

Democracy and Intensity of Preferences

A Test of Storable Votes and Quadratic Voting on Four California

Propositions

Alessandra Casella*

Luis Sanchez†

Short title: Storable Votes and Quadratic Voting

*Alessandra Casella is the corresponding author and can be reached at ac186@columbia.edu.

†Luis Sanchez can be reached at las497@cornell.edu.

Abstract

Can direct democracy overcome the "Problem of Intensity", treating everyone equally and yet allowing an intense minority to prevail if, but only if, the majority's preferences are weak? Storable Votes (SV) and Quadratic Voting (QV) propose possible solutions. We test their performance in two samples of California residents using data on four initiatives prepared for the 2016 California ballot. To estimate the probable effects of the two systems, we bootstrap the original samples and generate two sets of 10,000 multi-elections simulations. Both systems induce minority victories in 30-35 percent of the samples and yet result in higher expected welfare relative to majority voting, appropriating between one third and two thirds of the remaining surplus.

Keywords: Intensity of preferences; minority, direct democracy, majority voting.

Supplementary material for this article is available in the appendix in the online edition. Replication files are available in the JOP Data Archive on Dataverse. The archive can be found at <http://thedata.harvard.edu/dvn/dv/jop>. The research was approved by Columbia University Institutional Review Board. Support for this research was provided by National Science Foundation grant SES-0617934.

In *Preface to Democratic Theory* (1956), Dahl discussed the "Problem of Intensity" as a fundamental challenge to majoritarian principles: on both ethical and pragmatic grounds, an intense minority should be able to prevail over an indifferent majority, but only if the minority is indeed intense and the majority indeed indifferent. Dahl concluded that no such provision was realistically available, because of the difficulty of observing intensities, of the challenge of deciding when the case applies, of the different objectives of American constitutional rules. The "radical majoritarianism" of today's populism (Urbinati 2019) raises new concerns on how to protect the legitimate voice of the minority at the voting booth.

Two recent proposals suggest that a possible answer lies in the voting rule. Imagine a voter faced with multiple referenda, as indeed is typically the case in many US states. Each referendum can either pass or fail and is decided according to the majority of votes cast; all voters are treated equally and given the same number of total votes. The only deviation from usual voting practices is that voters choose how many votes to cast on any individual referendum, out of the total at their disposal. The number of votes becomes a measure of intensity, and the minority can indeed prevail, but only over those decisions on which it feels strongly and the majority feels weakly. Storable Votes (SV) (Casella 2012) work exactly as described; Quadratic Voting (QV) (Goeree and Zhang 2017, Lalley and Weyl 2018) imposes a penalty on concentrating votes: "effective" votes cast on any one referendum equal the square root of the number of original votes dedicated to the referendum by the voter.

Both proposals have been studied theoretically, in the laboratory, in simulations, and in opinion polls, and both have been found promising. But their final properties depend on the distribution of preferences in the electorate, and the two schemes have never been tested in the context of actual political decisions. In this paper we report on the results of an incentivized survey that applies SV and QV to four actual initiatives in California.¹

¹Summaries of the theory behind the two voting schemes, as well as a brief overview of the literature, additional details on the experiment, and screenshots of the survey can all be found in the appendix.

1 The Survey

In May 2016 we selected four propositions that were being prepared for inclusion in the November 2016 California ballot:

(1) *Bilingual education (BE)*: re-instate the possibility of bilingual classes in public schools. (The proposition was included in the November 2016 ballot and passed.)

(2) *Immigration (IM)*: require all state law enforcement officials to verify immigration status in case of an infraction and report undocumented immigrants to federal authorities. (The proposition was not included in the final ballot.)

(3) *Teachers' tenure (TT)*: increase required pre-tenure experience for teachers from two to five years. (The proposition was not included in the ballot.)

(4) *Public Vote on Bonds (PB)*: require voters' approval for all public infrastructure projects of more than \$2 billion. (The proposition was included in the ballot and failed.)

We then recruited 647 California subjects via Amazon Mechanical Turk (MTurk). We first asked each subject how (s)he would vote on each of the four propositions, presented in random order, allowing for the option to abstain. Answers to this part of the survey allowed us to compute outcomes under majority voting. We then elicited measures of intensity of preferences. Each subject was asked to distribute 100 points among the four propositions, with the number of points used as scale of the importance attributed to each proposal ("How important is this issue to you?"). We used examples to clarify that importance is independent of whether the respondent is in favor or against a proposition, and summarized responses in terms of priorities, allowing for revisions and asking for a final confirmation.²

After this first part of the survey, common to all respondents, subjects were randomly assigned either to the SV treatment (324 subjects; 306 after data cleaning) or to the QV treatment (323 subjects; 313 after cleaning). We used two simplified versions of SV and

²The report was not incentivized. The simplest procedure—a bonus proportional to the value attributed to propositions in which the subject is on the winning side—distorts replies towards least contentious propositions. We concluded that incentive compatible methods would be too cumbersome for MTurk.

QV, well-suited to practical implementation in large electorates. In the SV treatment, we exploited theoretical results showing that, in large elections, the optimal SV design couples regular votes, one per election, with a single *bonus vote* to be cast as desired (Casella and Gelman 2008). Thus subjects in the SV sample were told that each was granted one extra vote, in addition to the regular votes cast earlier, and were asked to choose the proposition in which to use it. The vote was cast in the direction indicated in the first part of the survey, and the final outcome under SV was calculated summing regular and bonus votes.

The design of the QV scheme required some innovation. Existing opinion surveys using QV rely on proprietary software as well as a training video (Quarfoot et al. 2017). We chose instead to simplify the QV scheme. We asked respondents to choose one of four classes of votes, distinguished by color and weight. Blue votes are regular votes, four in number; a person choosing blue votes casts one vote on each proposition. Green votes are only three, but each is worth more than a regular blue vote and beats a blue vote if the two are opposed. Yellow votes are two, each stronger than a green vote. Finally, a subject can choose to cast a single red vote, stronger than a yellow vote. The weights we assigned to the different votes are 1 for blue votes, 1.2 for green votes, 1.5 for yellow votes, and 2 for the red vote. A subject who chooses green/yellow/red votes casts votes on only three/two/one proposition(s). The simple four-class classification respects the convex cost of concentrating votes at the heart of QV. A voter casting votes on all four propositions—choosing blue votes—has a total weight of 4, but the total weight declines as votes are concentrated: the total weight corresponding to the three green votes is 3.6, to the two yellow votes is 3, and to the single red vote is 2. The decline is increasing with concentration, and increasing at an increasing rate, capturing the core feature of QV.

We asked each subject to choose a class of votes, and then select the proposition(s) on which to cast the vote(s). As with SV, votes were then cast automatically according to the preferences indicated in the first part of the survey. The final outcome was calculated on the basis of the QV votes cast.

Under both voting systems and in all calculations we report, ties were resolved randomly. Outcomes were computed using simple majority, and either SV or QV. We incentivized voting choices by promising \$250 for an organization working in favor of any proposal that passed under either SV or QV, depending on the sample.

Our performance criterion is an empirical approximation to utilitarian welfare, based on the points assigned by respondents to each proposition at the end of the first part of the survey. Allocating points within a common budget encourages truthfulness in reporting preferences, and prevents factors of scale from distorting welfare.³ Denoting by b_{ik} the number of points attributed to proposition k by individual i , we define aggregate welfare W^S as $W^S = \sum_i \sum_{k:i \in M_k^S} b_{ik}$, where $S \in \{majority\ voting, SV, QV\}$ indicates the voting scheme, and M_k^S the side casting the majority of votes on k under S . Points are interpreted as proxy for intensity—or more precisely as proportional to the value attributed to winning a proposition over losing it. The measure W^S reports how well the outcome of a proposition mirrors the aggregate intensity on the two sides of each proposition. Utilitarian efficiency requires that each proposition be won by the side which collectively values it most, or $W^* = \sum_i \sum_{k:i \in B_k} b_{ik}$ where B_k denotes the side with higher total points on k . Thus a voting scheme S resolves disagreement over proposition k efficiently if $M_k^S = B_k$, i.e. if the winning side under scheme S is also the side with higher total intensity (higher total points).

If the two opposite sides attribute similar aggregate values to a proposition, any outcome for that proposition is close to efficient. To control for this, we normalize the welfare measures by a floor corresponding to expected welfare under random decision making, where either side of any proposition has equal probability of winning: $R = \sum_i \sum_k b_{ik}/2$. For each voting scheme S , we call the ratio $(W^S - R)/(W^* - R)$ S 's *realized share of surplus* and use it as our primary performance measure.

³Relative utilitarianism, such that each individual's preferences are normalized to range between 0 and 1, uniquely satisfies desirable axioms when intensities matter (Dhillon and Mertens, 1999). It is fragile to misreporting, a problem alleviated in part by the imposition of a common budget. In our approximation, intensities can only be read as relative to the four initiatives included.

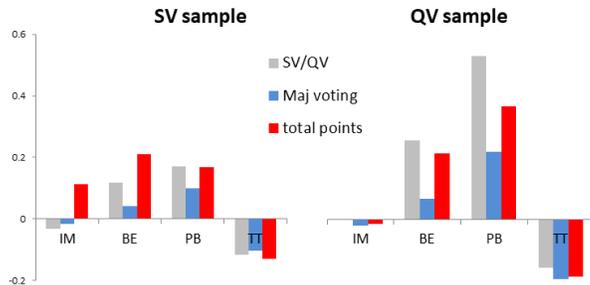


Figure 1: *Margins in favor*. Two-sided KS tests bootstrapped to correct for discreteness and assessing whether the distributions of points are drawn from the same population yield p-values equal to 0.629 (IM), 0.66 (BE), 0.092 (PB), 0.384 (TT).

We reproduce in the appendix the histograms of respondents’ intensities over each proposition, distinguishing supporters and opponents, as well as detailed information on voting choices. In SV, the bonus vote was primarily but not exclusively cast in the proposal to which the respondent attributed highest value (74% of subjects did so). In QV, a full 40% of subjects chose the Red class; the corresponding share recommended by the theory, given reported preferences, is 24%; the disparity reflects the respondents’ bias towards vote classes with fewer, heavier votes.

Figure 1 summarizes preferences and voting choices by reporting percentage margins in favor of each proposition, in terms of number of votes (under either SV or QV), number of voters (majority voting), and aggregate points. In both samples, a majority of respondents is in favor of BE and PB and against TT and IM, although the margin in the IM proposition is very small. In both samples and all propositions, the outcome is unchanged whether using majority voting, SV, or QV. When the margins under the three voting schemes have the same sign as the aggregate point margin, all three schemes deliver the utilitarian-efficient outcome. Thus both majority voting and QV appropriate the full surplus in the QV sample, while both majority and SV fall short in the SV sample because of the IM proposition.

The IM proposition stands out under several dimensions. It is the most contested. Although it fails in both samples and with all three voting systems, it always does so with very small vote margins: the vote tallies under majority are 129 to 125 (SV sample) and 136 to

130 (QV sample); under SV the tally is 181 to 170, and under QV a bare 124.6 to 124.4. It is also the most salient: it receives the highest number of total points in both samples, the highest number of bonus votes in the SV sample, and the highest number of red votes and of total votes in the QV sample.

Two simple statistical models summarize voting choices and help to classify behavior. With SV, we conjecture that each subject follows one of four mutually exclusive behaviors: with probability p_{Max} , the bonus vote is cast on the subject's highest value proposition; with probability p_{Close} on the one with closest outcome (IM); with probability p_{Fam} on the most familiar (either BE or TT, the two education propositions, with equal probability); and with probability p_{Rand} according to some other criterion that appears to us fully (uniformly) random. Each choice observed in the data, matched with the individual's reported valuations, can be expressed as function of the four probabilities. We report the estimated probabilities, as well as the estimation details, in the Appendix. We find that the bonus vote is cast primarily on the proposition that the voter considers her highest priority, but a relatively large role is left for randomness.

With QV, we posit a noisy two-step process that starts with choosing a vote class and then, given the class, proceeds to casting the vote(s). The choice of vote class depends on relative intensities: if one proposition is attached much higher value than any of the others, the voter chooses the red vote; if such is not the case but two propositions are attached much higher value than the remaining two, the voter chooses yellow votes, etc. The statistical model estimates the thresholds guiding the choice of vote class and the probability with which such a reasoning is followed (as opposed to choosing the class randomly), and finally, given the vote class, the probability that votes are cast monotonically, i.e. on the highest intensity propositions. Detailed results are in the Appendix. We find a strong tendency towards cumulating votes and frequent deviations from the threshold rules estimated by the model.⁴ Given the vote class, however, the tendency towards monotonicity is strong.

⁴As described in the appendix, a full 40 percent of subjects choose the single red vote.

2 Results

On all four propositions, both SV and QV confirmed the outcome reached with simple majority voting. The result, however, is not very informative: because the votes cast across propositions are tied by a budget constraint, each sample reduces to a single data point. To evaluate the potential impact of SV and QV, we would want to replicate the same elections many times, with different electorates all drawn from the same population distribution. We cannot rerun the elections, but, as in Casella 2012 (ch. 6), we can approximate such iterations by bootstrapping our data.

The objective is to estimate the impact of the voting rules in a population for which our samples are representative. The maintained assumption is that preferences are independent across individuals, but not necessarily across propositions for a single individual. We sample with replacement N individuals from each of our datasets, where $N = 306$ for SV and $N = 313$ for QV. For each individual, we sample the direction of preferences over each proposition, the number of points assigned to each, and the votes cast according to either the SV or the QV scheme. We replicate this procedure 10,000 times for each original dataset, SV or QV. A replication generates a distribution of preferences over each proposition and a voting decision for all N voters, and thus a voting outcome for all four propositions. The focus is on the fraction of simulations in which the two voting systems reach different results from majority voting, and on their welfare properties.

Simulations allow to investigate different rules-of-thumb governing the use of the votes, and thus to evaluate the robustness of the voting schemes to a range of plausible behaviors. Matching individuals with their SV or QV choices is the obvious option, and the first one we consider—we call it rule A. Individual choices, however, tend to be noisy, and the statistical models organize the data into more systematic behavior. Given respondents' allocation of points and voting scheme, we implement the voting choices by having each drawn subject behave according to the estimates of the statistical model—we call it rule B. If individuals had voted as theory says is optimal, welfare gains would have been maximal. We simulate such

behavior under a rule we call C.⁵ The distance in the results under A and C gives us a measure of impact of the deviations from optimal behavior. One possibility is that such deviations are purely random—we simulate a fourth rule, which we call D, such that individuals behave either optimally or randomly, with equal probability. The distance between B and D teaches us whether the deviations from optimality are random or have systematic patterns.

With all four rules, both SV and QV resulted in frequent minority victories (Figure 2:A). More than one fourth of the 10,000 simulations in each of the two data sets, using any rule, had at least one minority victory: the average across rules was 30% for QV and 35% for SV. Remarkably, under all four rules both voting systems consistently delivered welfare gains over majority voting, and this even though majority voting works well in these data, especially in the QV samples. Averaging across rules and simulations, the realized share of surplus was 85% for SV and 98% for QV, compared to 71% and 94% for majority in the two sets of simulations (Figure 2:B). Focusing on rule A, SV appropriated about one third of the surplus left on the table by majority (29%); QV, about two thirds (64%). However, many minority victories also came with welfare losses. Averaging across all rules, SV causes welfare losses in 11% of all simulations in which it delivers at least one minority victory, while the percentage rises to 31% for QV (Figure 2:C). Under rule A, the numbers are 28% for SV and 36% for QV.⁶

Reporting the realized share of surplus over all simulations (Figure 2:B), whether or not any outcome differs from simple majority, gives weight not only to realized but also to foregone efficiency gains—to minority victories that would have been efficient but did not occur. But only a fraction of simulations include a minority victory, and within each sample none of the expected surplus measures are statistically different from one another.

In the SV simulations, the outlier is rule A, which implements the actual voting choice indicated by the subject drawn in the simulation (Figure 2:C). The problem comes from the

⁵Rule C is optimal under some simplifying assumptions. Under SV, it corresponds to casting the bonus vote on the highest intensity proposition; under QV the optimal number of votes is proportional to the voter's intensities. See the Appendix for details.

⁶See the appendix for the numerical values corresponding to Figure 2, as well as Figure 2 below.

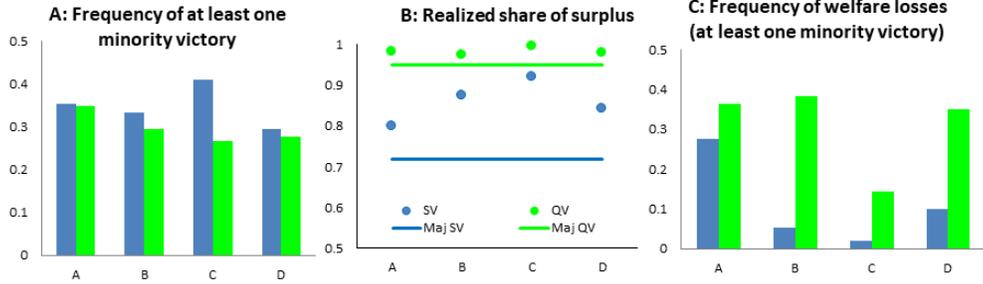


Figure 2: *Bootstrap results.* In panel A, the frequency is significantly higher for SV under rules B, C, and D ($p < 0.01$; one-sided Z test). In panel B, none of the differences in means are statistically different from one another. Note the difference in surplus under majority in the SV and QV samples. In panel C, all frequencies are significantly positive, but smaller than 0.05 for SV-C (0.019) and SV-B (0.047).

IM proposition, where bonus votes are predominantly cast against the proposition, while high points are predominantly attributed by subjects in favor. However, the asymmetry in behavior concerns a small number of subjects, and may reflect pure noise (see the appendix).

The difference in performance between SV and QV is largely driven by the different potential for improvement over majority voting (Figure 2:B). The two samples were populated randomly during the MTurk survey, and the point distributions over IM are statistically similar (see the legend of Figure 1). The discrepancy reflects small sampling noise when the distribution of values is symmetric, as is the case for IM in both samples. To compare SV and QV directly, we can combine the two MTurk samples. We lose the ability to evaluate the voting schemes according to rule A, since only the SV (QV) sample was exposed to SV (QV), but we can simulate voting behavior according to rules B, C and D.

Three regularities emerge clearly. First, QV results in a consistently higher fraction of minority victories than SV (Figure 2:A): averaging across rules, 34% of QV simulations have at least one minority victory, vs. 18% for SV. Second, under any rule, both voting systems continue to appropriate a higher share of surplus than majority does (Figure 2:B). QV captures 97% of surplus on average, and SV 94% (vs. 89% with majority). Third, the frequency of welfare losses induced by QV and SV becomes more similar (Figure 2:C). Averaging over the three rules, 13% of all simulations with at least one minority victory

Simulations based on the two samples, joint.

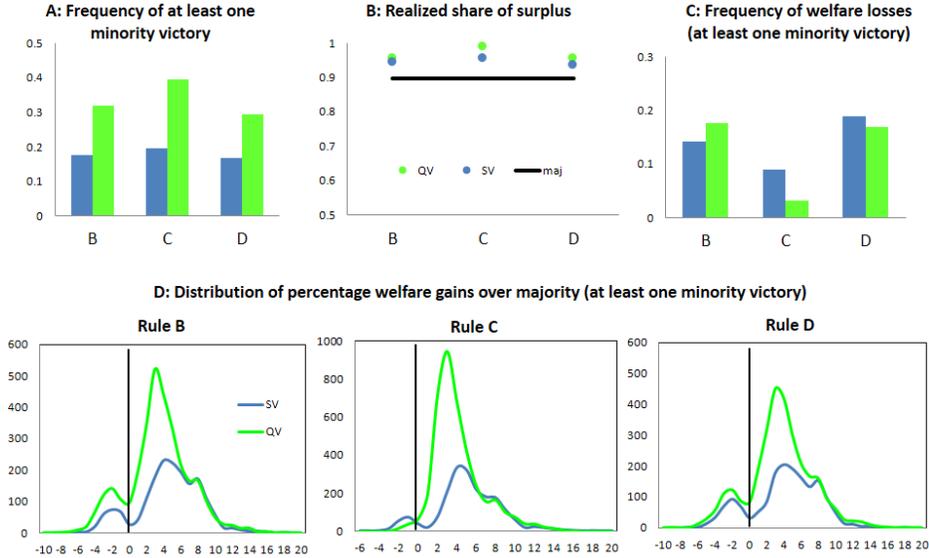


Figure 3: *Joining the two samples*. A: the fraction of samples with at least one minority victory is consistently higher under QV ($p < 0.001$). B: none of the differences in means are statistically different from one another. C: The frequency of welfare losses is significantly higher under QV for rule B, but significantly lower for rules C and D ($p < 0.01$ for rules B and C, $p = 0.046$ for D). All frequencies are significantly positive but less than 0.10 only for QV-C (0.03) and SV-C (0.09) (One-sided Z tests). D: the vertical axis is absolute numbers; the horizontal axis are percentage gains, and the vertical line is at 0.

induce welfare losses in QV, and 14% in SV.

The defining difference between the two voting schemes is the high frequency of minority victories under QV at positive but small welfare gains (Figure 2:D).⁷ Does this higher sensitivity reflect the specific parametrization we have implemented? As we show in the appendix, QV behaves better in our data if it is complemented by regular votes—votes that must be cast one on each initiative and that move the voting scheme towards majority voting. This is advantageous under QV because respondents select higher weight vote classes much more often than theory prescribes. Adding regular votes to QV also makes it closer to SV, however, inviting questions on the trade-off between complexity and surplus gains.

Finally, our data allow us to construct measures not only of aggregate surplus but also of inequality, defined as ex post disparity in the number of propositions on which each voter

⁷However, a puzzling aspect of our data is the weak performance of SV-A.

is on the winning side, weighted by the importance (the number of points) the voter assigns to each. An important tenet of democracy is that the composition of winning coalitions shifts across issues, ensuring that no group is disenfranchised. We show in the appendix that under this dimension too in our data SV and QV perform well: because, ceteris paribus, both voting schemes increase the probability of being on the winning side on issues the voter considers higher priorities, ex post inequality is reduced.

Our study confirms the theoretical promise of the two voting schemes. SV and QV allow for occasional minority victories on those issues over which the minority's intensity of preferences is sufficiently stronger than the majority's to make a minority victory desirable.

Acknowledgements

We thank Bora Erdamar, Andrew Gelman, Antonin Mace, this journal's editor, two anonymous referees, and participants to numerous conferences and seminars for their comments.

References

- Casella**, Alessandra, 2012, *Storable Votes*, Oxford University Press: New York, NY.
- Casella**, Alessandra. and Andrew Gelman, 2008, "A Simple Scheme to Improve the Efficiency of Referenda", *Journal of Public Economics*, 92, 2240-2261.
- Dahl**, Robert, 1956, *A Preface to Democratic Theory*, Chicago: University of Chicago Press.
- Dhillon**, Amrita and Jean-Francois Mertens, 1999, "Relative Utilitarianism", *Econometrica*, 67, 471-498.
- Goeree**, Jacob, and Jingjing Zhang, 2017, "One Person, One Bid", *Games and Economic Behavior*, 101, 151-171.
- Lalley**, Steven and Glen Weyl, 2018, "Nash Equilibria for Quadratic Voting", unpublished.
- Quarfoot**, David, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein and Ellen Konar, 2017, "Quadratic Voting in the Wild: Real People, Real Votes", *Public Choice*, 172(1), 283-303.
- Urbinati**, Nadia, 2019, "Political Theory of Populism", *Annual Review of Political Science*, 22: 111-127.

Alessandra Casella (ac186@columbia.edu) is Professor of Economics and Professor of Political Science at Columbia University, New York, NY 10027. Luis Sanchez (las497@cornell.edu) is a PhD candidate in Economics at Cornell University, Ithaca, NY 14853.

A Appendix

A.1 Very Brief Notes on the Literature

The question of intensity has been the focus of a large literature in both political science and economics, in fields ranging from social choice to voting theory to mechanism design. One fundamental prior question—whether the very notion of intensity is legitimate in collective decision problems—has given rise to passionate debate. To give weight to intensities in a collective decision requires being able to evaluate and compare them across different individuals, a problem that has no fully satisfactory answer. Without notions of intensity however, social welfare functions cannot give rise to rigorous discussions of inequality (Sen, 1973) and Dahls’s pragmatic position continues to seem very wise: ”We shall continue to believe not only that we can guess intelligently but that we must guess intelligently about such things” (Dahl, 1956, p.100).

If intensities are accepted and grounded on a common numeraire, a major challenge is how to elicit them truthfully. A large literature in mechanism design is devoted to this question. The main answers are mechanisms with side-payments, where individuals are asked to state their willingness to pay for a public decision and are rewarded or taxed according to the impact of their statement on the final decision. The subsidy or tax is calculated in such a way that honesty is the best strategy. VCG mechanisms take their acronym from the authors who first proposed them: Vickrey (1961), Clarke (1971) and Groves (1973).¹ The mechanisms are remarkably clever, but rely on some restricting assumptions and have problems with collusion, with bankruptcy, and with individuals’ willingness to participate if budget balance is required (Green and Laffont, 1980, Mailath and Postlewaite 1990).

A different take on the same question comes from the literature on vote markets. Here the focus is directly on voting, as the collective decision-making procedure, and, prior to voting, on purchases and sales of votes as channels for transfers, whether in the form

1. d’Apremont and Gerard-Varet (1979) proposed an important extension. See Krishna (2002) for a very clear description of the different mechanisms and their underlying logic.

of money or of broader favors. Political scientists and economists have long conjectured that, in the absence of binding budget constraints, markets for votes would allow voters to differentiate themselves according to the intensity of their preferences and lead to decisions that reflect preferences more accurately (for example, Buchanan and Tullock, 1962; Coleman, 1966; Haeefele, 1971; Mueller, 1973; Parisi, 2003). The conjecture however turned out to be misleading. Even ignoring other critiques on distributional and philosophical grounds, markets for votes would not mimic good markets. For one thing, vote trading imposes externalities on third parties; more fundamentally, votes have no value in themselves: their value depends on the influence they provide on the final decision, and such influence depends on the allocation of votes among all other voters. Hence the value of a vote is positive only if the vote is pivotal, and falls to zero when it is not. Casella et al (2012) show that once these aspects are taken into account, with a large electorate a competitive vote market followed by majority voting would lead to lower welfare than in the absence of the market.²

The assumption of a not-binding budget constraint allows these studies to ignore distributional issues that would be of great relevance in practice: voters have unequal access to resources, and markets for votes, or more generally mechanisms with side-payments, tie individuals' influence on collective decision-making to their economic power. Research has then focused on mechanisms without side-payments. Within the perspective of vote trading, scholars studied log-rolling, the possibility to give away votes on issues over which a voter has weak preferences in exchange for votes on issues felt deeply. However, after receiving a lot of attention in the 60's and 70's (see, for example, Park, 1967; Tullock, 1970; Bernholz, 1973; Riker and Brams, 1973; Ferejohn, 1974; Koehler, 1975; Schwartz, 1975; Miller, 1977) the lack of an agreed upon framework and shared conclusions lead to the loss of interest in the question. Recently, Casella and Palfrey (2019) have suggested analyzing vote trading

2. The literature claimed better results for centralized vote trading, mediated either by a market-maker or by party leaders (Koford, 1982, and Philipson and Snyder, 1996) because centralized trading can address the externalities caused by individual trades on voters who are not part of the transaction. Here too, however, very strong assumptions are required. As discussed below, the original version of QV also falls under the heading of a centralized market (an auction system) for the purchase of votes via money.

as a sequential dynamic process and reached precise but on the whole discouraging conclusions: in general, trading votes for votes need not converge to the Condorcet winner (the alternative that is majority preferred to any other), even when such an alternative exists, and, echoing Riker and Brams, if trading by the coalition of the whole is restricted, Pareto inferior outcomes are possible.

Beyond vote trading, and focussing more precisely on the question of eliciting truthful revelation of intensity of preferences, the literature has suggested other mechanisms without side-payments. The attention to intensity and the common focus on decisions with two alternatives only, as in this paper, correspond to restrictions both on preferences and on applications. One implication is that Gibbard and Satterthwaite’s general result on the manipulability of all decision rules does not apply. Nevertheless, in the collective decision-making setting that are of interest here, mechanisms without transfers that induce both truthful revelation and efficient outcomes are rare. The simplest scenario—when voting is costly and optional—does lead to abstention for weaker preferences (for example, Börgers, 2004; Krishna and Morgan, 2015), but biases will result if the cost of voting is correlated with voters’ preferences (Campbell, 1999). Limit results help: Ledyard and Palfrey (2002) show that, as the electorate becomes very large, welfare can be maximized via a simple voting rule if the threshold for approval reflects correctly the distribution of preferences in the population; Jackson and Sonnenschein (2007) show that truthfulness can be achieved without sacrificing any surplus if individuals are asked to state their priorities over several similar issues, as the number of issues becomes very large and the possible answers are constrained to reflect the distribution of preferences. Both mechanisms have high ambition—appropriating the full surplus—but rely on large numbers and on knowledge of the preference distribution.

In this scenario, the versions of SV and QV studied in this paper have both more modest goals and weaker requirements.³ The realistic ambition is to improve over simple

3. Qualitative Voting (Hortala-Vallve, 2012) was developed independently and is very similar to SV.

majority voting, with no strong claim to full efficiency, but, on the positive side, the voting schemes are set without reference to the distributions of preferences. Indeed, this is the main motivation for the study: do SV and QV perform well when specific distributions of preferences are neither assumed, as often in theoretical analyses, nor induced, as in the laboratory, nor in fact known?

A.2 The Theory

A large number N of voters are asked to vote, contemporaneously, on a set of $K > 1$ unrelated proposals. Each proposal can either pass or fail. Voter i 's preferences over proposal k are summarized by a valuation v_{ik} , where $v_{ik} > 0$ indicates that i is in favor of the proposal, and $v_{ik} < 0$ that i is against. If the proposal is decided in i 's preferred direction, then i 's realized utility from proposal k , denoted u_{ik} , equals $v_{ik} = |v_{ik}|$, otherwise it is normalized to 0. Thus the sign of v_{ik} indicates the direction of i 's preferences, and v_{ik} their intensity, of the voter's differential utility from winning the proposal over losing it. Preferences are separable across proposals, and the voter's objective is to maximize total utility U_i , where $U_i = \sum_k u_{ik}$.

Each individual's valuations $\{v_{i1}, \dots, v_{iK}\}$ are privately known. They are a random sample from a joint distribution $\mathcal{F}(v_1, \dots, v_K)$ which is common knowledge. There is no cost of voting, and voters vote sincerely. We consider three voting systems: majority voting, SV, and QV. In all three, each proposal is decided in the direction preferred by a majority of the votes cast. The voting systems differ in the rules under which votes are cast.

Under majority voting, each voter has K votes and casts a single vote on each proposal. The voting scheme gives weight to the *extent* of support for a proposal. Storable votes and quadratic voting allow voters to express not only the direction of their preferences but also their intensity.

A.2.1 Storable votes

SV grants each voter a budget of "bonus votes" to be distributed freely over the different proposals. We summarize here the main results of Casella and Gelman (2008), to which we refer the reader for details. The theoretical analysis assumes that valuations are independent across voters and propositions and restricts attention to symmetric Bayesian equilibria in undominated strategies where, conditional on their set of valuations, all voters vote sincerely. The only decision is the proposition on which to cast the bonus vote.

If voters are endowed with multiple bonus votes to distribute over multiple proposals, in a large electorate with independent values, the optimal strategy is to cumulate all bonus votes on a single proposal (section 7.11 in Casella and Gelman). Thus, in a large electorate a simple design becomes desirable. Each voter is asked to cast one vote on each proposition, and in addition is given one extra bonus vote. The bonus vote is modeled as having value $\theta > 0$, relative to a regular vote, with θ part of the optimal design of the mechanism, and dependent on the distribution of valuations. In the parametrization we use in the experiment we set $\theta = 1$: the bonus vote is equivalent to a regular vote.

With valuations independent across voters and proposals, we can phrase the problem in terms of the marginal distributions $F_k(\mathbf{v})$. Casella and Gelman show that SV behaves well, in the precise sense that ex ante expected utility improves over majority voting under multiple scenarios, as summarized by different assumptions on the marginal distributions. The result holds in the following environments. (1) If $F_k(\mathbf{v}) = F(\mathbf{v})$ for all k , where $F(\mathbf{v})$ is a distribution with known median (the median can be 0, if the distribution is symmetric, or differ from 0, if the distribution is asymmetric). (2) If $F_k(\mathbf{v})$ varies with k , but for all k $F_k(\mathbf{v})$ is symmetric around a zero median. (3) If $F_k(\mathbf{v}) = G(\mathbf{v})$ for all k , where $G(\mathbf{v})$ is symmetric around a random median with expected value at 0.

With independent voters and large N , assumptions about the shape of the distributions $F_k(\mathbf{v})$ have immediate implications about the results of the referenda. In particular, assuming specific medians for the distributions $F_k(\mathbf{v})$ amounts to assuming that a random voter's

probability of approval of each proposition is effectively known ex ante. It is then possible to predict the majority voting outcome with accuracy that converges to 1 as N becomes large. The literature has remarked that allowing for a random median, as in environment (3) above, is a better assumption (Good and Mayer 1975, Margolis 1977, Chamberlain and Rothschild 1981, Gelman et al. 2002). We report here in more detail the results that refer to that case.

Suppose that ex ante each voter i has a probability ψ_k of being in favor of proposal k ($v_{ik} > 0$), and $1 - \psi_k$ of being against ($v_{ik} < 0$). The probability ψ_k is distributed according to some distribution H_ψ defined over the support $[0, 1]$ and symmetric around $1/2$: the probability of approval is uncertain and there is no expected bias in favor or against the proposition. Each realized ψ_k is an independent draw from H_ψ .

Recall that $|v_{ik}| \equiv v_{ik}$ is i 's intensity over proposal k . To rule out systematic expected biases in intensities, both within and across proposals, assume that, regardless of the direction of preferences, the distribution of intensities is described by $Q_k(v)$, defined over support $[0, 1]$, with $Q_k(v) = Q(v)$ for all k .

We want to evaluate the welfare impact of the bonus vote, relative to a scenario with majority voting. We construct the measure:

$$\omega \equiv \frac{EW^{SV} - ER}{EW - ER} \quad (1)$$

where EW is a voter's ex ante expected utility under majority voting, ER is a floor, given by expected utility under random decision making (when any proposal passes with probability $1/2$), and EW^{SV} is ex ante expected utility under SV.

In equilibrium voters cast their bonus vote in the proposition to which they attach the highest intensity. Denoting by Ev the expected intensity over any proposal, and by $Ev_{(j)}$ the expected j th order statistic among each individual's k intensities, it is then possible to

derive:⁴

$$\omega = \frac{k(Ev) + \theta Ev_{(k)}}{(Ev)(k + \theta)} \quad (2)$$

It follows that $\omega > 1$ for all $\theta > 0$, for all distributions $H_\psi(\psi)$ and $Q(v)$, and for all $K > 1$.

By using the bonus vote to give weight to the intensity of their preferences, voters' actions work towards increasing the probability of achieving their preferred outcome in the proposition they consider their highest priority, at the cost of some reduced influence over the resolution of the other proposals. The result is an increase in expected welfare.

The conclusion, with some minor qualifications, holds in the different environments listed earlier.

A.2.2 Quadratic voting

QV is an auction-type mechanism designed for a large population faced with a single binary proposal (Goeree and Zhang 2017, Lalley and Weyl 2018a). Each voter is endowed with a numeraire and bids for the direction in which the proposal is decided. The winning side is the one with the larger total bid. The important innovation is that each voter's bid is proportional to the *square root* of the numeraire the voter commits. If values are independent across voters and the distribution F is common knowledge, the literature shows that the equilibrium strategy for almost all voters is to bid an amount proportional to one's valuation. It then follows that the decision must be efficient in utilitarian terms: it mirrors the preferences of the side with higher total valuation.⁵

4. Equation (2) follows from:

$$\begin{aligned} ER &= kEv/2 \\ EW &= kEv\pi \\ EW^{SV} &= Ev_{(k)}p_\theta + \sum_{j=1}^{k-1} Ev_{(j)}p \end{aligned}$$

where π is the ex ante probability of a desired outcome in any referendum under majority voting, and p_θ and p are the corresponding probabilities under SV when casting and when not casting the bonus vote. The challenge is characterizing these probabilities in the assumed stochastic environment.

5. If F is symmetric, bidding in proportion to one's values is the unique equilibrium strategy for all voters. If F is not symmetric, the characterization of the equilibrium is more delicate, and bids in the tails of

In the case of multiple elections, QV could be implemented by paying for votes in an artificial currency: "voices", which can be translated into votes at a quadratic cost. Casting x_k votes on proposal k requires spending x_k^2 voices on k (Posner and Weyl 2015, Lalley and Weyl 2018b). QV becomes similar to SV, but for the quadratic cost, and the quadratic cost limits the incentive to cumulate votes.

There is no theoretical analysis of the equilibrium properties of QV in multiple elections. However, a simple model shows that efficiency can extend to this case if voters believe that, on any election, the marginal impact of their votes on the probability of their preferred side prevailing is constant. We know from Lalley and Weyl (2018a) that the condition is generally not satisfied in equilibrium, but the deviations may be too subtle for voters to take into account.

The following model, similar but more transparent than the model in Lalley and Weyl (2018b), was suggested to us by Glen Weyl. There are $K > 1$ independent binary proposals, and voters values over each proposal are randomly drawn from marginal distribution $F(v)$; each voter is endowed with a budget of "voices" y_i , for simplicity set equal to 1 and fully divisible. Voices are allocated across proposals and are transformed into a number of votes on each proposal equal to the square root of the dedicated voices. Note that votes too are fully divisible. If x_{ik} denotes the votes cast on proposal k by voter i , and y_{ik} the corresponding voices, then $x_{ik} = \sqrt{y_{ik}}$, or $\sum_{k=1}^K (x_{ik})^2 = \sum_{k=1}^K y_{ik} = 1$. Each voter i faces the constrained maximization problem:

$$\text{Max}_{\{x_{ik}\}} 2 \sum_{k=1}^K p_{ik}(x_{ik}) v_{ik} \text{ subject to } \sum_{k=1}^K (x_{ik})^2 = 1$$

where 2 is a normalizing constant and $p_{ik}(x_{ik})$ is the probability that proposal k is decided as i prefers when casting x_{ik} votes. Voters adopt weakly undominated strategies and thus vote sincerely over each proposal.

distribution need not be proportional to values. Nevertheless the efficiency results continues to hold (Lalley and Weyl 2018a).

Suppose now that the marginal impact of any additional vote is constant:

$$\frac{\partial p_{ik}(x_{ik})}{\partial x_{ik}} \equiv q \quad (3)$$

Then for each proposal k , the first order condition yields:

$$x_{ik} = \frac{qv_{ik}}{\lambda_i}$$

where λ_i is the Lagrange multiplier linked to the budget constraint. Substituting the budget constraint $\sum_{k=1}^K (x_{ik})^2 = 1$, we obtain:

$$\frac{q}{\lambda_i} = \sqrt{\frac{1}{\sum_{k=1}^K (v_{ik})^2}}$$

and thus:

$$x_{ik} = \frac{1}{\sqrt{\sum_{k=1}^K (v_{ik})^2}} v_{ik} \quad (4)$$

Equation 4 says that the optimal number of votes cast on each proposal equals the voter's value, normalized by the Euclidean norm of the voter's values across all proposals. If such norms are similar across voters—for example because the number of issues is very large—or if each individual's value norm is used to normalize cardinal values in the welfare criterion, then utilitarian efficiency follows immediately by equation 4: because the number of votes cast in each proposal is proportional to the voter's value (or equal to the voter's normalized value), each proposal is won by the side with larger total values.

The model relies on two approximations. First, voices and votes are assumed to be fully divisible. Theoretically, the assumption simplifies the analysis by avoiding the complications caused by discrete vote distributions. In practice, it suggests giving voters a large number of voices. Experiments, on the other hand, routinely suggest that subjects have difficulties making decisions when the set of options is large. In our experimental implementation, we

take a different route and simplify the subjects’ problem by limiting the number of options.

The second approximation is more substantive and is the assumption of constant marginal impact of additional votes, the simplification embodied in equation 3 above.⁶ Theoretically the simplification is strong and unlikely to hold in general. The practical question is how large the deviation is and how is it reflected in voters’ actual choices. As long as voters *believe* that votes have constant marginal impact, the characterization of their behavior follows correctly.

A.3 Implementation of QV in the experiment.

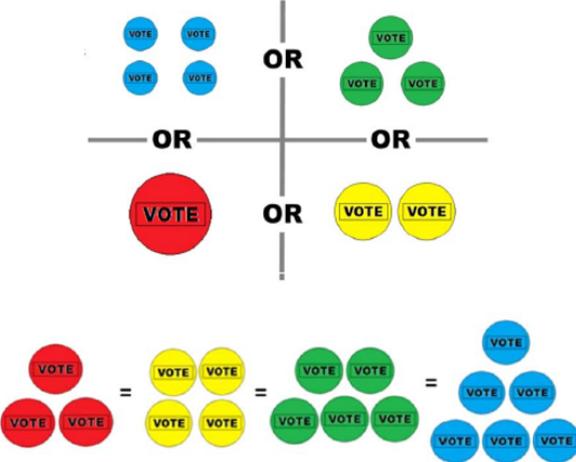


Figure A1: *The design of the QV scheme in the MTurk survey.*

A.4 Experimental data

We collected the data in May 2016. Two months earlier, in March, we had presented an original set of ten propositions, all with the potential to reach the November ballot, to a sample

6. In this simple model, with equal marginal distributions of values, marginal pivotality is constant across issues, for given number of votes. The iid assumption could be relaxed and, with a large number of independent voters and a large number of issues, constant marginal pivotality across issues may conceivably arise as an equilibrium result. Constant marginal pivotality across the number of votes cast, for given issue, however will not hold in equilibrium.

of 94 California MTurk subjects. Given the responses, we selected the four propositions we used in the final survey on the basis of three criteria: we needed propositions whose outcome was unlikely to be a landslide, about which some voters would feel strongly, and that would be clear enough to the average MTurk subject. The March responses also yielded a poll we reported at the end of the May survey before asking if respondents wanted to change any of their answers. Very few did, with no impact on aggregate results, and we ignore it.

A.4.1 Cleaning procedures

In designing the survey, we added an attention check to both samples. The check took the form of a fictitious fifth proposition, titled the "Effective Workers Initiative", whose accompanying text asked the reader *not* to hit any of the three "For", "Against" and "Abstain" buttons and continue directly to the next screen. The order of this fifth "initiative" was random.

Before analyzing the data, we excluded all subjects who either did not conclude the survey or failed the attention check. In addition, we excluded subjects in the QV sample who chose the red vote and cast it on a proposition on which they abstained—these subjects effectively abstained on all propositions under the QV scheme, and left us no alternative. We also excluded all subjects in the SV sample who cast the bonus vote on a proposition on which they abstained—a behavior that may correspond to rejecting the use of the bonus vote, but seems more likely to denote confusion or lack of interest, as in the QV sample. (Results are effectively unchanged if we maintain these subjects in the sample). These exclusions reduced the two samples to 306 (from 324) subjects for SV, and 313 (from 323) for QV.

In both samples, we set to zero the number of points assigned by a subject to a proposition on which the subject abstained (again note that we have no alternative since we do not know the direction of the subject's preferences on such a proposition). Finally, we set to +1 (or -1) the points attached to a proposal on which a subject voted in favor (or against) but to which the subject assigned zero points. Out of 100 total points, this very

minor adjustment allows us to give at least minimal weight to the direction of preferences expressed by the subject.

A.4.2 Subjects' preferences

Subjects' preferences over each proposition are summarized in histograms reporting the number of respondents who assign different numbers of points to a proposition. Points are coded as negative when the subject voted against the proposal, and as positive when the subject voted in favor, with bins of size 10 (0, colored light blue in the figures, corresponds to abstentions). Figure [A2](#) below reports the histograms relative to the IM proposition for the two samples. It says, for example, that in the SV sample 38 respondents assigned it between 1 and 10 points and voted against it, while 27 assigned to the proposition equally low points but voted in favor (the corresponding numbers for QV are 29 and 20). The figure also reports, for each sample, the total number of subjects for and against, the abstentions, and the total number of points, for and against (bold indicates the larger number).

Figures [A3](#), [A4](#), and [A5](#) report the histograms for the other propositions.

Note in particular the strong symmetry of the distributions of value in the IM initiative, in both samples. Thus both the mean number of points (which determines the welfare measure) and the median (which determines the majority winner) are close to zero: on a possible range of $[-100, 100]$, they are 2.85 and -2 in the SV sample, and -0.42 and -4 in the QV sample. In addition, symmetry implies that random decision-making would perform well, in welfare terms, over this initiative, and thus the surplus left to exploit through the voting rule is small. The difference in sign between mean and median in the SV sample, and thus the inefficient outcome under majority voting, is heavily penalized.

A.4.3 The voting choices

SV The optimal selection of the proposition on which to cast one's bonus vote is not trivial. If there are asymmetries among the propositions, it should reflect not only relative

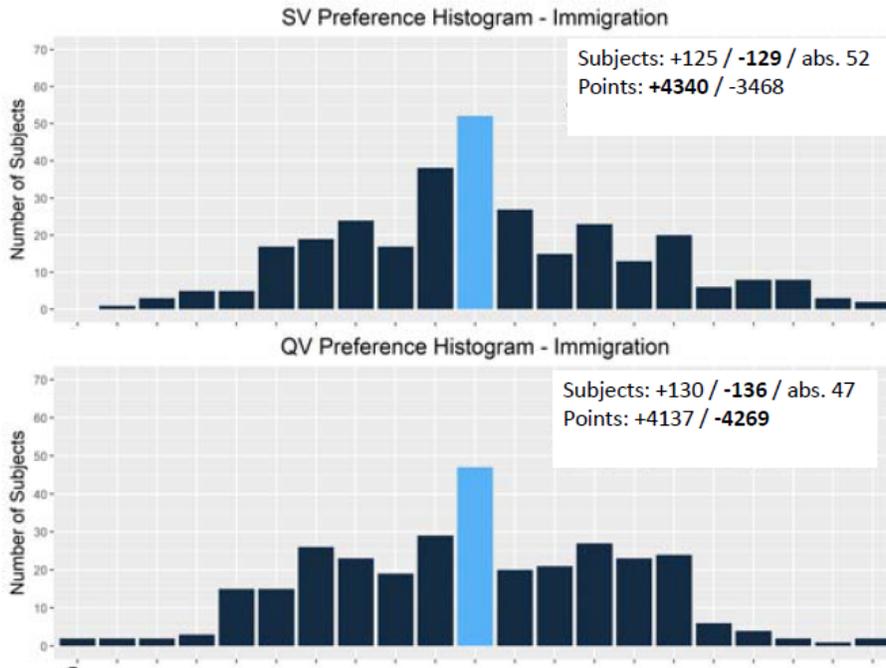


Figure A2: *Distribution of preferences: the IM proposition*

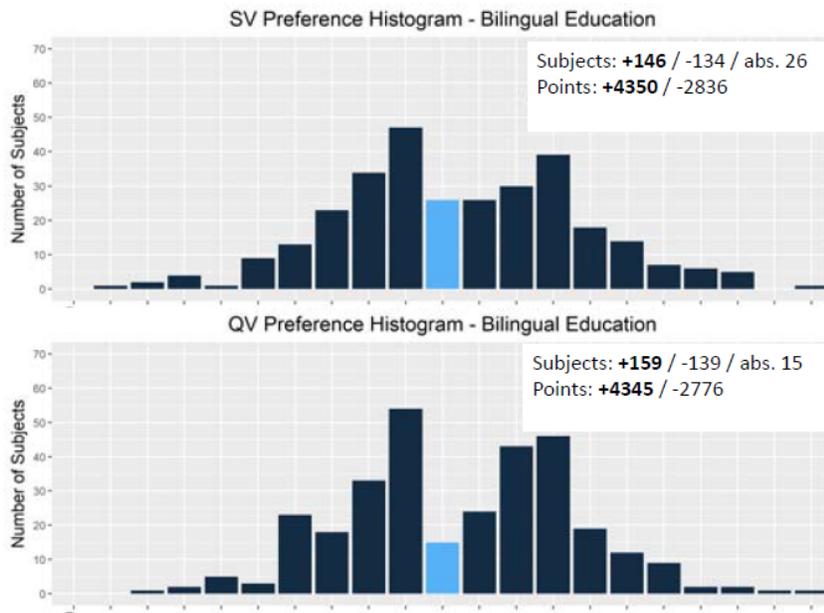


Figure A3: *Distribution of preferences: the BE proposition.*

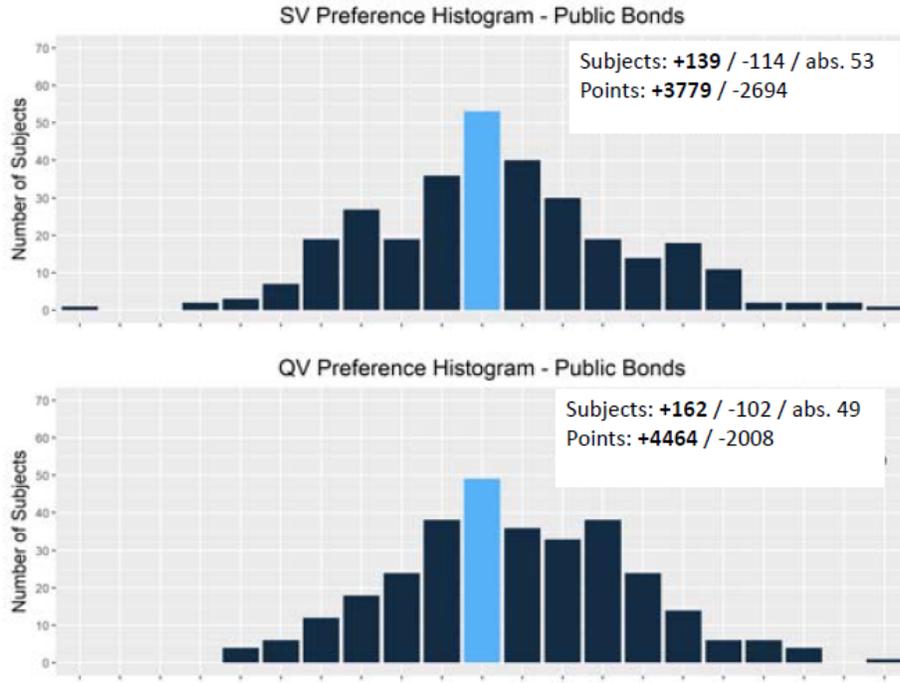


Figure A4: *Distribution of preferences: the PB proposition*

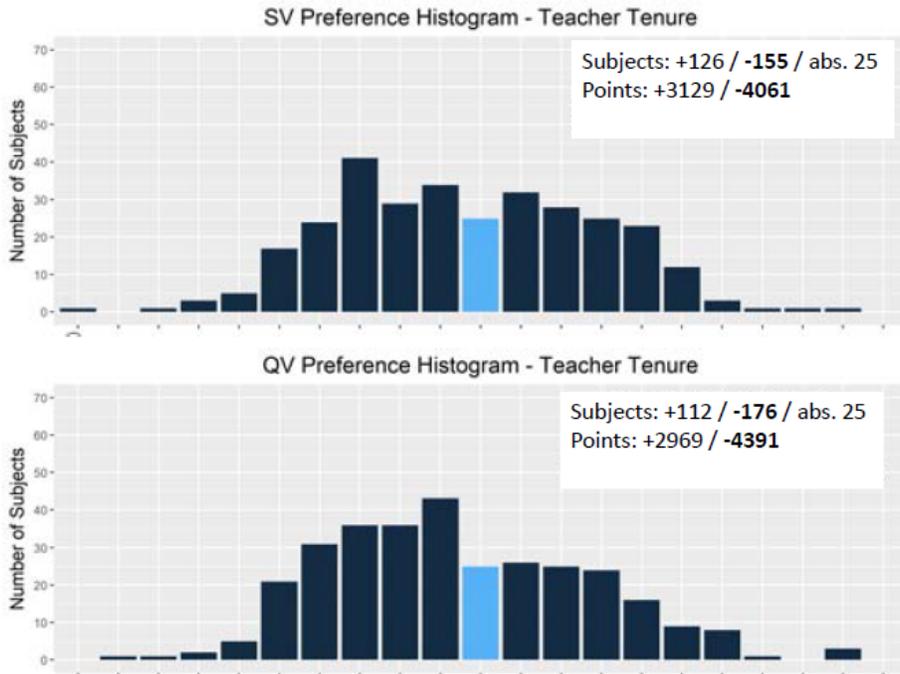


Figure A5: *Distribution of preferences: the TT proposition*

valuations, but also pivotality (all else equal, pivotality is higher if the proposition is expected to be close and lower if it is salient). If voters are not well-informed, however, or unable or unwilling to compute equilibria, a plausible rule-of-thumb is to treat all propositions equally and cast the bonus vote on the proposal on which one’s preferences are most intense. The simple rule corresponds to optimal behavior if voters believe that valuations v are iid across proposals according to some marginal distribution F symmetric around zero (Casella and Gelman 2008). Figure A6 shows, for each of the four propositions, a measure of the relative intensity of preferences for all voters who cast their bonus vote on that proposition (with points slightly jittered for visibility). The vertical axis is the number of points assigned to the proposition; the horizontal axis is the maximum number of points assigned to any other.⁷

If all voters had cast their bonus vote on the proposition to which they assigned the highest number of points, all dots in each panel would be above the 45 degree line. In total, three fourths of all subjects (74%) did so.

The salience of the IM proposition is supported by the high number of bonus votes (97 vs. 76 for BE, 61 for PB and 72 for TT) . As noted in the text, when accounting for bonus votes the margin of victory for opponents of the proposition increases, although IM supporters report higher average and total intensity. The result reflects two sources of asymmetry. Of the 23 subjects who identify the IM proposition as their first priority and yet do not target it with their bonus vote, more than twice (16) are supporters rather than opponents (7); of the 20 subjects who cast their bonus vote on IM and yet do not identify it as their priority, more than twice (14) are opponents rather than supporters (6). However these differences are numerically very small and in neither case are the differences in proportions statistically significant.

As we describe briefly in the text, we summarize the voting choices through a statistical model, to be read as compact representation of the data.

7. A few subjects cast the bonus vote on a proposition to which they had not assigned any points. As described earlier, if they nevertheless voted on that proposition they are recoded as assigning +/- 1 point, depending on the direction of preferences.

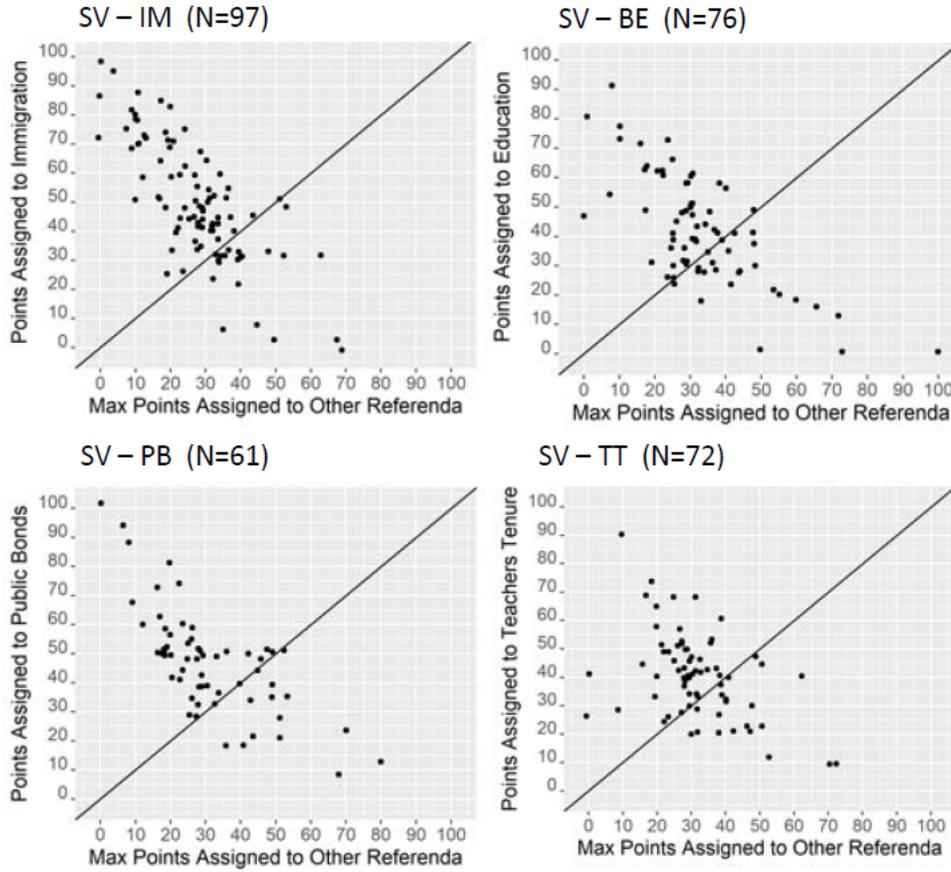


Figure A6: *The Bonus Vote Decision*

We conjecture that each subject follows one of four mutually exclusive behaviors: with probability p_{Max} , the bonus vote is cast on the subject's highest value proposition; with probability p_{Close} on the one with closest outcome (IM); with probability p_{Fam} on the most familiar (either BE or TT, the two education propositions, with equal probability); and with probability p_{Rand} according to some other criterion that appears to us fully (uniformly) random. Each choice observed in the data, matched with the individual's reported valuations, can be expressed as function of the four probabilities. Assuming independence across subjects, the probabilities are estimated by MLE are reproduced in Table A1.

SV

		95% CI
p_{Max}	0.63	[0.56, 0.76]
p_{Close}	0.04	[0, 0.14]
p_{Fam}	0.09	[0.04, 0.21]
p_{Rand}	0.23	[0.15, 0.37]

Table A1. *SV voting choices: a statistical model.* MLE estimates under the constraints that all probabilities be non-negative and sum to 1. The confidence intervals are obtained by bootstrapping and reflect the distribution of the estimated probabilities in 10,000 simulations.

In the SV sample, the bonus vote is cast primarily on the proposition that the voter considers the highest priority, but a relatively large role is left for randomness.

QV Under QV, voters need to make two choices: the class of votes, and, given the class, the propositions on which the votes are cast. Figure A7 reports the frequencies with which subjects chose the different classes.

As the figure shows, even with the convex penalty from cumulating voting power, a full 40 percent of subjects chose the red vote, and thus cast their vote on a single proposition; less than 10 percent cast votes on all four propositions. In the data, the frequency of selection of a vote class is monotonic in the votes' weight, contrary to what theory predicts. Across propositions, IM received the highest number of red votes (45 vs. 19 for BE, 26 for PB and 34 for TT), as well as the highest total number of unweighted or weighted votes (178 vs. 153 for BE, 120 for PB, and 156 for TT (unweighted), or 249 vs. 210.4 for BE, 165.8 for PB, and 218.7 for TT (weighted)), confirming its salience.

As for SV, we can summarize observed voting choices through a simple statistical model. We posit a noisy two-step process that starts with choosing a vote class and then, given the vote class, proceeds to casting the vote(s) on the different propositions in order of

QV: fraction of vote classes

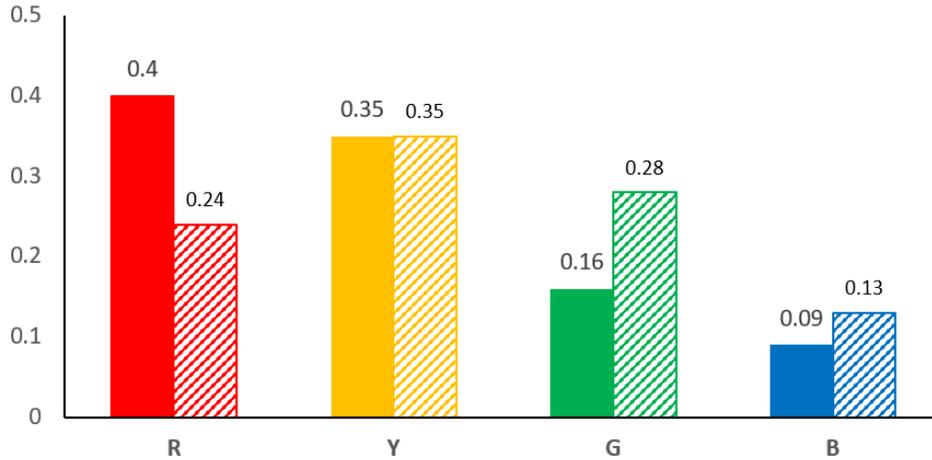


Figure A7: *QV sample. Frequency of vote classes.* The solid columns correspond to the observed frequencies in the MTurk sample. The striped columns correspond to optimal choices, given the observed distributions of preferences, if voters perceive the marginal pivotal probability to be constant (as under rule-of-thumb QV-C, described below).

intensity.

Denoting by $b_{(k)}$ the voter's k th highest number of assigned points, we summarize behavior through five parameters: $\{\rho, \gamma, \xi, \varepsilon, \mu\}$. With probability $(1 - \varepsilon)$, the voter chooses the vote class that reflects her relative priorities: red if $b_{(4)}/b_{(3)} \geq \rho$, yellow if $b_{(4)}/b_{(3)} < \rho$ but $b_{(3)}/b_{(2)} \geq \gamma$, green if $b_{(4)}/b_{(3)} < \rho$, $b_{(3)}/b_{(2)} < \gamma$, but $b_{(2)}/b_{(1)} \geq \xi$, and blue otherwise; with total probability ε , the voter chooses one of the other classes (uniformly). Given a vote class, with probability $(1 - \mu)$ all votes are cast monotonically, i.e. on the highest intensity propositions; with probability μ , votes are cast non-monotonically, with uniform probability over the different options. Parameters are estimated by MLE, assuming independence across subjects, and reproduced in Table A2.

QV		
		95% CI
ρ	1.33	[1.21, 2.10]
γ	1.19	[1.04, 1.21]
ξ	1.39	[1.18, 1.79]
ε	0.50	[0.44, 0.54]
μ	0.21	[0.17, 0.26]

Table A2. *QV voting choices: a statistical model.* MLE estimates under the constraints that all threshold parameters be larger or equal to 1 and all probabilities be non-negative and smaller than 1. The confidence intervals are obtained by bootstrapping and reflect the distribution of the estimated parameters in 10,000 simulations.

The thresholds determining the choice of vote class are significantly higher than 1 but quantitatively not far, reflecting a strong tendency towards cumulating voting power. About half of the time, the vote class does not obey the threshold rules estimated by the model. Given the vote class, however, the tendency towards monotonicity is strong.

A.5 Bootstrap simulations

A.5.1 Rules-of-thumb

As described in the text, rule A corresponds to the observed behavior of the subject drawn when populating the bootstrapping sample. Rule B implements the statistical models in Tables A1 and A2, given the subject’s allocation of points and voting system. Rule C corresponds to optimal voting behavior under the maintained assumption that individuals take their probability of pivotality as constant (across propositions in SV, and across the number and weight of votes in QV). Under SV, the optimal rule is then to cast the bonus vote on one’s highest intensity proposition. Under QV, individuals choose a vote class so as to minimize the distance between the weights of the votes they cast and their normalized

values. Rule D suppose that individuals act as under C with probability 1/2, and randomly with probability 1/2. Because rules C and D are less transparent under QV, we describe them in some more detail.

QV: rules C and D As shown in section A.2.2, assuming constant marginal pivot probabilities, the optimal number of votes under QV is proportional to the voter’s value, or $x_{ik} = \alpha_i v_{ik}$. The budget constraint $\sum_{k=1}^K (x_{ik})^2 = 4$ then implies $\alpha_i = 2/\beta_i$, where β_i is the Euclidean norm of the voters’ values, or $\beta_i = \sqrt{\sum_{k=1}^K (v_{ik})^2}$. Under QV, rule C attributes to each subject a vote class by selecting the vector, out of $\{2, 0, 0, 0\}$, $\{1.5, 1.5, 0, 0\}$, $\{1.2, 1.2, 1.2, 0\}$, $\{1, 1, 1, 1\}$ that minimizes the distance from $\{2v_{i(4)}/\beta_i, 2v_{i(3)}/\beta_i, 2v_{i(2)}/\beta_i, 2v_{i(1)}/\beta_i\}$.⁸ Given a vote class, votes should be assigned monotonically.

As a check, we reestimated the QV statistical model reported in Table A2 after having imposed the normative choice on the MTurk sample. As shown in Table A3 below, estimated thresholds ρ and γ are significantly higher than in Table A2, supporting the hypothesis that MTurk respondents concentrated votes excessively.⁹

	95% CI	
ρ	2.35	[2.33, 2.56]
γ	1.95	[1.85, 2.04]
ξ	1.63	[1.42, 1.79]
ε	0.05	[0.03, 0.08]
μ	0.0	[0.0, 0.0]

Table A3. *Reestimating the QV statistical model imposing rule C.* MLE estimates. The confidence intervals are obtained by bootstrapping and reflect the distribution of the estimated parameters in 10,000 simulations.

8. Requiring subjects to indicate intensity by allocating 100 points among the four initiatives implies that values are normalized linearly. It is easy to verify however that the linear normalization does not affect the transformation described here: if $\tilde{v}_{ik} = v_{ik} / (\sum_{k=1}^K v_{ik})$, then $\tilde{v}_{ik} / \sqrt{\sum_{k=1}^K (\tilde{v}_{ik})^2} = v_{ik} / \sqrt{\sum_{k=1}^K (v_{ik})^2}$.

9. Note that the constraints imposed on the problem—the four vote classes and the fixed number of votes per initiative in each class—imply that when estimating intensity thresholds as allocation criteria, some errors ($\varepsilon > 0$) are to be expected. Monotonicity, on the other hand, can be fully respected ($\mu = 0$).

Rule D leaves a larger role to randomness. With QV, it means choosing the vote class according to rule C with probability 1/2, and choosing any of the four classes with probability 1/8 each. Given a vote class, the voter casts the available votes on the propositions with highest values with probability 1/2, and randomly, treating all propositions equally, otherwise.

Random behavior To evaluate the welfare costs of misusing the voting systems, we have simulated, for both SV and QV, a fifth possible behavior capturing randomness. We posit that voters vote in the direction of their preferences but choose randomly the proposition on which to cast the bonus vote under SV, and both the vote class and the propositions under QV. As intuition suggests, over the 10,000 simulations, average welfare under both SV and QV replicates average welfare under majority voting. (The results are available upon request.)

A.5.2 Numerical values for Figure 2 (in the text)

We report here the numerical values corresponding to the three panels of Figure 2 (in the text).

A: Minority victories			B: Share of surplus			C: Welfare losses		
	SV	QV		SV	QV		SV	QV
<i>A</i>	0.35	0.35	<i>A</i>	0.80	0.98	<i>A</i>	0.28	0.36
<i>B</i>	0.33	0.29	<i>B</i>	0.87	0.975	<i>B</i>	0.05	0.38
<i>C</i>	0.41	0.27	<i>C</i>	0.92	0.996	<i>C</i>	0.02	0.14
<i>D</i>	0.29	0.28	<i>D</i>	0.84	0.98	<i>D</i>	0.10	0.35
<i>avg</i>	0.35	0.30	<i>avg</i>	0.86	0.98	<i>avg</i>	0.11	0.31
			<i>maj SV</i>	0.72				
			<i>maj QV</i>	0.95				

Table A4. Simulations based on the original samples. Panel (A): Frequency of simulations such that the minority wins at least one proposition; Panel (B) Share of appropriated surplus over the full 10,000 simulations; Panel (C): Frequency of welfare losses, relative to majority, conditional on at least one minority victory.

As discussed in the text, majority voting fails to appropriate 28 percent of available surplus in the simulations drawn from the SV sample, and only 5 percent in those drawn from the QV sample. And yet, even with such little leeway, QV succeeds in improving over majority under each of the four rules (averaging over all simulations). If we use majority as reference, we can calculate the fraction of surplus *above* what is appropriated by majority that the two voting rules capture, according to each of the four rules. For example, under rule A, SV captures $(0.80-0.72)/(1-0.72)$ of such surplus, or 29 percent, while QV captures $(0.982-0.95)/(1-0.95)$ or 64 percent. Table A5 reports these numbers for each of the four rules.

	Fraction of surplus over majority	
	SV	QV
<i>A</i>	0.29	0.64
<i>B</i>	0.55	0.50
<i>C</i>	0.72	0.915
<i>D</i>	0.44	0.59
<i>avg</i>	0.50	0.66

Table A5. *Simulations based on the two samples separately.* Fraction of the surplus left on the table by majority appropriated by the two voting schemes, under each of the four rules.

A.5.3 Differences across propositions

Because both SV and QV constrain the use of the votes across propositions, each bootstrapped sample corresponds to an outcome for all four propositions. The discussion in the text focuses on the frequency of samples in which at least one minority victory is observed,

without distinguishing among propositions. Yet, large differences exist. As the preference histograms in Figures A2, A3, A4, and A5 show, the potential for minority victories is largest in IM and BE, while SV and QV have a much smaller effect on the other two propositions. Figures A8-A9 below report the results.

SV. Frequency of minority victories by initiative

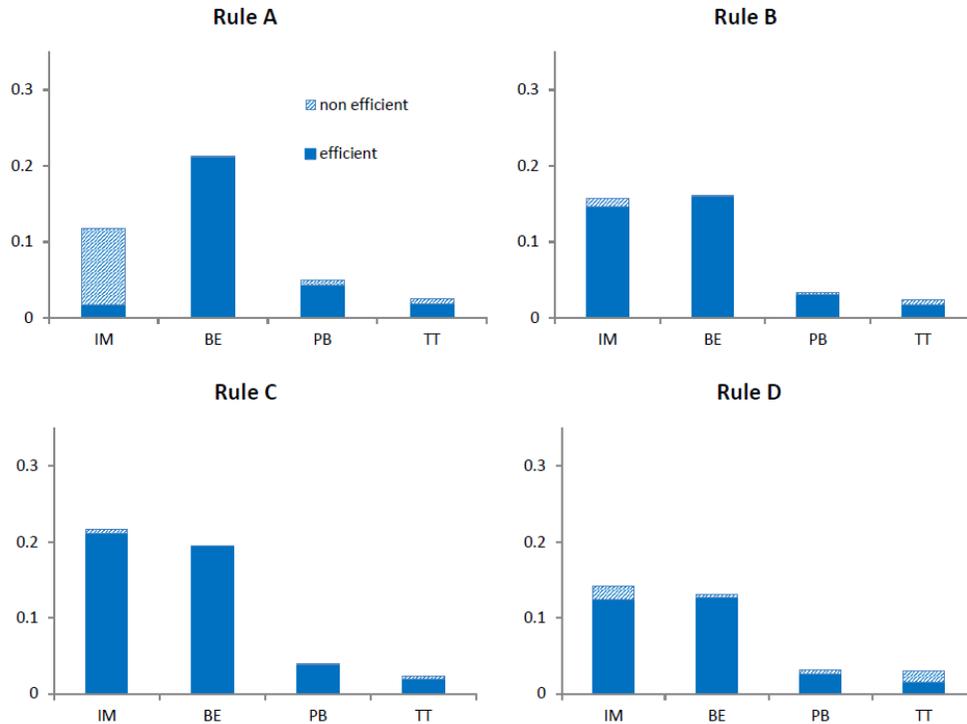


Figure A8: *Minority victories by initiative. SV simulations.* In each panel, the solid (dashed) part of each column is the fraction of minority victories that are efficient (inefficient). The efficient frequency of minority victories by initiative is: 0.52 (IM), 0.24 (BE), 0.05 (PB), and 0.04 (TT).

QV. Frequency of minority victories by initiative

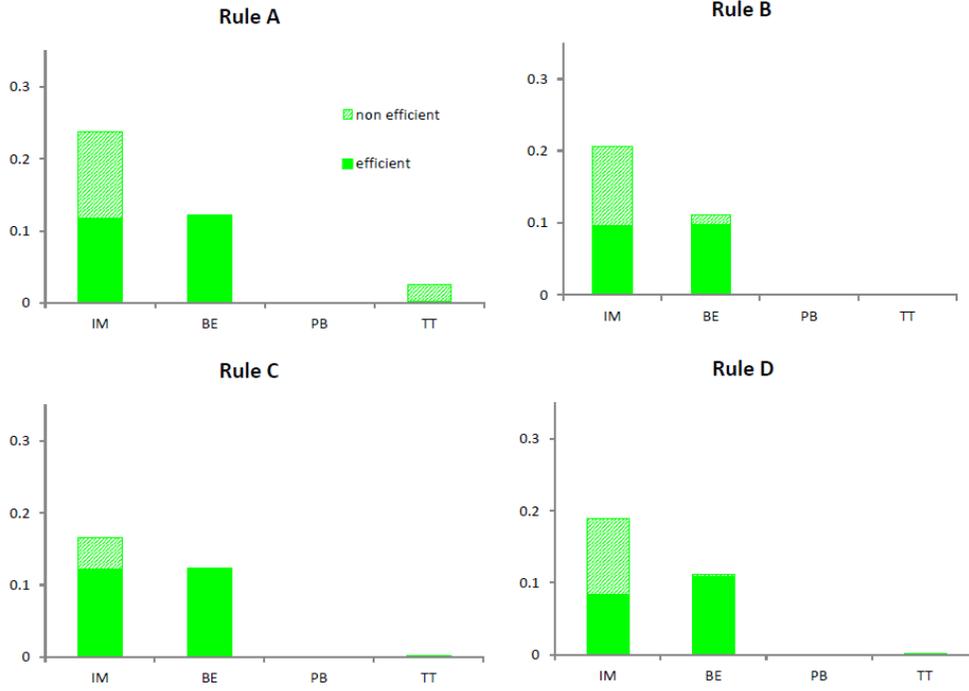


Figure A9: *Minority victories by initiative. QV simulations: bootstrapping the original QV sample.* In each panel, the solid (dashed) part of each column is the fraction of minority victories that are efficient (inefficient). The efficient frequency of minority victories by initiative is: 0.18 (IM), 0.12 (BE), 0.00 (PB and TT).

A.5.4 Numerical values for Figure 3 (in the text)

We report here the numerical values corresponding to Figure 3 (in the text).

A: Minority victories			B: Share of surplus			C: Welfare losses		
	SV	QV		SV	QV		SV	QV
<i>B</i>	0.18	0.32	<i>B</i>	0.94	0.96	<i>B</i>	0.14	0.18
<i>C</i>	0.195	0.39	<i>C</i>	0.95	0.99	<i>C</i>	0.09	0.03
<i>D</i>	0.17	0.29	<i>D</i>	0.93	0.95	<i>D</i>	0.19	0.17
<i>avg</i>	0.18	0.34	<i>avg</i>	0.94	0.97	<i>avg</i>	0.14	0.13
			<i>maj</i>	0.896	0.896			

D: Fraction of surplus over majority

	SV	QV
<i>B</i>	0.45	0.575
<i>C</i>	0.56	0.89
<i>D</i>	0.37	0.55
<i>avg</i>	0.46	0.67

Table A6. *Simulations based on the two samples combined.* Panel (A): Frequency of simulations such that the minority wins at least one proposition; Panel (B) Share of appropriated surplus over the full 10,000 simulations; Panel (C): Frequency of welfare losses, relative to majority, conditional on at least one minority victory; Panel (D): Fraction of the surplus left on the table by majority appropriated by the two voting schemes.

A.6 Adding regular votes to QV

The QV survey did not include regular votes—votes that must be cast one each on each proposition. However, according to the theory in section A.2.2, neither a voter’s optimal strategy nor QV’s efficiency properties would change. On this basis, we have conducted an additional set of 10,000 simulations adding regular votes to QV, cast over each initiative according to the preferences elicited at the start of the survey, as for SV. We call this voting scheme QVV–QV *with Vote*.

We begin by reporting here data comparing QV and QVV when bootstrapping the original QV sample. As noted in the text, in the QV simulations based on the original sample majority voting appropriates a full 95% of the available surplus. On average, QV improves over majority under each of the four rules-of-thumb, but the small margin for improvement translates into a relatively high frequency of inefficient minority victories. QVV addresses this problem effectively: averaging over all rules, the frequency of minority victories declines from 30% to 19% (Fig. A10: A), with a corresponding decline (from 31% to 18%) in the frequency of welfare losses, in the simulations with at least one minority victory (Fig. A10: C). The impact on the average realized share of surplus is negligible (Fig. A10: B).

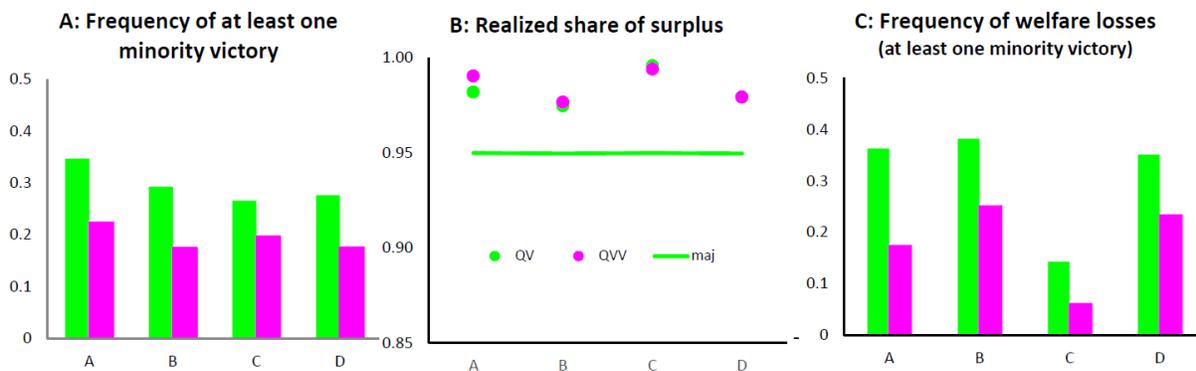


Figure A10: *Bootstrapping the original QV sample: QV and QVV.* The fractions of samples with at least one minority victory (Panel A) or with welfare losses, conditional on at least one minority victory (Panel C) are consistently lower under QVV ($p < 0.001$ in all cases). In Panel B, the average shares of surplus are not statistically different from majority (based on bootstrapped CI's). All tests of proportions are one-sided Z tests.

Table A7 reports the numerical values corresponding to the figure (including the numbers for QV for comparison).

A: Minority victories			B: Share of surplus			C: Welfare losses		
	QVV	QV		QVV	QV		QVV	QV
<i>A</i>	0.22	0.35	<i>A</i>	0.99	0.98	<i>A</i>	0.17	0.36
<i>B</i>	0.175	0.29	<i>B</i>	0.98	0.975	<i>B</i>	0.25	0.38
<i>C</i>	0.20	0.27	<i>C</i>	0.99	0.996	<i>C</i>	0.06	0.14
<i>D</i>	0.18	0.28	<i>D</i>	0.98	0.98	<i>D</i>	0.23	0.35
<i>avg</i>	0.19	0.30	<i>avg</i>	0.99	0.98	<i>avg</i>	0.18	0.31
			<i>maj</i>	0.95	0.95			

D: Fraction of surplus over majority

	QVV	QV
<i>A</i>	0.81	0.64
<i>B</i>	0.54	0.50
<i>C</i>	0.88	0.91
<i>D</i>	0.59	0.59
<i>avg</i>	0.70	0.66

Table A7. *Comparison of QVV and QV in the simulations drawn from the original QV sample.* Panel (A): Frequency of simulations such that the minority wins at least one proposition; Panel (B) Share of appropriated surplus over the full 10,000 simulations; Panel (C): Frequency of welfare losses, relative to majority, conditional on at least one minority victory; Panel (D): Fraction of the surplus left on the table by majority appropriated by the two voting schemes.

We find similar results in the simulations based on the two samples combined, although the gains from QVV, relative to QV, are reduced by the lower proportion of QV errors. Relative to QV, QVV forfeits minority victories with small welfare changes, whether in terms of gains or losses (Figure [A11](#)).

Simulations based on the two samples, joint.

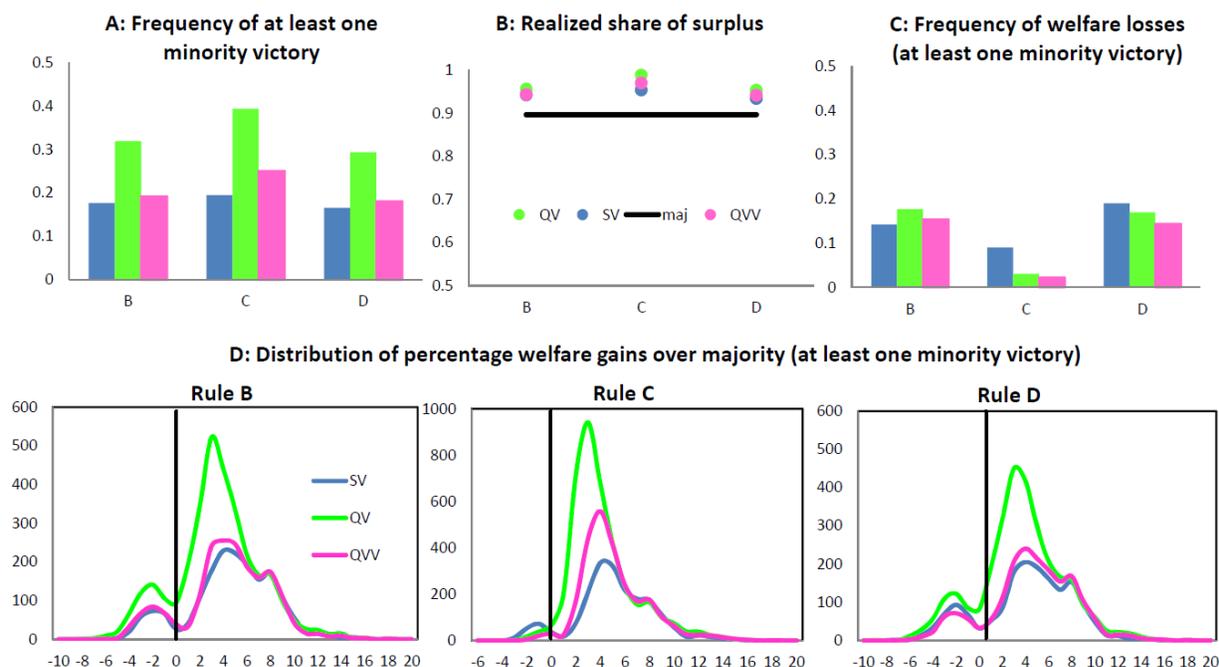


Figure A11: *Comparing SV, QV and QVV over the two samples, joint.* In Panel A, the fraction of samples with at least one minority victory remains consistently lower under SV than under QVV ($p < 0.01$). In Panel B, none of the differences in mean surplus are statistically significant from one another (bootstrapped CI's), but Panel D reports detailed information on percentage welfare gains over all samples with at least one minority victory (the vertical axis is absolute numbers; the horizontal axis are percentage gains, and the vertical line is at 0). Over those samples, the frequency of welfare losses under QVV is significantly lower than under QV ($p = 0.018$ (B), $p = 0.019$ (C), $p = 0.0096$ (D)), and than SV for rules C and D ($p < 0.01$), but not for rule B ($p = 0.15$) (Panel C). In Panel C, the frequency of welfare losses under QVV-C is 0.02. All tests of proportions are one-sided Z tests.

Table A8 reports the numerical values for QVV (including QV for comparison).

A: Minority victories			B: Share of surplus			C: Welfare losses		
	QVV	QV		QVV	QV		QVV	QV
<i>B</i>	0.19	0.32	<i>B</i>	0.94	0.96	<i>B</i>	0.15	0.18
<i>C</i>	0.25	0.39	<i>C</i>	0.97	0.99	<i>C</i>	0.02	0.03
<i>D</i>	0.18	0.29	<i>D</i>	0.94	0.95	<i>D</i>	0.14	0.17
<i>avg</i>	0.21	0.34	<i>avg</i>	0.95	0.97	<i>avg</i>	0.11	0.13
			<i>maj</i>	0.896	0.896			

D: Fraction of surplus over majority

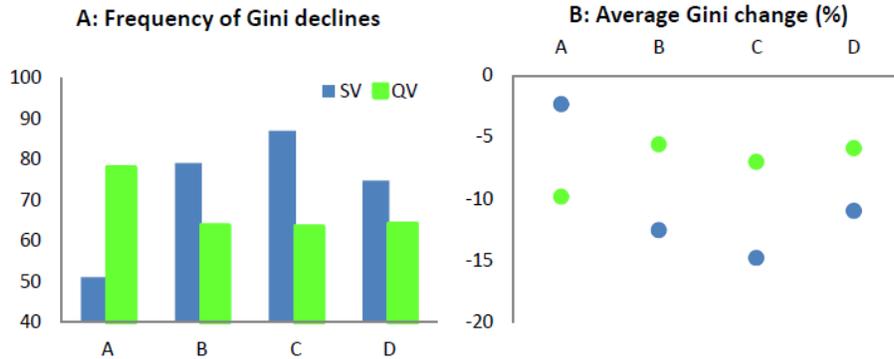
	QVV	QV
<i>B</i>	0.45	0.575
<i>C</i>	0.72	0.89
<i>D</i>	0.44	0.55
<i>avg</i>	0.54	0.67

Table A8. *Comparison of QVV and QV in the simulations drawn from the two samples combined.* Panel (A): Frequency of simulations such that the minority wins at least one proposition; Panel (B) Share of appropriated surplus over the full 10,000 simulations; Panel (C): Frequency of welfare losses, relative to majority, conditional on at least one minority victory; Panel (D): Fraction of the surplus left on the table by majority appropriated by the two voting schemes.

A.7 Inequality

As we constructed a measure of aggregate welfare, we can construct a measure of individual welfare. Denoting by b_{ik} the number of points attributed to proposition k by individual i , and M_k^S the side casting the majority of votes on k under voting system S , as in the text, individual ex post utility under S corresponds to $U_i^S = \sum_{k:i \in M_k^S} b_{ik}$. U_i^S reflects both the frequency with which voter i is on the winning side of a proposition and the importance that i attributes to it. By construction, a voter who loses all propositions has an ex post utility of 0; and one who wins them all of 100.

Simulations based on the two separate samples
(at least one minority victory)



Simulations based on the two samples, joint
(at least one minority victory)

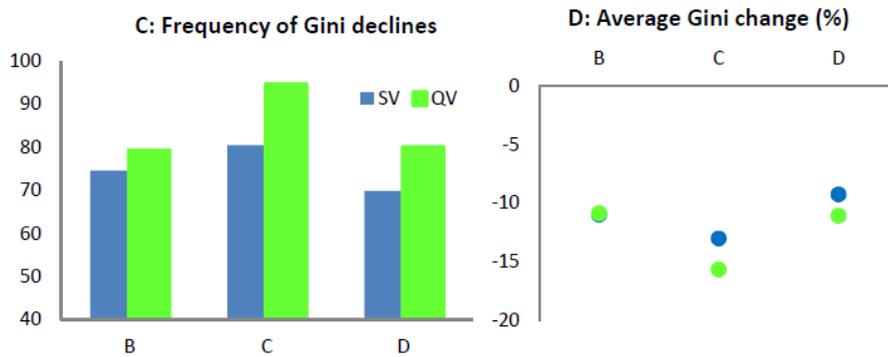


Figure A12: *Frequency and magnitude of Gini declines*. Panels A and C: With the exception of SV-A, the probability of a Gini decline is statistically higher than 50% for both SV and QV, and the differences between SV and QV are all significant ($p < 0.01$; one sided Z-tests). Panels B and D: The average magnitudes of the Gini declines are not statistically significant (based on bootstrapped CI's. See SI).

To evaluate SV and QV's impact on inequality, we calculate the Gini coefficient of the realized utility distribution for each of our simulations and under each rule-of-thumb, under the relevant voting system and under simple majority

Focusing on simulations with at least one minority victory and averaging across the rules-of-thumb, the frequency of Gini declines, relative to majority voting, is 73% for SV and 67% for QV when bootstrapping separately the two original samples (Figure A12:A)

and increases to 75% for SV and 85% for QV in the simulations based on the two samples combined (Figure A12:C). Over all rules, the average Gini change is a decline of 10% for SV and 7% for QV in the simulations based on the two separate samples (Figure A12:B), and a decline of 11% and 12% respectively when bootstrapping the two samples jointly (Figure A12:D). In both sets of simulations, the impact of QVV is statistically indistinguishable from QV (Tables A9 and A10). The summary message is that both SV and QV have a positive impact on ex post inequality.

Average Gini coefficients are reported in Tables A9 and A10.¹⁰

SV			QV		
		95% CI			95% CI
<i>maj</i>	0.37	[0.32, 0.43]	<i>maj</i>	0.32	[0.27, 0.37]
<i>A</i>	0.35	[0.28, 0.40]	<i>A</i>	0.29	[0.26, 0.345]
<i>B</i>	0.32	[0.28, 0.40]	<i>B</i>	0.30	[0.26, 0.35]
<i>C</i>	0.315	[0.28, 0.39]	<i>C</i>	0.30	[0.26, 0.35]
<i>D</i>	0.33	[0.28, 0.40]	<i>D</i>	0.30	[0.26, 0.35]

QVV		
		95% CI
<i>maj</i>	0.33	[0.27, 0.37]
<i>A</i>	0.29	[0.26, 0.35]
<i>B</i>	0.30	[0.26, 0.35]
<i>C</i>	0.30	[0.26, 0.35]
<i>D</i>	0.30	[0.26, 0.35]

10. Given the 10,000 simulations, each rule, A, B, C, or D, corresponds to a somewhat different realization of minority victories, and thus to somewhat different subsamples. Thus results under majority voting vary slightly across rules. The differences are very small. The numbers we report for majority voting in Tables A9 and A10 correspond to the subsamples with at least one minority victory under rule C, but the changes are very minor and the conclusions identical if we use any other rule.

Table A9. *Gini coefficient: Simulations based on the two separate samples (at least one minority victory)*. The confidence intervals are obtained by bootstrapping.

SV			QV		
		95% CI			95% CI
<i>maj</i>	0.35	[0.29, 0.38]	<i>maj</i>	0.35	[0.30, 0.38]
<i>B</i>	0.31	[0.27, 0.36]	<i>B</i>	0.31	[0.27, 0.36]
<i>C</i>	0.30	[0.27, 0.36]	<i>C</i>	0.29	[0.27, 0.35]
<i>D</i>	0.31	[0.27, 0.37]	<i>D</i>	0.31	[0.28, 0.36]

QVV		
		95% CI
<i>maj</i>	0.35	[0.31, 0.38]
<i>B</i>	0.31	[0.27, 0.37]
<i>C</i>	0.30	[0.27, 0.36]
<i>D</i>	0.31	[0.28, 0.36]

Table A10. *Gini coefficient: Simulations based on the two samples joint (at least one minority victory)*. The confidence intervals are obtained by bootstrapping.

References in the appendix

- Bernholz, P., 1973, "Logrolling, Arrow Paradox and Cyclical Majorities," *Public Choice*, 15, 87-96.
- Börger, T., 2004, "Costly Voting", *American Economic Review*, 94, 57-66.
- Buchanan, J. and G. Tullock, 1962, *The Calculus of Consent*, Ann Arbor: Univ. Michigan Press.
- Campbell, C., 1999, "Large Electorates and Decisive Majorities", *Journal of Political Economy*, 107, 1199-1217.
- Casella, A. and A. Gelman, 2008, "A Simple Scheme to Improve the Efficiency of Referenda", *Journal of Public Economics*, 92, 2240-2261.
- Casella, A., A. Llorente-Saguer and T. Palfrey, 2012, "Competitive Equilibrium in Markets for Votes", *Journal of Political Economy*, 120, 593-658.
- Casella, A. and T. Palfrey, 2019, "Trading Votes for Votes. A Dynamic Theory", *Econometrica*, 87, 631-652.
- Chamberlain, Gary and Michael Rothschild, 1981, "A Note on the Probability of Casting a Decisive Vote", *Journal of Economic Theory*, 25, 152-162.
- Clarke, E.H., 1971, "Multipart Pricing of Public Goods", *Public Choice*, 17-33.
- Coleman, J., 1966, "The Possibility of a Social Welfare Function", *American Economic Review*, 56, 1105-22.
- Dahl, R., 1956, *A Preface to Democratic Theory*, Chicago: University of Chicago Press.
- d'Aspremont C. and L.A. Gérard-Varet, 1979, "Incentives and incomplete information", *Journal of Public Economics*, 11, 25-45.
- Ferejohn, J., 1974, "Some Notes on the Theory of Vote Trading," *Social Science Working Paper No. 41*, California Institute of Technology.
- Gelman, Andrew, Jonathan N. Katz and Francis Tuerlinckx, 2002, "The Mathematics and Statistics of Voting Power", *Statistical Science*, 17, 420-435.
- Goeree, J., and J. Zhang, 2017, "One Person, One Bid", *Games and Economic Behav-*

ior, 101, 151-171.

Good I.J. and Lawrence S. Mayer, 1975, "Estimating the Efficacy of a Vote", *Behavioral Science*, 20, 25-33.

Green J. and J.J. Laffont, 1979, *Incentives in Public Decision-Making*, Amsterdam: North-Holland.

Groves, T., 1973, "Incentives in Teams", *Econometrica*, 41, 617-631.

Haefele, E., 1971, "A Utility Theory of Representative Government", *American Economic Review*, 61, 350-65.

Hortala-Vallve, R., 2012, "Qualitative Voting", *Journal of Theoretical Politics*, 24, 526-54.

Jackson, M. and H. Sonnenschein, 2007, "Overcoming Incentive Constraints by Linking Decisions", *Econometrica* 75, 241-57.

Koehler, D., 1975, "Vote Trading and the Voting Paradox: A Proof of Logical Equivalence," *American Political Science Review*, 69, 954-960.

Koford, K., 1982, "Centralized Vote-Trading", *Public Choice*, 39, 245-68.

Krishna, V., 2002, *Auction Theory*, Cambridge, Ma: Academic Press.

Krishna, V. and J. Morgan, 2015, "Majority Rule and Utilitarian Welfare", *American Economic Journal: Microeconomics*, 7, 339-375.

Lalley, S. and G. Weyl, 2018a, "Nash Equilibria for Quadratic Voting", unpublished, Microsoft Research.

Lalley, S. and G. Weyl, 2018b, "Quadratic Voting. How Mechanism Design Can Radicalize Democracy", *American Economic Review, P&P*, 108, 33-37.

Ledyard, J., and T. Palfrey, 2002, "The Approximation of Efficient Public Good Mechanisms by Simple Voting Schemes", *Journal of Public Economics*, 83, 153-72.

Mailath G. and A. Postlewaite, 1990, "Asymmetric Information Bargaining Problems with Many Agents", *Review of Economic Studies*, 57, 351-67.

Margolis, Howard, 1977, "Probability of a Tie Election", *Public Choice*, 31, 135-138.

Miller, N., 1977, "Logrolling, Vote Trading and the Paradox of Voting: A Game Theoretical Overview," *Public Choice*, 30, 51-75.

Mueller, D., 1973, "Constitutional Democracy and Social Welfare", *Quarterly Journal of Economics*, 87, 61-79.

Parisi, F., 2003, "Political Coase Theorem", *Public Choice*, 115, 1-36.

Park, R., 1967, "The Possibility of a Social Welfare Function: Comment", *American Economic Review*, 57, 1300-1304.

Philipson, T. and J. Snyder, 1996, "Equilibrium and Efficiency in an Organized Vote Market", *Public Choice*, 89, 245-65.

Posner, E. and G. Weyl, 2015, "Voting Squared: Quadratic Voting in Democratic Politics", *Vanderbilt Law Review*, 68(2), 441-499.

Riker, W. and S. Brams, 1973, "The Paradox of Vote Trading", *American Political Science Review*, 67, 1235-47.

Sen A. *On Economic Inequality*. Oxford: Clarendon Press; 1973.

Schwartz, T., 1975, "Vote Trading and Pareto Efficiency," *Public Choice*, 24, 101-110.

Tullock, G., 1970. A simple algebraic logrolling model, *American Economic Review*, 60, 419-426.

Vickrey, 1961, "Counterspeculation, auctions, and competitive sealed-tenders", *Journal of Finance*, 16, 8-37.