Preferential voting prevails as most legitimate in polarized decision-making contexts

Carina I. Hausladen[1,*], Regula Hänggli Fricker[2], Dirk Helbing[1,3], Renato Kunz[1], Junling Wang[1], and Evangelos Pournanas[4]

[1]Computational Social Science, ETH Zurich
[2]Department of Communication and Media Research, University of Fribourg
[3]Complexity Science Hub Vienna
[4]School of Computing, University of Leeds
[*]Corresponding Author, now at Caltech. Email: carinah@caltech.edu

April 6, 2023

Abstract

The intricate and multifarious challenges of our era demand a willingness to make compromises and modify behaviors, which is unfeasible without the perception of collective decisions as legitimate. This study provides novel insights into how voters perceive the legitimacy of a range of voting methods. Using human subject experiment, 120 participants cast their votes using four voting methods: majority voting, combined approval voting, score voting, and modified Borda count. They represent a design space of preference elicitation: from low to high complexity and flexibility. The experiment was conducted in two contexts: a non-consequential setting (voting on color preferences) and a consequential context (voting on *COVID-19*-related questions). The results indicate that the perceived legitimacy of a voting method is context-dependent. Specifically, preferential voting methods are perceived as more legitimate than majority voting in complex decision-making situations, but only for individuals with well-defined preferences. Our study highlights that in highly polarized situations, preferential voting methods are more effective than majority voting. However, voter education and information campaigns are crucial to promote legitimacy and consensus-building in times of crisis.

Keywords: Social choice theory—Democracy—Legitimacy—Voting rules—Human subject experiment—COVID-19.

JEL Codes: D70, D71

1 Introduction

The challenges faced by our society are increasingly complex and multifaceted. With the ongoing process of globalization and urbanization, infectious diseases can spread more rapidly and widely than ever before. Combating these diseases requires not only scientific and technological advances, but also a willingness to compromise and change behavior, even if it may be inconvenient or go against personal preferences. But how can we promote such behavioral change? We argue that one important factor is the perception of collective decisions as legitimate. Therefore, in this paper, we investigate how legitimacy can be enhanced through the democratic mechanisms itself, particularly voting.

In fact, seeking ways to make decision-making methods *legitimate* to citizens is essential [1]: Legitimacy can be understood as the cornerstone of both social choice theory and democratic initiatives. Voting enables people to express their interests and treat everyone as equal [1]. A detailed elicitation of citizen preferences offers room for participation and therefore builds a solid basis for the legitimacy of the resulting laws [2]. Voting mechanisms differ in their potential to elicit detailed preferences and their incentives to state the latter truthfully [see, e.g., 3]. In many situations, both political and non-political, the majority vote is often seen as decisive, reflecting the importance placed on the majority of people getting their preferred outcome [4]. In contrast, other voting methods, such as multi-option preferential voting, emphasize consensus — with the intention to avoid a "tyranny of the majority" [5].

Although the relationship between legitimacy and voting methods is crucial for the functioning of democratic societies, there is a significant gap in our practical knowledge due to the lack of experimental studies in this area. As a result, there is a pressing need for more research to explore how different voting methods affect the perceived legitimacy of voting outcomes, and how such knowledge can inform policy-making and campaign design in real-world contexts.

One experimental study found that in comparison to representative decision making (the status quo), participatory processes were associated with higher perceptions of fairness, even when the outcomes were unfavorable [6]. Rather than solely focusing on the relationship between outcome and process effects in evaluating the potential of participatory processes to increase perceptions of legitimacy, we test whether different procedural settings can facilitate better outcomes for those who do not win, while also ensuring that those who do win are satisfied.

Hence, the present paper aims to close this gap. To that end, we have developed an open-source smartphone application called "VoteApp". This application provides a comparison of four different voting methods: Majority vote, combined approval voting, score voting and the modified Borda count. Note that the method to *aggregate* preferences, that is the rule for calculating the elected winners after voting, may also be relevant for the dimension of legitimacy, which, however, is not investigated in our experiment. Furthermore, the app is designed to be user-friendly while maintaining high standards of privacy protection. In a controlled laboratory experiment, participants vote with four different voting methods and provide legitimacy ratings of the respective method.

To elicit legitimacy ratings, we build on the psychological perspective in line with Tyler [7]. He defines legitimacy as "the belief that authorities, institutions, and social arrangements are appropriate, proper, and just" [7, p.376]. This belief can be multi-dimensional:

- *Input legitimacy* refers to the extent to which citizens feel represented in the process, their opportunities to participate, or the procedures introducing their preferences into the political decision-making process [8]. In our study, the voting method influences this dimension of legitimacy through different opportunities to express opinions.

- *Output legitimacy* is contingent on the substantive outputs of governing authorities (or other socially or individually desirable goals) [8]. In our study, the outcome is the result of the vote, for instance, chosen *COVID-19*-related measures. Thus, outcome legitimacy reflects the extent to which one is expected to comply with the result.

- *Throughput legitimacy* refers to the quality of the voting mechanism, it is a performance criterion [9]. Within our research, the fairness of the voting method corresponds to throughput legitimacy.

Questions on decision-related acceptance, fairness, trust, and representation can load on one factor of legitimacy [10]. That is, these dimensions are not independent, belong together (correlate highly) and characterize the more abstract concept of legitimacy in an ordinary political-decision making process. This allows us to address them with a single question.

Legitimacy is also theorized to be context-dependent: Accordingly, legitimacy may depend on (1) the criticality (i.e., recognized, imminent, serious circumstances), (2) the point in time, and (3) the motivational landscape (i.e., the level of interest) [11]. It is further found that legitimacy judgments are influenced by factors such as the stability of the institutional framework, the stage of the legitimacy judgment process, and the methods used to form such judgments [12, 13]. In our study, we introduce polarized voting questions related to *COVID-19* to better distinguish the effect of input voting methods on legitimacy. We compare this critical, highly relevant context with a more neutral context, in which participants are asked to choose their favorite color.

2 Results

Our study finds that commonly used voting methods, such as the majority vote, are perceived as less legitimate than less commonly used methods, such as score voting. Especially in highly polarized contexts, voters value the ability to express their preferences in detail.

Table 1: Winning options of COVID-19-related questions using different voting methods. This table presents the results of four COVID-related questions (columns) that were voted upon using four different voting methods (rows): majority voting ($mv = \{0, 1\}$), combined approval voting ($cav = \{-1, 0, 1\}$), score voting ($sv = \{0, 1, 2, 3, 4\}$), and modified Borda count ($mbc = \{0, 1, 2, 3, 4\}$). The winning option out of $\{o_1, ..., o_5\}$ was determined by the highest sum of scores.

Voting Method	Question *vaccine*	*icu*	*protection*	*lockdown*
mv	o_2	o_5	o_4	o_4
cav	o_4	o_2	o_4	o_4
sv	o_4	o_1	o_4	o_4
mbc	o_4	o_1	o_4	o_1

Winning Options Vary by Input Method. First, we find that voting on different input formats often leads to different outcomes: Depending on the input method, the winning option for some questions changes (*vaccine*, *icu*, and *lockdown*). Only *protection* did not show a change in the winning options (Table 1 Winners for the *color* context can be found in Table A3). Applying another aggregation rule—the Condorcet method—gives similar results (Table A2).

Why does the input method induce outcome variation for some questions, but not for others?

To this end, we investigate the sequence of scores assigned to each option. We refer to this sequence as preference profiles for each individual i, question q, and input voting method vm: $pp_i^{q,vm}$ (see details in Section 4.4). We compare variations in $pp_i^{q,vm}$ by calculating the standard deviation $\sigma(pp_i^{q,vm})$ and divisiveness D_{im}^q.

As is evident in Figure 1, *protection* shows the lowest median $\eta_\sigma^{protection} = .395$ and the second lowest divisiveness $\eta_D^{protection} = .45$. A Wilxocon ranksum test Both, $\sigma(pp_i^{q,\cdot})$ and D_{im}^q are not normally distributed

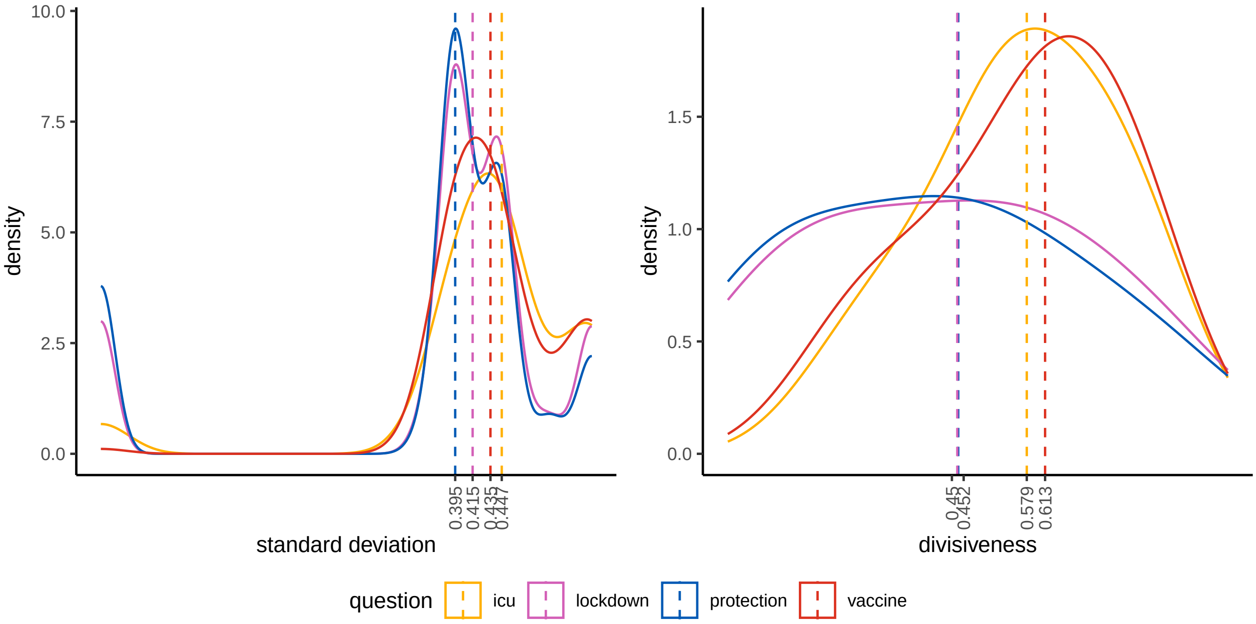


Figure 1: The figure displays two measures of variation in preferences. The left panel displays probability densities and standard deviations of preference profiles, $\sigma(pp_i^q)$. The right panel shows densities of divisiveness D^q. We compare the medians η by questions q (dashed lines), using a non-parametric test.

(details in Table A13 and Table A14). Therefore, we apply a non-parametric test, the Wilcoxon signed rank test for paired samples [14], to compare the median $\eta(\sigma(pp_i^{q,}))$ and $\eta(D_{im}^q)$ by question. confirms significance for pairwise comparisons regarding σ (Table A15) and D (Table A16). This test result describes why question *protection* shows the lowest outcome dependency on the input method. We conclude that, the higher the standard deviation and divisiveness, the more dependent is the aggregate result on the input method. In other words, the choice of the voting method is especially crucial when preference profiles are highly polarized.

Another way to characterize the questions is to investigate how voters rate the same choice option over the four input methods. In other words, do voters' preference profiles differ by the question? To this end, we introduce the notation of max-choice profiles. Max-choice profiles, broadly speaking, follow the rationale that the highest rated option is of special importance. Its score is remapped to 1, while the remaining scores are mapped to 0. An example of a max-choice profile for one participant, one question, and one option, is 1111. It means that this particular option was rated highest across all four input methods. More details can be found in Section 4.6.

Figure 6 shows the five most common max choice profiles: 0110, 1111, 0100, and 0000. The grey vertical lines represent the actual count of each choice profile. The dots on each grey bar show the count per question. The orange dashed lines above two of the max choice profiles indicate the theoretical maximum of that profile. The choice profile "1111," which signifies a fully consistent voter, is of particular interest. If all votes had been cast consistently, in Figure 6 the grey vertical line would reach the orange one. However, this is not the case, indicating that a significant portion of voters, namely 32.8%, did not vote consistently. Furthermore, we test for significant differences in counts across questions for those four choice profiles (Poisson regression models as summarized by Table A18).

The number of counts per question exhibit a significant disparity between the max-choice profiles 0110, 0100, and 0000. However, this difference is not observed in the entirely consistent profile 1111. This suggests that consistency is not influenced by the specific question asked, but rather reflects an individual's unique trait. As we will describe later, this personal attribute has a correlation with the individual's perception of legitimacy.

Flexibility in Voting Methods is Perceived as More Legitimate in Critical Contexts. In this paragraph, the investigation focuses on how voting methods and context impact perceived legitimacy. Participants in our experiment provided legitimacy ratings for every input method across two contexts on a Likert scale from zero to four.

Figure 3 displays a comparison of legitimacy ratings. The left panel is analyzed first. The dots within each box indicate the mean μ, while the horizontal lines represent the median η of legitimacy ratings. We focus on interpreting the median since the data is not normally distributed (as shown in Table A4). In the *color* context, we found that the medians ($\eta = 3$) for all voting methods were identical. However, in the *COVID-19* context, the median perception of legitimacy for majority voting ($\eta = 1$) was lower than that of the other three voting methods, which had equal median values ($\eta = 3$).

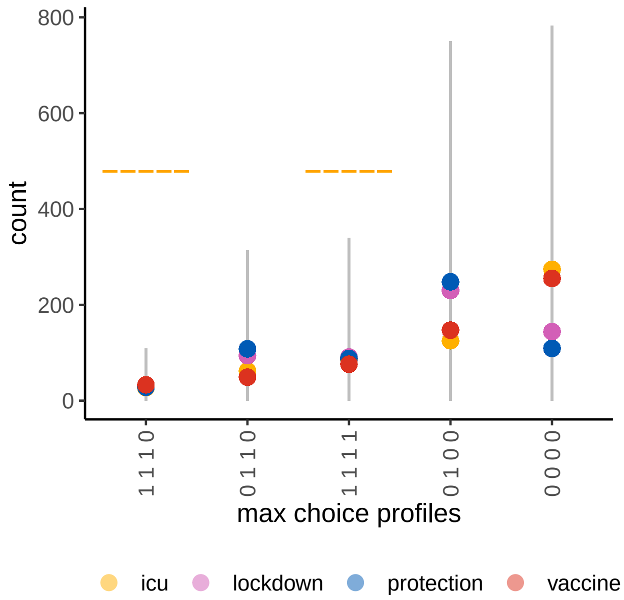


Figure 2: Absolute frequencies (grey vertical lines) per max-choice profile for all four questions. Profiles (x-axis ticks) with $n > 25$ are displayed. The colored dots show counts by question. The orange horizontal line depicts the maximum frequency for that specific choice profile. Theoretical maxima differ by profile; see Section 4.6.

Pairwise comparisons of legitimacy ratings between input methods within each context are indicated by orange and pink p-values and brackets. These comparisons are based to the full array—neither μ nor η—of legitimacy ratings, using the paired Wilcoxon signed rank test with Holms adjustment (see Appendix A.6 for more details). In the *color* context (in orange), all but one comparison (ca-mbc) are significantly different from each other. Consequently, majority voting was considered the least legitimate, followed by combined approval and modified Borda count, both of which were considered less legitimate than score voting In the *COVID-19* context (in pink), all but two comparisons (mv-ca and sv-mbc) are significantly different from each other. Consequently, majority and combined approval voting were perceived as the least legitimate voting methods, followed by the modified Borda count, being considered less legitimate than score voting.

In summary, the results show that score voting is perceived as more legitimate than majority voting in both the *color* and *COVID-19* contexts. However, in the *COVID-19* context, score voting and the modified Borda count were rated as equally legitimate. This is likely because the modified Borda count requires a voter to exclude a choice, which may not make sense when considering colors, as disliking a color to the point of not wanting to vote on it is quite uncommon. On the other hand, in the *COVID-19* context, issues such as strict lockdown measures can be strongly disliked, leading to a similar level of legitimacy for both voting methods.

Furthermore, we investigate whether the perceived legitimacy of a specific voting method varies by context. Two significant results emerge (black p-values in Figure 3): The majority vote is rated as more legitimate when voting on colors compared to *COVID-19*-related issues (paired Wilcoxon signed-rank test, Table A9, p-value=7.02×10^{-6}). The opposite is true for score voting (p-value= .043). These results suggest that the legitimacy of a voting method depends on the context. In other words, our results support the notion that legitimacy is context-dependent and not a universal property.

Less Complex Voting Methods are Perceived as More Legitimate by Inconsistent Voters. Another crucial question is whether an individual's personal traits affect their legitimacy ratings. To examine this, we focus on the right panel of Figure 3. Here, we compare those who consistently vote to those who do not vote consistently. Please recall that the grey vertical line in the choice profile "1111" of Figure 6 represents the fully consistent voters. These voters are compared to the inconsistent voters, who are responsible for the discrepancy between the grey and orange lines. Now, let's return our focus to Figure 3.

Our analysis of consistent voters (shown in blue) reveals that the median legitimacy rating for majority voting ($\eta = 1$) was lower than that of combined approval and the modified Borda count ($\eta = 3$), which in turn was lower than the median legitimacy rating for score voting ($\eta = 4$). In contrast, for voters with inconsistent preferences, we found that the legitimacy rating for majority voting ($\eta = 2$) was significantly lower than for the other three voting methods, which had the same median legitimacy rating ($\eta = 3$). The exact values for μ and η can be found in Table A6.

Pairwise comparisons of legitimacy ratings between input methods across consistent and inconsistent voters are indicated by blue and red p-values and brackets (calculated using the paired Wilcoxon signed rank test with Holms adjustment, see Appendix A.6 for more details). For consistent voters (in blue), all but one comparison (ca-mbc) are significantly different from each other and thereby mirror the results of the whole population. For inconsistent voters (in red), two additional comparisons turn out to be non-significant (mv-mbc, cav-sv,cav-mbc).

The results of the Wilcoxon signed rank test reveal an interesting phenomenon. Despite the difference in the number of scores allowed, inconsistent voters regard the exclusive three-score scale (cav) as equally legitimate as the non-exclusive five-score scale (sv). Furthermore, it is noteworthy that, on a lower level, the restrictive majority voting is perceived as equally legitimate as the more flexible modified Borda count. This could suggest that the complexity of the modified Borda count may have been challenging for individuals with uncertain preferences.

Having analyzed the comparisons within the two groups of consistent and inconsistent voters, we will now focus on comparisons between these two groups.

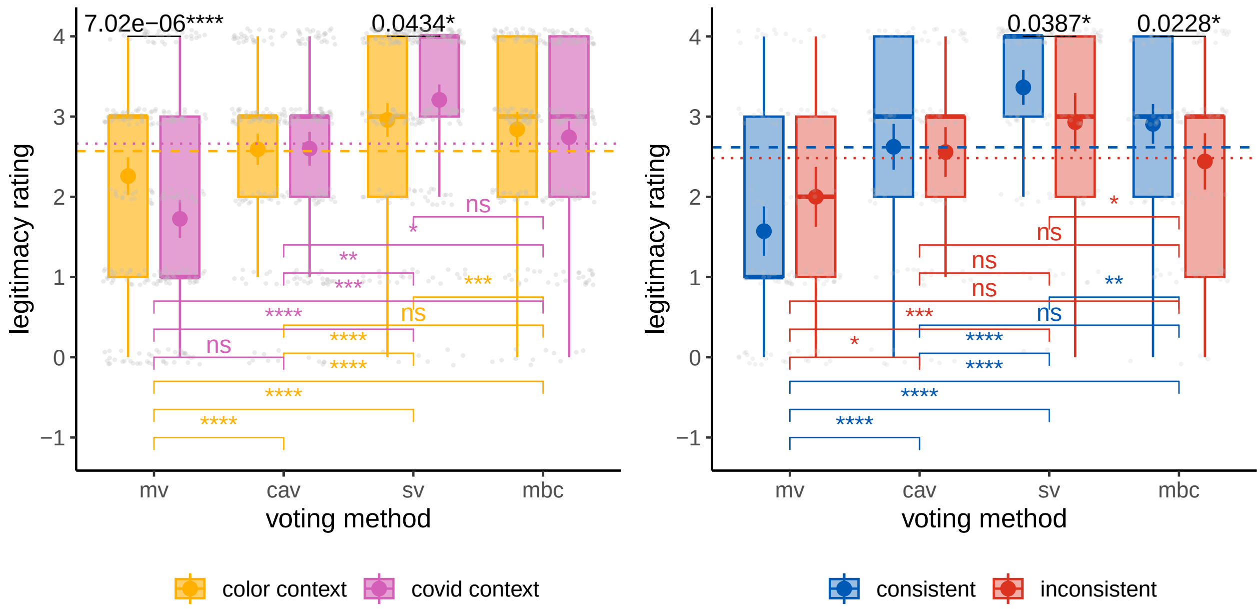


Figure 3: Legitimacy Ratings by Context (Left) and Inconsistency (Right). P-values are calculated using the Wilcoxon signed rank test with Holms adjustment. Black p-values and brackets denote comparisons between dimensions (*color* vs. *COVID-19*, consistent vs. inconsistent). P-values in *color* indicate comparisons within a dimension (*color*, *COVID-19*, consistent, inconsistent) across voting methods. The scatter plot shows individual ratings, with mean values represented by dots inside the box, median values indicated by horizontal lines, standard deviation by vertical lines, and outliers indicated by dots outside the box. The dashed horizontal lines represent the mean of all voting methods. *ns* indicates no significance.

The results of the Wilcoxon rank sum test (as seen in Table A10) reveal a statistically significant difference in the legitimacy ratings assigned by fully consistent and partially consistent voters. The results indicate that consistent voters tend to assign higher ratings to score voting and the modified Borda count, with p-values of .0387 and .0228 respectively. These findings are noteworthy. Consistent voting behavior across four voting methods indicates that preferences are well-defined and distinct from one another, making it feasible for these voters to express themselves in a detailed manner, as required by score voting and the modified Borda count. On the other hand, for voters with less stable preferences, the requirement for more detail in these voting methods may feel burdensome, leading to a lower perceived legitimacy of these methods. This highlights the importance of considering the varying levels of preference stability among voters when evaluating the legitimacy of different voting methods.

Perception of Legitimacy is Linked to Appreciation of Flexibility in Score Voting. Figure 4 illustrates the relationship between options, preference profiles, and perceived legitimacy. The left panel displays how preference profiles are connected to legitimacy ratings. The orange lines represent individual preference profiles, sorted from highest to lowest ratings on the right. The rows in the figure represent the voting method, while the columns show the legitimacy rating. The area under the preference curve (AUC) is greatest when multiple options receive high ratings and smallest when only one option is rated highly. The right panel of Figure 4 presents the relationship between the AUC and the legitimacy rating for each voting method. It can be observed that the relationship between the two is clearly inverse U-shaped for score voting. As far as score voting is concerned, when few options receive high scores, resulting in a low AUC, the legitimacy ratings are also low, indicating that score voting is seen as only slightly legitimate. However, as AUC increases and more options receive high ratings, the legitimacy ratings also increase, showing that voters who appreciate the flexibility offered by score voting are more likely to rate the method as legitimate. On the other hand, when all options receive very high scores with no distinction between them, resulting in a high AUC, the legitimacy ratings decrease. This suggests that, when voters are unable to distinguish between different options, score voting is perceived as less legitimate. Hence, the perceived legitimacy of a voting method depends on the complexity of the voter's preference profile. If a voter's preferences are complex and nuanced, they require a more complex voting method, and vice versa.

The Independence of Perception of Legitimacy from COVID-19-Related Topics Confirms the Validity of Our Legitimacy Framework. Additionally, our results suggest that the legitimacy ratings of a particular voting method do not significantly vary based on a voter's preference cluster for *COVID-19*-related topics. In other words, legitimacy ratings

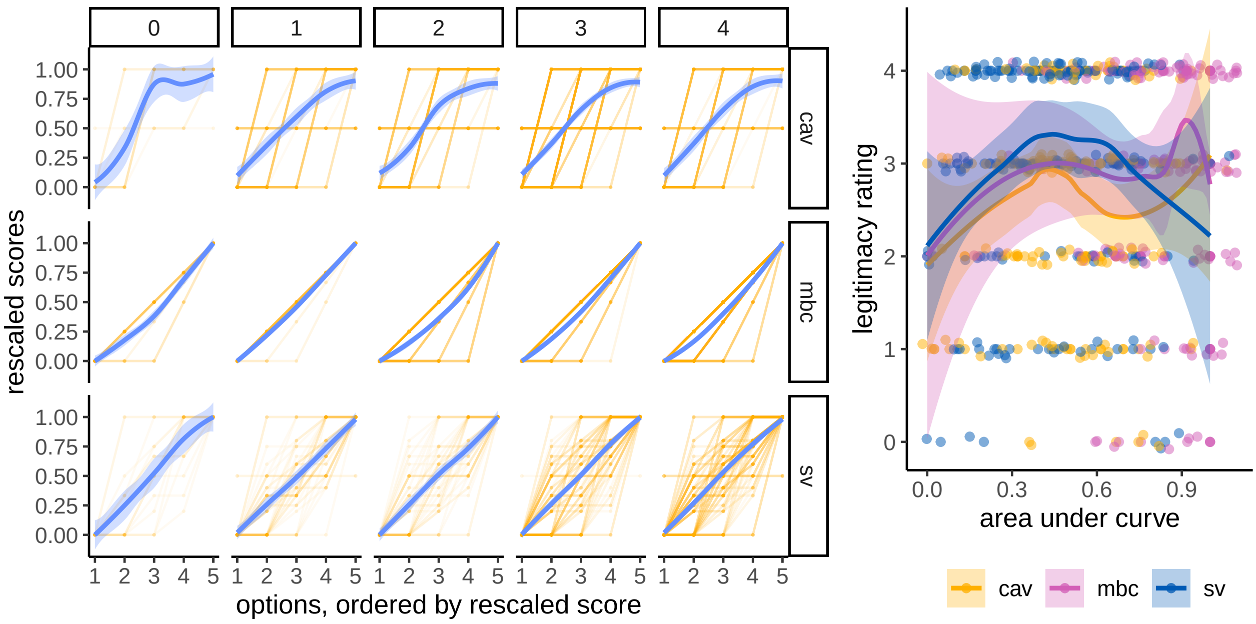


Figure 4: The left subfigure in the figure illustrates the preference profiles of all participants, with the rows representing three distinct voting methods and columns referring to the legitimacy ratings. The preference profiles are depicted within each panel, and the options are sorted in such a way that the highest-rated option is on the right and the lowest-rated option is on the left. The blue line represents the Local Polynomial Regression Fitting with a smoothing function $y \sim x$, and the grey area represents the 95 percent confidence interval. The right subfigure displays the average area under the preference profile curves (AUC) across questions for each participant, which has been rescaled within each voting method. The AUC is plotted against the legitimacy rating for each voting method, providing a visual representation of the relationship between the AUC and the perceived legitimacy of the voting methods.

are independent of *COVID-19* preferences. To that end, we cluster a voter's preference profile (more details can be found in subsection 4.7). Through the clustering process, we were able to control for any potential influences of the satisfaction with the options upon which the rating was given on the legitimacy ratings. If participants' ratings of voting method legitimacy were influenced by their preferences, we would have to contend with the challenge of controlling for the potential biasing effect of the randomization of question order, where the last question answered by a participant may have been overrepresented in their memory and may, therefore, have skewed their ratings. This analysis was crucial in determining the validity of our method for investigating legitimacy, as it aimed to verify whether participants were rating the intended concept. As our analysis showed, out of the 16 comparisons made, none had a statistically significant p-value (Kruskal-Wallis test, presented in Table 3). This indicates that the preference profiles on *COVID-19*-related topics do not structurally impact how participants rate the legitimacy of voting methods. Therefore, our questions regarding the legitimacy of voting methods were valuable and accurate indicators of the participants' views.

3 Discussion

Our key findings can be summarized as:

1. Our study provides empirical evidence that different voting methods can lead to different outcomes, even when the same group of individuals vote on the same set of questions.
2. In particular, the choice of voting method is particularly important in contexts where preferences are highly polarized.
3. Our study also found that the perceived legitimacy of a voting method is not a universal property, but rather context-dependent. Specifically, we found that the legitimacy gain for preferential voting methods over majority voting is maximized when more complex questions are being asked.
4. However, the latter statement is only true for individuals with certain preferences. Those with uncertain preferences tend to conflate their own undecidedness with the perceived legitimacy of the method.
5. It is not only uncertainty but also nuance that matters: If a voter's preferences are nuanced, they perceive a more nuanced voting method as more legitimate, and vice versa.

Our findings highlight the critical importance of carefully selecting a voting method in highly polarized contexts, such as those related to COVID-19, as the chosen outcome is likely to vary depending on the mechanism employed. While consistent voters prefer a voting method that allows for a more nuanced expression of preferences, voters with unstable or less well-defined preferences may find such requirements burdensome and perceive these methods as less legitimate. In other words, a voter who perceives the process to be illegitimate and has no clear preference on the outcome is challenging to satisfy. Conversely, a voter who perceives the process to be legitimate but is dissatisfied with the outcome may become disillusioned with the voting process if consistently suboptimal outcomes are chosen due to unstable preferences. Therefore, it is crucial to inform and educate voters to make informed decisions. Like a referendum, which is a powerful instrument that can lead to contradictory results if used in an uninformed way (as seen in the Brexit referendum), the voting process must be chosen thoughtfully and supported by information campaigns to maintain its legitimacy and effectiveness.

The study opens up several avenues for *future research*, including the relationship between outcome favorability and perceived legitimacy, which was not explicitly evaluated in this study. Future studies could investigate the extent to which the outcome favorability of a particular voting method influences the perceived legitimacy of the election results, especially in polarized or crisis situations.

In this study, we shed light on the crucial relationship between the choice of voting method and the perceived legitimacy of decisions. Our findings show that the perceived legitimacy of a voting method is influenced by a variety of factors, including the context of the decision-making process, the complexity of voter preferences, and the flexibility of the method.

Our study represents a significant advancement in the field of voting mechanism research, as it is one of the first to focus on the human perception and experience of the voting process, rather than on the theoretical properties of voting methods. In addition, our study makes a valuable contribution to the literature on legitimacy . Understanding the perception of legitimacy is crucial to both social choice theory and democratic initiatives, because it determines the willingness to support (implementation of) the voting outcome. Yet, a systematic and straightforward method for measuring it has remained elusive. We address this issue by distilling four key elements of legitimacy and developing a practical framework for assessing it, providing practitioners in the field with a valuable tool.

Based on previous research in Switzerland's direct democracy [15, 16], we know that personal preferences as well as party heuristics also influence the decision. In polarized contexts or when the issue at stake is complex and emotional, this could result in (few) voters casting their vote against their own preferences. However, it has been shown that inconsistencies in voting behavior can be reduced in real-world contexts when voters are provided with high-quality information [17] Thus, it is crucial to pay attention on what information the voters are exposed to, who is involved in preparing them, to which degree and how these inconsistencies are resolved, and how the decision-making process fits the voting method.

4 Methods

We begin this section by presenting the various voting methods evaluated by participants in a human subject experiment. The primary dependent variable, legitimacy, is introduced, and participants' ratings of different input methods on its legitimacy are discussed. We refer to the resulting data as "preference profiles," which we measure rigorously.

4.1 Input Methods

A voting method consists of (1) an input mechanism (the voting process) and (2) an aggregation rule (the evaluation process). We vary the input mechanism. In our behavioral experiment, we implement four input methods that differ in their scale s, the framing, and whether ranking is required. Table A1 provides an overview, and the following list details on the four input methods we implemented:

1. **Majority voting:** This requires the selection of one out of two or more options, i.e. $s_{mv} = \{0, 1\}$.

2. **Combined approval voting:** This requires disapproval (-1), indication of neutrality (0), or approval (+1) of the voting options, i.e. $s_{cav} = \{-1, 0, 1\}$.

3. **Score voting:** This requires assigning a numerical score to each alternative option, to reflect the degree of the preference. We assume $s_{sv} = \{0, 1, 2, 3, 4\}$.

4. **Modified Borda count:** This gives no points to unranked options, 1 point to the least preferred of the ranked options, etc. The choices are: $s_{mbc} = \{0, 1, 2, 3, 4\}$. If a voter ranks A above B and leaves other options unranked, A will receive 2 points, B will receive 1 point, and the remaining options will receive none.

In our lab experiments, the order of the input methods was the same for all participants and questions, i.e., the input method became increasingly complex (going from method 1 to method 4 above). This was done to address the additional cognitive effort required from voters to express their preferences. That is, the next method augments the previous ones.

Majority voting ("choose one option") offers exactly one approval option and no option to reject. *Combined approval voting* requires assigning one of three different scores ("approve, stay neutral, disapprove"). *Score voting* ("assign points to options") adds another two levels, i.e., it has overall five levels to choose from. It allows for assigning the same score multiple times.

Therefore, a voter can express indecisiveness. In contrast, the *modified Borda count* ("choose and rank options") does not offer this option. It is cognitively even more challenging, as explicit ranking is required over a selected number of options.

4.2 Measuring Legitimacy

To obtain a comprehensive proxy of legitimacy that covers the relevant theoretical grounding, we ask the following question:

> "You voted in four different ways. Now, please assess the following statement for each voting method applied. — I would comply with the result and accept it as fair, reflecting my and others' opinions."

For every input method $im \in \{mv, cav, sv, mbc\}$, participants in our experiment provide legitimacy ratings LR on a Likert scale $LR = \{0, ..., 4\}$ across two contexts $c \in \{color, COVID\text{-}19\}$.

4.3 Behavioral Experiment

To test our hypotheses, we conducted a preregistered human subject experiment in a controlled laboratory environment. The preregistration link is `https://doi.org/10.1257/rct.7871-1.0`. Our experiment had three stages (see Table 2 for an overview): During Stage I, participants were introduced to both the *COVID-19* questions and the input methods that would be used during the experiment. Stage II was the main focus of our study, where participants were asked to vote on a set of questions related to *COVID-19* and provide legitimacy ratings for their chosen voting methods. Specifically, each of the four *COVID-19* questions ($q = vaccine, icu, protection, lockdown$) presented participants with five options ($o = o_1, o_2, o_3, o_4, o_5$) from which they were asked to assign a score (s). Finally, Stage III was dedicated to answering a set of control questions.

Figure 5 shows screenshots of the four *COVID-19*-related questions, their options, and how the user interface for the four input methods looked like.

Overall, 121 subjects (share of females: .36) participated in the experiment. Their mean age was 25.47, coming from 22 different countries. Most participants' highest level of education was a Bachelor's degree (share: .37).

4.4 Preference Profile

We call the choice data collected via the human subject experiment (within one context) a preference profile pp for individual i, question q and voting method vm. In our experiment, preferences are expressed through four voting methods, each of which assigns a different set of scores s to a voter's choice (scores are summarized in Table A1). The preference profile is defined as $pp_i^{q,vm} = \{s_{o_1}, s_{o_2}, s_{o_3}, s_{o_4}, s_{o_5}\}$, where $s_{o_{vm}} = \{s_{mv_{o_j}}, s_{ca_{o_j}}, s_{sv_{o_j}}, s_{mbc_{o_j}}\}$

Table 2: Overview of the three stages participants encountered in our behavioral experiment.

I	**Introduction** • read *COVID-19* questions • understand voting methods
II	**Voting and legitimacy rating** • vote on *color*; provide legitimacy rating • vote on *COVID-19*; provide legitimacy rating
III	**Answer control questions** • socio-demographic • strategic voting • polarization

4.5 Variation in Preferences

Input methods vary in their ability to capture the properties of a distribution of preferences, $pp_i^{q,im}$. Our aim is to comprehend how voters use the input methods to express their preferences on the various questions they answer. We propose two ways to capture the variation in preferences: Standard deviation σ and divisiveness D [18]. A full distribution of scores is displayed by Figure A1.

In statistics and probability research, quantitative data are summarized via various *measures of spread.* Some of these measures were proposed to capture political polarization, for example, standard deviation σ and variance σ^2 [19]. Therefore, we calculate $\sigma(pp_i^{q,im})$. To compare $\sigma(pp_i^{q,im})$ across questions, we calculate the median $\eta(\sigma(pp_i^q))$. For the median, we exclude the majority vote from σ as $\sigma_{mv} = \sigma^2 = 1$ for all individuals i and questions q.

Additionally, we use a measure for polarization referred to as *divisiveness* [18]. Divisiveness D is defined for all option pairs by the mean difference in scores s of voters who prefer one option o_m over another o_n versus those who prefer o_n over o_m. Furthermore, divisiveness D_{im}^q is calculated for each input method im and for each question q. In other words, D provides an intuition on how divisive a question is expressed through a particular input method. It is defined as

$$D_{im}^q = \frac{1}{n-1} \sum_{j!=i} \|s(o_m, o_n) - s(o_n, o_m)\| \quad (1)$$

To compare D_{im}^q across questions, we calculate $\eta(D^q)$. More details on D are presented in Appendix A.8.

4.6 Max-Choice Profile

An interesting question to investigate is how a voter rates the same option o for one specific question q over the four input methods im.

To that end, Figure 2 visualizes preferences pp_i^q by a *five*-dimensional series (five options o) with *four* time steps (four input methods im) each.

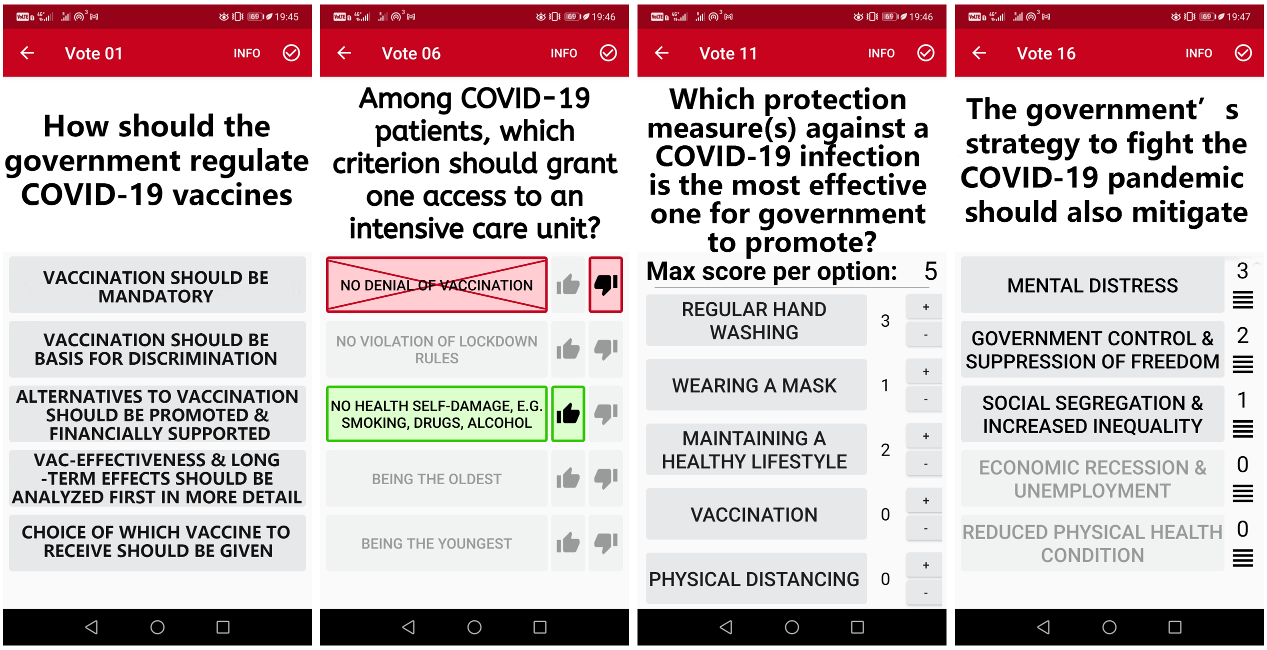


Figure 5: Screenshots of the mobile application used in the human subject experiment, displaying four different questions on which participants voted via four different input methods. In this paper, we refer to questions via the following short labels (from left to right): *vaccine*, *icu*, *protection*, and *lockdown*.

The theoretical number of possible pp_i^q is extremely large (see Appendix A.5). To compare pp_i^q across questions q, we follow the rationale that the option with the highest score is of special importance. The originally assigned score is mapped to a binary scale, where 1 represents the highest score of the options and 0 otherwise. im refers to any input method. We define $s_{im} = \{s_{im_{o_1}}, s_{im_{o_2}}, s_{im_{o_3}}, s_{im_{o_4}}, s_{im_{o_5}}\}$

$$\hat{s}_{im_{o_i}} = \begin{cases} 1, & \text{if } s_{im_{o_i}} = max(s_{im}). \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

We refer to $\hat{s}_{o_j}$ as the *max-choice-profile*: $\hat{s}_{o_j} = \{\hat{s}_{mv_{o_j}}, \hat{s}_{ca_{o_j}}, \hat{s}_{sv_{o_j}}, \hat{s}_{mbc_{o_j}}\}$

Figure 6 provides an example of how a voter's choices pp_i^q are mapped to a multi-dimensional time series.

Participants state preferences in the following order: majority vote, combined approval, score voting, and modified Borda count. The first and last input methods required ranking. The second and third methods, by contrast, allowed (winning) draws. Therefore, only those two input methods allow for expressing multi-peaked preferences. In other words, combined approval and score voting allow (winning) draws in contrast to majority voting and modified Borda count, which are based on ranking.

Whether ranking was required or not has consequences for interpreting the max-choice profile. Let's look at an example to clarify this point: A voter ranks example option A first in all four voting methods. The resulting max-choice profile is 1111. Let's further assume that the voter's preferences are multi-peaked; her max-choice profile for option B is 0110. Both profiles can be interpreted as *consistent*.

Consistency in voting choices implies that no matter the scale of the input method, the voter should rank the favorite option first in all four voting methods. Consequently, any max-choice profile with patterns $1xx0$ and $0xx1$ implies that an option was not ranked first in the two input methods requiring explicit ranking. In other words, those profiles can be interpreted as *inconsistent*.

Furthermore, to interpret the max-choice profile counts correctly, it is necessary to understand the theoretical maximum of certain profile types: Any profile with a 1 for an exclusive voting method (first and fourth), for example, 1110 and 1111, can reach a theoretical maximum count of 1 $option \times 4$ $questions \times 120$ $voters = 480$. By contrast, any profile with two 0 for the ranking-based voting methods, such as 0110, 0100, and 0000, can reach a theoretical maximum of $(5 - 1)$ $options \times 4$ $questions \times 120$ $voters = 1920$.

4.7 Clustering Expressed Preferences

The voter is asked *after* voting to rate the input method according to its legitimacy. A voter might be unsatisfied with the options voted upon. This dissatisfaction might prevail and carry over to the legitimacy rating. In this case, the voter would fail to disentangle the voting method from the subject voted upon. In other words, satisfaction with the proposed option to vote could influence legitimacy ratings. Furthermore, the question order was randomized. The last question would be overrepresented in the participant's memory; Therefore, when providing legitimacy ratings, the voter might think about the last question answered, which could carry over to legitimacy ratings. In this case, there is variation between participants we need to control for. Therefore, we cluster preferences and compare legitimacy ratings across clusters. If ratings significantly differ by cluster, legitimacy ratings are mixed

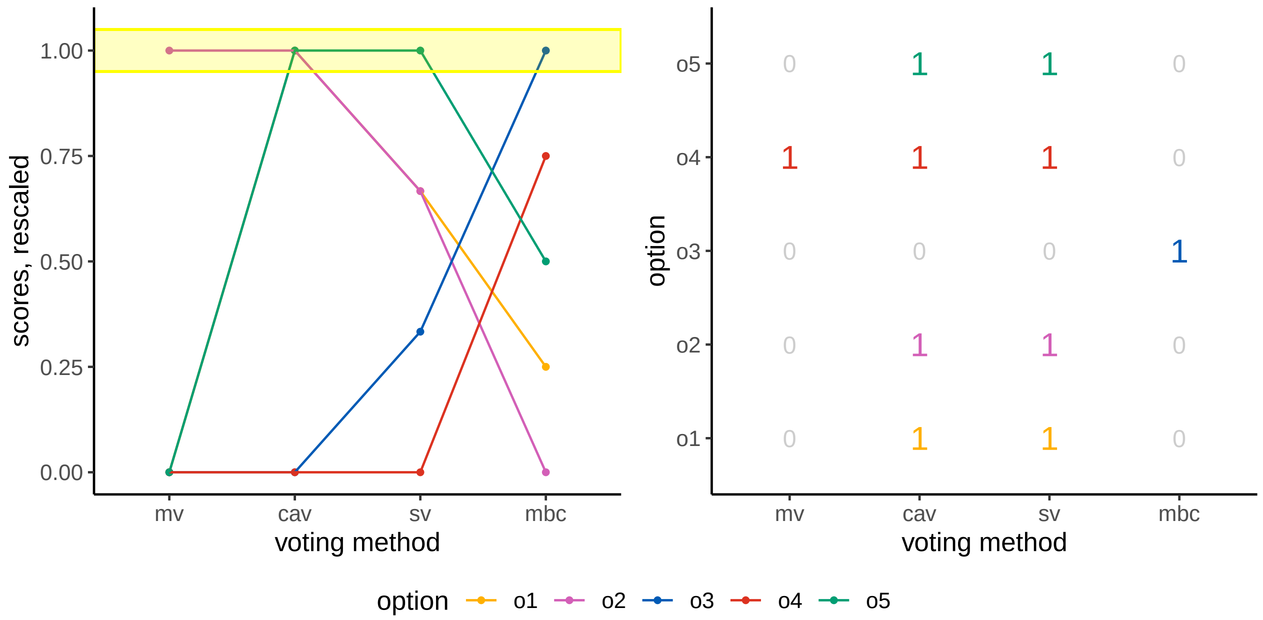


Figure 6: The highest valued choice—highlighted in yellow in the left panel—is of special importance; therefore, we reduce the dimensionality of a participant's full choice profile (left panel) to its first derivation (right panel). Plotted are sample values of one participant to one question. Padding open spaces with 0 forms the max-choice profile. For each voting method, the options with the highest score are encoded by 1. The remaining ones are encoded with 0. These encodings will be used to compare the preference profiles of different voting methods.

with satisfaction.

The basis for clustering is a voter's *pp*. For clustering, a similar problem as described in Section 4.6 emerges: The number of possible combinations is extremely large and exceeds the number of observations by far. Therefore, we need to reduce dimensionality. The max-choice profile is a highly reduced representation. To retain more information, a less reduced form is obtained by calculating the mean. Specifically, we average the (rescaled) scores per option over the four input methods as follows:

$$\mu_{o1} = \frac{1}{4}(s_{mv} + s_{cav} + s_{sv} + s_{mbc}) \qquad (3)$$

We cluster $pp_i = \{\mu_{o1}, \mu_{o2}, \mu_{o3}, \mu_{o4}\}$.

To determine the number of clusters, we calculate nine cluster evaluation indices and deploy the most frequent number of clusters. For all questions, three clusters are proposed, as listed in Table A20.

To increase the robustness of the clustering results, we deploy nine clustering methods from various categories [20], as listed in Table A21.

We investigated the relationship between legitimacy ratings and *COVID-19* preference clusters. The Kruskal-Wallis test p-values, presented in Table 3, suggest that there is no significant difference between the groups. As a result, we conclude that participants' evaluations of the voting methods were not influenced by their preferences for the *COVID-19* related choices being rated.

Table 3: Kruskal-Wallis Test p-values of legitimacy ratings by voting method, question, and cluster: evaluating the significance of differences among groups.

	vaccine	icu	protection	lockdown
mv	0.163	0.721	0.548	0.583
cav	0.458	0.501	0.139	0.361
sv	0.0576	0.253	0.198	0.839
mbc	0.531	0.36	0.22	0.951

Author Contributions

CIH: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. RH: Review & Editing, Funding acquisition. DH: Encouraged and supported/enabled the study on different voting methods, editing and review, funding acquisition. RK: Software. JW: Cluster Analysis. EP: Conceptualization, Methodology, Software, Investigation, Writing - Review & Editing, Supervision, Funding acquisition.

Acknowledgments

This study was financed by the Swiss National Science Foundation (SNSF) as part of the National Research Programme NRP77 Digital Transformation, project no. 187249 [Applicants: Regula Hänggli Fricker, Evangelos Pournaras, Dirk Helbing].

The work of EP is supported by a UKRI Future

Leaders Fellowship (MR-/W009560/1): *Digitally Assisted Collective Governance of Smart City Commons–ARTIO*, the Swiss National Science Foundation NRP77 'Digital Transformation' project (#407740_187249): *Digital Democracy: Innovations in Decision-making Processes*, the White Rose Collaboration Fund: *Socially Responsible AI for Distributed Autonomous Systems* and a 2021 Alan Turing Fellowship.

References

[1] Mikael Persson, Peter Esaiasson, and Mikael Gilljam. "The effects of direct voting and deliberation on legitimacy beliefs: An experimental study of small group decision-making". In: *European Political Science Review* 5.3 (2013), pp. 381–399. DOI: 10.1017/S1755773912000173.

[2] Cillian McBride. "Consensus, Legitimacy, and the Exercise of Judgement in Political Deliberation". In: *Critical Review of International Social and Political Philosophy* 6.3 (2003), pp. 104–128. DOI: 10.1080/1369823032000233573.

[3] Shmuel Nitzan. "The vulnerability of point-voting schemes to preference variation and strategic manipulation". In: *Public Choice* 47.2 (1985), pp. 349–370. ISSN: 00485829. DOI: 10.1007/BF00127531/METRICS.

[4] Peter Emerson. "The Consequences of Consequences". In: *Democratic Decision-making*. Ed. by Peter Emerson. Springer, 2021, pp. 41–51. ISBN: 978-3-030-52808-9.

[5] Peter Emerson. *Majority voting as a catalyst of populism*. Springer, 2020. ISBN: 978-3-030-20219-4.

[6] Hannah Werner and Sofie Marien. "Process vs. outcome? How to evaluate the effects of participatory processes on legitimacy perceptions". In: *British Journal of Political Science* 52.1 (2022), pp. 429–436.

[7] Tom R. Tyler. "Psychological Perspectives on Legitimacy and Legitimation". In: *Annual Review of Psychology* 57.1 (2006), pp. 375–400. DOI: 10.1146/annurev.psych.57.102904.190038.

[8] Fritz W Scharpf. *Regieren in Europa: Effektiv und demokratisch?* Campus Verlag, 1999. ISBN: 3593361116.

[9] Vivien A Schmidt. "Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'". In: *Political Studies* 61.1 (2013), pp. 2–22. DOI: 10.1111/j.1467-9248.2012.00962.x.

[10] Leopold Weil and Regula Hänggli. "How innovation in participation could increase legitimacy". 2021.

[11] Pietro Maffettone and Luke Ulaş. "Legitimacy, metacoordination and context-dependence". In: *International Theory* 11.1 (2019), pp. 81–109. DOI: 10.1017/S1752971918000258.

[12] Alex Bitektine and Patrick Haack. "The "Macro" and the "Micro" of Legitimacy: Toward a Multilevel Theory of the Legitimacy Process". In: *The Academy of Management Review* 40.1 (2015), pp. 49–75. DOI: 10.5465/amr.2013.0318.

[13] Leigh Plunkett Tost. "An Integrative Model of Legitimacy Judgments". In: *Academy of Management Review* 36.4 (2011), pp. 686–710. DOI: 10.5465/AMR.2010.0227.

[14] Frank Wilcoxon and American Cyanamid. "Individual Comparisons by Ranking Methods". In: *Breakthroughs in statistics* (1992), pp. 196–202. DOI: 10.1007/978-1-4612-4380-9{_}16.

[15] Regula Haenggli. *The Origin of Dialogue in the News Media (Challenges to Democracy in the 21st Century*. Palgrave Macmillan, 2020. ISBN: 3030265811.

[16] Hanspeter Kriesi. *Direct democratic choice: The Swiss experience*. Lexington Books, 2005.

[17] Peter Selb, Hanspeter Kriesi, Regula Hänggli, and Mirko Marr. "Partisan choices in a direct-democratic campaign". In: *European Political Science Review* 1.1 (2009), pp. 155–172. ISSN: 1755-7747. DOI: 10.1017/S175577390900006X.

[18] Carlos Navarrete et al. *Understanding Political Agreements and Disagreements: Evidence from the 2022 French Presidential Election*. 2022. DOI: 10.48550/arxiv.2211.04577.

[19] Johannes Schmitt. "How to Measure Ideological Polarization in Party Systems". In: *ECPR Graduate Student Conference* 2016 (2016).

[20] Dongkuan Xu and Yingjie Tian. "A Comprehensive Survey of Clustering Algorithms". In: *Annals of Data Science 2015 2:2* 2.2 (2015), pp. 165–193. DOI: 10.1007/S40745-015-0040-1.

[21] Leon Festinger. "Cognitive dissonance". In: *Scientific American* 207.4 (1962), pp. 93–106. ISSN: 0036-8733.

[22] Matías Núñez. "The strategic sincerity of Approval voting". In: *Economic Theory* 56.1 (2014), pp. 157–189. DOI: 10.1007/s00199-013-0775-x.

A Supplementary Materials

A.1 Details on Input Methods

Table A1: Overview of implemented input methods. The *score* refers to the numerical value assigned to a choice, *framing* refers to the verbal explanation of the voting method provided, and *exclusive* indicates whether the same score s can be assigned multiple times.

	score s	exclusive	framing
majority voting (*mv*)	[0,1]	yes	no
combined approval voting (*cav*)	[-1,0,1]	no	disapprove, neutral, approve
score voting (*sv*)	[0,1,2,3,4]	no	no
modified Borda count (*mbc*)	[0,1,2,3,4]	yes	approve

A.2 Distribution of Scores

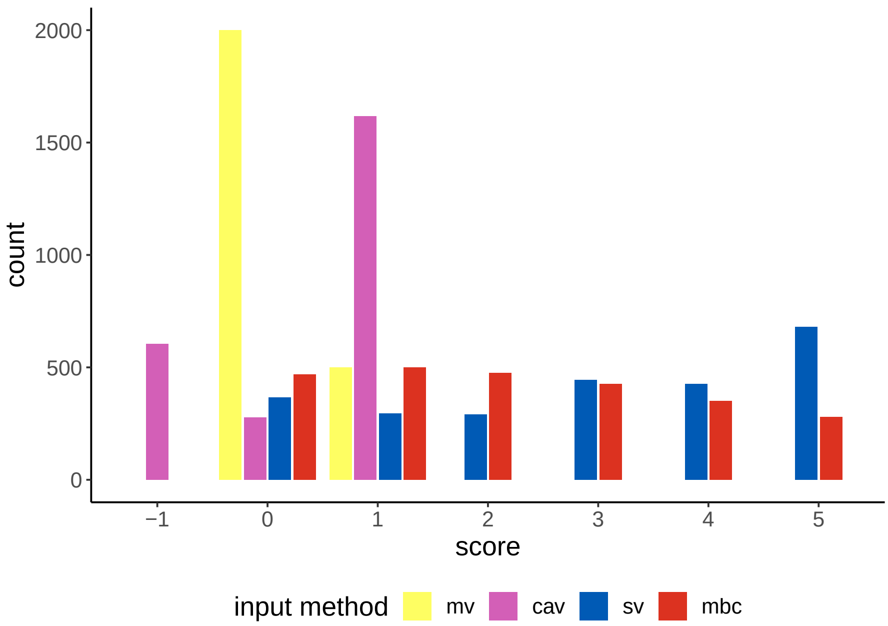


Figure A1: The plot displays the count of participants' votes, mapped to a numerical score s, by input method.

A.3 Condorcet Alternative

Table A2: Winning options as determined by the Condorcet method, separated by input format (rows) and question (columns). Protection experiences the least amount of change (two different winning options), whereas vaccine, icu, and lockdown experience three different options as winners.

Voting Method	Question			
	vaccine	*icu*	*protection*	*lockdown*
mv	o_1	o_3	o_3	o_2
cav	o_2	o_5	o_4	o_4
sv	o_4	o_3	o_3	o_5
mbc	o_2	o_1	o_3	o_2

A.4 Winning Options for *color*

Table A3: Winning options of the color context, identified by the highest total score, by voting method (columns).

	mv	cav	sv	mbc
option	blue	yellow	blue	red

A.5 Calculating the Number of Distinct Preference Profiles

Preferences pp_i^q can be interpreted as a *five*-dimensional series (five options o) with *four* time steps (four input methods im) each. The theoretical number of possible pp_i^q is extremely large. It is calculated as follows:

- Majority Vote: o
- Combined Approval Voting: s^o
- Score Voting: s^o
- modified Borda Count: $\sum_s^o$

where o stands for the options to rate and s identifies the dimension of the respective scale. Therefore, the possible number of distinct pp_i^q can be calculated as follows:

$$5 \times 3^5 \times 6^5 \times (1! + 2! + 3! + 4! + 5!) = 1'445'519'520 \quad (4)$$

A.6 Details on Legitimacy

Table A4: Shapiro-Wilk test for assessing the normality of the legitimacy ratings across different input methods.

which_method	statistic	p
mv	c(W = 0.88)	1.78e-08
cav	c(W = 0.88)	1.55e-08
sv	c(W = 0.75)	4.48e-13
mbc	c(W = 0.86)	2.05e-09

Table A5: Mean and median legitimacy ratings by context and input method.

methods	mean covid	median covid	mean color	median color
mv	1.725	1	2.258	3
cav	2.600	3	2.592	3
mbc	2.742	3	2.842	3
sv	3.208	4	2.958	3

Table A6: Mean and median legitimacy ratings by consistent and inconsistent voters.

methods	mean consistent	mean inconsistent	median consistent	median inconsistent
mv	1.570	2	1	2
mbc	2.910	2.440	3	3
cav	2.620	2.560	3	3
sv	3.360	2.930	4	3

Table A7: Paired Wilcoxon signed-rank test to compare legitimacy ratings within the covid context.

group1	group2	p.adj	p.signif
mv	cav	4.8e-07	****
mv	sv	1.6e-11	****
mv	mbc	1.7e-07	****
cav	sv	0.00012	****
cav	mbc	0.43	ns
sv	mbc	0.00073	***

The p-values are adjusted using the Holm-Bonferroni correction method.
*p<0.05; **p<0.01; ***p<0.001; ****p<1e-04.

Table A8: Paired Wilcoxon signed-rank test to compare legitimacy ratings within the color context.

group1	group2	p.adj	p.signif
mv	cav	0.12	ns
mv	sv	6.7e-05	****
mv	mbc	0.0017	***
cav	sv	0.011	**
cav	mbc	0.11	*
sv	mbc	0.45	ns

The p-values are adjusted using the Holm-Bonferroni correction method.
*p<0.05; **p<0.01; ***p<0.001; ****p<1e-04.

Table A9: A paired Wilcoxon signed-rank test is used to compare the legitimacy ratings across the color and covid context.

which_method	group1	group2	statistic	p	p.signif
mv	color	covid	c(V = 1890.5)	7.02e-06	****
cav	color	covid	c(V = 1019.5)	0.759	ns
sv	color	covid	c(V = 648)	0.0434	*
mbc	color	covid	c(V = 1457)	0.404	ns

The p-values are adjusted using the Holm-Bonferroni correction method.
*p<0.05; **p<0.01; ***p<0.001; ****p<1e-04.

Table A10: Wilcoxon rank sum test to compare legitimacy ratings across consistent and inconsistent voters.

which_method	statistic	p	p.signif
mv	c(W = 1330.5)	0.0667	ns
cav	c(W = 1794.5)	0.431	ns
sv	c(W = 1998)	0.0387	*
mbc	c(W = 2053)	0.0228	*

The p-values are adjusted using the Holm-Bonferroni correction method.
*p<0.05; **p<0.01; ***p<0.001; ****p<1e-04.

Table A11: Wilcoxon signed-rank test to compare legitimacy ratings across gender.

which_method	group1	group2	n1	n2	statistic	p
mv	female	male	43	75	c(W = 1589.5)	0.897
cav	female	male	43	75	c(W = 1619.5)	0.97
sv	female	male	43	75	c(W = 1643)	0.854
mbc	female	male	43	75	c(W = 1644)	0.856

Table A12: Wilcoxon signed-rank test to compare legitimacy ratings across levels of education.

which_method	group1	group2	n1	n2	statistic	p	p.adj	p.adj.signif
mv	high school	Abitur	10	31	c(W = 149)	0.861	1	ns
mv	high school	college	10	4	c(W = 17.5)	0.769	1	ns
mv	high school	bachelor	10	39	c(W = 145)	0.206	1	ns
mv	high school	master	10	32	c(W = 115)	0.173	1	ns
mv	high school	PhD	10	4	c(W = 12.5)	0.303	1	ns
mv	Abitur	college	31	4	c(W = 58)	0.85	1	ns
mv	Abitur	bachelor	31	39	c(W = 469.5)	0.1	1	ns
mv	Abitur	master	31	32	c(W = 372)	0.078	1	ns
mv	Abitur	PhD	31	4	c(W = 46.5)	0.42	1	ns
mv	college	bachelor	4	39	c(W = 64)	0.561	1	ns
mv	college	master	4	32	c(W = 52)	0.549	1	ns
mv	college	PhD	4	4	c(W = 7)	0.877	1	ns
mv	bachelor	master	39	32	c(W = 609)	0.863	1	ns
mv	bachelor	PhD	39	4	c(W = 83.5)	0.83	1	ns
mv	master	PhD	32	4	c(W = 65.5)	0.958	1	ns
cav	high school	Abitur	10	31	c(W = 183)	0.386	1	ns
cav	high school	college	10	4	c(W = 32)	0.088	1	ns
cav	high school	bachelor	10	39	c(W = 206.5)	0.777	1	ns
cav	high school	master	10	32	c(W = 150.5)	0.777	1	ns
cav	high school	PhD	10	4	c(W = 18)	0.819	1	ns
cav	Abitur	college	31	4	c(W = 86.5)	0.2	1	ns
cav	Abitur	bachelor	31	39	c(W = 543.5)	0.461	1	ns
cav	Abitur	master	31	32	c(W = 388.5)	0.125	1	ns
cav	Abitur	PhD	31	4	c(W = 45)	0.376	1	ns
cav	college	bachelor	4	39	c(W = 41)	0.116	1	ns
cav	college	master	4	32	c(W = 25.5)	0.044	0.667	ns
cav	college	PhD	4	4	c(W = 2.5)	0.137	1	ns
cav	bachelor	master	39	32	c(W = 560.5)	0.448	1	ns
cav	bachelor	PhD	39	4	c(W = 66.5)	0.634	1	ns
cav	master	PhD	32	4	c(W = 62)	0.936	1	ns
sv	high school	Abitur	10	31	c(W = 153.5)	0.974	1	ns
sv	high school	college	10	4	c(W = 17.5)	0.752	1	ns
sv	high school	bachelor	10	39	c(W = 182)	0.728	1	ns
sv	high school	master	10	32	c(W = 169.5)	0.772	1	ns
sv	high school	PhD	10	4	c(W = 23)	0.701	1	ns
sv	Abitur	college	31	4	c(W = 54.5)	0.689	1	ns
sv	Abitur	bachelor	31	39	c(W = 570.5)	0.659	1	ns
sv	Abitur	master	31	32	c(W = 528)	0.637	1	ns
sv	Abitur	PhD	31	4	c(W = 71)	0.631	1	ns
sv	college	bachelor	4	39	c(W = 84.5)	0.775	1	ns
sv	college	master	4	32	c(W = 74.5)	0.578	1	ns
sv	college	PhD	4	4	c(W = 9)	0.868	1	ns
sv	bachelor	master	39	32	c(W = 689)	0.408	1	ns
sv	bachelor	PhD	39	4	c(W = 91)	0.557	1	ns
sv	master	PhD	32	4	c(W = 65)	0.978	1	ns
mbc	high school	Abitur	10	31	c(W = 110.5)	0.165	1	ns
mbc	high school	college	10	4	c(W = 4)	0.021	0.295	ns
mbc	high school	bachelor	10	39	c(W = 95.5)	0.01	0.152	ns
mbc	high school	master	10	32	c(W = 98)	0.053	0.685	ns
mbc	high school	PhD	10	4	c(W = 22.5)	0.767	1	ns
mbc	Abitur	college	31	4	c(W = 35)	0.152	1	ns
mbc	Abitur	bachelor	31	39	c(W = 481)	0.127	1	ns
mbc	Abitur	master	31	32	c(W = 468.5)	0.696	1	ns
mbc	Abitur	PhD	31	4	c(W = 79.5)	0.362	1	ns
mbc	college	bachelor	4	39	c(W = 94)	0.488	1	ns
mbc	college	master	4	32	c(W = 91)	0.15	1	ns
mbc	college	PhD	4	4	c(W = 13)	0.178	1	ns
mbc	bachelor	master	39	32	c(W = 732)	0.187	1	ns
mbc	bachelor	PhD	39	4	c(W = 111)	0.152	1	ns
mbc	master	PhD	32	4	c(W = 86.5)	0.242	1	ns

The Abitur is the German version of a high school diploma after 12 or 13 years of schooling.

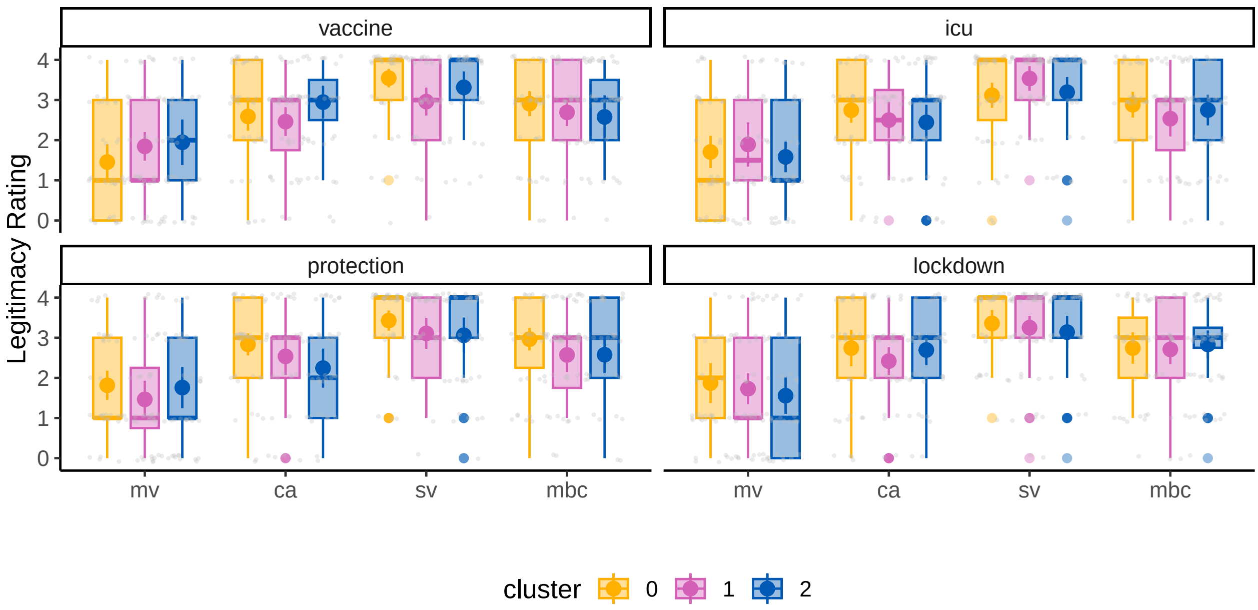


Figure A2: Boxplot comparison of legitimacy ratings by voting method, question, and cluster of *COVID-19* preference profiles. The plot provides insights into the dependence of legitimacy ratings on the interaction of voting method, question, and *COVID-19* preference profile.

Furthermore, we test whether legitimacy ratings vary by socio-demographic characteristics. We find that ratings are unaffected by gender (Table A11) and level of education (Table A12). However, correlations with age are interesting: Figure A3 shows that legitimacy ratings on majority vote are low among the youngest and high among the oldest. The majority votes require the lowest cognitive load; therefore, this observation seems reasonable. We do not have enough observations to test for significance and is therefore left for future research to investigate.

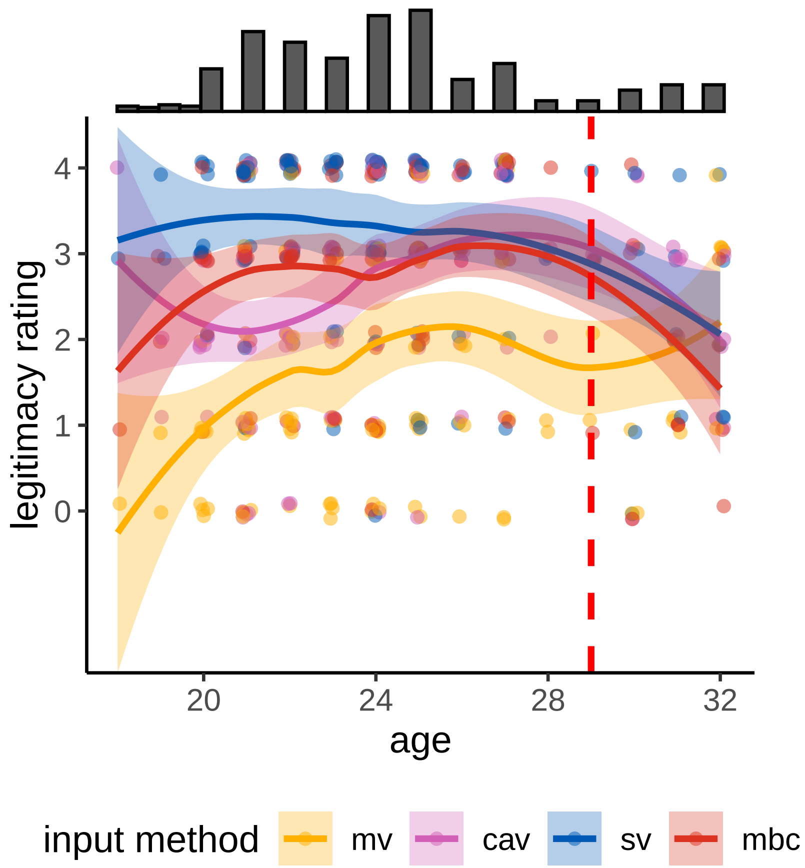


Figure A3: Dots represent a participant's age and her/his respective legitimacy rating. Lines represent Local Polynomial Regression Fitting (y x) with 95 percent confidence intervals. Age label > 32 refers to age 32 or older. The Red dashed line indicates a continuous raise for "mv": Participants aged 28 or older consider mv as increasingly legitimate, whereas the remaining three voting methods decrease in their legitimacy ratings. Figure legend shows abbreviations for the voting methods used: Majority voting (mv), combined approval voting (cav), score voting (sv), and modified Borda Count (mbc).

A.7 Details on σ and D

Table A13: Shapiro-Wilk test for normality of the standard deviation of the votes.

variable	statistic	p
sd	0.669	4.16e-47

Table A14: Shapiro-Wilk test for normality of divisiveness of the voting results.

variable	statistic	p
divisiveness	0.925	1.29e-11

Table A15: A paired Wilcoxon signed-rank test is used to compare the standard deviation of the voting results across questions.

group1	group2	statistic	p.adj	p.adj.signif
icu	lockdown	c(V = 23320.5)	2.51e-08	****
icu	protection	c(V = 26080)	2.83e-13	****
icu	vaccine	c(V = 13583)	0.595	ns
lockdown	protection	c(V = 13576.5)	0.031	*
lockdown	vaccine	c(V = 9043)	2.51e-08	****
protection	vaccine	c(V = 7343.5)	1.3e-14	****

The p-values are adjusted using the Holm-Bonferroni correction method.
*p<0.05; **p<0.01; ***p<0.001; ****p<1e-04.

Table A16: A paired Wilcoxon signed-rank test is used to compare the Divisiveness of the voting results across questions.

question I	question II	statistic	p.adj	signif
vaccine	icu	c(V = 857)	0.766	ns
vaccine	protection	c(V = 1525)	3.62e-05	****
vaccine	lockdown	c(V = 1385)	0.002	**
icu	protection	c(V = 1581)	5.78e-06	****
icu	lockdown	c(V = 1424)	0.000728	***
protection	lockdown	c(V = 796)	0.766	ns

The p-values are adjusted using the Holm-Bonferroni correction method.
*p<0.05; **p<0.01; ***p<0.001; ****p<1e-04.

A.8 Details on Divisiveness

Divisiveness D is calculated as follows: For each pair of options o_i, o_j (subsetted by question and voting method), we split the population into those who do and do not prefer one particular option. Subsequently, we calculate the absolute differences in scores s. This process is repeated for all possible pairs, absolute differences are summed and then divided by the number of votes/voters (which is equal when subsetting by question and voting method) n.

$$D = \frac{1}{n-1} \sum_{j!=i} \|s(o_i, o_j) - s(o_j, o_i)\| \tag{5}$$

where $s(o_i, o_j)$, $s(o_j, o_i)$ are the scores from the population of users that preferred option i and j respectively, while n is the number of voters.

Under majority voting, $D = 1$ for all questions, as the voting method allows for only two scores $\in 0, 1$. Consequently, the within-person (absolute) difference $\|s(o_i, o_j) - s(o_j, o_i)\|$ is always 1. Similarly to the majority vote, the MBC displays similar values of divisiveness across questions. The method requires ranking and therefore does not allow assigning the same score twice. This reduces the variation of the distribution of differences in scores across participants. Therefore, mean divisiveness scores by question are similar.

The distribution of divisiveness scores across questions is interesting for the remaining two voting methods. Combined approval and score voting allow for assigning the same score multiple times. Therefore, the distribution of differences can vary tremendously across participants. Figure A4 shows that under both voting methods, questions *lowckdown* and *protection* show lower divisiveness scores than questions *icu* and *vaccine*. This means that options for questions *lockdown* and *protection* more often received similar scores. This can be interpreted as it was hard for participants to rank options for these two questions.

The interpretation of divisiveness shows that the metric is meaningful for voting methods that do not enforce explicit ranking of options. In this case, divisiveness indicates whether participants' preferences are ordered or similar in value. The lower the divisveness score the less participants' have ordered preferences.

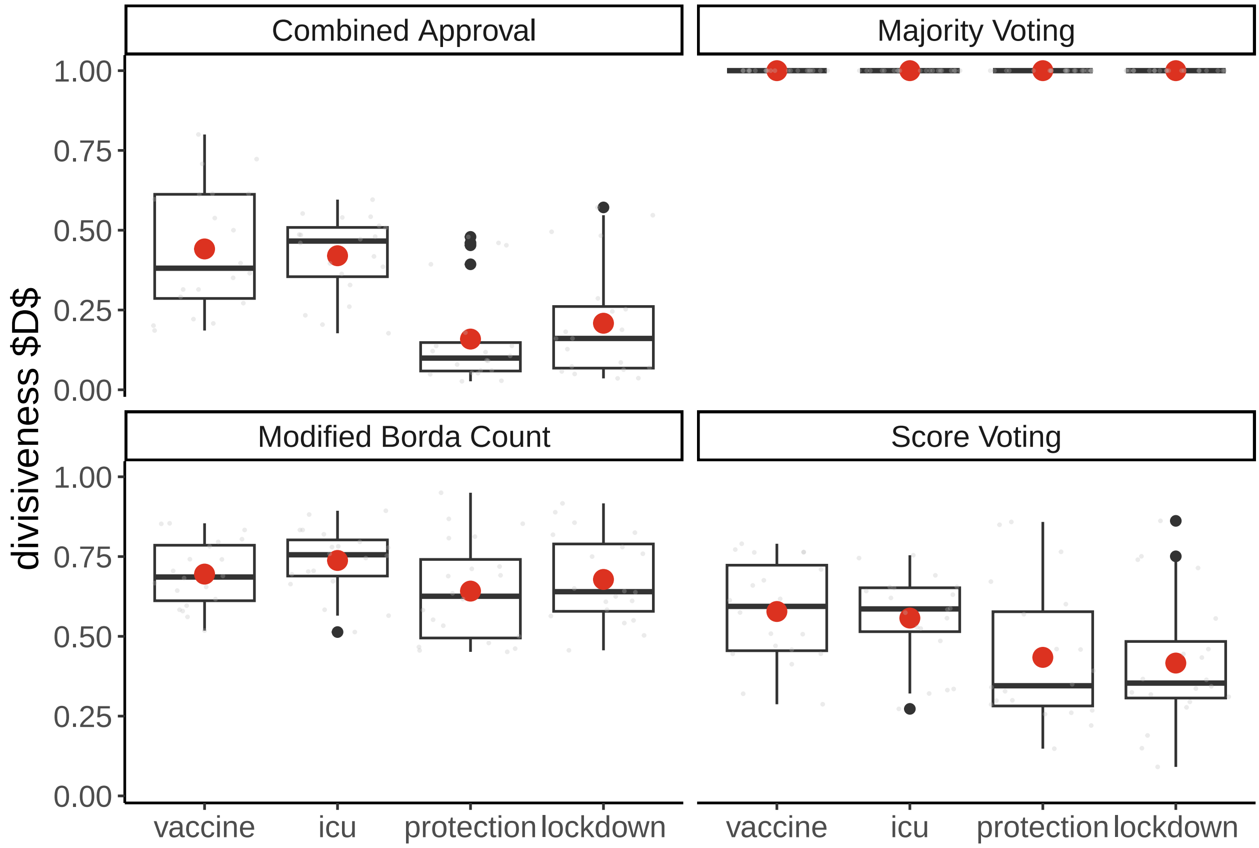


Figure A4: Divisivess (scaled within method) by question. Jitter represents divisiveness scores by option-pair.

A.9 Details on the Experiment

All experimental protocols were approved by the ETH ethics commission and carried out in accordance with the relevant guidelines and regulations. Furthermore, informed consent was obtained from all participants. The experiment was conducted online via Qualtrics, in collaboration with the ETH Decision Science Laboratory (DeSciL). It was conducted in three sessions in July 2021. Our experiment had three stages: Introduction and practice (I), vote (II), and answer control questions (III). Stages I and III were answered on Qualtrics, stage II on VoteApp found at `https://play.google.com/store/apps/details?id=ch.ethz.coss.votedifferent.Theapplicationfocusesoncomparingdifferentvotingmethods,collectsmeta-dataaboutthevotingprocess(timeandclickbehavior),offersabuilt-infeedbacksystemthatguidesthe andpreservesanonymityusingnon-reversible,deterministichashingmethods.` an application specifically developed for that purpose.

In stage (I) introduction and practice, participants first read the *COVID-19* questions and then subsequently saw a schematic presentation of the voting methods. This separate introduction of the method and question is expected to facilitate comprehension. In stages I and II, participants encountered four different voting methods.

Figure 5 shows screenshots of the interface VoteApp presented to the participants.

The same aggregation method across all input formats was communicated to participants: The option with the highest sum wins. To check for the stability of results in dependence of aggregation, we will also calculate the Condorcet method.

On every screen, VoteApp and Qualtrics track time. Time gives us an indirect estimation of the cognitive effort needed to apply one voting mechanism. We reach the most accurate time measurement if the participants are familiar with both the questions and the voting method. Therefore, we introduce the questions and the voting methods separately. The cognitive effort allows us to predict whether citizens are willing to vote via methods that require more mental resources. It could be that consensus-based voting mechanisms increase legitimacy, but the citizens are unwilling to engage in them because it is cognitively too demanding.

The legitimacy question is to be answered twice for every voting method: (1) Context independent and (2) context-dependent. To get a context-independent answer to the legitimacy question, participants are introduced to the four voting mechanisms within a neutral setting: They should choose their favorite color. *Color* is a topic with a low impact on societal welfare compared to policies concerning epidemics. Subsequently, they answer

	mv	cav	sv	mbc	**all**
Color	LR	LR	LR	LR	LR
COVID-19	LR	LR	LR	LR	LR
all	LR	LR	LR	LR	LR

Table A17: Structured overview of elicited legitimacy ratings. Participants in our experiment provided legitimacy ratings LR for every input method $im \in \{mv, cav, sv, mbc\}$ across two contexts $c \in \{color, COVID-19\}$ on a Likert scale $LR = \{0, ..., 4\}$.

the legitimacy question. Participants answer the legitimacy question directly after voting on the COVID-19 questions to get a context-dependent answer.

Table A17 provides an overview of all legitimacy ratings provided.

Survey theory shows that the ordering of questions and answering options does matter. The most crucial impact of the sequence is that the first and last questions/options get more attention than those in-between. We expect a similar effect in our experiment, too. Therefore, we randomize both, the questions and the options. Humans have the urge to be consistent [21]; therefore, rating via the first voting mechanisms will influence subsequent voting. For this reason, we let participants start with the "easiest" vote (plurality) and proceed with voting mechanisms, gradually increasing their requirements to score and order all options listed. The order will only need to be randomized if we theorize an unambiguous impact on legitimacy. However, we do not.

A.10 Details on the Poisson Regression Model

A Poisson Regression model is a Generalized Linear Model (GLM) that is used to model count data. The dependent variable y follows the Poisson distribution. To transform the non-linear relationship to linear form, a link function is used which is the *log* for Poisson Regression. The mathematical form of Poisson Regression model is:

$$\log(y) = \alpha + \sum_{i=1}^{p} \beta_{i,x_i} \tag{6}$$

Where, y is the dependent variable, α is the intercept and β are numeric coefficients, and x is the explanatory variable. The iteratively reweighted least squares method is used to find the maximum likelihood estimation of the coefficients.

Table A18: P-Values of Poisson regression analysis: Comparison of different choice profiles with respect to omitted category 'icu access'. The counts for four choice profiles differ significantly by question.

	0110	0100	0000	1111
(Intercept)	1.17e-231	0.00e+00	0.00e+00	0.00e+00
lockdown	1.10e-02	4.08e-08	4.10e-10	6.51e-01
infection	4.96e-04	4.22e-10	3.96e-16	8.20e-01
vaccine	2.18e-01	1.83e-01	4.09e-01	4.78e-01

A.11 Details on Strategic Voting

Some input methods we deploy are vulnerable towards strategic voting: Approval voting, alongside other rank scoring rules, is shown not necessarily to trigger sincere voting behavior [22]. Therefore, we ask participants whether they voted strategically (Appendix A.11).

To determine if participants engaged in strategic voting, we asked them directly through an open-ended textbox. Of the 120 participants, 62 provided a response. Of these, 77.4% stated that they either had no expectations or voted sincerely based on their beliefs. To identify instances of sincere voting, we searched for key words such as "not adjust," "my opinion," "my belief," "my true," and "my individual." The remaining responses, which couldn't be classified using these keywords, are included below. Out of the responses, only one instance of strategic voting was reported.

Table A19: The open-ended response to the question regarding whether the participant voted strategically. Only answers that did not contain keywords indicating negation are displayed.

Answers
i thought others should have same opinions as i have, though the outcome surprised me a little
for my second or third choice depending on my vote i can ”push” my first choice
i expected more or less the same voting results
others voted the complete opposite as me
as those are discussed themes i think that different opinions will come.
i favoured the outcome for the greater good deed
sometimes it was difficult to choose and i went by my gut feeling
i expect others will have similar voting.
i think most of the people agree with my choices
i just chose the answer which was right one
they make opinions based on media

A.12 Duration

The duration of time spent on casting each vote was analyzed to assess participants' understanding of the voting methods and questions. As shown in Figure A5, the time spent by participants on each vote is displayed in seconds, with consideration given to the fact that a practice round was conducted beforehand to equalize reaction times. A noticeable increase in the duration can be observed for Vote 04, which was the first time the MBC method was introduced. Additionally, Vote 13, which was the first time participants casted their vote on Question 4 ”The government's strategy to fight the COVID-19 pandemic should also mitigate”, shows a higher duration, indicating that this question might have been more difficult to comprehend and potentially poorly posed.

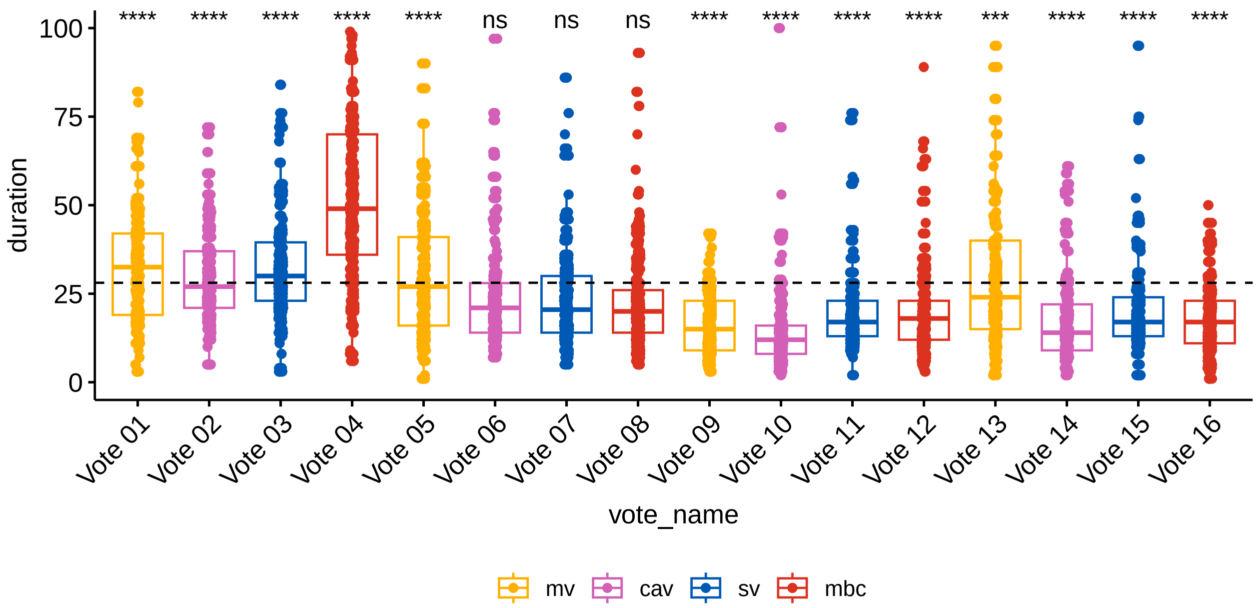


Figure A5: The plot displays the results of a Wilcoxon test comparing the mean time spent by participants on each question and voting method with the overall group mean (dashed line). Significance stars indicate the level of significance of the pairwise comparisons.

A.13 *COVID-19* Preference-Clusters

Regarding the question 'What are you most concerned about with regards to the COVID-19 vaccines,' the first cluster exhibits a favorable attitude towards vaccination and expresses a desire to receive the vaccine promptly. The second cluster expresses concern about potential side effects but still desires to be vaccinated. The third cluster is primarily worried about the potential misuse of the vaccines.

Table A20: Results of cluster number detection. We choose the most frequent number of clusters of all cluster evaluation indices.

Method/Question	**Q1**	**Q2**	**Q3**	**Q4**
Elbow	3	3	3	3
AIC	3	8	3	7
BIC	3	3	3	3
Calinski-Harabasz Index	3	3	3	3
Davies Bouldin Index	4	7	6	4
Dendrogram	4	5	5	3
Gap Statistics	4	6	8	3
Silhouette Scores	5	4	3	5
PCA	3	1	1	3
Most frequent	3	3	3	3

Table A21: Comparison of clustering methods utilized in the study. The table showcases the categories of clustering techniques applied in the research, including fuzzy c-means (FCM) and Gaussian mixture models (GMM)

Method Category	**Method**
Partition-based	K-means, MiniBatch K-means
Hierarchy-based	BIRCH, Hierarchical Clustering (Ward, Average, Complete)
Fuzzy Theory-based	FCM
Distribution and Model-based	GMM
Spectral Graph Theory-based	Spectral

Regarding the question 'Among COVID-19 patients, which criteria should grant access to an intensive care unit?' the first cluster prioritizes following established rules, with the belief that those who abide by the rules should be protected. The second cluster places a strong emphasis on protecting the lives of young individuals, whereas the third cluster prioritizes the elderly.

Regarding the question 'Which is the most effective protection measure against a COVID-19 infection?' the first cluster favors the use of medical technology products, such as vaccines. The second cluster holds the view that the spread of the virus is independent of people's lifestyles. The trend of the third cluster is unclear and, as a result, we choose not to provide an interpretation.

Regarding the fourth question, "What is the most significant problem caused by the lockdown?" the first cluster highlights the economic impact of the lockdown, with a focus on financial difficulties and job losses. The second cluster focuses on broader societal issues, such as inequalities and class differentiation, resulting from the lockdown. Meanwhile, the third cluster expresses concern about the impact of lockdown on mental health and the trade-off between government control and personal freedom.

A.14 Qualtrics Survey

The following pages showcase the introduction of the incentivized experiment and voting methods to participants, which took place on Qualtrics. Participants were given this introduction before they cast their actual votes via the voting app. The screenshots depict the Qualtrics section of the experiment.

Intro

Dear participant, thank you for taking part in our research study. Before we begin, please read the study information carefully. If you decide to participate, we ask you to give informed consent.

Conducting person (full name): Dr. Carina I. Hausladen
Contact project team: carina.hausladen@gess.ethz.ch
Data Protection Officer ETHZ: Tomislav Mitar, tomislav.mitar@sl.ethz.ch

Welcome! Before the experiment begins, please read the text below carefully and ask the conducting person about anything you do not understand or would like to know.

What is the purpose of this study? What is investigated and why?
The study aims to gain insights into human behavior. For that purpose, the choices and statements you make are investigated. More precisely, we want to know, to what extent different voting mechanisms influence your decision and your satisfaction with the final outcome.

What am I asked to do?
During this study, you will be asked to vote upon four different statements in four different ways. Subsequently, you will be asked to rate your satisfaction regarding the outcome. Finally, we ask you to answer a short questionnaire.

Who can participate?
All participation is voluntary. You may discontinue participating at any time if you wish with no penalty or impact on your future relationship with ETH Zurich. You must be 18 or older to participate and own or have access to an Android phone. You can only participate in this very study once.

Will I be compensated for participating?
You can make money during the study. We expect that it will take you (at maximum) 30 minutes to finish the tasks. Based on this duration, you will receive a fixed payment of 15 CHF. The payout is anonymous: other participants will not see how much you receive.

How much time will I be expected to invest?
This study is expected to take 30 minutes.

Are there any risks or benefits?
There are no risks concerning your physical, mental, or psychical health above and beyond what could be expected to occur in daily living.

How am I insured?
Adverse health effects that are directly caused by the study and can be demonstrated to be attributable to fault on the part of the project team or ETH Zurich are covered by ETH's liability insurance (Basler Versicherungen, policy no. 30/4.078.362). You are responsible for insuring yourself against any other adverse health effects such as might occur, for instance, in connection with the trip to or from the place where the study is conducted.

What are my rights during participation?
Your participation in this study is voluntary. You may cancel your participation at any time without specifying reasons and without any disadvantages.

Which data is collected?
The experimental data of your decisions and possibly metadata, such as the time it takes to arrive at your decision, are anonymously stored. Your recruitment data are kept separately from the experimental data. That is, your identity or recruitment data will not be revealed to the experimenter, to researchers analyzing the data, or to other persons not authorized by you.

How are my data treated?
All experimental data is stored and processed in agreement with applicable national and EU laws. The experimental data are intended for scientific analysis and publication.

Scientific journals may publish the experimental data for the sake of reproducibility. This implies the possibility of secondary use of the experimental data. However, in no case will data about your identity and the experimental data be combined.
Additional building block: Members of the ETH Zurich Ethics Commission may access the original data for examination purposes. Strict confidentiality will be observed at any time.

What are my rights to the data?
You may request comprehensive information about the personal data that were collected from you in the study at any time. You also have the right to have them corrected, handed over to you, barred for processing, or deleted. You may revoke your consent to the processing of your data at any time without giving reasons. Your contact person for related requests is given above.

Who funds the study?
The study is funded by the National Research Programme “Digital Transformation” (NRP 77).

Who examined the study?
This study was approved by the ETH Zurich Ethics Commission as proposal 2021-N-28.

I, the participant, confirm that:

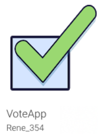
- I have read and understood the study information.
- My questions have been answered completely and to my satisfaction.
- I am aware of the requirements and restrictions to be observed during the study.
- I have had enough time to decide on my participation.
- I participate in this study voluntarily and consent that my data be used as described above.

○ I agree

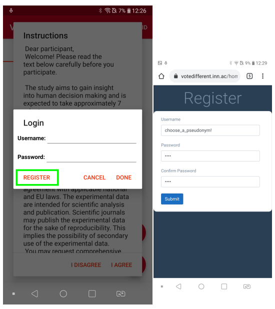
- I do not agree

install

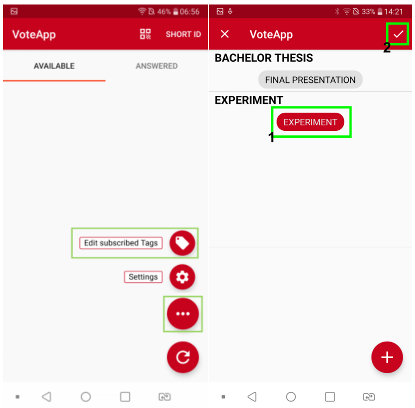
This study involves participation via an application — VoteApp.
Install VoteApp on your Android phone. You find the app on Google Play: https://play.google.com/store/apps/details?id=ch.ethz.coss.votedifferent&gl=CH



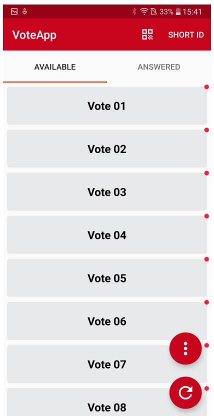
Register via VoteApp's login screen. Please choose a username that **does not resemble your name** and a password.



Subscribe to the tag "Experiment"

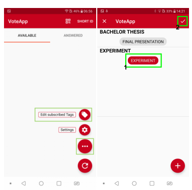


Can you see all 16 votes under the tag AVAILABLE?



- yes
- no

As you can not see all of the 16 votes, please make sure that you 1) subscribed to the tag EXPERIMENT. The button turns red if selected. Colours might differ if your phone is in dark mode.



2) pressed the refresh button (at the bottom right).

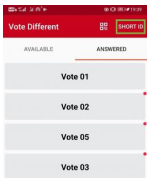
3) Closed and opened the app and repeat 1) and 2).

Can you now see all of the 16 votes under the tag AVAILABLE?

- ○ yes
- ○ no

If you can still not see all 16 votes, please contact the experimenter via Zoom.

Please enter your ID. You find the 5-digit number at the top right corner of VoteApp.



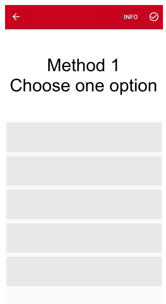
practice_methods

During this experiment, we will ask you to vote about **four questions** in **four ways**. Before you vote, we want to introduce first the voting mechanisms and then the questions to you.

In the following, you'll practice four different voting methods. You will encounter these methods later during this experiment, as well.

Now, please imagine, that together with the other participants of this experiment, you have to **choose a color** that you will encounter on an everyday basis.

Method 1
Please choose the color that you like most.



- ○ yellow
- ○ black
- ○ green
- ○ blue
- ○ red

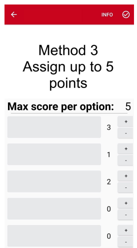
Method 2
Please approve (1), stay neutral (0), or disapprove(-1) of each of the colors listed below.

	disapprove (-1)	neutral (0)	approve (1)
yellow	○	○	○
black	○	○	○
green	○	○	○
blue	○	○	○

red	○	○	○

Method 3

Please score each color between 0 (like least) and 5 (like most).



	0	1	2	3	4	5
yellow	○	○	○	○	○	○
black	○	○	○	○	○	○
green	○	○	○	○	○	○
blue	○	○	○	○	○	○
red	○	○	○	○	○	○

Method 4

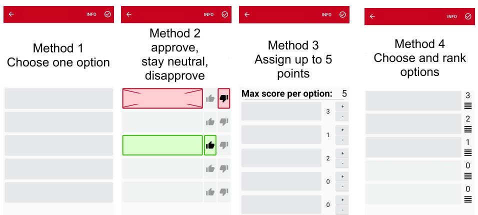
1. Please select the colors you would like to vote upon.
2. Subsequently, please order your selected options. The number next to the option is the number of votes you give to the respective option. The more options you selected, the more votes you can assign.

- ☐ yellow
- ☐ black
- ☐ green
- ☐ blue
- ☐ red

Method 4
Please order your selected options. The number next to the option is the number of votes you give to the respective option. The more options you selected, the more votes you can assign.

- » yellow
- » black
- » green
- » blue
- » red

You voted in four different ways. Now, please assess the following statement for each of the voting methods applied.



I would comply with the result and accept it as fair reflecting my and others' opinions.

	strongly disagree	somewhat disagree	neutral	somewhat agree	strongly agree
choose one option	○	○	○	○	○
approve, stay neutral, disapprove	○	○	○	○	○
assign points to options	○	○	○	○	○
choose and rank options	○	○	○	○	○

practice_questions

Later in this experiment, you will vote upon 4 different questions. Please read these questions on the following screens.

How should the government regulate COVID-19 vaccines?

- vaccination should be mandatory
- vaccination should be the basis for discrimination e.g. access to facilities and services
- alternatives to vaccination should be promoted and financially supported
- vaccination-effectiveness and long-term effects should be analyzed first in more detail
- choice of which vaccine to receive should be given

Among COVID-19 patients, which criteria should grant access to an intensive care unit?

- no denial of vaccination
- no violation of lock down rules
- no health self-damage, e.g. smoking, drugs, alcohol
- being the oldest
- being the youngest

Which protection measure(s) against a COVID-19 infection is the most effective one for governments to promote?

- regular hand washing
- wearing a mask
- maintaining a healthy lifestyle
- vaccination
- physical/social distancing

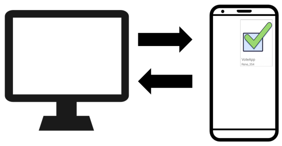
The government's strategy to fight the COVID-19 pandemic should also mitigate

- mental distress
- government control and suppression of freedom
- social segregation and increased inequality
- economic recession and unemployment
- reduced physical health condition

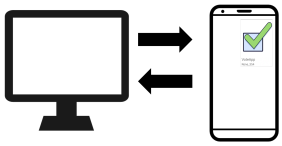
switch1

Now, please open **VoteApp** on your **phone** and participate in all of the 16 votes.

Subsequently, please proceed with this questionnaire.

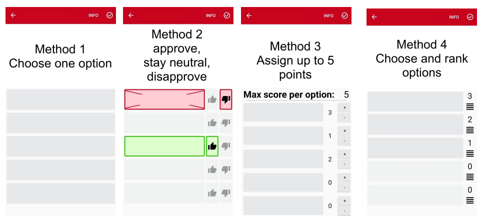


Did you participate in all of the 16 votes within VoteApp?



- ◯ yes
- ◯ no

You voted in four different ways. Now, please assess the following statement for each of the voting methods applied.



I would comply with the result and accept it as fair reflecting my and others' opinions.

	stronlgy disagree	somewhat disagree	neutral	somewhat agree	strongly agree
choose one option	◯	◯	◯	◯	◯
approve, stay neutral, disapprove	◯	◯	◯	◯	◯

assign points to options	○	○	○	○	○
choose and rank options	○	○	○	○	○

switch2

How polarized do you find the following statements in terms of how opinions in society vary from each other?

	not polarized	low polarized	neutral	somewhat polarized	strongly polarized
The long-term side effects of COVID-19 vaccines are concerning.	○	○	○	○	○
Governments and companies instrumentalize the COVID-19 vaccines.	○	○	○	○	○
The COVID-19 vaccines are used in a discriminatory way, e.g. travels, access facilities and services.	○	○	○	○	○
Early vaccination for COVID-19.	○	○	○	○	○
The overall effectiveness of the COVID-19 vaccines.	○	○	○	○	○
Denial to vaccination should be a criterion for a COVID-19 patient to grant him/her access to an intensive care unit.	○	○	○	○	○

No violation of lockdown rules should be a criterion for a COVID-19 patient to grant him/her access to an intensive care unit.	○	○	○	○	○
No health self-damage, e.g. smoking, drugs, alcohol, should be a criterion for a COVID-19 patient to grant him/her access to an intensive care unit.	○	○	○	○	○
Being the oldest should be a criterion for a COVID-19 patient to grant him/her access to an intensive care unit.	○	○	○	○	○
Being the youngest should be a criterion for a COVID-19 patient to grant him/her access to an intensive care unit.	○	○	○	○	○
Regular hand washing is an effective protection measure against COVID-19 infection.	○	○	○	○	○
Wearing a mask is an effective protection measure against COVID-19 infection.	○	○	○	○	○
Maintaining a healthy lifestyle is an effective protection measure against COVID-19 infection.	○	○	○	○	○
Vaccination is an					

effective protection measure against COVID-19 infection.	○	○	○	○	○
Physical/social distancing is an effective protection measure against COVID-19 infection.	○	○	○	○	○
The lockdown has caused mental distress.	○	○	○	○	○
The lockdown has caused government control and suppression of freedom.	○	○	○	○	○
The lockdown has caused social segregation and increased inequality.	○	○	○	○	○
The lockdown has caused economic recession and unemployment.	○	○	○	○	○
The lockdown has reduced physical health.	○	○	○	○	○

With which gender do you identify?

- ○ male
- ○ female
- ○ other

What is your age? (in years)

What is your country of origin?

What is the highest level of education that you obtained?

Before you voted on any of the 16 votes, did you vote solely based on your individual beliefs, or did you adjust your vote based on your expectation of how others voted?

- I voted solely based on my beliefs.
- I adjusted my vote based on my expectation of how others voted.

Which expectations did you have about how others voted and how did you adjust your vote?
For example, you did not indicate your true preferences but favored different options to alter the outcome.

Powered by Qualtrics

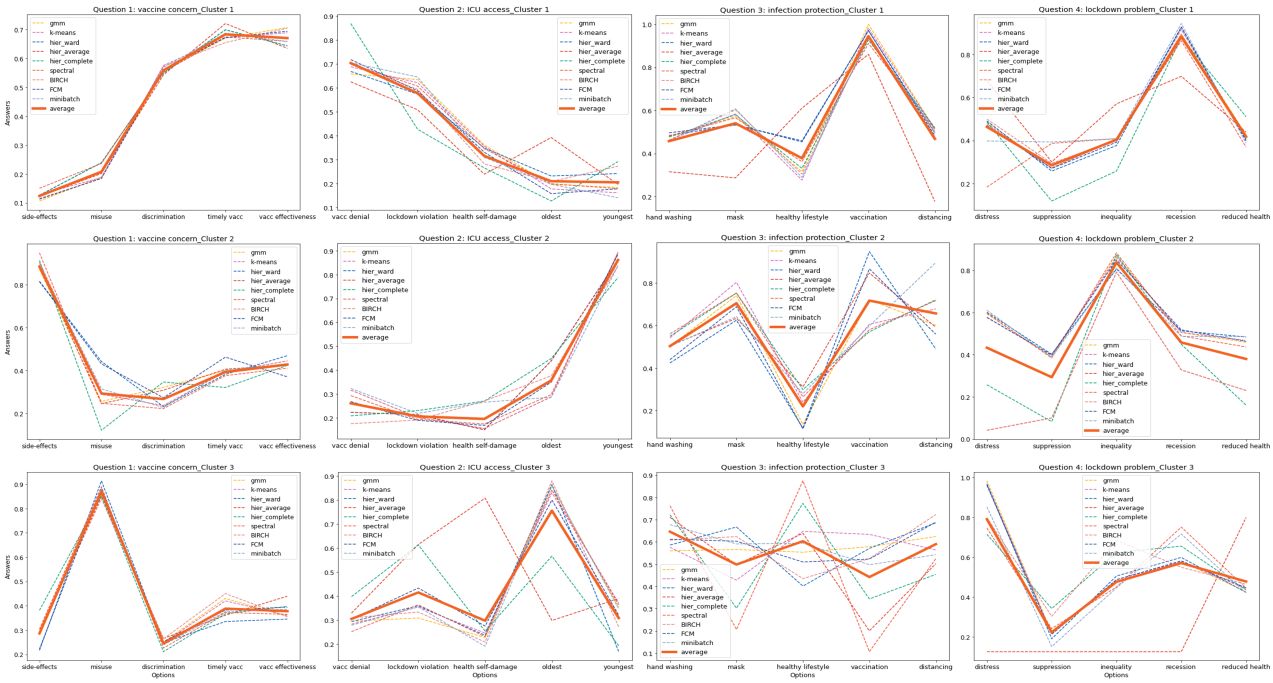


Figure A6: Clustering of individual participants' preference profiles. Each column represents a question, and each row represents one of three distinct clusters identified. The centroids of the various clustering algorithms are displayed within each panel.