ARTICLE

Check for updates

How voting rules impact legitimacy

Carina I. Hausladen [1✉], Regula Hänggli Fricker[2], Dirk Helbing[1,3], Renato Kunz[1], Junling Wang[1] & Evangelos Pournaras[4]

Collective action is essential for addressing the grand challenges of our time. However, for such action to be successful, decision-making processes must be perceived as legitimate. In this study, we investigate the legitimacy of different voting methods. Using a pre-registered human subject experiment, 120 participants cast their votes using four voting methods: majority voting, combined approval voting, range voting, and the modified Borda count. These methods represent a range of preference elicitation designs, from low to high complexity and flexibility. Furthermore, we developed a legitimacy scale upon which the participants rate the voting methods. The experiment was conducted in a non-political setting (voting on color preferences) and a political context (voting on COVID-19-related questions). Our findings suggest that the perceived legitimacy of a voting method is context-dependent. Specifically, preferential voting methods are seen as more legitimate than majority voting in a political decision-making situation, but only for individuals with well-defined preferences. Furthermore, preferential voting methods are more legitimate than majority voting in a highly polarized situation.

[1] Computational Social Science, ETH Zurich, Zürich, Switzerland. [2] Department of Communication and Media Research, University of Fribourg, Fribourg, Switzerland. [3] Complexity Science Hub Vienna, Vienna, Austria. [4] School of Computing, University of Leeds, Leeds, England. ✉email: carinah@ethz.ch

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 quickly and widely than ever before. Combating such problems requires scientific and technological advances and collective action. How can one promote behavioral change needed for this? We argue that one important factor is the perception of collective decisions as legitimate. Therefore, this paper investigates how legitimacy can be enhanced through democratic mechanisms, particularly the voting method applied (Helbing et al., 2023).

In fact, seeking ways to make decision-making methods *legitimate* is essential (Persson et al., 2013; Wellings et al., 2023): Legitimacy can be understood as cornerstone of both social choice theory and democratic initiatives. Voting enables people to express their interests while treating everyone equally (Persson et al., 2013). A detailed elicitation of citizen preferences offers room for participation and, therefore, builds a solid basis for the legitimacy of the results (McBride, 2003).

Voting mechanisms differ in their potential to elicit detailed preferences and the incentives to state the latter truthfully (see, e.g., Nitzan, 1985). 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 (Emerson, 2021). In contrast, multi-option preferential voting harnesses collective intelligence, thereby allowing for more informed and nuanced choices. In any case, one important goal of voting is to avoid a "tyranny of the majority" (Emerson, 2020).

Although the relationship between legitimacy and voting methods is crucial for the functioning of democratic societies, there is still a significant gap in our knowledge about the implications of alternative voting methods due to the relatively small number of experimental studies in this area. For instance, Bol et al. (2023) investigates two voting rules, aiming to distinguish between value-driven and self-interest-driven choices. 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 in real-world contexts.

Compared to representative decision making (the status quo), participatory processes appear to be associated with higher perceptions of fairness, even when the outcomes are unfavorable (Wellings et al., 2023; Werner and Marien, 2022). 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 ensure that those who win are satisfied while facilitating better outcomes for those who do not win as well.

To study this, we have developed an open-source smartphone application called "Votelab" (Kunz et al., 2023). This application compares four different voting methods: Majority vote, combined approval voting, range voting, and the modified Borda count. Note that the method to *aggregate* preferences, i.e., the rule to determine the resulting voting outcome, may also be relevant for the dimension of legitimacy, but is not investigated in our experiment. Furthermore, the app is designed to be user-friendly, while maintaining high privacy protection standards. In a controlled online experiment, participants vote with four different voting methods and provide legitimacy ratings of the respective method.

To elicit legitimacy ratings, we build on a psychological perspective in line with Tyler (Tyler, 2006). He defines legitimacy as "the belief that authorities, institutions, and social arrangements are appropriate, proper, and just" (Tyler, 2006, 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 (Scharpf, 1999). 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) (Scharpf, 1999). In our study, the outcome, for instance, chosen *COVID-19*-related measures, is the result of a vote. 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 (Schmidt, 2013). Within our research, the fairness of the voting method corresponds to throughput legitimacy.

Questions on decision-related acceptance, fairness, trust, and representation can all load on one and the same factor of legitimacy (Weil and Hänggli, 2021). These dimensions are not independent, seem to belong together (i.e., 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: Specifically, 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) (Maffettone and Ulaş, 2019). It is further found that legitimacy judgments are influenced by factors such as the institutional framework's stability, the legitimacy judgment process stage, and the methods used to form such judgments (Bitektine and Haack, 2015; Tost, 2011). In our study, we introduce polarized voting questions related to *COVID-19* to better distinguish the effect of input voting methods on legitimacy (Fig. 1). We compare this critical, highly relevant context with a more neutral context, in which participants are asked to choose their favorite color.

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.

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. The following list details the four input methods we implemented:

1. *Majority voting (mv)*: This requires the selection of one out of two or more options. That is, $s_{mv} \in \{0, 1\}$, where 1 stands for a chosen option.
2. *Combined approval voting (cav)*: This requires disapproval (-1), an indication of neutrality (0), or approval $(+1)$ of the voting options, i.e., $s_{cav} \in \{-1, 0, 1\}$.
3. *Range voting (rv)*: This requires assigning a numerical rating to each alternative option to reflect the degree of the preference. We assume $s_{sv} \in \{0, 1, 2, 3, 4\}$.
4. *Modified Borda count (mbc)*: This gives no points to unranked options, 1 point to the least preferred of the ranked options, etc. The choices are: $s_{mbc} \in \{0, 1, 2, 3, 4\}$. If a

Fig. 1 Screenshots of the mobile application used in the human subject experiment, displaying four questions on which participants voted via four different input methods. In this paper, we refer to these four questions with the following short labels (from left to right): *vaccine, icu, protection,* and *lockdown.*

Table 1 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

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 online experiment, 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 addressed the additional cognitive effort required of voters to express their preferences. Accordingly, the next method augments the previous ones:

Majority voting ("choose one option") offers just one approval option and no option to reject. *Combined approval voting* allows one to assigning one of three different ratings ("approve, stay neutral, disapprove"). *Range voting* ("assign points to options") adds another two levels, as it has overall five levels to choose from. It allows for assigning the same rating multiple times. Therefore, a voter can express indecisiveness. The *modified Borda count* ("choose and rank options") does not offer this option, but it is cognitively even more challenging, as it requires an explicit ranking of all options.

In our choice of these four voting methods, we were driven by several key considerations. Majority voting, being the most historically recognized and widely used method, serves as a

benchmark for comparison. Our focus on varied input mechanisms is evident: while majority voting provides a binary choice, combined approval voting adds depth with a scale capturing preference, disapproval, or neutrality. Range voting and the modified Borda count offer granularity in representing voter preferences. Beyond their individual merits, integrating these methods with the legitimacy scale presents a novel research contribution, highlighting the impact of the input method on perceptions of legitimacy and preference expression.

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 $LR \in \{0, ..., 4\}$ across two contexts $c \in \{color, COVID-19\}$.

Behavioral experiment. To test our hypotheses, we conducted a preregistered human subject experiment in a controlled online environment. 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 performed online via Qualtrics in collaboration with the ETH Decision Science Laboratory (DeSciL) in Zurich, Switzerland. It was conducted in three sessions in July 2021. The pre-registration link is https://doi.org/10.1257/rct.7871-1.0. Our experiment had three stages (see Table 1 for an overview): During Stage I, participants were introduced to both the *COVID-19* questions and the input methods 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 all voting methods. Specifically, each of the four *COVID-19* questions ($q \in \{vaccine, icu, protection, lockdown\}$) presented participants with five options ($o \in \{o_1, o_2, o_3, o_4, o_5\}$) from which they were asked to assign a rating (s). Finally, Stage III was dedicated to answering a set of control questions.

Figure 1 shows screenshots of the four *COVID-19*-related questions, their options, and how the user interface for the four input methods looked like.

Overall, 120 subjects (share of females: 0.36) 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: 0.37).

The sample size of $n = 120$ is based on the computed point of stability for the legitimacy ratings across various voting methods (Fig. S.7).

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 assigning a different set of ratings *s* to a voter's choice. The preference profile is defined as $p_i^{q,v} = \{s_{o_1}, s_{o_2}, s_{o_3}, s_{o_4}, s_{o_5}\}$, where $s_{o_v} = \{s_{mv_{o_j}}, s_{ca_{o_j}}, s_{sv_{o_j}}, s_{mbc_{o_j}}\}$.

Variation in preferences. Input methods vary in their ability to capture the properties of a distribution of preferences, $p_i^{q,v}$. We aim 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* (Navarrete et al., 2022). Both metrics, by quantifying score differences, allow for consistent comparisons across voting methods. Supplementary Fig. S.8 displays a full distribution of ratings. 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(Schmitt, 2016). Therefore, we calculate $\sigma(pi^{q,v})$. To compare $\sigma(pi^{q,v})$ across questions, we calculate the median $\eta(\sigma(pi^q))$. 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* (Navarrete et al., 2022). Divisiveness *D* is defined for all option pairs by the mean difference in ratings *s* between voters who prefer one option o_m over another o_n and those who prefer o_n over o_m. Furthermore, divisiveness D_v^q is calculated for each voting method *v* and each question *q*. In other words, *D* provides an intuition on how divisive a question is expressed through a particular input method.

For each pair of options o_m and o_n, let:

$s(o_m, o_n)$ be the rating when o_m is preferred over o_n

$s(o_n, o_m)$ be the rating when o_n is preferred over o_m

Then, divisiveness for each voting method *vm* for each question *q* is defined as:

$$D_v^q = \frac{1}{n(n-1)} \sum_{m=1}^{n} \sum_{\substack{n=1 \\ n \neq m}}^{n} \| s(o_m, o_n) - s(o_n, o_m) \| . \quad (1)$$

To compare D_v^q across questions, we calculate $\eta(D^q)$. More details on *D* are presented in S.11.

Max-choice profile. An interesting question to investigate is how a voter rates option *o* across four voting methods (*vm*) for a specific question *q*.

To that end, Fig. 4 visualizes preferences p_i^q by a *five-dimensional series* (five options *o*) with *four* time steps (four voting methods *vm*) each.

The theoretical number of possible p_i^q is extremely large (S.10). To compare p_i^q across questions *q*, we follow the rationale that the option with the highest rating is of special importance. The originally assigned rating is mapped to a binary scale, where 1 represents the highest rating of the options and 0 otherwise. *vm* refers to any voting method. We define $s_v = \left\{ s_{v_{o_1}}, s_{v_{o_2}}, s_{v_{o_3}}, s_{v_{o_4}}, s_{v_{o_5}} \right\}$

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

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

Figure 2 provides an example of how a voter's choices p_i^q are mapped to a multi-dimensional time series.

Participants stated preferences in the following order: majority vote, combined approval, range voting, and the modified Borda count. The first and last input methods required ranking. By contrast, the second and third methods allowed for expressing multi-peaked preferences.

Whether ranking was required or not has consequences for interpreting the max-choice profile. In the following, we provide 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. We further assume that the voter's preferences are multi-peaked; the 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 among all four voting methods. Consequently, any max-choice profile with patterns 1*xx*0 and 0*xx*1 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* × 4 *questions* × 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* × 4 *questions* × 120 *voters* = 1920.

Clustering expressed preferences. The voter is asked *after* voting to rate the legitimacy of the input method. Note, however, that 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 options to vote on could influence the legitimacy rating.

Furthermore, when rating legitimacy, the voter might think about the last question answered. As the last question could be over-represented in the participant's memory, the question order was randomized.

In this case, there is variation between participants we need to control for.

Therefore, we cluster preferences and compare legitimacy ratings across these clusters. If ratings significantly differ by cluster, it suggests that legitimacy ratings are influenced by satisfaction.
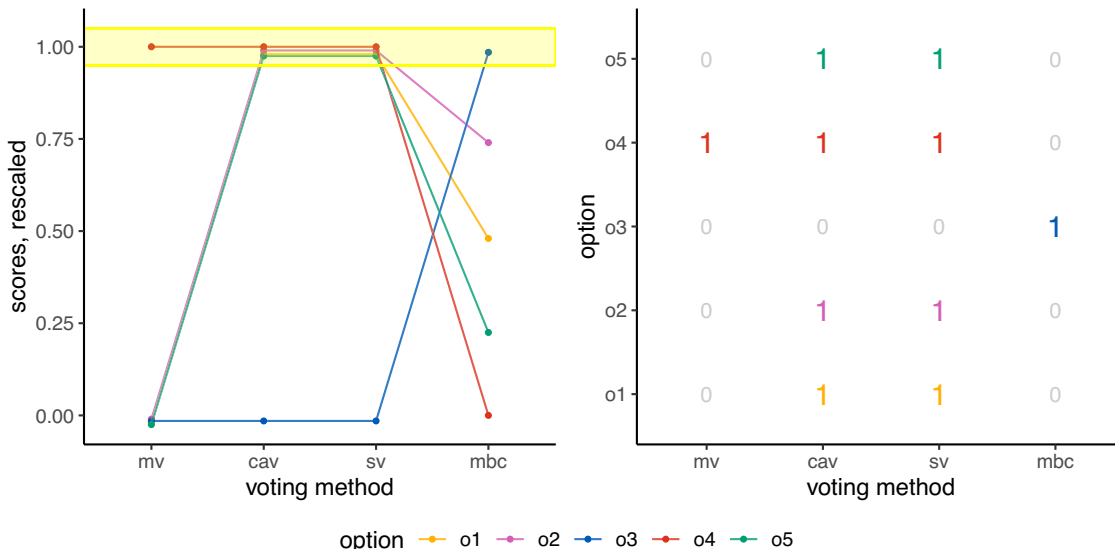
Fig. 2 The highest valued choice—highlighted in yellow in the left panel—is of special importance. Therefore, we reduce the dimensionality of a participant's full choice profile (left panel) to its first derivation (right panel). Plotted are sample values of one participant and one question. For each voting method, the options with the highest rating are encoded by 1. The remaining ones are encoded with 0.

The basis for clustering is a voter's preference profile pp. However, the number of possible combinations is extremely large and far exceeds the number of observations. Therefore, we reduce dimensionality. The max-choice profile is a highly reduced representation of pp. A less reduced form is obtained by averaging the (rescaled) ratings per option over the four input methods as follows:

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

Subsequently, we cluster $p_i = \{\mu_{o1}, \mu_{o2}, \mu_{o3}, \mu_{o4}\}$.

To determine the number of clusters, we calculate nine cluster evaluation indices, suggesting three clusters are proposed (Supplementary Table S.15). To increase the robustness of the clustering results, we deploy nine clustering methods from various categories (Xu and Tian, 2015) (Supplementary Table S.16). We investigated the relationship between legitimacy ratings and *COVID-19* preference clusters. The Kruskal-Wallis test p-values suggest no significant difference between the groups (Table 3). 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.

Results
Our study finds that the frequently used method of majority vote is perceived as less legitimate than range voting. Especially in highly polarized contexts, voters value the ability to express their preferences in more detail.

Winning options vary by input method. First, we find that voting with different input methods often leads to different outcomes, i.e., for some questions, the winning option changes (Table 2). This is also true for the *color* context (Supplementary Table S.4). Applying another aggregation rule—the Condorcet method—gives similar results (Supplementary Table S.5). Why does the input method induce outcome variation for some questions, but not for others?

To answer this question, we investigate the sequence of ratings assigned to each option. We refer to this sequence as a preference

Table 2 The results of four COVID-19 related questions (columns), which were voted upon using four different voting methods (rows).

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
rv	o_4	o_1	o_4	o_4
mbc	o_4	o_1	o_4	o_1

The winning option $\{o_1, ..., o_5\}$ out of five possible options was determined by the highest sum of ratings. Options and questions are detailed in Fig. 1.
Majority voting ($mv \in \{0, 1\}$), combined approval voting ($cav \in \{-1, 0, 1\}$), range voting ($rv \in \{0, 1, 2, 3, 4\}$), and modified Borda count ($mbc \in \{0, 1, 2, 3, 4\}$).

profile for each individual, question, and input method (details in "Preference profile"). We compare variations in preference profiles by calculating the standard deviation σ and divisiveness D.

As is evident in Fig. 3, *protection* shows the lowest standard deviation $\eta_\sigma^{protection} = 0.395$ and the second lowest divisiveness $\eta_D^{protection} = 0.45$. When testing pairs of ratings for *protection* against other questions, we find that *protection* shows lower standard deviations σ (Wilcoxon signed rank test for paired samples, Supplementary Table S.8) and divisiveness D (Supplementary Table S.9) than most other questions. These results explain why question *protection* shows the lowest outcome dependency on the input method. We conclude that the higher is the standard deviation and divisiveness, the more dependent is the aggregate result on the input method.

Our use of standard deviation and divisiveness in examining preference profiles aligns with the principles of spatial logic (Downs (1957); Enelow and Hinich (1989)). Spacial logic visualizes preferences in a hypothetical space, where each option and individual occupies a unique position. The closer an option is to an individual's position, the higher the rating or preference
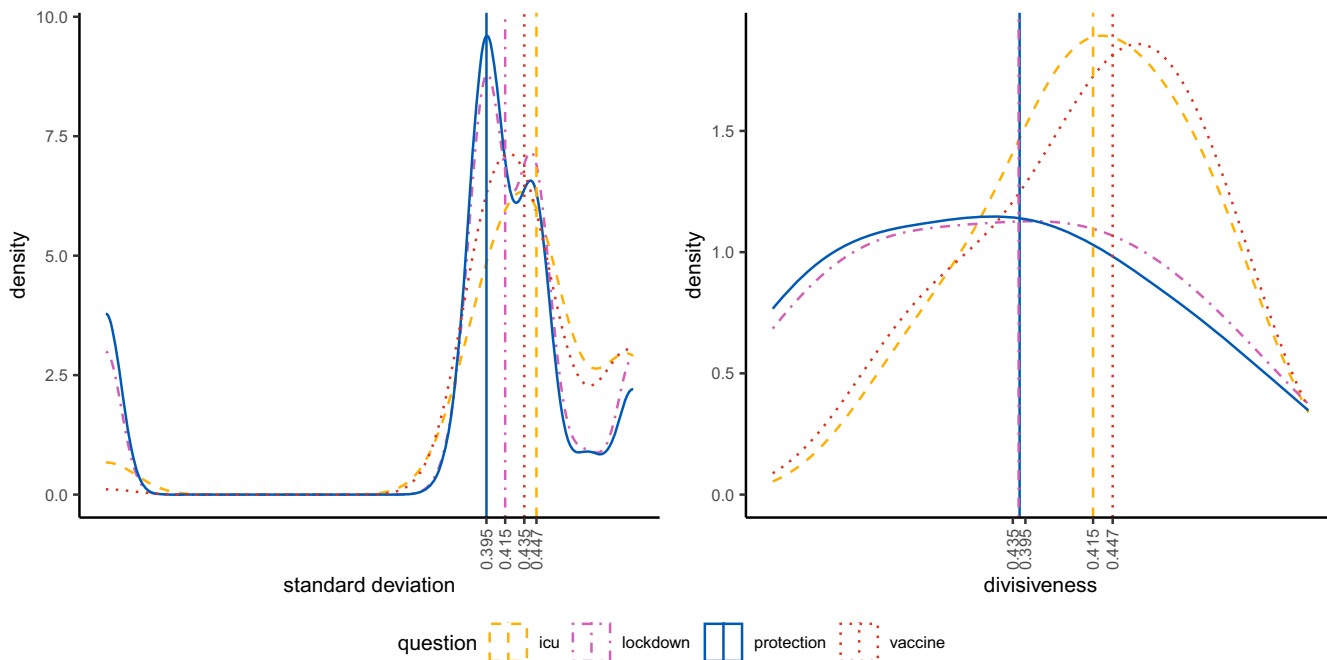
Fig. 3 Protection shows lower standard deviation and divisiveness than most other questions. Left panel: Densities of standard deviations σ of preference profiles by question. Right panel: Densities of divisiveness D. For both ratings, we compare the medians η, represented as dashed vertical lines, of question-pairs via non-parametric tests.

assigned to that option. When analyzing preference profiles, a high standard deviation suggests that an individual perceives distinct differences in the "distance" between their position and that of the various options. Thus, when many individuals display high standard deviations and their rankings are polarized—with certain options consistently being ranked high and others low—this indicates a pronounced spatial polarization in preferences. Such spatial polarization serves as a crucial metric, underscoring the significance of the voting method in contexts where preferences diverge sharply. In other words, the choice of 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 voter preferences differ by the question? To this end, we introduce "max-choice profiles". Following the rationale that the highest rated option is of special importance, we construct max-choice profiles by coding the highest rated option as 1 and coding the remaining options as 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.

Figure 4 shows the five most common max choice profiles: 1110, 0110, 1111, 0100, and 0000. The gray vertical lines represent the actual count of each choice profile. The markers show the count per question. The orange dashed segments 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 particularly interesting. If all votes had been cast consistently, in Fig. 4 the gray vertical line would reach the orange segment. However, this is not the case, indicating that a significant portion of voters (32.8%) did not vote consistently.

We test for significant differences in counts across questions for four choice profiles (Supplementary Table S.10). The number of counts per question exhibits a significant disparity between the max-choice profiles 0110, 0100, and 0000. However, this
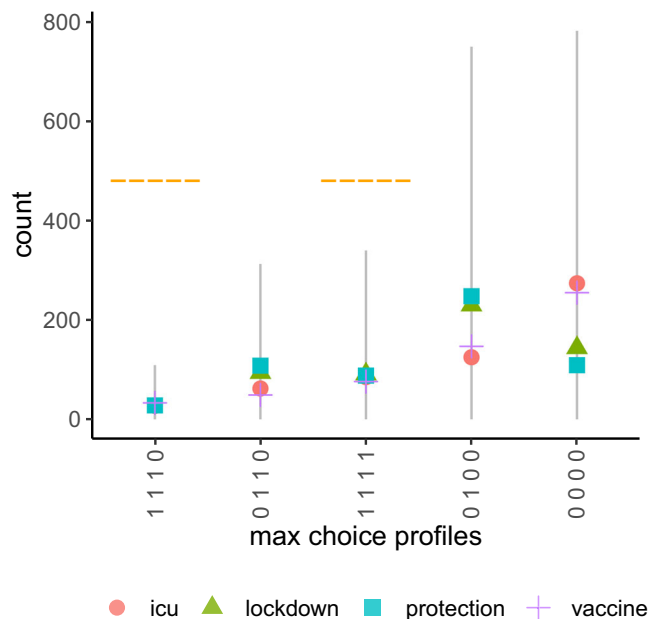

Fig. 4 Only the entirely consistent choice profile 1111 shows no significant difference in counts per question. The x-axis lists max-choice profiles, where the highest rated option per question is coded as 1, while the remaining options are coded as 0. The gray vertical lines show absolute frequencies per max-choice profile. Only profiles with $n > 25$ are displayed. The colored markers show mean counts by question. The orange horizontal dashed segments depict the maximum frequency for that specific choice profile.

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 trait. As we will describe later, this personal attribute correlates with how an individual perceives the legitimacy of a voting method.
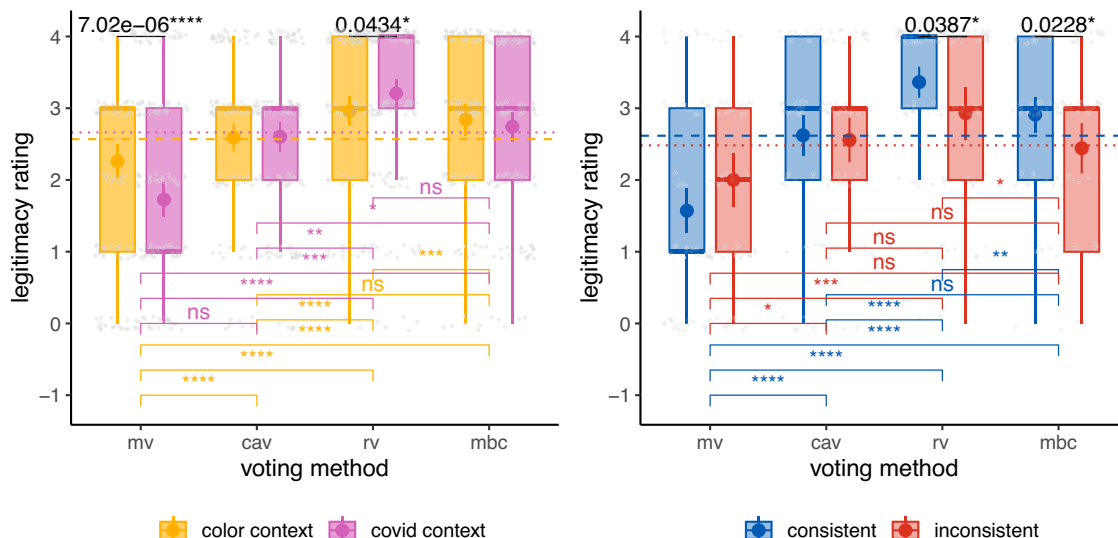
Fig. 5 The perceived legitimacy of a voting method varies by context. Legitimacy ratings by context (left) and inconsistency (right). *p*-values are calculated using the Wilcoxon signed rank test with Holms adjustment. Black *p*-values and brackets denote comparisons between dimensions (*color* vs. *COVID-19*, consistent vs. inconsistent). *p*-values in *color* indicate comparisons within a dimension but across voting methods. The gray scatter shows individual ratings, with mean values represented by dots inside the box, median values indicated by horizontal lines. The dashed horizontal lines represent the mean of all voting methods by consistency or context.

Flexibility in voting methods is perceived as more legitimate in a political context. In this paragraph, the investigation focuses on how voting method 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 5 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. In the *color* context, we find that the medians ($\eta = 3$) for all voting methods are identical. However, in the *COVID-19* context, the median perception of legitimacy for majority voting ($\eta = 1$) is lower than that of the other three voting methods, which have equal median values ($\eta = 3$).

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 on the full array of legitimacy ratings, using the paired Wilcoxon signed rank test with Holms adjustment. In the *color* context (in orange), all but one comparison (ca–mbc) are significantly different from each other. Consequently, majority voting is considered the least legitimate, followed by combined approval and modified Borda count, both of which are considered less legitimate than range 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 are perceived as the least legitimate, followed by the modified Borda count, which is considered less legitimate than range voting.

In summary, the results show that range voting is perceived as more legitimate than majority voting in both the *color* and *COVID-19* contexts. However, in the *COVID-19* context, range voting, and the modified Borda count are 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. In comparison, in the *COVID-19* context, issues such as strict lockdown measures may 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 Fig. 5): The majority vote is rated as more legitimate when voting on colors compared to *COVID-19*-related issues (paired Wilcoxon signed rank test, $p = 7.02 \times 10^{-6}$, Supplementary Table S.12. The opposite is true for range voting ($p = 0.043$). These results suggest that the legitimacy of a voting method is context-dependent to some degree. In other words, our results suggest that legitimacy is context-dependent.

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 Fig. 5. Here, we compare those who vote consistently to those who do not vote consistently. Please recall that the gray vertical line in the choice profile 1111 of Fig. 4 represents the fully consistent voters. These voters are compared to the inconsistent voters, who are responsible for the discrepancy between the gray and orange lines. Now, let us return our focus to Fig. 5: Our analysis of consistent voters (shown in blue) reveals that the median legitimacy rating for majority voting ($\eta = 1$) is lower than that of combined approval and the modified Borda count ($\eta = 3$), which in turn is lower than the median legitimacy rating for range voting ($\eta = 4$). In contrast, for voters with inconsistent preferences, the legitimacy rating for majority voting ($\eta = 2$) is significantly lower than for the other three voting methods, which have the same median legitimacy rating ($\eta = 3$, Supplementary Table S.13). Pairwise comparisons of legitimacy ratings between input methods across consistent and inconsistent voters are indicated by blue and red *p*-values and brackets (paired Wilcoxon signed rank test with Holms adjustment). For consistent voters (in blue), all comparisons but one (ca–mbc) are significantly different from each other. For inconsistent voters (in red), two additional comparisons turn out to be non-significant (mv–mbc, cav–sv, cav–mbc).

The results reveal an interesting phenomenon. Despite the difference in the number of ratings allowed, inconsistent voters consider the exclusive three-rating scale (cav) as equally legitimate compared to the non-exclusive five-rating scale (rv). Furthermore, it is noteworthy that, on a lower level, the restrictive
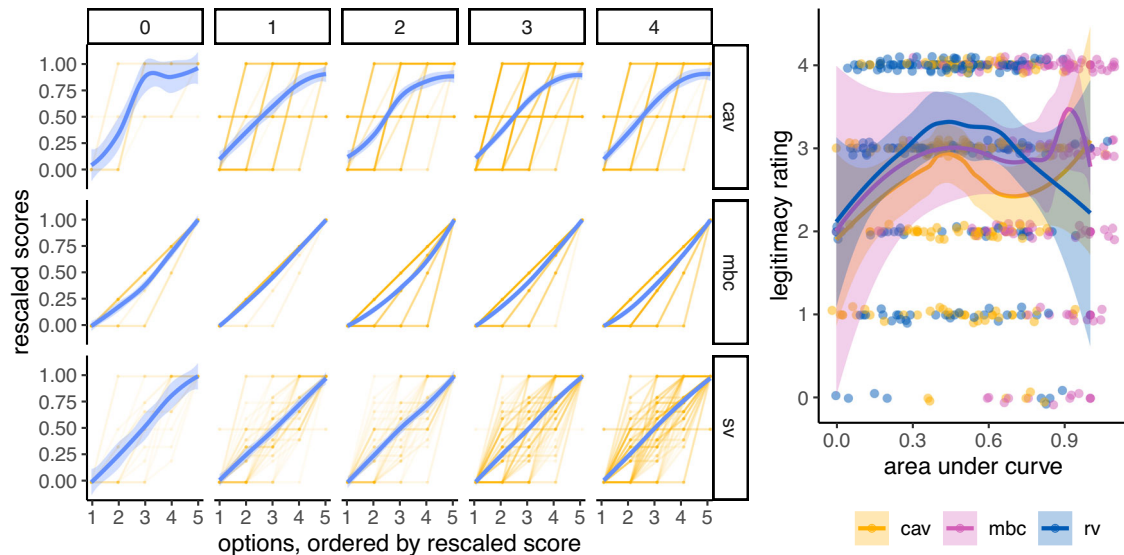
Fig. 6 The area under the preference curve (AUC) is greatest when multiple options receive high ratings. Left panel: Illustration of the preference profiles of all participants, with the rows representing three distinct voting methods and columns referring to the legitimacy ratings. The options are sorted in such a way that the highest-rated option is on the right, and the lowest-rated option is on the left. The blue curve is estimated via Local Polynomial Regression Fitting; its shaded area represents the 95 percent confidence interval. Right panel: Illustration of the average area under the preference profile curves (AUC) across questions for each participant, 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.

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 (black p-values). A Wilcoxon rank sum test reveals a statistically significant difference in the legitimacy ratings assigned by fully consistent and inconsistent voters. The results indicate that consistent voters tend to assign higher ratings to range voting and the modified Borda count, with p-values of 0.0387 and 0.0228, respectively (Supplementary Table S.14). 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 range 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 appreciating flexibility in range voting. Figure 6 illustrates the relationship between options, preference profiles, and perceived legitimacy. In the left panel, the rows in the figure represent the voting method, while the columns show the legitimacy rating. The orange lines illustrate individual preference profiles, with options organized in ascending order based on their ratings. To highlight underlying trends, a blue curve, generated using Local Polynomial Regression Fitting, has been superimposed. The area under the preference profiles, denoted as (AUC), is greater if more options receive high ratings. For three voting methods, multiple options can attain similar rankings. Consequently, no method inherently prescribes a specific AUC. This is not true for majority voting, which is the reason for not plotting this method. The right panel of Fig. 6

depicts the relationship between AUC and legitimacy ratings across different voting methods. Each point signifies an individual's AUC for a particular voting method, along with the associated legitimacy rating provided by that participant. Overlaying these points, the curves generated via Local Polynomial Regression Fitting serve to aid interpretation.

One curve is particularly noteworthy: For range voting, the relationship between AUC and legitimacy rating exhibits a distinct inverse U-shape. The most left part of the "U" indicates low values of AUC, which in turn indicate that *few* options receive high ratings. In this case, range voting receives low legitimacy ratings. However, as AUC increases and more options receive high ratings, the legitimacy ratings also increase. This shows that voters, who appreciate the flexibility offered by range voting because they assign high ratings to *many* options, are more likely to rate the method as legitimate. However, when *all* options receive very high ratings, resulting in a high AUC, the legitimacy ratings decrease. This suggests that, when voters are unable to distinguish between different options, range voting is perceived as less legitimate. Hence, the perceived legitimacy of a voting method depends on the complexity of a voter's preference profile. If voter preferences are complex and nuanced, they require a more complex voting method, and vice versa.

The independence of the perceived legitimacy from COVID-19-related topics confirms the validity of our legitimacy framework. Our final analysis was motivated by the question of whether a participant's set of opinions concerning COVID-19 would systematically influence legitimacy ratings. In such a scenario, we would need to control for the specific question. If the question was randomized, the most recent question answered before rating legitimacy could be more prominent in the participant's memory, potentially skewing their ratings. This analysis is essential to ascertain the validity of our method for studying legitimacy, as it aims to confirm whether participants are indeed rating the intended concept.

To test this hypothesis, we clustered voter preference profiles. We interpreted the centroid based on nine clustering methods (details in "Clustering expressed preferences") and identified

Table 3 The *p*-values of a Kruskal-Wallis Test indicate no differences in legitimacy ratings by voting method and question across the three preference clusters.

	vaccine	icu	protection	lockdown
mv	0.163	0.721	0.548	0.583
cav	0.458	0.501	0.139	0.361
rv	0.058	0.253	0.198	0.839
mbc	0.531	0.36	0.22	0.951

three preference clusters per question (Fig. S.9). Subsequently, we examined whether there was a difference in legitimacy ratings across the three clusters for each question and voting method. Out of the 16 tests, none had a statistically significant *p*-value (Kruskal-Wallis test, Table 3). This suggests that the preference profiles on COVID-19-related topics do not significantly influence how participants rate the legitimacy of voting methods. Therefore, our questions about the legitimacy of voting methods should be reliable and accurate indicators of participants' views.

Conclusion and discussion

The key findings of our experimental study can be summarized as follows:

1. Different voting methods can lead to different outcomes, even when the same group of individuals votes on the same set of questions.
2. The choice of voting method is particularly important in contexts where preferences are highly polarized.
3. The perceived legitimacy of a voting method is not a universal property, but context-dependent. Specifically, the legitimacy gain of preferential voting methods over majority voting is bigger when more complex questions are being asked.
4. However, the latter statement is only true for individuals with clear preferences. Those with uncertain preferences tend to conflate their 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 study underscores the significance of selecting an appropriate voting method, particularly in polarized contexts (Alós-Ferrer and Buckenmaier, 2021) such as those we examined relating to COVID-19. The choice of voting method can significantly influence the outcomes in such situations.

Additionally, our research elucidates how color preference embodies a position issue, distinct in nature from valence issues, as observed in the context of COVID-19 health questions. While the critical importance of health transcends partisan divides, the paths to achieving health objectives highlighted by the varied responses to COVID-19-underscore the complexity of what might initially appear as valence issues. This delineation enriches our understanding of issue salience and its impact on the perceived legitimacy of different voting methods. Moreover, our findings reveal that the dichotomy between valence and position issues alone does not fully capture the variations in legitimacy ratings. Specifically, within the context of COVID-19, range voting was consistently deemed more legitimate, pointing to the significance of both the issue's nature and the voting method's characteristics in shaping legitimacy perceptions.

Our findings suggest that, while voters with consistent preferences favor detailed voting methods, those with less defined preferences may view these methods as unfavorably complex and less legitimate.

What are the *policy implications* of our paper? We found that voters value flexibility in the voting method, and this preference intensifies in decisions with societal relevance. Before implementing range voting as the baseline choice, however, policymakers should consider an additional dimension: the correlation between voters' clarity of opinion and their preferred voting method. Those with clear opinions view more flexible voting systems as more legitimate. Based on these findings, what is our practical recommendation for policymakers? To enhance perceived legitimacy, consider adopting the study's phased approach: select and communicate the deciding voting method beforehand, then progressively engage voters with the issue through multiple rounds, from simple majority voting to more complex methods like range voting. This graduated approach could help voters, especially the undecided ones, to crystallize their preferences without feeling overwhelmed by complex voting systems.

Our research highlights important areas for future investigation. One key aspect not explicitly addressed in this study is the relationship between outcome favorability and perceived legitimacy. Future studies could investigate how the favorability of an election result, as determined by different voting methods, affects its perceived legitimacy. This is particularly pertinent in polarized or crisis situations, where the acceptance of election outcomes may be particularly crucial. Additionally, future research could test the legitimacy perception of an additional way to vote, namely, offering voters the choice of their voting method. Accordingly, different sets of people would vote according to different voting methods. To determine the winning option, each voting procedure's outcome could be weighted by the respective number of voters. In this way, democratic elections would be open to an evolutionary process, and voters would be driving it. Furthermore, it is likely that people with well-defined preferences —let us call them "decided people"—prefer to express their preferences in a more nuanced way than "undecided people," who would need less complex voting methods. While certain cognitive factors might influence both well-defined preferences and the grasp of complex voting methods, cognitive factors, for example, may not necessarily determine how decided people are. These kinds of questions deserve to be addressed by future research. Additionally, expanding the scope to various political contexts such as participatory budgeting, constitutional agendas, or policy-making in the context of protests, presents other valuable research directions. Investigating how different voting systems interact with diverse political issues would provide a more comprehensive understanding of the legitimacy of these methods.

References

Alós-Ferrer C, Buckenmaier J (2021) Voting for compromises: alternative voting methods in polarized societies. Department of Economics, Working paper, University of Zurich

Bitektine A, Haack P (2015) The "Macro" and the "Micro" of legitimacy: toward a multilevel theory of the legitimacy process. Acad Manag Rev 40(1):49–75

Bol D, Blais A, Coulombe M, Laslier J-F, Pilet J-B (2023) Choosing an electoral rule: values and self-interest in the lab. J Econom Psychol 95:102602

Downs A (1957) An economic theory of democracy. Vol 28, Harper and Row

Emerson P (2020) Majority voting as a catalyst of populism. Springer

Emerson P (2021) The consequences of consequences. In: Emerson, P. (ed.) Democratic decision-making. Springer, p 41–51

Enelow JM, Hinich MJ (1989) A general probabilistic spatial theory of elections. Public Choice 61(2):101–113

Helbing D, Mahajan S, Hänggli Fricker R, Musso A, Hausladen CI, Carissimo C et al. (2023) Democracy by design: perspectives for digitally assisted, participatory upgrades of society. J Comput Sci 71:102061

Kunz R, Banaie F, Sharma A, Hausladen CI, Helbing D, Pournaras, E (2023) Votelab: a modular and adaptive experimentation platform for online collective decision making. arXiv preprint arXiv:2307.10903

Maffettone P, Ulaş L (2019) Legitimacy, metacoordination and context-dependence. Int Theory 11(1):81–109

McBride C (2003) Consensus, legitimacy, and the exercise of judgement in political deliberation. Crit Rev Int Soc Polit Philos 6(3):104–128

Navarrete C, Ferrada N, Macedo M, Colley R, Zhang J, Grandi U et al. (2022) Understanding political agreements and disagreements: evidence from the 2022 French presidential election

Nitzan S (1985) The vulnerability of point-voting schemes to preference variation and strategic manipulation. Public Choice 47(2):349–370

Persson M, Esaiasson P, Gilljam M (2013) The effects of direct voting and deliberation on legitimacy beliefs: an experimental study of small group decision-making. Eur Polit Sci Rev 5:381–399

Scharpf FW (1999) Regieren in Europa: Effektiv und demokratisch? Campus Verlag

Schmidt VA (2013) Democracy and legitimacy in the European Union revisited: input, output and 'throughput'. Polit Stud 61(1):2–22

Schmitt J (2016) How to measure ideological polarization in party systems. ECPR graduate student conference 2016 https://ecpr.eu/Events/Event/PaperDetails/28307

Tost LP (2011) An integrative model of legitimacy judgments. Acad Manag Rev 36(4):686–710

Tyler TR (2006) Psychological perspectives on legitimacy and legitimation. Ann Rev Psychol 57(1):375–400

Weil L, Hänggli Fricker R (2023) "How innovation in participation could increase legitimacy." Paper presented at the Swiss Political Science Conference. February 2021. Working Papers SES 532

Wellings TS, Majumdar S, Hänggli Fricker R, Pournaras E (2023) Improving city life via legitimate and participatory policy-making: a data-driven approach in Switzerland. In: Proceedings of the 24th annual international conference on digital government research, pp 23–35

Werner H, Marien S (2022) Process vs. outcome? how to evaluate the effects of participatory processes on legitimacy perceptions. Br J Polit Sci 52(1):429–436

Xu D, Tian Y(2015) A comprehensive survey of clustering algorithms Ann Data Sci 2:165–193

Acknowledgements

Author contributions

CIH: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing—Original Draft, Writing—Review & Editing, Visualization. RHF: Review & Editing, Funding acquisition. DH: Encouraged, supported, and enabled the study on different voting methods, contributed to the review and editing of the manuscript, and acquired funding with some of the co-authors. RK: Software. JW: Cluster Analysis. EP: Conceptualization, Methodology, Software, Investigation, Writing—Review & Editing, Supervision, Funding acquisition.

Competing interests

The authors declare no competing interests.

Ethical Approval

The questionnaire and methodology for this study was approved by the Human Research Ethics Committee of ETH Zurich (Ethics approval number: 2021-N-28).

Informed consent

Informed consent was obtained from all participants before the study began.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-03056-8.

Correspondence and requests for materials should be addressed to Carina I. Hausladen.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.