Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

Dichotomous Preferences: Concepts, Measurement, and Evidence

Salvatore Barbaro and Anna-Sophie Kurella

2nd Oktober 2025

Discussion paper number 2506

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

Prof. Dr. Salvatore Barbaro
Chair of Social Choice
Johannes-Gutenberg University
Jakob-Welder-Weg 4
55128 Mainz
Germany
sbarbaro@uni-mainz.de

Anna-Sophie Kurella
Institut für Politikwissenschaft
Leibniz Universität Hannover
Schneiderberg 50
30167 Hannover
Germany
a.kurella@ipw-uni-hannover.de

Dichotomous Preferences: Concepts, Measurement, and Evidence

SCWE-D-25-00070

Salvatore Barbaro[∓] and Anna-Sophie Kurella[*⊗]

[∓] *Johannes-Gutenberg University, Interdisciplinary Public Policy , 55122 Mainz, Germany sbarbaro@uni-mainz.de*
[⊗] *Leibniz University, Institute of Political Science, 30167 Hannover, Germany a.kurella@ipw-uni-hannover.de*

Date of resubmission: 2nd October 2025

Dichotomous preferences are a widely assumed feature in social choice theory. Despite their prominence in theoretical models, the empirical validity of this assumption has remained largely unexplored. Nor is it always clear how dichotomous preferences are defined across different research contexts. This paper introduces two new concepts that weaken the strict dichotomy assumption and can each be tested empirically. Using CSES data and three experimental datasets—two from French presidential elections and one from a regional election in Austria—we examine how frequently the different forms of dichotomous preferences occur. In addition, the paper provides evidence on the relationship between ranking and approval ballots. The results suggest that while dichotomous preferences do not offer a perfect representation of voter preferences, they constitute an acceptable approximation, particularly among voters who approve more than two alternatives and among respondents with higher educational attainment levels.

Keywords: Preferences, Dichotomous Preferences, Inequality Measures, Cluster Analysis
JEL-Code: D71

1 Introduction

Dichotomous preferences are a widely used assumption in the study of preference aggregation. Under this assumption, elections can be simplified while retaining several normatively desirable properties (Vorsatz, 2007, 2008; Ju, 2010; Maniquet and Mongin, 2015; Sato, 2019; Brandl and Brandt, 2020; Brandl and Peters, 2022; Komatsu, 2024). Approval Voting (AV) is a prime example of a simple and effective method. Dichotomous preferences also support normatively appealing designs of parliamentary and committee representation (Brill et al., 2018, 2022; Skowron et al., 2016; Brams et al., 2018), feature in the analysis of randomisation mechanisms (Bogomolnaia et al., 2005), and are widely applied in computational social choice to ensure tractability (Elkind and Lackner, 2016).

Many of the advantages of single- and multi-winner approval rules can arise even without dichotomous preferences. Several experimental studies that do not explicitly test for dichotomous structures in preferences (Alós-Ferrer and Granić, 2012; Gehrlein and Lepelley, 1998; Gehrlein et al., 2016; Laslier and Van der Straeten, 2008; Alós-Ferrer et al., 2025) shed light on a general well-functioning of Approval Voting. However, in theoretical contributions, a number of desirable properties no longer hold when the dichotomous preference assumption is dropped (Brams and Fishburn, 2007, Theorems. 2.3, 2.4).

Given the extensive theoretical and experimental research—and the early expressed view that dichotomous preferences are a "critical" assumption (Niemi, 1984; Saari and Van Newenhizen, 1988)—it is striking that, to our knowledge, no study has systematically examined the empirical validity of this widely used assumption.

Preferences are considered dichotomous when voters divide the available alternatives (e.g. candidates or parties) into two groups (Regenwetter and Grofman, 1998; Dehez and Ginsburgh, 2019), evaluating alternatives within the same group identically while differentiating between groups (Brams and Fishburn, 2007, Def. 2.1).[1] In this strict form, dichotomous preferences are unlikely to be observed empirically. A more flexible and widely cited assumption is that voters can transform a ranking into an approval ballot by choosing a threshold alternative, assigning all candidates above the threshold to the approved set and all below to the disapproved set (Dehez and Ginsburgh, 2019; Terzopoulou et al., 2025; Brams and Fishburn, 2007, Ch. 2). This assumption is frequently referred to as the *threshold approach.* The same concept is also known as *size-independent model of approval voting* (Falmagne and Regenwetter, 1996; Regenwetter et al., 2002, Sec. 3). Below, we will investigate how often this requirement is met in empirical data.

At the core of this paper we introduce two new concepts, both representing weakenings of dichotomous preferences and both amenable to empirical testing. The first, which we term *weakly dichotomous preferences* (WDP), arises when a voter divides the candidates into two groups and

*The authors are indebted to Herrade Iggersheim for valuable comments and Christian Klamler for providing the data from the *Graz* experiment.

[1] To the best of our knowledge, the term *dichotomous preferences* originates from Inada (1969). A few years earlier, however, Inada (1964) had already described such preferences, though without explicitly coining the term, instead referring to "indifferent groups of alternatives".

the dispersion of ratings within groups is smaller than the dispersion between groups. This follows a common approach in the income-inequality research, where overall dispersion is decomposed into *within*- and *between*-group components. A limitation of this concept is that its empirical assessment requires complex data: for each voter, information is needed both on how they rate the alternatives (ranking ballots[2]) and on which alternatives they approve or disapprove (approval ballots). In the empirical part of our study we rely on three datasets that contain both types of ballots, collected in different surveys and described in detail in Section 3.

Our second weakening of dichotomous preferences relies exclusively on ranking ballots. We refer to this concept as *quasi-dichotomous preferences* (QDP). Preferences are classified as QDP when clustering the rating data (through cluster-analytical methods) yields an optimal number of two clusters. In this case, representing preferences by two groups provides a superior description compared to three or more groups. We present the analytical methods used to identify QDP and apply them not only to the three just-mentioned datasets but also to a large comparative dataset from CSES (2024), which allows us to analyse ranking ballots from over 200,000 respondents across 172 elections.

We find that both weakenings of dichotomous preferences (WDP and QDP) provide good, though not perfect, representations of individual preferences. WDP can be identified for around 50–70% of respondents. These number increase when restricting attention to respondents that approve more than 2 alternatives. With respect to quasi-dichotomous preferences, we find that clustering into two groups offers by far the best approximation for 50–65% of respondents.

Another line of research closely related to preference dichotomy concerns the nature of approval preferences, ranking ballots, and their comparison (Terzopoulou et al., 2025). One question in this context is which alternatives can become approval winners when (available) ranking ballots are converted into (non-available) approval ballots (Regenwetter and Grofman, 1998). Our cluster-analytical approach, used to assess QDP, is closely related to this research string. Conversely, when approval ballots are available, the second question is which alternatives can become winners under a ranking-based rule once they are converted into ranking ballots.

On the latter question, Terzopoulou et al. (2025) have recently provided valuable insights in this journal, while at the same time leaving the further treatment of the "foundational question" (Terzopoulou et al., 2025) on how approval and ranking ballots compare to future research. Since we have access to three datasets containing both types of ballots, we are able to contribute empirically to this debate. Specifically, we address the following two questions:

1. How often do respondents rate at least one disapproved alternative as highly as, or higher than, an approved alternative?
2. How often does the optimal cluster allocation coincide with the approval data? We refer to this number as the *matching value*.

The first question immediately relates to the 'threshold approach'. Our data show that, depending on the dataset, between 10 and 30% of respondents rate a disapproved candidate at least as highly as an approved candidate. The figures are considerably lower, however, when only strictly higher ratings are considered (i.e. excluding ties). Regarding the second question, we find an average matching value of around 80%. Matching values are particularly pronounced when respondents evaluate and classify candidates from the extreme right as approved or disapproved.

In addition to its conceptual and methodological contributions, this paper addresses the question of how valid it is to assume dichotomous preferences. Taken together, our analyses suggest that dichotomous preferences provide a generally good approximation of voter preferences. They can certainly be defended more convincingly than trichotomous, fourfold, or multichotomous structures. At the same time, we find that for roughly 30 to 40% of respondents, the assumption of dichotomous preferences—even in the weaker versions we propose—does not provide an adequate representation of their preferences.

2 Dichotomous preferences: conceptions and measurement

We begin by formally defining dichotomous preferences. Definition 1 follows the formulation of Brams and Fishburn (2007, Def. 2.1). Let $\mathscr{B} = \{x_1, x_2, \ldots, x_m\}$ denote a finite set of m alternatives.

Definition 1 (Dichotomous preferences).
A preference relation $\succsim_i$ of voter i over $\mathscr{B}$ is called dichotomous *if there exists a partition of $\mathscr{B}$ into two disjoint sets A_i (approved) and D_i (disapproved) such that $A_i \cup D_i = \mathscr{B}$, $A_i \cap D_i = \varnothing$, and $\forall x \in A_i$, $\forall y \in D_i : x \succ_i y$, while $\forall x, y \in A_i : x \sim_i y$ and $\forall x, y \in D_i : x \sim_i y$.*

Criticism of the dichotomous preference assumption dates back to the 1980s. Niemi (1984) demonstrated that Approval Voting loses several normatively attractive properties with only minor deviations from this assumption. Saari and Van Newenhizen (1988) reinforced and extended this argument. In response, Brams et al. (1988) rejected their claim that all analyses of Approval Voting depend on dichotomous preferences, pointing out that much of the literature "goes well beyond the highly specialised dichotomous case" (Brams et al., 1988, Sec. 2.1). Indeed, it has often been assumed that voters can translate their preference orderings into a "compatible approval set" (Brams and Fishburn, 2007, Ch. 2) by designating a particular alternative as a threshold ('threshold approach'). Brams et al. (1988) further argued that this assumption imposes weaker requirements on voters' preference structures, since it does not presuppose the existence of a complete preference ordering.

[2] By "ranking ballots" we mean data on respondents' preference orderings. In most datasets, however, respondents are not asked to rank but to rate the available alternatives. A large empirical literature (e.g. Lachat and Laslier (2024)) converts such rating data into preference orders. For example, if respondent i assigns $+4$ to party A, $+1$ to B, and $+2$ to C, this yields the ordering $A \succ_i C, C \succ_i B$ (and, by transitivity, $A \succ_i B$). Thus the underlying data are typically ratings, but because of the one-to-one mapping into orderings we refer to as "ranking ballots."

Let $r_i : \mathscr{B} \to \mathbb{R}$ denote a *rating function* that represents the preferences $\succsim_i$ of voter i over $\mathscr{B}$. For any $x, y \in \mathscr{B}$, we have $x \succsim_i y$ if and only if $r_i(x) \geq r_i(y)$.

Definition 2 (Threshold approach). *A preference relation $\succsim_i$ of voter i over $\mathscr{B}$ is called* threshold-dichotomous *if there exists a partition of $\mathscr{B}$ into two disjoint sets A_i and D_i such that $A_i \cup D_i = \mathscr{B}$, $A_i \cap D_i = \varnothing$, and $\min_{x \in A_i} r_i(x) \geq \max_{y \in D_i} r_i(y)$.*

For some empirical applications, it will be necessary to adopt a variant of the threshold approach in which the highest-rated disapproved candidate is strictly lower rated than the lowest-rated approved candidate. In empirical data, indifference relations and focal points around particular values are common, and such clustering can cause the strict inequality to yield substantially different outcomes from the weak inequality.

We raise two objections to the threshold approach. The first concerns its focus: the threshold approach reflects the consistency of a voter's decision rather than the existence of genuinely dichotomous structures. Consider a voter who evaluates six candidates $(x_1, \ldots, x_6)$ on a $[0, 1]$ scale: $r(x_1) = 0.1$, $r(x_2) = 0.2$, $r(x_3) = 0.4$, $r(x_4) = 0.6$, $r(x_5) = 0.8$, $r(x_6) = 0.9$. To classify this voter as having threshold-dichotomous preferences, it is sufficient that they place the threshold at some point along the scale and include all candidates with ratings above this threshold in the approval set. For example, the threshold could be set between x_5 and x_6. This satisfies the definition of threshold-dichotomous preferences, but the example makes clear that what is being captured is a form of internal consistency in the approval decision, not dichotomous structures as such. The disapproved set contains too many internal dissimilarities for the ratings to be considered a homogeneous group.

The second, and main objection is that the threshold approach can also classify trichotomous or multichotomous preferences as threshold-dichotomous. Trichotomous (multichotomous) preferences are defined analogously to Definition 1, except that there are three (more than three) groups in which the voter is indifferent among alternatives (Manjunath and Westkamp, 2021; Brams and Fishburn, 2007, Def. 2.1). This objection can likewise be illustrated easily. Suppose the voter's evaluations are $(r(x_1) = r(x_2) = 0.1, r(x_3) = r(x_4) = 0.5, r(x_5) = r(x_6) = 0.8)$. If the threshold is set at 0.2, threshold-dichotomous preferences would be assigned, yet the underlying structure is clearly trichotomous.

In light of these objections, it appears warranted to introduce two alternative definitions of dichotomous preference structures: *weakly dichotomous preferences* and *quasi-dichotomous preferences.*

Weakly dichotomous preferences are defined as preferences in which the dissimilarities of ratings within each group (approved and disapproved alternatives) is lower than the dispersion between the two groups.

Let $\delta : \mathscr{B} \times \mathscr{B} \to \mathbb{R}_{\geq 0}$ denote a dispersion function that for all $x, y \in \mathscr{B}$ is non-negative ($\delta(x, y) \geq 0$), satisfies identity ($\delta(x, x) = 0$), and is symmetric ($\delta(x, y) = \delta(y, x)$).

Definition 3 (Weakly dichotomous preferences). *A preference relation $\succsim_i$ of voter i over $\mathscr{B}$ is called* weakly dichotomous *if there exists a partition of $\mathscr{B}$ into two disjoint sets A_i and D_i such that $A_i \cup D_i = \mathscr{B}$, $A_i \cap D_i = \varnothing$, and the overall dispersion of preferences can be decomposed into a within-subsets component δ_w and a between component δ_b with $\delta_w < \delta_b$.*

To operationalise the concept of dispersion, one can choose from a wide range of decomposable inequality measures (Cowell and Victoria-Feser, 1996). In this context it is worth noting that inequality comparisons are sensitive to the choice of the inequality measures such that replacing one measure will almost always change the relative significance of the within and between components (Shorrocks, 1980). In our empirical part we mainly rely on the Gini coefficient as a widely used measure of inequality and dispersion in social sciences (Fleurbaey et al., 2025). However, we test the robustness of our results by utilising three other voting rules.

Let still $r_i : \mathscr{B} \to \mathbb{R}$ denote the rating function of voter i over the $m = |\mathscr{B}|$ alternatives, with mean rating $\mu_i = \frac{1}{m} \sum_{x \in \mathscr{B}} r_i(x)$. The Gini coefficient of voter i over $\mathscr{B}$ is defined as $\mathcal{G}_i = \frac{1}{2m^2 \mu_i} \sum_{x,y \in \mathscr{B}} \delta(x, y)$, where the dispersion function in this particular case is given by $\delta(x, y) = |r_i(x) - r_i(y)|$.

Given a partition of $\mathscr{B}$ into the groups A_i and D_i, the Gini coefficient can be decomposed into a within-group component and a between-group component:[3]

$$\mathcal{G}_i = \mathcal{G}_{w,i} + \mathcal{G}_{b,i} + \omega_i \tag{1}$$

Here,

$$\mathcal{G}_{w,i} = \frac{1}{2m^2 \mu_i} \left(\sum_{x,y \in A_i} \delta(x, y) + \sum_{x,y \in D_i} \delta(x, y) \right) \tag{2}$$

measures the dispersion within the approved and within the disapproved sets, while

$$\mathcal{G}_{b,i} \;=\; \frac{1}{2m^2 \mu_i} \sum_{x \in A_i} \sum_{y \in D_i} \delta(x, y) \tag{3}$$

captures the dispersion between approved and disapproved alternatives.

The within-group inequality captures the weighted sum of the group rating dispersions. The between-component measures the dissimilarities between both groups by considering a smoothed distribution for each group.[4]

The Gini index is scale-invariant. This is advantageous for our analysis, as the rating values in the three datasets were collected on different scales, which we transformed into a common scale for comparability (see the next section). Its drawback is that the decomposition needs not to be additively decomposable (Shorrocks, 1980), meaning that the within- and between-group components do not necessarily sum to the overall Gini. The residual, known as the "overlapping index" (Yitzhaki and Lerman, 1991), may bias comparisons between within- and between-group

[3] Note that, unlike in income distribution research, we focus on each respondent's ratings, which allows us to assess whether they exhibit weakly dichotomous preferences.

[4] Note that Eq. (2) and Eq. (3) represent a general form of the Gini decomposition, for which different specifications exist in the literature (Bhattacharya and Mahalanobis, 1967; Lerman and Yitzhaki, 1989).

components (Yitzhaki, 1994). The overlap index is denoted by ω in Eq. (1).

The overlap index equals zero when the observations of each group are confined to distinct ranges and these ranges do not overlap (Yitzhaki, 1994; Costa, 2016). Applying this property to our setting reveals a connection between the Gini decomposition and the notion of threshold-dichotomous preferences:

Remark 1 (Threshold-dichotomous preferences and additive decomposability.)**.**
If a voter has threshold-dichotomous preferences, then the Gini decomposition of their preferences is perfectly additively decomposable into within- and between-group components.

As can be seen from Eq. (2) and Eq. (3), the analysis of weakly dichotomous preferences requires data both on respondents' ratings of the available alternatives and on their assignment of alternatives into two disjoint sets (approval ballots). While rating data are available in many datasets, data containing both ranking ballots and approval ballots are relatively rare. It is therefore useful to introduce a further concept that relies solely on the more widely available rating values. We refer to this concept as quasi-dichotomous preferences and define it as follows:

Definition 4 (Quasi-dichotomous preferences (QDP))**.**
A preference relation $\succsim_i$ of voter i over $\mathscr{B}$ is called quasi-dichotomous *if the optimal clustering of i's ratings r_i yields exactly two robust clusters.*

Beyond its less demanding data requirements, this approach directly addresses our second critique of the threshold approach. We test explicitly whether two groups provide a better representation of preferences than three or more groups.

QDP can be identified empirically using cluster-analytical methods. Cluster analysis is widely applied in the social sciences, particularly when observations are to be grouped—much like the mapping of ranking ballots onto approval ballots. As we explain in the empirical section (Section 6), grouping observations into two clusters may provide a better representation than grouping into three or more, while still being insufficiently robust. This is the aspect captured in the final part of Definition 4, which we also discuss in detail in Section 6.

Remark 2 (Interrelations between variants of dichotomous preferences)**.**

1. *If a voter has dichotomous preferences in the sense of Definition 1, then these preferences are also threshold-dichotomous, weakly dichotomous, and quasi-dichotomous. The converse implications, however, do not hold.*
2. *Threshold-dichotomous, weakly dichotomous, and cluster-analytical dichotomous preferences are mutually independent of one another.*

The first part of Remark 2 is straightforward: if the preference relation of voter i complies with Definition 1, then the dispersion within each group is zero and therefore smaller than the dispersion between A_i and D_i, so that WDP holds. In this case a compatible threshold necessarily exists, and if the ratings take only two distinct values, they can be best represented by exactly two clusters.

Table 1 illustrates the preferences of six voters over $m = 6$ candidates. The columns labelled "Ap" indicate whether a voter approves (Ap = 1) or disapproves (Ap = 0) of a candidate. For instance, Voter 1 approves candidates x_4, x_5, x_6. The columns labelled $r_i(j)$ display voter i's rating of candidate $j \in \mathscr{B}$ on a $[1, 2]$ scale. Voter 1 exhibits perfectly dichotomous preferences, as they assign equal ratings to all approved candidates and equal ratings to all non-approved candidates.

The preferences of Voter 2 resemble those of Voter 1. However, they do not comply with Definition 1, as their ratings vary within the two groups. Nevertheless, each approved candidate (x_4, x_5, x_6) is rated higher than any non-approved candidate.

For Voter 3, the rating values indicate that within-group dispersion is smaller than between-group dispersion. Applying the numbers to Eqs. (2) and (3) yields $\mathcal{G}_{w,3} = 0.017 < 0.046 = \mathcal{G}_{b,3}$, so we classify these preferences as weakly dichotomous. They also satisfy threshold consistency (Def. 2), since $\min_{j \in A_3} r_3(j) = 1.7 > \max_{j \in D_3} r_3(j) = 1.6$. Although the detailed procedure for calculating the optimal number of clusters is presented in Section 6, it suffices here to note that two clusters provide the best grouping of Voter 3's ratings. Thus, the preferences of Voter 3 satisfy all three properties defined in Definitions 2–4.

Instead of providing a formal proof for Statement 2 in Remark 2, we illustrate the mutual independence of the concepts using Voters 4 to 6. The ratings of Voter 4 are best grouped into two clusters, so that QDP is satisfied. However, the within-group dispersion exceeds the between-group dispersion (0.031 vs. 0.020), meaning that WDP does not hold. Moreover, the disapproved candidate x_2 is rated at 1.9, which is clearly higher than, for example, x_6, an approved candidate.

By contrast, Voter 5 satisfies WDP but not QDP. Voter 6 satisfies neither WDP nor QDP, but their preferences do meet threshold consistency.

3 Data

The first dataset originates from an in-situ experiment conducted alongside the 2017 French presidential election (Bouveret et al., 2019; Baujard et al., 2020). In various cities, survey participants indicated which of the eleven candidates they approved or disapproved of. We refer to these responses as *approval data*, also known as *approval ballots*. In one of these cities, Grenoble, respondents were additionally asked to rate the candidates on a $[0, 1] \in \mathbb{R}$ scale. We henceforth denote these numbers as *ranking ballots*. As a result, we have both approval and rating data for a total of 1,069 respondents. The dataset also includes socioeconomic information such as age, gender, and educational attainment level.

To prevent inconsistencies arising from log-transformation in the calculation of inequality measures indices and to avoid asymmetric ratings (Morrisson and Murtin, 2012), the data were transformed to a $[2, 3]$ scale.

Table 1: *Numerical examples of approval preferences.*

Candidate	Voter 1 Ap	Voter 1 $r_1(j)$	Voter 2 Ap	Voter 2 $r_2(j)$	Voter 3 Ap	Voter 3 $r_3(j)$	Voter 4 Ap	Voter 4 $r_4(j)$	Voter 5 Ap	Voter 5 $r_5(j)$	Voter 6 Ap	Voter 6 $r_6(j)$
x_1	0	1.3	0	1.2	0	1.3	0	1.3	0	1.1	0	1.07
x_2	0	1.3	0	1.3	0	1.5	0	1.9	0	1.1	1	1.61
x_3	0	1.3	0	1.4	0	1.6	0	1.5	1	1.4	1	1.13
x_4	1	1.8	1	1.7	1	1.7	1	1.6	0	1.5	1	1.93
x_5	1	1.8	1	1.8	1	1.7	1	1.9	1	1.9	0	1.12
x_6	1	1.8	1	1.9	1	1.9	1	1.6	1	1.9	1	1.15
Thresh	✓		✓		✓		✗		✗		✓	
WDP	✓		✓		✓		✗		✓		✗	
QDP	✓		✓		✓		✓		✗		✗	

Note: Artificially generated examples illustrating how approval decisions (Ap) relate to individual ratings $r_i(j) \in (0, 1)$. Here, $Ap = 0$ denotes disapproval and $Ap = 1$ denotes approval. The last three rows report whether the respective profile satisfies threshold-dichotomous (Thresh), weakly dichotomous (WDP), or quasi-dichotomous (QDP) preferences, indicated by ✓ if satisfied and ✗ if not.

Note that the Gini-decomposition as well as the variance decomposition are scale-invariant, such that the rescaling do not affect the results based on these measures. The rescaling can have an effect on Generalized Entropy measures, which we introduce in the Section 5. However, the results are similar with altering the rescaling range.

A descriptive data evaluation leads to conclusions that align with observations made in several other experiments on Approval Voting (AV), namely that voters handle AV seemingly without being overwhelmed by the task of categorizing candidates into two groups (Laslier and Van der Straeten, 2008; Baujard et al., 2014; Alós-Ferrer and Granić, 2012). In the specific *Grenoble experiment*, only 21 individuals did not approve any candidate, and only 10 persons approved more than six out of eleven candidates. 187 individuals approved only one candidate, and the overwhelming majority approved either two or three candidates. On average, the approved candidates were rated significantly better than the non-approved ones (2.80 vs. 2.19).

Experimental studies frequently report that the Approval winner does not coincide with the actual (plurality-based) winner (e.g., Alós-Ferrer and Granić, 2012). This can also be observed in the *Grenoble* data: while Emmanuel Macron emerged as the best vote-getter (plurality winner), Jean-Luc Mélenchon got the second most votes but had a head-to-head majority against Macron (i.e., would emerge as plurality run-off winner), the Approval winner was Benoît Hamon with a significant lead to Macron. The approval winner coincides with the Borda winner as well as with the Condorcet winner.

The second data set was conducted by Darmann and Klamler (2023) during the election day for the state parliament in Styria (Austria) in 2019. The data were collected in the state's capital, Graz, by exit polls at nine voting stations. Unlike the Grenoble data, respondents rated six parties instead of candidates. Respondents approved/disapproved parties and rated each on a $[-20, 20] \in \mathbb{N}$ scale, which we transformed for the reasons mentioned above and for the sake of comparability to a $[2, 3]$ range. To avoid redundancies, we summarize the main descriptive summary statistics for this and the other surveys in Table 2.

The third dataset, which we refer to as *France22*, closely resembles the first dataset from *Grenoble*. This dataset originates from an open online experiment conducted during the French presidential election, this time in 2022 (Delemazure and Bouveret, 2024). In total, twelve candidates ran in the 2022 presidential election, seven of whom had also competed in 2017. Respondents rated the candidates on a $[0, 100] \in \mathbb{N}$ scale.

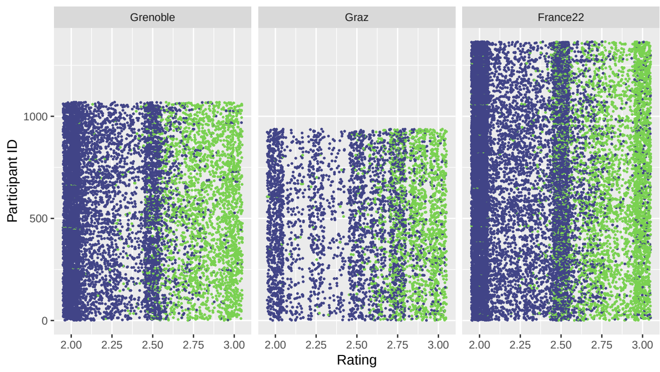

Figure 1: *Approval Data and Ratings*

Note: Each row corresponds to one survey participant, displaying their ratings on a $[2, 3]$-scale and approval choices. Green points mark approved alternatives, blue points mark disapproved ones.

A particular feature of the *France22* dataset is the absence of missing values for candidate ratings. This is due to the fact that when respondents did not provide a rating, the dataset records a value of 50. As a result, a rating of 50 may have two distinct interpretations: respondents may genuinely be indifferent between a positive and a negative evaluation, or they may simply have no opinion about the candidate in question. Unfortunately, it is not possible to distinguish between these two groups. For this reason, in our analyses we employ a variant of the *France22* dataset, denoted $France22_{NA}$, in which *all* ratings of 50 are treated as missing values. While this imputation does not materially affect the results of the weakly dichotomous preference analysis or the examination of the relationship between approval and rating ballots, it does have a noticeable impact on the cluster analysis.

Figure 1 presents the data for the three datasets. Each survey participant's ratings and approval data are represented by a single row. Blue points indicate the rating of a non-approved candidate/party, while green points repre-

Table 2: *Summary statistics for the data sets.*

	Grenoble	Graz	France22
Year	2017	2019	2022
# Respondents	1,069	937	1,365
Alternatives	Candidates	Parties	Candidates
# Alternatives	11	6	12
# Approved			
1	17.5 %	21.8%	8.5%
2	38.0%	47.9%	22.9%
3+	42.6 %	28.7%	66.2%
Avg. Rating ($\bar{r}$)			
Approved	2.80	2.84	2.81
Disapproved	2.19	2.40	2.21

Note: Descriptive statistics for the three datasets used in the analysis. The table reports the number of respondents, the type and number of alternatives, the distribution of approval decisions, and the average ratings. Rating values were rescaled from their original formats to a common $[2, 3]$ scale to ensure comparability across datasets.

sent the ratings of approved alternatives. In all panels, we observe that voters apply distinct cut-off levels—i.e., specific rating values that serve as thresholds to distinguish between approved and non-approved alternatives.

Notably, the *Graz* and *France22* datasets over-represent younger respondents and those with left-leaning political orientations. Darmann and Klamler (2023), referring to the *Graz* dataset, describe this as a "liberal bias." Our focus, however, is on dichotomous structures rather than on predicting electoral outcomes under different voting rules. For this reason, what respondents would actually vote for is not of direct relevance here. Drawing on the CSES dataset, which we also use and introduce below, we argue that the "liberal bias" does not lead to a systematic distortion in our context, although it cannot be entirely ruled out.

We have no indication that the data are distorted by strategic considerations of survey respondents. As Brams and Fishburn (2007, Ch. 7) note, misreporting in a survey requires a reason, and in the context of rating tasks it is difficult to identify any plausible incentive for respondents to deviate from their genuine views. With regard to the *Graz* dataset, Darmann and Klamler (2023) likewise conclude that sincere preferences were recorded.

As shown in Figure 1, focal points are evident particularly in the *French* datasets. They occur at the extremes of the respective scales and at the midpoint. With respect to the concentration of medium values in the *France22* dataset, we have already noted above the issue related to missing values. The focal points, however, are likely to be primarily attributable to the way in which the rating data were collected. In the *Grenoble* study, respondents were asked to mark each candidate's position on a continuous line. At the extremes the labels "contre" and "pour" were provided, and in the middle "indifférent". These labels could, and apparently did, serve as focal points. Data collection for the *France22* dataset followed a very similar design. Here, ratings were recorded via a slider on a scale illustrated with three icons ("emojis"), representing hostile, indifference, and 'in favour'. Again, it can be assumed—and the data strongly suggest—that these icons functioned as focal points.

Such focal points are quite common in experimental data collection. A standard way of assessing their effect is to add a random number to the recorded values. For the *France22* dataset, we added to each of the 16,830 ratings a random draw from the interval $[-5, +5] \in \mathbb{R}$. This random-noise procedure shows that the results are essentially robust. The detailed results are provided in the replication files accompanying this article.

A key advantage of the cluster-analytical approach to assess QDP is its less demanding data requirements: it does not require approval data but only respondents' ratings of the available alternatives. This allows us to use the CSES (2024) data for the cluster-analytical part of the analysis. The CSES is a global research programme where election study teams from participating countries include a common set of survey questions in their post-election studies. The research agenda, questionnaires, and study design are developed by an international committee of experts and implemented by leading social scientists in each country. We use this cumulated data set consisting of 172 nationally representative post-election studies fielded in 54 countries from 1996 onwards.

With this dataset, we can analyse the rating values of 212,729 respondents, each of whom evaluated between six and nine parties or candidates. The cluster-analytical results based on the CSES data can be compared with those derived from the three original datasets. In Section 6.2, we show that the results from the three original datasets fall within the range of those obtained from the CSES data. We interpret this as evidence that the socio-economic bias in the original datasets does not lead to substantial distortions.

4 Threshold approach

We start the empirical part of this paper by evaluating how many respondents exhibit threshold-dichotomous preferences. As depicted in Figures 1, some respondents rate non-approved parties or candidates higher than approved alternatives. This can be seen from the green dots appearing to the left of the blue dots.

Recall that a voter is classified as having threshold-dichotomous preferences if their highest-rated disapproved candidate is rated lower than their lowest-rated approved candidate. As discussed in Definition 2, we argued that the strict criterion is more suitable for empirical applications due to the presence of focal points and the large share of indifferences in respondents' preferences. Table 3 reports the share of survey respondents with threshold-dichotomous preferences. Under the strict version (lower part of the table), the values range from 80 to 90 percent. The values in square brackets indicate bootstrap-generated confidence intervals (on a 5% confidence level and by applying the percentile method). These are relatively high figures, but they should not be over-interpreted. They primarily capture a consistent allocation into approval ballots and provide limited information about whether respondents genuinely evaluate candidates or parties as either "good" or "bad."

A related question is where respondents place their threshold. Saari and Van Newenhizen (1988) showed that electoral outcomes under AV depend critically on this choice. In a three-candidate setting, if all voters use the same threshold, AV coincides with plurality voting when the cut is between the first- and second-ranked candidates, and with anti-plurality rule when it is between the second and third. The Condorcet efficiency of AV heavily depends on the threshold-setting behaviour of voters (Gehrlein et al., 2016), as well the resilience of AV against the electoral success of polarizing candidates (Barbaro, 2025), also coined 'exclusive candidates' vs. 'inclusive candidates' (Baujard et al., 2014, 2020)

For the threshold calculation, we use the midpoint between the lowest rating among approved and the highest rating among disapproved candidates, defined for each respondent i as

$$t_i = \tfrac{1}{2}\left(\min_{x \in A_i} r_i(x) + \max_{y \in D_i} r_i(y)\right).$$

Figure 2 depicts the distribution of individual thresholds across datasets. The violin plots illustrate the density of observations over the rating scale, while the embedded boxplots indicate the median and interquartile range.

Three observations from the data analysis are noteworthy:

1. The distribution of thresholds lies predominantly in the upper half of the rating scale (between 2.5 and 3). This corresponds to the observation that individuals generally set their threshold above their median rating. As shown in Fig. 2, the violins widen between 2.4 and 2.8. Around 90% of respondents in the French datasets set their threshold above their respective median rating, while in the *Graz* dataset the share is about two thirds.

2. The median values of the thresholds are similar across all datasets, at around 2.65. In Fig. 2 the median is indicated by the central vertical line in the boxplots.

3. The distribution of threshold values is also comparable across the datasets. The boxplots display a similar level and a comparable range.

As noted by Brams et al. (1988), the threshold setting can reflect preference intensities and thus does not represent a strategically motivated decision that would render the election outcome idiosyncratic, as claimed by Saari and Van Newenhizen (1988). In a three-candidate setting, if the top-ranked candidate is viewed very positively, the last-ranked very negatively, and the middle candidate as neutral, then placing the threshold between the first and second candidate may indicate a stronger concern with ensuring *who is* elected. By contrast, setting it between the second and third candidate suggests a motivation to exclude particularly unsuitable candidates. Following this perspective, voters appear more inclined to use their threshold to indicate and reinforce their preferred candidates.

5 Weakly dichotomous preferences

In this section, we present results on the share of respondents exhibiting weakly dichotomous preferences as defined in Definition 3. Our main focus lies on the Gini decomposition, which we set out in Section 2.

Table 4 presents the share of survey respondents for whom the within-group dispersion is smaller than the respective between-group dispersion. The values in square brackets indicate bootstrap-generated confidence intervals (on a 5% confidence level and by applying the percentile method) based on 1,000 replications each.

We find that between 50 and 70 per cent of respondents can be characterised as having weakly dichotomous preferences. For this majority, the assumption of dichotomous preferences provides a reasonable approximation of their actual preferences. However, for 30 to 50 per cent of respondents this assumption does not hold, which may pose a challenge for theoretical models that rely critically on dichotomous preferences.

Table 4 also presents the share of respondents with weakly dichotomous preferences calculated using three alternative inequality measures, in order to strengthen the robustness of the results obtained with the Gini coefficient and to address common criticisms of its use. We consider the additively decomposable Theil's T-measure (Ebert, 1988, 2010), the Atkinson-measure[5] (Atkinson, 1970), and the squared coefficient of variation (Shorrocks, 1980). We use the weights $s_{g,i} \equiv \frac{|g|\,\mu_{g,i}}{m\,\mu_i}$, $g \in \{A_i, D_i\}$, which reflect the relative group shares in voter i's distribution.

The Theil-measure

$$\mathcal{T}_i = \tfrac{1}{m}\sum_{j=1}^{m} \tfrac{r_i(j)}{\mu_i} \ln\!\left(\tfrac{r_i(j)}{\mu_i}\right) = \underbrace{\sum_{g \in \{A_i, D_i\}} s_{g,i}\, \mathcal{T}_{g,i}}_{\text{within}} + \underbrace{\sum_{g \in \{A_i, D_i\}} s_{g,i} \ln\!\left(\tfrac{\mu_{g,i}}{\mu_i}\right)}_{\text{between}},$$

[5] We set the inequality-aversion parameter to $\varepsilon = 1$, the limiting case in which the Atkinson index compares the geometric mean with the arithmetic mean. This specification gives the measure particular sensitivity to differences at the lower end of the distribution.

Table 3: *Share of survey respondents with threshold-dichotomous preferences (values in %)*

Criteria	Grenoble	Graz	France22	France22$_{NA}$
$\min_{x\in A_i} r_i(x) \geq \max_{y\in D_i} r_i(y)$	89.2	67.2	63.7	76.9
	[87.6, 90.7]	[64.8, 69.6]	[61.8, 65.8]	[75.3, 78.7]
$\min_{x\in A_i} r_i(x) > \max_{y\in D_i} r_i(y)$	90.2	86.0	81.3	85.1
	[88.8, 91.7]	[84.1, 87.8]	[79.6, 82.8]	[83.6, 86.6]

Note: The table reports the share of respondents with threshold-dichotomous preferences under a weak (top) and a strict (bottom) definition. Values in square brackets are bootstrap confidence intervals (5% level, percentile method).

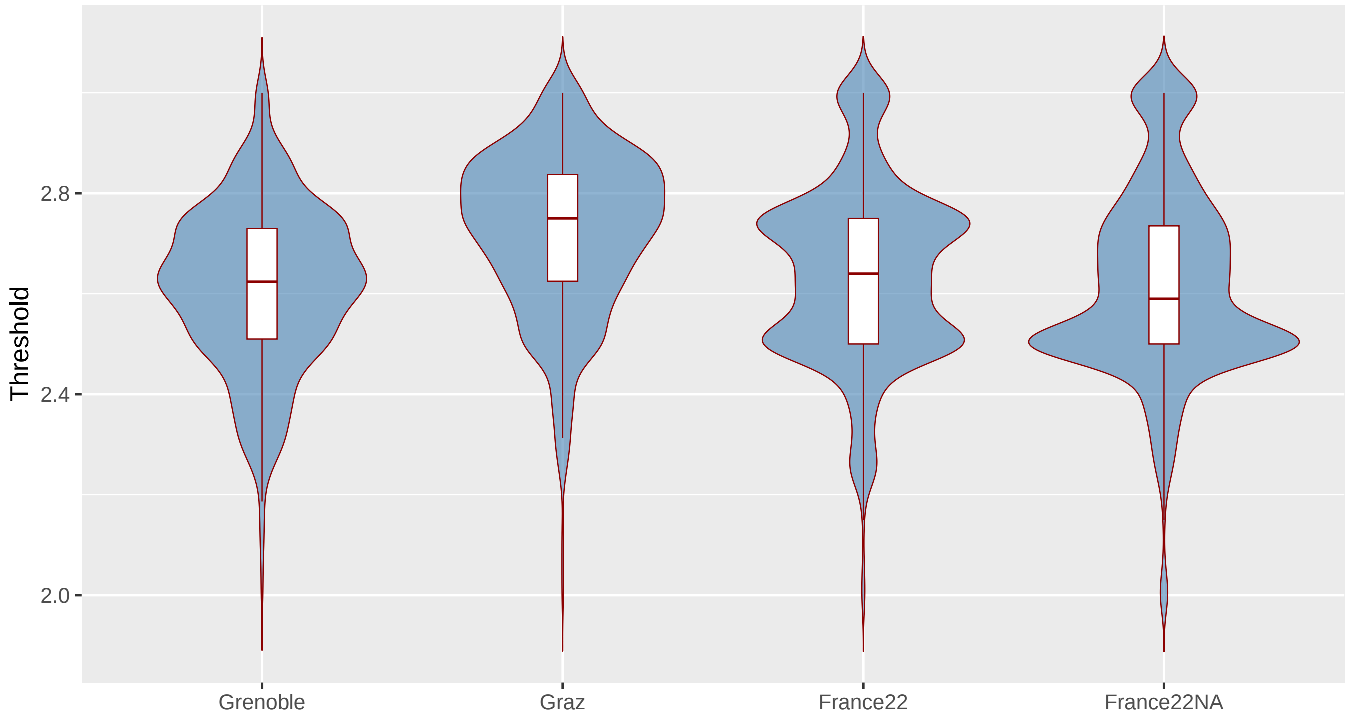

Figure 2: *Distribution of thresholds set by respondents.*
Note: The plot depicts the distribution of individual approval thresholds across datasets. Violin plots show the density of threshold values, with embedded boxplots marking the median and interquartile range.

and the Atkinson-measure

$$\begin{aligned}\mathcal{A}_i &= 1 - \frac{\exp\left(\frac{1}{m}\sum_{j=1}^m \ln r_i(j)\right)}{\mu_i} \\ &= \underbrace{\sum_{g\in\{A_i,D_i\}} s_{g,i}\,\mathcal{A}_{g,i}}_{\text{within}} \\ &+ \underbrace{\left[\sum_{g\in\{A_i,D_i\}} s_{g,i}\frac{g_{g,i}}{\mu_{g,i}} - \frac{g_i}{\mu_i}\right]}_{\text{between}},\end{aligned}$$

both belong to the class of Generalized Entropy (GE) measures (Shorrocks, 1980; Cowell and Victoria-Feser, 1996). Their drawback is that the common scaling may distort the values, as they are not scale-invariant.

As a fourth measure, we use the squared coefficient of variation (SCV), which through the ANOVA identity is perfectly additively decomposable into within- and between-group components (Shorrocks, 1980). For voter i, let $p_{g,i} = \frac{|g|}{m}$ $(g \in \{A_i, D_i\})$ denote the relative size of each group in voter i's partition of the alternatives, and let $\sigma_i^2(r)$ denote the variance of respondent's i ratings. Then the SCV is

$$\begin{aligned}\text{SCV}_i &= \frac{\sigma_i^2(r)}{\mu_i^2} \\ &= \underbrace{\frac{1}{\mu_i^2}\sum_{g\in\{A_i,D_i\}} p_{g,i}\,\sigma_{g,i}^2}_{\text{within}} \\ &+ \underbrace{\frac{1}{\mu_i^2}\sum_{g\in\{A_i,D_i\}} p_{g,i}\left(\mu_{g,i}-\mu_i\right)^2}_{\text{between}},\end{aligned}$$

A limitation of the SCV is the absence of an upper bound ($\text{SCV}_i \in [0,\infty)$). This reduces interpretability and can let extreme observations dominate, in contrast to bounded measures such as the Gini or GE indices.

The application of the three additional inequality measures corroborates the overall picture derived from the Gini decomposition. The share of respondents with weakly dichotomous preferences generally lies between 50 and 70

Table 4: *Share of respondents with weak dichotomous preferences according to the inequality-measure approach. Values in %.*

Index	Grenoble	Graz	France22	France22$_{NA}$
Gini	68.4	50.8	55.2	68.4
	[66.1, 70.8]	[48.1; 54.4]	[53.1; 57.2]	[66.5; 70.3]
Theil-T	73.6	47.2	68.0	75.73
	[71.5, 75.8]	[44.6; 49.7]	[66.1; 69.8]	[74.1; 77.2]
Atkinson	70.3	45.1	65.3	73.9
	[68.1, 72.6]	[42.6; 47.8]	[63.4; 67.2]	[72.3; 75.5]
SCV	73.3	38.6	63.8	72.3
	[71.1, 75.6]	[35.9; 41.1]	[61.8; 65.8]	[77.1; 74.5]

Note: Share of respondents with weakly dichotomous preferences based on inequality decompositions (values in per cent). Within-group dispersion (within approved and disapproved sets) is compared to between-group dispersion. Square brackets report bootstrap confidence intervals (5% level, percentile method, 1,000 replications). Results are shown for Gini, Theil-T, Atkinson, and squared coefficient of variation (SCV) indices.

per cent.

The share of respondents to whom we cannot attribute weakly dichotomous preferences is largely driven by those who approve only one or two alternatives. The more alternatives a respondent approves, the higher the likelihood of observing weakly dichotomous preferences. We depict this observation in Fig. 3. Restricting attention to respondents who approve at least three alternatives, we obtain a robust share of around three quarters who exhibit weakly dichotomous preferences.

It should be noted that for the Gini decomposition we used the procedure proposed by Lerman and Yitzhaki (1989) (henceforth: YL). The YL approach decomposes overall inequality into within- and between-group components by linking the between-group term to the covariance between group means and the fractional rank of individuals. An alternative procedure is the decomposition by Bhattacharya and Mahalanobis (1967) (henceforth: BM), which differs from YL in that the between-group component is defined as the Gini coefficient of group means. In the BM approach, inequality arising from cross-group rank overlap is reported separately as an overlap component, whereas in the YL approach such overlap is mostly absorbed into the within-group component.

In our view, the YL procedure is better suited to our setting. Unlike BM, which defines the between component solely in terms of differences in group means, YL links the between component to the rank ordering of individuals across groups. This makes the decomposition sensitive to the extent to which approval categories structure the ranking of candidates. In our context, where the key question is whether approved and disapproved candidates form distinct strata in respondents' evaluations, this feature of the YL approach provides a conceptually meaningful measure of between-group inequality.

Because of the absorption to the within component, the YL approach is more conservative in the sense that it yields lower WDP shares than BM. To assess how sensitive our results are to the choice of decomposition, we also calculated the WDP shares under the BM approach. The results are 84.5% for *Grenoble*, 76.25% for *Graz*, 77.8% for *France22*, and 84.4% for *France22*$_{NA}$.

Both YL and BM belong to the class of additive decomposition measures and stand in contrast to so-called path-independent decompositions (Foster and Shneyerov, 2000). In a recent contribution to this journal, Fleurbaey et al. (2025) emphasise that additive decompositions attribute part of the between-group variation to the within component, whereas path-independent decompositions attribute part of the within-group variation to the between component. This implies that the choice of method should be guided by the analytical focus: studies concerned with within-group inequality are better served by a path-independent procedure, while analyses centred on between-group inequality are more appropriately conducted using an additive decomposition. Since our analysis is concerned with whether approval categories account for systematic differences in candidate ratings—that is, with the magnitude of between-group inequality—the additive approach is the appropriate choice in our setting.

However, using a path-independent approach to calculate the within component and an additive approach to the between component, as proposed by Fleurbaey et al. (2025),[6] would cause the WDP shares to drop significantly (about 49% in the *Grenoble* dataset, to merely 6% in *Graz*, in to about 35% by focusing on the *France22* dataset).

6 Quasi-dichotomous preferences

In the preceding section, we treated respondents' group assignments (approvals and disapprovals) as given and examined the dissimilarities within and between these groups. In contrast, the present approach sets aside the approval data and instead determines the optimal number of clusters based on respondents' rating data. While the previous section relaxed the assumption of indifference within each group, the current analysis questions the expectation that voters categorise the alternatives into exactly two groups. Thus, this approach allows us to determine whether the ratings are best grouped into two clusters (indicating a dichotomy), three clusters (trichotomy), or even more. In other words, irrespective of the approval data, we aim to identify how many respondents express ratings that can be optimally partitioned into two subsets (clusters).

Example 1 (Trichotomous preference pattern)**.**

[6] We thank Domenico Moramarco, a co-author of Fleurbaey et al. (2025), for sharing the R code implementing their proposed method.

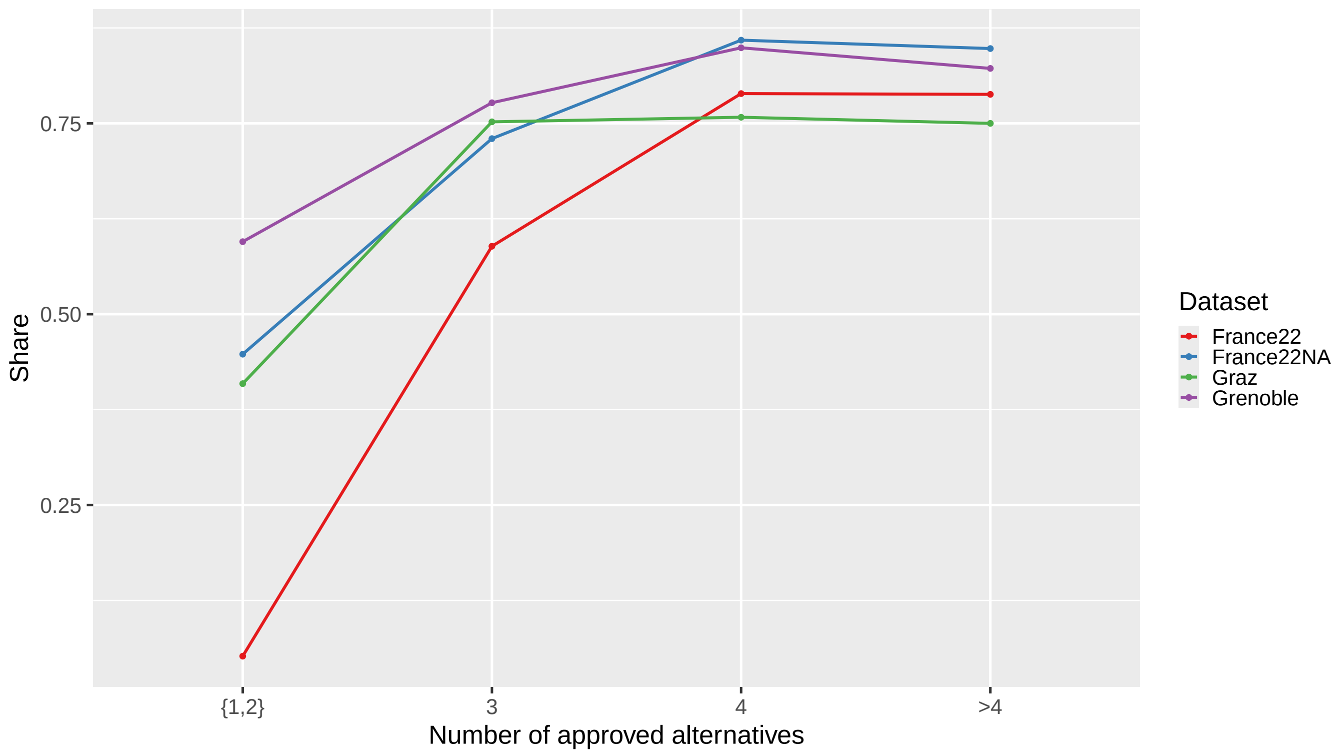

Figure 3: *Share of respondents with weakly-dichotomous preferences based on the Gini decomposition*
Note: Share of respondents with weakly dichotomous preferences by number of approved alternatives. Categories are $\{1, 2\}$ (one or two approvals), 3, 4, and > 4 (denoting four or more approvals).

Consider a voter who assigns the following ratings to six candidates on a $[1, 2]$ scale: $r_i = (1.1, 1.1, 1.5, 1.5, 1.9, 1.9)$. Dichotomous preferences are characterised by the ability of voters to divide the candidates into two groups, with identical ratings within each group. In this example, however, the voter appears to divide the candidates more naturally into three rather than two groups. Within each group, the ratings are indeed identical. The violation of the dichotomy assumption therefore does not arise from a lack of indifference within groups, but rather from the assumption of exactly two groups.

The approach presented here is related to a line of research concerned with mapping available ranking ballots onto inferred approval ballots. Our approach is reminiscent of the probabilistic model introduced by Regenwetter and Grofman (1998). According to their model, the probability that a voter approves the set $\mathcal{X} \subset \mathscr{B}$ equals the probability that they approve of for $|\mathcal{X}|$ candidates multiplied by the probability that they rate all alternatives in $\mathcal{X}$ higher than all alternatives in $\mathscr{B} \setminus \mathcal{X}$ (Regenwetter and Grofman, 1998, p. 426). Our approach differs in that we do not ask which alternatives are approved, but whether partitioning into two groups provides the best approximation of preferences.

As indicated by Def. 4, the QDP approach follows two sequential steps: first, we determine the optimal number of clusters for each survey participant; second, for those with an optimal cluster count of two, we assess the quality of the clustering. With regard to the empirical strategy, both steps rely on silhouette scores as the evaluation criterion.

The silhouette score (Rousseeuw, 1987) for each data point r_j is defined as:

$$s(r_j) = \frac{b(r_j) - a(r_j)}{\max\{a(r_j), b(r_j)\}} \in [-1, 1] \subset \mathbb{R}, \tag{4}$$

where $a(r_j)$ is the average intra-cluster distance, i.e., the average distance of $r_i(j)$ to all other vectors in the same cluster (cohesion), and $b(r_j)$ is the average nearest-cluster distance (separation).

To identify the optimal number of clusters, $\tilde{k}$, we apply a k-means clustering analysis[7] with the *average silhouette width* (ASW, Batool and Hennig, 2021, also known as the 'silhouette coefficient') as indicator. In particular, we compute the average silhouette width for different numbers of clusters, $k = 1, 2, \ldots$, to evaluate clustering quality across varying cluster counts. The ASW represents the mean silhouette score across all observations within a given clustering solution. The optimal number of clusters, $\tilde{k}$, is determined by selecting the value of k that corresponds to the maximum ASW (Kaufman and Rousseeuw,

[7] By using the k-means clustering algorithm (Hartigan and Wong, 1979), we split each respondent's rating values into k clusters based on their distance to the mean value of all ratings of the respective respondent (*hard clustering*). The rating values are moved between clusters, one at a time, based on their closeness to the mean of each cluster (measured by Euclidean distance). The algorithm finishes when no rating value can be moved between clusters without increasing the average Euclidean distance between rating values and the means of their respective clusters. The PAM ('Partitioning around medoids') is an alternative to k-means clustering. Like k-means, it partitions data into k clusters. But instead of means (centroids), it uses the actual observations that minimise total dissimilarity to points in their cluster ('medoids'). On discrete scales with many ties it is sensitive to tie-breaking procedures and tends to over-segment. As our aim is parsimonious partitioning rather than outlier-robust clustering, k-means is more appropriate here.

2005, p. 86).

To illustrate the connection between our QDP (Def. 4) and the WDP (Def. 3), consider that in Eq. (4) $a(r_j)$ denotes the within-group dissimilarity, while $b(r_j)$ represents the between-group dissimilarity.

For high silhouette scores ($s(r_j) \to 1$), the within-group dissimilarity is much smaller than the smallest between-group dissimilarity. In this sense, high silhouette values indicate strong clustering, as the second-best cluster is not nearly as close as the actual cluster assignment. Conversely, when $s(r_j) \sim 0$, within-group and between-group dissimilarities are approximately equal, meaning that r_j cannot be sharply assigned to one cluster or another.

Example 2 (Optimal cluster number based on maximum average silhouette width)**.**
Consider the rating values of Voter 4 in Table 1. We calculate the ASW for different numbers of clusters, $k = 1, \ldots, 5$. *The ASW is zero for* $k = 1$, *0.72 for* $k = 2$, *and 0.596 for* $k = 3$. *Since the ASW value is highest at* $k = 2$, *clustering Voter 4's rating values into two clusters is optimal.*

Furthermore, the cluster analysis indicates that the two rating values equal to 1.9 form one cluster (the approved cluster), while the remaining values constitute the other (disapproved) cluster.

We depict the example in Table 5. The columns $r_j \equiv r_4(j)$ *and 'Ap' are identical to those in Table 1 (column 'Voter 4'). The column 'Opt. Cluster' indicates the assignment of* r_j*-values to the two clusters, while the column 'Silhouette' provides the corresponding silhouette scores.*

For example, the silhouette score is $s(i) = 1$ *for* $i \in \{r_2, r_5\}$ *because the mean rating in their cluster is 1.9, resulting in a distance to the cluster mean of zero,* $a(i) = 0$. *The average distance to the neighbouring cluster (the disapproved cluster) is* $b(i) = 0.4$. *Since* $b(i) > a(i)$, *the numerator and denominator of Eq. (4) are equal, yielding* $s(i) = 1$.

Regarding the disapproved cluster, consider r_1. *Its average distance to the other elements in its cluster,* (r_3, r_4, r_6), *is* $|1.3 - 1.566| = 0.266 = a(1)$.

The distance to the approved-cluster mean is $b(1) = 0.6$. *Thus,* $s(r_1) = \frac{0.6-0.266}{0.6} = 0.556$.

Even if $\tilde{k} = 2$ is the optimal number of clusters for a survey participant, this alone provides little information about the quality of the clustering. It merely indicates that a two-cluster solution offers a better grouping than three or more clusters. To assess clustering quality, we once again rely on the silhouette score.

Following the categorization proposed by Kaufman and Rousseeuw (2005, Tab. 4), we classify a clustering as "strong" if the silhouette score satisfies $\bar{s}(r_i) > 0.70$, as "moderate" if $\bar{s}(r_i) \in [0.5, 0.7]$, and as "weak" if the average silhouette score is below 0.5. For instance, the average silhouette score for Voter 4 in the above example is 0.723, which would be categorized as a strong clustering.

We pick the *Graz* dataset to describe our results in depth and present all results in Table 6. We find that $\tilde{k} = 2$ applies to 78.4% of respondents. An optimal cluster count of three is observed for 20.6%, while only 1% of respondents exhibit an optimal cluster count of four.

As an intermediate result, we conclude that more than two clusters provide a better grouping for roughly one-fifth of survey participants, suggesting that their preferences align more closely with trichotomous rather than dichotomous preferences. Among respondents with $\tilde{k} = 2$, we find strong clustering for less than half, while approximately half of all respondents exhibit only moderate clustering quality.

Table 6 summarizes the results. The second to fourth columns ('Optimal Cluster ($\tilde{k}$)') denote the share of respondents with 2, 3, or more optimal clusters. The fourth to sixth columns $({}'\tilde{k} = 2')$ show the share of observations within the individuals with an optimal $k = 2$ and subsume them into three categories according to the respective categories. The column 'QDP' finally denotes the share of respondents holding preferences that align with Def. 4.

The data from the *Grenoble* experiment indicate a lower share of respondents with optimal clustering in two groups compared to the *Graz* dataset. However, within this group, for three-quarters this grouping is highly robust. We find the lowest share of respondents with $\tilde{k} = 2$ in the *France22* dataset. The number of 46.4% reported in Table 6 appears underestimated because, as mentioned in Section 3, respondents who did not report a rating were recorded with the median value of 50. When we convert all observations with a rating of 50 to missing values (*France22*$_{NA}$), we came to somehow higher values.

According to our above-described procedure, we denote the preferences of a respondent as quasi-dichotomous if they optimally group their ratings in two groups and if this grouping is strong. In this sense, we assign quasi-dichotomous preferences to (.618 × .754 =)46.6% of the respondents in the *Grenoble* experiment and to 34% in *Graz*. By weakening this requirement by including those with moderate clustering, we get around 60% in *Grenoble* and around 74% in *Graz*. In any case, a significant share of the voters we cannot attribute (quasi-)dichotomous preferences.

6.1 Robustness check: fuzzy clustering

To further validate our results, we assess the quality of the clustering using an alternative approach, known as fuzzy clustering ((*FANNY*), Kaufman and Rousseeuw, 2005, Ch. 4).

Thus far, each observation is assigned to exactly one cluster (hard clustering, see Footnote 7). This approach inherently excludes the possibility that voters may evaluate individual candidates or parties somewhere between approval and disapproval. For example, a voter might strongly agree with certain parts of a party's platform while disagreeing with others, leading to an ambiguous stance between approval and disapproval. To account for this, we incorporate fuzzy clustering as a robustness check. In this approach, each rating is distributed across multiple clusters, and the degree of membership to a cluster is measured using a membership index[8] (Dunn, 1973). The more

[8] Formally, fuzzy clustering involves minimizing the following objective function:

$$\sum_{k=1}^{m} \frac{\sum_{i,j=1}^{n} u_{ik}^2 \cdot u_{jk}^2 d(i,j)}{2 \cdot \sum_{j=1}^{n} u_{jk}^2}.$$

Here, u_{ik} indicates that object i belongs to cluster k, and each u_{ik} is strictly positive, with the constraint that for all $i \in \{1, \ldots, n\}$,

Table 5: *Optimal clustering and silhouette calculations for Voter 4*

r_j	Silhouette	Ap	Opt. Cluster	Cluster mean
1.3	0.556	0	0	1.5
1.9	1.000	0	1	1.9
1.5	0.667	0	0	1.5
1.6	0.556	1	0	1.5
1.9	1.000	1	1	1.9
1.6	0.556	1	0	1.5

Table 6: *Cluster Analysis: Distribution of optimal clusters across respondents and robustness.*

	Optimal Cluster $(\tilde{k})$			$\tilde{k}=2$			QDP
Survey	2	3	4+	Strong	Moderate	Weak	
Grenoble	61.8	21.5	16.7	75.4	23.4	1.1	46.6
Graz	78.4	20.6	0.9	43.7	50.9	5.4	34.2
France22	46.4	26.9	26.7	68.8	31.1	0.4	31.9
France22$_{NA}$	54.8	22.2	23.0	61.0	38.5	0.5	33.4
	Robustness check: fuzzy clustering (FANNY)						
Grenoble	57.6	28.1	14.3	74.5	23.7	1.8	42.9
Graz	–	–	–	42.8	50.6	6.7	
France22	38.0	26.4	35.5	70.2	28.8	0.9	26.6
France22$_{NA}$	52.05	21.51	26.44	56.58	41.92	1.5	36.54

Note: The first three columns report the share of respondents with an optimal number of clusters $\tilde{k} = 2, 3$, or 4+. Columns under $\tilde{k} = 2$ classify those cases by strength (strong, moderate, weak). The final column ('QDP') shows the share of respondents whose preferences satisfy the definition of quasi-dichotomous preferences. The lower panel presents robustness checks using fuzzy clustering (FANNY). All values in per cent.

similar an observation is to others within a given cluster, the higher its membership value for that cluster. If significant overlap is detected, this suggests that the assumption of strictly separable clusters may be overly rigid.

Example 3 (Ambiguous rating and membership value)**.** *Consider the following tuple of rating values:* $(1.1, 1.1, 1.1, 1.5, 1.9, 1.9, 1.9)$*. A voter holding these ratings would be forced in an Approval Voting environment to assign the candidate they rated 1.5 to either the approval or the disapproval group. However, it is reasonable to assume that this candidate essentially lies somewhere between these two groups. This inbetweenness is measured with the membership values in fuzzy clustering.*

In a second step, we calculate the *silhouette values*—analogous to the previous analysis—for those survey participants with an optimal cluster number of $\tilde{k} = 2$, but this time based on *fuzzy clustering* instead of *k-means clustering*. The interpretation of these values follows the same logic as in the previous analysis.

One limitation of fuzzy clustering is that determining the optimal number of clusters is only valid for $k < n/2$. In the *Graz* dataset, which includes only six parties, this restricts the interpretability of the results. To address this issue, we analyse silhouette values only for those voters for whom the k-means-based cluster analysis identified exactly two optimal clusters.

The analysis of the data using fuzzy clustering suggests a high degree of robustness regarding the previously obtained results. All values essentially correspond to those obtained in the main analysis. The bottom rows in Table 6 present the respective values. Because of the limitation mentioned in the preceding paragraph regarding the fuzzy analysis in *Graz*, we left the respective cells empty.

With very few exceptions (each around 5%), we observe high membership values in all three surveys. There is little evidence to support the previously stated concern regarding intermediate evaluations by the voters.

6.2 Quasi-dichotomous preferences in CSES data

The less-demanding data requirement of the cluster-analytical assessment allows us to draw on sources beyond the three datasets primarily employed. The analysis on a broader dataset thereby provides indication on whether the findings obtained in the previous subsection are generalisable.

We draw on data from the Comparative Study of Electoral Systems (CSES, 2024). The CSES is a collaborative research programme that provides harmonised survey data from post-election studies conducted in a wide range of countries. The CSES integrates nationally representative surveys with contextual information on electoral rules, political institutions, and party systems, thereby enabling systematic cross-national comparisons of individual political behaviour. Its design ensures both temporal and spatial comparability, making it one of the most widely used sources for the study of electoral attitudes, preferences, and participation in comparative politics.

We draw on rating data from 212,729 survey respondents, covering 172 elections across 54 countries. Each respondent evaluated at least six different parties, and the

$\sum_k u_{ij} = 1$ holds (Kaufman and Rousseeuw, 2005, p. 170f).

elections comprise an average of 1,237 respondents (minimum = 150, maximum = 3,611). We apply the same procedure to this large dataset as to the three original datasets. For each election, we obtain the number of respondents for whom the optimal number of clusters is two, three, or four, respectively.

On average across all elections, the optimal number of clusters is $\tilde{k} = 2$ for 57.4% of respondents. For one quarter of respondents the optimal number is three, while for 17.8% it is four.

Figure 4 illustrates these results. Each point represents the share of respondents for whom a given number of clusters is optimal, and the boxplots summarise the corresponding distributions.

Overall, the results derived from the CSES data closely resemble those presented in Table 6. The CSES data tend to reveal a lower incidence of individuals for whom the optimal number of clusters is two. Taken together, these analyses indicate that the preference ratings in the three primary datasets do not appear to be systematically biased.

7 Concordance between reported and cluster-derived approval ballots

Our cluster-analytical procedures provide a way of mapping ranking ballots onto approval ballots, provided that clustering is reasonable in two subgroups. A natural question is how closely the approval sets generated through cluster analysis correspond to the approval sets reported in the three datasets. Focusing on respondents who display QDP, we can ask: how often does an alternative appear both in the reported approval set (from the datasets) and in the cluster-analytically generated approval set? We measure this correspondence by means of a *matching value*.

Let $m_i(j)$ be a binary variable. It is equal to 1 if an alternative $j \in \mathscr{B}$ is contained both in respondent i's reported approval set and in the cluster-analytically generated approval set, and 0 otherwise. The matching value regarding an alternative j is $\mathcal{M}(j) = \sum_i {}^{m_i(j)}/_N$, where N denotes the number of respondents. The matching value over all respondents and all candidates is then the proportion of candidate–voter pairs for which reported and cluster-based approval coincide.

Over all candidates, respondents, and data sets we find a remarkably high coincidence between reported approval ballots and those resulting from the cluster analysis of 78%. This value is similar across the data sets (80.8% in *Grenoble*, 71.8 in *Graz*, and 78.1% in *France22*). Given that around 20% do not comply with threshold-dichotomy (as shown in Sec. 4), we interpret the high matching values as indication for a very good approximation of preferences to approval ballots.

Next we focus on the matching values for each candidate or party. We expect the matching values to become higher for polarizing candidates/parties. Polarizing political figures are often either rated very highly or strongly rejected, leading to their ratings being disproportionately located at the extremes of the scale (Barbaro, 2025). Accordingly, voters tend to hold either a very positive or a very negative opinion of polarizing candidates/parties, such that they can be more readily assigned to the categories of approved or disapproved. A high matching value therefore reflects polarisation rather than political quality or the capacities of supporters.

Table 7 reports the matching values for each candidate. As in the previous analyses, the number in brackets denote bootstrap confidence intervals.

The matching values differ remarkably across the candidates and political parties.[9]

As expected, significantly higher matching values are observed for polarizing candidates and parties on the right edge of the political spectrum. Both Marine Le Pen in the 2017 presidential election (*Grenoble*) and the right-wing FPÖ in Austria exhibit the highest values. In the 2022 election, Le Pen ran for president again, and the previously observed high matching value was confirmed. The only candidate with a higher value was Éric Zemmour, the candidate of the far-right *Reconquête* party. The third candidate with a matching value above 0.9 was Nicolas Dupont-Aignan, who also belonged to the group of radical-right candidates. We also see that – contrary to our expectation – we cannot assign higher matching values for candidates from the left edge of the political spectrum. Examples in this regard are the Communist Party in Austria (KPÖ), and candidates from the left like Jean-Luc Mélenchon, Nathalie Artaud, Philippe Poutou, and Fabien Roussel.

We would, however, like to qualify this relationship between the matching values and a far-right orientation with a caveat. As discussed in Sec. 3, the respondents are not representative in terms of either age or political convictions. Instead, the sample over-represents a younger, left-leaning population. This may affect the results, and with the data at hand we are not in a position to control for this potential bias.

Nota bene: These results do not challenge the well-documented finding that extreme and polarizing candidates tend to perform worse under Approval Voting than under plurality rule (Brams and Fishburn, 1978; Laslier and Van der Straeten, 2008; Alós-Ferrer and Granić, 2012). Rather, our observations indicate that voters distinctly classify extreme candidates into approved or disapproved categories, whereas lower matching values reflect a more centrist perception.

Next, we show that respondents with quasi-dichotomous preferences exhibit a significantly higher coincidence between reported and cluster-analytically assigned approved candidates or parties. To measure this coincidence, we use the Pearson correlation (ϕ) coefficient for two binary variables.[10]

We run OLS regressions with the ϕ-values as dependent variables. The independent variables is a binary variable indicating whether the respective respondent has QDP or not. As control variables, we consider socio-demographic characteristics (age, gender, and educational level achieve-

[9] Replacing the rating values of 50 by missings in the *France22* dataset (*France22$_{NA}$*) does not change the values much (for example: Nicolas Dupont-Aignan get 0.942 instead of 0.922, Le Pen 0.953 instead of 0.939, Macron 0.859 instead of 0.842).

[10] For a respondent, let $R \subseteq \mathscr{B}$ denote the set of approved alternatives and $C \subseteq \mathscr{B}$ the set of cluster-assigned alternatives. With $p_R = \frac{|R|}{m}$, $p_C = \frac{|C|}{m}$, and $p_\cap = \frac{|R \cap C|}{m}$, the Pearson correlation coefficient is $\phi(R,C) = \frac{p_\cap - p_R p_C}{\sqrt{p_R(1-p_R)\,p_C(1-p_C)}}$.

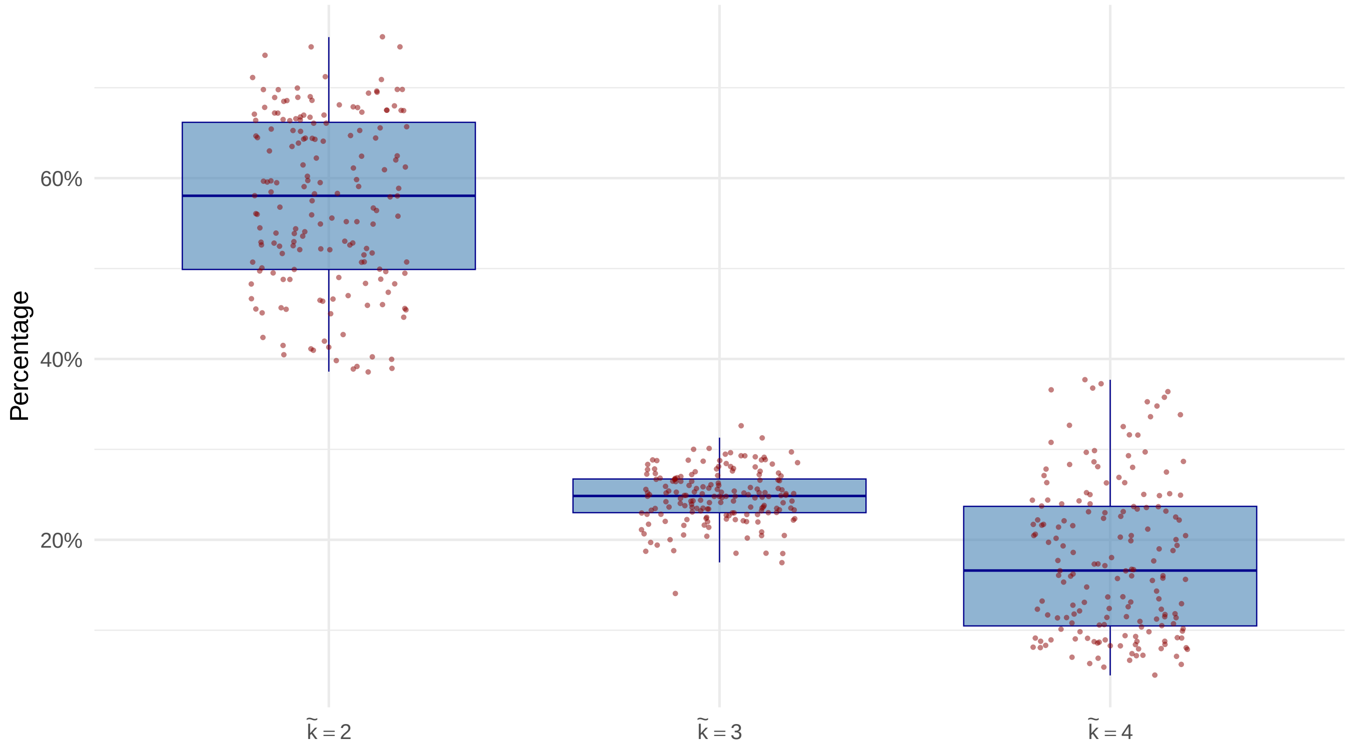

Figure 4: *Distribution of average optimal clusters across 172 elections*

Note: Each point denotes one of the 172 elections from the CSES dataset, covering rating data from around 210,000 respondents. Each election contributes three points: one in the distribution for $\tilde{k} = 2$, one for $\tilde{k} = 3$, and one for $\tilde{k} = 4$. The boxplots summarise these distributions across elections, showing the median, interquartile range, and overall spread.

ment). Table 8 displays the regression results.

For both surveys conducted during the French presidential elections, we find significant effects for the dependent variable. The negative signs indicate that individuals with a 'beyond-two' optimal number of clusters less often exhibit that their approved candidates/parties coincide with the cluster-analytical approval set. The coefficient in the *Graz* dataset is also negative but on a lower significant level.

To further assess the robustness of our results, we conducted four robustness checks (RC). In the first RC, we incorporated individual characteristics that exhibit a strong affinity for alternatives on the extreme right or left. We categorize a respondent to have such an affinity if they rated highest a candidate from the political edges.[11] In the second RC, we computed robust standard errors instead of conventional standard errors. This adjustment was motivated by strong evidence of heteroscedasticity in the regressions using the French data.

For the third RC, we generated a pooled dataset that integrates all three surveys. We applied the main model to this dataset and accounted for survey differences by including fixed effects and calculating robust standard errors. In the fourth RC, we followed the same approach but additionally included the dummy variables from the first RC (affinity to the political extremes).

For the main specification and all four robustness checks, we present the coefficients and corresponding confidence intervals for the variable QDP in Figure 5. As can be observed, our key finding from the main specification remains robust: respondents with quasi-dichotomous preferences exhibit a significantly higher coincidence between reported and analytically generated approval subsets.

8 Dichotomous preferences and socio-demographic factors

In this final section of the empirical analysis, we examine which factors may account for the emergence of dichotomous preference structures. For example, are quasi-dichotomous preferences more frequently observed among female respondents than among males, or do they become more or less prevalent with increasing age? To investigate the relationship between dichotomous preferences and socio-demographic characteristics, we draw on the CSES dataset introduced in Section 3 and employed in Subsection 6.2 for the analysis of quasi-dichotomous preferences. From the results presented there, we know whose preferences are best represented by two clusters and whose by more than two.

The dataset also records socio-economic characteristics of respondents, including age, gender, educational attainment (five categories: illiterate, lower secondary, upper secondary, post-secondary non-university, and university education), household income (quintiles), ideological self-placement (Likert scale from 0 = "left" to 10 = "right"), and satisfaction with democracy (five categories from 1 = "not at all satisfied" to 5 = "very satisfied").

[11] Specifically, we classify respondents as radical right supporters if they rated the FPÖ party exceptionally highly in the *Graz experiment* or reported a plurality vote in favour of one of the aforementioned radical-right candidates in the French experiments. Similarly, we identify respondents as radical-left supporters if they rated the Austrian Communist Party (KPÖ) very highly or cast a plurality vote for Jean-Luc Mélenchon, Philippe Poutou, or Fabien Roussel.

Table 7: *Matching values for different candidates and parties*

Grenoble		Graz		France22	
Candidate	Value / CI	Party	Value / CI	Candidate	Value / CI
Dup.-Aignan	0.819 [0.791; 0.846]	SPÖ	0.606 [0.572; 0.642]	Dup.-Aignan	0.922 [0.908; 0.935]
Le Pen	0.971 [0.957; 0.983]	ÖVP	0.700 [0.668; 0.730]	Le Pen	0.939 [0.926; 0.951]
Macron	0.838 [0.809; 0.865]	FPÖ	0.965 [0.952; 0.977]	Macron	0.842 [0.822; 0.861]
Hamon	0.832 [0.805; 0.858]	Green	0.798 [0.771; 0.827]	Hildago	0.628 [0.601; 0.656]
Arthaud	0.669 [0.634; 0.701]	KPÖ	0.670 [0.638; 0.703]	Arthaud	0.610 [0.586; 0.635]
Poutou	0.704 [0.672; 0.736]	NEOS	0.578 [0.544; 0.615]	Poutou	0.746 [0.724; 0.769]
Cheminade	0.764 [0.734; 0.793]			Zemmour	0.967 [0.957; 0.976]
Lassalle	0.713 [0.681; 0.746]			Lassalle	0.675 [0.651; 0.699]
Mélenchon	0.843 [0.816; 0.870]			Mélenchon	0.873 [0.854; 0.890]
Asselineau	0.809 [0.779; 0.837]			Pecresse	0.860 [0.841; 0.878]
Fillon	0.932 [0.914; 0.949]			Jadot	0.727 [0.703; 0.751]
				Roussel	0.585 [0.556; 0.610]
Mean	.808		.718		.781

Note: Matching values indicate the degree of coincidence between reported approval ballots and cluster-analytically generated approval sets. Higher values denote stronger similarity. Numbers in brackets are bootstrap confidence intervals (95% level, percentile method, 1,000 replications).

Educational attainment can be regarded as a central variable, as it is linearly correlated with several others. Income and democracy satisfaction both increase with higher levels of education. Respondents with higher educational attainment also tend to position themselves somewhat further to the left on the ideological scale, although this effect is minimal. Moreover, educational attainment is associated with two effects in the rating data. First, the number of parties actually evaluated increases with education (viz., we observe fewer missing ratings among more highly educated respondents). Higher levels of education thus appear to facilitate, or at least render easier, the evaluation of the available parties.

A second noteworthy relation is that the propensity to assign extreme ratings (minimum or maximum values) decreases with higher levels of education, while respondents with higher educational level instead display a greater tendency to use the entire scale for their ratings. The coefficient of variation stands at 0.80 and 0.75 among the lower educational groups, compared to 0.71 and 0.69 among those with higher educational attainment.

We do not expect dichotomous structures to arise in cases where respondents rate parties as either very poor, very good, or—out of indifference—at the median value (hence exhibiting three focal points). In such instances, it is difficult to meaningfully divide the observations into exactly two groups. By contrast, the more differentiated the ratings, the less we observe focal points around a median rating value, the more we expect dichotomous structures in preferences. We therefore expect that the likelihood of observing quasi-dichotomous preferences increases with rising levels of education. On the other hand, we expect a rating behaviour with three focal points to be observable more frequently across individuals with higher dissatisfaction with democracy.

We estimate a series of logistic regression models in which QDP serves as the dependent variable. Specifically, $\mathcal{D}_{i,e}$ denotes the dichotomous-preference value for respondent i in election e (recall that we evaluate data from 172 elections), which takes the value $\mathcal{D}_{i,e} = 1$ if a respondent exhibits dichotomous preferences and zero otherwise. The main explanatory variable of interest is the respondent's educational level, denoted by `Educ.lvl`. To assess the robustness of this relationship, we extend the baseline specification by including the aforementioned covariates (age, gender, income, ideology, and satisfaction with democracy). Moreover, to account for unobserved heterogeneity across elections, we incorporate election-specific

Table 8: *OLS regression results.*

Dep. Var.: ϕ	Grenoble	Graz	France22
QDP	-0.15 (0.02)***	-0.03 (0.02)	-0.06 (0.01)***
Covariates:			
Age	-0.00 (0.00)	-0.02 (0.01)***	-0.00 (0.00)
Gender (Male)	0.02 (0.02)	0.02 (0.02)	0.02 (0.01)
Education Level	0.02 (0.03)	0.01 (0.01)	0.04 (0.02)
Intercept	0.67 (0.08)***	0.60 (0.04)***	0.55 (0.07)***
R^2	0.09	0.03	0.03
Adj. R^2	0.09	0.02	0.03
Num. obs.	692	756	1266

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: OLS regressions with Pearson's correlation coefficient (ϕ) as the dependent variable. ϕ serves as a proxy measure for coincidence between reported and analytically generated approval ballots. The main independent variable is a dummy for $\tilde{k} > 2$ (optimal number of clusters greater than two), with age, gender, and education included as controls. Negative coefficients indicate that respondents with more than two optimal clusters show lower coincidence between approval ballots and cluster-analytical approval sets.

fixed effects.

$$\Pr(\mathcal{D}_{i,e} = 1 \mid X_{i,e}, \alpha_e) = \frac{\exp\left(\beta_0+\beta_1 \, \texttt{Educ.lvl}_{i,e}+X'_{i,e}\gamma+\alpha_e\right)}{1+\exp\left(\beta_0+\beta_1 \, \texttt{Educ.lvl}_{i,e}+X'_{i,e}\gamma+\alpha_e\right)} \quad (5)$$

In the model, formally represented by Eq. (5), $\texttt{Educ.lvl}_{i,e}$ denotes the educational level of respondent i in election e. $X'_{i,e}\gamma$ denotes the inner product of the covariate vector $X_{i,e}$ and its associated coefficient vector γ. The covariate vector consists of additional control variables (or is empty). The term α_e captures the election-specific fixed effects. All coefficients will be presented as odds ratios $\left(\text{e.g., } e^{\beta}\right)$. An odds ratio below one indicates that an increase in the respective variable reduces the odds of exhibiting dichotomous preferences, while an odds ratio above one implies an increase in these odds. The confidence intervals are calculated with heteroscedasticity-robust standard errors.

Table 9 reports the regression results. In specification (1), the covariate vector is empty, while specifications (2) to (6) successively introduce additional controls. An odds ratio of 1.013 for educational attainment, as reported for model specification (1), implies that moving from a lower to a higher category of education (illiterate, lower secondary, upper secondary, post-secondary non-university, university) increases the odds of exhibiting quasi-dichotomous preferences by about 1.3 per cent, holding other factors constant.

A striking finding is the highly robust effect of educational level. The odds-ratio estimator exceeds unity in all model specifications and statistically significant at the 95 per cent level. We thus find consistent evidence of a slightly, but robustly positive association between educational attainment and dichotomous preference structures, in line with our prior expectations. A second effect arises from the variable measuring satisfaction with democracy. When included, it shows a statistically significant, albeit weak, association: lower satisfaction with democracy is linked to a higher likelihood of dichotomous preferences. This result can be explained by rating behaviour observed among respondents with high levels of dissatisfaction. Such individuals tend disproportionately to evaluate candidates either very positively, very negatively, or not at all. To illustrate with descriptive figures: respondents who report being dissatisfied with democracy assign, on average, twice as many best or worst ratings (i.e., lowest and highest possible rating values) compared with those who declare themselves satisfied. By contrast, none of the remaining covariates—age, gender, income, and ideology—show statistically significant effects. This strengthens the conclusion that educational level is the principal explanatory factor.

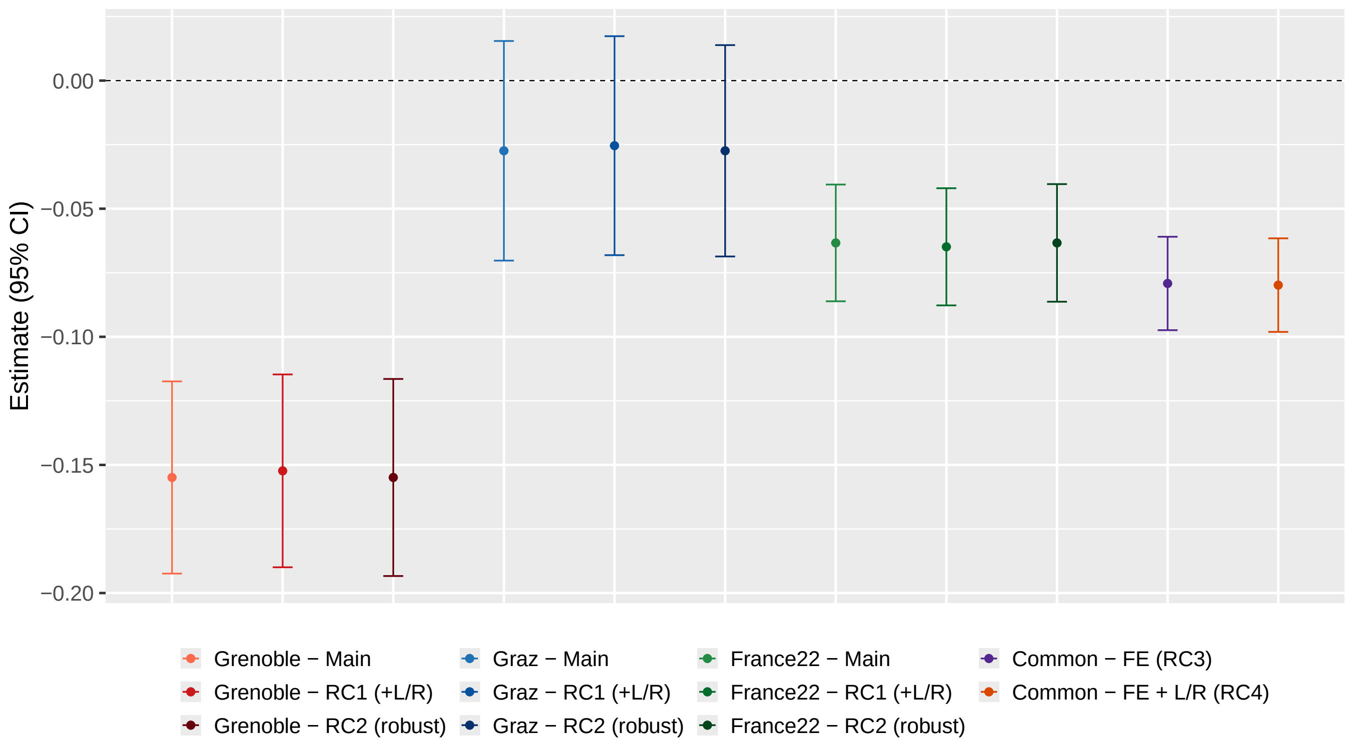

Figure 5: *OLS Regressions (Main model and robustness checks)*

Notes. Points show the estimated coefficient on the optimal number of cluster $\tilde{k} > 2$ from multiple specifications; vertical bars are 95% confidence intervals. Main estimates separately for *Grenoble*, *Graz*, and *France22*. RC1 augments the main specification with ideological dummies (+L/R). RC2 repeats the main specification but reports heteroscedasticity-robust standard errors. RC3 pools all datasets and estimates a fixed-effects (within) model with dataset fixed effects. RC4 adds the ideological dummies to RC3.

Table 9: *Logistic regression results with QDP as dependent variable. Data: CSES*

	(1)	(2)	(3)	(4)	(5)	(6)
Educ.lvl	1.013*	1.013*	1.012*	1.014*	1.012*	1.018*
	[1.003, 1.022]	[1.004, 1.023]	[1.002, 1.022]	[1.004, 1.023]	[1.002, 1.023]	[1.004, 1.032]
Satisfaction.Dem		0.988*			0.990*	0.981*
		[0.979, 0.998]			[0.980, 0.999]	[0.968, 0.994]
Ideology			1.002		1.002	1.003
			[0.998, 1.006]		[0.998, 1.007]	[0.997, 1.009]
Age				1.000		1.000
				[1.000, 1.001]		[0.999, 1.001]
Gender(Male)				1.014		1.017
				[0.993, 1.035]		[0.989, 1.047]
Income Qunitile						0.994
						[0.983, 1.005]
Num.Obs.	155544	150571	139967	153050	136261	83076
AIC	208222.0	201649.3	187336.3	204903.3	182455.7	110835.0
BIC	209964.1	203365.9	189040.2	206632.6	184135.4	112028.9

Note: Odds ratios are reported. Educational level shows a consistent and statistically significant positive association with dichotomous preferences across all model specifications (1) to (6). Satisfaction with democracy has a weak but significant effect, indicating that lower satisfaction is linked to a higher likelihood of dichotomous preferences. Age, gender, income, and ideology display no significant effects. Heteroscedasticity-robust standard errors are used to calculate the confidence intervals (exponentiated values, in square brackets). A '*' denotes that the respective confidence interval does not overlap unity.

The strong positive effect of educational attainment remains robust if we replace the variable capturing the self-assessed position on a ideological left-right $[0, 10]$-scale (`Ideology`) with a variable that captures the self-perceived distance to the ideological centre. Specifically, we replaced 'Ideology with $\zeta \equiv |\text{Ideology} - 5|$. Thus, ζ is high if someone self-declares to belong to the radical left or right. We re-run all six regressions and find that the odds ratio for the educational level attainment remain significantly positive across all model specification. The same applies when considering ζ^2 to stronger emphasize a distance to the ideological centre. However, ζ and ζ^2 both are statistically significant across the model specification with odds ratios consistently around 1.05. We regard these regressions more as a robustness check than as an alternative, because ζ is highly correlated to the dissatisfaction-with-democracy measure.

We obtain similar results when applying a closely related model specification to the three experimental datasets. In these cases, fewer covariates and, naturally, fewer observations are available. Age and gender are included as covariates, as these variables are jointly available and comparable across all three datasets. Educational attainment is also recorded in each dataset, though not perfectly comparable by construction (for instance, the Austrian education system differs from the French one). Nonetheless, it is possible to distinguish between lower and higher levels of education in all datasets and to code them accordingly.

We estimated five logistic regression models differing in the dependent variable to investigate the relationship between weakly and quasi-dichotomous preferences on one hand, and key socio-demographic factors on the other side. The first three models measure weak dichotomy using Theil's T, Gini, and the Atkinson index. The fourth model includes all respondents for whom the optimal number of clusters is two (denoted by $\tilde{k} = 2|all$). The fifth model employs quasi-dichotomy as the dependent variable. Given that the data originate from three different surveys, we included survey fixed effects. The results are presented in Table 10. Again, the reported values represent odds ratios, while the values in square brackets indicate the corresponding ranges of 95% confidence intervals.

The analysis reveals that educational attainment has a positive explanatory effect in the weak dichotomous preference models (WDP.Theil, WDP.Gini, WDP.Atkinson). This suggests that higher levels of education are associated with a greater likelihood of exhibiting weak dichotomous preferences. However, the effect size diminishes in the cluster-analytical models and loses statistical significance in the specification using the QDP (column '$\tilde{k} = 2|\text{strong}$').

9 Conclusion

This paper has revisited the widely used assumption of dichotomous individual preferences. Empirically, strictly dichotomous preferences are virtually absent—a finding that is hardly surprising.

For many applications, strictly dichotomous preferences are not required. It is sufficient to view them as an approximation of actual preferences, though the nature of this approximation has often remained vague. We contend that the threshold approach–the standard approximation–is better seen as an indicator of decision consistency than as a model of dichotomous structures. This paper advances two alternative definitions of dichotomous preferences, each relaxing strict dichotomy in a different and independent way. Both are empirically testable, and we demonstrate how data can be connected to these concepts.

Many theoretical contributions rely on simplified representations of complex preferences. A central finding of our analysis is that dividing preferences into two groups provides a closer approximation than representations with three or more groups. Depending on the method, weakened forms of dichotomous preferences plausibly describe between 50 and 70 per cent of respondents' ratings.

There is mounting evidence that a significant share of individuals exhibit preferences that do not comply with the fundamental transitivity requirement (Anand, 1993), particularly when they are confronted with complex, multi-attribute alternatives (Fishburn, 1982), as is the case in political elections. Given that even transitivity often fails to hold, this result may be seen in the proper light. Moreover, it lies in the very nature of normative assumptions and models that they capture only part of observable reality—after all, one would not reject a model simply because, within a regression analysis, its R^2 falls short of unity.

If, however, key results in normative work pivot on the assumption of strict dichotomous preferences, the evidence presented here poses a genuine challenge. In microeconomic theory it has been shown that central economic results remain robust when the assumption of transitivity is slightly relaxed (Gale and Mas-Colell, 1975). It may therefore be fruitful to examine the extent to which models with dichotomous structures retain their robustness once weakened forms of dichotomous preferences are taken into account.

In recent years, experimental studies have become increasingly prominent and have attracted considerable attention. Most of this work examines behaviour under approval voting and the resulting electoral outcomes. Such studies typically generate complex data that permit inferences about underlying preference orders, thus offering richer information than the approval ballots usually collected alongside them. This paper encourages future experimental research to focus more closely on these underlying preference structures. Outcomes may differ substantially between respondents with weakly or quasi-dichotomous preferences and those without such patterns. Analysing these differences would deepen our understanding of how sensitive experimental findings are to the nature of the underlying preferences.

We observe in 10 to 20 per cent of respondents that their approval ballots do not comply with the threshold approach, which is commonly regarded as a fundamental requirement of choice consistency (Terzopoulou et al., 2025). Since Tversky (1969), it has been known that respondents often wish to revise their answers once confronted with intransitivities in their own judgements. Our findings, together with this early research on intransitivity, suggest that future experiments should give respondents the opportunity to revise their assignments into approval and disapproval ballots in case it conflicts with their ratings.

Table 10: *Dichotomy and socio-demographic factors (Experiment-data)*

	(1) WDP.Theil	(2) WDP.Gini	(3) WDP.Atkinson	(4) $\tilde{k} = 2$|all	(5) $\tilde{k} = 2$|strong
Educ.lvl	1.18*	1.15*	1.18*	1.07*	0.98
	[1.01; 1.31]	[1.05; 1.26]	[1.07; 1.30]	[1.025; 1.18]	[0.89; 1.07]
Age	0.85*	0.87	0.86	0.99	1.04
	[0.73; 0.98]	[0.71; 1.06]	[0.74; 1.009]	[0.91; 1.08]	[0.98; 1.10]
Gender	1.23*	1.15	1.20	1.19*	0.96
(Male)	[1.09; 1.39]	[0.98; 1.34]	[1.01; 1.43]	[1.04; 1.36]	[0.91; 1.03]
Num. obs.	3120	2995	3120	3195	3195
AIC	3861.03	3972.73	3978.19	4304.78	3973.65
BIC	3897.30	4008.76	4014.46	4341.20	4010.07

* Null hypothesis value outside the confidence interval.

Note: Logistic regression models with different operationalisations of dichotomous preferences as dependent variables. Models (1)–(3) use weak dichotomy measured by Theil's T, Gini, and the Atkinson index. Model (4) includes all respondents with an optimal number of clusters $\tilde{k} = 2$, while Model (5) employs the quasi-dichotomy (QDP). Explanatory variables are age, gender (male), and educational attainment. Coefficients are reported as odds ratios with 95% confidence intervals in square brackets.

It would be highly interesting to see whether such an revision option leads to results that differ from those presented here.

In this paper, we have employed several empirical approaches, both to ensure robustness and to illustrate the breadth of available methods. The choice of method should always depend on the research question at hand and, above all, on the characteristics of the data. To facilitate further work, we provide replication files and invite future research to build on them—whether to extend, refine, or challenge our findings. In this way, we hope to advance a more precise understanding of dichotomous preferences and their role in the analysis of social choice.

References

Alós-Ferrer, C., Granić, Đ.G., 2012. Two field experiments on Approval Voting in Germany. Social Choice and Welfare 39, 171–205. doi:10.1007/s00355-011-0550-5.

Alós-Ferrer, C., Garagnani, M., García-Segarra, J., 2025. The framing of elections: cooperation vs. competition. Social Choice and Welfare 65, 363–385. doi:10.1007/s00355-024-01580-9.

Anand, P., 1993. The philosophy of intransitive preference. The Economic Journal 103, 337–346. doi:10.2307/2234772.

Atkinson, A.B., 1970. On the measurement of inequality. Journal of Economic Theory 2, 244–263. doi:10.1016/0022-0531(70)90039-6.

Barbaro, S., 2025. Election methods and political polarization. Mathematical Social Sciences 135, 102417. doi:10.1016/j.mathsocsci.2025.102417.

Batool, F., Hennig, C., 2021. Clustering with the average silhouette width. Computational Statistics & Data Analysis 158, 107190. doi:10.1016/j.csda.2021.107190.

Baujard, A., Igersheim, H., Lebon, I., 2020. Some regrettable grading scale effects under different versions of evaluative voting. Social Choice and Welfare 56, 803–834. doi:10.1007/s00355-020-01300-z.

Baujard, A., Igersheim, H., Lebon, I., Gavrel, F., Laslier, J.F., 2014. Who's favored by evaluative voting? An experiment conducted during the 2012 French presidential election. Electoral Studies 34, 131–145. doi:10.1016/j.electstud.2013.11.003.

Bhattacharya, N., Mahalanobis, B., 1967. Regional disparities in household consumption in India. Journal of the American Statistical Association 62, 143–161. doi:10.2307/2282918.

Bogomolnaia, A., Moulin, H., Stong, R., 2005. Collective choice under dichotomous preferences. Journal of Economic Theory 122, 165–184. doi:10.1016/j.jet.2004.05.005.

Bouveret, S., Blanch, R., Boujard, A., Durand, F., Igersheim, H., Lang, J., Laruelle, A., Laslier, J.F., Lebon, I., Merlin, V., 2019. Voter autrement 2017 for the French presidential election. The in situ experiments voting material. doi:10.5281/ZENODO.3548303.

Brams, S.J., Fishburn, P.C., 1978. Approval voting. American Political Science Review 72, 831–847. doi:10.2307/1955105.

Brams, S.J., Fishburn, P.C., 2007. Approval Voting. Second ed., Springer, New York, NY.

Brams, S.J., Fishburn, P.C., Merrill, S., 1988. The responsiveness of Approval Voting: Comments on Saari and van Newenhizen. Public Choice 59, 121–131. doi:10.1007/bf00054448.

Brams, S.J., Kilgour, D.M., Potthoff, R.F., 2018. Multiwinner approval voting: An apportionment approach. Public Choice 178, 67–93. doi:10.1007/s11127-018-0609-2.

Brandl, F., Brandt, F., 2020. Arrovian aggregation of convex preferences. Econometrica 88, 799–844. doi:10.3982/ecta15749.

Brandl, F., Peters, D., 2022. Approval voting under dichotomous preferences: A catalogue of characterizations. Journal of Economic Theory 205, 105532. doi:10.1016/j.jet.2022.105532.

Brill, M., Gölz, P., Peters, D., Schmidt-Kraepelin, U., Wilker, K., 2022. Approval-based apportionment. Mathematical Programming 203, 77–105. doi:10.1007/s10107-022-01852-1.

Brill, M., Laslier, J.F., Skowron, P., 2018. Multiwinner approval rules as apportionment methods. Journal of Theoretical Politics 30, 358–382. doi:10.1177/0951629818775518.

Costa, M., 2016. Overlapping component and inequality decomposition: a simulation study for the Gini index. METRON 74, 193–205. doi:10.1007/s40300-016-0092-9.

Cowell, F.A., Victoria-Feser, M.P., 1996. Robustness properties of inequality measures. Econometrica 64, 77–101. doi:10.2307/2171925.

CSES, 2024. CSES integrated module dataset (IMD) [dataset and documentation]. February 27, 2024 version. The Comparative Study of Electoral Systems. doi:10.7804/cses.imd.2024-02-27.

Darmann, A., Klamler, C., 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, 63–87. doi:https://doi.org/10.1007/s11127-023-01071-y.

Dehez, P., Ginsburgh, V., 2019. Approval voting and Shapley ranking. Public Choice 184, 415–428. doi:10.1007/s11127-019-00729-w.

Delemazure, T., Bouveret, S., 2024. Voter Autrement 2022 - the online experiment ("un autre vote"). doi:10.5281/ZENODO.10568798.

Dunn, J.C., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics 3, 32–57. doi:10.1080/01969727308546046.

Ebert, U., 1988. Measurement of inequality: An attempt at unification and generalization. Social Choice and Welfare 5, 147–169. doi:10.1007/bf00735758.

Ebert, U., 2010. The decomposition of inequality reconsidered: Weakly decomposable measures. Mathematical Social Sciences 60, 94–103. doi:10.1016/j.mathsocsci.2010.05.001.

Elkind, E., Lackner, M., 2016. Structure in dichotomous preferences. URL: https://arxiv.org/abs/1505.00341, arXiv:1505.00341.

Falmagne, J.C., Regenwetter, M., 1996. A random utility model for Approval Voting. Journal of Mathematical Psychology 40, 152–159. doi:10.1006/jmps.1996.0014.

Fishburn, P.C., 1982. Nontransitive measurable utility. Journal of Mathematical Psychology 26, 31–67. doi:10.1016/0022-2496(82)90034-7.

Fleurbaey, M., Lambert, P., Moramarco, D., Peragine, V., 2025. Inequality decomposition analysis, the Lorenz curve and the Gini coefficient. Social Choice and Welfare doi:10.1007/s00355-025-01612-y.

Foster, J.E., Shneyerov, A.A., 2000. Path independent inequality measures. Journal of Economic Theory 91, 199–222. doi:10.1006/jeth.1999.2565.

Gale, D., Mas-Colell, A., 1975. An equilibrium existence theorem for a general model without ordered preferences. Journal of Mathematical Economics 2, 9–15. doi:10.1016/0304-4068(75)90009-9.

Gehrlein, W.V., 2005. Probabilities of election outcomes with two parameters: The relative impact of unifying and polarizing candidates. Review of Economic Design 9, 317–336. doi:10.1007/s10058-005-0132-z.

Gehrlein, W.V., Lepelley, D., 1998. The Condorcet efficiency of Approval Voting and the probability of electing the Condorcet loser. Journal of Mathematical Economics 29, 271–283. doi:10.1016/s0304-4068(97)00020-7.

Gehrlein, W.V., Lepelley, D., Moyouwou, I., 2016. A note on Approval Voting and electing the Condorcet loser. Mathematical Social Sciences 80, 115–122. doi:10.1016/j.mathsocsci.2016.02.009.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108. doi:10.2307/2346830.

Inada, K.i., 1964. A note on the simple majority decision rule. Econometrica 32, 525–531. doi:10.2307/1910176.

Inada, K.i., 1969. The simple majority decision rule. Econometrica 37, 490–506. doi:10.2307/1912796.

Ju, B.G., 2010. Collective choice for simple preferences, in: Laslier, J., Sanver, M.R. (Eds.), Handbook on Approval Voting. Springer Berlin Heidelberg, pp. 41–90. doi:10.1007/978-3-642-02839-7.

Kaufman, L., Rousseeuw, P.J., 2005. Finding groups in data. Wiley, Malden, MA.

Komatsu, H., 2024. Characterizations of approval ranking. Mathematical Social Sciences 128, 18–24. doi:10.1016/j.mathsocsci.2024.01.001.

Lachat, R., Laslier, J.F., 2024. Alternatives to plurality rule for single-winner elections: When do they make a difference? European Journal of Political Economy 81, 102505. doi:10.1016/j.ejpoleco.2024.102505.

Laslier, J.F., Van der Straeten, K., 2008. A live experiment on Approval Voting. Experimental Economics 11, 97–105. doi:10.1007/s10683-006-9149-6.

Lerman, R.I., Yitzhaki, S., 1989. Improving the accuracy of estimates of Gini coefficients. Journal of Econometrics 42, 43–47. doi:10.1016/0304-4076(89)90074-2.

Maniquet, F., Mongin, P., 2015. Approval voting and Arrow's impossibility theorem. Social Choice and Welfare 44, 519–532. doi:10.1007/s00355-014-0847-2.

Manjunath, V., Westkamp, A., 2021. Strategy-proof exchange under trichotomous preferences. Journal of Economic Theory 193, 105197. doi:10.1016/j.jet.2021.105197.

Morrisson, C., Murtin, F., 2012. The Kuznets curve of human capital inequality: 1870–2010. The Journal of Economic Inequality 11, 283–301. doi:10.1007/s10888-012-9227-2.

Niemi, R.G., 1984. The problem of strategic behavior under Approval Voting. American Political Science Review 78, 952–958. doi:10.2307/1955800.

Regenwetter, M., Grofman, B., 1998. Choosing subsets: a size-independent probabilistic model and the quest for a social welfare ordering. Social Choice and Welfare 15, 423–443. doi:10.1007/s003550050115.

Regenwetter, M., Grofman, B., Marley, A., 2002. On the model dependence of majority preference relations reconstructed from ballot or survey data. Mathematical Social Sciences 43, 451–466. doi:10.1016/s0165-4896(02)00022-7.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65. doi:10.1016/0377-0427(87)90125-7.

Saari, D.G., Van Newenhizen, J., 1988. The problem of indeterminacy in approval, multiple, and truncated voting systems. Public Choice 59, 101–120. doi:10.1007/bf00054447.

Sato, N., 2019. Approval voting and fixed electorate with dichotomous preferences. Mathematical Social Sciences 97, 51–60. doi:10.1016/j.mathsocsci.2018.10.001.

Shorrocks, A.F., 1980. The class of additively decomposable inequality measures. Econometrica 48, 613–625. doi:10.2307/1913126.

Skowron, P., Lackner, M., Brill, M., Peters, D., Elkind, E., 2016. Proportional rankings. URL: arxiv.org/abs/1612.01434, arXiv:1612.01434.

Terzopoulou, Z., Lang, J., Zwicker, W.S., 2025. Approval compatible voting rules. Social Choice and Welfare doi:10.1007/s00355-025-01588-9.

Tversky, A., 1969. Intransitivity of preferences. Psychological Review 76, 31–48. doi:10.1037/h0026750.

Vorsatz, M., 2007. Approval voting on dichotomous preferences. Social Choice and Welfare 28, 127–141. doi:10.1007/s00355-006-0149-4.

Vorsatz, M., 2008. Scoring rules on dichotomous preferences. Social Choice and Welfare 31, 151–162. doi:10.1007/s00355-007-0270-z.

Yitzhaki, S., 1994. Economic distance and overlapping of distributions. Journal of Econometrics 61, 147–159. doi:10.1016/0304-4076(94)90081-7.

Yitzhaki, S., Lerman, R.I., 1991. Income stratification and income inequality. Review of Income and Wealth 37, 313–329. doi:10.1111/j.1475-4991.1991.tb00374.x.