STORABLE VOTES AND QUADRATIC VOTING. AN EXPERIMENT ON FOUR
CALIFORNIA PROPOSITIONS

Alessandra Casella
Luis Sanchez

Storable Votes and Quadratic Voting. An Experiment on Four California Propositions
Alessandra Casella and Luis Sanchez
NBER Working Paper No. 25510
January 2019
JEL No. D02,D71,D72

ABSTRACT

Storable Votes and Quadratic Voting are voting systems designed to account for voters' intensity of preferences. We test their performance in two samples of California residents using data on four initiatives prepared for the 2016 California ballot. We bootstrap the original samples and generate two sets of 10,000 multi-elections simulations. As per design, both systems induce minority victories and result in higher expected welfare relative to majority voting. In our parametrization, quadratic voting induces more minority victories and achieves higher average welfare, but causes more frequent inefficient minority victories. The results are robust to different plausible rules-of-thumb in casting votes.

Alessandra Casella
Department of Economics
Columbia University
420 West 118 Street
New York, NY 10027
and NBER
ac186@columbia.edu

Luis Sanchez
Department of Economics
Cornell University
404 Uris Hall
Ithaca, NY 14853
las497@cornell.edu

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

Alessandra Casella and Luis Sanchez[*]

January 19, 2019

Abstract

Storable Votes and Quadratic Voting are voting systems designed to account for voters' intensity of preferences. We test their performance in two samples of California residents using data on four initiatives prepared for the 2016 California ballot. We bootstrap the original samples and generate two sets of 10,000 multi-elections simulations. As per design, both systems induce minority victories and result in higher expected welfare relative to majority voting. In our parametrization, quadratic voting induces more minority victories and achieves higher average welfare, but causes more frequent inefficient minority victories. The results are robust to different plausible rules-of-thumb in casting votes.

1 Introduction

Traditional political parties in Western democracies are in decline. From the *Gilets Jaunes* in France to *Cinque Stelle* in Italy, voters call for direct democracy and continuous electronic polling. If these calls are heard, we need to rethink how to protect the system from the *tyranny of the majority,* a classic problem in political design.[1] The protection of basic rights 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:

[1]See discussions in Dahl (1956, 1989), Guinier (1994), Issacharoff et al. (2002).

fairness and legitimacy are foremost, but straightforward utilitarian efficiency is an additional one, and is the object of this study. The numerical size of the two opposing groups does not account for varying intensities of preferences and even in binary decisions the outcome of majority voting can be inefficient.

An ideal solution does not superimpose additional rules–quota systems, preset rotation in positions of power–but amends the voting system to allow for the expression of preference intensity. Outcomes then reflect preferences endogenously, mirroring changes over time, and reverting to majority voting when intensities do not differ substantially. Two voting schemes with such properties have been proposed recently: storable votes (SV) (Casella, 2005, 2012, Hortala-Vallve, 2015), and quadratic voting (QV) (Posner and Weyl, 2015, Goeree and Zhang, 2017, Lalley and Weyl, 2018a). This paper analyzes their performance by eliciting preferences and voting behavior over four state-wide propositions in two samples of California respondents. We find that both voting schemes generate non-negligible minority victories and result in consistently higher efficiency than majority voting.

Both schemes induce voters to express the intensity of their preferences by allowing them to distribute a budget of votes across different binary decisions. In each decision, the side with more votes then prevails. In large elections, SV advocates granting voters a "regular" vote to cast on each decision, complemented by a single "bonus vote" to be spent as desired. QV grants voters a budget of "voices" to be distributed freely, with the defining feature 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 incentive is tempered: with SV by the existence of regular votes and the single bonus vote; with QV by the convex cost of bundling.

In our experiment, we polled California subjects registered on Amazon Mechanical Turk (MTurk) on their preferences and voting choices over four propositions that, at the time of polling, were being considered for the 2016 ballot. The four propositions 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 teachers' tenure. We collected data simultaneously from two samples of about 300 respondents each, exposing one to SV and one to QV.

Because the voting systems rely on the single budget constraint tying the

different voting decisions, the unit of analysis is the joint four-proposition election. Our data then correspond to two data points, one for storable votes, and one for quadratic voting, inadequate for any extrapolation. Ideally, we would 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 sets of 10,000 simulated multi-elections, one with SV and one with QV. This is the dataset on which we base our analysis.[2]

2   The Theory

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

Each individual's valuations $\{v_{i1}, .., v_{iK}\}$ are privately known. They are a random sample from a joint distribution $\mathcal{F}(v_1, \ldots, v_K)$ 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.

2.1   Storable votes

SV grants each voter a budget of "bonus votes" to be distributed freely over the different proposals. In large elections, it is typically optimal to cumulate all bonus votes on a single proposal, and the simplest practical implementation

---

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

3

is then to endow each voter with one bonus vote to spend as desired, complemented with K "regular" votes, one per proposal, that cannot be bundled.[3] The addition of the bonus vote, simple as it is, grants some weight to intensities of preferences: the voter's choice of where to cast it signals the voter's priorities.

If voters' valuations are independent across voters and proposals, 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 marginal distributions F_k.[4] 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 the minority holds more intense preferences– the case in which the desirability of majority voting is most disputed.

2.2 Quadratic voting

QV has been studied 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 (2018a) show that if valuations are independent across voters and the distribution F satisfies some plausible conditions, then any symmetric Bayesian Nash equilibrium converges asymptotically to 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.

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). 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[5] 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.

---

[3] Casella and Gelman (2008).

[4] Casella and Gelman (2008). We summarize the arguments in the appendix.

[5] The model was suggested by Glen Weyl and is reported in the appendix. It is similar but more transparent than the one in Lalley and Weyl (2018b).

3   The Experiment

In May 2016 we selected four propositions that were being prepared for inclusion in the November 2016 California ballot. We then recruited 647 California subjects via MTurk, and randomly assigned them either to the SV treatment (324 subjects; 306 after data cleaning[6]) or to the QV treatment (323 subjects; 313 after cleaning). We asked how each subject would vote on each of the four propositions, presented in random order, allowing for the option to abstain, and how important each of these propositions was to the subject. Respondents assigned to the SV (QV) treatment were then presented with a version of SV (QV) and asked how they would cast their votes. 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. The questions were programmed in Qualtrics, and copies of all screenshots are reproduced in the appendix.

3.1   The four propositions

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. In March 2016 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. Based on their responses, we selected for the experiment the following four propositions:[7]

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

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

---

[6] See the appendix for details.

[7] The March survey also yielded a small poll we reported at the end of the experiment. It had no effect and we ignore it in what follows.

the 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. We asked each subject to distribute 100 points among the four propositions, with the number of points a scale of the importance attributed to each proposal ("How important is this issue to you?"). We used examples to clarify that importance is unrelated to preferred direction and 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 preventing factors of scale from distorting welfare.[8]

Denoting by b_{ik} the number of points attributed to proposition k by individual i, we set $v_{ik} = b_{ik}$. Thus $U_i^S = \sum_{k:i\in M_k^S} b_{ik}$, 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. Realized aggregate welfare W^S is given by $W^S = \sum_i \sum_{k:i\in M_k^S} b_{ik}$. 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.

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 consider it our primary performance measure.

3.3 The voting schemes

The second part of the survey tests the new voting schemes. In the SV sample, subjects are told that each is granted a *bonus vote*–one extra vote–in addition

---

[8]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.

to the regular votes cast earlier. Each subject is asked to choose the proposition in which to use it; the vote is then cast automatically in the direction indicated 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.[9] Our survey presents the subjects with four different classes of votes, distinguished by color, and asks 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 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. 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. A subject who chooses green/yellow/red votes casts votes on only three/two/one propositions.

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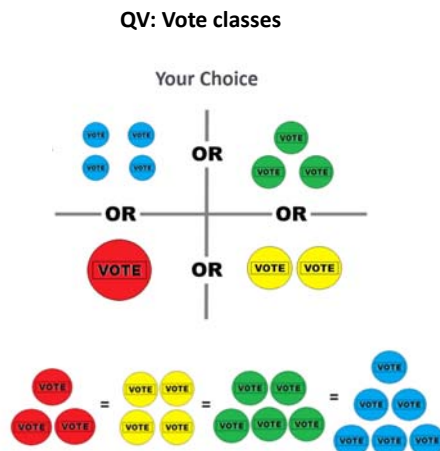
Figure 1: *The design of the QV scheme in the MTurk survey.*

The simple four-class classification respects the convex cost of concentrating votes. A voter using all four votes on different propositions–choosing blue votes–

---

[9]Quarfoot et al. (2017) employed QV to elicit opinions in a large electronic survey, but did so with proprietary software developed for the purpose, and after showing respondents a training video. We chose instead to simplify the QV scheme.

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 propositions, 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 propositions on which the subject casts votes, a constraint that is not part of the original idea. In the uncontrolled and fast MTurk environment, the hope is that our simplified QV benefits from its easier comprehension more than it loses in flexibility.

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.

3.4   The experimental data

3.4.1   Preferences and outcomes

In each sample, the survey yielded a distribution of preferences–the empirical counterpart to the theoretical distribution $\mathcal{F}$–and voting outcomes. We reproduce the histograms of respondents' preferences over each proposition in the appendix. Here we summarize preferences and voting choices by reporting percentage margins in favor of each proposition, in terms of aggregate points, number of voters, and number of votes (Figure 2).[10]

---

[10]Because each subject must allocate 100 points, within each sample the distributions of points across propositions are not independent, but with four initiatives no individual distribution is redundant.
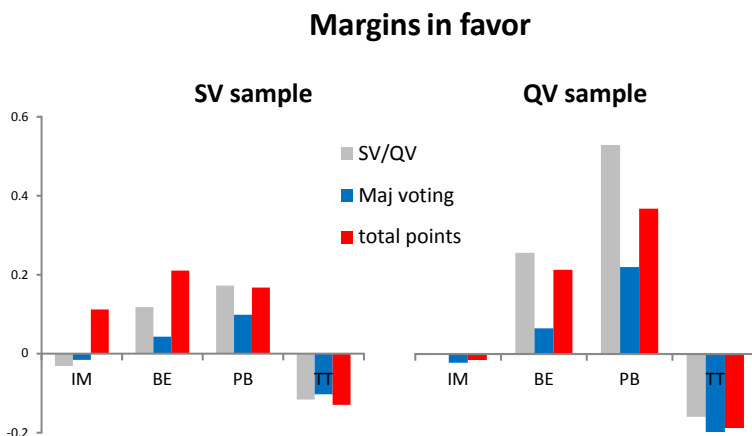
8

Margins in favor

Figure 2: *Margins in favor.* Two-sided KS tests assessing whether the distributions of points are drawn from the same population yield bootstrapped p-values equal to 0.629 (IM), 0.66 (BE), 0.092 (PB), 0.384 (TT).

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 aggregate point margin has the same sign as the majority voting outcome (all propositions in the QV sample; all but IM in the SV sample), the outcome is efficient. 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 (aggregating over all propositions, they realize 60.4% of full surplus).[11]

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.[12] It is also the most salient: as reported in the appendix, 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. In addition, the voting result shows that under SV in the IM proposition the bonus vote was not used

---

[11] The inefficient victory in the IM proposition is heavily penalized because in the SV sample all margins of victory are small, and thus welfare under random outcomes, the floor of our welfare measure, is relatively high.

[12] 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.

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.[13]

3.4.2   Voting choices: SV and QV

We report details on the voting choices in the appendix. Here we summarize them through two statistical models, to be read as compact representations of the data.

With SV, inspired by Casella et al. (2010), 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.

With QV, 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 $v_{(k)}$ the voter's kth highest value, 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 $v_{(4)}/v_{(3)} \geq \rho$, yellow if $v_{(4)}/v_{(3)} < \rho$ but $v_{(3)}/v_{(2)} \geq \gamma$, green if $v_{(4)}/v_{(3)} < \rho$, $v_{(3)}/v_{(2)} < \gamma$, but $v_{(2)}/v_{(1)} \geq \xi$, 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.

For both SV and QV, parameters are estimated by MLE, assuming independence across subjects. We reproduce them in Table 1.

---

[13]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 ($p < 0.01$)).

	SV				QV	
		95% CI				95% CI
p_{Max}	0.63	[0.56, 0.76]		ρ	1.33	[1.21, 2.10]
p_{Close}	0.04	[0, 0.14]		γ	1.19	[1.04, 1.21]
p_{Fam}	0.09	[0.04, 0.21]		ξ	1.39	[1.18, 1.79]
p_{Rand}	0.23	[0.15, 0.37]		ε	0.50	[0.44, 0.54]
				μ	0.21	[0.17, 0.26]

Table 1. *Voting choices: two simple statistical models.* MLE estimates of the statistical models.[14] The confidence intervals are obtained by bootstrapping and reflect the distribution of the estimated parameters in 10,000 simulations.

In the SV sample, the bonus vote is cast primarily on the proposition that the voter considers her highest priority, but a relatively large role is left for randomness. In the QV sample, the thresholds in relative values determining the choice of vote class are significantly higher than 1 but quantitatively not far, indicating a strong tendency towards cumulating votes.[15] 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.

4 10,000 Multi-election Samples

On all four propositions, both SV and QV confirmed the outcome reached with simple majority voting. The observation, 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. 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 we can approximate such iterations by bootstrapping our data.[16]

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 elections for a single individual. We sample with replacement N individuals from

---

[14] Under the constraints that all probabilities be non-negative and smaller than 1, and all threshold parameters be larger or equal to 1. In the SV model, the estimated probabilities must sum to 1.

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

[16] See the discussion in Casella et al. (2010). For a discussion of bootstrapping, see Davidson and MacKinnon (2006).

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.[17] 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 the SV or QV voting choices they reported is an obvious option, and the first one we consider. But an additional goal is to evaluate the robustness of the two voting schemes to a range of plausible behaviors, and with this in mind, we posit four alternative rules-of-thumb governing the use of the votes: A, as just mentioned, i.e., as the individual did in the original sample; B, as in the descriptive models of the original samples described above; C, as optimal in a simplified environment; D, introducing randomness in rule C. If the MTurk sample is not fully representative, the welfare implications of the voting choices reported by our respondents are difficult to evaluate because they presumably reflect beliefs about the full electorate. By abstracting from reported choices, rules C and D allow a comparative evaluation of SV and QV that is not affected by the possible lack of representativeness of the MTurk sample.

In the simulations, we implement rule B by having each drawn subject behave according to the corresponding estimates in Table 1, given the subject's allocation of points and voting system. Rule C supposes 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 $1/2$.[18]

---

[17] We verified that with 10,000 simulations the sampling error is negligible. Doubling the number of simulations to 20,000 has barely detectable effects.

[18] See the appendix for details.

Simulation Results

With all four rules, both SV and QV resulted in frequent minority victories (Figure 3: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.[19] 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 86% for SV and 98% for QV, compared to 72% and 95% for majority in the two sets of simulations (Figure 3: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 3:C).
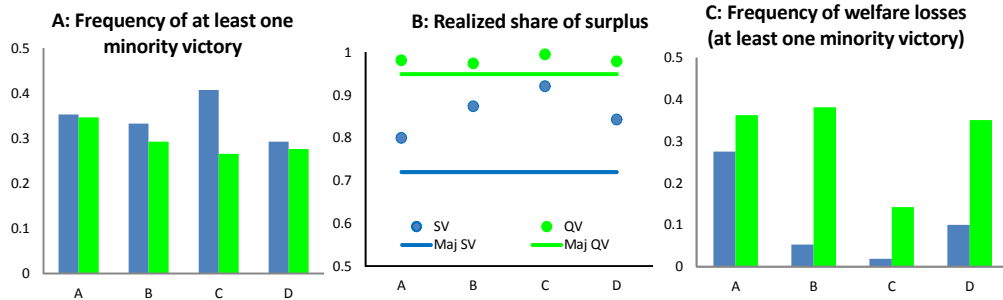

Figure 3: *Bootstrap results*. In panel B, note the difference in majority welfare in the SV and QV samples.

Reporting the realized share of surplus over all simulations, whether or not any outcome differs from simple majority, (as in Figure 3:B) allows us to give weight to foregone efficiency gains. But only a fraction of simulations include a minority victory, and none of the differences in means in panel B are statistically significant. If we consider only simulations with at least one minority victory, the probability of a positive increase in welfare is the complement to 1 of the numbers shown in Figure 3:C. Such probability is significant at conventional levels for SV-B ($p = 0.047$), SV-C ($p = 0.019$), and SV-D ($p = 0.10$), but in no

---

[19] Most minority victories were concentrated on the IM and BE initiatives. See the appendix for details.

other case.

In the SV simulations, the outlier is rule A, which implements the actual voting choice indicated by the subject drawn in the simulation (Figure 3:C). The problem comes from the IM proposition, where, as noted earlier, bonus votes are predominantly cast against the proposition, while high points are predominantly attributed by subjects in favor. The asymmetry in behavior concerns a small number of subjects, and we cannot rule out that it reflects pure noise[20], but is sufficient to affect the results. Rules B, C, and D implement voting behavior that is symmetric relative to the direction of preferences, and inefficient minority victories are fewer.

The difference in performance between SV and QV is largely driven by the different potential for improvement over majority voting: such potential is smaller under QV, and as a result errors are more likely and more frequent. The discrepancy in the performance of majority voting across the two samples is surprising because the two samples were populated randomly during the MTurk survey. Again, it is due primarily to the IM proposition, where, as we saw, in the MTurk data efficiency favors the majority outcome in the QV sample but not in the SV sample. Over the 10,000 simulations, considering the IM proposition alone, majority captures 24% of the surplus in the SV sample, versus 85% in the QV sample. The difference between the two original samples is random, but persists in the simulations, which draw from the original samples. We can attempt to correct the problem by recalibrating the bootstrap exercise.

4.1   Comparing SV and QV: Recalibrating the samples

The data are not sufficiently rich to restrict simulations to subjects with similar preferences over all propositions. We must limit ourselves to simulate electorates with more similar distributions of preferences over the proposition that appears most problematic. More precisely, we construct two sets of 10,000 recalibrated samples by restricting the distributions of preferences over the IM proposition so as to replicate, on average, the distribution of preferences over IM in the total population of our original data (that is, over the two MTurk samples combined): we draw subjects, with replacement, from the IM preference histogram of the relevant dataset–SV to construct the SV samples, and QV for the QV samples– but constrain each draw to come from any bin of the histogram with probability equal to the corresponding frequency in the population IM preference histogram.

---

[20]See the appendix.

The resulting SV and QV samples are more closely comparable: over 10,000 simulations, the average share of surplus appropriated by majority voting in the IM proposition is now 49% for SV and 54% for QV–a difference of 5%, as opposed to more than 60% before recalibration. Averaging over all propositions, majority voting now appropriates 84% of the surplus on average in the SV simulations (vs. 72% in the non-recalibrated data), and 91% in the QV simulations (vs. 95%).[21]

We have verified that the recalibration does not distort the distributions of preferences over the other propositions, creating new disparities across SV and QV data: majority voting's average performance becomes more similar across both datasets in three of the four propositions; in the fourth (PB), majority appropriates on average more than 90% of the available surplus in both datasets, even as the share declines slightly with recalibration in the SV samples.

Three regularities emerge clearly from the recalibrated simulations (Figure 4). First, QV now results in a consistently higher fraction of minority victories than SV (averaging over all rules, 38% of QV simulations have at least one minority victory, vs. 29% per SV).[22] Second, under any rule, both voting systems continue to appropriate a higher share of surplus than majority does: QV captures 98% of surplus on average (vs. 91% with majority), and SV 91% percent (vs. 84% with majority). Third, as expected, the frequency of welfare losses mirrors the available room for improvement over simple majority. In the recalibrated simulations, majority becomes more efficient in the SV sample, and the frequency of inefficient minority victories increases (averaging over all rules, 22% of all simulations with at least one minority victory induce welfare losses, up from 11% in the original simulations); majority becomes less efficient in the QV sample, and the frequency of errors declines (down to 17%, from 30% originally). Conditional on at least one minority victory, the probability of welfare improvement over majority is significant at the 10% level for SV-C ($p = 0.08$) and for QV-C ($p = 0.06$).

---

[21] In the population, the majority of voters is against the IM proposition, but aggregate points are higher among voters in favor: majority voting is inefficient.

[22] Under any rule, the difference is statistically significant at the 1% level.
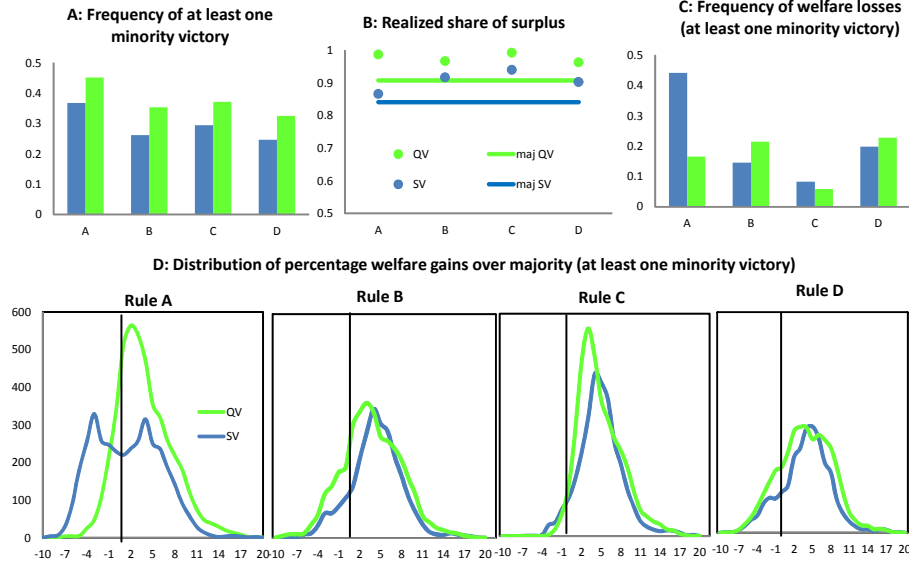
Figure 4: *Bootstrap results with recalibrated samples.* In the D panels, the scale is constant. The black vertical line separates losses and gains.

SV-A remains an outlier, and again the reason is the impact of the bonus vote on the IM proposition. The recalibration leads to samples where defeats of the proposition are more frequent, whether under majority voting, or under SV-A, or under efficiency. Overall, the frequency with which SV-A leads to efficient outcomes over the IM proposition increases. However, increasing the frequency of voters who oppose the proposition also increases the frequency of voters who, under SV-A, cast their bonus vote against IM at relatively low intensities, and thus the frequency of inefficient minority victories. As before, the problem arises from the asymmetry of behavior in the original SV sample between voters favoring or opposing the IM proposition. Rules B, C, and D, on the other hand, perform noticeably better, and under these rules SV yields results comparable to QV (Figure 4, panels B, C, and D).

The QV performance is remarkably consistent across rules, and appropriates, on average, close to the totality of the available surplus. QV appears more sensitive to potential improvements than SV, inducing more minority victories than SV even though majority performs better in the QV samples. A natural

question is the extent to which this sensitivity reflects the specific parametrization we have implemented. For SV we have followed the theory in Casella and Gelman (2008), where background "regular" votes dull 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. With multiple decisions, however, and the maintained assumption of constant marginal pivot probability, neither a voter's optimal choice nor the efficiency properties of the voting system would change when adding regular votes.[23]

We report in the appendix how the QV simulations results are affected by the addition of regular votes. As expected, regular votes strengthen the majority side and reduce the frequency of minority victories. Because this is particularly true when the total intensities on the two sides are comparable, the result is a strong decline in inefficient minority victories, with little if any impact on average extracted surplus. With our data, the modification is to be recommended.

Conclusions

We report the results of a Mechanical Turk experiment designed to test the performance of Storable Votes (SV) and Quadratic Voting (QV), two voting systems designed to incentivize the expression of voters' intensities of preferences and allow for minority victories. We elicited subjects' preferences and voting choices over four propositions intended for the 2016 California ballot. By bootstrapping our data, we generated two samples of 10,000 joint elections (i.e. 10,000 instances of voting on all four propositions) on which the performances of the two voting systems can be compared, and compared to majority voting.

We find that both systems indeed result in non-negligible minority victories: depending on different details of the bootstrapping procedure, between 30 and 40% of simulations in both samples have at least one proposition won by the minority side. And while making minority victories possible, both systems consistently raise our measure of total welfare, appropriating at least 80 and as much as 99% of the available surplus. In the main parametrizations we implement, QV is more sensitive to minority preferences than SV, inducing more minority victories and appropriating a larger share of the surplus. When the scope for improvement over majority voting is small, however, such sensitivity can have a cost and induce frequent inefficient minority victories. SV is simpler,

---

[23]The model sketched in the appendix accommodates background votes with no change in behavior.

a strong argument in its favor. However, its performance is less consistent across plausible subjects' behaviors.

Minority victories can be desirable not only for their direct impact on welfare, but also because, if preferences are correlated across decisions, the same part of the electorate can be on the losing side on multiple decisions and come to feel disenfranchised, questioning the legitimacy of the voting system. An additional question, important in its own right, is the impact of SV and QV on the inequality of realized ex post utilities. We show in the appendix that in our data both SV and QV not only increase welfare but also reduce ex post inequality. Under both aggregate welfare and inequality criteria, we find their performance to be desirable.

There is widespread demand for more reliance on direct democracy, together with justified concerns about possible misuses and abuses. It is important to understand what the weaknesses of direct democracy are, and to think of better, more robust designs. Both SV and QV contribute to this discussion and are sufficiently well understood to be ready for testing. This paper is one step in that direction.

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

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

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

13. 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.

14. 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.

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

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

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

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

19. 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.

5 Online Appendix

5.1 Theory

5.1.1 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 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. 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 proposition. The theoretical analysis in CG 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.

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 of values over each proposition. CG show that the result holds in the following environments. (1) If $F_k(v) = F(v)$ for all k, where $F(v)$ 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 $F_{k''}(v) \neq F_{k'}(v)$ for $k'' \neq k'$, but $F_k(v)$ is symmetric around a known median of value 0 for all $k \in \{1,..,K\}$. (3) If $F_k(v) = G(v)$ for all k, where $G(v)$ is symmetric around a random median with expected value at 0.

With independent voters and large N, assumptions about the shape of the distributions $F_k(v)$ have immediate implications about the results of the referenda. In particular, assuming specific medians for the distributions $F_k(v)$ 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.[24] We report here in more detail the results that refer to that case.

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

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

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

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

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.[25]

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

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

---

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

[25] 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:

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

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.

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.

5.1.2 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, 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. 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.[26]

The model relies on the use of a 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 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

---

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

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 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

24

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.

5.2 Experimental data

5.2.1 Cleaning procedures

In designing the survey, we added an attention check to both samples. The check took the form of a fictitious fifth 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.

5.2.2 Subjects' preferences

Subjects' preferences over each proposition are summarized in histograms reporting the number of respondents assigning to a proposition different numbers of points. Points are coded as negative when the subject voted against the proposal, and as positive when the subject voted in favor, with bins of size 10 (0, colored light blue in the figures, corresponds to abstentions). Figure 5 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).

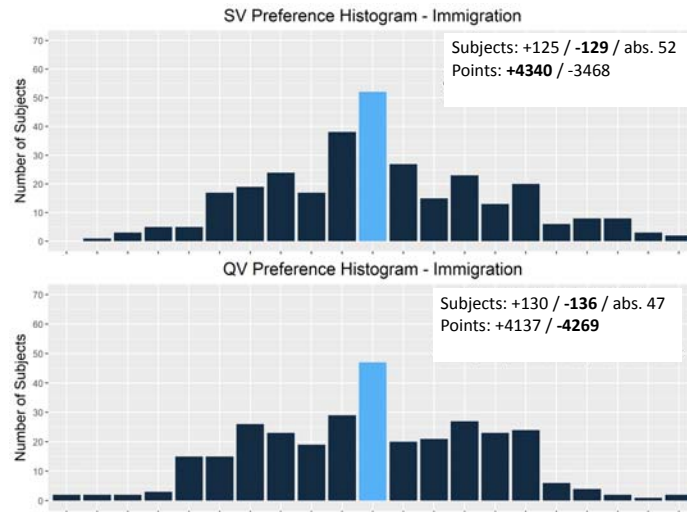Figures 6, 7, and 8 report the histograms for the other propositions.

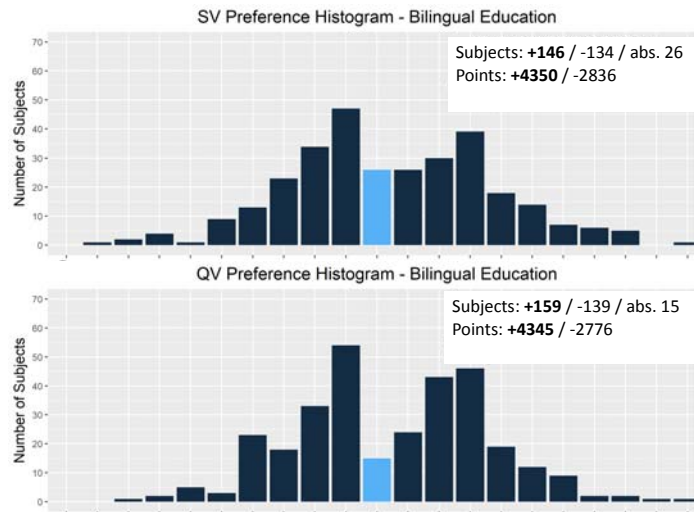Figure 5: *Distribution of preferences: the IM proposition.*

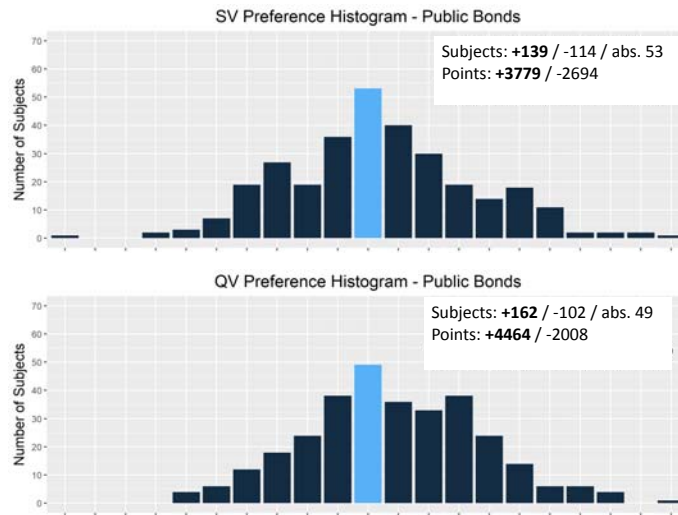Figure 6: *Distribution of preferences: the BE proposition.*

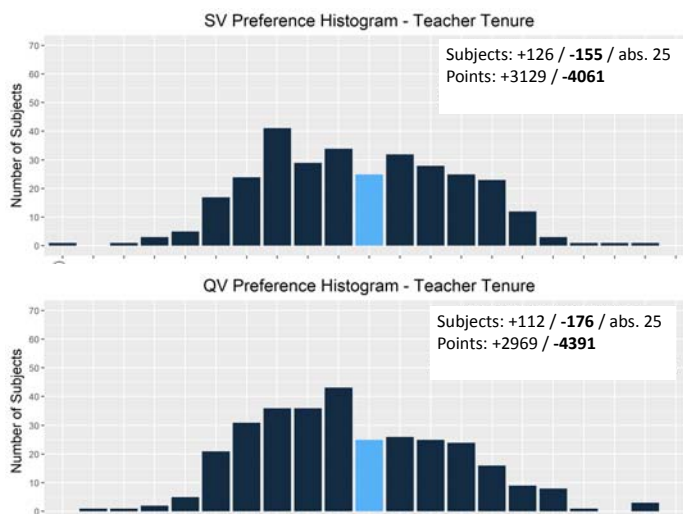Figure 7: *Distribution of preferences: the PB proposition.*

SV Preference Histogram - Teacher Tenure

Subjects: +126 / **-155** / abs. 25
Points: +3129 / **-4061**

QV Preference Histogram - Teacher Tenure

Subjects: +112 / **-176** / abs. 25
Points: +2969 / **-4391**

Figure 8: *Distribution of preferences: the TT proposition.*

5.2.3 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, 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 approximation is analogous to the posited behavior under QV. Figure 9 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.[27]

---

[27] 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.
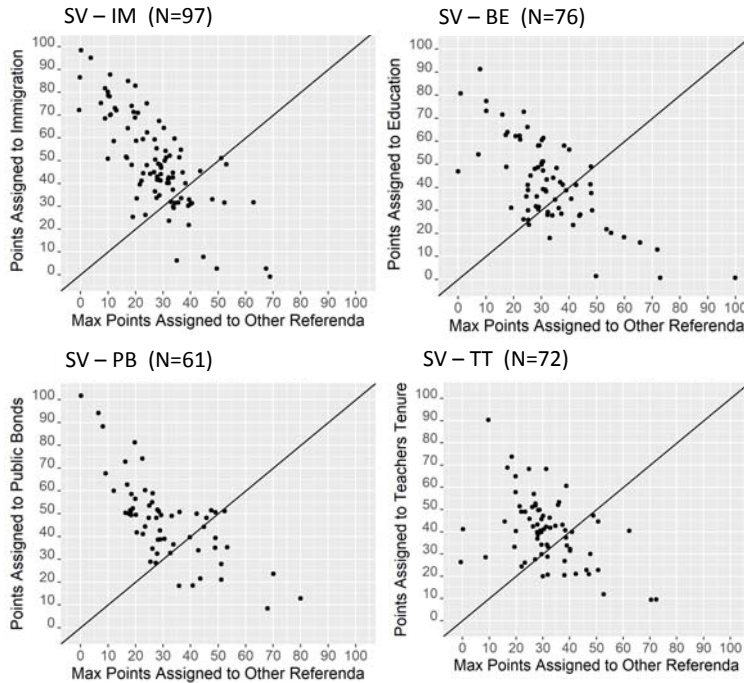
Figure 9: *The Bonus Vote Decision.*

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

The salience of the IM proposition is supported by the high number of bonus votes (97, vs. 76 for BE, 61 for PB and 72 for TT) . As noted in the text, when accounting for bonus votes the margin of victory for opponents of the proposition increases, although IM supporters report higher average and total intensity. The result reflects two sources of asymmetry. Of the 23 subjects who identify the IM proposition as their first priority and yet do not target it with their bonus vote, more than twice (16) are supporters rather than opponents (7); of the 20 subjects who cast their bonus vote on IM and yet do not identify it as their priority, more than twice (14) are opponents rather than supporters (6). However these differences are numerically very small and in neither case are the differences in proportions statistically significant.
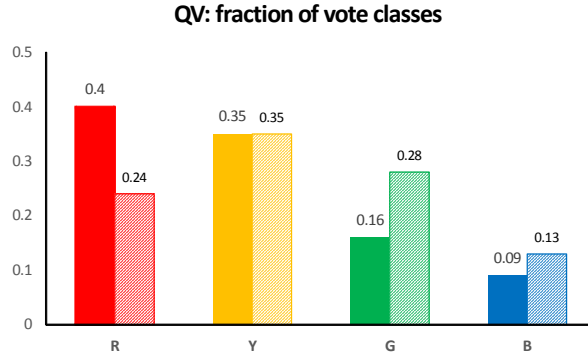
29

Figure 10: *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).

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 10 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. As described in Table 2 below, we calculated the shares of the different vote classes subjects would have chosen had they followed the theoretical model sketched in Section 5.1.2. The corresponding values, with lower vote cumulation, are in the shaded columns. The theory suggests a lower frequency of red votes, and a higher frequency of green and blue votes.

Across propositions, IM received the highest total number of votes (97 vs. 76 for BE, 61 for PB and 72 for TT), as well as the highest number of red votes (45 vs. 19 for BE, 26 for PB and 34 for TT). confirming its salience.

5.3 Bootstrap Simulations

5.3.1 QV: rules C and D

Under rule C, voters act optimally under the belief that the marginal impact of their votes is constant. As shown in section 5.1.2, 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 $\{2v_{i(4)}/\beta_i, 2v_{i(3)}/\beta_i, 2v_{i(2)}/\beta_i, 2v_{i(1)}/\beta_i\}$.[28] Given a vote class, votes should be assigned monotonically.

We reestimated the QV statistical model reported in Table 1 after having imposed the normative choice on the MTurk sample. As shown in Table 2 below, estimated thresholds ρ and γ are significantly higher than in Table 1, supporting the hypothesis that MTurk respondents concentrated votes excessively.

		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 2. *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.

5.3.2 Rule E

To evaluate the welfare costs of misusing the voting systems, we have simulated, for both SV and QV, a fifth rule, E, that captures random behavior. Following rule E, voters vote in the direction of their preferences but choose randomly

---

[28]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 $\widetilde{v}_{ik}/\sqrt{\sum_{k=1}^{K}(\widetilde{v}_{ik})^2} = v_{ik}/\sqrt{\sum_{k=1}^{K}(v_{ik})^2}$.

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.)

5.3.3 Results: 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 5, 6, 7, and 8 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 11-14 below report the results.
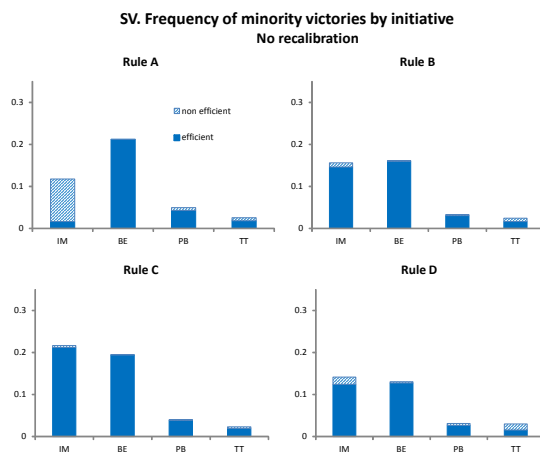

Figure 11: *Minority victories by initiative. SV simulations.*The efficient frequency of minority victories by initiative is: 0.52 (IM), 0.24 (BE), 0.05 (PB), and 0.04 (TT).
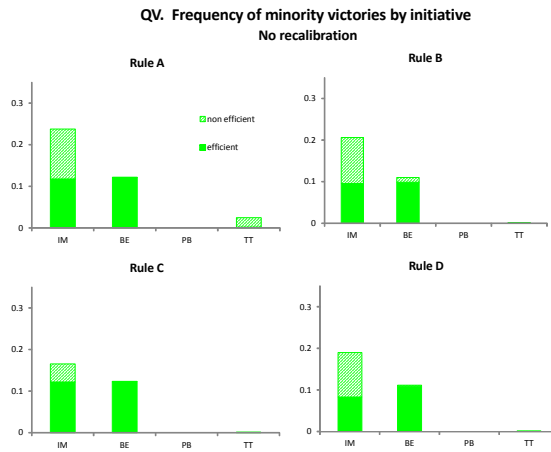
Figure 12: *Minority victories by initiative. QV simulations: bootstrapping the original QV sample.* The efficient frequency of minority victories by initiative is: 0.18 (IM), 0.12 (BE), 0 (PB), and 0.002 (TT).
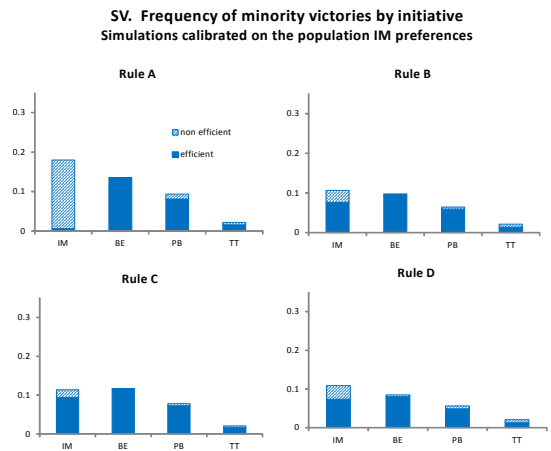

Figure 13: *Minority victories by initiative. SV simulations calibrated on the population IM preferences.* The efficient frequency of minority victories by initiative is: 0.37 (IM), 0.16 (BE), 0.09 (PB), and 0.03 (TT).

QV. Frequency of minority victories by initiative
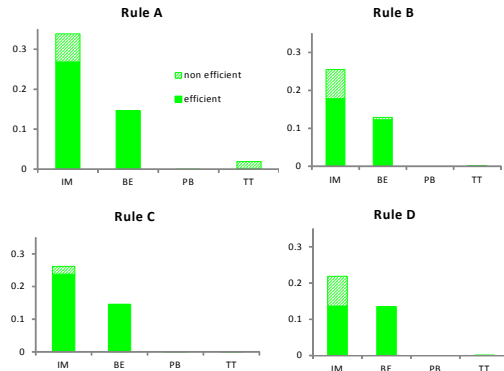**Simulations calibrated on the population IM preferences**

Figure 14: *Minority victories by initiative. QV simulations calibrated on the population IM preferences.* The efficient frequency of minority victories by initiative is: 0.34 (IM), 0.14 (BE), 0 (PB), and 0.002 (TT).

5.4 QV cum regular votes

We report here the simulations results when we add to QV four regular votes, to be cast one on each proposition. Because the MTurk survey described QV as consisting exclusively of the classes of votes chosen by the respondents, applying the subjects' responses to a different environment is not fully legitimate. We proceed with this caveat in mind, but noting again that if marginal pivot probability is perceived to be constant, behavior should not change. We suppose that each subject casts the regular votes 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.*

In both sets of simulations, whether bootstrapping the original raw data or recalibrating the samples to reflect the population preferences over the IM proposition, QVV behaves as predicted: relative to QV, the frequency of minority victories declines; so does the frequency of welfare losses, and the share of appropriated surplus barely changes.

In the QV simulations based on the original samples, without recalibration, majority voting appropriates a full 95% of the available surplus. As noted in the text, although on average QV improves over majority under each of the four rules-of-thumb, 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%, with a corresponding decline (from 31 to 18%) in the frequency of welfare losses, in the simulations with at least one minority victory. (All differences in frequencies, for any rule, are significant at the 1 percent level). The impact on the average realized share of surplus is positive, if very small and not statistically significant (averaging over all simulations and rules, an improvement of one percentage point, barely noticeable in Figure 15).

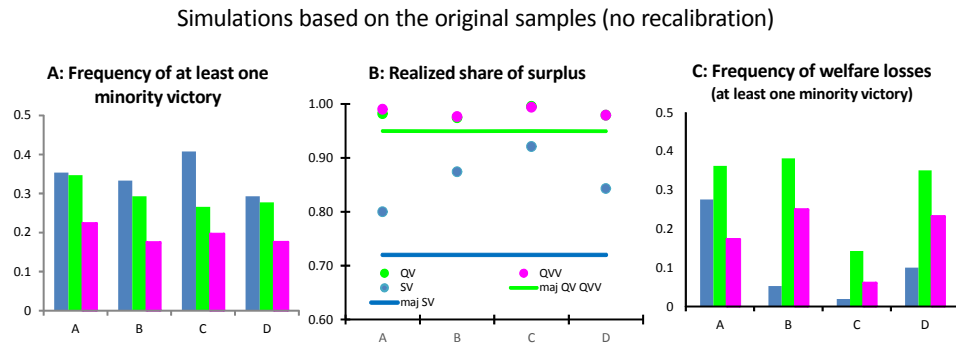Simulations based on the original samples (no recalibration)

Figure 15: *QV, SV, and QVV. Simulations based on the original QV sample.*

Figure 16 depicts the results for the recalibrated simulations. Relative to QV, QVV forfeits minority victories with small welfare changes, whether in terms of gains or losses. On the whole, it is quite comparable to SV–but for SV's weak performance under rule A.
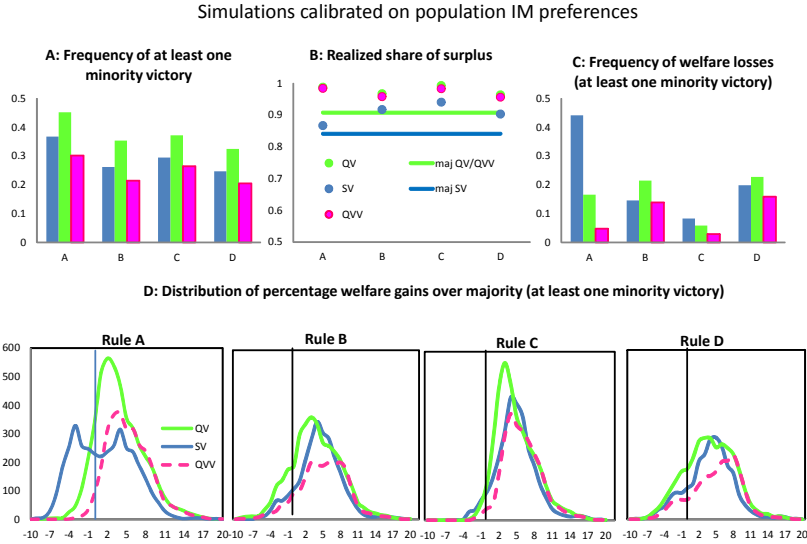
Figure 16: *QV, SV and QVV. Simulations calibrated on the population IM preferences.*

5.5 Inequality

Majority voting becomes problematic when the composition of majority and minority does not change across decisions. An important question is whether voting systems that make minority victories possible also reduce the inequality in voters' probability of achieving their preferred outcomes. In our data, the frequency with which voters' preferences are satisfied is reflected in their ex post utility: by construction, a voter who loses all propositions has an ex post utility of 0; and one who wins them all of 100. Thus, under majority voting (and supposing disagreement), correlation in individual voters' directions of preferences across proposals would result into high inequality in ex post utility, reflecting the disenfranchisement of voters who find themselves on the minority side disproportionally often.

To evaluate SV and QV's impact on inequality, we calculate the Gini coefficient of the realized utility distribution for each of our simulations, and under each of the four rules-of-thumb, under the relevant voting system and simple majority. Focusing on simulations with at least one minority victory and averaging across the four rules, the frequency of Gini declines, relative to majority

voting, is 73% for SV and 69% for QV when bootstrapping the original samples, and 65% for SV and 84% for QV in the simulations recalibrated on the population IM preferences. The average Gini decline is 10% for SV and 8% for QV when the simulations are not recalibrated, and 6% and 12% respectively, when they are. Figures 17 and 18 report the results.
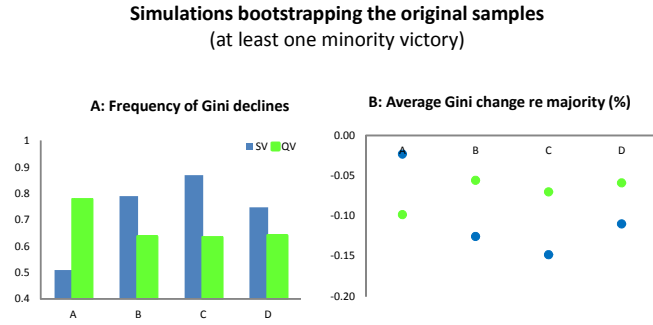

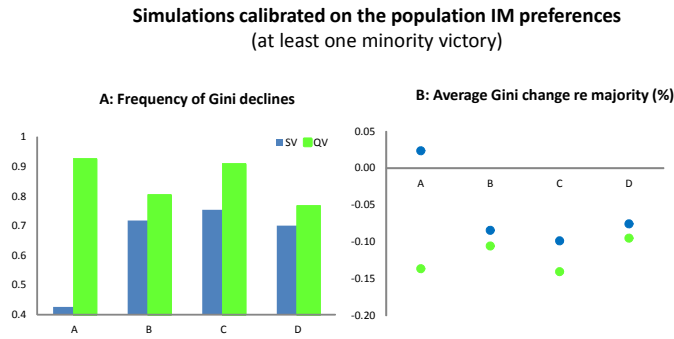Figure 17: *Gini coefficients. Simulations based on the original samples.*

Figure 18: *Gini coefficients. Simulations recalibrated on the population IM preferences.*

The average Gini coefficients are reported in Tables 3 and 4. (The results for QVV, not reproduced here, are effectively identical to QV).

37

SV			QV		
		95% CI			95% CI
maj	0.37	[0.30, 0.43]	*maj*	0.325	[0.27, 0.37]
A	0.35	[0.28, 0.40]	*A*	0.29	[0.26, 0.345]
B	0.32	[0.28, 0.40]	*B*	0.30	[0.26, 0.35]
C	0.315	[0.28, 0.39]	*C*	0.30	[0.26, 0.35]
D	0.33	[0.28, 0.40]	*D*	0.30	[0.26, 0.35]

Table 3. *Gini coefficient: Simulations based on the original samples (at least one minority victory).* The confidence intervals are obtained by bootstrapping.

SV			QV		
		95% CI			95% CI
maj	0.36	[0.30, 0.43]	*maj*	0.33	[0.28, 0.37]
A	0.36	[0.29, 0.40]	*A*	0.28	[0.26, 0.34]
B	0.33	[0.28, 0.39]	*B*	0.29	[0.26, 0.35]
C	0.33	[0.28, 0.39]	*C*	0.28	[0.26, 0.34]
D	0.34	[0.28, 0.40]	*D*	0.30	[0.26, 0.35]

Table 4. *Gini coefficient: Simulations recalibrated on the population IM preferences (at least one minority victory).* The confidence intervals are obtained by bootstrapping.

The summary message is that both SV and QV have a positive impact on ex post inequality. With the exception of SV-A, the probability of a decline in Gini is statistically significant at any level, for either system and with any rule in both sets of simulations. The magnitude of the effect, however, is not statistically significant. Both SV and QV can be compared to majority, but comparing inequality across the two voting system is more problematic because the two samples differ. All Gini indices are higher in the SV sample, whether under SV or majority, reflecting the smaller margins of victory in that sample.

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