 37, 67
18	19, 26, 36, 68
19	5, 17, 56, 69
20	70, 78, 79, 84
21	71
22	72
23	2, 5, 10, 13, 14, 15, 47, 48, 73, 88
24	47, 48, 53, 55, 74, 77, 78
25	2, 5, 11, 17, 45, 47, 48, 60, 75, 79
26	0, 2, 5, 7, 12, 14, 47, 48, 75, 76, 78, 79, 83
27	11, 77, 81
28	2, 3, 5, 16, 45, 47, 48, 75, 78, 88
29	2, 3, 5, 16, 45, 47, 79, 88
30	3, 11, 13, 16, 45, 78, 80, 82, 91
31	1, 6, 11, 16, 53, 77, 79, 81
32	2, 5, 6, 15, 42, 45, 48, 82, 85, 86, 88
33	2, 3, 5, 11, 17, 42, 47, 48, 53, 60, 66, 83

Continuation of Table A.8	
Attacker	Comments Attacked
34	16, 17, 48, 64, 66, 84, 91
35	2, 11, 17, 18, 47, 48, 73, 74, 81, 85
36	2, 3, 5, 11, 16, 17, 42, 45, 47, 48, 73, 78, 86, 88, 91
37	2, 3, 5, 6, 17, 45, 47, 48, 73, 79, 87, 88, 90
38	2, 5, 11, 15, 16, 17, 45, 47, 48, 84, 88, 89
39	2, 5, 11, 48, 73, 79, 89, 90
40	48, 52, 69, 90
41	5, 47, 48, 73, 78, 88, 91
42	14, 92, 97, 98, 99
43	38, 68, 93, 97, 98
44	55, 94, 98
45	23, 26, 41, 95, 97, 98
46	14, 16, 38, 61, 96, 98
47	14, 16, 26, 96, 97, 98
48	5, 10, 14, 16, 26, 29, 37, 38, 41, 69, 97, 98
49	26, 33, 36, 97, 98, 99
50	0
51	1
52	2, 47, 48, 54
53	3, 35, 60
54	4, 52
55	5, 45, 48, 56
56	6, 76
57	7, 69
58	8
59	9
60	2, 5, 10
61	6, 8, 11
62	12, 17, 18
63	2, 11, 13, 47
64	14, 24, 41
65	15, 17, 18
66	16
67	17, 48
68	17, 18
69	19, 26, 29
70	20
71	21
72	22
73	23, 24, 28
74	24
75	25
76	26

Continuation of Table A.8	
Attacker	Comments Attacked
77	27
78	28, 36, 38
79	29, 33, 36, 38, 41
80	30
81	3, 10, 31, 35
82	29, 32, 33
83	33, 36, 38
84	34
85	27, 31, 35
86	36
87	37
88	38
89	39
90	40
91	41
92	42
93	43
94	44
95	45
96	46
97	15, 17, 18, 47, 48, 51, 66
98	1, 2, 5, 17, 48, 52
99	49

Appendix B

Omitted Plots

Here, we include the corresponding plots for the questions that were omitted from our discussion in the body of our work.

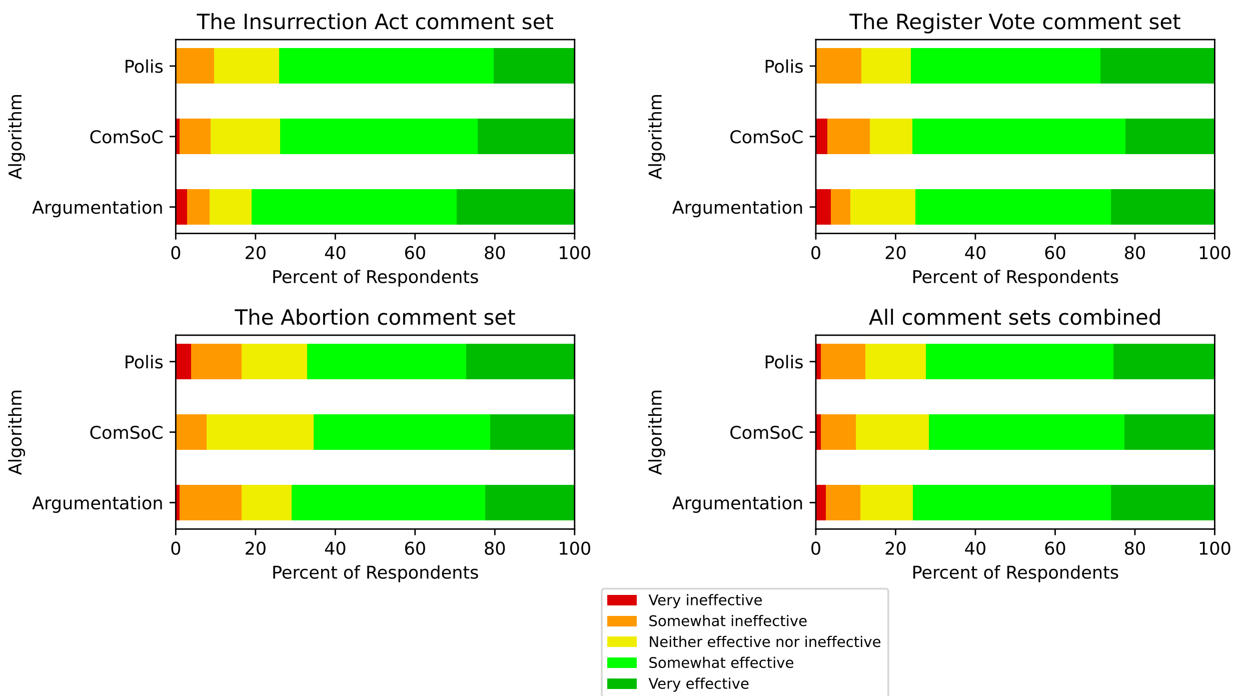


Figure B.1: How effective participants found the comments for addressing the most important dimensions of the issue.

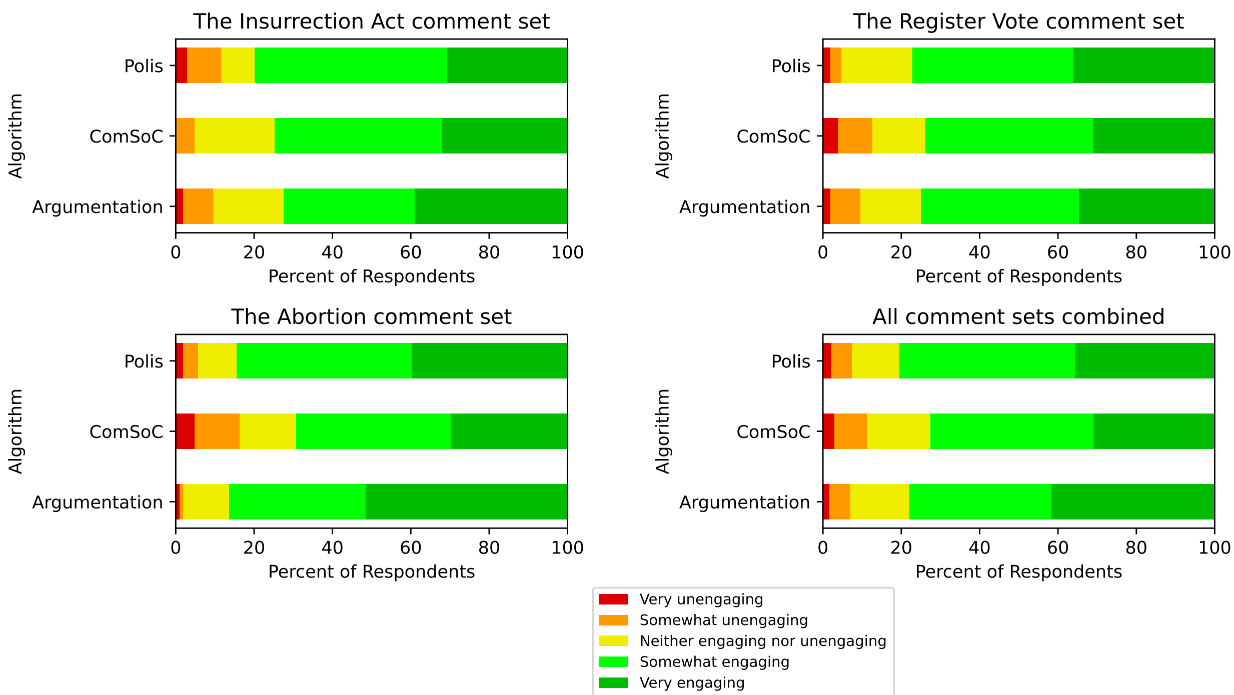


Figure B.2: How engaging participants found the voting process.

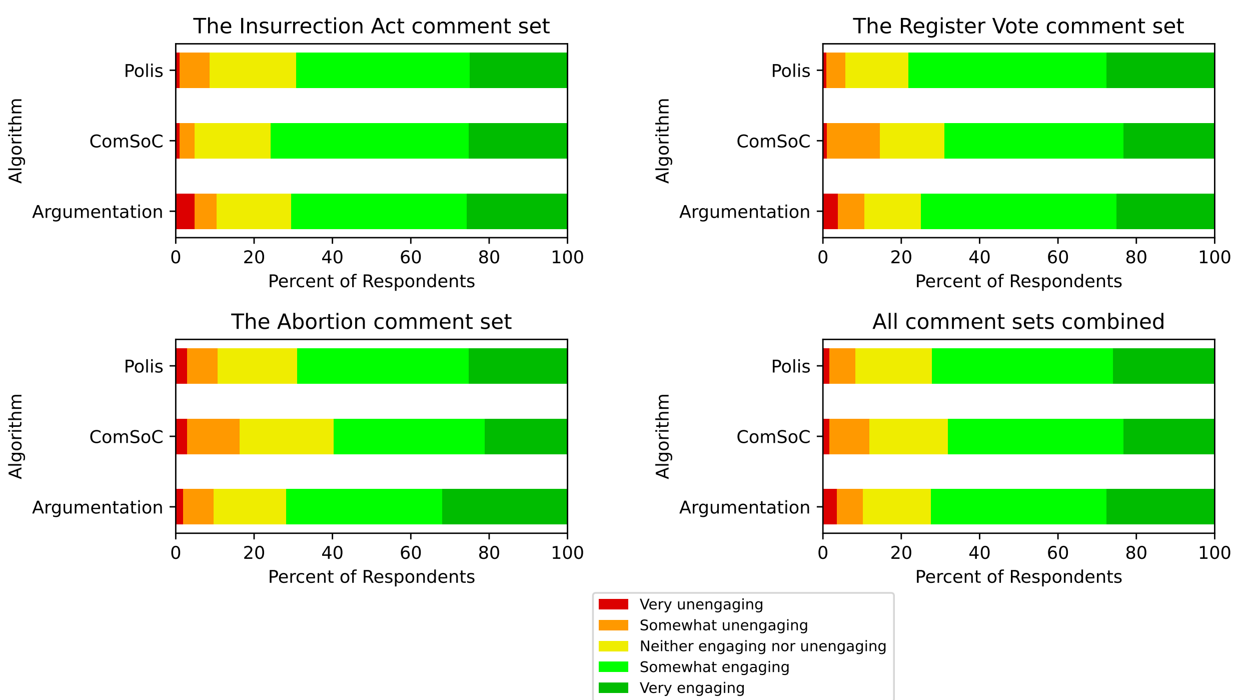


Figure B.3: How engaging participants found the comments.

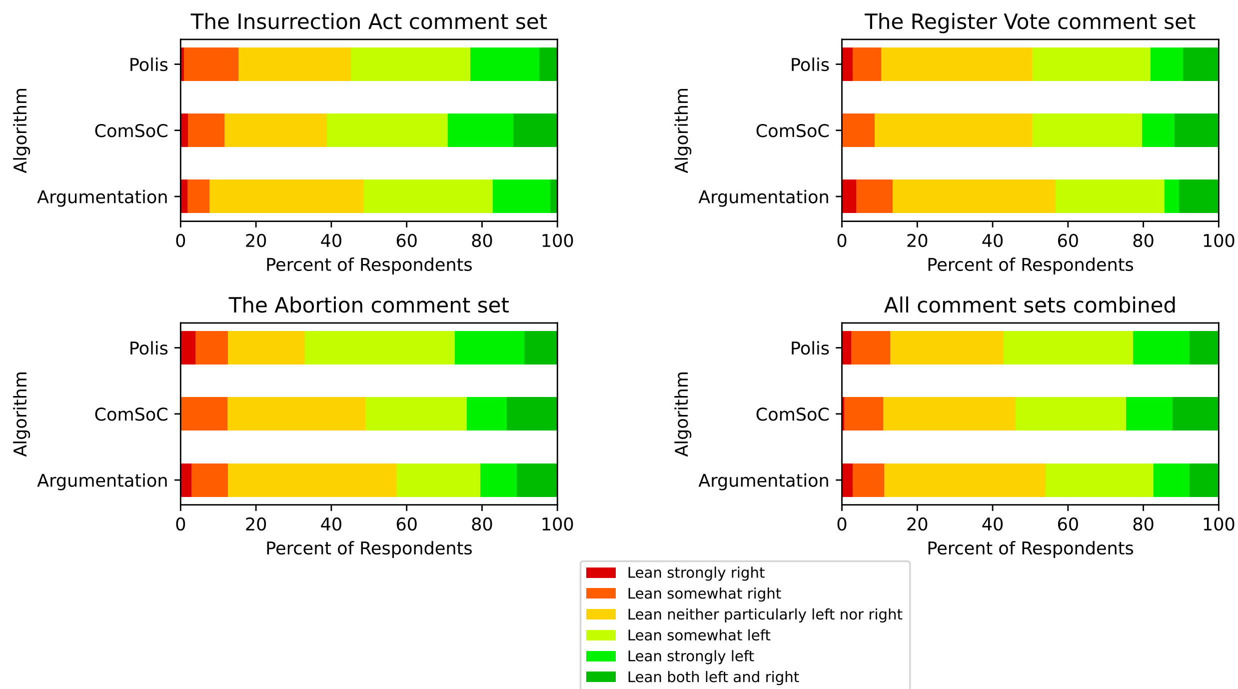


Figure B.4: The political leaning that participants felt of the group that made the comments.

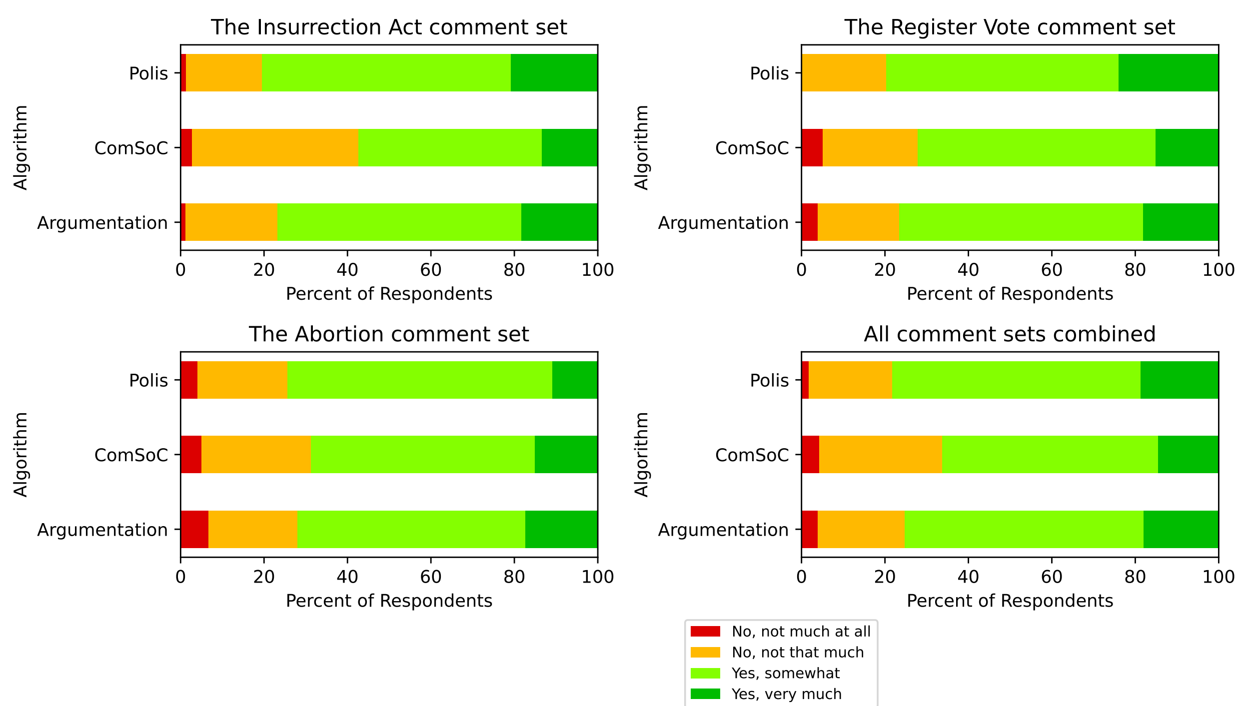


Figure B.5: If participants felt that the summarization was representative of different viewpoints.

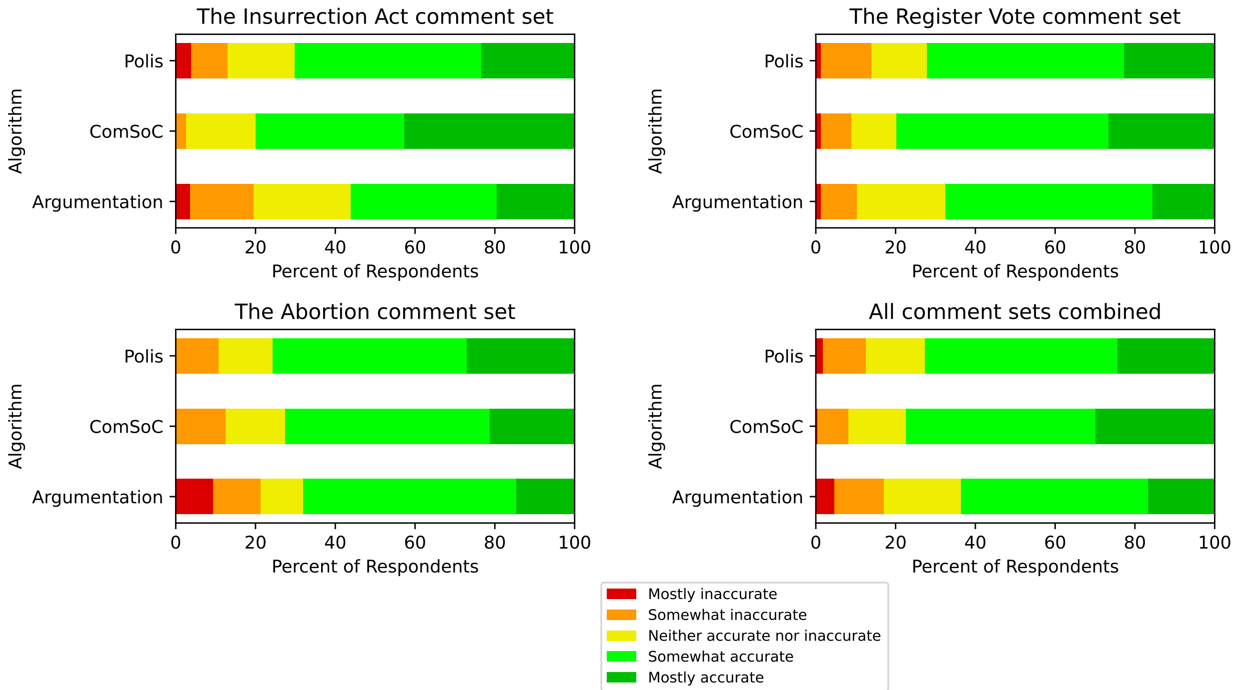


Figure B.6: The extent to which participants felt that the summarization accurately represents the general public.

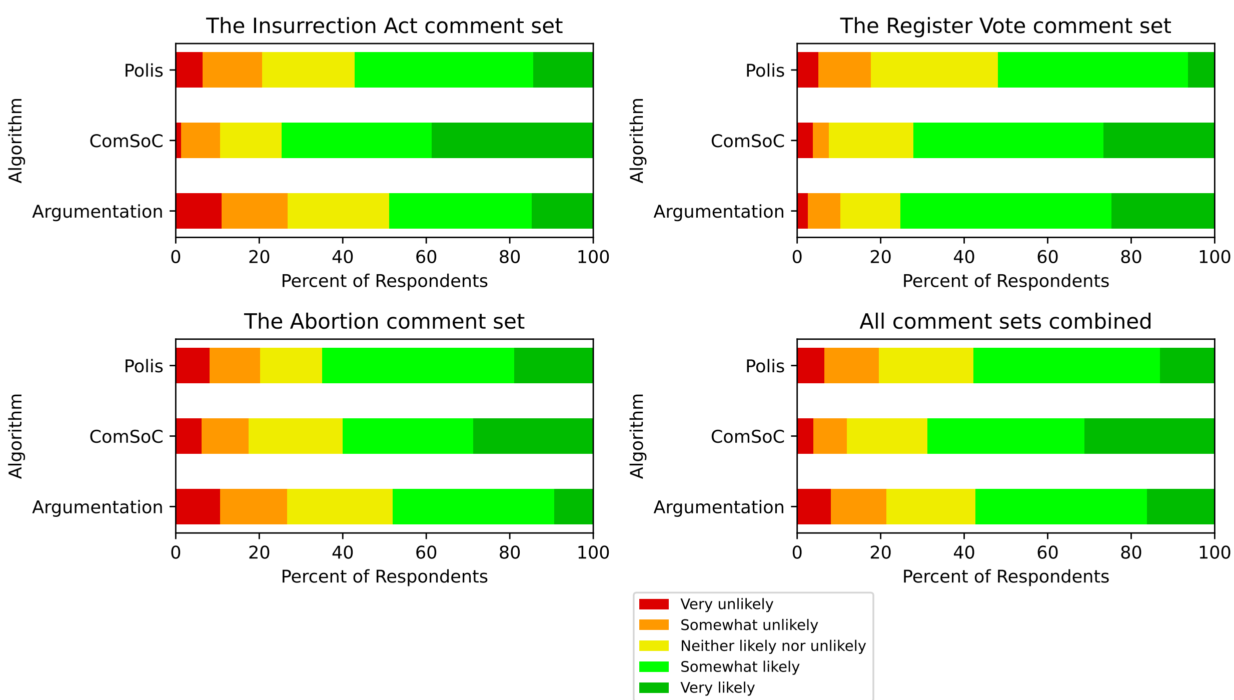


Figure B.7: How participants would have accepted that the algorithmically-generated summarization was the main feedback taken from this discussion as input by policy-makers.

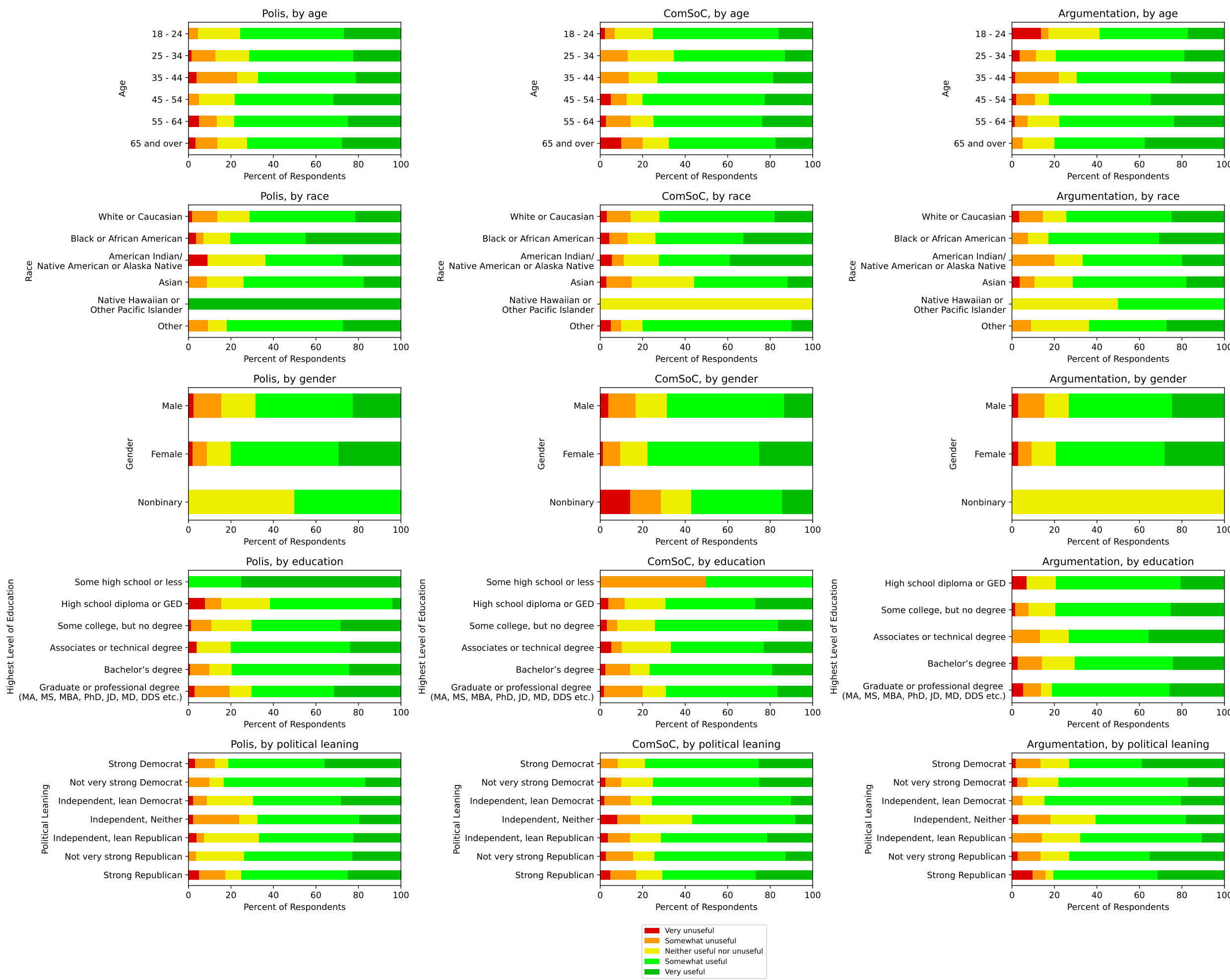


Figure B.8: Usefulness of comment routing algorithms for expressing personal position, by demographic category.

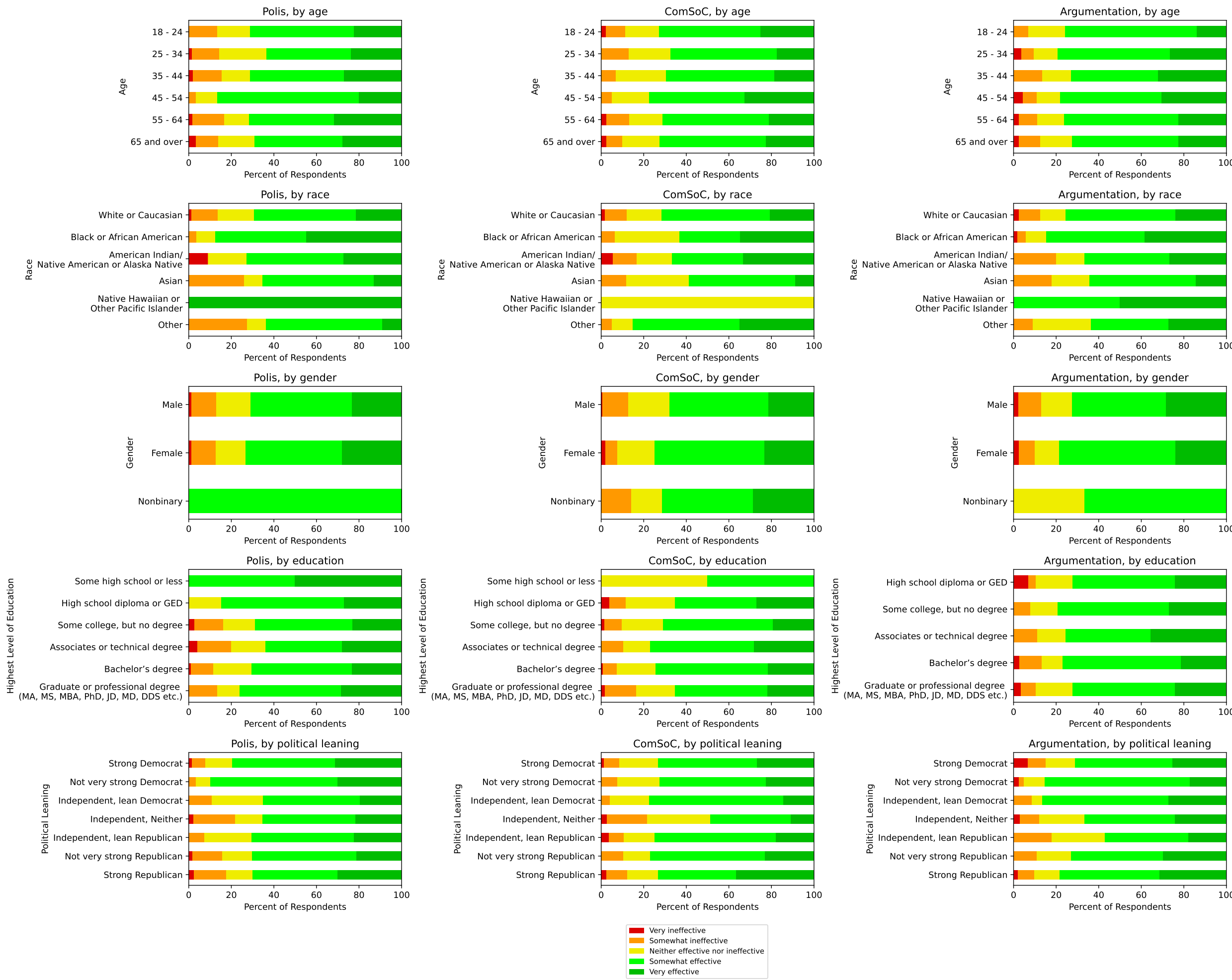


Figure B.9: How effective participants found the comments for addressing the most important dimensions of the issue, by demographic category.

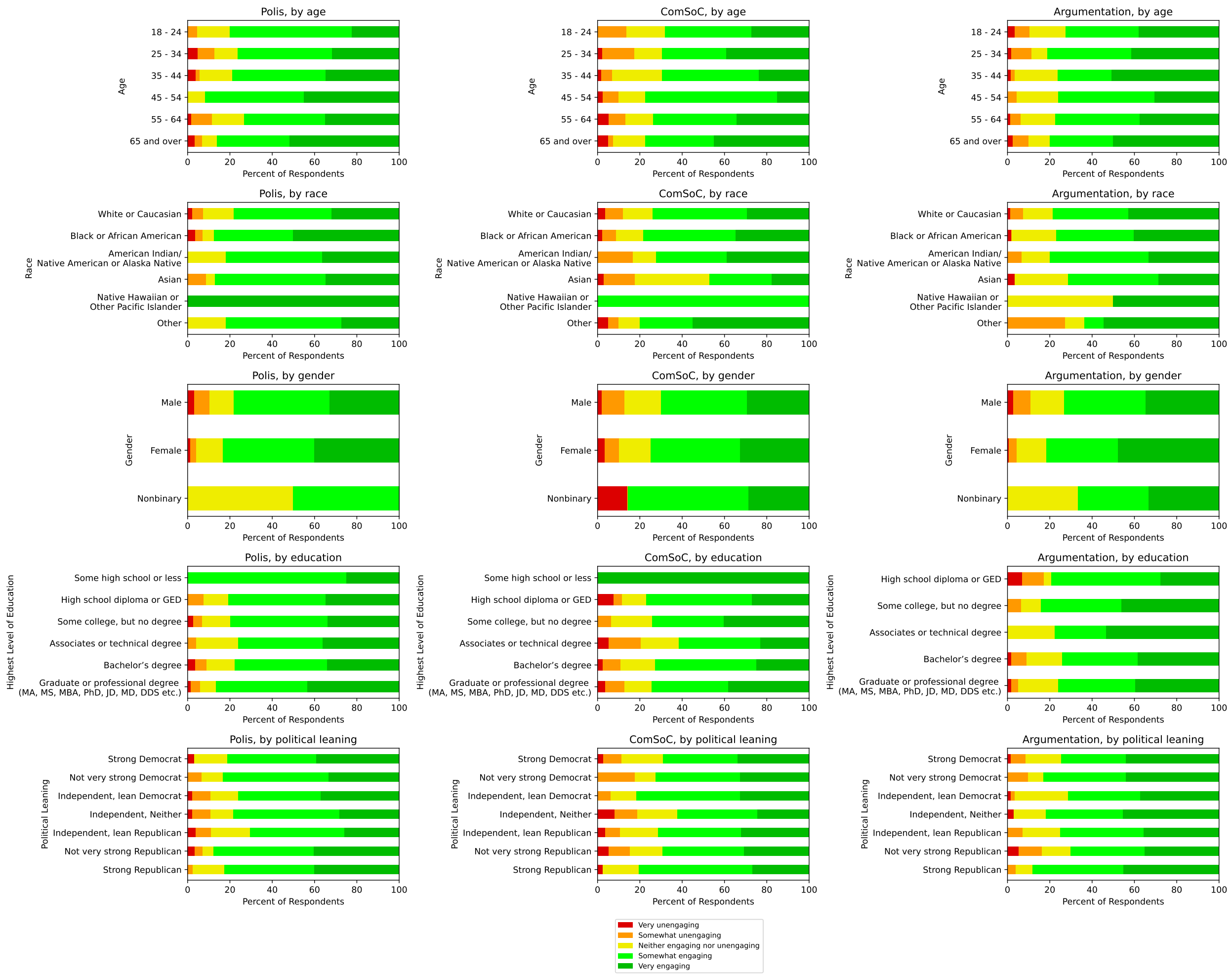


Figure B.10: How engaging participants found the voting process, by demographic category.

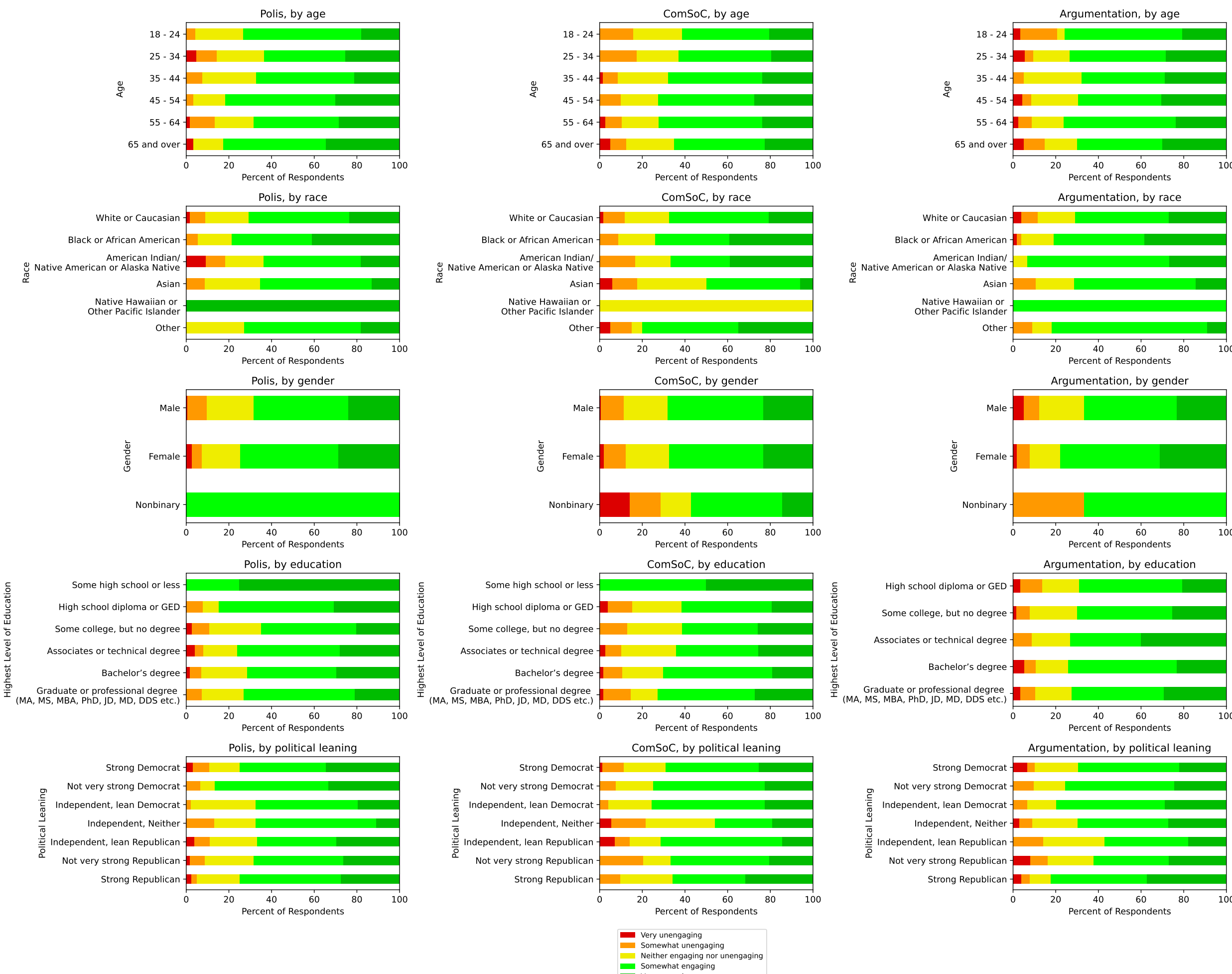


Figure B.11: How engaging participants found the comments, by demographic category.

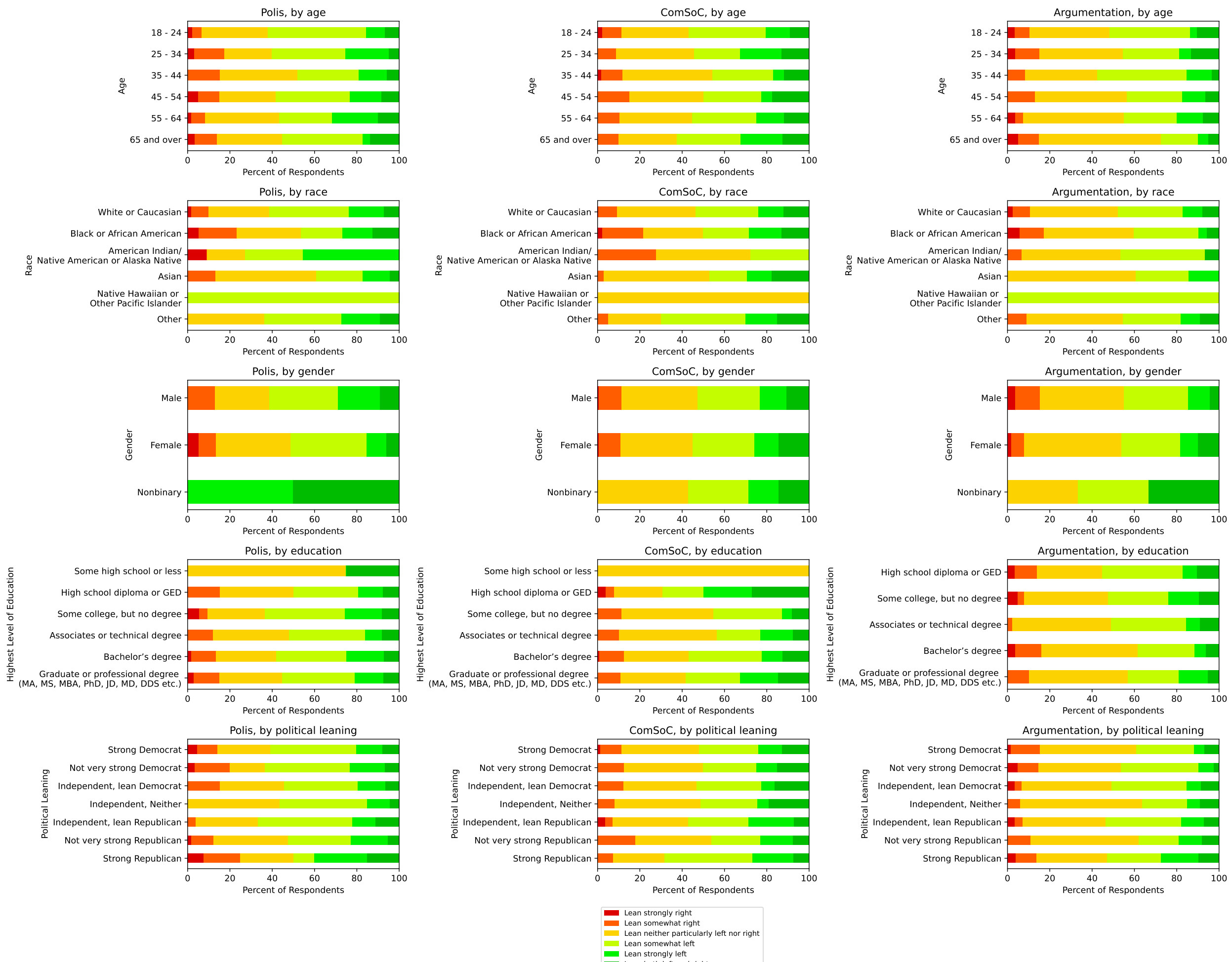


Figure B.12: The political leaning that participants felt of the group that made the comments, by demographic category.

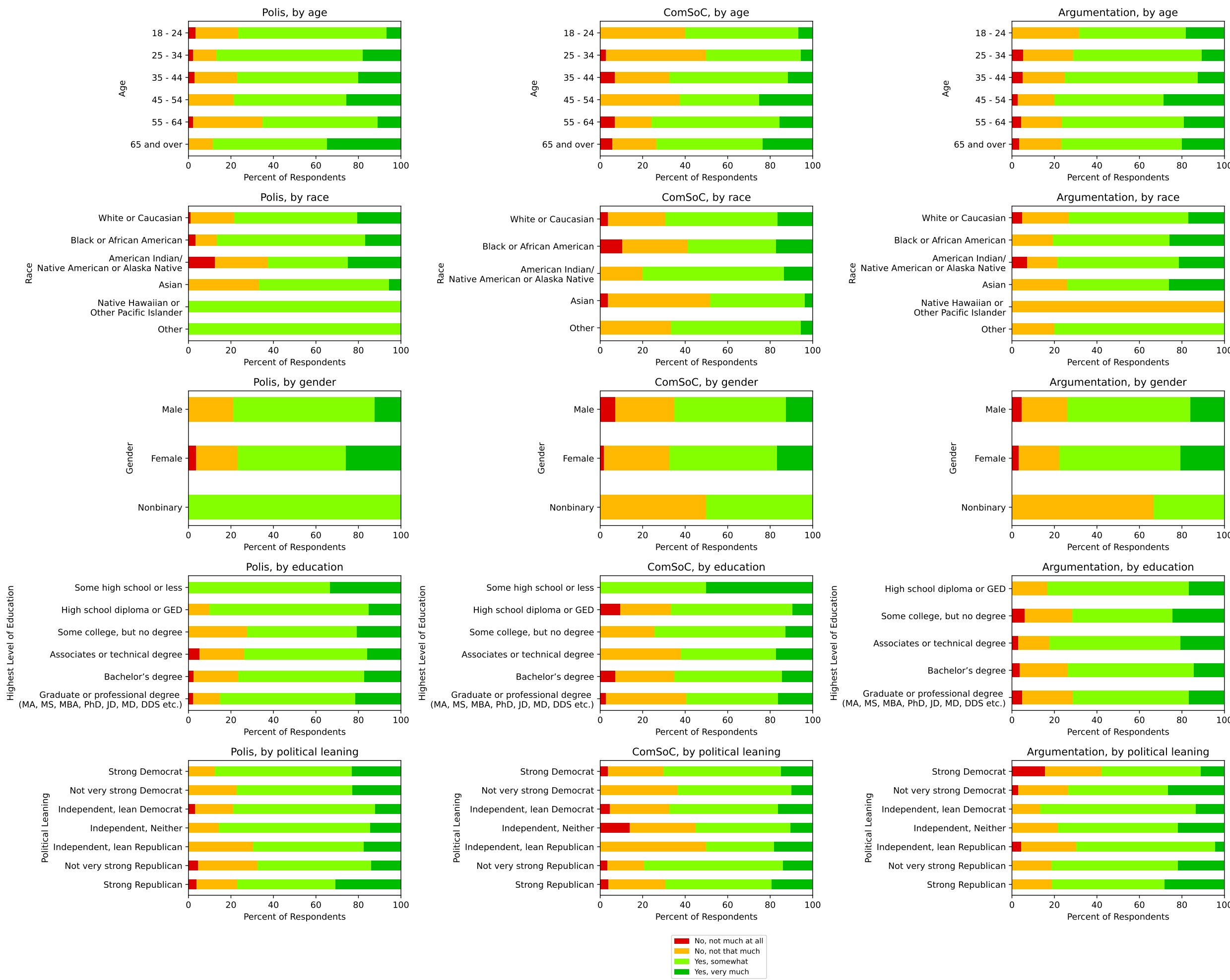


Figure B.13: If participants felt that the summarization was representative of different viewpoints, by demographic category.

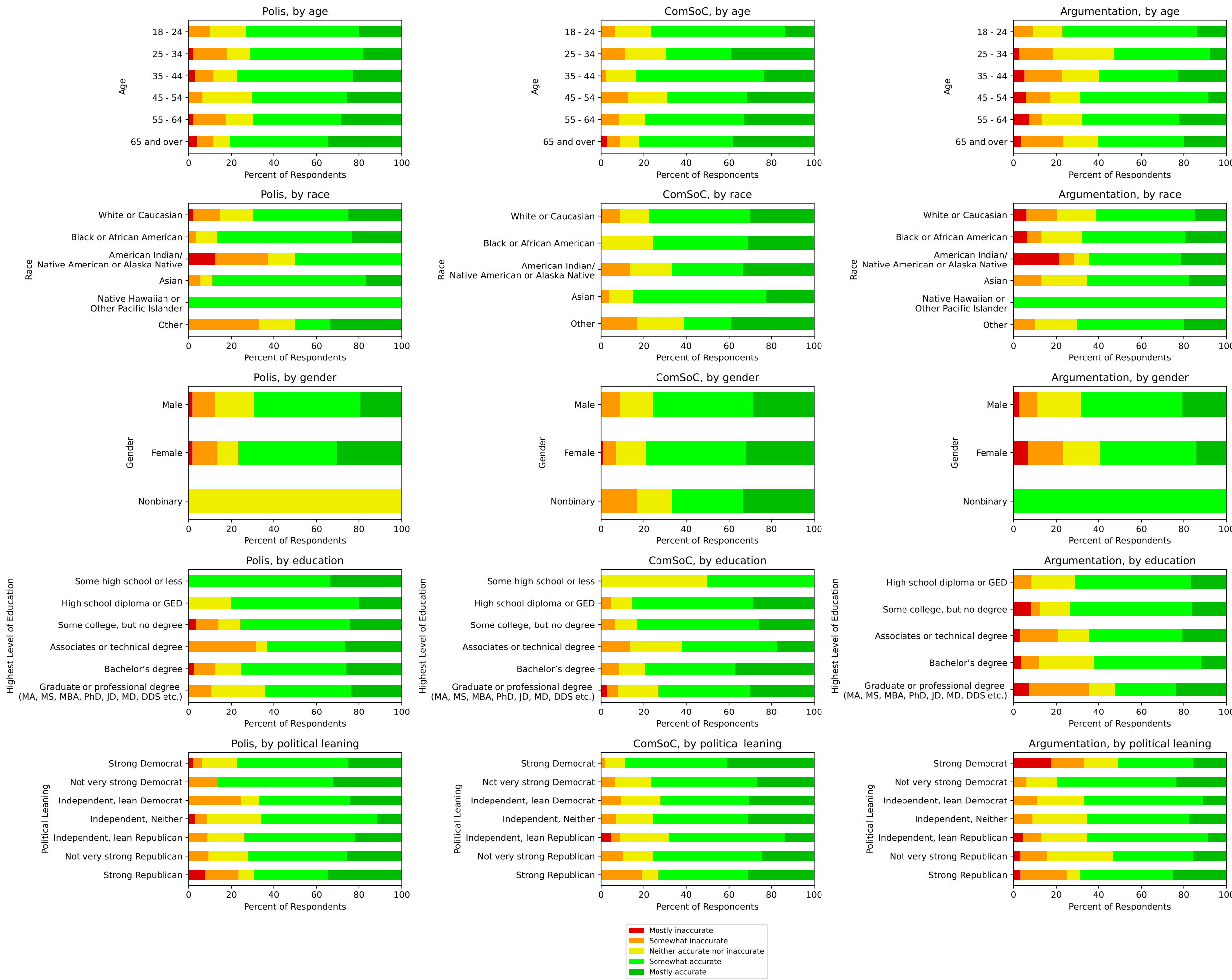


Figure B.14: The extent to which participants felt that the summarization accurately represents the general public, by demographic category.

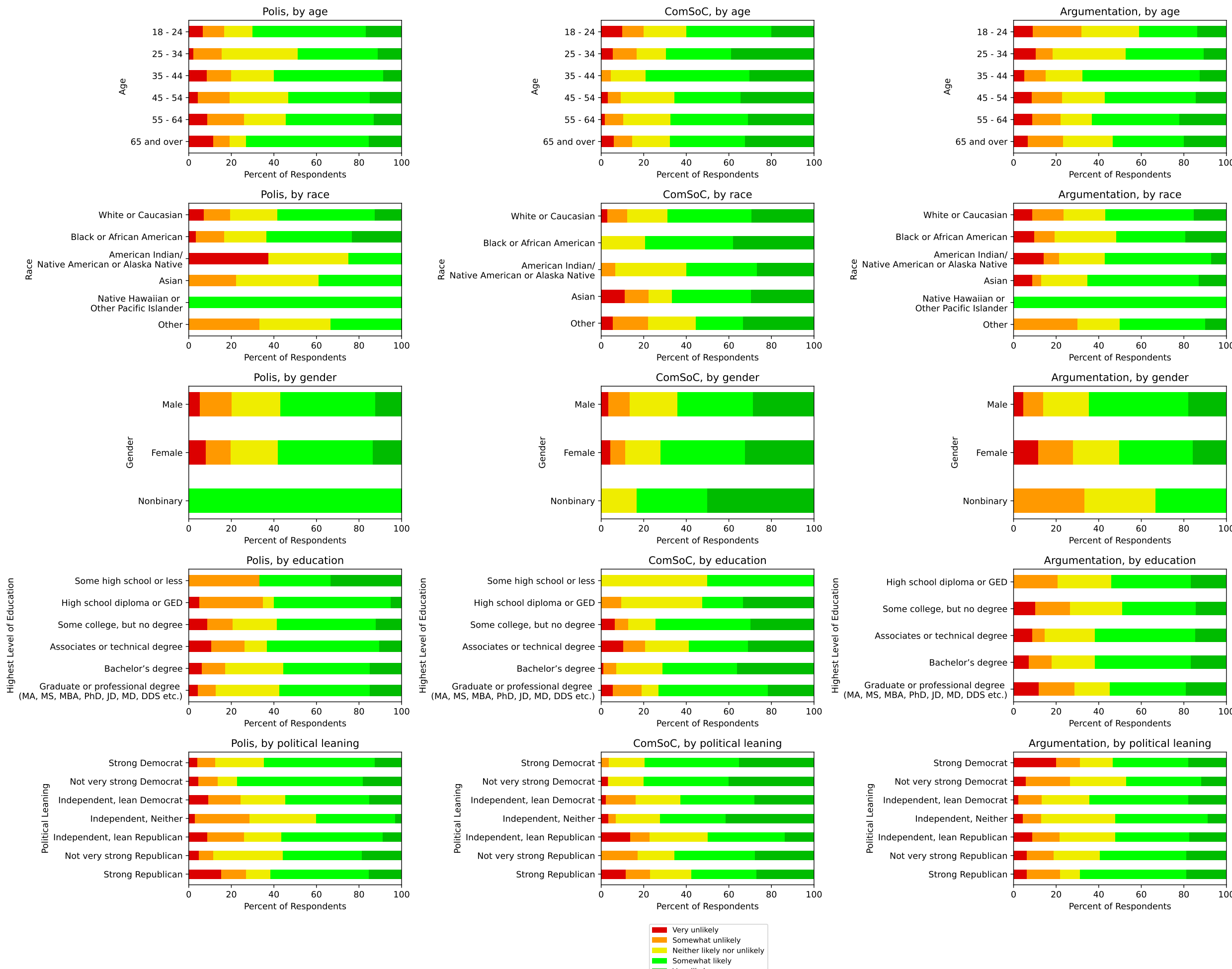


Figure B.15: How participants would have accepted that the algorithmically-generated summarization was the main feedback taken from this discussion as input by policy-makers.

References

[1] B. Barber, *Strong democracy: Participatory politics for a new age*. Univ of California Press, 2003.

[2] V. Weerakkody and C. G. Reddick, *Public sector transformation through e-government: experiences from Europe and North America*. Routledge, 2012.

[3] C. C. F. DeTar, "Intertwinkles: Online tools for non-hierarchical, consensus-oriented decision making," Ph.D. dissertation, Massachusetts Institute of Technology, 2013.

[4] J. Fishkin and P. Laslett, *Debating Deliberative Democracy* (Philosophy, Politics & Society). Wiley, 2008, ISBN: 9780470680469. URL: https://books.google.com/books?id=ffhWeWFOrIMC.

[5] G. Smith, "Democratic innovations: Designing institutions for citizen participation," *Cambridge University*, 2009.

[6] S. Coleman and J. G. Blumler, *The Internet and democratic citizenship: Theory, practice and policy*. Cambridge University Press, 2009.

[7] P. Gerbaudo, "Are digital parties more democratic than traditional parties? evaluating podemos and movimento 5 stelle's online decision-making platforms," *Party Politics*, vol. 27, no. 4, pp. 730–742, 2021.

[8] R. Fuller, *Principles of Digital Democracy: Theory and Case Studies*. Walter de Gruyter GmbH & Co KG, 2023, vol. 8.

[9] A. Macintosh, "Characterizing e-participation in policy-making," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, 2004, 10 pp.-. DOI: 10.1109/HICSS.2004.1265300.

[10] R. Medaglia, "Eparticipation research: Moving characterization forward (2006–2011)," *Government Information Quarterly*, vol. 29, no. 3, pp. 346–360, 2012.

[11] Ø. Sæbø, J. Rose, and L. S. Flak, "The shape of eparticipation: Characterizing an emerging research area," *Government information quarterly*, vol. 25, no. 3, pp. 400–428, 2008.

[12] I. Susha and Å. Grönlund, "Eparticipation research: Systematizing the field," *Government Information Quarterly*, vol. 29, no. 3, pp. 373–382, 2012.

[13] B. W. Wirtz, P. Daiser, and B. Binkowska, "E-participation: A strategic framework," *International Journal of Public Administration*, vol. 41, no. 1, pp. 1–12, 2018.

[14] B. N. Hague and B. D. Loader, *Digital democracy: Discourse and decision making in the information age*. Routledge, 2005.

[15] J. Åström, M. Karlsson, J. Linde, and A. Pirannejad, "Understanding the rise of e-participation in non-democracies: Domestic and international factors," *Government Information Quarterly*, vol. 29, no. 2, pp. 142–150, 2012.

[16] M. Steinbach, J. Sieweke, and S. Süß, "The diffusion of e-participation in public administrations: A systematic literature review," *Journal of organizational computing and electronic commerce*, vol. 29, no. 2, pp. 61–95, 2019.

[17] M. Toots, "Why e-participation systems fail: The case of estonia's osale. ee," *Government Information Quarterly*, vol. 36, no. 3, pp. 546–559, 2019.

[18] M. R. Vicente and A. Novo, "An empirical analysis of e-participation. the role of social networks and e-government over citizens' online engagement," *Government Information Quarterly*, vol. 31, no. 3, pp. 379–387, 2014.

[19] N. Bharosa, "The rise of govtech: Trojan horse or blessing in disguise? a research agenda," *Government Information Quarterly*, vol. 39, no. 3, p. 101 692, 2022.

[20] J. Saldivar, C. Parra, M. Alcaraz, R. Arteta, and L. Cernuzzi, "Civic technology for social innovation: A systematic literature review," *Computer Supported Cooperative Work (CSCW)*, vol. 28, pp. 169–207, 2019.

[21] L. Hennen, I. Van Keulen, I. Korthagen, G. Aichholzer, R. Lindner, and R. Ø. Nielsen, *European e-democracy in practice*. Springer Nature, 2020.

[22] M. Janssen and N. Helbig, "Innovating and changing the policy-cycle: Policy-makers be prepared!" *Government Information Quarterly*, vol. 35, no. 4, S99–S105, 2018.

[23] T. R. Coelho, M. A. Cunha, and M. Pozzebon, "Eparticipation practices and mechanisms of influence: An investigation of public policymaking," *Government Information Quarterly*, vol. 39, no. 2, p. 101 667, 2022.

[24] A. Chadwick, "Web 2.0: New challenges for the study of e-democracy in an era of informational exuberance," *Isjlp*, vol. 5, p. 9, 2008.

[25] T. Vedel, "The idea of electronic democracy: Origins, visions and questions," *Parliamentary Affairs*, vol. 59, no. 2, pp. 226–235, 2006.

[26] G. Aichholzer and G. Rose, "Experience with digital tools in different types of e-participation," *European E-democracy in practice*, pp. 93–140, 2020.

[27] B. Shin, M. Rask, and P. Tuominen, "Learning through online participation: A longitudinal analysis of participatory budgeting using big data indicators," *Information Polity*, vol. 27, no. 4, pp. 517–538, 2022.

[28] V. Palacin, M. Nelimarkka, P. Reynolds-Cuéllar, and C. Becker, "The design of pseudo-participation," in *Proceedings of the 16th Participatory Design Conference 2020-Participation (s) Otherwise-Volume 2*, 2020, pp. 40–44.

[29] B. Shin, J. Floch, M. Rask, P. Bæck, C. Edgar, A. Berditchevskaia, P. Mesure, and M. Branlat, "A systematic analysis of digital tools for citizen participation," *Government Information Quarterly*, vol. 41, no. 3, p. 101 954, 2024.

[30] R. Borge, J. Balcells, and A. Padró-Solanet, "Democratic disruption or continuity? analysis of the decidim platform in catalan municipalities," *American Behavioral Scientist*, vol. 67, no. 7, pp. 926–939, 2023.

[31] P. Aragón, A. Kaltenbrunner, A. Calleja-López, A. Pereira, A. Monterde, X. E. Barandiaran, and V. Gómez, "Deliberative platform design: The case study of the online discussions in decidim barcelona," in *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II 9*, Springer, 2017, pp. 277–287.

[32] S. Royo, V. Pina, and J. Garcia-Rayado, "Decide madrid: A critical analysis of an award-winning e-participation initiative," *Sustainability*, vol. 12, no. 4, p. 1674, 2020.

[33] M. Deseriis, D. Vittori, *et al.*, "The impact of online participation platforms on the internal democracy of two southern european parties: Podemos and the five star movement," *International Journal of Communication*, vol. 13, pp. 5696–5714, 2019.

[34] P. Gerbaudo, *The digital party: Political organisation and online democracy*. Pluto Press, 2018.

[35] R. Borge, J. Balcells, and A. Padró-Solanet, "A model for the analysis of online citizen deliberation: Barcelona case study," *International Journal of Communication, 2019, 13*, 2019.

[36] M. Deseriis, "Reducing the burden of decision in digital democracy applications: A comparative analysis of six decision-making software," *Science, Technology, & Human Values*, vol. 48, no. 2, pp. 401–427, 2023.

[37] N. Carpentier, *Media and participation: A site of ideological-democratic struggle*. Intellect, 2011.

[38] L. L. Tsai, A. Pentland, A. Braley, N. Chen, J. R. Enríquez, and A. Reuel, "Generative AI for Pro-Democracy Platforms," *An MIT Exploration of Generative AI*, Mar. 2024, https://mit-genai.pubpub.org/pub/mn45hexw.

[39] D. Halpern, G. Kehne, A. D. Procaccia, J. Tucker-Foltz, and M. Wüthrich, "Representation with incomplete votes," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 5657–5664.

[40] M. Bernreiter, J. Maly, O. Nardi, and S. Woltran, "Combining voting and abstract argumentation to understand online discussions," *arXiv preprint arXiv:2402.05895*, 2024.

[41] M. Lopez-Sanchez, M. Serramia, and J. A. Rodríguez-Aguilar, "Improving on-line debates by aggregating citizen support 1," in *Artificial Intelligence Research and Development*, IOS Press, 2021, pp. 425–434.

[42] M. Serramia, J. Ganzer, M. López-Sánchez, J. A. Rodríguez-Aguilar, N. Criado, S. Parsons, P. Escobar, and M. Fernández, "Citizen support aggregation methods for participatory platforms 1," in *Artificial Intelligence Research and Development*, IOS Press, 2019, pp. 9–18.

[43] J. A. Rodriguez-Aguilar, M. Serramia, and M. Lopez-Sanchez, "Aggregation operators to support collective reasoning," in *International Conference on Modeling Decisions for Artificial Intelligence*, Springer, 2016, pp. 3–14.

[44] S. Fish, P. Gölz, D. C. Parkes, A. D. Procaccia, G. Rusak, I. Shapira, and M. Wüthrich, "Generative social choice," *arXiv preprint arXiv:2309.01291*, 2023.

[45] A. Konya, L. Schirch, C. Irwin, and A. Ovadya, "Democratic policy development using collective dialogues and ai," *arXiv preprint arXiv:2311.02242*, 2023.

[46] C. T. Small, I. Vendrov, E. Durmus, H. Homaei, E. Barry, J. Cornebise, T. Suzman, D. Ganguli, and C. Megill, "Opportunities and risks of llms for scalable deliberation with polis," *arXiv preprint arXiv:2306.11932*, 2023.

[47] C. Small, M. Bjorkegren, T. Erkkilä, L. Shaw, and C. Megill, "Polis: Scaling deliberation by mapping high dimensional opinion spaces," *Recerca: revista de pensament i anàlisi*, vol. 26, no. 2, 2021.

[48] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559–572, 1901.

[49] S. Roweis, "Em algorithms for pca and spca," *Advances in neural information processing systems*, vol. 10, 1997.

[50] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA, vol. 1, 1967, pp. 281–297.

[51] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[52] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[53] R. A. Fisher, "On the interpretation of χ 2 from contingency tables, and the calculation of p," *Journal of the royal statistical society*, vol. 85, no. 1, pp. 87–94, 1922.

[54] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh, "Justified representation in approval-based committee voting," *Social Choice and Welfare*, vol. 48, no. 2, pp. 461–485, 2017.

[55] J. L. Fernández-Martínez, M. Lopez-Sanchez, J. A. R. Aguilar, D. S. Rubio, and B. Z. Nemegyei, "Co-designing participatory tools for a new age: A proposal for combining collective and artificial intelligences," *International Journal of Public Administration in the Digital Age (IJPADA)*, vol. 5, no. 4, pp. 1–17, 2018.

[56] H. Aziz, E. Elkind, S. Huang, M. Lackner, L. Sánchez-Fernández, and P. Skowron, "On the complexity of extended and proportional justified representation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[57] P. Skowron, "Proportionality degree of multiwinner rules," in *Proceedings of the 22nd ACM Conference on Economics and Computation*, 2021, pp. 820–840.

[58] H. Aziz, S. Gaspers, J. Gudmundsson, S. Mackenzie, N. Mattei, and T. Walsh, "Computational aspects of multi-winner approval voting," in *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[59] P. M. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games," *Artificial intelligence*, vol. 77, no. 2, pp. 321–357, 1995.

[60] J. Cohen, "Deliberative democracy," in *Deliberation, participation and democracy: Can the people govern?* Springer, 2007, pp. 219–236.

[61] J. C. Harsanyi, "Interpersonal utility comparisons," in *Utility and Probability*, Springer, 1990, pp. 128–133.

[62] A. Sen, *Collective choice and social welfare: Expanded edition*. Penguin UK, 2017.

[63] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[64] U. Egly, S. A. Gaggl, and S. Woltran, "Aspartix: Implementing argumentation frameworks using answer-set programming," in *International Conference on Logic Programming*, Springer, 2008, pp. 734–738.

[65] A. Niskanen and M. Järvisalo, "μ-toksia: An efficient abstract argumentation reasoner," in *International Conference on Principles of Knowledge Representation and Reasoning*, International Joint Conference on Artificial Intelligence, Inc, 2020, pp. 800–804.

[66] R. Baumann, G. Wiedemann, M. Heinrich, A. D. Hakimi, and G. Heyer, "The road map to fame: A framework for mining and formal evaluation of arguments," *Datenbank-Spektrum*, vol. 20, pp. 107–113, 2020.