Democracy and Intensity of Preferences. A Test of Storable Votes and Quadratic Voting on Four California Propositions

Alessandra Casella and Luis Sanchez*

September 16, 2019

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 and Quadratic Voting propose possible solutions. We test their performance in two samples of California residents using data on four initiatives prepared for the 2016 California ballot. As per design, both systems induce some minority victories while our measure of aggregate welfare increases, relative to majority voting, and ex post inequality in welfare declines.

*Columbia University, ac186@columbia.edu, and Cornell University, las497@cornell.edu. For useful comments, we thank Bora Erdamar, who gave the initial impetus to the project, Andrew Gelman, Antonin Mace', and participants to the Columbia Experimental Lunch, the 2018 ETH Workshop on Democracy, and the 2019 LSE Workshop on Behavioral Political Economy. We thank the National Science Foundation (grant SES-0617934) for financial support. The research was approved by Columbia University Institutional Review Board.

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. In our present times, populism's "radical majoritarianism" (Urbinati 2019) raises new concerns on how to express and 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.

Summaries of the theory behind the two voting schemes, as well as additional details on the experimental implementation and on the data, screenshots of the survey, and additional results can all be found in the Appendix.

1 The Survey

In May 2016 we selected the following 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. (Finally, the proposition was not included in the 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.[1]

After this first part of the survey, common to all respondents, subjects were randomly

[1] The 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.

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 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 automatically 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 simple utilitarian welfare. We constructed it on the basis of the points assigned by respondents to each proposition at the end of the first part of the survey. The common total of 100 points is a normalization preventing factors of scale from distorting welfare. 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 number of 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.

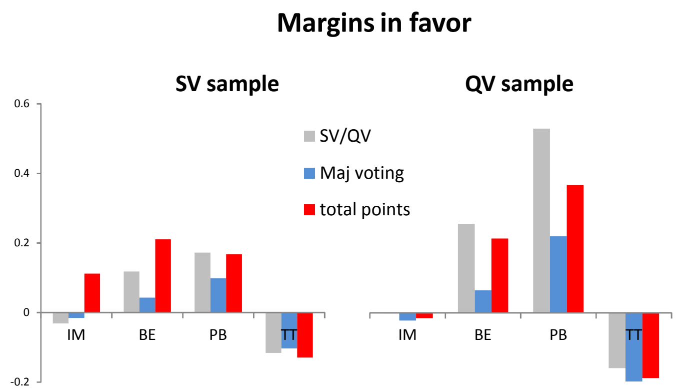


Figure 1: *Margins in favor.* Two-sided KS tests assessing whether the distributions of points are drawn from the same population yield p-values equal to 0.629 (IM), 0.66 (BE), 0.092 (PB), 0.384 (TT).

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 (aggregating over all propositions, both voting schemes realize 60.4% of full surplus).

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.

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 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.

Generating voting outcomes by matching individuals with their SV or QV choices is the obvious option, and the first one we consider. An additional goal, however, is to evaluate the robustness of the voting schemes to a range of plausible behaviors. With this in mind, we posit four alternative rules-of-thumb governing the use of the votes (see the appendix for details): A, as just mentioned, i.e. as the individual did in the original sample; B, according to two reduced-form statistical models, one for SV and one for QV, that summarize the regularities in the data as probabilities of making different choices; C, as optimal under some simplifying assumptions; D, introducing randomness in rule C.

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). 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).

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

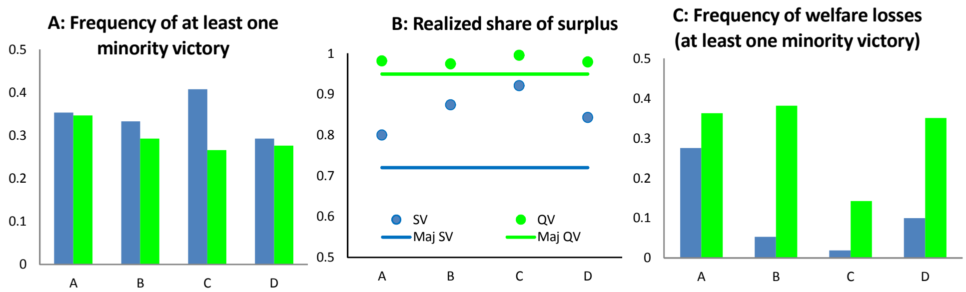


Figure 2: *Bootstrap results*. In panel A, the frequency is significantly higher for SV under rules B, C, and D ($p < 0.01$; one-sided Z test). In panel B, none of the differences in means are statistically different from one another. Note the difference in surplus under majority in the SV and QV samples. 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 discrepancy is surprising because the two samples were populated randomly during the MTurk survey, and yet majority voting performs better in the QV sample, and as a result welfare losses are more frequent under QV. To compare SV and QV, we need to eliminate such discrepancy, something we can achieve by combining 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 3: 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 3: 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 3: C).

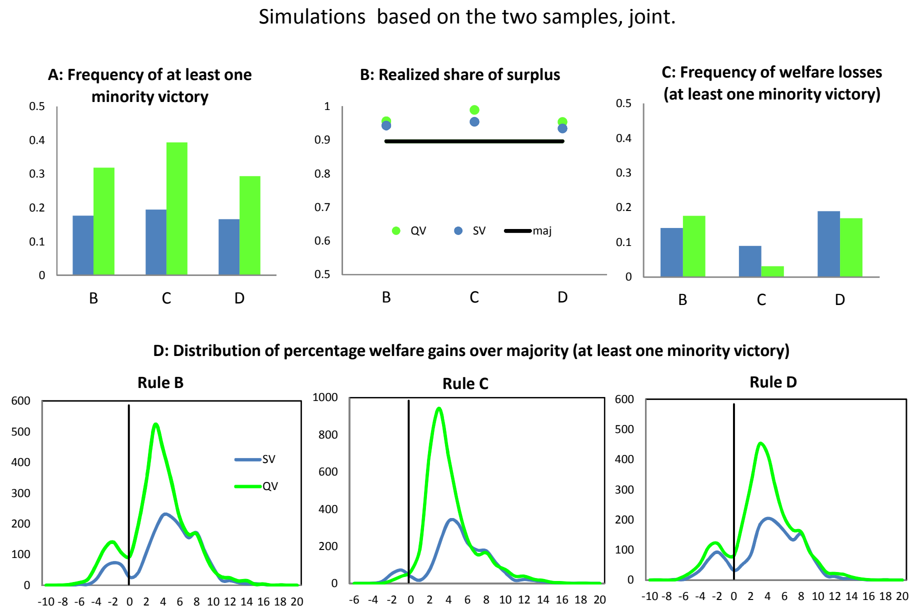


Figure 3: *Joining the two samples.* A: the fraction of samples with at least one minority victory is consistently higher under QV ($p < 0.001$). B: none of the differences in means are statistically different from one another. C: The frequency of welfare losses is significantly higher under QV for rule 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.

Averaging over the three rules, 13% of all simulations with at least one minority victory 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 3: D). A natural question is the extent to which this higher sensitivity reflects 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. Regular votes move the voting scheme towards majority voting. In our data this is advantageous under QV because respondents select higher weight vote classes much more often than theory prescribes. But adding regular votes to QV also makes it closer to SV, inviting questions on the trade-off between complexity and surplus gains.[2]

[2]However, a puzzling aspect of our data is the weak performance of SV-A, for which we have no expla-

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 is on the winning side, weighted by the importance (the number of points) the voter assigns to each. Contrary to populism, 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.

On the whole then our data confirm 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 normatively desirable.

References

Casella, A., 2012, *Storable Votes,* Oxford Un. Press: New York, NY.

Casella, A. and A. Gelman, 2008, "A Simple Scheme to Improve the Efficiency of Referenda", *Journal of Public Economics*, 92, 2240-2261.)

Dahl, R., 1956, *A Preface to Democratic Theory*, Chicago: University of Chicago Press.

Goeree, J., and J. Zhang, 2017, "One Person, One Bid", *Games and Economic Behavior,* 101, 151-171.

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

Quarfoot, D., D. Kohorn, K. Slavin, R. Sutherland, D Goldstein and E. Konar, 2017, "Quadratic Voting in the Wild: Real People, Real Votes", *Public Choice*, 172(1), 283-303.

Urbinati, N., 2019, "Political Theory of Populism", *Annual Review of Political Science*, 22: 111-127.

nation.

3 Appendix

3.1 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 $\mathrm{v}_{ik} > 0$ indicates that i is in favor of the proposal, and $\mathrm{v}_{ik} < 0$ 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 $v_{ik} = |\mathrm{v}_{ik}|$, 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 $\{\mathrm{v}_{i1},..,\mathrm{v}_{iK}\}$ are privately known. They are a random sample from a joint distribution $\mathcal{F}(\mathrm{v}_1, \ldots, \mathrm{v}_K)$ 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.

3.1.1 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 $F_k(\mathrm{v})$. 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 $F_k(\mathrm{v}) = F(\mathrm{v})$ for all k, where $F(\mathrm{v})$ 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 $F_k(\mathrm{v})$ varies with k, but for all k $F_k(\mathrm{v})$ is symmetric around a zero median. (3) If $F_k(\mathrm{v}) = G(\mathrm{v})$ for all k, where $G(\mathrm{v})$ is symmetric around a random median with expected value at 0.

With independent voters and large N, assumptions about the shape of the distributions $F_k(\mathrm{v})$ have immediate implications about the results of the referenda. In particular, assuming specific medians for the distributions $F_k(\mathrm{v})$ 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 ($\mathrm{v}_{ik} > 0$), and $1-\psi_k$ of being against ($\mathrm{v}_{ik} < 0$). 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 $|\mathrm{v}_{ik}| \equiv v_{ik}$ 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:

$$\omega \equiv \frac{EW^{SV} - ER}{EW - ER} \tag{1}$$

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 jth order statistic among each individual's k intensities, it is then possible to derive:[3]

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

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.

3.1.2 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.[4]

[3] 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.

[4] 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 for any number of votes they cast. 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; 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:

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

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.

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

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

of votes cast:

$$\frac{\partial p_{ik}(x_{ik})}{\partial x_{ik}} \equiv q_k \tag{3}$$

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

$$x_{ik} = \frac{q_k v_{ik}}{\lambda_i}$$

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:

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

and thus:

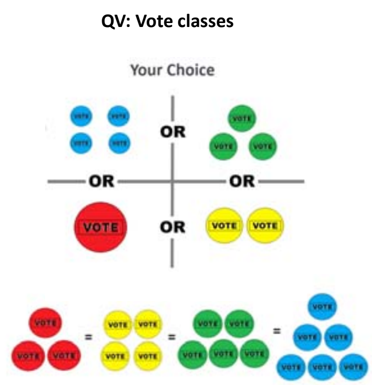
$$x_{ik} = \frac{1}{\sqrt{\sum_{k=1}^{K}(v_{ik})^2}} v_{ik} \tag{4}$$

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 they are 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. 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.

3.2 Implementation of QV in the experiment.



The design of the QV scheme in the MTurk survey.

3.3 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 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.

3.3.1 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.

3.3.2 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

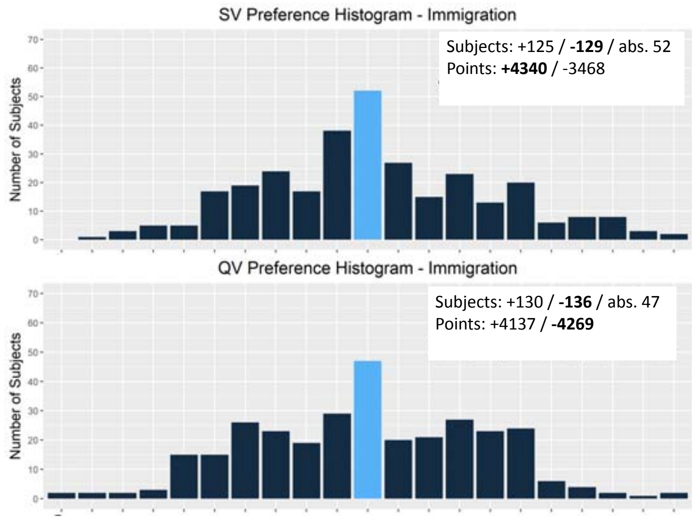


Figure 4: *Distribution of preferences: the IM proposition.*

voted in favor, with bins of size 10 (0, colored light blue in the figures, corresponds to abstentions). Figure 4 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 5, 6, and 7 report the histograms for the other propositions.

3.3.3 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 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

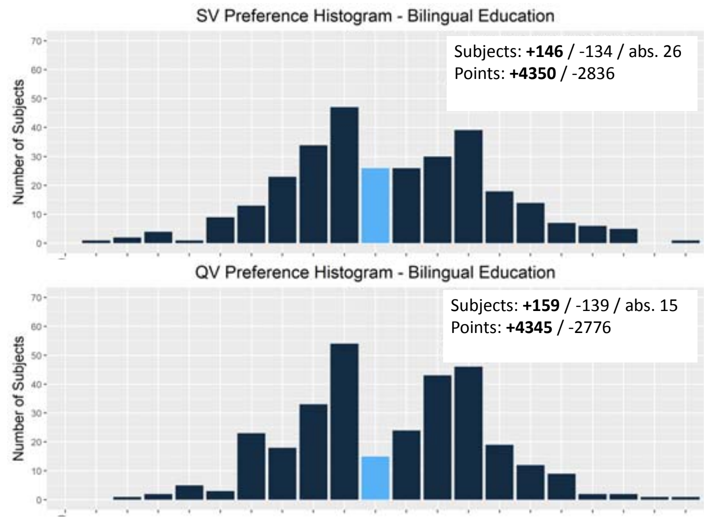


Figure 5: *Distribution of preferences: the BE proposition.*

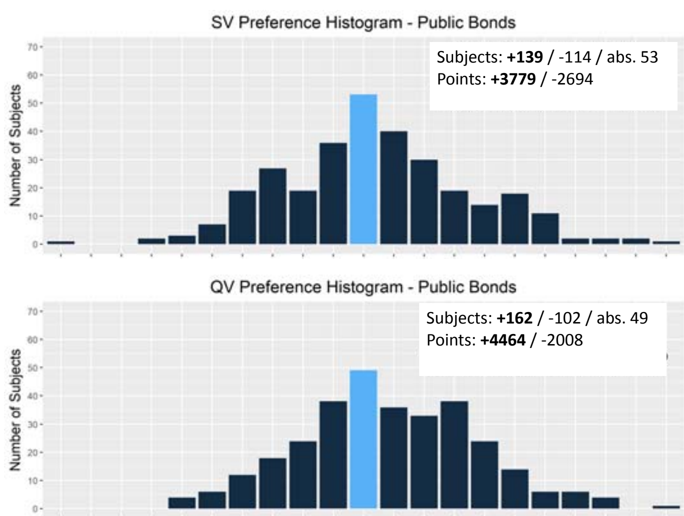


Figure 6: *Distribution of preferences: the PB proposition.*

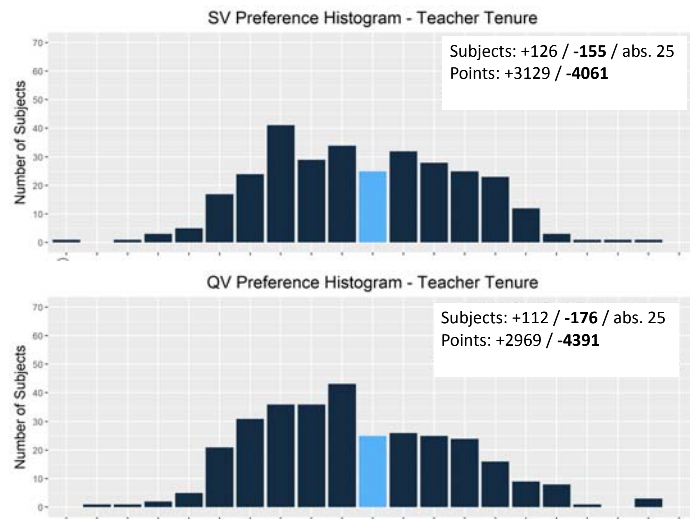


Figure 7: *Distribution of preferences: the TT proposition.*

proposals according to some marginal distribution F symmetric around zero (Casella and Gelman 2008). Figure 8 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.[5]

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,

[5]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.

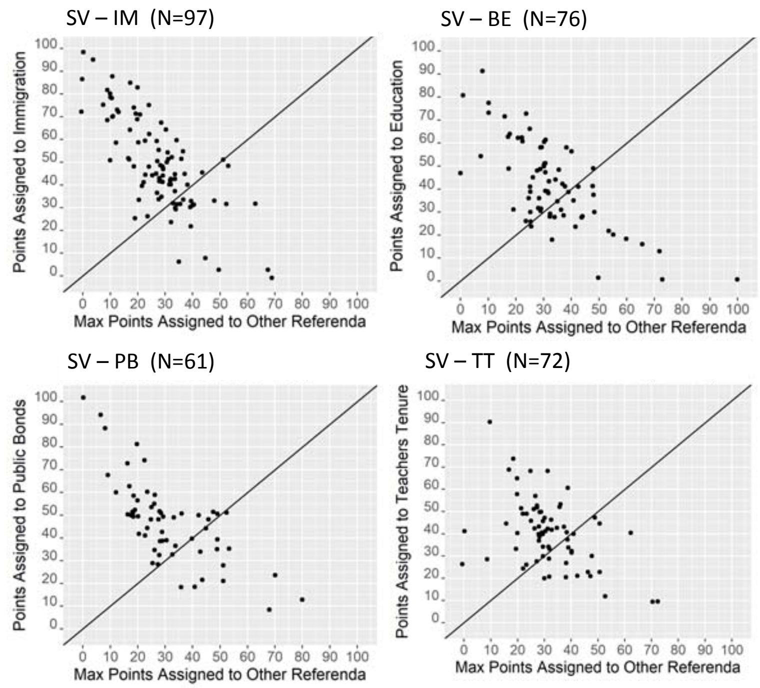


Figure 8: *The Bonus Vote Decision.*

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.

We summarize the voting choices through a statistical model, to be read as compact representations of the data. 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 1.

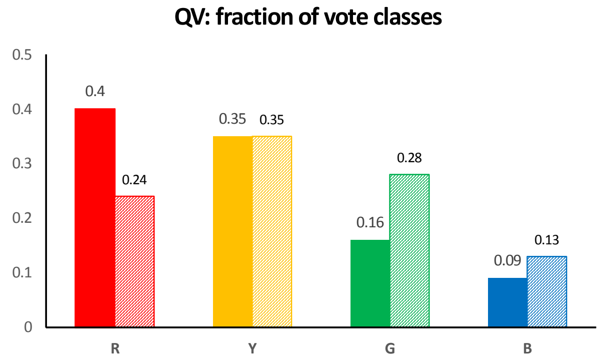


Figure 9: *QV sample. Frequency of vote classes.* The solid columns correspond to the observed frequencies in the MTurk sample. The striped columns correspond to optimal choices, given the observed distributions of preferences, if voters perceive the marginal pivotal probability to be constant (as under rule-of-thumb QV-C, described below).

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 1. *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 9 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 intensity.

Denoting by $b_{(k)}$ the voter's kth 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 2.

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 2. *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.

3.4 Bootstrap simulations

3.4.1 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 1 and 2, 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 3.1.2, 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 $\left\{2v_{i(4)}/\beta_i, 2v_{i(3)}/\beta_i, 2v_{i(2)}/\beta_i, 2v_{i(1)}/\beta_i\right\}$.[6] Given a vote class, votes should be assigned monotonically.

As a check, we reestimated the QV statistical model reported in Table 2 after having imposed the normative choice on the MTurk sample. As shown in Table 3 below, estimated thresholds ρ and γ are significantly higher than in Table 2, supporting the hypothesis that MTurk respondents concentrated votes excessively.[7]

		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 3. *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.

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

[6] 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 $\widetilde{v}_{ik} = v_{ik}/\left(\sum_{k=1}^{K} v_{ik}\right)$, then $\widetilde{v}_{ik}/\sqrt{\sum_{k=1}^{K}(\widetilde{v}_{ik})^2} = v_{ik}/\sqrt{\sum_{k=1}^{K}(v_{ik})^2}$.

[7] 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 ($\mu = 0$).

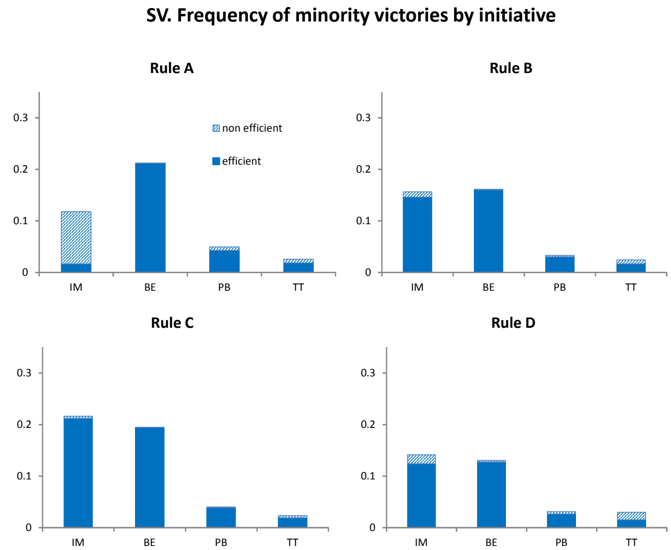


Figure 10: *Minority victories by initiative. SV simulations.* In each panel, the solid (dashed) part of each column is the fraction of minority victories that are efficient (inefficient). The efficient frequency of minority victories by initiative is: 0.52 (IM), 0.24 (BE), 0.05 (PB), and 0.04 (TT).

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.)

3.4.2 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 4, 5, 6, and 7 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 10-11 below report the results.

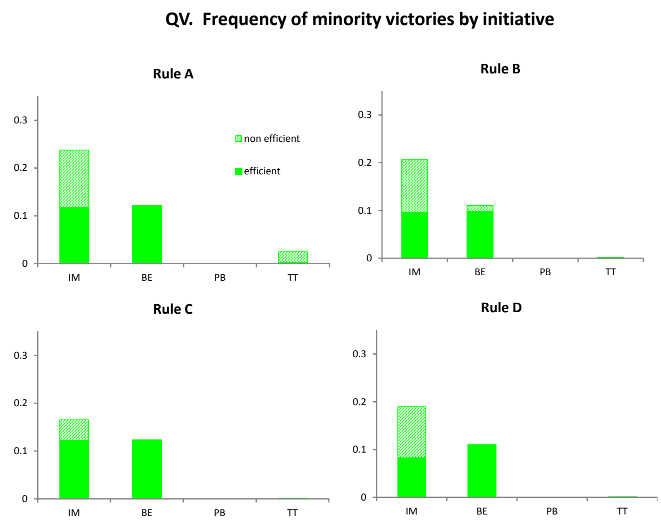


Figure 11: *Minority victories by initiative. QV simulations: bootstrapping the original QV sample*. In each panel, the solid (dashed) part of each column is the fraction of minority victories that are efficient (inefficient). The efficient frequency of minority victories by initiative is: 0.18 (IM), 0.12 (BE), 0.00 (PB and TT).

3.5 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 3.1.2, 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. **??**: A), with a corresponding decline (from 31% to 18%) in the

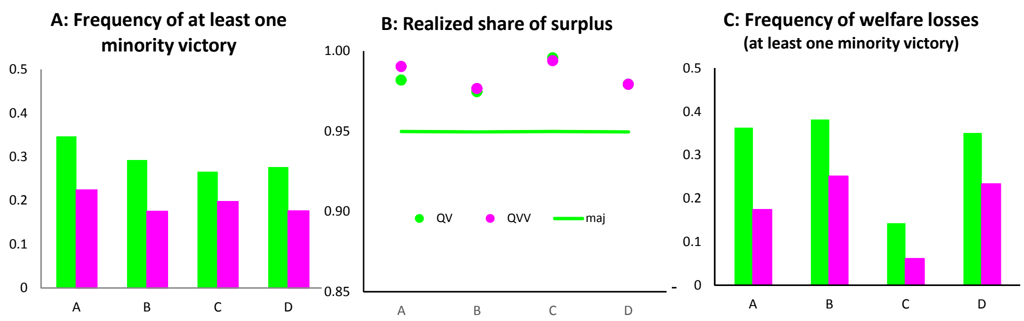


Figure 12: *Boostrapping the original QV sample: QV and QVV.* The fractions of samples with at least one minority victory (Panel A) or with welfare losses, conditional on at least one minority victory (Panel C) are consistently lower under QVV ($p < 0.001$ in all cases). In Panel B, the average shares of surplus are not statistically different from majority (based on bootstrapped CI's). All tests of proportions are one-sided Z tests.

frequency of welfare losses, in the simulations with at least one minority victory (Fig. **??**: C). The impact on the average realized share of surplus is negligible (Fig. **??**: B).

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 13).

3.6 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.

To evaluate SV and QV's impact on inequality, we calculate the Gini coefficient of the

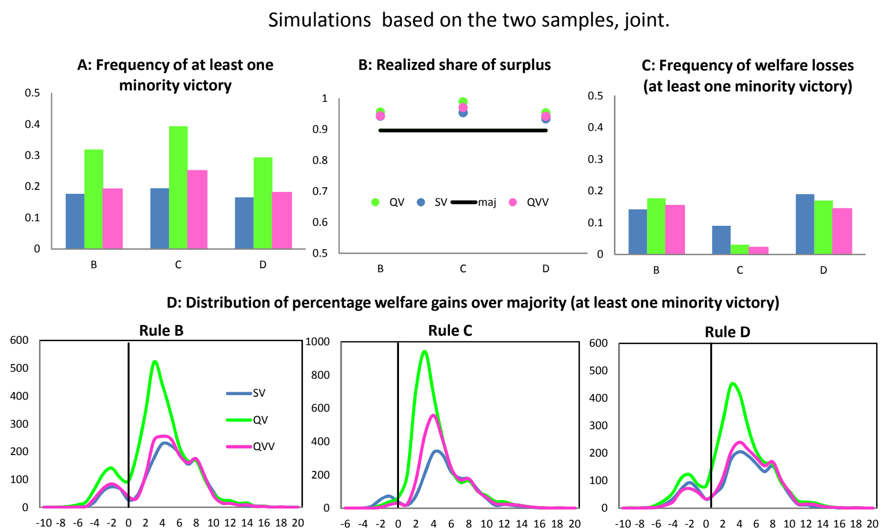


Figure 13: *Comparing SV, QV and QVV over the two samples, joint.* In Panel A, the fraction of samples with at least one minority victory remains consistently lower under SV than under QVV ($p < 0.01$). In Panel B, none of the differences in mean surplus are statistically significant from one another (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 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.

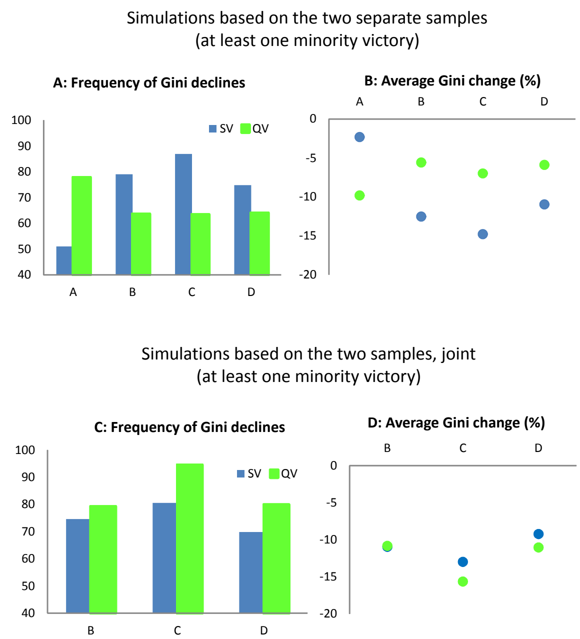


Figure 14: *Frequency and magnitude of Gini declines*. Panels A and C: With the exception of SV-A, the probability of a Gini decline is statistically higher than 50% for both SV and QV, and the differences between SV and QV are all significant ($p < 0.01$; one sided Z-tests). Panels B and D: The average magnitudes of the Gini declines are not statistically significant (based on bootstrapped CI's. See SI).

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 14:A) and increases to 75% for SV and 85% for QV in the simulations based on the two samples combined (Figure 14: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 14:B), and a decline of 11% and 12% respectively when bootstrapping the two samples jointly (Figure 14:D). In both sets of simulations, the impact of QVV is statistically indistinguishable from QV (Tables 4 and 5). The summary message is that both SV and QV have a positive impact on ex post inequality.

Average Gini coefficients are reported in Tables 4 and 5.[8]

SV

		95% CI
maj	0.37	[0.32, 0.43]
A	0.35	[0.28, 0.40]
B	0.32	[0.28, 0.40]
C	0.315	[0.28, 0.39]
D	0.33	[0.28, 0.40]

QV

		95% CI
maj	0.32	[0.27, 0.37]
A	0.29	[0.26, 0.345]
B	0.30	[0.26, 0.35]
C	0.30	[0.26, 0.35]
D	0.30	[0.26, 0.35]

QVV

		95% CI
maj	0.33	[0.27, 0.37]
A	0.29	[0.26, 0.35]
B	0.30	[0.26, 0.35]
C	0.30	[0.26, 0.35]
D	0.30	[0.26, 0.35]

Table 4. *Gini coefficient: Simulations based on the two separate samples (at least one minority victory).* The confidence intervals are obtained by bootstrapping.

SV

		95% CI
maj	0.35	[0.29, 0.38]
B	0.31	[0.27, 0.36]
C	0.30	[0.27, 0.36]
D	0.31	[0.27, 0.37]

QV

		95% CI
maj	0.35	[0.30, 0.38]
B	0.31	[0.27, 0.36]
C	0.29	[0.27, 0.35]
D	0.31	[0.28, 0.36]

[8] 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 Table 4 and 5 correspond to the subsamples with at least one minority victory under rule C, but the changes are vey minor and the conclusions identical if we use any other rule.

QVV

		95% CI
maj	0.35	[0.31, 0.38]
B	0.31	[0.27, 0.37]
C	0.30	[0.27, 0.36]
D	0.31	[0.28, 0.36]

Table 5. *Gini coefficient: Simulations based on the two samples joint (at least one minority victory).* The confidence intervals are obtained by bootstrapping.

References to the appendix

Casella, A. and A. Gelman, 2008, "A Simple Scheme to Improve the Efficiency of Referenda", *Journal of Public Economics*, 92, 2240-2261.

Chamberlain, Gary and Michael Rothschild, 1981, "A Note on the Probability of Casting a Decisive Vote", *Journal of Economic Theory*, 25, 152-162.

Gelman, Andrew, Jonathan N. Katz and Francis Tuerlinckx, 2002, "The Mathematics and Statistics of Voting Power", *Statistical Science*, 17, 420-435.

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

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.

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

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