Storable Votes and Quadratic Voting. An Experiment on Four California Propositions

Alessandra Casella and Luis Sanchez

August 25, 2018

Abstract

Storable Votes and Quadratic Voting are two voting systems that elicit and reward voters' intensity of preferences. Both apply when voters face multiple, binary choices and work by granting a total budget of votes and allowing some discretion in the number cast on each election. We test their performance in a Mechanical Turk experiment where two samples of California residents reported their preferences and voting choices over four initiatives being prepared for the 2016 California ballot. Both voting systems make minority victories possible and both result in higher welfare relative to majority voting. Quadratic voting induces more minority victories and achieves higher welfare, but also causes more inefficient minority victories. Testing new voting systems is hampered by the scarcity of data. We overcome the problem by bootstrapping the original samples and generating two sets of 10,000 multi-elections simulations, a methodology applicable more broadly.

1 Introduction

From the Brexit referendum in the UK to Trumpian politics in the US, to the *En Marche* movement in France, to *Cinque Stelle* in Italy, Western democracies show a clear decline of traditional political parties. If, as Cinque Stelle for one theorizes, we move to a world where political choices are subject to immediate electronic polling, we need to rethink how to protect the minority: increased reliance on direct democracy invokes new fears of the *tyranny of the majority,* a classic problem in political design.[1] [**Refs in fn**] The protection of basic rights

[1] Federalist Papers (Hamilton and Adams), Mill, Dahl(?), Guinier.

is the domain of the judicial branch. Here we are concerned instead with the expression and weighing of political preferences. There are different reasons for such concern–fairness and legitimacy are foremost, in diverse societies where racial, ethnic or religious identities can make some groups disproportionately likely to be on the minority side. Straightforward utilitarian efficiency is an additional one, and is the object of this study: in a binary decision, the numerical size of the two opposing groups does not account for varying intensities of preferences and majority voting can lead to an inefficient outcome (**refs?**)

An ideal solution would not superimpose a different set of rules or institutions–quota systems, pre-set rotation in positions of power, vetoes–but amend the voting system to allow for the expression of preference intensity. Outcomes would then reflect preferences endogenously, mirroring changes in preferences over time, and reverting to simple majority voting when in fact intensities do not differ substantially. Recently, two such voting schemes have been proposed: Storable Votes (Casella, 2005, 2012), and Quadratic Voting (Posner and Weyl xx, Lalley and Weyl, 2018). This paper compares their performance, and compares them to simple majority voting, by eliciting preferences and voting behavior over four state initiatives in two samples of California respondents. According to our results, both voting schemes do indeed generate non-negligible minority victories and both result in consistently higher efficiency than simple majority.

Both schemes induce voters to express the intensity of their preferences by allowing them to distribute a total budget of votes across different binary decisions. In each decision, the side with more votes then prevails. Storable votes is simpler: it advocates granting voters one "regular" vote to cast on each decision, as well as a budget of "bonus votes" that can be spent as desired. In large elections, the best design grants voters a single bonus vote, in addition to regular votes. Quadratic voting was originally designed as an auction system in which voters use money to bid for the outcome of a single binary election (see also Goeree and Zhang, 2017). When applied to multiple decisions, quadratic voting grants voters a budget of "voices" to be used as numeraire–voices acquire value through their possible use in different elections. The central feature is that the cost of bundling votes is quadratic: casting one vote in an election costs one voice, but casting ten votes cost 100 voices. Under both systems, voters have an incentive to cumulate votes to express intensity, but the mechanism is somewhat tempered: with storable votes, by the existence of regular votes and, in large elections, by the single bonus votes; with quadratic voting by the convex cost of bundling.

In our experiment, we polled subjects registered on Amazon Mechanical Turk and **resident (?)** in California on their preferences and voting choices over four initiatives that, at the time of polling, were being prepared for the 2016 ballot. The four initiatives concern: (1) the reestablishment of bilingual education in public schools; (2) the cooperation of the state's enforcement agencies with federal immigration agents; (3) new requirements for voters' approval of large public infrastructure projects; and (4) longer required teaching experience for teacher's tenure. We collected data simultaneously from two samples of about 300 respondents each, exposing one to storable votes and one to quadratic voting.

Because the voting systems rely on the single budget constraint tying the different voting decisions, the unit of analysis is the joint four-initiative election. Our data then correspond to two data points, one for storable votes, and one for quadratic voting, obviously too few for any meaningful extrapolation. We would ideally want a large number of identical joint elections, all with electorates drawn from our population. We cannot have such data, but we can bootstrap our original samples to generate two samples, one with storable votes one with quadratic votes, of 10,000 simulated multi-elections each. This is the data set on which we base our conclusions.[2]

As mentioned earlier, both voting systems perform well: they make minority victories possible and consistently result in higher total welfare (according to our utilitarian measure) than majority voting. There are however differences between them. In our data, quadratic voting leads to more frequent minority victories and higher realized efficiency. However, quadratic voting is more likely to also cause a substantial number of inefficient minority victories. Storable votes is a more conservative system, missing some potential efficiency gains, but also overthrowing majority's decisions more rarely.

Our conclusions apply to the specific parametrizations of the two systems used in our survey. Based on what we see, however, it is tempting to suggest complementing quadratic voting with regular votes, one to be cast in each election, as we do (and the theory advocates) for storable votes. The addition of regular votes would dampen the working of quadratic voting, making outcomes less sensitive to minority preferences. In our data, they would cause negligible declines in efficiency, in exchange for fewer "mistaken" minority victories.

[2] We borrowed the methodology from Casella et al. 2010.

Whether the efficiency gains resulting from a system of "quadratic voting with regular votes" are worth the higher complexity, relative to a simple system with storable votes, we leave to the reader to evaluate.

2 The Theory

A large number N of voters are asked to vote, contemporaneously, on a set of K unrelated proposals (with $K > 1$). 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. Preferences are separable across proposal and the voter's objective is to maximize total utility U_i from the set of proposals, where $U_i = \sum_k u_{ik}$.

Each individual's valuations $\{\mathrm{v}_{i1},\mathrm{v}_{i2},..,\mathrm{v}_{iK}\}$ 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 adopt weakly undominated strategies and vote sincerely with no abstention. We consider three voting systems: majority voting, Storable Votes and Quadratic Voting. In all three, each proposal is decided in the direction preferred by a majority of the votes cast on the proposal. 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.

2.1 Storable Votes

Storable Votes (SV) is a voting system that grants each voter a budget of "bonus votes" to be distributed as desired over the different proposals (Casella, 2005, 2012; Hortala-Vallve, 2012). In large elections, it is typically optimal to cumulate all bonus votes on one proposal[3], and the simplest implementation is then to endow each voter with a single bonus vote, to spend as desired, and complement it with K "regular" votes, one per proposal. If voters' valuations are independent across voters and proposals, Casella and Gelman (2008) shows

[3] Casella and Gelman (2008).

that ex ante welfare in the unique symmetric Bayesian Nash equilibrium is higher than ex ante welfare under majority voting, under a range of plausible assumptions about the distributions $F_k(\mathrm{v})$. (We summarize the arguments in the online appendix). In particular, as is to be expected, the benefits of SV increase if there are asymmetries in intensities between voters on the opposite side of a proposal, and especially so if the minority holds more intense preferences–not coincidentally the case in which the desirability of simple majority voting is most disputed.

2.2 Quadratic Voting

Quadratic Voting (QV) has been proposed in the context of a single binary election: voters purchase votes with a numeraire and pay a price that equals the square of the number of votes purchased. Goeree and Zhang (2017) and Lalley and Weyl (2018) show that if valuations are independent across voters, the symmetric Bayesian Nash equilibrium, conjectured unique, converges asymptotically to full utilitarian efficiency. As in the case of SV, the possibility of casting more than one vote, at a cost, allows for the expression of a voter's intensity of preferences and gives it weight over the final outcome.

There is no theoretical analysis of the efficiency properties of QV in multiple elections, where bidding can be expressed in an artificial currency. In this design, first described by Posner and Weyl (2015), each voter is granted a budget of "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. QV becomes similar to SV, but the quadratic cost limits the incentive to cumulate votes. Although no rigorous results exist, a simple model (suggested by Glen Weyl and reported in the online appendix) shows that efficiency extends 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 that the condition is generally not satisfied in equilibrium, but the deviations may be too subtle for voters to take into account.

3 The Experiment

In May 2016 we selected four initiatives submitted for inclusion in the November 2016 California ballot. We then recruited 647 subjects in California via Amazon's Mechanical Turk (MTurk), and randomly assigned them either to the SV treatment (324 subjects; 306 remained after data cleaning[4]) or to the QV treatment (323 subjects; 313 after cleaning). We asked each subject how she would vote on each of the four initiatives, presented in random order, allowing for the option to abstain, and how important each of these initiatives was to her. The sample assigned to the SV treatment was then presented with a version of SV; the second sample with a simplified QV scheme; both were asked on which initiatives they would cast these votes. Outcomes were computed using simple majority, and either SV or QV. In both samples, 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. The questions were programmed in Qualtrics, and copies of all screenshots are reproduced in the online appendix.

3.1 The four initiatives

Our goal was to identify multiple initiatives whose outcome was unlikely to be a landslide, about which some voters at least would feel strongly, and that would be clear enough to the average MTurk subject. In March 2016 we presented an original set of ten initiatives, all with the potential to reach the November ballot, to a sample of 94 California MTurk subjects. Based on their responses, we selected the following four initiatives:[5]

(1) *Bilingual education (BE):* abolish the requirement that basic education be offered in English only and re-instate the possibility of bilingual classes in public schools. The initiative was included in the November 2016 ballot and passed.

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

(3) *Teachers' tenure (TT)*: increase the required years of pre-tenure experience for teachers from two to five. The initiative 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, including those financed by tolls

[4] See the online Appendix for details on the data cleaning criteria.

[5] The March survey were used exclusively to select the four initiatives we submitted to the May samples. The exact wording with which we summarized the four initiatives in the May survey is on the screenshots in the online appendix.

and fees generated by the projects themselves. The initiative was on the November 2016 ballot and failed.

3.2 Direction and intensity of preferences

The two novel voting rules appeared only in the second part of the survey. The first part–eliciting voters' direction and intensity of preferences–allowed us to compute outcomes under majority voting and construct a measure of utilitarian welfare. The latter is our performance meter. We constructed it as follows.

In both samples, we asked each subject to distribute 100 points among the four initiatives, with the number of points meant as a scale of the importance attributed to each proposal ("How important are these issues to you?"). We used examples to make clear that importance is unrelated to preferred direction and we summarized responses in terms of priorities, allowing for revisions and asking for a final confirmation. We interpret points as proxy for valuations, and the common total of 100 points as a normalization required to prevent factors of scale from distorting welfare.[6]

Denoting by r_{ik} the number of points attributed to initiative k by individual i, we set $v_{ik} = r_{ik}$. Thus $U_i^S = \sum_{k:i\in M_k^S} r_{ik}$, where $S \in \{Majority\ Voting, SV, QV\}$ indicates the voting scheme and M_k^S the side casting a majority of votes on k under S. Realized aggregate welfare W^S is then given by $W^S = \sum_i \sum_{k:i\in M_k^S} r_{ik}$. In this perspective, full utilitarian efficiency requires that each initiative be won by the side which collectively values it most, or $W^* = \sum_i \sum_{k:i\in R_k} r_{ik}$ where R_k denotes the side with higher total number of points on k. For each voting scheme S, our primary performance measure is thus the ratio W^S/W^*. We should note here that although this is the most natural–in fact the only–measure we can construct, it remains imperfect even if our questions are understood correctly and answered with no error. Respondents presumably use their votes thinking of the broad Californian electorate, while we can only compare the voting schemes' effects on the MTurk samples.

3.3 The voting schemes

The second part of the survey was designed to test the new voting rules, under specific parametrizations. In the SV sample, subjects are told that each is

[6] In the same spirit to the constraint that all valuations be drawn from the same distribution in the theoretical model.

granted a *bonus vote*–one extra vote–in addition to the regular votes cast earlier. With no reference to voting weights, the bonus vote is understood to be equivalent to a regular vote. Each subject is asked to choose the initiative in which to use it; the vote is then cast automatically in the direction chosen in the first part of the survey. The final outcome is calculated summing regular and bonus votes.

The design of the QV scheme is less straightforward because it must convey clearly that cumulating votes has a quadratic cost. We present the subjects with four different classes of votes, distinguished by color, and ask them to choose one class. Blue votes are regular votes, four in number. Votes cannot be cumulated, and thus a person choosing blue votes casts one vote in each initiative. 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. A subject who chooses green votes must abstain in one of the initiatives. Yellow votes are two, each stronger than a green vote, but choosing yellow votes means abstaining on two initiatives. Finally, a subject can choose to cast a single red vote, stronger than a yellow vote, but forcing the subject to abstain in three of the four initiatives. More precisely, the weights of the different votes are 1 for blue votes, 1.2 for green votes, 1.5 for yellow votes, and 2 for the red vote.

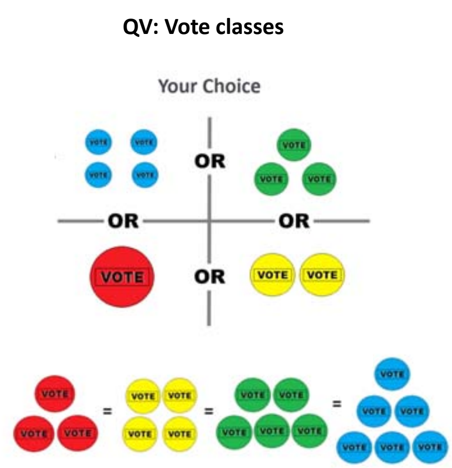


Figure 1: *The design of the QV scheme in the MTurk survey.*

The simple four-class classification respects the convex cost of concentrating votes. A voter using all four votes on different initiatives–choosing blue votes–has a total weight of 4, but the total weight declines as votes are concentrated: a voter using all her voting power on a single referendum has a total weight of 2, the square root of 4, exactly in line with QV logic. For the intermediate cases of concentration on three or two initiatives, the total weights are 3.6 and 3. The decline in total voting weight is increasing with concentration, and increasing at an increasing rate, capturing the core feature of QV. On the other hand, our streamlined version of QV forces subjects to cast votes of equal weight on all referenda on which the subject casts votes, a constraint that is not part of the original idea. Especially in the uncontrolled and fast MTurk environment, the hope is that our simplified QV may benefit from its easier comprehension more than it loses in flexibility.

We asked each subject to choose a class of votes, and then select the initiative(s) on which the vote(s) were to be cast. As with SV, the 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.[7]

3.4 The experimental data

The survey yielded distributions of preferences–empirical counterparts to the theoretical distributions $F_k(\mathrm{v})$–and voting outcomes. We reproduce the histograms of respondents' preferences in the online appendix. Here we summarize both preferences and voting choices by reporting the percentage margin in favor of each initiative, in terms of aggregate points, average points, number of voters, and number of votes under either SV or QV (Figure 2).[8]

[7] At the end of both the SV and QV survey, we reported the voting results over the four initiatives from our March poll of 96 subjects, including the poll's margins of errors, and asked whether subjects wanted to change their allocation of the bonus vote (or choice of QV class and vote allocation). At the 95% confidence level, margins of errors were +/- 10%, and no outcome could be ruled out. About 10% of subjects in each sample changed their vote but the net vote changes in each initiative were negligible. We base the analysis that follows on the data collected before showing the poll.

[8] Representing results in terms of margins accentuates the differences between the two samples. For all four initiatives, two-sided KS tests cannot reject that the distributions of points are drawn from the same population, with boostrapped p-values equal to 0.629 (IM), 0.66 (BE), 0.092 (PB), 0.384 (TT). Because each subject must allocate 100 points, within each sample the distributions are not independent, although with four initiatives no individual distribution is redundant.

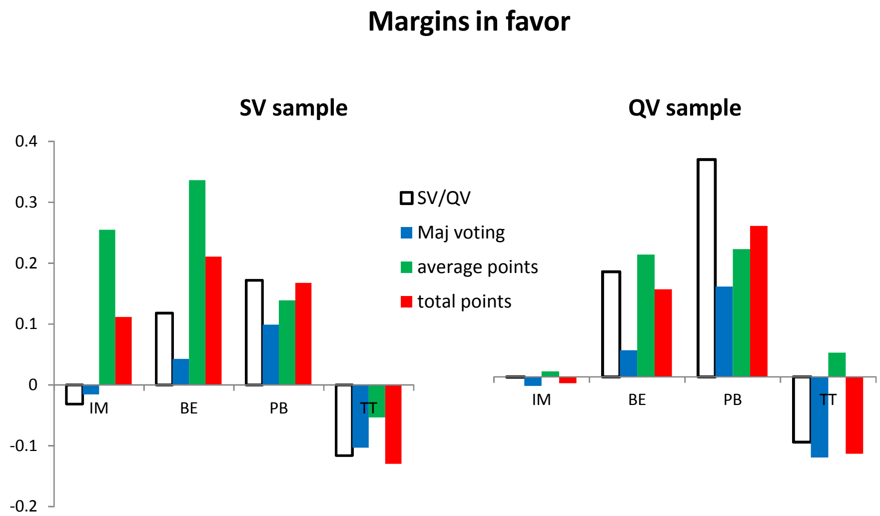


Figure 2: *Margins in favor.*

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 initiative is very small. In both samples and all initiatives, the outcome is unchanged whether using majority voting, SV, or QV. When the aggregate point margin has the same sign as the majority voting outcome (all four initiatives in the QV sample; BE, PB and TT in the SV sample), the outcome is efficient. Thus both majority voting and QV achieve full efficiency in the QV sample, while both majority and SV fall short in the SV sample (both appropriate 94.7 percent of efficiency). When the margin in terms of average points also has the same sign as the majority voting outcome, the more numerous side has higher average intensity, and since SV and QV are expected to reflect intensities, not only should both voting schemes confirm the majority voting outcome, but they should do so with a larger margin (in absolute value)–as is the case for BE, PB and TT in the SV sample, and BE and PB in the QV sample. If, on the contrary, the average points margin has the opposite sign, as in IM in both samples and TT in the QV sample, then the margin should be smaller than under majority–as indeed occurs for both initiatives in the QV sample, but not for IM in the SV sample.

The IM initiative 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.[9] It is also the most salient: as reported in the online appendix, it is allocated 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. In addition, the voting result under SV shows that in the IM initiative the bonus vote was not used symmetrically by supporters and opponents: supporters' higher total intensity is not reflected in the SV results, and the inefficiency of majority voting is not corrected. [10]

As noted, on all four initiatives, both SV and QV confirmed the outcome reached with simple majority voting. The observation, however, is not very informative: with both voting systems, the votes cast across initiatives are tied by a budget constraint. Effectively, we have a single data point in each sample. Evaluating the impact of the two voting systems requires much larger samples– ideally we would want to replicate the same elections many times, with many different electorates whose preferences are all drawn from the same underlying distribution. We cannot rerun the elections, but, following the methodology proposed by Casella et al. (2010), we can approximate such iterations by bootstrapping our data.[11]

4 10,000 Multi-election Samples

The objective is to estimate the impact of the voting rules in a population for which our samples are representative.[12] The maintained assumption of the bootstrapping exercise is that preferences are independent across individuals, but not necessarily across elections for a single individual. We sample with replacement N individuals from each of our data sets, where $N = 306$ for SV and $N = 313$ for QV. For each individual, we sample the direction of preferences over each initiative, the number of points assigned to each, and the votes cast

[9] 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.

[10] On average, opponents who cast their bonus vote on IM did so at lower intensity than supporters. (43.7 points for opponents versus 55.9 for supporters, a significant difference (p=0.0034)).

[11] The classical references are Efron (1979) and Efron and Tibshirani (1993). For a more recent treatment, see Davidson and MacKinnon (2006).

[12] Of the two initiatives that did appear on the ballot in November 2016, one (BE) was decided as in our samples, the other (PB) was not. The representativeness of MTUrk workers has been debated (Bartneck et al., 2015, Clifford et al., 2015, Huff and Tingley, 2015), and our results confirm such doubts. However, for our purposes–the performance and comparison of SV and QV–the imperfect representativeness of the samples is not by itself problematic.

over all initiatives, according to either the SV or the QV scheme, depending on the data set. We then replicate this procedure 10,000 times for each original data set, SV or QV. Each replication generates a distribution of preferences over each initiative and a voting decision for each voter, and thus a voting outcome for all four initiatives. The focus is on the fraction of simulations in which the two voting rules reach different results from majority voting, and on their welfare properties.

Generating voting outcomes by matching individuals with their SV or QV voting choices in the original samples is a natural option, and the first one we consider in our simulations, but an important goal of the simulations is to test the robustness of the two voting schemes to a range of plausible behaviors in casting votes. Without claims of optimality or micro-founded sources of bounded rationality, we posit four alternative rules-of-thumb governing the use of SV and QV. They are: (A): as the individual did in the original sample; (B) as in a descriptive statistical model of the original samples; (C) as optimal in a simplified environment; (D) introducing randomness in rule C.

Rule (B) imposes some structure on the data we collected and adds some noise. Rule (C) supposes, for both SV and QV, that individuals take their probability of pivotality as constant, across initiatives and, in QV, across the number and weight of the votes they cast. Under SV, the optimal rule is then to cast the bonus vote on one's highest intensity initiative. 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. (See the online appendix for detailed descriptions of all rules).

Voters could potentially be fully random–i.e., although voting according to their preferences, they could choose randomly the initiatives on which to cast the votes at their disposal (and with QV, they could choose randomly the vote class too). The scenario provides a plausible floor to the performance of the voting systems, relative to majority voting. We do not report the results in what follows, but have verified what intuition suggests: only close elections would be overturned, with equal frequency of efficiency gains and losses. Over the 10,000 simulations, average welfare under both SV and QV replicates average welfare under majority voting. We consider this case a useful check, both on the properties of SV and QV and on our simulations.

Simulation Results

In both sets of simulations, with all four rules, both SV and QV resulted in frequent minority victories (Figure 3:A).[13]

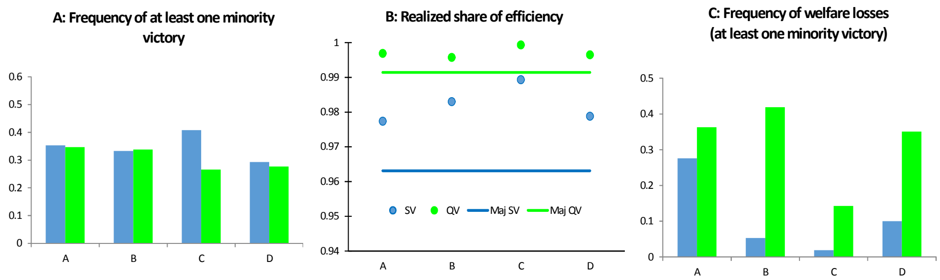


Figure 3: *Bootstrap results.* Panel A reports the frequency of samples with at least one minority victory. Minority victories were concentrated in the IM and BE initiatives (see online appendix for details). Panel B reports the average share of efficiency over the full 10,000 simulations. Note the difference in majority welfare in the SV and QV samples. Panel C reports the frequency of SV and QV samples with lower welfare than majority voting among those with at least one minority victory.

More than one fourth of the 10,000 simulations in each data set, using any rule, had at least one minority victory: the average across rules was 31 percent for QV and 35 for SV. Remarkably, both voting systems consistently delivered welfare gains over majority voting (Figure **3:**B). Averaging across rules, the realized share of efficiency was 98.2% for SV and 99.7% for QV, compared to 96.3 and 99.14% for majority in the two sets of simulations. Majority does very well in these data, and particularly so in the QV simulations, and the fact that both voting systems succeed in improving over majority is impressive, when so little foregone surplus is available.

However, many minority victories come 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, a percentage that rises to almost a third for QV (32%) (Figure 3:C). SV is more conservative, foregoing potential gains (on average, SV

[13] Standard errors do not appear in the figures because with 10,000 bootstrap samples the sampling error is negligible. As a test, we recalculated the results with 20,000 independent bootstrap samples, with barely detectable differences.

appropriates 55% of available improvement over majority, v/s 66% for QV), but also inducing less frequent losses.

A proper comparison between the two voting systems is difficult because majority voting does substantially better in the QV simulations [**test of diff in means?**]. The reason is the difference in the original samples in the preferences over the IM initiative. As we described, in the SV sample the majoritarian outcome in the IM initiative is inefficient: the proposal fails but the intensity of preferences among supporters is sufficiently strong to overcome the numerical superiority of the opponents. Such is not the case in the QV sample. This disparity carries over to the bootstrapped samples.

We can correct the problem by recalibrating the samples in the bootstrap exercise.

4.1 Comparing SV and QV: Recalibrating the QV samples

We construct 10,000 recalibrated QV samples by restricting the distributions of preferences over the IM initiative so as not to differ, on average, from the average distribution generated when populating the SV samples. For each QV simulation, as before, we draw with replacement 313 subjects from the QV sample, but now constrain each draw to come from each bin of the QV IM preference histogram with a probability equal to the corresponding frequency in the SV IM preference histogram. Since a draw is a subject, each simulation corresponds to a sample of 313 preferences and voting choices over all four initiatives.[14] [**How do we check that the other distrs are not distorted?**]

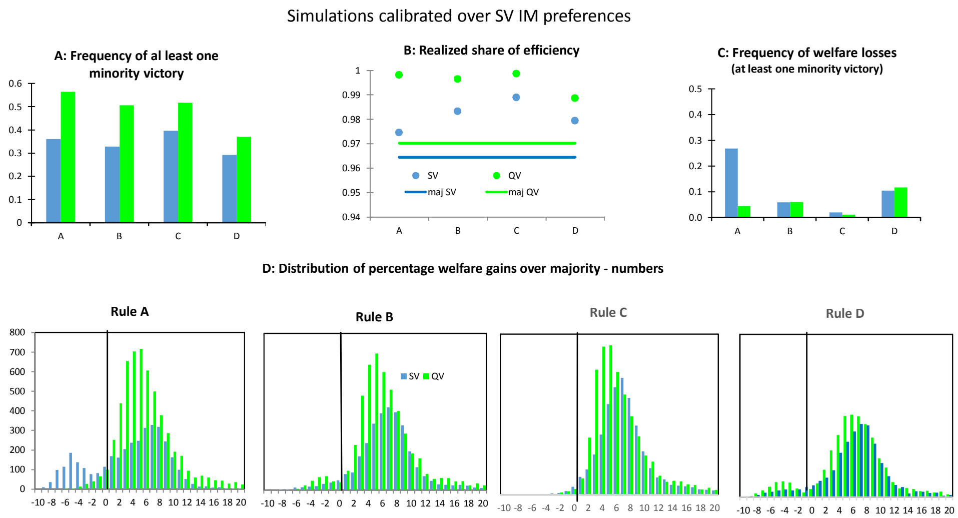
The recalibrated simulations lead to a substantially higher frequency of minority victories under QV–now consistently more frequent than under SV (Figure **??**: A). For three of the rules–A, B, and C–under QV the minority wins at least one initiative in more than half of all simulations, a frequency we never observe with SV. In terms of appropriated surplus, QV now reaches, on average, 99.5% of efficiency, with majority voting at 97% (versus 98 for SV, with majority at 96.5). Again, majority voting does very well in these samples and outperforming it across all rules-of-thumb is remarkable. With QV's recalibrated sample, the room for improvement over majority voting, if still small, is larger than in the original sample, and the average frequency of welfare losses under QV falls drastically to 6% of all simulations with minority victories (v/s

[14] Note that because subjects are drawn with their full set of preferences over all initiatives, by imposing constraints over the IM distribution, we could be distorting the distributions in the QV samples over the other initiatives.[**We verified that such is not the case. How?**]

11% with SV). Over all 10,000 samples, the frequency of losses is small for both voting systems–at 3.9% for QV and 2.6% for SV (minority victories are less frequent under SV).

SV's somewhat weaker performance is due to rule A. As noticed earlier, in the original SV data the use of the bonus vote is biased against the IM initiative, relative to the number of points assigned: supporters assign more points than opponents, but cast fewer bonus votes. In the simulations that allocate the bonus votes according to the subjects' original choice (rule A), the IM initiative fails disproportionately– it fails more than efficiency dictates (in 68% of the samples, v/s less than 10% under efficiency), and more than under majority (60%). Under rule A, close to 90% of all simulations in which SV induce losses relative to majority voting are due to the inefficient minority victory of opponents of the IM initiative. The anomaly appears clearly if we plot the distribution of welfare gains over majority voting (Figure **??**:D). No such problem arises with QV, where results are more consistent across rules.

If the difference between SV and QV under rule A is driven by asymmetries in the original SV sample, under the other rules the difference is more representative of the underlying properties of the two voting systems. Consider rule C in Figure **??**:D. We drew the histograms in terms of absolute numbers, as opposed to proportions, to highlight the higher number of minority victories under QV, as well as the values of the welfare changes corresponding to such higher number. The two distributions differ almost exclusively because of the higher spikes under QV corresponding to small welfare gains: 85 percent of difference in the number of samples in which minority victories are realized under QV correspond to welfare gains over majority voting not larger than 5%. The two voting systems differ in their sensitivity to intensity of preferences. If we draw the histograms in terms of relative frequency of welfare gains, SV under rule C FOSD's QV: whenever minority victories are realized, they tend to be associated with large gains.



Bootstrap results with recalibrated QV sample. In panel D the scale of the histograms is constant. The black vertical line separates losses and gains.

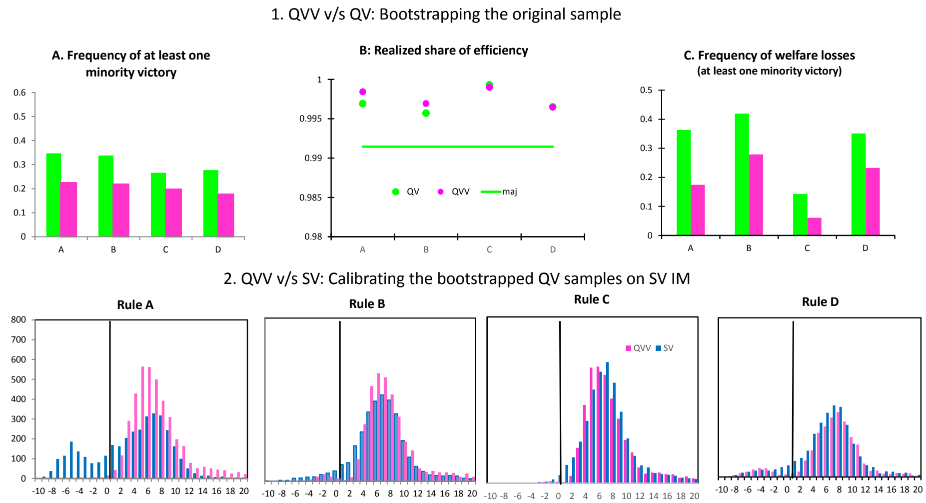
The observation invites two questions. First, both voting systems should be evaluated according to the welfare criterion we have chosen, and QV does very well in these simulations. From the point of view of practical implementation, however, the legitimacy of majority decisions has value, and one may ask whether reversing a majority decision for small welfare gains is wise. Second, and related, how much does QV's higher frequency of minority victories depend on the exact parametrization employed in the experiment? For SV, we have followed closely Casella and Gelman (2008), where background "regular" votes–one vote per decision–are added to blunt the impact of the bonus vote. With QV, the convex cost by itself moderates the incentive to cumulate votes, and formal QV models do not include background votes. But formal QV models do not address multiple decisions, and the intuitive model in the online appendix would accommodate background votes with no change [**Verify**]. Adding such votes would favor the majority and reduce the fraction of minority victories. Could reducing QV's sensitivity to intensity be desirable?

4.2 QV cum regular votes

The MTurk survey described QV as consisting exclusively of the classes of votes chosen by the respondents, without additional ordinary votes. Applying the subjects' responses to a different environment with regular votes is thus not fully legitimate. It is true however that under the maintained assumption of constant marginal pivot probability neither a voter's optimal choice nor the efficiency properties of the voting system would change. In line with our focus on robustness and rules-of-thumb, we investigate in this section how the simulations results would be affected by endowing each voter with four additional votes. These votes, one on each initiative, are cast according to the voter's declared preferences in the first part of the survey, as we do for SV. We call the voting system QVV–*QV with Vote*.

We begin by comparing QVV to QV in the simulations obtained from bootstrapping the original sample, where QV induced frequent welfare losses. Predictably, under QVV minority victories become less frequent: averaging across all rules, the frequency falls from 31 to 20.5% (Figure **??**: 1A). And yet the realized share of efficiency does not fall (the average moves from 99.7 to 99.8) while the frequency of samples with lower welfare than under majority does decline by more than a third (from 32 to 19% on average) (Figures **??**: 1B and 1C).[15] (**Need fn**?)

[15] The impact of adding regular votes is confirmed in the re-calibrated simulations: minority victories are less frequent than under QV and when they happen, the frequency of welfare losses declines. Efficiency remains higher than under majority under every rule, with the average realized ratio of full efficiency effectively unchanged relative to QV (at 98.5 versus 98.8%, compared to 97% for majority voting).



QVV: comparison to QV and to SV.

Our results thus suggest that tempering QV's sensitivity to intensity of preferences by adding ordinary votes is probably desirable: by attributing some weight to majority preferences, QVV reduces the frequency of minority victories, but especially of inefficient minority victories, without substantial effects on overall efficiency.

But if that is the case, isn't then SV, with its much simpler design, a preferable choice? The difficulty in answering the question is that SV's simulations are disturbingly sensitive to the rule we posit in casting the bonus vote. Under rule A, the asymmetric behavior of voters on the IM initiative translates into a frequency of welfare losses that is a multiple of what we observe under QV or QVV (Figure **??**: 2A. Under the other rules, on the other hand, especially under C and D, the performances of SV and QVV are very similar (Figure **??**: 2B,2C and 2D).

QV and QVV's performances are more consistent across rules. It could be that QV's design, with or without background votes, is itself responsible for the more consistent outcomes: the convex cost in concentrating votes may focus voters' attention, relative to the cost-free allocation of the bonus vote under SV–and in this case QV's observed consistency recommends it, especially when accompanied by regular votes. Or it may be that the surprising asymmetries in SV-rule A are just noise: the IM initiative is very close and changing a handful of data points would affect the results. And in this case, the simplicity of SV, together with its rather conservative behavior, may make it a more pragmatic choice.

Conclusions

The institutions of representative democracy are in decline. If, as seems plausible, the future brings more reliance on direct democracy, electronic and not, the protection of minority interests assumes new urgency. Simple voting schemes that allow minority victories when, and only when, minority preferences are strong and majority preferences are not hold the potential to improve both legitimacy and efficiency. Storable votes and quadratic voting are two such schemes. Both are designed to incentivize and reward the expression of voters' intensities of preferences when multiple, independent binary elections are on the ballot and function by endowing voters with a budget and allowing discretion in choosing the number of votes to cast on each of the elections. Ceteris paribus, the two schemes encourage casting more votes on decisions on which a voter feels more strongly. And because decisions are resolved in favor of the side with more *votes*, as opposed to more *voters*, minority victories are possible. In both cases, the theory is promising, but is also complex and incomplete.

This paper reports the results of a Mechanical Turk experiment that elicited subjects' preferences and voting choices over four initiatives intended for the 2016 California ballot. After having expressed their preferences, subjects were asked to vote over the four initiatives, using either storable votes or quadratic voting. By bootstrapping our data, we generated two samples of 10,000 joint elections (i.e. 10,000 instances of voting on all four initiatives) on which the performances of the two voting systems can be compared, and compared to simple majority voting.

We find that both systems result in non-negligible minority victories: between 30 and 50% of all samples have at least one initiative won by the minority side. And while making minority victories possible, both systems consistently raise our measure of total welfare, appropriating between one half and two thirds of the available welfare improvement over majority voting. In the parametrizations we implement, quadratic voting is more sensitive to minority preferences than storable votes: it induces more minority victories and appropriates a larger share of the surplus. Such sensitivity however comes at a cost: inefficient minority victories are more frequent too. If blunting the sensitivity of quadratic voting to minority preferences is desirable, this can be achieved by adding to the scheme a set of regular votes, one to be cast in each election, as in the storable votes design. The result is a smaller frequency of minority victories, and in our data, on balance, a negligible impact on efficiency.

In detail in the text and briefly here, we have described the impact of storable votes and quadratic voting on aggregate welfare. A second possible dimension of analysis, however, is the impact of the voting systems on ex post inequality, measured as inequality in realized utilities, or satisfied preferences weighted by declared intensity. A particularly strong motivation for voting schemes that allow minority victories is the possibility of systematic minorities–groups that are regularly confined to the losing side of most public decisions. Majority voting effectively disenfranchises systematic minorities, a state of affairs that can not only damage efficiency but radically limit the meaning of democracy.[16] (**Check fn**) In our experiment, such a concern is measurable: if individual preferences are correlated across initiatives, majority voting will lead to inequality in the distribution of realized utilities. Voting schemes that allow for occasional minority victories are predicted to reduce such ex post inequality. We report our results in the online appendix, and indeed this is what we find. Both storable votes and quadratic voting consistently lead to declines [**numbers!**] in the Gini coefficient measured on the distribution of ex post utilities. The decline correlates with the frequency of minority victories, and thus is larger for quadratic voting, relative to storable votes, while similar for storable votes and quadratic voting with regular votes.

On the whole, the voting systems we have tested perform well, even very well, better than we had grounds to expect. But two related questions remain. The first is the cost of making the voting rule more complex, and thus less transparent. Our subjects on the whole did well, and surprisingly acted more consistently with the more complicated quadratic voting scheme than with storable votes. (We did implement a very streamlined version of quadratic voting and spent effort and time in making it clear.) But are voters more widely ready to trust new voting rules? And are we ready to trust that such rules will be used correctly? Second, and related, majority voting worked very well in both of our samples, appropriating more than 96% of full efficiency in the storable votes sample, and more than 99% in the quadratic voting sample. The fact that both storable votes and quadratic voting still outperformed majority voting is

[16] Indeed, doubts about the legitimacy of democracy typically arise from the failure of representation of systematic minorities (see for example, Dahl xxx)

impressive. But, at such high realized efficiency, the need for new voting systems, clever, well-meaning, and effective, but also less transparent, less tested, and most probably less trusted is not self-evident. How much weight should be given to the established legitimacy of majority voting?

5 References

1. Bartneck, C., A. Duenser, E. Moltchanova and K. Zawieska, 2015, "Comparing the Similarity of Responses Received from Studies in Amazon's Mechanical Turk to Studies Conducted Online and with Direct Recruitment", *PLoS ONE* 10(4): e0121595. doi:10.1371/journal.pone.0121595.

2. Casella, A., 2005, "Storable Votes", *Games and Economic Behavior*, 51, 391-419.

3. Casella, A., 2012, "*Storable Votes. Protecting the Minority Voice*", Oxford Un. Press: Oxford and New York.

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

5. Casella, A, S. Ehrenberg, A. Gelman and J. Shen, 2010, "Protecting Minorities in Binary Elections: A Test of Storable Votes Using Field Data", *The B.E. Journal of Economic Analysis & Policy (Advances)*, 10(1).

6. Clifford, S, R. Jewell and P. Waggoner, 2015, "Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology?", *Research & Politics*, 2(4), 1-9.

7. Dahl, R., 1956, *A Preface to Democratic Theory*, Chicago: University of Chicago Press (third edition in 2006).

8. Dahl, R., 1989, *Democracy and Its Critics*, New Haven: Yale University Press.

9. Davidson, R. and J. G. MacKinnon, 2006, "Bootstrap Methods in Econometrics", in K. Patterson and T. C. Mills (eds.), *Palgrave Handbook of Econometrics: Vol. 1 Econometric Theory,* Palgrave Macmillan Ltd, Houndmills, Basingstoke, Hampshire RG61 6XS, UK.

10. Efron, B., 1979, "Bootstrap Methods: Another Look at the Jackknife", *Annals of Statistics*, 7, 1-26.

11. Efron B. and R. J. Tibshirani, 1993, *An Introduction to the Bootstrap*, Chapman and Hall: New York.

12. Gerber, E. R., 1999, *The Populist Paradox: Interest Group Influence and the Promise of Direct Legislation*, Princeton University Press.

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

14. Guinier, L., 1994, *The Tyranny of the Majority*, Free Press: New York.

15. Hortala-Vallve, R., 2012, “Qualitative Voting”, *Journal of Theoretical Politics,* 24(4), 526-554.

16. Huff C and D. Tingley, 2015, “Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents”, *Research & Politics,* 2(3): 1–12.

17. Issacharoff, S., P. Karlan and R. H. Pildes, 2002, *The Law of Democracy: Legal Structure and the Political Process*, Foundation Press: New York, 2nd edition.

18. Lalley, S. and G. Weyl, 2018, "Nash Equilibria for Quadratic Voting", unpublished, Un. of Chicago.

19. Matsusaka, J. G., 2004, *For The Many or The Few: The Initiative Process, Public Policy, and American Democracy,* Chicago: University of Chicago Press.

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

6 Online Appendix

6.1 Theory

6.1.1 Storable Votes

We summarize here the main results of Casella and Gelman (2008) (CG), to which we refer the reader for details. As mentioned in the text, if voters are endowed with multiple votes to distribute over multiple proposals, in a large electorate with independent values, the optimal strategy is to cumulate all votes on a single proposal (section 7.11 in CG). Thus in the case of referenda a simple design becomes desirable. Each voter is asked to cast one vote in each referendum, and in addition is given one extra bonus vote. It is natural to think of the bonus vote as equivalent to a regular vote–and that is indeed the parametrization we use in the experiment–but we can suppose, more generally, that the bonus vote is worth θ regular votes, with $\theta > 0$ and either an integer or the inverse of an integer. The optimal value of θ is part of the design of the mechanism.

A voter's strategy is a mapping from the voter's set of valuations to the vote or votes cast on each referendum. The theoretical analysis in CG restricts attention to symmetrical Bayesian equilibria in undominated strategies where, conditional on their set of valuations, all voters select the same optimal strategy. Because there can be no gain from voting against one's preferences, in these equilibria voters vote sincerely. The only decision is the referendum on which to cast the bonus vote.

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 distributions of values over each referendum. CG show that 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 $F_k(\mathrm{v}) \neq F_{k'}(\mathrm{v})$ if $k \neq k'$, but $F_t(\mathrm{v})$ is symmetric around 0 for all $t \in \{1, .., K\}$. (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 referendum 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, seems a better assumption.[17] 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 referendum.[18] 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 simple majority voting. We construct the measure:

$$\omega_s \equiv \frac{EU_s - ER}{EW_s - ER} \tag{1}$$

where the subscript s denotes the stochastic probability of approval in each referendum, EW_s 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 EU_s is ex ante expected utility under SV.[19]

[17] See for example Good and Mayer (1975), Margolis (1977) and Chamberlain and Rothschild (1981). Gelman, Katz and Tuerlinckx (2002) discuss the implications of a number of alternative models.

[18] Ruling out expected biases in voters' preferences is likely to be the correct empirical assumption because legislatures defer to the popular vote predominantly those measures whose popularity is difficult to predict (Matsusaka, 1992).

[19] If we denote by Ev the expected intensity over any proposal, and by $Ev_{(j)}$ the expected jth order statistic among each individual's k intensities, we have:

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

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 (i.e. given the properties of the value distributions).

CG show that in equilibrium voters cast their bonus vote in the referendum to which they attach the highest intensity. It is then possible to derive:

$$\omega_s = \frac{k(Ev) + \theta E v_{(k)}}{(Ev)(k+\theta)} \tag{2}$$

It follows that $\omega_s > 1$ 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 referendum 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.

6.1.2 Quadratic Voting

As described in the text, QV is an auction-type mechanism designed for a large population faced with a single binary proposal (Goeree and Zhang (2017), Lalley and Weyl (2018)). 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. Goeree and Zhang, and Lalley and Weyl show that if values are independent across voters and the distribution F is common knowledge, 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.[20]

The model relies on the use of the numeraire, valuable in the private market, and the absence of credit constraints. However, Posner and Weyl (2015) first suggested that the numeraire could be substituted by an "artificial currency" whose value derives from being the bidding currency over multiple binary proposals. In this formulation, QV becomes equivalent to SV, but with a square root rule translating the artificial currency budget into bids (or votes) over each

[20] If F is symmetric, bidding in proprtion to one's values is the unique equilibrium strategies for all voters. If F is not symmetric, the characterization of the equilibrium is more delicate, and bids in the tails of distribution need not be proportional to values. Nevertheless the efficiency results continues to hold (Lalley and Weyl, 2018).

proposal.

The theoretical extension to this case has not been worked out and is likely to be complex. Still, under some approximation its intuition is simply captured by the following model, 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 use 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 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 conditions yield:

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

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 referendum equals the voter's value, normalized by the Euclidean norm of the voter's values across all proposals.

With a welfare criterion depending on cardinal measures of utility, values need to be normalized to prevent arbitrary scaling from affecting welfare considerations. In the experiment, values are normalized by constraining each individual's values over the four initiatives to sum up to the same total of 100 points, a constraint equivalent to dividing each value by the sum of the individual's total values. That is, in the experiment, the relevant component of the welfare criterion is $v_{ik}/\sum_{k=1}^{K}(v_{ik})$. But other normalizations are plausible too, including normalizing by the Euclidean norm of the voter's values. If the relevant welfare component is $v_{ik}/\sqrt{\sum_{k=1}^{K}(v_{ik})^2}$, equation 4 immediately delivers utilitarian efficiency: because the number of votes cast in each proposal equals the voter's (normalized) value, each proposal is won by the side with larger total (normalized) values. Relative to the linear normalization, normalizing by the norm discounts extreme values. The experiment pre-dated the model and, by using a different normalization, does not capture its intuition exactly.

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.

6.2 Experimental data

6.2.1 Cleaning procedures

In designing the survey, we added an attention check to both samples. The check took the form of a fictitious fifth initiative, 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 an initiative on which they abstained–these subjects effectively abstained on all initiatives under the QV scheme, and left us no alternative. We also excluded all subjects in the SV sample who cast the bonus vote on an initiative on which they abstained–a behavior that may correspond to rejecting the use of the bonus vote, in principle a plausible choice, 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 an initiative 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 an initiative). 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 to give at least minimal weight to the direction of preferences expressed by the subject.

6.2.2 Subjects' preferences

Subjects' preferences are summarized in histograms reporting the number of respondents assigning to an initiative different number of points. Points are coded as negative values when the subject voted against the proposal, and as positive values when the subject 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 initiative for the two samples, expressed in number of subjects. 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 initiative equally low points but voted in favor (the corresponding numbers for QV are 29 and 20).

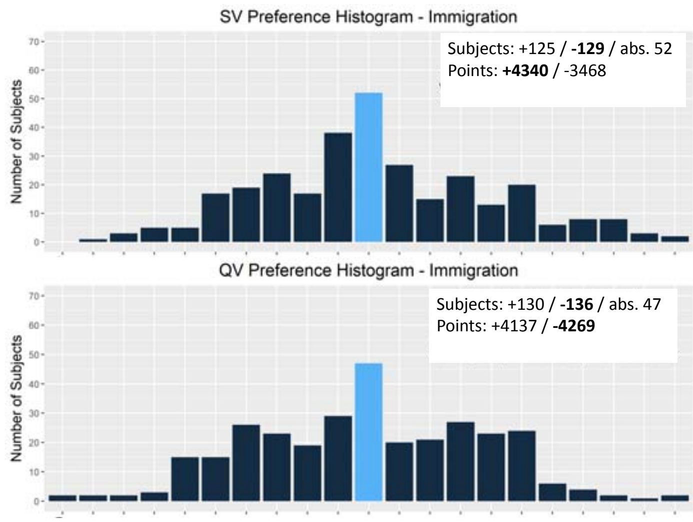


Figure 4: *Distribution of preferences: the IM initiative*

The figure also reports, in 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 initiatives.

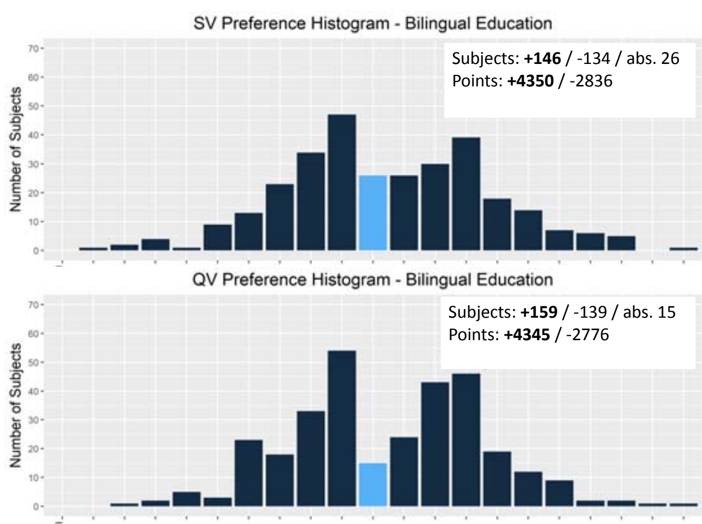


Figure 5: *Distribution of preferences: the BE initiative.*

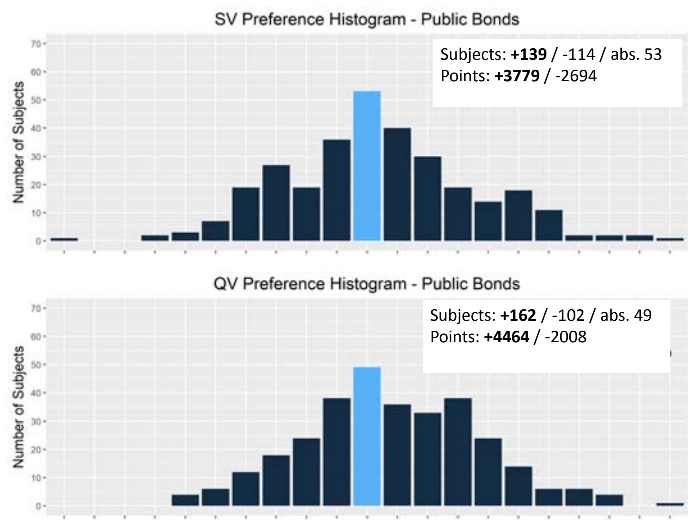


Figure 6: *Distribution of preferences: the PB initiative.*

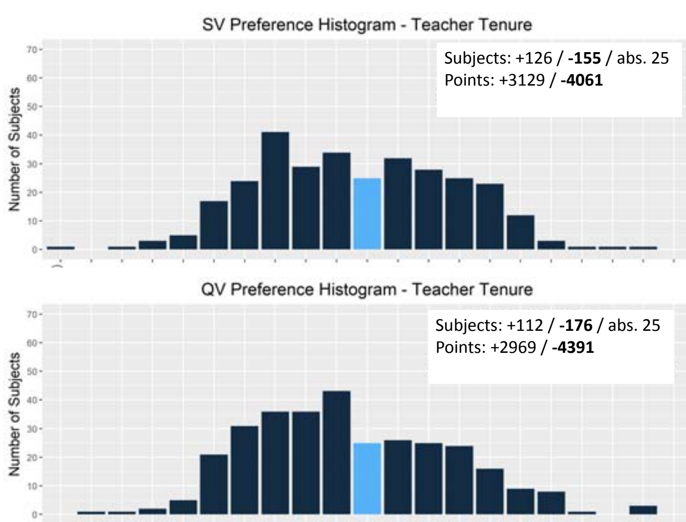


Figure 7: *Distribution of preferences: the TT initiative.*

6.2.3 The SV voting choice

In general, the optimal selection of the initiative on which to cast one's bonus vote is a complicated problem. If there are asymmetries among the initiatives–if some are expected to be more salient than others, or the vote is expected to be closer–then considerations other than intensity of preferences should influence voting choices.[21] On the other hand, if voters are not well-informed, or unable or unwilling to compute equilibria, a plausible rule-of-thumb is to treat all initiatives equally and cast the bonus vote on the proposal on which one's preferences are most intense. The approximation is analogous to the posited behavior under QV. Figure 8 shows, for each of the four initiatives, a measure of the relative intensity of preferences for all voters who cast their bonus vote on that initiative. The vertical axis is the number of points assigned to the initiative; the horizontal axis is the maximum number of points assigned to any other.[22]

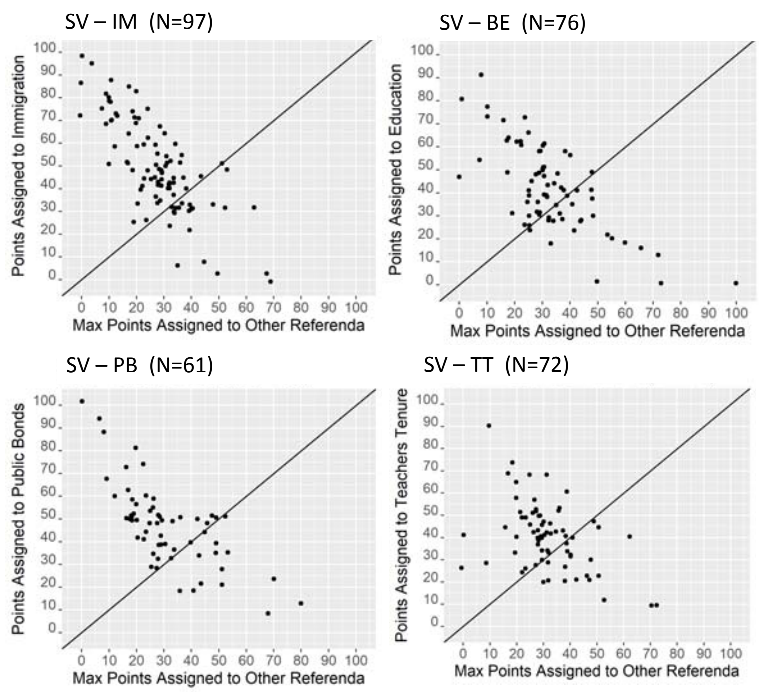


Figure 8: *The Bonus Vote Decision.*

[21] See Casella and Gelman (2008). The decision of where to cast the bonus vote reflects the probability of pivotality, which is higher in close elections and, all else equal, smaller in salient elections.

[22] A few subjects cast the bonus vote on an initiative to which they had not assigned any point. As described earlier, if they had voted on such initiative they are recoded as assigning +/- 1 point, depending on the direction of preferences.

If all voters had cast their bonus vote on the initiative to which they assigned the highest number of points, all dots in each panel would be above the 45 degree line. In total, three fourth of all subjects (74%) did so. Across initiatives, the differences are small, although the two education initiatives see a smaller fraction of bonus votes coming from highest priority voters (68% for TT and 71% for BE), relative to the other two initiatives (77% for IM and 78% for PB).

The salience of the IM initiative is supported by the high number of bonus votes (97, v/s 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 initiative increases, although IM supporters report higher average and total intensity. The result reflects two sources of asymmetry. Among those who identify the IM initiative as their first priority and yet do not target it with their bonus vote, more than twice are supporters rather than opponents; among those who cast their bonus vote on Immigration and yet do not identify it as their priority, more than twice are opponents rather than supporters[23].

6.2.4 The QV voting choices

Under QV, voters need to make two choices: the class of votes, and, given the class, the initiatives on which the votes are cast. Figure 9 reports the frequencies with which subjects chose the different classes.

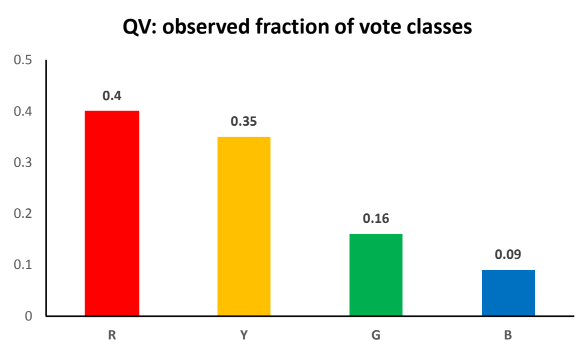


Figure 9: *Frequency of vote classes. QV sample.*

[23] 23 subjects identify the IM initiative as their highest intensity proposal and yet cast their bonus vote elsewhere; of these 23, 16 are supporters and 7 are opponents. 97 subjects cast their bonus vote on IM, but 20 of these assign to it a strictly lower number of points than to some other initiative; of these 20, 14 are opponents and 6 are supporters.

As the figure shows, even with the convex penalty from cumulating voting power, a full 40% of subjects chose the red vote, and thus cast their vote on a single initiative; less than 10% cast votes on all four initiatives. As described in detail below, we calculated the shares of the different vote classes subjects would have chosen had they followed the theoretical model skectched earlier. We find: Red=0.24, Yellow=0.35, Green=0.28, and Blue=0.13–relative to the theory, subjects chose the red vote too often.[24]

Figure 10 reports, for all subjects who cast a vote on the relevant initiative, the number of points assigned (on the vertical axis) as well as the maximum number of points assigned to any other initiatives. Each dot is a voter, and the color of the dot matches the chosen class of votes. Thus red dots denote voters who only voted on that initiative, and for whom we expect that the assigned number of points be higher than for any other initiative. For the most part, indeed, red dots are above the 45 degree line. The salience of the IM initiative is reflected by several features of subjects' vote choices: IM received the highest total number of votes among all initiatives (97 v/s 76 for BE, 61 for PB and 72 for TT), as well as the highest number of red votes (45 v/s 19 for BE, 26 for PB and 34 for TT); looking at red votes across all initiatives, those cast on IM correspond to higher average points (54.9 v/s 46.4 for BE, 50.4 for PB and 46.2 for TT)

[24] These numbers are derived applying rule (C), described in section 6.3.1 below, to the QV sample.

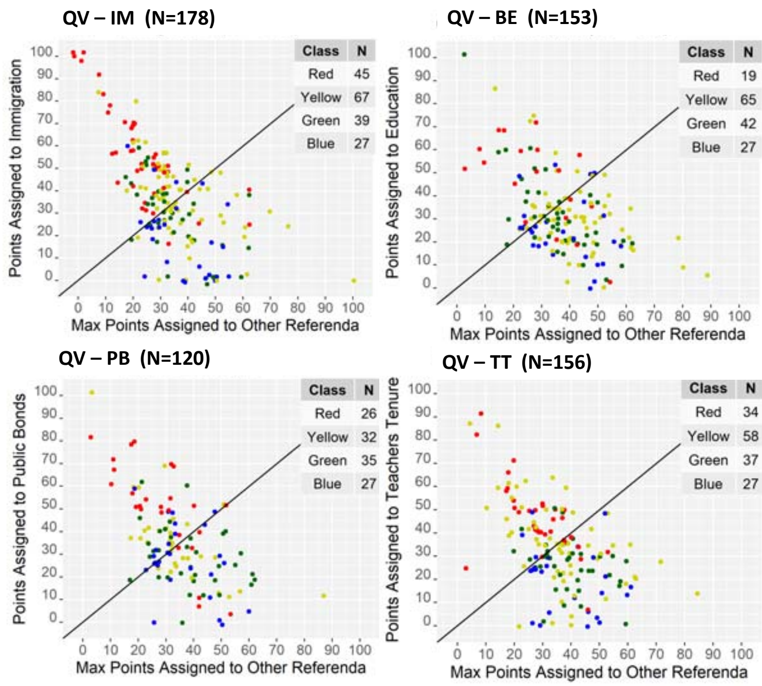


Figure 10: *The QV Vote Allocation.*

6.3 Bootstrap Simulations

6.3.1 The four rules-of-thumb

SV. Rule (A) replicates the individual's choice in the original MTurk sample and needs no further description.

Rule (B) is obtained by replicating the best-fit frequencies of reasonable but exogenously given behaviors estimated from the original samples. For SV, we conjecture the following alternative criteria: casting the bonus vote on the highest value initiative (with probability p_{Max}); or on the one with closest outcome (IM) (p_{Sal}); or on the most familiar (either BE or TT, the two education initiatives, with equal probability) (p_{Fam}); or randomly, with equal probability on all initiatives (p_{Rand}).[25] Thus, for example, in terms of our criteria the choice of a respondent who attributed maximal points to the IM initiative and cast on it the bonus vote has probability $p_{Max} + p_{Sal} + (1/4)p_{Rand}$. Each choice

[25] Casella et al. (2010) propose a similar descriptive model.

observed in the data, matched with the individual's reported valuations, can be expressed as function of the four probabilities. Assuming that respondent's choices are independent, the likelihood of observing the data is then the product of the probabilities of the individual choices. We estimated the underlying probabilities by MLE and report them in Table 6.3.1. By construction, the estimated probabilities are better understood as population frequencies of the four behavioral criteria.

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

Table

SV: MLE estimates of the statistical model. The confidence intervals are obtained by bootstrapping. They are derived from the distribution of the estimated probabilities in 10,000 simulations

In the SV simulations, we implement rule B by having each drawn subject cast the bonus vote according to the probabilities in Table 6.3.1, given the subject's allocation of points.[26]

Rule (C) is the normative reference in a simplified scenario in which voters believe that the marginal impact of their votes is equal over all initiatives. With SV, as noted in the text, this implies casting the bonus vote on the highest value initiative.

Finally under rule (D) each individual casts the bonus vote on the highest value initiative (as under (C)) with probability 1/2, and on any of the four initiatives with equal probability otherwise.

QV. Again, rule (A) replicates the individual's choice in the original MTurk sample.

Under QV, rule (B) posits that behavior is shaped by intensity of preferences only, and the thresholds determining the choice of vote class are estimated from

[26] Note that the model does not allow for asymmetries based on the direction of voters' preferences. This is why rule B typically outperforms rule A, in welfare terms: although based on the same data, it induces similar behavior for supporters and opponents of any initiative, and of the IM initiative in particular.

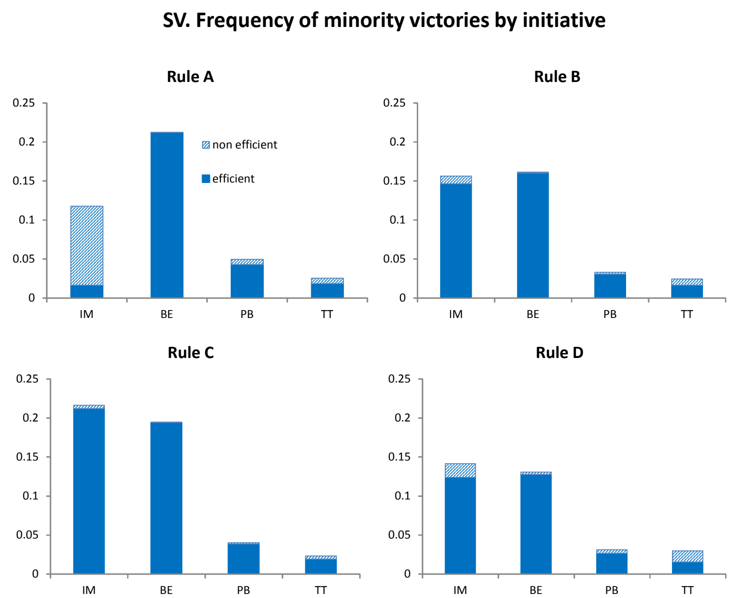
the MTurk sample. Denoting by $v_{(k)}$ the voter's kth highest value, we conjecture that behavior can be summarized by three parameters $\{\rho, \gamma, \xi\}$: the voter chooses the single red vote if $v_{(4)}/v_{(3)} \geq \rho$, the two yellow votes if $v_{(4)}/v_{(3)} < \rho$ but $v_{(3)}/v_{(2)} \geq \gamma$, the three green votes if $v_{(4)}/v_{(3)} < \rho$, $v_{(3)}/v_{(2)} < \gamma$, but $v_{(2)}/v_{(1)} \geq \xi$, and the four blue votes otherwise; given a vote class, all votes are cast on the highest intensity initiatives. We estimate $\{\rho, \gamma, \xi\}$ so as to minimize the number of misclassified individuals. A grid search (varying each parameter between 1 and 3, in intervals of 0.05) yields: $\rho = 1.25$; $\gamma = 1.05$; $\xi = 1$.

Under rule (C), as noted, voters act optimally under the belief that the marginal impact of their votes is constant. As shown in section 6.1.2, 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 $\left\{2v_{i(4)}/\beta_i, 2v_{i(3)}/\beta_i, 2v_{i(2)}/\beta_i, 2v_{i(1)}/\beta_i\right\}$.

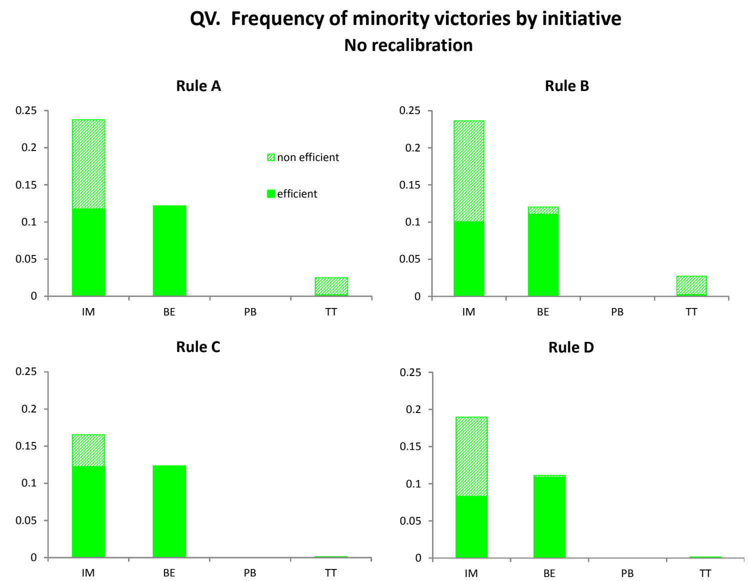
Finally, rule (D) leaves a larger role to randomness. Under rule (D) each voter behaves according to rule (C) with probability 1/2, and randomly otherwise. With SV, that corresponds to casting the bonus vote on the highest value initiative with probability 1/2, and on any of the four initiatives with probability 1/8. 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, under (D) and QV, the voter casts the available votes on the initiatives with highest values with probability 1/2, and randomly, treating all initiatives equally, otherwise.

6.3.2 Experimental results: Differences across initiatives

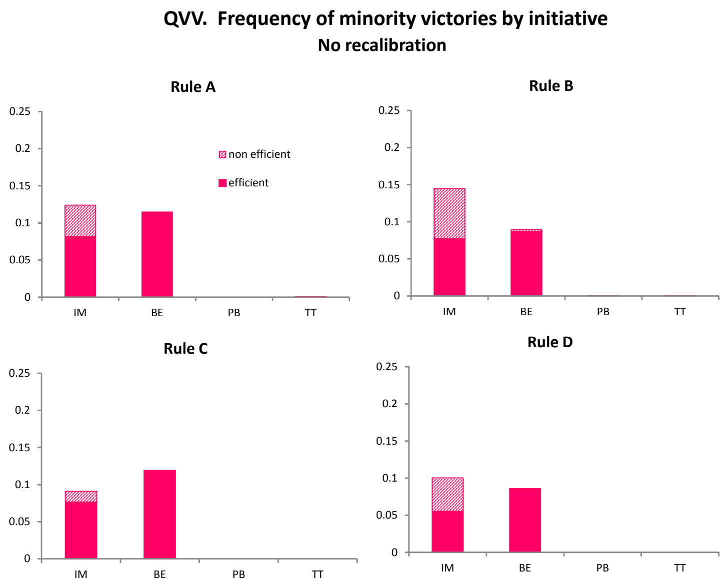
Because both SV and QV constrain the use of the votes across initiatives, each bootstrapped sample corresponds to an outcome for all four initiatives. The discussion in the text focuses on the frequency of samples in which at least one minority victory is observed, without distinguishing among initiatives. Yet there are large differences across initiatives. As the preference histograms above show, the potential for minority victories is largest in IM and BE, while SV and QV have little effect on the other two initiatives. There are also differences across the two voting systems in their impact on individual initiatives, and in particular on the IM initiative: comparing SV and QV (recalibrated), the frequency of minority victories in the IM initiative is about double under QV. Figures ??-11 below report the results.



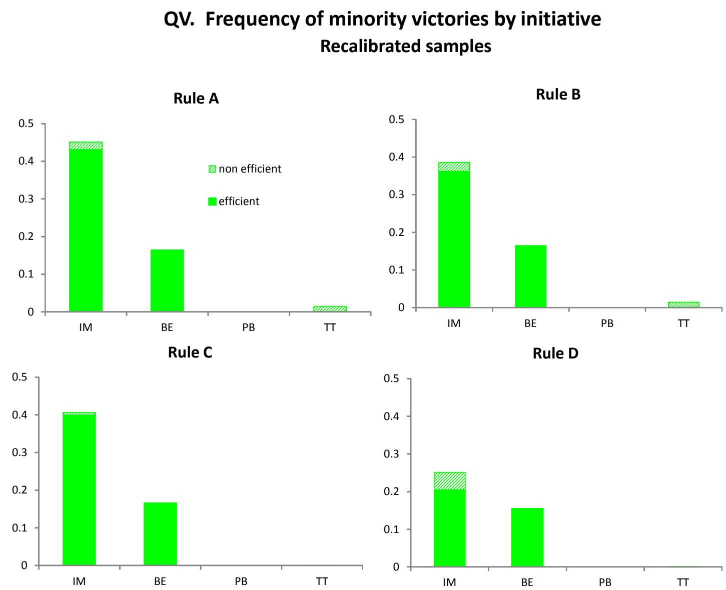
Minority victories by initiative. SV simulations. Frequency of minority victories obtained by SV in 10,000 simulations constructed by bootstrapping the SV sample, distinguishing between efficient (solid color) and inefficient (shaded color) ones. The efficient frequency of minority victories by initiative is: 0.52 (IM), 0.24 (BE), 0.05 (PB), and 0.04 (TT).



Minority victories by initiative. QV simulations. Frequency of minority victories obtained by QV in 10,000 simulations constructed by bootstrapping the QV sample, distinguishing between efficient (solid color) and inefficient (shaded color) ones. The efficient frequency of minority victories by initiative is: 0.18 (IM), 0.12 (BE), 0 (PB), and 0.002 (TT).



Minority victories by initiative. QVV simulations. Frequency of minority victories obtained by QVV in 10,000 simulations constructed by bootstrapping the QV sample, distinguishing between efficient (solid color) and inefficient (shaded color) ones. The efficient frequency of minority victories is reported in the caption of Fgure **??**.



Minority victories by initiative. QV simulations on recalibrated sample. Frequency of minority victories obtained by QV in 10,000 simulations constructed by bootstrapping the QV sample recalibrated on SV IM preferences, distinguishing between efficient (solid color) and inefficient (shaded color) ones. The efficient frequency of minority victories by initiative in the recalibrated sample is: 0.51 (IM), 0.16 (BE), 0 (PB), and 0.001 (TT). Note the difference in scale relative to the preceeding figures.

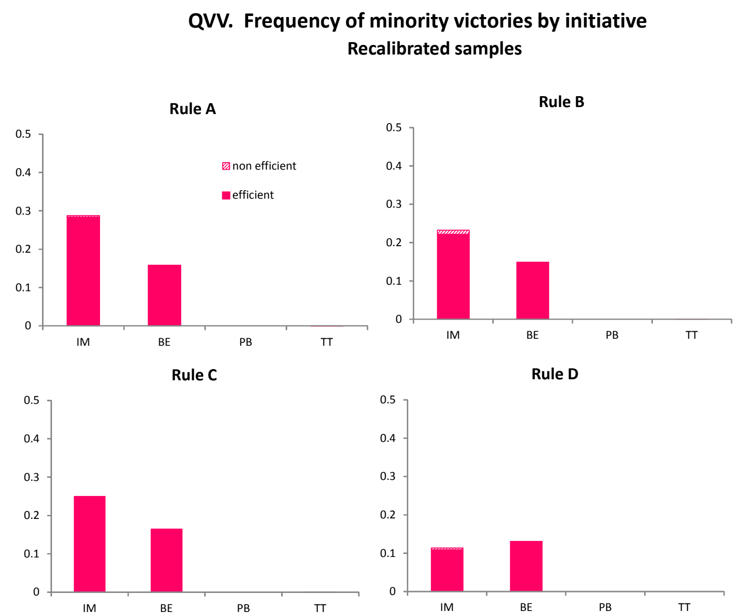
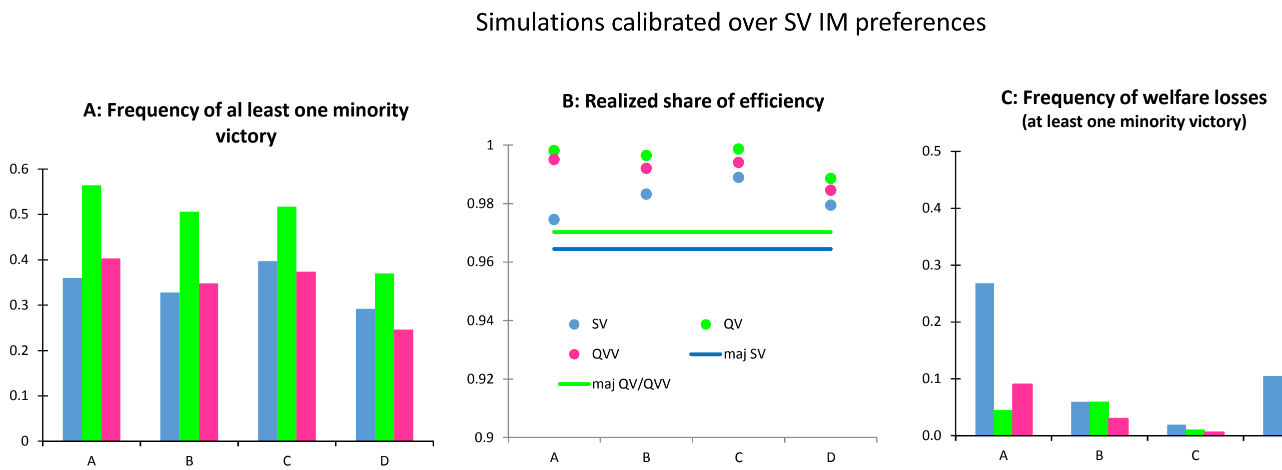


Figure 11: *Minority victories by initiative. QVV simulations on recalibrated sample.* Frequency of minority victories obtained by QVV in 10,000 simulations constructed by bootstrapping the QV sample recalibrated on SV IM preferences, distinguishing between efficient (solid color) and inefficient (shaded color) ones. The efficient frequency of minority victories by initiative is reported in the caption of Figure **??**. Again, note the difference in scale.

6.3.3 QVV and recalibrated samples. Comparisons to QV and to SV.



SV, QV and QVV comparisons; 10,000 simulations. The QV and QVV simulations are based on recalibrated samples.

6.4 Inequality

7 References for the online appendix

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

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

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

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

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

6. Lalley, S. and G. Weyl, 2018, "Nash Equilibria for Quadratic Voting", unpublished, Un. of Chicago.

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

8. Matsusaka, John G., 1992, "Economics of Direct Legislation", *Quarterly Journal of Economics*, 107, 541-571.

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