Public Choice (2017) 172:283–303
DOI 10.1007/s11127-017-0416-1

Quadratic voting in the wild: real people, real votes

David Quarfoot[1] · Douglas von Kohorn[2] · Kevin Slavin[3] · Rory Sutherland[2] · David Goldstein[2] · Ellen Konar[4]

Received: 31 August 2016 / Accepted: 30 September 2016 / Published online: 14 February 2017

Abstract Since their introduction in 1932, Likert and other continuous, independent rating scales have become the de facto toolset for survey research. Scholars have raised significant reliability and validity problems with these types of scales, and alternative methods for capturing perceptions and preferences have gained traction within specific domains. In this paper, we evaluate a new, broadly applicable approach to opinion measurement based on quadratic voting (QV), a method in which respondents express preferences by 'buying' votes for options using a fixed budget from which they pay quadratic prices for votes. Comparable QV-based and Likert-based survey instruments designed by Collective Decision Engines LLC were evaluated experimentally by assigning potential respondents randomly to one or the other method. Using a host of metrics, including respondent engagement and process-based metrics, we provide some initial evidence that the QV-based instrument provides a clearer measure of the preferences of the most intensely motivated respondents than the

✉ Douglas von Kohorn
dougvk@aya.yale.edu

David Quarfoot
dquarfoot@ucsd.edu

Kevin Slavin
slavin@media.mit.edu

Rory Sutherland
rory.sutherland@ogilvy.com

David Goldstein
david@collectivedecisionengines.com

Ellen Konar
ellen.konar@gmail.com

[1] University of California, San Diego, San Diego, CA, USA

[2] Collective Decision Engines, LLC, Jackson Heights, NY, USA

[3] Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

[4] Stanford Center on Longevity, Stanford, CA, USA

Likert-based instrument does. We consider the implications for survey satisficing, a key threat to the continued value of survey research, and discuss the mechanisms by which QV differentiates itself from Likert-based scales, thus establishing QV as a promising alternative survey tool for political and commercial research. We also explore key design issues within QV-based surveys to extend these promising results.

Keywords Social choice · Collective decisions · Survey methods · Intensity of preference · Preference elicitation · Budgeted voting

JEL Classification C42 · C93 · D71 · D78

1 Introduction

National surveys long have played unique and critical roles in reflecting public opinion to inform the workings of democratic processes (Krosnick et al. 2009). Surveys also have served important commercial functions, improving the odds that companies understand consumers or potential business customers and, hence, invest in the most desirable new products and product features. Regardless of domain, the value of surveys depends on their accuracy.

Remarkably, survey and questionnaire design has changed little in the past 50 years. The Likert scale, the nearly universal method of attitude measurement, was developed initially and tested by Rensis Likert in the 1930s. In his 1932 dissertation, Likert demonstrated a simple and versatile approach to measuring opinions on subjects as diverse as evolution, war, birth control and even the existence of God. His standard multi-point, unidimensional scale—on which respondents indicated their opinions on a set of positions on a scale from 'strongly disagree' to 'strongly agree' or 'strongly disapprove' to 'strongly approve'—became and still is the most common tool of survey research in the twenty-first century.

Studies of the Likert approach have established that it satisfies many criteria of reliability and validity. Attitudes on policy issues, as thereby measured, especially those that are salient, are stable over time (Krosnick 1988) and resistant to change (Gorn 1975). The more importance people attach to a policy preference, the better their survey responses predict their choices in the voting booth (e.g., Anand and Krosnick 2003; Fournier et al. 2003; Visser et al. 2003).

Despite these satisfying results, Likert-based approaches to opinion measurement do suffer from significant limitations. First, as the research above suggests, reliability and validity appear to be limited to those issues of significant importance. Yet, this limitation rarely is accounted for in practice, and respondents often are asked to indicate opinions on issues well beyond those of personal importance to them or on which their knowledge is adequate. Satisficing, or the decision-making strategy in which a respondent reviews available alternatives until an acceptability threshold is met, is now an all-too-common response mode for survey respondents (Krosnick et al. 2015). Krosnick et al. (2009) touch on a more fundamental issue of relative weight: "It is especially difficult to predict the importance a person attaches to one issue without knowing how much importance that person attaches to another" (Anand and Krosnick 2003; Krosnick 1988). The two approaches routinely applied in garnering relative assessments of options, rank-order questions and conjoint studies, have raised the burden on survey designers and survey takers without compelling evidence that such methods generate better (i.e., more reliable) results.

In this paper, we evaluate a new approach based on quadratic voting (QV). In the original formulation due to Weyl (2012) (see also Lalley and Weyl 2016), QV is used for collective decision making by a group of people. Each person has a budget of money (real or artificial), which he/she uses to "buy" votes for or against a proposal, paying a quadratic price for the vote. For example, a city may propose building bicycle lanes. A resident who favors the proposal could buy one vote (for $1), two votes (for $4), three votes (for $9), and so on. Similarly, a resident who opposes the proposal could buy one or more votes against it. The proposal passes if a voting majority favors it. The money either is returned to residents pro rata or used to fund the project. This system gives residents strong incentives truthfully to reveal their preferences, while ensuring (with increasing and very high degree of probability as the number of residents increases) that the project maximizes the public's welfare. At its heart, QV forces the marginal cost of voting to be proportional to the number of votes purchased, and as a consequence, creates an efficient space in which voting is proportional to value.

To date, the work on quadratic voting has not focused on its application as a tool in market research. Theoretically, the applicability of QV in a market research context is strong given the proofs of its value in other collective choice contexts. The major concern in market research is understanding the opinions of various groups of actual or potential consumers (e.g., on features of products/services in development or brand attributes) and QV's applications in other settings, such as voting on political candidates or policy options. We can apply the theory of QV to market research by giving respondents a budget of artificial currency and asking them to use it to buy votes for or against alternatives of interest to marketers. If the theory of QV is valid, then respondents should be given strong incentives to reveal their preferences truthfully.

It is critical to emphasize a key difference between the Likert and QV approaches. At its heart, Likert is a *methodology of abundance*—respondents express whatever opinions they wish to express, and no respondent's opinion affects anyone else's. As in the case of abundant resources (operating on the economist's extensive margin), careful tradeoffs are not necessary because no incentive to economize exists: enough "and as good", in John Locke's words, remains for others to consume. In contrast, QV is a *methodology of scarcity* (it operates on the intensive margin)—voters are constrained by their credit banks, and this forces voters to make tradeoffs across proposals. This difference is crucial, for as Mullainathan and Shafir (2013, p. 7) write in their book on scarcity: "[W]hen we experience scarcity of any kind, we become absorbed by it. The mind orients automatically, powerfully, toward unfulfilled needs… It changes how we think. It imposes itself on our mind". Moreover, "Scarcity is not just a physical constraint. It is also a mindset… By staying top of mind, it affects what we notice, how we weigh our choices, how we deliberate, and ultimately what we decide and how we behave" (Ibid., p. 12). In this paper, we investigate whether QV successfully creates a scarcity mindset relative to Likert and, if so, document the value of that feature as well.

2 Methods

2.1 Participants

Participants for the study were provided by Toluna, one of the largest market research providers in the United States. Initially, requests to participate were sent to all members of

the Toluna system. As results came in, additional requests were sent in order to better align the sample's demographic characteristics with those of the United States. In total, a sample of 4850 individuals was drawn between February 5 and 15, 2016. Panelists were compensated in "Toluna points" for completed surveys, which they could redeem for various items. Only those participants who completed a "screener" section successfully, which confirmed that they were at least 18 years old and registered to vote, were directed to an online survey questionnaire.

2.2 Experimental design

Participants were assigned randomly to one of three categories: Likert (referred to as Likert-only), QV (QV-only), or Likert and then QV (Likert-QV). See Table 1 for a demographic summary of the Likert-only group; results were similar for the other groups.

All three subsets of respondents were asked to express opinions about the same policy proposals. These appeared on the survey as:

1. Requirement for background checks for all gun purchasers.
2. Elimination of the Affordable Care Act of 2010 (aka 'Obamacare').
3. Raise the minimum wage from $7.25 to $10.10 over the next 3 years.
4. Nationwide ban on abortion in nearly all circumstances.
5. Sharp reduction of federal and state programs intended to help the poor with costs of food, shelter and education.
6. Immediate deportation of any person who is found to be living in the United States illegally.
7. Raise taxes on corporations and the wealthy to combat income inequality.
8. Send large numbers of US ground troops to fight ISIS militants in Syria and Iraq.
9. Do not allow Syrian refugees into the United States.
10. Legally require employers generally to pay women the same salary as men receive for the same work.

In the Likert-only condition, respondents were asked to complete a survey that presented the ten public policy proposals listed above in randomized order. Each proposal allowed the respondent to select one of seven possible response choices consistent with the ranges recommended by Krosnick and Fabrigar's (2012) review of the relevant literature: "Very strongly against"; "Strongly against"; "Somewhat against"; "Neutral"; "Somewhat in favor"; "Strongly in favor"; "Very strongly in favor". The online system enabled respondents to navigate between questions, to change previous answers and to skip questions.

The QV-only participants viewed a 90-s introduction to the mechanics of the QV software as well as an explanation of the hypothetical budget they would be able to allocate and a preview of the proposals on which the respondent was about to vote. The QV survey itself presented the ten proposals in randomized order. Respondents were to allocate 100 credits across the ten proposals with the quadratic pricing mechanism active. So, for example, if a respondent decided to purchase four votes in favor of the first proposal, 16 credits would be deducted from his/her budget, thus leaving 84 credits to allocate across the remaining nine proposals. The user interface allowed the respondents to scroll up and down through the proposals as they wished, particularly if they wanted to adjust the votes cast on an earlier proposal to free up credits to use elsewhere in the survey.

In the Likert-QV condition, respondents were presented with both the Likert items and the QV survey, in that order.

Table 1 Descriptive statistics for the Likert-only condition. The QV-only and Likert-QV participants were demographically similar

Participant demographics	
Gender (%)	
Male	49
Female	50
Age (%)	
18–20	4
21–29	18
30–39	17
40–49	17
50–59	16
60–69	19
70+	9
Highest education level attained (%)	
Didn't finish high school	2
High school degree	20
Some college	24
Associate degree	13
Bachelor degree	26
Graduate degree	15
Political affiliation (%)	
Strong Democrat	27
Weak Democrat	9
Independent Democrat	14
Independent	21
Independent Republican	8
Weak Republican	6
Strong Republican	11
Other	4
Race (%)	
White	64
Latino	10
Black	19
Asian	4
Other	2
Region of residence (%)	
Midwest	21
Northeast	25
Southeast	27
Southwest	11
West	16
Annual income (25–35K means $25,000–$34,999) (%)	
<25K	17
25–35K	12
35–50K	15
50–75K	18
75–100K	17

Table 1 continued

100–150K	10
150–250K	4
250–500K	0.4
500K–1M	0.5
1M+	0.5
Refuse to answer	4

Results may not add to 100% because of rounding or participants refusing to answer

After completing their respective tasks, respondents in all three conditions were requested to provide basic demographic information, asked for survey feedback and given a multiple-choice follow-up question: "Which of the following issues would you be interested in getting more information for how you might take action or influence public policy? Please check all that apply." The ten topics were presented in the order in which they appeared in the QV or Likert portion of the survey.

In addition to the survey responses, the software captured metadata to reflect participant actions during the course of survey completion. This information included actions such as movement among and between survey sections, time spent in each section and time lapsing between responses.

2.3 Survey software

The most important difference between this study and all past work on quadratic voting was the development of proprietary software called "weDesign", tailored specifically for our research aims. It is capable of capturing both participants' responses and the metadata of how the respondent engages with the survey itself. For comparative purposes, we also built a tool to conduct standard survey research on the same platform used in the QV condition so as to ensure uniformity in gathering data and metadata. We also added a further improvement by allowing quadratic vote allocations to take place across several different items at once, rather than the single-issue approach on which much of the past QV literature has focused. Also, because requiring respondents to commit their own personal resources to the test raised a host of thorny issues, including income effects, participants instead were allotted a fixed budget of virtual "credits" to spend on individual policy proposals. As discussed earlier, respondents were then allowed to cast votes either for or against each proposal at the quadratic cost dictated by QV theory. Each participant was endowed with an overall budget of 100 hypothetical credits which could be spent in any way the respondents' chose, allowing them to allocate 10 credits to each proposal, 100 credits to one proposal and none to the nine others, or any other way the respondent desired. This virtual credit system, which places an opportunity cost on point allocation, is similar to the mechanism developed by Hylland and Zeckhauser (1979) where virtual credits are allocated in proportion to their marginal utility (Tideman and Plassmann 2016).

2.4 Completion rate

From the 4850 panelists who qualified, we realized the following sample sizes for each test condition:

- Likert only (n = 1095)
- QV only (n = 1048)
- Likert-QV (n = 1055)

Researchers are advised to maximize completion rates by reducing the burden on respondents, which they can best do by keeping surveys simple and short. Likert, perhaps the simplest and expected form of survey questioning, is relatively quick and easy to understand and complete; it imposes little or no burden on respondents. QV, which demands that tradeoffs be made across items, increases the participant's cognitive load and requires more time. And, as a new methodology, we added to the time demanded by including an introductory video to explain the approach and make clear the nature of the process. Yet, despite the greater demands of and unfamiliarity with the QV approach, completion rates were comparable for Likert and QV. The completion rates for Likert-only, QV-only, and Likert-QV were almost identical: 68, 65 and 65%, respectively. This suggests that QV does not impact the representativeness of participants negatively and, if streamlined, could even provide some added value to our study.

Furthermore, we checked the demographic characteristics of the respondents in each condition and found no statistically significant differences. It is thus reasonable to conclude that abstention was demographically consistent across test conditions. As such, differences seen in voter behavior can be attributed to the test conditions (i.e., the methodologies), rather than to participants' personal characteristics.

3 Results

This section outlines results from the study organized along three major themes: an exploration of the macro level (all Likert voters compared to all QV voters), a look at the micro level (the average Likert voter and the average QV voter), and insights into the predictive powers of the two methodologies. In the first two sections below, we frame our argument around the "scarcity mindset" mentioned above, and then use the third section to show the value in moving from abundance to scarcity.

3.1 Changing the group landscape

The most powerful macro effect of QV, relative to Likert, is one of moderation. Because QV voters can no longer express their preferences at the extreme ends of the opinion scale without consequence, as they can in Likert, they cast fewer votes there. Interestingly, not only do voting patterns migrate from the extremes, but they also normalize in the process. Figure 1 shows the distribution of votes on two proposals for the Likert-only and QV-only conditions. Note that similar results are present when voters are compared to themselves in the Likert-QV condition. Here, we have converted the language of the Likert scale to a numerical range, namely $\{-3, \ldots, 3\}$. These are the 7 "vote levels" of the Likert scale. In contrast, QV respondents were allowed to vote on a scale that ranged from -10 to 10, or 21 different vote levels.

As the Fig. 1 suggests, what appear to be strong ideological differences, as seen in the skewed and multi-modal graphs for Likert-only participants, tend toward more balanced opinions (symmetric, quasi-normal distributions) for QV-only participants. These claims can be made more rigorous by exploring the skewness (degree of symmetry about the mean) and excess kurtosis (heaviness of the tails relative to the normal distribution) for each proposal and methodology. Note that perfectly symmetric distributions are not skewed at all, and that normal distributions exhibit an excess kurtosis (defined as kurtosis—3) of 0. The results for all 10 proposals are reported in Table 2.

In all cases but one (proposal 2), the skewness of the QV results is smaller than those of the Likert scale, suggesting more symmetry around the mean. When this finding is coupled

Fig. 1 Voting distributions on two proposals (paying women equally and banning abortion) in the Likert-only (n = 1095) and QV-only (n = 1048) test conditions

Table 2 Skew and excess kurtosis for each proposal and methodology

Proposal	Likert skew	QV skew	Likert excess kurtosis	QV excess kurtosis
1. Gun background checks	−1.61	−0.31	1.82	1.33
2. Repeal Obamacare	−0.01	0.14	−1.42	−0.64
3. Up minimum wage	−0.88	−0.48	−0.27	0.66
4. Ban abortion	0.45	0.31	−1.13	−0.17
5. Decrease govt. programs	0.40	0.20	−1.08	−0.55
6. Deport illegals	−0.27	0.02	−1.09	−0.31
7. Tax wealthy	−0.68	−0.36	−0.71	0.08
8. Fight ISIS	0.10	0.10	−1.03	0.07
9. Block Syrians	−0.19	0.06	−1.12	0.17
10. Pay women equally	−1.43	−0.24	1.18	1.18

Skew values near 0 suggest symmetric distributions; excess kurtosis values near 0 suggest tails that mimic a normal distribution

with excess kurtosis values getting closer to 0 as we move from Likert to QV, the distributions of responses for the latter become closer to normal. Proposals 1 and 10 have larger positive kurtosis values (leptokurtic distributions) for both Likert and QV, suggesting more consensus on those issues. Indeed, once voting distributions become quasi-normal, the kurtosis value can be used as a quick summary of the degree of consensus.

3.2 Changing the individual landscape

The previous results suggest that at a high level, QV changes "what we decide" (adopting Mullainathan and Shafir's 2013 language). But can the Likert-QV divergence be seen at the individual voter level? Is there evidence for differences in "how we deliberate"?

We begin with some summary statistics (Table 3) showing the average experience of voters with each methodology.

Table 3 reveals that QV voters tend to spend about 30 more seconds voting (or 29% longer) when they are seeing the proposals for the first time (single methodology groups), and about 14 s more (or 14% longer) when they have seen the proposals before and already had opportunities to think about them (Likert-QV condition). In addition, QV voters take more voting actions overall (15 compared to 11) because they are revising previous votes more often. In Likert, a revision is defined as selecting a vote intensity on a question for which the user already had selected a preference intensity (this is done by selecting a radio button). In QV, revisions occur in two ways: (1) If the voter votes on an issue, goes to a new issue, and then returns to the first issue, or (2) If the voter votes on an issue, stays on the issue, and casts a new vote with at least 2 s of time between casting the first vote and starting to cast a second. Each time a revision occurs, it can be classified as a revision stronger (away from 0), weaker (toward 0), or holding at the current level (e.g., in Likert, repressing a radio button at the original voting intensity; in QV, a move from 7 to 6 to 7 in a short period of time). Remarkably, on average, QV voters tend to make about five total vote revisions (combining the three types), while Likert voters make only one.

These revision statistics get to the heart of the QV experience: When resources are scarce, voters cannot assign extreme answers to all proposals, and so voters must confront the reality of expressing measured opinions that reflect the tradeoffs of life. Interestingly, these behavioral differences can be seen "in real time" by plotting a voter's total votes cast as he or she progresses through the survey. More specifically, for a given voter, we plot the accumulated total of votes cast (ignoring whether votes are for or against) versus the time elapsed since the start of the survey (as a percentage of the total time spent).

Table 3 Average time spent voting, number of voting actions, and numbers and types of revisions in Likert and QV surveys for each methodology and the joint (Likert-QV) one

Methodology	Likert-only	QV-only	Likert results from Likert-QV	QV results from Likert-QV
Time voting (s)	102.2	132.4	100.31	114.3
Voting actions	11.01	15	10.98	15.2
Revisions	1.01	5	0.98	5.2
Revisions stronger	0.39	3.5	0.34	3.6
Revisions weaker	0.33	1.3	0.29	1.5
Revisions hold	0.29	0.2	0.34	0.2

Two graphs of that type for each methodology are shown in Fig. 2. To avoid plotting the voting trajectory of 1000 voters per methodology, we sort all the voters based on the number of revisions made, break them into 50 quintiles (or about 20 voters each), and then

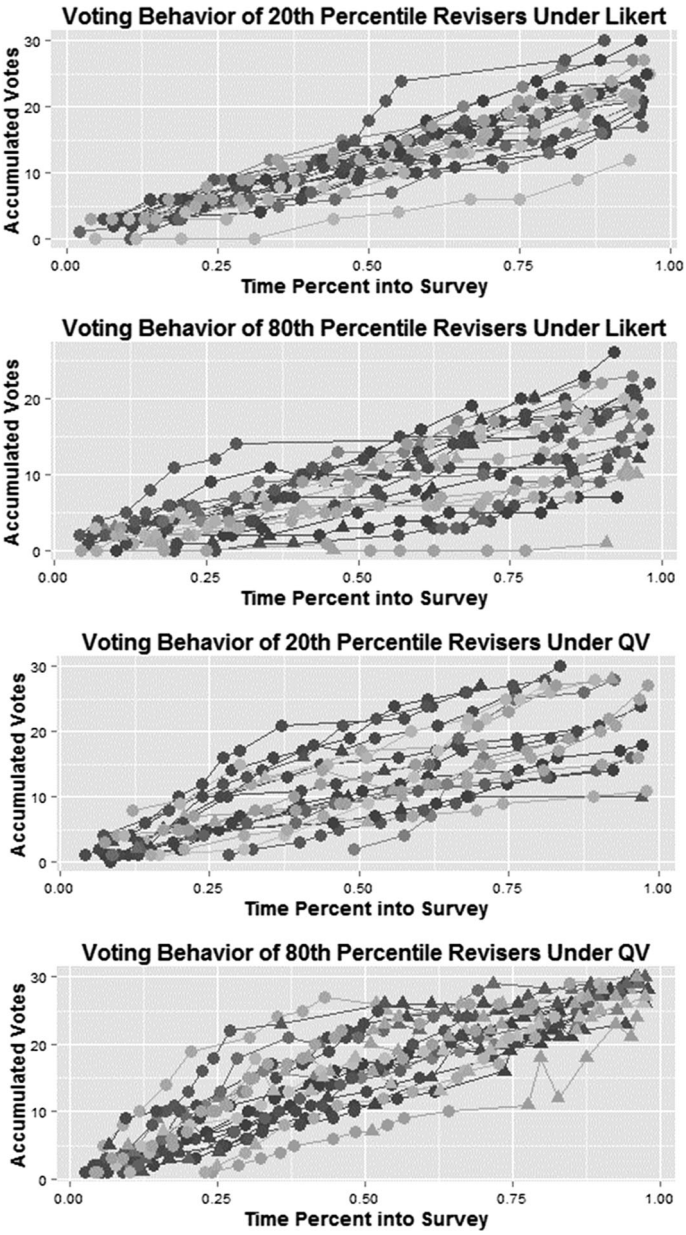

Fig. 2 Vote accumulation versus time into survey for Likert and QV methodologies using the Likert-only and QV-only conditions. A single voter's behavior is encoded in a jagged line. These *graphs* show the behavior of voters around the 20th and 80th percentiles of the number of revisions made. Results are similar for the Likert-QV condition. *Note* that voting actions are discrete events, but here, a *line graph* is used for presentational convenience

Table 4 The percentage of proposals voted on and later revised (on average) by voters using a given methodology

Data source	% of proposals that voters revised, on average
Likert-only	8.44
QV-only	34.69
Likert results from Likert-QV	8.14
QV results from Likert-QV	35.37

plot the data for voters in quintiles 10 and 40 (or the 20 and 80% most revision-heavy participants). That is, we explore the accumulating vote trajectories of low and high revisers for each methodology. In these diagrams, note that circles denote first-time votes on a proposal while triangles indicate revisions to previous votes.

Figure 2's graphs demonstrate that Likert voters in the 20th percentile of total revisions made (TRM) tend to make few revisions, and, as such, their accumulated votes generally rise monotonically as time elapses. By the 80th percentile of TRM, Likert voters begin to make some revisions, but, as before, most graphs are increasing monotonically. In contrast, the 20th percentile of TRM for the QV methodology looks quite similar to the 80th percentile of Likert: revisions are made occasionally (triangle icons), and some graphs begin to change direction. By the 80th percentile, more directional changes are evident and the total number of voting actions has risen visibly. In addition, voters tend to end up casting about the same total number of votes (in contrast to the other three graphs): QV voters are coaxing the most votes possible out of their limited credit banks.

For each voter and each methodology, we can calculate the percentage of proposals the voter revisited. As Table 4 suggests, the average Likert user revises about one in every 12 proposals they encounter (about 8%). In contrast, the average QV voter revisits about one in every three proposals (about 33.3%).

These findings are extended further in Fig. 3, which shows the distribution of voting revisions for the two methodologies in both the single (Likert-only versus QV-only) and dual methodology (Likert-QV) conditions.

Figure 3 suggests that almost half of all Likert voters never revise their opinions on a single policy proposal, and that it is rare for anyone ever to revise more than three. In contrast, QV voters often revise more than three proposals (Table 3 shows that the mean is between three and four proposals), and that revisions of seven or more proposals are not uncommon. Furthermore, these results are the same in single and dual methodology test conditions, so the differences revealed are inherent to the methodologies, not to the voting samples.

Having toured the macro- and micro-landscape of Likert versus QV, we turn now to issues of predictive power.

3.3 Insights on predictive power

Two critical questions remain for Likert and QV: (1) How do findings from the Likert and QV conditions relate to known qualitative trends in polling, and (2) how do these findings differ in helping researchers understand voters and predict their behavior? This second question is of particular importance, for differences can materialize between voters' espoused views and the actions they are willing to take based on those claims.

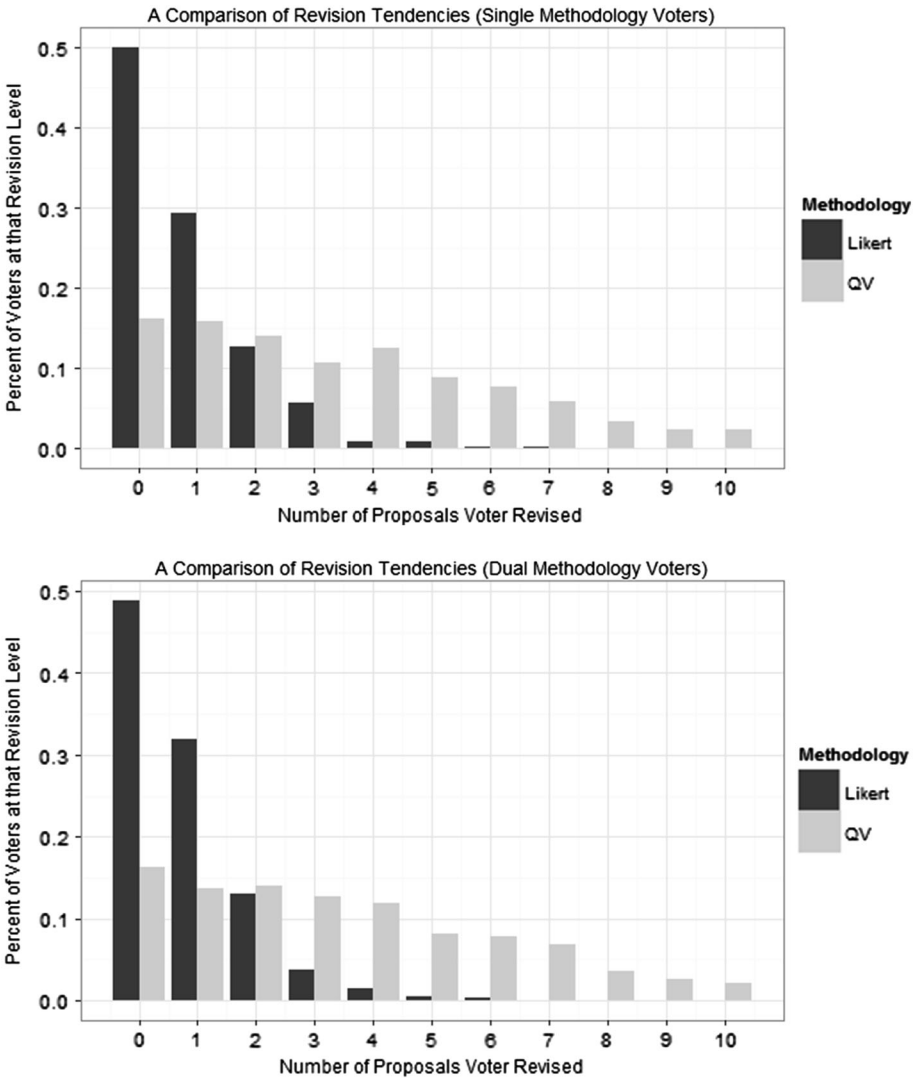

Fig. 3 Distribution of proposal revision counts for Likert versus QV methodologies in both the single and dual survey conditions. The means of these distributions are reported in Table 3

Turning first to question 1, we focus on a particular trend seen in political polling. With the rise of the independent voter in the last 30 years, many survey instruments have begun using a seven-category distinction for political affiliation: Strong Democrat, Weak Democrat, Independent Democrat, Independent, Independent Republican, Weak Republican, and Strong Republican (Sides 2014). Despite this seven-way distinction, it still is common practice in some circles to create only three distinctions: Democrats (including Strong Democrat, Weak Democrat, and Independent Democrat), Independents (including Independent), and Republicans (including Independent Republican, Weak Republican, and Strong Republican). This practice emerges from analyses of actual voters. As Sides (2014, p. 1) writes, "[I]ndependents who lean toward a party … behave like partisans, on average. They tend to be loyal to their party's candidate in elections."

Naturally, one might ask whether a voting methodology supports this clustering practice. To explore this issue, we take a given methodology (e.g., Likert) and a given proposal (e.g., Increasing the minimum wage), and turn all votes on that issue into z-scores. Then, we group together votes for the same political affiliation (e.g., Strong Democrat) and compute the mean. Repeating this process for the other party affiliations and methodologies, we can observe if the averages differ in Likert versus QV. Given that z-scores negate the methodologies' inherent scale differences, we get an accurate read on where participants can be placed along the seven partisan categories for each given issue. Figure 4 displays two examples.

The first graph above provides a clear example of the differences between Likert and QV. While the Strong Democrats, Weak Democrats, and Independent Democrats look like three distinct groups under Likert, they move to common ground under QV. As such, QV suggests that the attitudes of these groups are closer to one another than Likert does. A similar, but weaker, trend is seen on the Republican side. The second graph, concerning equal pay for women, displays similar patterns: the opinions of the three left-leaning party affiliations come together, as do the three right-leaning ones. Here, we also see that QV "pulls apart" Independents and Independent Republicans, with the latter group heading toward other Republican-leaning affiliations.

We turn now to question 2 above and explore which methodology helps researchers understand voters better and the possibility of turning opinions into action. To do this, we first label a set of votes in each methodology as "moderate" or "strong". Table 5 shows the percentages of votes cast in certain ranges for the two methodologies (Likert-only and QV-only conditions).

As Table 5 suggests, it is reasonable to consider the QV vote levels with absolute values of 3 or greater as "strong" because they map closely (in percentage terms) to the labels "Strongly Disagree" and "Strongly Agree" in Likert (-3 and 3, respectively). This definition of "strong" vote suggests one advantage of QV: while Likert collects strong votes at only two levels (-3 and 3), QV distributes this same percentage of votes across 16 different categories (-10 to -3 and 3 to 10). For future discussion, we define a "moderate" vote as one in the $\{-2, \ldots, 2\}$ range in either methodology. We argue that one of QV's greatest benefits is its ability to make sense of the "strong" vote (recorded about 42% of the time) and the voters who express those opinions.

We turn, first, to voters casting strong votes. Table 6 reports the percentage of voters that cast strong votes on at least b proposals, where b takes on the values 6 through 10. Note for $b = 6$ (and likewise for 7–10), this means that the voter is expressing strong opinions on the majority of proposals. These data are derived from the Likert-QV condition, so observed differences are inherent to the methodologies.

The previous two tables indicate that while the total number of "strong" votes is roughly the same across methodologies, the percentage of voters who *express such opinions frequently* is smaller in QV. That finding suggests that the information we get from a voter using QV will be less influenced by extreme opinions than when using Likert. For example, Fig. 5 shows the voting profile of two "extreme" voters. Here, an "extreme" voter is defined as one who gives a strong vote on eight or more of the ten proposals when using Likert (hence, by the above table, 13.1% of voters are extreme). In the first graph, we see that the Likert extremism is greatly reduced in QV. As such, we get a more accurate sense of the relative importance of the proposals.

As the top panel of Fig. 5 indicates, the Likert data for Extreme Voter #1 are of little value in deciding what matters to that voter: apparently everything but the deportation of illegal immigrants is viewed as of the highest degree of importance. Under QV, one sees

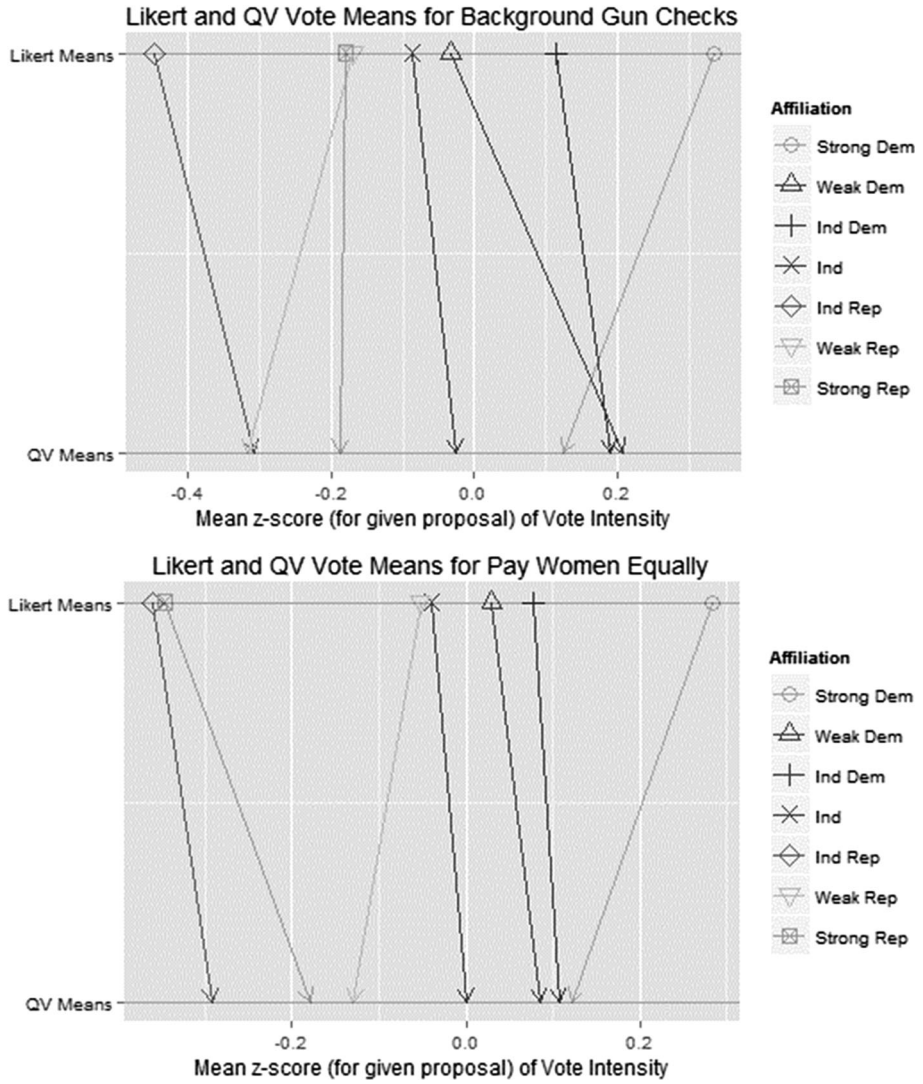

Fig. 4 Examples of two proposals for which QV shows greater support for the three-party clustering approach. Results are similar for the Likert-QV condition. *Note* that these results show relative, not absolute movement (in relation to the mean), that is, the *graphs* do not indicate whether participants aligning with the corresponding affiliation are voting for or against the various policy proposals

that guns, Obamacare, abortion, and ISIS carry more weight—with issues like minimum wage and government programs taking a back seat. The bottom panel, Extreme Voter #26, shows a respondent who cares about all issues but ISIS under Likert. Under QV, however, we see that only the minimum wage, taxation, and equal pay for women receive votes.

This behavioral change is not unique to the extreme voters depicted in Fig. 5. Figure 6 shows the voting profiles of 12 additional extreme voters. When taken together, Figs. 5 and 6 gives a broad sense of how extreme voters react when faced with the reality of tradeoffs found in QV.

Table 5 Percentages of votes in given ranges for both Likert and QV using the Likert-only and QV-only conditions

Vote levels	{−3}	{−2, …, 2}	{3}
Likert (%)	15	58	27
Vote levels	{−10, …, −3}	{−2, …, 2}	{3, …, 10}
QV (%)	13	59	28

Note the similar percentages in the three vote-level spans

Table 6 Percentages of voters casting strong votes on at least *b* proposals in the Likert-QV condition

Value of *b*	6	7*	8*	9*	10*
Likert (%)	33.2	22.6	13.1	7.7	3.0
QV (%)	31.2	16.4	7.2	3.6	1.4

Asterisks indicate the *b* values for which the Likert and QV proportions have a statistically significant difference

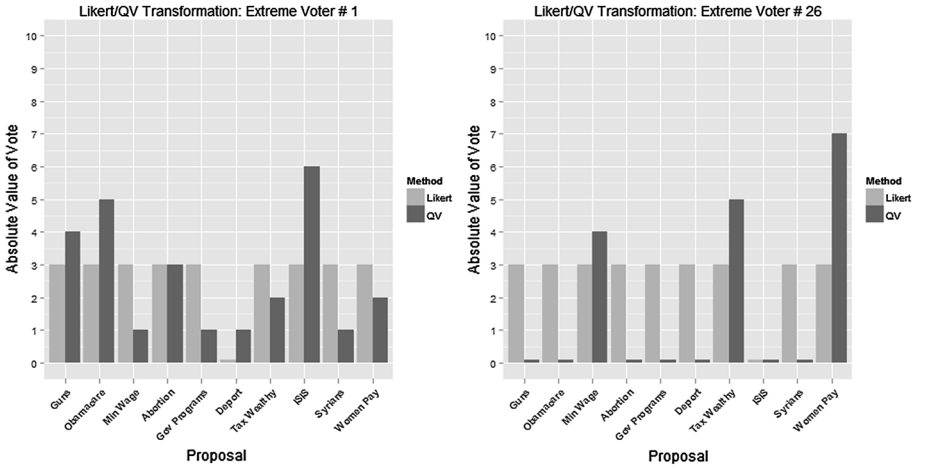

Fig. 5 An exploration of two "extreme" voters from the Likert-QV condition. Here, "extreme" is defined as casting strong votes on 8 or more proposals in Likert. Under QV, voting profiles are more nuanced. Absolute values of votes are used to simplify the presentation

To see in greater detail how QV brings meaning to a strong vote, we explore the relationship between vote strength and action. Recall that at the end of each session, respondents were asked a question that called for a commitment with apparent consequence beyond the test environment. They were asked which of the issues they would be interested in getting more information about so that they might take action or influence public policy (they were allowed to "check all that apply"). While such indications of interest are not a perfect measure of actions voters would take at the polls, on social media,

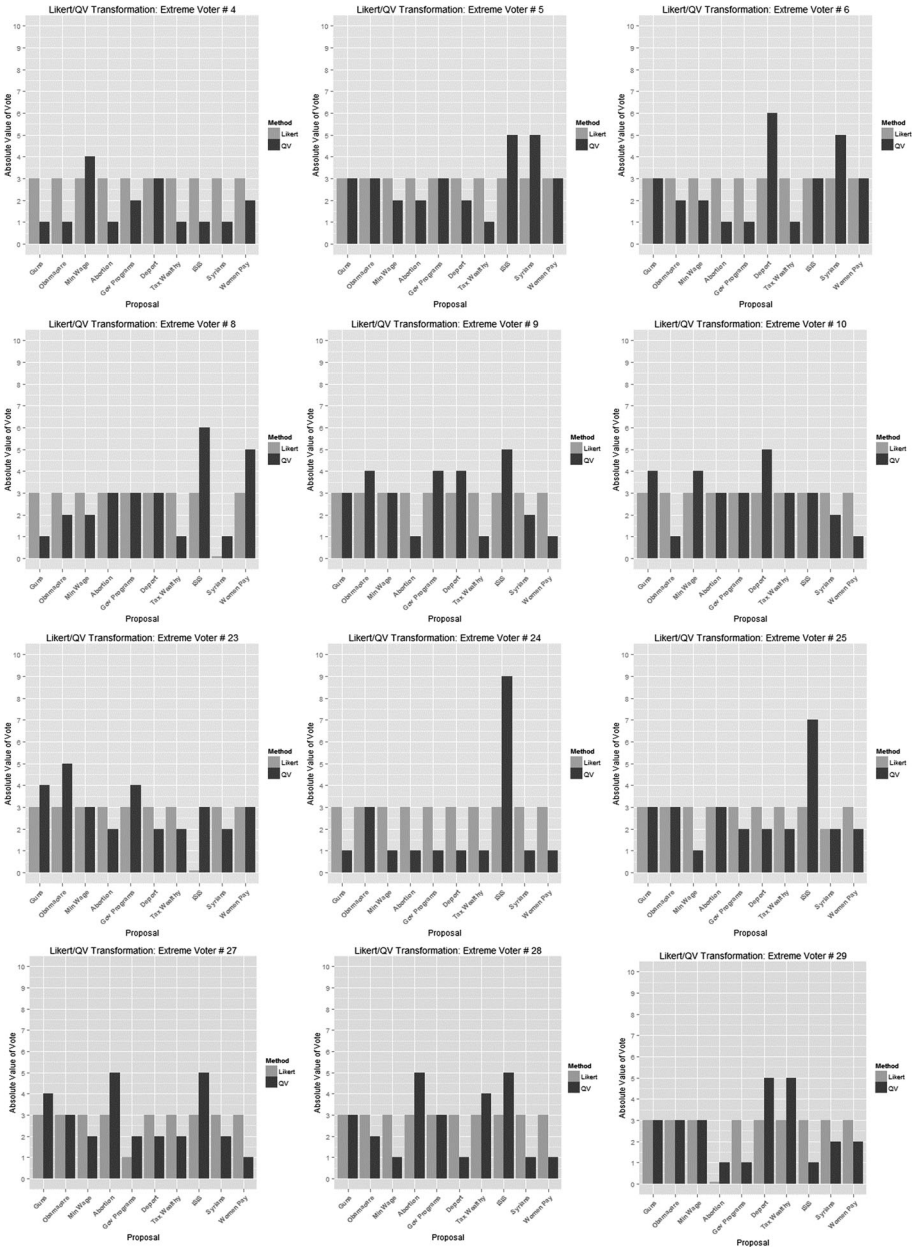

Fig. 6 *Graphs* showing the behavioral changes in some additional extreme voters. While voters appear to strongly care about nearly all issues in Likert, QV brings out nuance in their voting profiles

or in grassroots campaigning, they do provide a window into a world beyond the voters' stated preferences.

We note at the outset that the QV methodology did not cause voters to check boxes with any greater frequency than Likert did. Indeed, Table 7 shows that the percentage of box-

Table 7 Percentages of participants asking for more information on policy issues by intensity of Likert and QV voting in the Likert-only and QV-only conditions

Vote Intensity	Strong negative	Moderate	Strong positive
% Boxes checked under Likert	30.3	23.5	42.3
% Boxes checked under QV	30.4	23.4	42.8

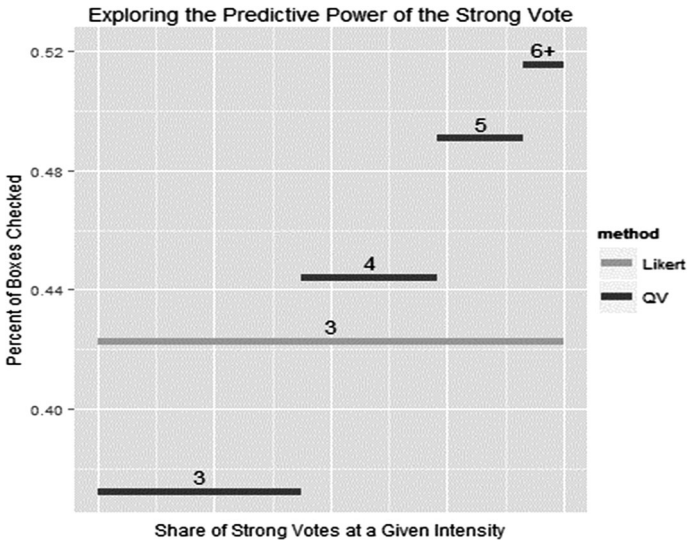

Fig. 7 The probability that participants will ask for more information on an issue when casting a certain "strong" positive vote in Likert and QV. The width of each line segment shows the fraction of "strong" votes at the given intensities. QV levels 6–10 have been grouped to ensure a sample size sufficient to produce a stable estimate. Similar results occur with negative "strong" votes

checking (that is, requests for more information on how to take action) is nearly identical for negative "strong" votes, "moderate" votes, and positive "strong" votes.

While these percentages essentially are identical, they hide a stark difference between the Likert and QV methodologies. Said simply, the "strong" vote in Likert aggregates a wide range of policy opinions—from the voter who espouses strong support initially, but caves under the pressure of the QV voting budget constraint, to the voter who redoubles his/her vote intensity under that constraint. One might expect that voters in these two extreme cases would *behave* differently as well. Unfortunately, Likert does not offer a voting milieu that aligns vote intensity and action. As seen in Fig. 7, the "strongest" voters in Likert are likely to express great interest: 42.3% want more information. In contrast, QV disaggregates these voters, identifying those who are less likely to act (the QV dark gray, horizontal line segment labelled "3" in Fig. 7) despite casting a "strong" vote, and those whose votes speak quite powerfully to their probability of taking action (the QV dark gray horizontal segments labelled 4, 5, and 6+). In essence, QV breaks the light gray Likert line into multiple pieces, sorting voters into groups that more fully reveal their true action-based preferences, both weaker (3) and stronger (4, 5, and 6+).

4 Discussion

This study shows that QV offers a promising alternative to Likert for opinion polls. The results demonstrate that QV achieves a quasi-normal distribution of preferences, something Likert has never done; engages a higher level of thoughtfulness; attenuates extremism to expose deeper insight; and predicts behavior better than Likert at high levels preference intensity.

Perhaps the most powerful macro effect of QV is one of moderation. Because voters no longer are able to express opinions at the extreme ends of the scale without consequence, owing to a vote-credit constraint, they cast fewer strong votes. Interestingly, not only do voting patterns migrate from the extremes, but they also normalize in the process. That moderation like this is reflective of deeper insight is validated by two factors: (1) it maps well onto a willingness to take action and (2) other constraint mechanisms, such as a linear constraint, have been shown to limit respondents to just one or two key issues, while QV allows for a greater diversity of expression (Louviere and Islam 2008).

In terms of engagement, the "total revisions made" (TRM) results and the revisions to votes on earlier proposals suggest that QV engenders greater thoughtfulness among those using it. Indeed, evidence drawn from qualitative user experience (UX) testing and the comments section available at the end of the survey indicate that QV voters often cite two main reasons for making revisions. First, they note that the methodology forces them to determine their preference intensities not in a vacuum, but rather in comparison to intensities on other issues. Often, these comparative intensity levels are quite different than those seen in isolated settings, and not something the voter has considered before. Second, while voters are tempering and ordering their preferences, they still want to reveal those preferences as strongly as possible. Thus, many voters work to use every last credit in expressing their opinions (hence, the large number of revision triangles in the upper right corner of the 80th percentile QV plot of Fig. 3). These two forces—preference realignment and voting profile maximization—are examples of what Mullainathan and Shafir (2013, p. 24) refer to as the "focus dividend": "the positive outcome of scarcity capturing the mind".

Taken together, the results on revision tendencies and accumulated votes as the opinion-gathering survey progresses suggest some of the changes that are going on at the individual voter level. Not only do QV voters spend longer with the QV instrument, they make five times the number of revisions, and revisit almost 1 in 3 proposals when making those changes. Furthermore, the accumulating vote graphs (Fig. 2) suggest that the 80th percentile of Likert voting is just beginning to look like the 20th percentile of QV voting, a possible consequence of the focus dividend inducing greater thoughtfulness in the presence of credit scarcity.

This aspect of respondent thoughtfulness is a powerful one for anyone utilizing survey research. The concept of "survey satisficing" has been a growing concern in research circles. Satisficing in online surveys is particularly common given that, unlike telephone or face-to-face interviews, the only person determining the interview's speed is the respondent him- or herself. Thus QV's demonstrated ability to engage respondents in such a way that they are willing to return to questions already asked and answered and to revise their previous responses seems a strong step in addressing this aspect of satisficing.

In the study reported herein, more engagement and less satisficing show great promise for understanding the populations from which survey samples are taken. QV and Likert approaches to understanding public opinion on topical and often divisive issues reveal

stark differences in results. For example, this study's data show that on several hot-button issues, seven separate political party identifications collapse together when placed in a QV context. This has important philosophical ramifications for questions like, "How divided are we as a country?" The answer to such a question is a function of both the voters themselves and the methodology used to elicit their opinions. Strong ideological divides in one methodology may dissolve in a different methodology, especially one like QV that forces then to make preference tradeoffs across issues.

Not only does QV reduce extreme ratings/voting, it tempers ratings such that they are more meaningful and predictive of relevant and even costly behavior. QV infuses the notion of a "strong" vote or opinion with greater granularity, disaggregating those who claim to care deeply about an issue but are unlikely to act, from those who care and are likely to act. Furthermore, these findings are not simply the result of arbitrary differences in the scales used by the Likert and QV methods. A review of years of research on the optimal number of Likert scale points by Krosnick and Fabrigar (2012) indicates a curvilinear pattern such that scales from five to seven points are more reliable than scales with either fewer or more points. Seven points were found to be optimal for bipolar scales such as those used here. As such, Likert (in its near-optimal configuration) is outperformed by QV (with no scale-size optimization).

5 Future considerations

Despite the value-added by QV demonstrated here, clear challenges lie ahead. For example, the histograms of QV data show an unexpected dip at vote intensity zero. An initial hypothesis to account for the observed drop off is that, given the quadratic pricing mechanism we implemented, some respondents were left with unused credits at the end of the survey and then sought to redistribute their remaining credits in such a way that allowed them to spend their credits fully, namely by purchasing one vote for issues on which they had originally wanted to remain neutral (the zero vote level). Given the economic theory of consumer behavior, this hypothesis is reasonable given that rational actors wish to maximize utility. Unfortunately, it remains unclear whether purchasing that vote reflected a genuine sentiment on the issue from the respondent, or if the simple psychological satisfaction of achieving zero credits remaining is so strong that the respondent is willing to indicate that they hold a belief on an issue that they truly do not have—an example of what Mullainathan and Shafir (2013) call "tunneling". Future testing should address this question. One path would be to move from discrete to continuous vote allocations. This would allow respondents to purchase tenths or even hundredths of votes, as QV theory contemplates. Another would be to "refund" or allow respondents to "bank" otherwise unused credits for future use.

Also, while this study demonstrates the promise of QV in the wild, more work is needed is understand and optimize the parameters available within the QV framework. For example, experimentation should take place varying the number of items on the survey, the types of issues included, the number of vote choices in the QV scale, and allowing more extensive qualitative inquiry into respondents' interactions with QV content.

While our analysis of QV in the context of public opinion surveys validates its potential empirically, fully realizing that potential will require pursuing at least two avenues of additional experimentation:

1. Demonstrating or evaluating QV's applicability to other types of market research, e.g., product concept testing, brand assessment.
2. Improving user experience, e.g., streamlining the onboarding process and dealing with leftover credits.

Addressing the second issue presents possibly the most interesting challenge for the future. As part of the survey, we gathered participants' ratings of the two methods and requested feedback on the survey experience. Respondents rated the Likert-only experience 8.5 out of 10, on average (10 being the most positive rating) and the QV-only experience as 7.9. Though the difference is small, it is statistically significant (using the Mann–Whitney U test, $p = 4.5e{-}10$). A basic sentiment analysis of the comments left by the fraction of the participants who wrote ten words or more in feedback (260 total respondents) showed that "highly positive" comments exceeded "highly negative" comments by 3–1. Among those "highly negative" comments, some respondents seemed to struggle with the limit imposed by the quadratic pricing mechanism and felt they were unable to express how they truly felt. That sentiment is an expected feature of inducing a scarcity mindset since respondents lose the freedom available in an "abundance" situation. Further work is needed to fully understand the benefits and costs of transitioning survey participants from an abundance to scarcity mindset using methodologies such as quadratic voting.

References

Anand, S., & Krosnick, J. A. (2003). The impact of attitudes toward foreign policy goals on public preferences among presidential candidates: A study of issue publics and the attentive public in the 2000 U.S. presidential election. *Presidential Studies Quarterly, 33,* 31–71.

Fournier, P., Blais, A., Nadeau, R., Gidengil, E., & Nevitte, N. (2003). Issue importance and performance voting. *Political Behavior*, *25*(1), 51–67. http://www.jstor.org/stable/3657314. Accessed 21 March 2016.

Gorn, G. J. (1975). The effects of personal involvement, communication discrepancy, and source prestige on reactions to communications on separatism. *Canadian Journal of Behavioral Science, 7,* 369–386.

Hylland, A., & Zeckhauser, R. (1979). A mechanism for selecting public goods when preferences must be elicited. Working paper, Kennedy School of Government, Harvard University.

Krosnick, J. A. (1988). Attitude importance and attitude change. *Journal of Experimental Social Psychology, 24*(3), 240–255.

Krosnick, J. A., & Fabrigar, L. R. (2012). Designing rating scales for effective measurement in surveys. *Lyberg/Survey Survey Measurement and Process Quality*, pp. 141–164.

Krosnick, J. A., Presser, S., Fealing, K. H., & Ruggles, S. (2015). The future of survey research: Challenges and opportunities. *NSF*. The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research. https://www.nsf.gov/sbe/ac_materials/the_future_of_survey_research.pdf. Accessed 21 March 2016.

Krosnick, J. A., Visser, P. S., & Harder, J. (2009). The psychological underpinnings of political behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 1288–1342). New York, NY: Wiley.

Lalley, S. P., & Weyl, E. G. (2016). Quadratic voting. *SSRN Electronic Journal SSRN Journal*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2003531. Accessed 1 Aug 2016.

Louviere, J. J., & Islam, T. (2008). A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *Journal of Business Research, 61*(9), 903–911.

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. New York: Times Books.

Sides, J. (2014). Most political independents actually aren't. *Washington Post*. The Washington Post. https://www.washingtonpost.com/news/monkey-cage/wp/2014/01/08/most-political-independents-actually-arent/. Accessed 21 March 2016.

Tideman, T. N., & Plassmann, F. (2016), Efficient collective decision-making, marginal cost pricing, and quadratic voting. *SSRN Electronic Journal SSRN Journal*. http://ssrn.com/abstract=2836610. Accessed 11 September 2016.

Visser, P. S., Krosnick, J. A., & Simmons, J. P. (2003). Distinguishing the cognitive and behavioral consequences of attitude importance and certainty: A new approach to testing the common-factor hypothesis. *Journal of Experimental Social Psychology, 39*(2), 118–141.

Weyl, E. G. (2012). Quadratic vote buying. *SSRN Electronic Journal SSRN Journal*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2003531. Accessed 1 August 2016.