

Baujard, Antoinette; Brunetti, Roberto; Lebon, Isabelle; Marsilio, Simone

Article — Published Version

## How people understand voting rules

Public Choice

*Suggested Citation:* Baujard, Antoinette; Brunetti, Roberto; Lebon, Isabelle; Marsilio, Simone (2025) : How people understand voting rules, Public Choice, ISSN 1573-7101, Springer US, New York, Vol. 207, Iss. 3-4, pp. 365-395, <https://doi.org/10.1007/s11127-025-01344-8>

This Version is available at:

<https://hdl.handle.net/10419/342696>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# How people understand voting rules

Antoinette Baujard<sup>1</sup> · Roberto Brunetti<sup>2</sup> · Isabelle Lebon<sup>3</sup> · Simone Marsilio<sup>4</sup>

Received: 12 July 2024 / Accepted: 27 September 2025 / Published online: 27 November 2025  
© The Author(s) 2025, modified publication 2025

## Abstract

If individuals are to be empowered in their selection or use of a voting rule, it is necessary that they understand it. This paper analyzes people's understanding of two voting rules: evaluative voting and majority judgment. We first distinguish three components of understanding in this context: how to fill in the ballot; how votes are aggregated; and how to vote strategically. To measure each component, we draw on results from a lab experiment on incentivized voting where participants are exogenously assigned single-peaked preferences and answer comprehension questions on the rules employed. We find that most participants understand how to fill in the ballot with both voting rules. However, participants' understanding of vote aggregation under majority judgment is lower and, crucially, more heterogeneous. While some participants correctly understand its aggregation property, a sizable group fails to grasp it. We also observe no difference in voting behavior between evaluative voting and majority judgment: the data confirm the theoretical prediction that under evaluative voting there will be a high incidence of strategic voting through the use of extreme grades, but contradict the prediction that under majority judgment voters will vote less strategically. Finally, we find that with majority judgment, the better voters understand how votes are aggregated, the more they vote strategically, hence resulting in inequality in voter agency.

**Keywords** Voting rules · Understanding · Evaluative voting · Majority judgment · Agency · Laboratory experiment

---

We are grateful to the participants at the Social Choice and Welfare meeting (Paris), the GATE on the Docks workshop (Lyon), and the Amsterdam/Saint-Etienne Workshop on Social Choice (Saint-Etienne) for their insightful comments and discussions, as well as to Jean-François Laslier for his comments on the first draft of this paper. We would also like to thank Sylvain Bouveret, G  rald Chatagnon, Herrade Igersheim, Rida Laraki, R  my Poulain, Elven Priour, and Quentin Thevenet for their help and support at various stages of the design conception. This research was jointly funded by the Research Grant "Voter Autrement 2022" from the University Jean Monnet, by CREM Caen, and by the French National Research Agency within the Citizens project "ANR-22-CE26-0019-01". For the purpose of Open Access, a CC-BY public copyright licence (<https://creativecommons.org/licenses/by/4.0/>) has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission.

---

Extended author information available on the last page of the article

JEL codes: A13 · C92 · D71 · D72 · O35

## 1 Introduction

The social choice literature presents a wide range of alternative voting rules intended to counter the shortcomings of traditional ones. Yet while these rules may have desirable theoretical properties (Sen 1995; Brams and Fishburn 2002), their validity in practice depends on assumptions about voters' behavior. One crucial assumption is that voters properly understand the rules in question, which may not always be true, especially where they involve unusual balloting systems and vote aggregation methods (Laslier 2011). Unexpected issues can therefore arise: for instance, with rules where voters assign grades to every candidate, voting inconsistencies increase when the range of available grades widens (Baujard et al. 2018). Moreover, the complexity of a rule can be a barrier to its usability (Bourgeois-Gironde and Ferreira 2023)<sup>1</sup> as well as its capacity to meet other desirable properties such as voter satisfaction (Blais et al. 2021). In the end, then, understanding is a normative property *per se*, as it determines their agency, *i.e.*, the degree to which they are empowered to use a given rule to reach their objectives. Since voter confusion can only be observed in the real world, this issue calls for empirical study. To our knowledge, however, only a few papers have sought to tackle this issue, and in particular we do not know of any attempt to offer a systematic method to measure citizens' understanding of voting rules.

This paper aims to fill this gap by scrutinizing individuals' understanding of two alternative voting rules: evaluative voting<sup>2</sup> and majority judgment. These two rules offer an ideal testing ground for exploring voters' understanding for at least three reasons. First, both rules feature one-round elections with multi-nominal ballots—*i.e.*, voters assign grades to every candidate—yet with two differences: in evaluative voting, grades are numerical and aggregated by their sum, while in majority judgment, grades are verbal and aggregated by the median; we can therefore leverage the subtle yet relevant differences between these rules to study people's understanding of their specific characteristics. Second, evaluative voting and majority judgment are increasingly the subject of study among academics (since Hillinger 2006; Balinski and Laraki 2007), but there is limited empirical evidence on how individuals use these rules. Third, even though both rules are already used in practice,<sup>3</sup> many people lack experience with them. Therefore, we can study people's understanding absent any previous knowledge of the rule or any *status quo* bias.

To do this, we run an incentivized laboratory experiment based on Der Straeten et al. (2010), where people are assigned single-peaked preferences and vote for five candidates.

<sup>1</sup> In a more general context, Oprea (2020) considers rules as algorithms to be implemented by humans and finds that most subjects in a lab-experiment are willing to pay to avoid implementing complex rules.

<sup>2</sup> Evaluative voting is also known as utilitarian voting, range voting, score voting, or point voting. The expression *vote par note* (grade voting) is also used in French.

<sup>3</sup> For instance, an equivalent of three-level evaluative voting is used in Latvia's parliamentary elections, where voters can give a negative, null, or positive grade to each candidate in the chosen party list (Laslier et al. 2015). A variation of evaluative voting in the form of cumulative voting is used in Luxembourg (Laslier et al. 2015) and the German states of Bremen and Hamburg (Bowler et al. 2018), where voters can assign a fixed total of points to different candidates. Majority judgment was used in France for the local elections of the French political party "La République en Marche", and for the primary elections of the left-wing coalition (*primaire populaire*) in 2022.

In our case, participants vote using either evaluating voting or majority judgment depending on the treatment to which they are assigned. The laboratory conditions offer time and incentives for individuals to encounter the two voting rules, learn them, and practice them. We conform to the common practice in experimental economics of reinforcing the experimental instructions with examples and context-rich practice rounds (Freeman et al. 2018). This allows us to scrutinize participants' understanding of voting rules in an ideal learning condition and at different stages. Most importantly, we can study people's understanding of the aspects of specific voting rules by observing their voting behavior and incentivized responses to targeted questions. This approach overcomes the metacognitive issue of individuals' illusion of understanding (Rozenblit and Keil 2002; Sloman and Fernbach 2017), which can bias their *self-assessed* and *general* knowledge of voting rules.

We distinguish and measure three components of understanding: First, we investigate whether people understand how to fill in a ballot, which involves participants' knowledge of what grades mean and how to use them to provide support to the desired candidate(s). This component is measured by the number of inconsistent votes (Igersheim et al. 2016; Baujard et al. 2018; Darmann and Klamler 2023), *i.e.*, choices such that participants rate a more-preferred candidate lower than a less-preferred one. Second, we observe peoples' understanding of how votes are aggregated, which concerns the aggregation methods (sum/mean vs. median) used in each voting rule. This component is measured using a set of incentivized questions where participants compute the results of mock elections with evaluative voting and majority judgment, and where they also answer some true/false questions. Finally, we study whether people vote strategically, which is related to participants' knowledge of how to vote to maximize their payoffs. To do this, we study the number of extreme grades assigned to candidates by voters, which signals a departure from sincere voting, and look at the difference when voting with one or the other rule.

We present four main results: First, only a few voters vote inconsistently with both the multinomial rules under test, suggesting that people understand how to fill in a ballot and can easily express their preferences with grades. Second, participants' understanding of how votes are aggregated is, on average, lower with majority judgment than evaluative voting. Third, their understanding is crucially more heterogeneous. While some participants understand this feature well, others fail almost completely to grasp it. Participants' misunderstanding arises from the definition of the median and its implications on the vote outcome. Fourth, and in contrast to theoretical predictions, we show that participants vote similarly with evaluative voting and majority judgment: indeed, voters mostly use extreme grades, which is in line with the prediction that there would be a high incidence of strategic voting in the case of evaluative voting (Núñez and Laslier 2014) but at odds with the prediction of less strategic voting with majority judgment (Balinski and Laraki 2020). Finally, we provide evidence that those participants who grasp the aggregation property of the majority judgment rule assign extreme grades even more often than participants who do not grasp it. We interpret this finding as indicating that participants—particularly those who understand the properties of the median—perceive majority judgment as manipulable. We ultimately highlight the possibility that differences in understanding may lead to variations in voter empowerment.

**Related literature** Our paper relates to three strands of literature. First, our paper speaks to the social choice literature that studies alternative voting methods, such as evaluative voting and majority judgment. This literature tends to focus on the theoretical properties

of these voting rules (Hillinger 2006; Balinski and Laraki 2007; Núñez and Laslier 2014; Macé 2018) and tests their applicability and empirical implications (Baujard and Igersheim 2009; Baujard et al. 2014, 2018, 2021; Balinski and Laraki 2010, 2012; Darmann et al. 2017, 2019; Igersheim et al. 2022; Darmann and Klamler 2023; Delemazure et al. 2025), and only a handful of papers probe voters' understanding of the rules. Few papers measure self-reported understanding through questions such as "How well do you think you understand evaluative voting?", to which most respondents answer positively (see Baujard and Igersheim (2009) for evaluative and approval voting, Kimball and Kropf (2016) for cumulative voting, and Donovan et al. (2019, 2022), for ranked-choice voting). This approach has two major limitations. First, the definition of what is meant by "understanding" is left to respondents' interpretations; second, participants who claim to understand the rule might not realize that they do not understand it, or might not want to admit it.<sup>4</sup>

A second method relies on assessing voters' errors when casting a ballot. For instance, in the case of ranked-choice voting, this implies checking for the presence of undervotes and overvotes (see, e.g., Cormack 2023).<sup>5</sup> In the case of multi-nominal rules, single-name ballots can signal voters' misunderstanding, especially since in most countries voting requires a single name.<sup>6</sup> However, single-name ballots can also derive from the characteristics of the political supply (where there are no similar candidates that voters would co-approve) or from an intrinsic preference for single-name voting rules (which could be reflected in a *status quo* bias). Hence, it is challenging to disentangle respondents' misunderstanding from other factors. Finally, some studies focus on inconsistent votes, *i.e.*, votes such that voters prefer X and not Y but give a strictly higher evaluation to candidate Y than to X (Igersheim et al. 2016; Baujard et al. 2018; Darmann and Klamler 2023). In these studies, inconsistent votes are generally rare but become more common with broader grading scales (Baujard et al. 2018).<sup>7</sup> Balinski and Laraki (2010) have conducted a similar analysis on inconsistent votes under majority judgment; however, the latter study only assessed the difference between the aggregate results with majority voting and the second round of the actual presidential election in Orsay, a suburban town near Paris.

A third method relies on the relationship between the complexity of a voting rule and voters' behavior in the laboratory (Herzberg and Wilson 1988; Der Straeten et al. 2010; Dellis and Kröger 2023).<sup>8</sup> In particular, Der Straeten et al. (2010) provide the experimental design we build on here, and study voting behavior with one-round plurality, two-round majority, approval voting, and single-transferable voting with Hare transfers. They find that fewer individuals vote strategically with more complex voting rules: specifically two-round

<sup>4</sup> Responses can be influenced by the experimenter demand effect (Zizzo 2010) so that participants might conform to what is viewed as appropriate behavior. Relatedly, participants might want to protect their self-image.

<sup>5</sup> In ranked-choice voting voters must rank every candidate: hence, undervotes imply ranking fewer than the existing candidates, while overvotes imply assigning a rank to the same candidate(s) more than once.

<sup>6</sup> In *in-situ* experiments with approval voting conducted in France in 2012, between 18.62% and 22.43% of voters surveyed approved only one candidate, depending on which scale of evaluative rules was under experiment (Baujard et al. 2012).

<sup>7</sup> Apart from inconsistencies, Baujard et al. (2021) found that introducing a negative grade in evaluative voting disadvantages polarizing candidates while benefiting minor candidates.

<sup>8</sup> In parallel to this literature, several studies in political science have investigated whether people "vote correctly". This concept varies depending on individuals' priorities—some may prioritize expression over instrumental voting to maximize their payoffs—and also hinges on the accuracy of available information (Lau and Redlawsk 1997; Blais et al. 2016).

majority and single-transferable voting. However, in these papers, the complexity of voting rules is assumed rather than investigated, leaving aside how complex each voting rule is from the voters' perspective.

Our contribution to the literature on the properties of voting rules is threefold. First, our experiment features *incentivized* questions that test people's understanding of the aggregation properties of the rules, which is an understudied topic. Second, by considering inconsistent and strategic voting as in the literature mentioned above, our paper combines three components of understanding in one design and tests their interrelation. Third, to our knowledge, our paper is the first to test majority judgment in the lab.

Our paper also contributes to a second strand of literature, namely the behavioral social choice literature initiated by Regenwetter (2006). This literature investigates the divergences between social choice theory and the empirical evidence, as in the case of the Condorcet paradox of cyclical majorities (Regenwetter et al. 2003, 2008). In general, considering how voters use voting rules in the field can help to moderate some pessimistic theoretical predictions (Regenwetter and Tsetlin 2004; Darmann et al. 2019; Lachat and Laslier 2024) and to scrutinize critical normative issues that are concealed by a standard theoretical approach (Baujard et al. 2014; Darmann et al. 2017; Darmann and Klamler 2023). In this paper, our data support the claim that evaluative voting can lead to strategic voting in the form of extreme grades being given to all candidates (Núñez and Laslier 2014). At the same time, the same behavior is observed with majority judgment, which contrasts with what has been claimed concerning its lower manipulability (Balinski and Laraki 2010, 2020).

Third, we contribute to the experimental methodological literature that studies subjects' understanding of instructions in economic experiments. Early studies inquired whether participants' confusion drove some common experimental results. For instance, Andreoni (1995) and Houser and Kurzban (2002) find that in a standard public good game, a participant's choice not to free-ride is partly driven by their confusion about the experimental task. This raises the question of whether participants understand the instructions as well as the implications of this possible confusion. Some papers focus on which types of instructions reduce confusion (Bigoni and Dragone 2012; Alekseev et al. 2017; Freeman et al. 2018; Ramalingam et al. 2018), others on which types of participants are likely to experience more confusion (Bigoni and Dragone 2012; Pulickal and Chakravarty 2023). Koppel et al. (2025) found that many players misunderstand classic games such as the dictator and the trust games, and their misunderstanding predicts their prosocial behavior. Although we do not focus on *how* to deliver instructions, nor on *who* understands them best, our paper brings a fresh perspective on *what* participants may fail to understand in the context of voting rules. In particular, we observe participants' confusion over the median and the implications of this in the case of majority judgment. This result acts as a warning for experiments where the instructions include statistical notions as simple as the median.<sup>9</sup>

The remainder of the paper is organized as follows. Section 2 presents the voting rules tested, namely evaluative voting and majority judgment, and offers a comprehensive definition of "understanding". Section 3 presents the experiment, the design, and the subjects. In Section 4, we set out our results and in Section 5 we discuss them. Section 6 concludes.

<sup>9</sup>For example, many studies offer statistics, including averages, ratios, and quintiles, in order to reduce peoples' misperceptions about society, but the effect is often muted. In a comprehensive literature review, Marino et al. (2024) argue that one possible explanation is that some people might fail to understand the information provided as their analytical reasoning skills are limited.

Detailed figures and the experiment description are provided in the appendices. Data are available on a remote depository under a Creative Commons license.

## 2 Understanding voting rules

### 2.1 Tested voting rules

We restrict our attention to voting rules which only elect one candidate, by contrast with voting rules, such as proportional rules, which are designed to elect a committee. Evaluative voting and majority judgment share the common feature of being a one-round voting rule based on multi-nominal ballots, *i.e.*, where voters assess every candidate with a grade. However, they differ concerning the grades available to voters and their aggregation method.

**Majority judgment.** With majority judgment, each voter assigns a verbal grade to each candidate, and the winner is the candidate who obtains the highest median grade, also called the “majority grade”. However, two or more candidates often receive the same majority grade (according to the pigeonhole principle), and a tie-breaking rule is needed. Unlike other best-median rules, the majority judgment is characterized by a specific tie-breaking rule (Balinski and Laraki 2007, 2010; Fabre 2021). This rule can be implemented in two ways, whose explanations correspond to two treatments in our design.

For a large population of voters, we can apply the *majority gauge method*. In the case of ties, the tie is broken by considering that the largest block of voters who disagree with the majority grade decides, *i.e.*, the largest group of voters who think the candidate is worth more (or less) than the majority grade. In the *iterative deletion method* (the one suited for a low number of voters), if two candidates share the same majority grade, a single majority grade is removed from each candidate until the tie is resolved and the candidates can be ranked.

In this paper, we used six verbal grades as typically suggested in most tests of majority judgment in political elections: “To reject” (in French, *A rejeter*), “Poor” (*Insuffisant*), “Acceptable” (*Passable*), “Fairly good” (*Assez bien*), “Good” (*Bien*), and “Very good” (*Très bien*). The reason why this list of verbal grade is used in majority voting is that they are “known to every French school child, to all those that sat *baccalauréat* examinations marking the end of secondary school, and to university students”, suggesting they allow a homogeneous interpretation of the grade for diverse individuals (Balinski and Laraki 2010).

**Evaluative voting.** With evaluative voting, each voter assigns every candidate a grade on a predetermined cardinal numerical scale, and the winner is the candidate who obtains the highest sum of grades (Hillinger 2006; Smaoui and Lepelley 2013; Macé 2018). This is equivalent to saying that the winner is the one who obtains the highest average grade.

In this paper, we have chosen an evaluation voting system with a 6-step scale ranging from  $-1$  to  $4$ . This choice facilitates comparisons with the scale used in the majority judgment, *e.g.*,  $-1$  captures the idea of “To reject”, and the best grade  $+4$  captures “Very good”.

### 2.2 The components of understanding

How to fill in a ballot. The first component of understanding concerns whether voters know how to fill in a ballot, which depends on what balloting information is required by each

voting rule: a single name, a selection of a subset of candidates, an assessment of every candidate, and so on. In elections requiring a single name to elect a committee or a single winner, the low ratio of nulls (a part of which may be voluntary) suggests that difficulties are rarely encountered. By contrast, when voters independently assess each candidate with a numerical or verbal grade, they may encounter problems in interpreting grades. Indeed, the ordinal meaning of grades is not necessarily similar for everybody, and a given grade may convey dissimilar cardinal meanings (Baujard et al. 2021). Yet the aforementioned evidence shows that most voters cast multi-nominal ballots and assign grades consistent with their preferences with both evaluative voting and majority judgment (Balinski and Laraki 2010; Igersheim et al. 2017; Baujard et al. 2018). Additionally, as long as the grading scale and the number of candidates are the same with both voting rules, as it will be in our experiment, there is no reason why we should observe more mistakes with one rule than the other. Thus, we propose the following:

**Hypothesis 1** Most participants understand how to fill in a ballot with both evaluative voting and majority judgment.

**How votes are aggregated.** The second component of understanding concerns the ability of voters to know how ballots are counted and aggregated to obtain the election outcome with a given voting rule.

This component of understanding helps guarantee the legitimacy of an election outcome, which will be acceptable to voters not only if they approve the rule employed, but also if they can compute the result by themselves and reproduce the same outcome as the officials. This is why scrutinizing the ballots and counting the cast votes is a public and transparent activity in many countries.

However, with complex voting rules the results cannot be computed at a decentralized level or with pen and paper,<sup>10</sup> and the issue of legitimacy becomes all the more important.

Past experiments have provided data on voters' use of more or less complex rules. However, as Laslier (2011) regrets, "Unfortunately, we do not learn from these experiments whether [participants] understand the way ballots are counted. This is not a problem for rules using simple counting schemes, but it is a problem for complex evaluative or ranking ballots."

Evaluative voting and majority judgment differ in their aggregation methods. Therefore, we study whether participants understand the implications of the use of the median and the sum/average. Regarding majority judgment, there are three potential sources of confusion. The first obstacle is the definition of the median itself: some individuals might not be used to computing medians in their daily lives. Second, as a "best median" rule, majority judgment does not respect the intuitive features of additive rules. In particular, when we compute the sum or the average of grades, assigning a higher number of better grades to one candidate than another always improves the outcome for the first candidate; however, this monotonicity property does not hold when candidates are ranked on the basis of their median grades in the case of a variable population. This violation of the monotonicity property in the context of variable population is a counter-intuitive feature of the median that is likely to create confusion. Third, the tie-breaking rule could be a potential (and major) reason for voters' misunderstanding. This paper does not tackle people's understanding of the tie-breaking

<sup>10</sup>For instance, the single-transferable vote used in Ireland requires the centralization of the ballots, while majority judgment and most participatory budgeting allocation rules require an algorithm to compute the outcome.

rule, as this could complicate the comparison between majority judgment and evaluative voting. Notably, these three obstacles are absent in evaluative voting: computing sums or averages is a daily task for most people; the rule respects the monotonicity property; and no tie-breaking rule is needed. Therefore, we propose the following:

**Hypothesis 2** More individuals understand how votes are aggregated with evaluative voting than with majority judgment.

**How to vote strategically.** The third component of understanding concerns voters' behaviors with respect to rules. First, rules have various properties.<sup>11</sup>

Second, each voting rule calls for different voting behavior that is adapted to the specifics of the rule. Understanding how to vote can be interpreted as the ability to vote strategically with the rule one is presented with, especially in the laboratory, where voting for fictitious candidates has no cost in terms of preference expression.

For evaluative voting, Núñez and Laslier (2014) show that strategic voters should overstate their preferences, using only extreme grades (either the lowest or the highest) and discarding intermediate grades. However, previous evidence shows that a significant proportion of voters use intermediate grades (Igersheim et al. 2016, 2022), although they do so less frequently in laboratory elections (Igersheim et al. 2016). By contrast, majority judgment is strategy-proof in grading, given that the aggregation method relies on the median grade (Balinski and Laraki 2010, 2020). Indeed, assigning an extreme or slightly lower (higher) grade to a candidate does not alter its majority grade as long as both grades are above or below the median. If voters vote strategically, there should be no change in the voting outcome. Hence, voters in this case ought to vote more sincerely than with the highly manipulable evaluative voting. Therefore, we propose the following:

**Hypothesis 3** Individuals assign extreme grades more frequently when using evaluative voting than majority judgment.

As measuring people's understanding is hardly straightforward, and because we are not aware of a specific experimental protocol targeting such an aim, we designed a protocol in which participants vote in elections with monetary incentives under evaluative voting and majority judgment, and are then given some exercises and questions to answer.

### 3 The experiment

#### 3.1 Experiment design

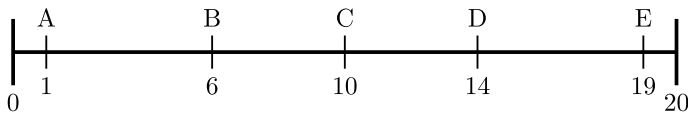
The experiment comprises three parts, as illustrated in Table 1: The first part is designed to teach participants various voting rules, among which the majority judgment and evaluative voting,<sup>12</sup> and test them in a given context. Depending on the treatment to which participants are assigned, the second part is devoted to voting with evaluative voting or majority judgment.

<sup>11</sup> Laslier (2011) also refers to the ability of voters to identify the consequences of a voting rule on voting outcome, such as whether a rule will favor consensual or minor candidates. It would be interesting to compare participants' understanding to the benchmark of the scientific knowledge. However, the definition of a benchmark requires both the existence of numerous theoretical and empirical publications forming a scientific consensus on how these rules impact electoral outcomes, which is still a work in progress for majority judgment and evaluative voting. For this reason, we do not tackle this issue here.

<sup>12</sup> Specifically, they also try the official two-round system, used in France in the general elections, and approval voting.

**Table 1** Sequence of the experiment

|          |                                                                                                                                                                                                   |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Part I   | – Learning of voting rules<br>– Vote in specific context (not incentivized)<br>– Satisfaction questions                                                                                           |
| Part II  | – First series of elections - Uniform preference distribution<br>– Second series of elections - Trimodal preference distribution<br>– Third series of elections - Bimodal preference distribution |
| Part III | – Comprehension questions<br>– Satisfaction questions<br>– Socio-demographic questionnaire                                                                                                        |

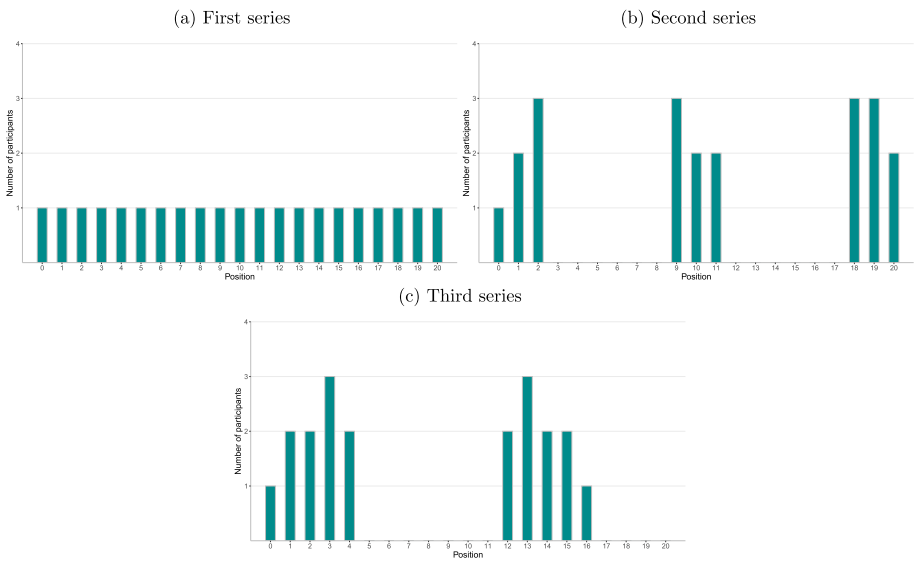
**Fig. 1** Position of the five candidates

ment in fictitious elections with controlled preferences. Finally, the third part involves a survey to gauge participants' comprehension and satisfaction with the voting rules to which they have been introduced. Some socio-demographic questions follow. The instructions, translated from French to English, are presented in the online Appendix. In the following paragraphs, we present each part of the experiment in detail.

**Part I.** In the first part, participants read the explanation of each voting rule and are asked to cast votes under evaluative voting and majority judgment in a hypothetical scenario in which they are a member of a housing assembly tasked with deciding on a single color for all the walls of the common area in a condominium. Moreover, participants answer some questions about their satisfaction with the different voting systems.

**Part II.** In the second part of the experiment, we use a standard design that has already been employed in many publications to study the impacts of voting rules, which candidates they favor, and voters' behavior (e.g. Der Straeten et al. 2010; Igersheim et al. 2016). Twenty-one participants vote for five candidates, labeled A, B, C, D, and E, presented on a 0 to 20 scale axis (see Fig. 1). The candidates are symmetrically positioned so that there is an extreme left candidate (A) located at 1, a moderate left (B) located at 6, a centrist (C) located at 10, a moderate right (D) located at 14, and an extreme right (E) located at 19. Participants are informed that the candidates' positions remain constant throughout the experiment.

Participants vote across three series of five elections each. At the start of each series, each participant is allocated a position on the  $[0, 20]$  axis, which remains fixed throughout the series. Their position represents their preferences. The distribution of participants' positions varies across the three-election series: In the first series, preferences are uniformly distributed, meaning that each participant is assigned one position between 0 and 20 with equal probability (Fig. 2a). In the second series, the distribution is trimodal, with the preference parameter taking a value in either the  $[0-2]$ , the  $[9-11]$ , or the  $[18-20]$  interval (Fig. 2b). This preference distribution resembles the French political spectrum, especially since 2022, and several other countries' political spectrum. Finally, the last series has a bimodal distribution, with preferences taking a value in the  $[0-5]$  or the  $[12-16]$  intervals (fig. 2c). This preference distribution mirrors the bi-partisan political spectrum, for instance in the US



**Fig. 2** Voters preferences distribution

case. Having three election series allows us to test voting behavior in a stylized case (first series) and two more realistic cases (second and third series) and assess whether the preference distribution impacts voting behavior and the election outcome (with the latter being a research question relegated to another paper).

Participants are made aware that their monetary reward depends linearly on the difference between their position and the elected candidate's position. The maximum gain, set at 20 tokens (10 euros), is achieved when there is no difference (a distance of zero) and decreases by one euro for each point of difference. Participants are informed of their payoff after each election, but they further know that their *actual* payoff depends only on the last election in each 5-elections series. Specifically, the final payoff for this section is equal to the average of the payoffs from the three fifth elections.

**Treatments:** Two treatments vary the voting system used in the elections in the second part: In the *Evaluative voting* treatment (EV, one-third of the sessions), participants vote by assigning a numerical grade between  $-1$  and  $+4$  to each candidate. Additionally, in the *Majority judgment* treatment (MJ, two-third of sessions), participants vote by assigning a verbal grade between “To reject”, “Insufficient”, “Acceptable”, “Fairly good”, “Good”, and “Very good” to each candidate.<sup>13</sup>

**Part III.** The last part consists of six questions on evaluative voting and six on majority judgment displayed to every participant, regardless of the treatment they are in. In the first three questions, we ask participants to solve some exercises to determine which of two candidates won an election based on some votes expressed through grades. There are specific

<sup>13</sup> Within the MJ treatment, the explanation of the tie-breaking rule relies on the majority gauge for half of the participants and on the iterative deletion process for the other half (see Section 2 for a description of the two methods). Since this paper does not examine participants' understanding of the tie-breaking rule, we pool these data in the analysis for simplicity.

circumstances in which calculating the election result for a profile works differently for the two rules, and must therefore be understood differently.

One particular feature among these differences inspires the exercise we propose. A rule respects the monotonicity property if the outcome in favor of a candidate is reinforced after a change, either because the intensity of the support by a voter improves in the fixed population case, or because there are additional voters in favor of this candidate. Because evaluative voting is additive, any improvement in the support of a candidate will improve their voting outcome, and never lower their ranking. Conversely, in the variable population case, majority judgment may violate the monotonicity property since the ranking of candidates is not derived from the (sum or) average grades, but from the median grades.

We specifically observe the monotonicity feature, by comparing participants' answers to the second and the third question. In the third question about evaluative voting, we add two voters to the nine voters from the previous question (see Table 2). The winner in the second question is candidate B (with 9 points, against A who only gets 8 points). In the third question, the two new voters assign high grades to both candidates: +3 to the previous winner B, and +2 to A. The candidate B receiving two +3 grades is favored and wins the new election (now with 15 points against 12 for A). Indeed, more intense support for a winner consolidates his position under evaluative voting.

Questions two and three are similar for majority judgment. However, the logic of vote aggregation with majority judgment is different. In the second question, Candidate B wins, because B's median grade "Fairly good" is better than A's median grade "Acceptable". With two new voters, we need to examine which candidate is supported by a group of six people, likely to be composed differently than the group of five from the previous majority. And indeed, even though candidate B receives new better grades than A does, he eventually loses in the new election. Indeed, in our case, the new grades are all higher than the previous median grades ("Very good" for the previous winner and "Good" for the previous loser), lifting A's median grade but not B's. After population change, A hence wins with "Good" against B whose median grade is still "Fairly good". Questions two and three are then aimed at assessing whether participants understand how a voting outcome is differently impacted by a higher number of better grades depending on the rule.<sup>14</sup>

**Table 2** Second and third questions

| Grades                | -1     | 0       | 1       | 2           | 3    | 4       |
|-----------------------|--------|---------|---------|-------------|------|---------|
| (a) Evaluative voting |        |         |         |             |      |         |
| Cand. A               | XXXX   | X       |         | XXX         | XX   | X       |
| Cand. B               | XX     | XX      | XX      | X           | XXX  | X       |
| Grades                | Reject | Insuff. | Accept. | Fairly Good | Good | V. Good |
| (b) Majority judgment |        |         |         |             |      |         |
| Cand. A               | X      | XXX     | X       |             | XXX  | XXX     |
| Cand. B               | XX     |         | XX      | XX          |      | XXXXXX  |

**Note:** Each black cross indicates a vote in the second question. Bold crosses represent the additional votes given in the third question

<sup>14</sup> As participants process exercises successively, first regarding evaluative voting and then majority judgment, they might remember their answers and a similar answer would not be meaningful. Hence, in order to avoid that they too easily fall into this trap, participants begin from a different vote distribution in question 2 for evaluative voting and for majority judgment. However, given that the number of votes is the same, the difficulty of computation of their voting outcome is equivalent.

In the last three questions, participants choose whether a statement is true or false. These statements are focused on the final evaluation obtained for each candidate and how they are computed—specifically, whether they are calculated as the mean or the median of the grades that voters assign. Therefore, we assess whether participants can disentangle the mean and median definitions and associate them with the correct voting rule.

These comprehension questions are incentivized, meaning that participants' payoffs depend on the number of correct answers they provide. The aim is to increase participants' concentration and reduce the likelihood of random answers. Each correct answer to the first three questions is rewarded with 2 points, a wrong answer receives 0 points, and answering "I do not know" is worth 0.5 points. For the last three questions, participants earn 2 points if all answers are correct, 0 points if at least one answer is incorrect, and 0.5 points if participants always choose not to answer.

### 3.2 Procedures

The experimental sessions were conducted in the experimental lab of the *Groupe d'Analyse et de Théorie Economique* (GATE-LAB), located in Ecully, next to Lyon, France, in 2023. We conducted nine sessions with 21 participants per session, for 189 participants. On average, participants took 1 hour and 6 minutes to complete the experiment and gained 20.03 euros, including a 5 euros show-up fee.

Women represent 53% of our sample. 91% of the participants are students, and this composition is reflected in the age composition of our sample, as 88% of participants are between 18 and 24 years old. Within the student group, 21.71% study economics, 33.71% management, and 30.29% engineering. For a comparison of descriptive statistics across treatments, see Table 6 in the Appendix.

## 4 Results

### 4.1 How to fill in a ballot

A first *non-incentivized* measure of participants' understanding of how to fill in a ballot involves a question in the first part of the experiment on how a participant can provide maximal support to a candidate when using grades or judgments. Of course, the correct answer is assigning the maximal grade (+4) or judgment (*Very good*) to the candidate. We find that 94.71% of participants answer the question correctly for evaluative voting and 95.24% answer the question correctly for majority judgment.

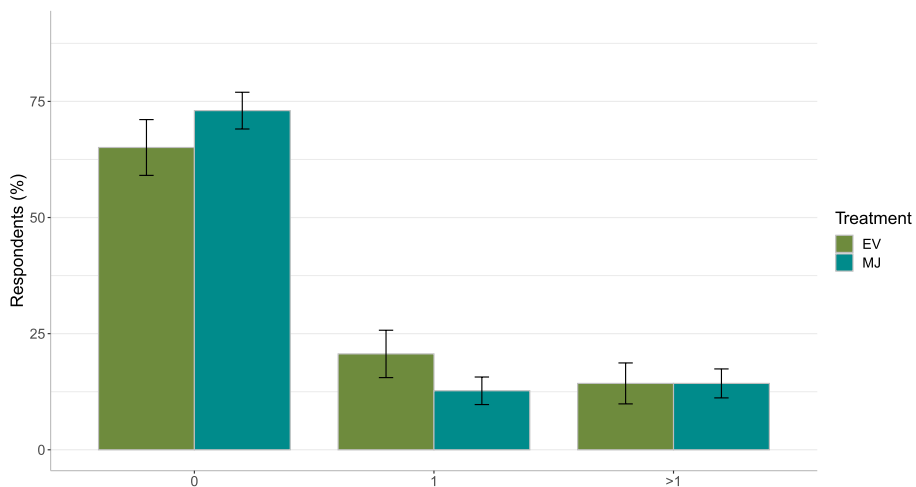
A second *incentivized* measure of this understanding component is to consider the consistency of participants' votes, similarly as in Baujard et al. (2018) and Darmann and Klamler (2023). We consider the vote participants give to all pairs of candidates in the last election of each series, where the election results impact voters' payoff. A vote is defined as inconsistent when a participant assigns a higher grade to a candidate more distant from his position than another candidate.<sup>15</sup> The total number of comparison is then 1890 in the EV treatment and 3780 in the MJ treatment. We find that 94.86% of votes in the EV treatment and 96.32% of

<sup>15</sup> In the case in which two candidates are equidistant (the situation of participants in positions 8, 10, and 12), a vote is defined as inconsistent if the participant gives different grades to the two candidates.

votes in the MJ treatment are consistent. These proportions are virtually the same in the first election of the series (94.87% in the EV treatment and 96.32% in the MJ treatment).<sup>16</sup> This suggests that the high level of participants' consistency is immediate from the beginning and remains constant throughout the experiment.

If we consider the total number of inconsistent votes per participant (see Fig. 3), we find that 65.08% of the sample always votes consistently in the EV treatment out of 30 candidates' comparisons, and the proportion is 73.02% in the MJ treatment. Of the remaining participants, some vote inconsistently only once (20.63% in the EV treatment, 12.70% in the MJ treatment) and some more than once (14.29% in both treatments). The difference between treatments is not statistically significant (Fisher's exact test:  $p = 0.348$ ) and is mainly driven by participants voting inconsistently only once, which might result from a minor lapse in attention. This leads to the average number of inconsistent equal to 1.54 in the EV treatment and 1.1 in the MJ treatment, with the difference not being statistically significant (two-sample t-test:  $p = 0.38$ ). Therefore, our results corroborate Hypothesis 1, and we can make the following statement:

**Result 1** Most participants understand how to fill in a ballot with both evaluative voting and majority judgment. There are few inconsistencies, and we cannot infer a difference in inconsistent votes between the two voting rules.



**Fig. 3** Number of inconsistent votes, by treatment. **Note:** Number of inconsistent votes per participant, where an inconsistent vote is defined as a voter assigning a higher vote to a candidate who is further from a candidate they prefer more. We consider the vote in only the fifth election in each election series. This means that, for each voter, there are 30 candidates' comparisons in total (10 for each series). Standard errors indicated by the bars. The difference between treatments is not statistically significant (Fisher's exact test:  $p = 0.348$ )

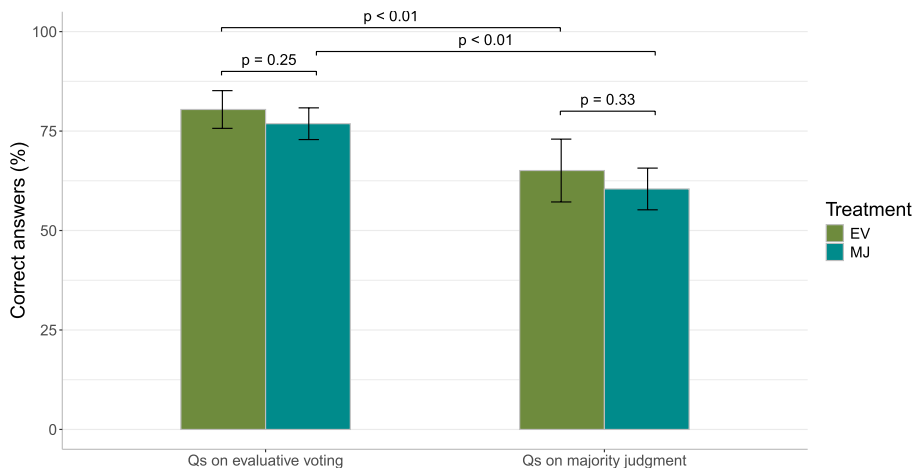
<sup>16</sup> Only four participants gave the same grade to every candidate, specifically, the same verbal grade in the MJ treatment.

## 4.2 How votes are aggregated

The second component of understanding pertains to how votes are aggregated. Figure 4 depicts the average percentage of correct answers to questions on this aspect for evaluative voting and majority judgment (see instructions in the online Appendix for questions on evaluative voting and majority judgment). Given that participants answer these questions after voting with one of the two rules, we split the results by treatment to analyze whether the previous experience impact understanding.

The average share of correct answers to questions on evaluative voting is 80.42% in the EV treatment and 76.85% in the MJ treatment, whereas for majority judgment questions it is 65.08% in the EV treatment and 60.45% in the MJ treatment.<sup>17</sup> Hence, evaluative voting exhibits a higher level of understanding than majority judgment across all treatments, with a statistically significant difference (in both treatments, two-sided paired t-test:  $p < 0.01$ ). Moreover, the comprehension of majority judgment does not improve in the MJ treatment in which participants have had the opportunity to experience this voting rule extensively in part II. If anything, these participants' comprehension is even worse, but the difference is not significant (two-sided t-test:  $p = 0.33$ ). By contrast, participants in the EV treatment score slightly better in the questions about evaluative voting, but the difference is again not significant (two-sided t-test:  $p = 0.25$ ).

Another notable difference in Fig. 4 is the larger confidence intervals in the share of correct answers for questions on majority judgment compared to those on evaluative voting. Indeed, the standard deviation of correct answers to the majority judgment questions



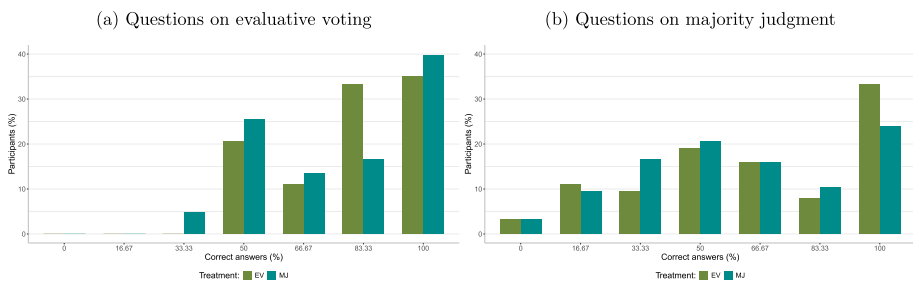
**Fig. 4** Average share of correct answers to the comprehension questions. **Note:** The treatment relates to the rule participants voted on before answering the comprehension questions in Part II. Confidence intervals at 95% indicated by the bars. In both treatments, the average share of correct answers to questions about evaluative voting is higher than the share of correct answers to questions about Majority Judgment (two-sided paired t-test:  $p < 0.01$ )

<sup>17</sup> In both cases, the remaining answers are nearly evenly distributed between incorrect responses and participants opting not to answer. Interestingly, participants who claimed to know MJ before the experiment (27 in the last five sessions) scored even worse (56% of correct answers) than other participants.

is significantly higher than in the evaluative voting questions in both treatments (F-test:  $p < 0.01$ ). To further explore this difference, Fig. 5 shows the distribution of correct answers to the questions on evaluative voting and majority judgment.<sup>18</sup> For both voting rules, the distribution mode is 100% correct answers. However, only 6 participants (3.17%) answer correctly less than 50% of the time to questions on evaluative voting, while this is the case for 52 participants (27.51%) for majority judgment questions. In other words, there is greater heterogeneity in the comprehension of majority judgment.

The next step in the analysis is to understand which questions were difficult for participants. In the first three questions on the results of an election where participants can see the voters' votes for two candidates, participants perform remarkably well when answering about evaluative voting (94% of correct answers, on average). In contrast, they perform less well answering questions on majority judgment (74.25% of correct answers, on average). In particular, subjects are much more likely to be wrong in their response to the third question on majority judgment (61.38% correct answers) than in the case of evaluative voting (94.71% correct answers). As explained in Sect. 3, the third question is aimed at evaluating participants' understanding of the monotonicity property when the set of voters varies. This indicates that subjects believe both evaluative voting and majority judgment rules to be monotonic, whereas in fact monotonicity is a feature only of evaluative voting.

Participants score generally worse in the True and False questions (see Table 7 in the Appendix). In the questions about evaluative voting, the average share of correct answers is 62.08%, while the share is 49.74% for questions on majority judgment. The worst score for questions on evaluative voting pertains to the statement “The grade corresponds to the average score given to this candidate by all voters” (47.62% correct answers). However, participants' confusion likely arises from the fact that, in the experiment, the grade is presented as the sum rather than the average, except once at the beginning of the experiment. The decision to use the average in the statement was influenced by the need to compare this statement to the equivalent one in the case of majority judgment. The worst score for majority judgment pertains to the last question, where only 43.39% of participants confirm that the majority grade given to a candidate “Corresponds to the judgment given by the voter who divides in two the set of voters ranked from the most favorable to the most unfavorable”. This suggests that participants often fail to understand the definition of the median, confirming our Hypothesis 2.



**Fig. 5** Distribution of share of right answers. **Note:** The treatment relates to the rule participants voted on before answering the comprehension questions in Part II

<sup>18</sup> Figures 7 and 8 also display the distribution for incorrect and “I don’t know” answers.

To summarize:

**Result 2.1** Participants' understanding of vote aggregation is, on average, lower under majority judgment than under evaluative voting.

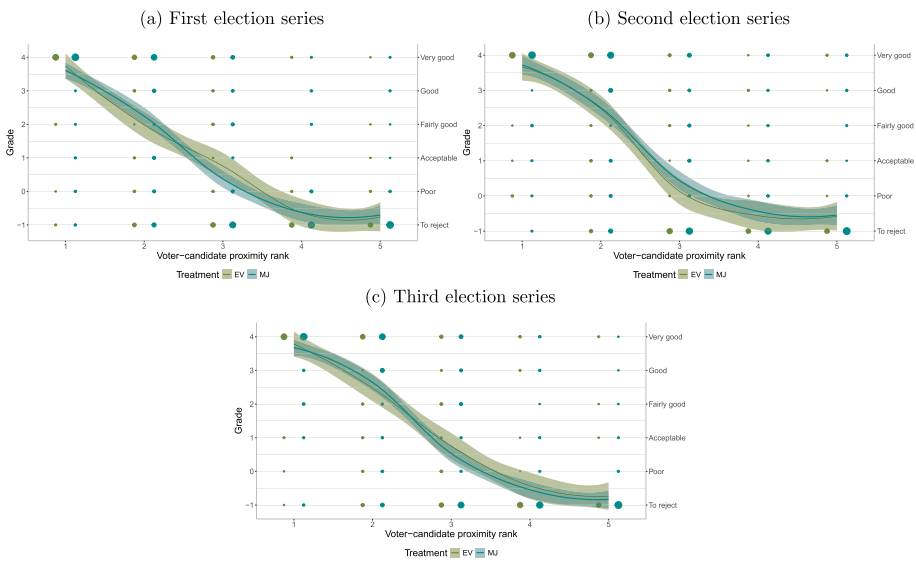
**Result 2.2** The most difficult barrier to understanding how votes are counted in majority judgment is its definition of median and its implication in terms of vote aggregation.

**Result 2.3** Participants' understanding of majority judgment is more unequal than their understanding of evaluative voting. Some participants have a strong grasp of how votes are aggregated, and others have little to no understanding.

Result 2.1 confirms Hypothesis 2 and is explained by the observation that the median's properties are hardly understood (Result 2.2). Non trivially, Result 2.3 highlights voters' understanding heterogeneity under majority judgment, which we will show in Sect. 4.4 to be important to predict their voting behavior.

### 4.3 How to vote strategically

To investigate participants' voting behavior, we first define a voter–candidate proximity rank (1 to 5) based on the distance between a participant and each candidate.<sup>19</sup> Figure 6



**Fig. 6** Voting behavior. **Note:** We consider the vote in only the fifth election in each election series. The voter–candidate proximity rank takes the value 1 if the candidate is the closest to a participant's position, 2 if the candidate is the second closest to the participant's position, etc. The lines represent the fitted values of local regressions. The shaded areas indicate confidence intervals at 95%

<sup>19</sup> If, for one participant, two candidates are equidistant, they display the same rank.

shows the relationship between the voter–candidate proximity rank and the grades participants assigned to each candidate in the fifth election of each series. We observe no differences in voting behavior across treatments. First, participants assign the highest grade to the closest candidate so that, by pooling the three election series, 92.06% of votes for the closest candidate are +4 in the EV treatment and 89.15% are “Very good” in the MJ treatment. Second, participants gradually decrease grades for candidates ranked second and third while providing intermediate grades. Finally, the majority of grades assigned to the fourth and fifth furthest candidates are the minimum, either −1 (83.07% to the fourth and 91.01% to the fifth candidate) or *To reject* (80.95% to the fourth and 89.42% to the fifth candidate).

Next, we count the number of extreme votes per election and assess participants’ learning throughout the election series. In Table 3, we show the share of extreme grades in the EV and MJ treatments in each election series and repetition in each series. In the EV treatment, the share of extreme grades evolves from 66.98 to 86.03% in the first series, from 77.78 to 85.08% in the second, and from 85.08 to 86.35% in the third. Similarly, in the MJ treatment, the share increases from 63.65 to 79.68% in the first series, from 76.51 to 81.90% in the second series, and from 80.95 to 82.22% in the third. Thus, with both voting rules, the number of extreme votes is already high from the first election in each series and increases as elections progress.

Finally, since participants’ voting behavior likely depends on their previous votes and the other participants’ voting behavior in their session, a robust test to examine differences in voting behavior across treatments involves regressing the number of times participants assign an extreme grade in the election on the MJ treatment variable while clustering standard errors at the individual and session level. We run this regression for the last election of every election series, and the three election series pooled to maximize statistical power. We also control for participants’ position fixed effects, participants’ share of correct answers to the questions on evaluative voting and majority judgment, and participants’ gender. The results of this analysis are presented in Table 4, indicating no significant effect of the treatment on the number of extreme grades. Besides the lack of statistical significance, the treatment effect is also low compared to the share of extreme grades in the EV treatment. Indeed, adding the MJ-treatment effect to this share gives us an average number of extreme grades between 4.05 and 4.15 in the MJ treatment. Summing up, our data do not provide support for more sincere voting with majority judgment (Hypothesis 3), hence:

**Result 3** Participants vote the same way in the EV and MJ treatments by assigning mostly extreme grades.

#### 4.4 Correlation between the components of understanding

We now study whether the three components of understanding are correlated, shedding light also on participants’ voting behavior. To do this, we use three variables at the participant level: the number of inconsistent choices in the last election of the three series, the share of correct answers to questions on either evaluative voting or majority judgment, and the number of extreme votes the candidate gave to each candidate in the three elections. Figure 9 in the Appendix presents the correlation matrix where we include these three variables and split the sample by treatment. The strongest correlation pertains to participants giving extreme grades, who tend to vote consistently in both the EV and MJ treatments. More-

**Table 3** Extreme votes and learning

|    | First series |       |       |       |       | Second series |       |       |       |       | Third series |       |       |       |       |
|----|--------------|-------|-------|-------|-------|---------------|-------|-------|-------|-------|--------------|-------|-------|-------|-------|
|    | 1            | 2     | 3     | 4     | 5     | 1             | 2     | 3     | 4     | 5     | 1            | 2     | 3     | 4     | 5     |
| EV | 66.98        | 75.87 | 82.22 | 84.44 | 86.03 | 77.78         | 81.90 | 83.49 | 86.03 | 85.08 | 85.08        | 85.71 | 84.13 | 85.71 | 86.35 |
| MJ | 63.65        | 69.21 | 73.49 | 76.83 | 79.68 | 76.51         | 78.89 | 80.79 | 81.27 | 81.90 | 80.95        | 80.48 | 81.27 | 81.11 | 82.22 |

Note: Each value represents the share of votes equal to +4 or -1 in the EV treatment, and “Very good” or “To reject” in the MJ treatment

over, in the EV treatment, there is no significant nor strong correlation between participants' understanding of the aggregation properties of evaluative voting and their consistency and extreme voting. This result is likely driven by the extremely low variation of the share of correct answers (*i.e.*, most participants understand well, see Fig. 5). By contrast, in the MJ treatment, participants answering correctly to the questions about vote aggregation by median are less likely to vote inconsistently and more likely to assign extreme grades.

The relationship between participants' understanding of vote aggregation and their voting behavior is further analyzed in Table 5, where we regress the number of extreme grades per election in all election series on participants' share of correct answers to the questions on the majority judgment and evaluative voting. A deeper understanding of vote aggregation under evaluative voting does not increase extreme grading. However, as already noted, the low variation in people's understanding of how votes are aggregated likely explains the absence of correlation. By contrast, in the MJ treatment, participants who understand vote aggregation under majority judgment assign extreme grades more frequently ( $p < 0.01$ ). This variation goes from an average of 3.4 extreme grades in the three elections for those participants answering well to less than half of questions, to 4.34 extreme grades for those answering well to more than half of the questions.

Finally, Fig. 10 illustrates the MJ treatment effect on the use of extreme grades (same analysis as in Table 4) for participants with low and high understanding of vote aggregation under majority judgment. We find that participants who have a low level of understanding vote less strategically in the MJ treatment than in the EV treatment. This highlights how the only differences between the rules are driven by those participants with low understanding of majority judgment who vote less strategically.

We conclude:

**Result 4** When voting with majority judgment, participants who do not understand its vote aggregation method use extreme grades less frequently than those who understand it.

## 5 Discussion

In this section, we critically discuss our primary findings regarding ballot casting, vote aggregation, and strategic voting.

**How to cast a ballot.** Our first result indicates that participants had no difficulty understanding how to cast a ballot, as only a few voted inconsistently under both evaluative voting and majority judgment. However, this result might stem from the limited number of grades available to voters—specifically, six in the experiment. For instance, Baujard et al. (2018) found that inconsistent votes in evaluative voting increase with broader grading scales, occurring in 3.41% of cases with the  $[-1, 0, 1]$  scale and 8.9% with the  $[0-20]$  scale. Inconsistent votes might become more likely with a wider voting scale under majority judgment as well, even though it is not clear whether this framing effect will be lower or higher than in evaluative voting. Comparing this framing effect with different rules is then an empirical question worth to be explored in the future.

**Understanding of vote aggregation.** Our second finding indicates that voters experience confusion regarding how votes are aggregated under majority judgment because of the difficulty in understanding the median, whereas this confusion is less prevalent under

**Table 4** Regression analysis - Extreme votes

|                            | 1st series     | 2nd series      | 3rd series     | All series pooled |
|----------------------------|----------------|-----------------|----------------|-------------------|
|                            | (1)            | (2)             | (3)            | (4)               |
| MJ                         | -0.25<br>(0.2) | -0.099<br>(0.1) | -0.15<br>(0.2) | -0.17<br>(0.1)    |
| Average extreme votes (EV) | 4.30           | 4.25            | 4.32           | 4.29              |
| Observations               | 189            | 189             | 189            | 567               |
| R-squared                  | 0.28           | 0.16            | 0.12           | 0.15              |

**Note:** OLS regressions. The dependent variable is the number of times a participant assigns an extreme grade in each election in columns (1) to (3), and in all elections in column (4). This ranges between 0 and 5 in columns (1) to (3) and between 0 and 15 in column (4). We control for participants' position fixed effects, gender, and share of correct answers to the questions on evaluative voting and majority judgment. Standard errors clustered at the individual and session level in parentheses. \* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%

**Table 5** Regression analysis - Extreme votes and vote aggregation understanding

|                       | Treatment: EV     | Treatment: MJ        |
|-----------------------|-------------------|----------------------|
|                       | (1)               | (2)                  |
| % correct answers EV  | 0.0029<br>(0.01)  | 0.0081<br>(0.005)    |
| % correct answers MJ  | 0.0039<br>(0.005) | 0.0083***<br>(0.002) |
| Average extreme votes | 4.29              | 4.06                 |
| Observations          | 189               | 378                  |
| R-squared             | 0.18              | 0.18                 |

**Note:** OLS regressions. The dependent variable is the number of times a participant assigns an extreme grade in the three elections. This ranges between 0 and 5. We control for participants' position fixed effects and gender. Standard errors clustered at the individual and session level in parentheses. \* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%

evaluative voting. The claim that the median's aggregation properties are harder to grasp than the sum's seems trivial *per se*. However, this highlights an often-overlooked aspect of understanding majority judgment, given that proponents frequently emphasize only the simplicity of casting a ballot.<sup>20</sup> Our contribution is indeed to decompose different components of understanding, to prevent such oversights. More importantly, our scrutiny crucially emphasizes the challenges associated with the different levels of understanding (Result 2.3). Some voters grasp the concept well, while many others do not. As a consequence, the heterogeneity in understanding brings unequal empowerment among voters. Indeed, this leads to different frequencies of strategic voting (Result 4), hence potentially contributing to unequal outcomes. These results echo Eggers and Vivyan (2020)'s finding that older voters and richer voters vote more strategically than younger and poorer voters. Additionally, they support Baujard and Lebon (2022)'s observation that more marginalized individuals are less likely to vote strategically, even when the cost of not doing so is relatively high for them.

Furthermore, the fact that participants misunderstood the properties of the median is significant, especially since our sample consisted mainly of bright students with a supposedly strong mathematical background. Indeed, among the students in our sample (91%), 54.29% of them come from the "École Management Lyon" (a management higher education institution) and 27.43% from the "École Centrale" (an engineering higher education institution), two institutions where students are required to have a high level of mathematics. Furthermore, the participants in our sample were in an ideal situation to learn voting rules, practice them several times, and to vote with clear monetary incentive. This situation is difficult to replicate in real life, as people vote only rarely and might pay little attention to information given about new rules. Finally, the current protocol did not measure any understanding of the tie-breaking rule. Should a protocol include the tie-breaking rule, which is a major component of vote aggregation in majority judgment, the measure of its misunderstanding would most likely be even higher. For these three reasons, our result of voters' misunderstanding represent a lower bound.

**Strategic voting under Majority Judgment.** Another fundamental result of this paper is that people do not vote more sincerely under majority judgment compared to evaluative voting, countering our hypothesis 3, based on Balinski and Laraki (2010, 2020). If voters vote strategically also under majority judgment, it means that they perceive it to be manipulable. The possibility to manipulate the majority judgment is discussed by Felsenthal and Machover (2008) and Laslier (2019): If there is a tie between two candidates, a voter can influence the outcome by exaggerating his preferences, exactly as in the evaluative voting case. In the majority judgment case, this means assigning a grade above the median to the

<sup>20</sup> Balinski and Laraki (2010) argue: "It has been contended by some that the majority judgment ballot is too complex, that it asks too much of voters. We do not believe so. On the contrary, the inputs are arguably the simplest for voters. First, because the ballot has greater cognitive simplicity: it is very natural to evaluate the merits of candidates in a scale defined by words. Second, because every electoral experiment to date has shown that voters were able to fill out the ballots quickly, and moreover appreciate being able to better express their opinions. It is clear that for voters it is easier to evaluate candidates in a natural scale of six grades than to rank-order candidates, and that it permits them to express their opinions more accurately. Some contend that the first- and two-past-the-post methods are by far the simplest for voters to understand. We do not believe so. Although it is easy enough to name one candidate, the subtleties of Arrow's and Condorcet's paradoxes engendered by these methods are not well understood by voters. In addition, voters are frustrated because they are unable to express their opinions concerning all the candidates. Instead they are faced with a stark strategic choice that is more difficult to resolve than it is to give one's honest evaluations of the several candidates".

favorable candidates and below the median to the others. Of course, this does not require using extreme grades, as long as the grades are above or below the median. However, using extreme grades increases the likelihood of a grade being above or below the median, provided voters do not have fixed beliefs about other voters' support for the candidates.

This intuition is reinforced by Result 4, showing that participants with a high understanding of the majority judgment aggregation properties tend to vote more strategically (*i.e.*, using more extreme grades) than those who understand less the median's aggregation properties. This might suggest that voters with a low understanding of vote aggregation follow a simple heuristic to face complexity relatively closer to a sincere expression of their preferences. Other papers already found that more complex voting rules lead to less strategic voting (Herzberg and Wilson 1988; Der Straeten et al. 2010; Aziz et al. 2015; Dellis and Kröger 2023). This paper shows that, for a given rule, depending on how people understand it, there can be differences in the recurrence of strategic voting.

Finally, we need to highlight a limit of our analysis: Assigning only extreme grades to fictitious candidates with assigned preferences in the lab has no cost in terms of preference expression, while in real-life elections voters might want to express their nuanced opinions. Hence, we might expect more sincere voting under both evaluative voting and majority judgment. However, it remains an empirical question whether this would also widen the gap in strategic behavior between the two systems.

## 6 Conclusion

Philosophers of science have pointed out that the epistemic value of scientific assertions should be scrutinized in practice and not just in theory (Ankeny et al. 2011). Economists often use normative criteria as inputs in scientific assertions and may conceive tools and devices in isolation of any operational constraints on their application. However, taking practice seriously requires scrutinizing how contextual information may affect the validity of such assertions. It is therefore important to examine whether, and how, the resulting scientific claims are still valid in other contexts, as illustrated by Baujard (2022) in the domain of public policy evaluation. This paper adapts this latter approach to the case of voting rules, highlighting the assumption that voters understand the rules as a crucial factor that needs to be tested empirically. We also argue for considering understanding as a key normative criterion in the selection of voting rules. To this end, we leverage an experimental design where participants learn and practice the evaluative voting and majority judgment to analyze their understanding of these voting rules.

This paper's first contribution is distinguishing three components of understanding—how to fill in a ballot, how votes are aggregated, and how to vote strategically—and providing experimental measures for each. We then analyze how these components of understanding differ for evaluative voting and majority judgment. The experimental data we gathered allow us to highlight four main results. First, participants vote consistently with both tested voting rules. Second, most voters understand how votes are counted. However, participants' understanding of majority judgment is, on average, lower and more heterogeneous than their understanding of evaluative voting because of their difficulties with defining the median and grasping its implications for the voting outcome. Third, we show that participants vote similarly with evaluative voting and majority judgment. The data confirm the theoretical prediction that under evaluative voting there would be more strategic voting through the use of extreme grades, but it contradicts the prediction of more sincere voting under majority judgment. Fourth, those participants who understand the system of vote aggregation with majority judgment are more likely to assign only extreme grades.

There are numerous potential avenues for future investigation. First, our methodology can be easily applied to other voting rules to test the relationship between voters' understanding of these rules' properties and their voting behavior. Second, it would be insightful to study how voting rules' complexity can impact participants' preferences over different rules.<sup>21</sup> Third, a dimension not tackled here is which normative properties individuals associate with different rules (*e.g.*, which candidates will be favored by each voting rule). This line of inquiry is promising as it could help explain voters' preferences for different voting rules and their possible reluctance to adopt new ones. It could help reveal how voters' wrong beliefs about rules can influence their support for them. Fourth, it would be possible to include a benchmark treatment with limited information on the rules before individuals get to vote, in order to study the causal impact of voters' understanding on their behavior. Finally, we remark that the whole of our society is pervaded by rules of all kinds: Some rules are designed to facilitate the assessment or analysis of states of affairs, others to aid in individual decisions, and others still to frame collective decision-making. We observe that there is a need for more research focused on the layman's understanding of rules.

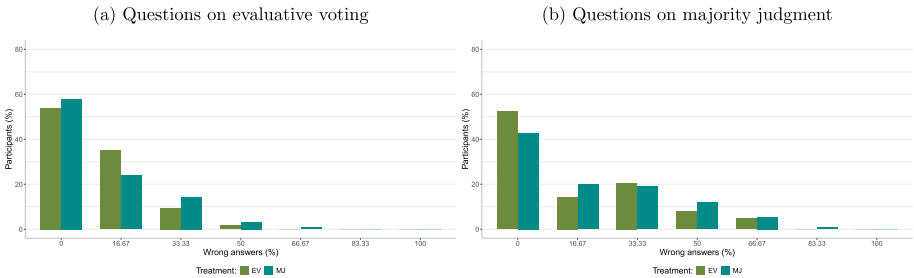
---

<sup>21</sup> As anecdotal evidence, we coded terms related to voter understanding (easy, difficult, complex, etc.) in an open-ended question concerning reasons on why they dislike a certain voting rule. Out of 189 respondents, 51 (26.98%) mentioned aspects related to the voting rule's complexity. While difficulties with evaluative voting mostly arose because of needing to weigh each option with numbers, majority judgment's complexity was addressed at different levels (tie-breaking rule, using words to evaluate, median). Interestingly, some participants evoked (a) the broader public receptance if it were implemented, suggesting that they were aware of the increased effort it took them to understand it and (b) its counterintuitiveness, suggesting that voters would start thinking more about strategies and over analyze their choices instead of being sincere. The latter concern points at the possibility that some voting rules are more intuitive than others independently of their expression power (Marsilio 2023), given that both majority judgment and evaluative voting include six labels.

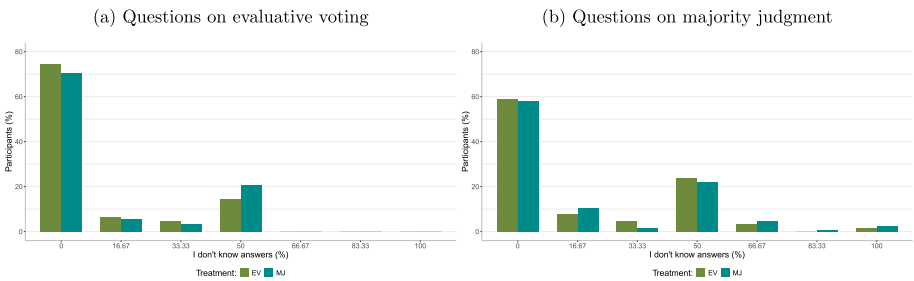
## Appendix

### Figures on understanding

See Figs. 7 and 8



**Fig. 7** Distribution of share of wrong answers. **Note:** The treatment relates to the rule participants voted on before answering the comprehension questions in Part II



**Fig. 8** Distribution of share of "I don't know" answers. **Note:** The treatment relates to the rule participants voted on before answering the comprehension questions in Part II

## Summary statistics

See Table 6

**Table 6** Summary statistics by treatment group

|                                     | Treatment<br>EV (N=63) | Treatment<br>JM-G<br>(N=63) | Treat-<br>ment<br>JM-I<br>(N=63) |
|-------------------------------------|------------------------|-----------------------------|----------------------------------|
| <i>Gender</i>                       |                        |                             |                                  |
| Male                                | 0.46                   | 0.49                        | 0.46                             |
| Female                              | 0.54                   | 0.51                        | 0.54                             |
| <i>Age</i>                          |                        |                             |                                  |
| [18-24]                             | 0.94                   | 0.83                        | 0.87                             |
| [24+]                               | 0.063                  | 0.17                        | 0.13                             |
| <i>Occupation</i>                   |                        |                             |                                  |
| Student                             | 0.95                   | 0.86                        | 0.92                             |
| Other                               | 0.048                  | 0.14                        | 0.079                            |
| <i>Field of study</i>               |                        |                             |                                  |
| Economics and business              | 0.47                   | 0.51                        | 0.33                             |
| Engineering                         | 0.35                   | 0.22                        | 0.33                             |
| Other                               | 0.18                   | 0.27                        | 0.33                             |
| <i>Higher education institution</i> |                        |                             |                                  |
| Ecole management Lyon               | 0.48                   | 0.71                        | 0.45                             |
| Ecole centrale                      | 0.37                   | 0.16                        | 0.28                             |
| Other                               | 0.15                   | 0.13                        | 0.27                             |
| <i>Agglomeration size</i>           |                        |                             |                                  |
| Rural municipality                  | 0.17                   | 0.063                       | 0.17                             |
| 2,000-19,999 inhabitants            | 0.14                   | 0.19                        | 0.21                             |
| 20,000-99,999 inhabitants           | 0.21                   | 0.13                        | 0.16                             |
| More than 100,000 inhabitants       | 0.30                   | 0.37                        | 0.32                             |
| Paris agglomeration                 | 0.17                   | 0.25                        | 0.14                             |

## Answers to True/False questions

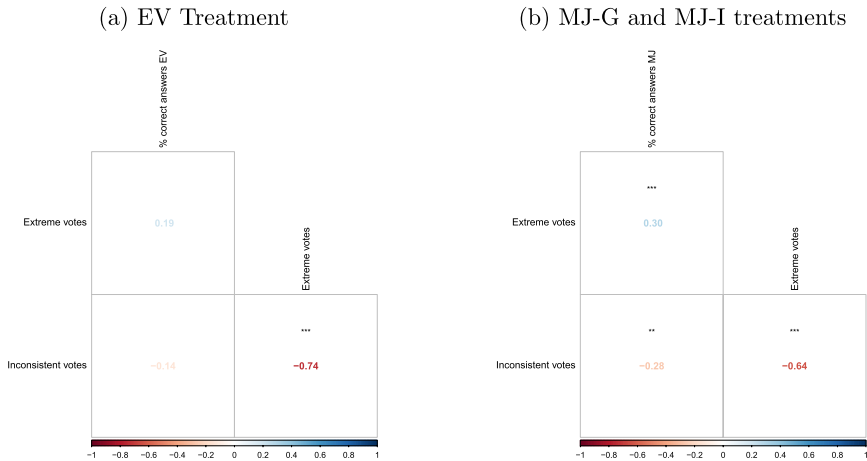
See Table 7.

**Table 7** True/false questions on preferences aggregation

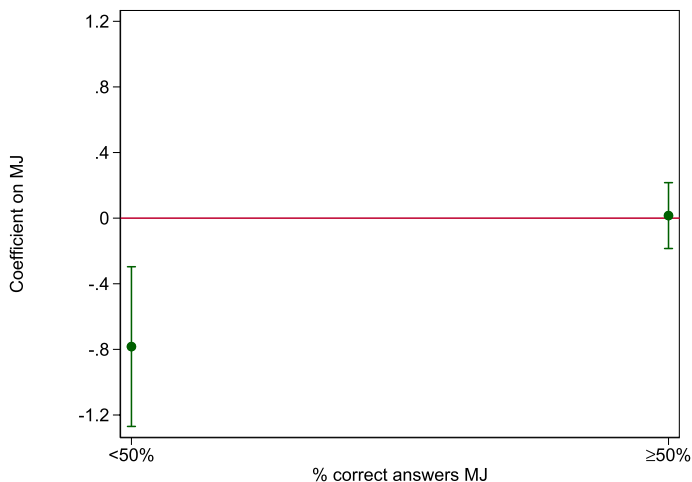
| Questions on evaluative voting                                                                                                                                | Correct answer | Correct | Wrong  | Did not know |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|---------|--------|--------------|
| The grade corresponds to the average score given to this candidate by all voters                                                                              | True           | 47.62%  | 32.28% | 20.11%       |
| The grade corresponds to the average rating given to this candidate by half plus one of the voters who are most in favor of him/her.                          | False          | 70.90%  | 5.29%  | 23.81%       |
| The grade corresponds to the score given by the voter who divides in two the set of voters ranked from the most favorable to the most unfavorable.            | False          | 67.72%  | 10.05% | 22.22%       |
| Questions on majority judgment                                                                                                                                | Correct answer | Correct | Wrong  | Did not know |
| The majority grade corresponds to the average rating given to this candidate by all voters                                                                    | False          | 51.85%  | 18.52% | 29.63%       |
| The majority grade corresponds to the average rating given to this candidate by half plus one of the voters most in favor of him/her                          | False          | 53.97%  | 14.29% | 31.75%       |
| The majority grade corresponds to the judgment given by the voter who divides in two the set of voters ranked from the most favorable to the most unfavorable | True           | 43.39%  | 25.40% | 31.22%       |

## Correlation among components of understanding

See Figs. 9 and 10



**Fig. 9** Correlation matrix - Components of understanding. **Note:** The variable “Inconsistent choices” represents the number of inconsistent choices in the three last elections of every series. “Extreme votes” represents the sum of the extreme votes given by each participant in the three last elections of every series. % correct answers EV (MJ) is the share of correct answers to the questions on how votes are aggregate in evaluative voting (majority judgment). \* significant at 10%, \*\* significant at 5%, \*\*\* significant at 1%



**Fig. 10** Differences in extreme votes between EV and MJ by vote aggregation understanding. **Note:** Marginal effect of the MJ treatment on the number of times a participant assigns an extreme grade in each election. The effect is computed for participants responding to less and more than half of questions on vote aggregation under majority judgment correctly. Standard errors are clustered at the individual and session level

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11127-025-01344-8>.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alekseev, Aleksandr, Gary Charness, and Uri Gneezy. 2017. Experimental Methods: When and Why Contextual Instructions are Important. *Journal of Economic Behavior & Organization* 134:48–59. <https://doi.org/10.1016/j.jebo.2016.12.005>.
- Andreoni, James. 1995. Cooperation in Public-Goods Experiments: Kindness or Confusion? *The American Economic Review* 85 (4): 891–904.
- Ankeny, Rachel, Hasok Chang, Marcel Boumans, and Mieke Boon. 2011. Introduction: Philosophy of Science in Practice. *European Journal for Philosophy of Science* 1 (3): 303. <https://doi.org/10.1007/s13194-011-0036-4>.
- Aziz, Haris, Serge Gaspers, Joachim Gudmundsson, Simon Mackenzie, Nicholas Mattei, and Toby Walsh. 2015. Computational Aspects of Multi-Winner Approval Voting. *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015)*.
- Balinski, Michel, and Rida Laraki. 2010. Election by Majority Judgment: Experimental Evidence. In *In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections*, edited by Bernard Dolez, Bernard Grofman, and Annie Laurent, 13–54. New York, NY: Springer. [https://doi.org/10.1007/978-1-4419-7539-3\\_2](https://doi.org/10.1007/978-1-4419-7539-3_2).
- Balinski, Michel, and Rida Laraki. 2007. A Theory of Measuring, Electing, and Ranking. *Proceedings of the National Academy of Sciences* 104 (21): 8720–8725. <https://doi.org/10.1073/pnas.0702634104>.
- Balinski, Michel, and Rida Laraki. 2012. Jugement majoritaire versus vote majoritaire. (via les présidentielles 2011-2012). *Revue française d'économie* 27 (4): 11–44. <https://doi.org/10.3917/rfe.124.0011>.
- Balinski, Michel, and Rida Laraki. 2020. Majority Judgment vs. Majority Rule. *Social Choice and Welfare* 54 (2): 429–461. <https://doi.org/10.1007/s00355-019-01200-x>.
- Baujard, Antoinette, Frédéric Gavrel, Herrade Igersheim, Jean-François Laslier, and Isabelle Lebon. 2012. Compte-rendu de l'expérimentation des votes par approbation et par note lors des élections présidentielles françaises le 22 avril 2012 à Saint-Etienne, Strasbourg et Louvigny. *HAL working paper - halshs-00713659*.
- Baujard, Antoinette. 2022. Ethics and Technique in Welfare Economics: How Welfarism Evolves in the Making. *Revue économique* 73 (6): 1039–1053. <https://doi.org/10.3917/reco.736.1039>.
- Baujard, Antoinette, and Herrade Igersheim. 2009. Expérimentation du vote par note et du vote par approbation le 22 avril 2007. *Premiers résultats. Revue économique* 60 (1): 189–201. <https://doi.org/10.3917/reco.601.0189>.
- Baujard, Antoinette, and Isabelle Lebon. 2022. Not-so-strategic voters. *Electoral Studies* 76 : 102458. <https://doi.org/10.1016/j.electstud.2022.102458>.
- Baujard, Antoinette, Herrade Igersheim, Isabelle Lebon, Frédéric Gavrel, and Jean-François Laslier. 2014. Who's favored by evaluative voting? An experiment conducted during the 2012 French presidential election. *Electoral Studies* 34:131–145. <https://doi.org/10.1016/j.electstud.2013.11.003>.
- Baujard, Antoinette, Frédéric Gavrel, Herrade Igersheim, Jean-François Laslier, and Isabelle Lebon. 2018. How voters use grade scales in evaluative voting. *European Journal of Political Economy* 55:14–28. <https://doi.org/10.1016/j.ejpoleco.2017.09.006>.
- Baujard, Antoinette, Herrade Igersheim, and Isabelle Lebon. 2021. Some regrettable grading scale effects under different versions of evaluative voting. *Social Choice and Welfare* 56 (4): 803–834. <https://doi.org/10.1007/s00355-020-01300-z>.

- Bigoni, M., and D. Dragone. 2012. Effective and efficient experimental instructions. *Economics Letters* 117 (2): 460–463. <https://doi.org/10.1016/j.econlet.2012.06.049>.
- Blais, André, Simon Labbé St-Vincent, Jean-Benoit Pilet, and Rafael Treibich. 2016. Voting correctly in lab elections with monetary incentives: The impact of district magnitude. *Party Politics* 22 (4): 544–551. <https://doi.org/10.1177/1354068814560933>.
- Blais, André, Carolina Plescia, and Semra Sevi. 2021. Do people want smarter ballots? *Research & Politics* 8:4. <https://doi.org/10.1177/20531680211062668>.
- Bourgeois-Gironde, Sacha, and João. V. Ferreira. 2023. The expressive power of voting rules. *Social Choice and Welfare* 62:233–273. <https://doi.org/10.1007/s00355-023-01486-y>.
- Bowler, Shaun, Gail McElroy, and Stefan Müller. 2018. Voter preferences and party loyalty under cumulative voting: Political behaviour after electoral reform in Bremen and Hamburg. *Electoral Studies* 51:93–102. <https://doi.org/10.1016/j.electstud.2017.10.001>.
- Brams, Steven J., and Peter C. Fishburn. 2002. Chapter 4 Voting procedures. In *Handbook of Social Choice and Welfare*, 1:173–236. Elsevier.
- Cormack, Lindsey. 2023. More Choices, More Problems? Ranked Choice Voting Errors in New York City. *American Politics Research* 52 (3): 306–319. <https://doi.org/10.1177/1532673X231220640>.
- Darmann, Andreas, and Christian Klamler. 2023. Does the rule matter? A comparison of preference elicitation methods and voting rules based on data from an Austrian regional parliamentary election in 2019. *Public Choice* 197 (1): 63–87. <https://doi.org/10.1007/s11127-023-01071-y>.
- Darmann, Andreas, Julia Grundner, and Christian Klamler. 2017. Election outcomes under different ways to announce preferences: an analysis of the 2015 parliament election in the Austrian federal state of Styria. *Public Choice* 173 (1): 201–216. <https://doi.org/10.1007/s11127-017-0472-6>.
- Darmann, Andreas, Julia Grundner, and Christian Klamler. 2019. Evaluative voting or classical voting rules: Does it make a difference? Empirical evidence for consensus among voting rules. *European Journal of Political Economy* 59:345–353. <https://doi.org/10.1016/j.ejpoleco.2019.04.003>.
- Delemazure, Théo, Roberto Brunetti, Antoinette Baujard, and Sylvain Bouveret. 2025. Do Grades Have Absolute Meaning? An Experiment on Majority Judgment. *HAL working paper - hal-05114129*.
- Dellis, Arnaud, and Sabine Kröger. 2023. How Do People Vote Under Instant Runoff Voting? An Experiment on Complexity and Voting Behavior. *Working paper*. <https://doi.org/10.2139/ssrn.4602773>.
- Der Straeten, Van, Jean-François Laslier, Karine, Nicolas Sauter, and André Blais. 2010. Strategic, sincere, and heuristic voting under four election rules: an experimental study. *Social Choice and Welfare* 35 (3): 435–472. <https://doi.org/10.1007/s00355-010-0448-7>.
- Donovan, Todd, Caroline Tolbert, and Kellen Gracey. 2019. Self-Reported Understanding of Ranked-Choice Voting. *Social Science Quarterly* 100 (5): 1768–1776. <https://doi.org/10.1111/ssqu.12651>.
- Donovan, Todd, Caroline Tolbert, and Samuel Harper. 2022. Demographic differences in understanding and utilization of ranked choice voting. *Social Science Quarterly* 103 (7): 1539–1550. <https://doi.org/10.1111/ssqu.13215>.
- Eggers, Andrew C., and Nick Vivyan. 2020. Who Votes More Strategically? *American Political Science Review* 114 (2): 470–485. <https://doi.org/10.1017/S0003055419000820>.
- Fabre, Adrien. 2021. Tie-breaking the Highest Median: Alternatives to the Majority Judgment. *Social Choice and Welfare* 56 (1): 101–124. <https://doi.org/10.1007/s00355-020-01269-9>.
- Felsenthal, Dan S., and Moshe Machover. 2008. The Majority Judgement Voting Procedure: A Critical Evaluation. *Homo oeconomicus* 25 (3/4): 319–334.
- Freeman, David J., Erik O. Kimbrough, Garrett M. Petersen, and Hanh T. Tong. 2018. Instructions. *Journal of the Economic Science Association* 4 (2): 165–179. <https://doi.org/10.1007/s40881-018-0059-0>.
- Herzberg, Roberta Q., and Rick K. Wilson. 1988. Results on Sophisticated Voting in an Experimental Setting. *The Journal of Politics* 50 (2): 471–486. <https://doi.org/10.2307/2131804>.
- Hillinger, Claude. 2006. The Case for Utilitarian Voting. *Homo Oeconomicus* 22:3.
- Houser, Daniel, and Robert Kurzban. 2002. Revisiting Kindness and Confusion in Public Goods Experiments. *American Economic Review* 92 (4): 1062–1069. <https://doi.org/10.1257/00028280260344605>.
- Igersheim, Herrade, Antoinette Baujard, Frédéric Gavrel, Jean-François Laslier, and Isabelle Lebon. 2016. Individual Behavior Under Evaluative Voting: A Comparison Between Laboratory and In Situ Experiments. In *Voting Experiments*, edited by André Blais, Jean-François Laslier, and Karine Van Der Straeten, 257–269. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-40573-5\\_13](https://doi.org/10.1007/978-3-319-40573-5_13).
- Igersheim, Herrade, Antoinette Baujard, and Jean-François. Laslier. 2017. La question du vote. Expérimentations en laboratoire et In Situ. *L'Actualité économique* 92 (1): 151–189. <https://doi.org/10.7202/1039875ar>.
- Igersheim, Herrade, François Durand, Aaron Hamlin, and Jean-François. Laslier. 2022. Comparing voting methods: 2016 US presidential election. *European Journal of Political Economy* 71 : 102057. <https://doi.org/10.1016/j.ejpoleco.2021.102057>.

- Kimball, David C., and Martha Kropf. 2016. Voter Competence with Cumulative Voting\*. *Social Science Quarterly* 97 (3): 619–635. <https://doi.org/10.1111/ssqu.12277>.
- Koppel, Lina, David Andersson, Magnus Johannesson, Eirik Strömland, and Gustav Tinghög. 2025. Comprehension in economic games. *Journal of Economic Behavior & Organization* 234 : 107039. <https://doi.org/10.1016/j.jebo.2025.107039>.
- Lachat, Romain, and Jean-François. Laslier. 2024. Alternatives to plurality rule for single-winner elections: When do they make a difference? *European Journal of Political Economy* 81 : 102505. <https://doi.org/10.1016/j.ejpoleco.2024.102505>.
- Laslier, Jean-François. 2011. Lessons from In Situ Experiments during French Elections. In *In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections*, ed. Bernard Dolez, Bernard Grofman, and Annie Laurent, 91–104. New York, NY: Springer.
- Laslier, Jean-François. 2019. The strange “Majority Judgment”. *Revue économique* 70 (4): 569–588.
- Laslier, Jean-François., André Blais, Damien Bol, Sona N. Golder, Philipp Harfst, Laura B. Stephenson, and Karine Van der Straeten. 2015. The EuroVotePlus experiment. *European Union Politics* 16 (4): 601–615. <https://doi.org/10.1177/1465116515580180>.
- Lau, Richard R., and David P. Redlawsk. 1997. Voting Correctly. *American Political Science Review* 91 (3): 585–598. <https://doi.org/10.2307/2952076>.
- Macé, Antonin. 2018. Voting with evaluations: Characterizations of evaluative voting and range voting. *Journal of Mathematical Economics* 79:10–17. <https://doi.org/10.1016/j.jmateco.2018.09.002>.
- Marino, Maria, Roberto Iacono, and Johanna Mollerstrom. 2024. (Mis-)Perceptions, information, and political polarization: A survey and a systematic literature review. *European Journal of Political Economy* 85 : 102578. <https://doi.org/10.1016/j.ejpoleco.2024.102578>.
- Marsilio, Simone. 2023. Nudge the voting method: Could an alternative voting design fit the nudging definition? *Nomopolis* 1 (1): 92–120. <https://doi.org/10.3917/nomop.001.0092>.
- Núñez, Matías, and Jean François Laslier. 2014. Preference intensity representation: strategic overstating in large elections. *Social Choice and Welfare* 42 (2): 313–340. <https://doi.org/10.1007/s00355-013-0728-0>.
- Oprea, Ryan. 2020. What Makes a Rule Complex? *American Economic Review* 110 (12): 3913–3951. <https://doi.org/10.1257/aer.20191717>.
- Pulickal, Anuvinda, and Sujoy Chakravarty. 2023. Subject confusion and task non-completion: Methodological insights from an artefactual field experiment with adolescents in India. *Journal of Behavioral and Experimental Economics* 103 : 101986. <https://doi.org/10.1016/j.soceec.2023.101986>.
- Ramalingam, Abhijit, Antonio J. Morales, and James M. Walker. 2018. Varying experimental instructions to improve comprehension: Punishment in public goods games. *Journal of Behavioral and Experimental Economics* 73:66–73. <https://doi.org/10.1016/j.soceec.2018.01.008>.
- Regenwetter, Michel, ed. 2006. *Behavioral social choice: probabilistic models, statistical inference, and applications*. Cambridge; NY: Cambridge University Press.
- Regenwetter, Michel, and Ilia Tsetlin. 2004. Approval voting and positional voting methods: Inference, relationship, examples. *Social Choice and Welfare* 22 (3): 539–566. <https://doi.org/10.1007/s00355-003-0232-z>.
- Regenwetter, Michel, A. A. J. Marley, and Bernard Grofman. 2003. General concepts of value restriction and preference majority. *Social Choice and Welfare* 21 (1): 149–173. <https://doi.org/10.1007/s00355-003-0210-5>.
- Regenwetter, Michel, Bernard Grofman, Anna Popova, William Messner, Clinton P. Davis-Stober, and Daniel R. Cavagnaro. 2008. Behavioural social choice: a status report. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1518): 833–843. <https://doi.org/10.1098/rstb.2008.0259>.
- Rozenblit, Leonid, and Frank Keil. 2002. The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science* 26 (5): 521–562. [https://doi.org/10.1207/s15516709cog2605\\_1](https://doi.org/10.1207/s15516709cog2605_1).
- Sen, Amartya. 1995. How to Judge Voting Schemes. *The Journal of Economic Perspectives* 9 (1): 91–98.
- Slooman, Steven, and Philip Fernbach. 2017. *The Knowledge Illusion: Why We Never Think Alone*. New York, NY: Riverhead Books.
- Smaoui, Hatem, and Dominique Lepelley. 2013. Le système de vote par note à trois niveaux: étude d’un nouveau mode de scrutin. *Revue d’économie politique* 123 (6): 827–850. <https://doi.org/10.3917/redp.236.0827>.
- Zizzo, Daniel John. 2010. Experimenter demand effects in economic experiments. *Experimental Economics* 13 (1): 75–98. <https://doi.org/10.1007/s10683-009-9230-z>.

## Authors and Affiliations

**Antoinette Baujard<sup>1</sup> · Roberto Brunetti<sup>2</sup>  · Isabelle Lebon<sup>3</sup> · Simone Marsilio<sup>4</sup> **

✉ Simone Marsilio  
s.marsilio@ipw.uni-hannover.de

Roberto Brunetti  
roberto.brunetti@assas-universite.fr

<sup>1</sup> Université Jean Monnet, Université Lyon 2, emlyon, GATE, CNRS, 42100 Saint-Etienne, France

<sup>2</sup> LEMMA Université Paris Panthéon-Assas, Paris, France

<sup>3</sup> CREM-CNRS, University Caen Normandy, UNICAEN, Caen, France

<sup>4</sup> Leibniz University Hannover, Hannover, Germany