

## Quadratic Voting in the Wild: Real People, Real Votes

David Quarfoot, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, and Ellen Konar<sup>1</sup>

### 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 a quadratic price for votes. Comparable QV-based and Likert-based survey instruments designed by Collective Decision Engines LLC were experimentally evaluated by randomly assigning potential respondents 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 intense respondents than the Likert-based instrument does. We consider the implications for survey satisficing, a key threat to the continued value of survey research, and reveal the mechanisms by which QV differentiates itself from Likert-based scales, thus establishing QV as a promising alternative survey tool for further political and commercial research. We also explore key design issues within QV-based surveys to extend these promising results.

### INTRODUCTION

National surveys have long played a unique and critical role in reflecting public opinion to inform the conduct of democracies (Krosnick, Visser, Harder, 2009). Surveys have also played a critical role in commerce, improving the odds that companies understand consumers or potential business customers and thereby invest in the most desirable new products and features. Regardless of domain, the value of surveys depends on their accuracy.

Remarkably, survey and questionnaire design has changed little in the last 50 years. The Likert scale, the nearly universal method of attitude measurement, was initially developed and tested by Rensis Likert in the 1930’s. 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 multi-point, unidimensional, standard scale — on which respondents indicated their opinion on a set of positions on a scale from ‘strongly disagree’ to ‘strongly agree’ or ‘strongly disapprove’ to ‘strongly approve’ — became and is still the most common tool of survey research in the 21st century.

---

<sup>1</sup> von Kohorn: Collective Decision Engines, LLC, [dougvk@aya.yale.edu](mailto:dougvk@aya.yale.edu); Sutherland: OgilvyOne, [rory.sutherland@ogilvy.com](mailto:rory.sutherland@ogilvy.com); Slavin: Media Lab, Massachusetts Institute of Technology [slavin@media.mit.edu](mailto:slavin@media.mit.edu); Konar: [Stanford Center on Longevity](http://Stanford Center on Longevity), [ellen.konar@gmail.com](mailto:ellen.konar@gmail.com); Quarfoot: Center for Research in Mathematics and Science Education, San Diego State University, [quarfoot@gmail.com](mailto:quarfoot@gmail.com)

Research on the Likert approach has established that it satisfies many criteria of reliability and validity. Attitudes on policy issues as measured, especially those that are important, 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 response predicts their choices in the voting booth (e.g., Anand & Krosnick, 2003; Fournier, Blais, Nadeau, Gidengil, & Nevitte, 2003; Visser, Krosnick, & Simmons, 2003).

Despite these 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 only those issues of significant importance. Yet, this constraint is rarely imposed, and respondents are asked to indicate opinions on issues well beyond those of personal significant importance to them. 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, Presser, Fealing & Ruggles, 2015). Krosnick, Visser, and Harder (2009) touch on a more fundamental issue of relative strength: “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 & Krosnick, 2003; Krosnick, 1990)”. The two approaches commonly used to garner relative assessments of options, rank order questions and conjoint studies, have raised the burden on survey designers and survey takers without compelling evidence of better results.

In this paper, we evaluate a new approach based on quadratic voting (QV). In the original formulation due to Weyl (2013) (see also Lalley & Weyl 2015), 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 that bike lanes be built. A resident who favors the proposal could buy one vote (for \$1), two votes (for \$4), three votes (for \$9), etc. Similarly, a resident who opposes the proposal could buy votes against it. The proposal passes if a majority of votes favor it. The money is returned to residents pro rata or used to fund the project. This system gives residents strong incentives to truthfully reveal their preferences while ensuring (with increasing and very high degree of probability as the number of residents increase) that the project maximizes the public good.

However, the previous 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 for it in other contexts. The major concern in market research is understanding the relationship between various groups (usually referred to as segments or subgroups) and items of interest, such as opinions on political issues, features of products/services in development, and brand attributes. 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 product features or policy proposals. If the theory of QV is correct, then respondents should be given strong incentives to truthfully reveal their preferences.

It is valuable to note a key difference between the Likert and QV approaches. At its heart, Likert is a *methodology of abundance* – voters may cast whatever votes they like, and no vote affects any other vote. As in the case of abundant resources, careful tradeoffs are not necessary because there is no incentive to adopt measured views. In contrast, QV is a *methodology of scarcity* – voters are

constrained by their credit banks, and this forces voting tradeoffs across proposals. This difference is critical, for as Mullainathan and Shafir 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” (Mullainathan & Shafir 2013, p. 7). 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” (Mullainathan & Shafir 2013, p.12). In this paper, we investigate whether QV successfully engenders a scarcity-mindset relative to Likert and document the value of this accomplishment as well.

## **METHODS**

### ***Participants***

Participants for the study were provided by Toluna, one of the largest market research sample providers in the U.S. A sample of 4850 individuals were drawn to reflect the age and gender balance of the U.S. population. Fieldwork was conducted February 5-15, 2016. Toluna panelists are compensated in “Toluna points” for completed surveys, which they can redeem for various prizes and items. Participants who successfully completed a “screener” section which affirmed they were 18 years or older and registered to vote in the US were directed to a survey link.

### ***Experimental Design***

Participants were randomly assigned to one of three conditions: Likert (referred to as Likert-only), QV (QV-only), or Likert and then QV (Likert-QV). The Fischer-Yates shuffle algorithm was implemented to be robust against floating point errors common in the Javascript programming language.

All three conditions contained the same set of proposals, presented in a randomized order:

- 1) Requirement for background checks for all gun purchasers.
- 2) Elimination of the Affordable Care Act of 2010 (aka ‘Obamacare’)
- 3) Rise in 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 US
- 10) Legally require employers to generally pay women an equal salary as men receive for the same work

In the Likert-only condition, respondents were given a survey that presented the ten proposals listed above presented in randomized order. Each proposal afforded the respondent the ability to select one of seven possible response choices consistent with the recommendations for 7 point bipolar ranges based on a review of research by Krosnick and Fabrigar (2012): “Very strongly against”; “Strongly against”; “Somewhat against”; “Neutral”; “Somewhat in favor”; “Strongly in favor”; “Very strongly in favor”. The system enabled respondents to navigate between questions and change their answers, and the ability to skip questions.

The QV-only participants viewed a 90 second introduction to the mechanics of the QV software, as well as a description of a budget and a preview of the number of proposals on which the respondent was about to vote. The QV survey itself presented the 10 proposals in randomized order. Respondents were to allocate 100 credits across these 10 proposals with the quadratic pricing mechanism active. So, for example, if a respondent decided to purchase 4 votes in favor of the first proposal, 16 credits would be deducted from their budget, thus giving the respondent 84 credits to allocate across the remaining proposals. The user interface design allowed for the respondents to scroll up and down through the proposals as they wished, particularly if they needed to adjust an earlier vote to free up credits to use elsewhere in the survey.

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

Respondents in all conditions were then asked for basic demographic information, survey feedback, and a multiple-choice followup question. The multiple choice question asked respondents, "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 response to the surveys, the software captured metadata to reflect participant actions during the course of survey completion. This included actions such as: movement among and between survey sections; time spent in each section; time between actions.

### ***Survey Software***

The most important change that took place between this study and all past work done on quadratic voting was the development of proprietary software called "weDesign" based on quadratic voting. It is capable of capturing both the data from respondents' 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 survey platform as QV in order to ensure uniformity in how we capture data and metadata. We also added a further improvement of allowing quadratic vote allocation to take place across several different items at once, rather than a single-issue approach that much of the past QV literature has focused on. Also, as the obstacles to having respondents commit their own personal resources to the test appeared too high in the near term, respondents were instead allotted a fixed budget of virtual "credits" to spend on individual proposals. As discussed earlier, respondents were then allowed to allocate votes either for or against each proposal at the quadratic cost dictated by QV theory. An overall budget of 100 credits was placed on each respondent in order to allow for maximal use of all credits on a single issue if the respondent desired.

### ***Completion Rate***

From the 4850 panelists who qualified, we attained 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 and should impose little to no burden. QV,

which demands trade offs across items, might increase the cognitive load and require more time. And, as a new methodology, we added to the time required by including an introductory video to explain the approach and make clear the nature of the process. Yet, despite the greater demands and unfamiliarity of 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 negatively impact representativeness of participants and if streamlined, could even provide some added value here.

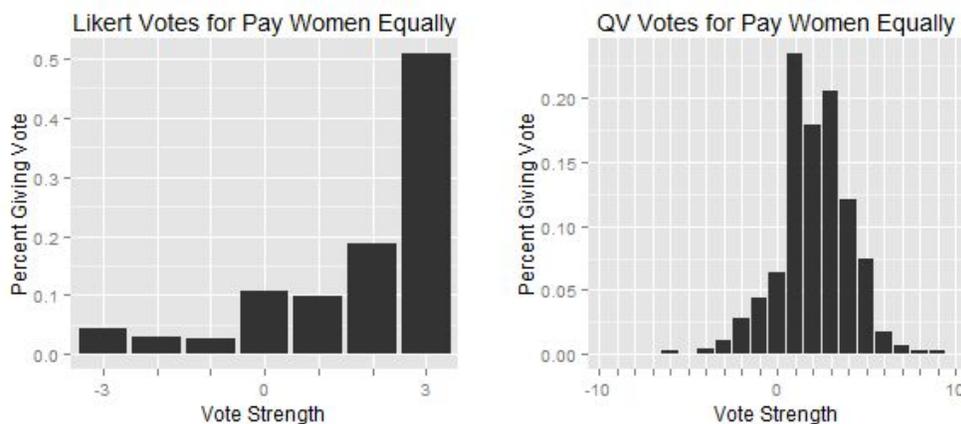
Furthermore, we checked the demographic characteristics of the respondents in each condition and found no statistically significant differences. It is thus reasonable to conclude that abandonment was demographically consistent across test conditions. As such, differences seen in voter behavior can be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

## RESULTS

This section outlines results from the study grouped into 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 on the predictive power of both 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.

### *Changing the Group Landscape*

The most powerful macro effect of QV, relative to Likert, is one of moderation. Because QV voters can no longer use the extreme ends of the scale without consequence, as they can in Likert, they cast fewer votes at the ends of the voting scale. 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 the numerical range  $\{-3, \dots, 3\}$ .



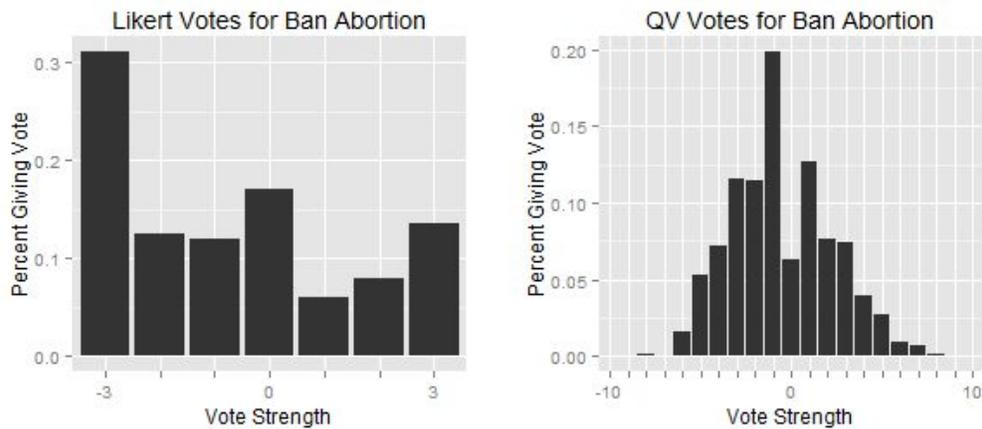


Figure 1. Voting distributions on two proposals (paying women equally and banning abortion) in the Likert-only (n=1095) and QV-only (n=1048) test conditions.

As the above figure suggests, what appear to be strong ideological differences, as seen through skewed and multi-modal graphs, tend toward more balanced opinions (symmetric, quasi-normal distributions). These claims can be made more rigorous by exploring the skew (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 have a skew of 0, and that normal distributions have an excess kurtosis (defined as kurtosis - 3) of 0.

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

Figure 2. Skew and excess kurtosis for each proposal and methodology. Skew values near 0 suggest symmetric graphs; excess kurtosis values near 0 suggest tails that mimic a normal distribution.

In all cases but one (proposal 2), the skew of the QV results has decreased, suggesting better symmetry around the mean. When this is coupled with excess kurtosis values getting closer to 0, this suggests that the QV distributions are closer to normal distributions than the Likert distributions are. It is particularly valuable to note those QV distributions with excess kurtosis values far from 0. Proposals 1 and 10 have larger positive kurtosis values (leptokurtic distributions) suggesting a higher degree of consensus on these issues. Indeed, once voting distributions become quasi-normal, the kurtosis value can be used as a quick summary of the degree of consensus around an issue.

### ***Changing the Individual Landscape***

The above section suggests that at a high level, QV changes “what we decide” (using Mullainathan & 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 that show the average experience of voters with each methodology:

<b>Methodology</b>	<b>Likert-only</b>	<b>QV-only</b>	<b>Likert results from Likert-QV</b>	<b>QV results from Likert-QV</b>
<b>Time voting (seconds)</b>	102.2	132.4	100.31	114.3
<b>Voting actions</b>	11.01	15	10.98	15.2
<b>Revisions</b>	1.01	5	0.98	5.2
<b>Revisions stronger</b>	.39	3.5	.34	3.6
<b>Revisions weaker</b>	0.33	1.3	0.29	1.5
<b>Revisions hold</b>	0.29	0.2	0.34	0.2

*Figure 3.* Average time spent voting, number of voting actions, and number and type of revisions in Likert and QV surveys for the single methodology and joint methodology (Likert-QV) conditions.

The above table reveals that QV voters tend to spend about 30 more seconds voting (or 29% longer) when they have not seen the proposals (single methodology groups), and about 14 seconds longer (or 14% longer) when they have seen the proposals and previously thought about them (Likert-QV condition). In addition, QV voters make more voting actions overall (15 compared to 11) because they are making more revisions. In Likert, a revision is defined as selecting a vote intensity on a question for which the user has already selected a vote intensity (this is done by selecting a radio button). In QV, a revision occurs 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 two seconds of time between the end of the first vote cast and the start of the 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 5 total revisions (combining the three types), while Likert voters make only 1.

These revision statistics get to the heart of the QV experience: When resources are scarce, voters cannot give extreme answers to all proposals, and so voters must face the reality of giving measured opinions that reflect the tradeoffs of life. Interestingly, these behavioral differences can be seen “in real time” by plotting a voter’s total cast votes over time. More specifically, for a given voter, we plot the sum of the absolute values of the votes versus the time percentage into the voter’s survey. Below are two graphs of this type for each methodology. To avoid plotting the voting trajectory of 1000 voters per methodology, we sort all the voters based on number of revisions made, break them into 50 quintiles (or about 20 voters each), and then plot the data for voters in quintiles 10 and 40 (or the 20% and 80% most revision-heavy users). 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 are revisions.

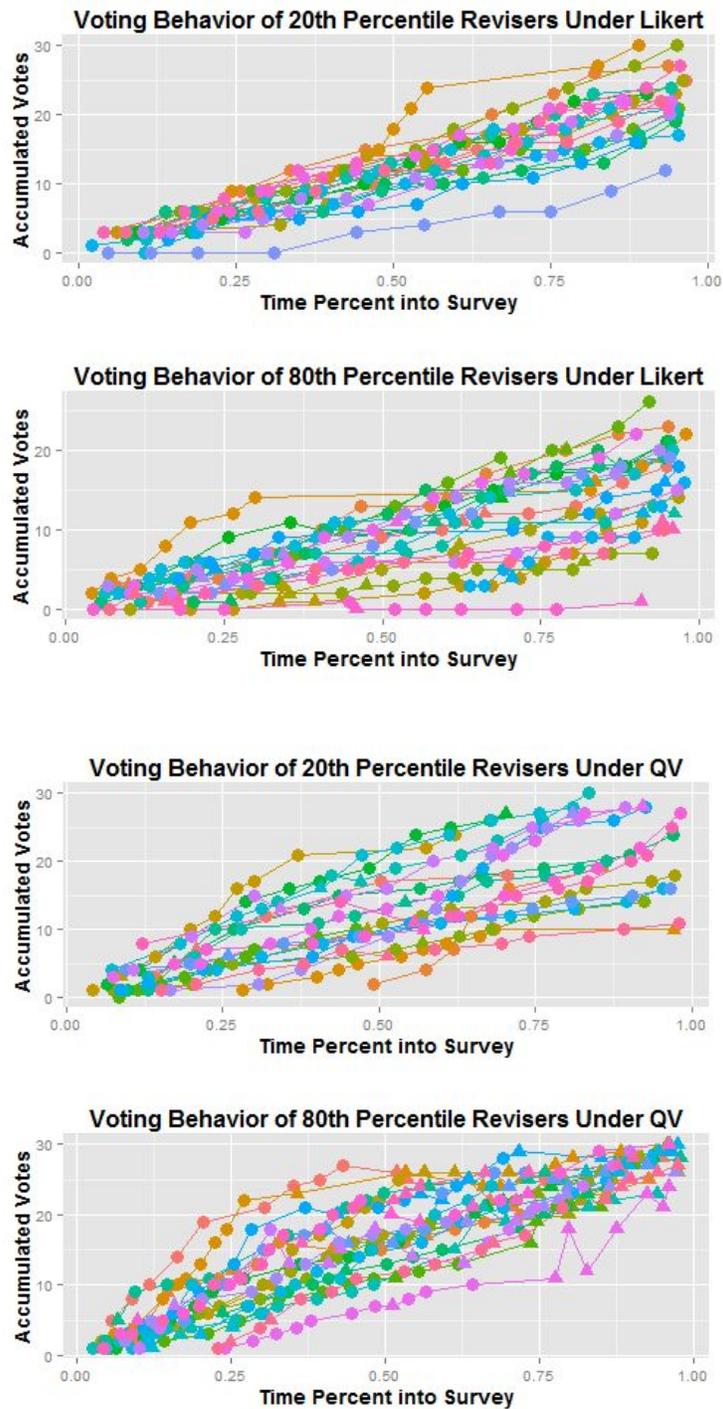


Figure 4. Vote accumulation graphs versus time into survey for Likert and QV methodologies using the Likert-only and QV-only conditions. A single voter's behavior is encoded in a jagged, colored line. These graphs show the behavior of voters around the 20<sup>th</sup> and 80<sup>th</sup> percentiles of 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.

The above images demonstrate that Likert voters in the 20<sup>th</sup> percentile of total revisions made (TRM) tend to make few revisions, and as such, their vote accumulation graphs are monotonically increasing. By the 80<sup>th</sup> percentile of TRM, Likert voters begin to make some revisions, but as before, most graphs are monotonically increasing. In contrast, the 20<sup>th</sup> percentile of TRM for the QV methodology looks quite similar to the 80<sup>th</sup> percentile of Likert: there are revisions on occasion (triangle icons), and some graphs begin to change direction. By the 80<sup>th</sup> percentile, more directional changes are present and the total number of voting actions has visibly increased. In addition, voters tend to end around the same total number of votes (in contrast with the other three graphs): voters are coaxing the most votes possible out of their limited credit banks.

For each voter and each methodology, we might explore what percentage of proposals the voter revisited. As the below table suggests, the average Likert user revises about 1 in every 12 proposals they encounter (about 8%). In contrast, the average QV voter revisits about 1 in every 3 proposals (about 33.3%).

<b>Data Source</b>	<b>% of proposals that voters revised, on average</b>
Likert-only	8.44
QV-only	34.69
Likert results from Likert-QV	8.14
QV results from Likert-QV	35.37

*Figure 5.* The percentage of proposals which are voted on and later revised (on average) by voters using a given methodology.

These findings are further developed in the below figure, which shows the distribution of proposal revision counts for the two methodologies in both the single (Likert-only vs. QV-only) and dual methodology (Likert-QV) conditions.

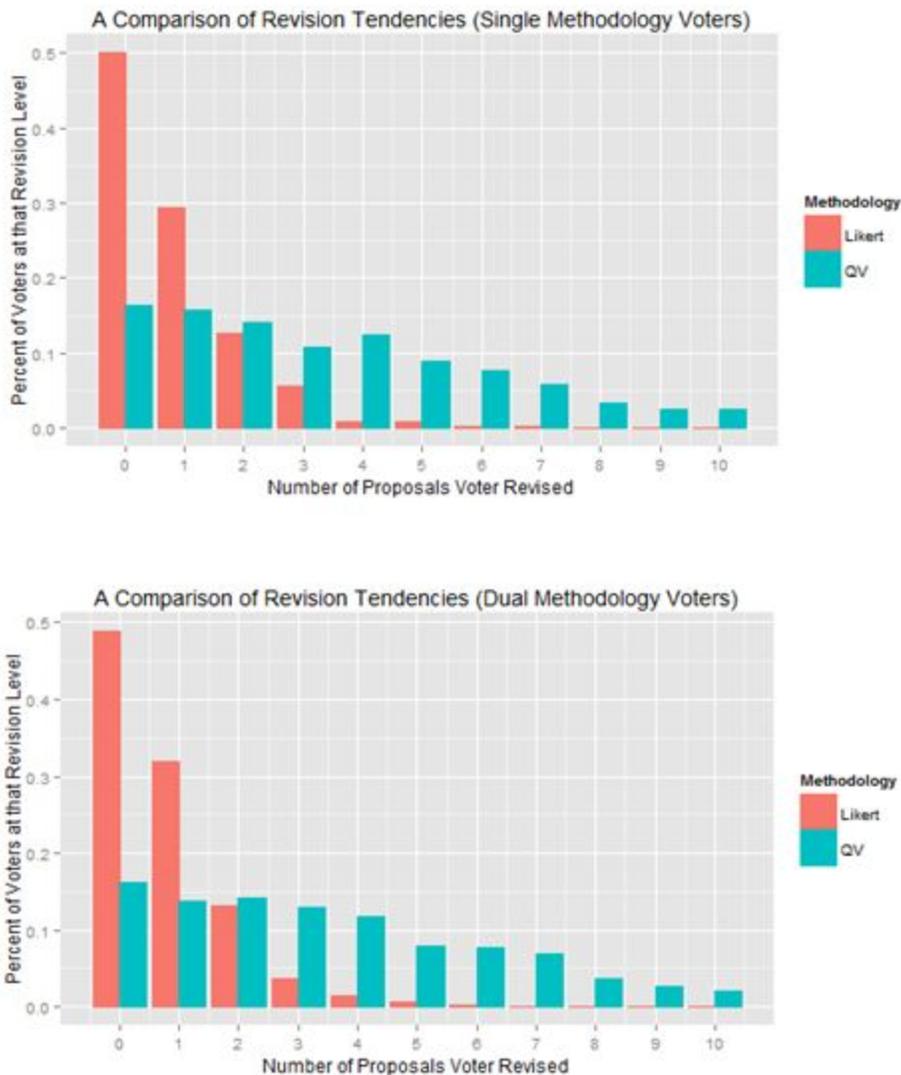


Figure 6. Distribution of proposal revision counts for Likert vs. QV methodologies in both the single and dual survey conditions. The means of these distributions are the results in Figure 5.

Figure 6 suggests that almost half of all Likert voters never revise a single proposal, and that it is rare to ever revise more than three. In contrast, QV voters often revise more than three proposals (Figure 5 shows the mean is between 3 and 4 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 these differences are inherent to the methodologies, not the voting samples.

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

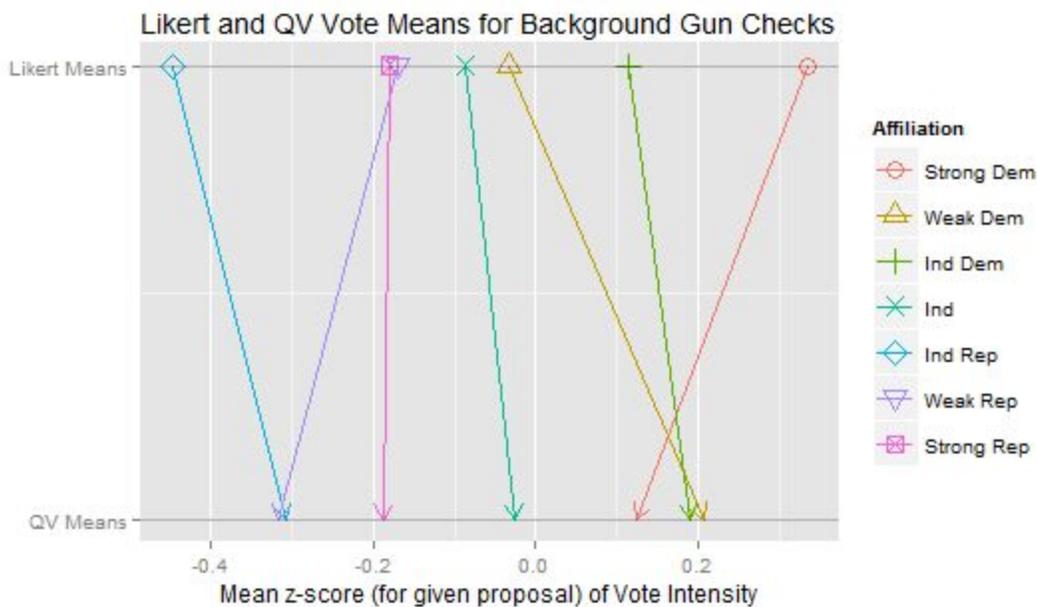
**Insights on Predictive Power**

Two critical questions remain for Likert and QV: 1) How do Likert and QV relate to known qualitative trends in polling, and 2) How do the results from Likert and QV studies differ in helping researchers

understand voters and predict their behavior? This second question is of particular importance, for there can be a divide between voters' espoused views and the actions they are willing to take based on those claims.

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 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 is born from analysis of actual voters. As Sides 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" (Sides, 2014).

Naturally, one might ask whether a voting methodology supports this clustering practice. To explore this, we take a given methodology (e.g. Likert) and a given proposal (e.g. Increasing the minimum wage), and turn all votes on this issue into z-scores. Then, we group together votes for the same affiliation (e.g. Strong Democrat) and take the mean. Repeating this for the other affiliations and the other methodology, we can see how affiliation means differ in Likert versus QV. Given that z-scores are used, inherent scale differences are negated, and we get an accurate read of the relative placement of these seven political affiliations on a given issue.



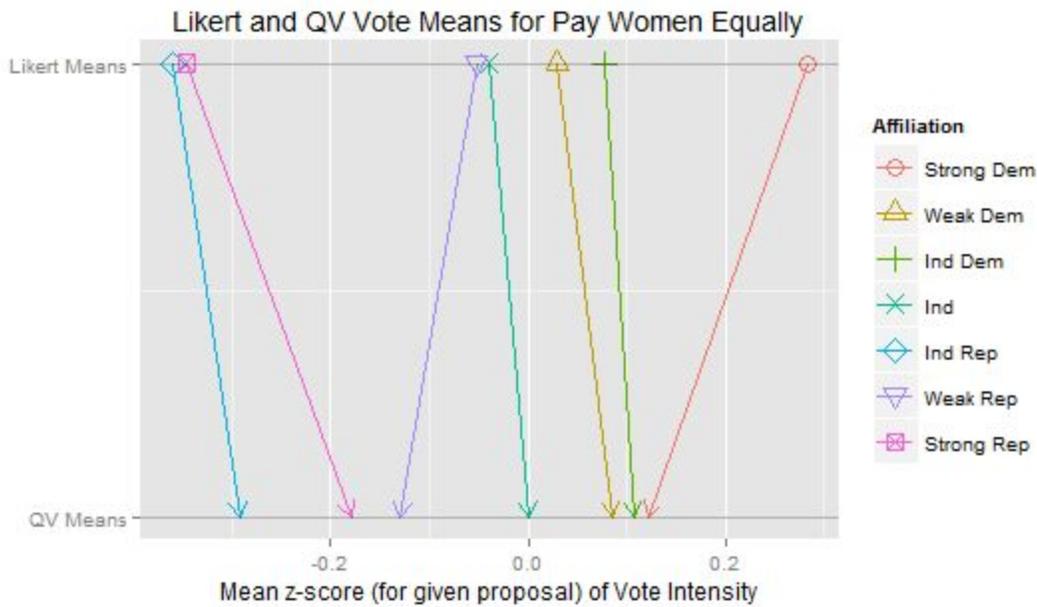


Figure 7. Examples of two proposals for which QV shows greater support for the three-party clustering approach. Results are similar for the Likert-QV condition. Note that these results show relative movement (in relation to the mean), not absolute: that is, there is no indication of whether these affiliations are voting for or against the above proposals.

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 separate groups under Likert, they move to common ground under QV. As such, QV suggests that the attitudes of these groups are closer than Likert does. A similar, but weaker, trend is seen on the Republican side. The second graph, concerning equal pay for women, shows similar trends: the three Democratic-leaning affiliations come together, as do the three Republican-leaning affiliations. 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 better understand voters and the possibility of action. To do this, we first label a set of voting levels in each methodology as “moderate” or “strong” votes. The below tables show the percentage of votes given in certain ranges for the two methodologies (Likert-only and QV-only conditions).

Vote Range	{-10, ..., -3}	{-2, ..., 2}	{3, ..., 10}
Likert %	15	58	27
QV %	13	59	28

Figure 8. Percent of votes in ranges for both Likert and QV using the Likert-only and QV-only conditions. For Likert, the range  $\{-10, \dots, -3\}$  simply means  $-3$ , and  $\{3, \dots, 10\}$  means  $3$ .

As this table suggests, it is reasonable to consider the QV vote levels with absolute value 3 or greater as “strong” because these closely map (in percentage) to the labels “Strongly Disagree” and “Strongly Agree” in Likert. This definition of “strong” vote suggests one advantage of QV: while Likert collects strong votes under 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 something in the range  $\{-2, \dots, 2\}$  in either methodology. We argue that one of the greatest benefits of QV is its ability to make sense of the “strong” vote (given about 42% of the time) and the voters that give these votes.

We turn first to voters giving strong votes. The below table shows the percentage of voters that give strong votes to at least  $b$  proposals, where  $b$  takes on the values 6 through 10. Note for  $b = 6$  (and 7-10), this means the voter is giving a strong vote on the majority of proposals. These data are built on the Likert-QV condition, so observed differences are inherent to the methodologies.

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

Figure 9. Percent of voters giving strong votes to at least  $b$  proposals in the Likert-QV condition.

The above two tables indicate that while the total number of “strong” votes is roughly the same across methodologies, the percent of voters who *frequently use* these is reduced in QV. This suggests that the information we get from a voter using QV will be more informative than when using Likert. For example, Figure 10 shows the voting profile of two “extreme” voters. Here, an “extreme” voter is defined as one who gives a strong vote on 8 or more of the 10 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.

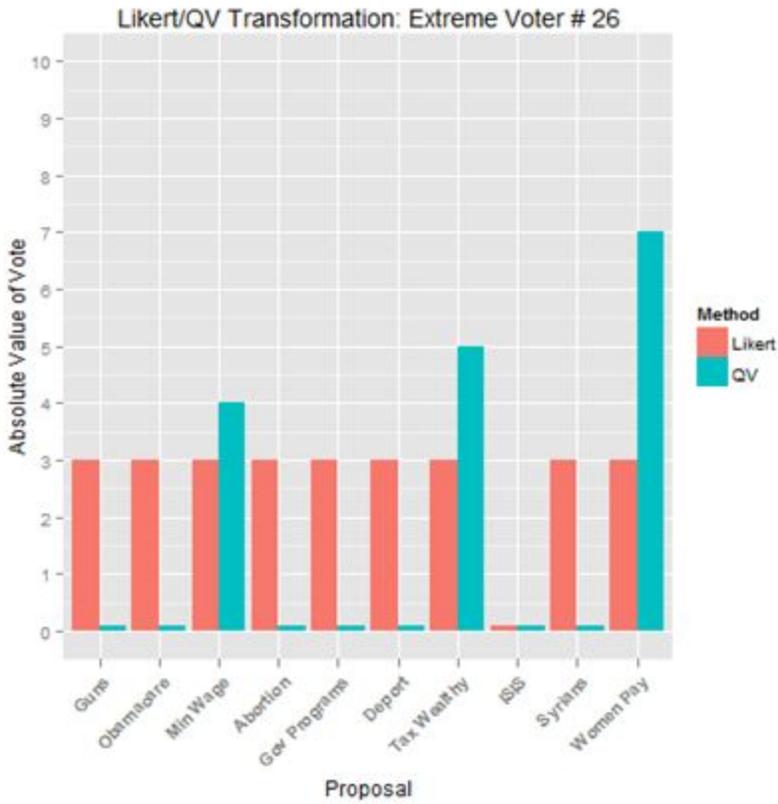
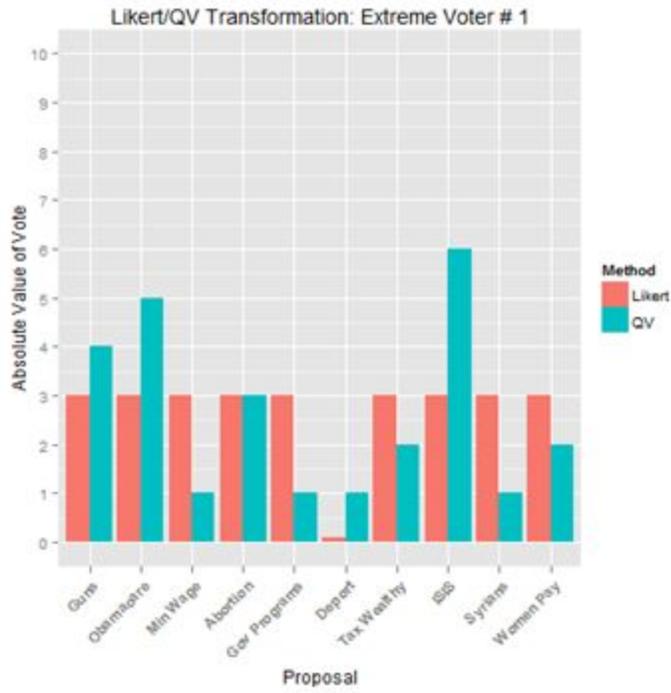


Figure 10. An exploration of two “extreme” voters from the Likert-QV condition. Here, “extreme” is defined giving 8 or more proposals a strong vote in Likert. Under QV, voting profiles are more nuanced. Absolute value of votes are used to simplify the presentation.

As the first graph above suggests, the Likert data for Extreme Voter # 1 is of little value in deciding what matters to this voter: apparently everything but the deportation of illegal immigrants is of the highest degree of importance. Under QV, one sees that guns, Obamacare, abortion, and ISIS carry more weight – with issues like minimum wage and government programs taking a back seat. The second graph, Extreme Voter # 26, shows a respondent who cares about all issues but ISIS under Likert. Under QV, however, we see that only minimum wage, taxation, and equal pay for women even receive votes.

This behavior change is not unique to the extreme voters seen above. The below figure is an animated GIF that shows the behavior of 42 of these extreme voters. When viewed in electronic format, it gives a broad sense of how extreme voters react when faced with the reality of scarcity. When viewed non-electronically, this image will appear as Extreme Voter # 1 seen above. Readers are encouraged to view this graph electronically.

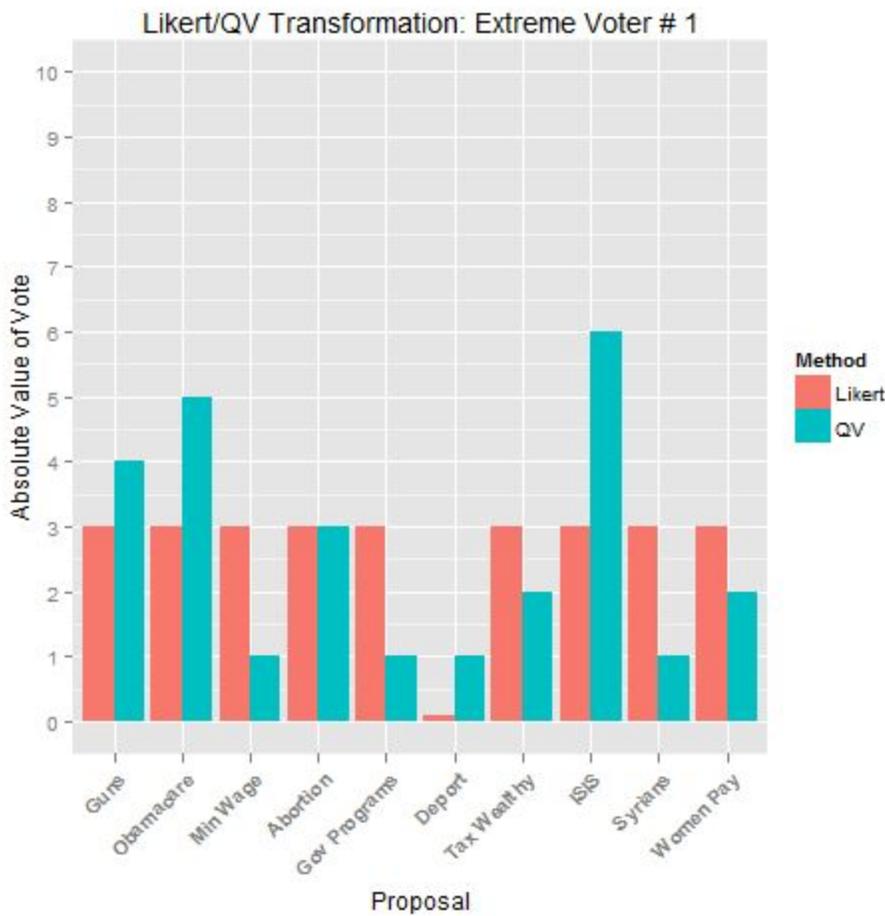


Figure 11. An animated GIF showing the voting profile transformation of 42 extreme voters in the Likert-QV condition.

To develop in greater detail how QV brings meaning to the strong vote, we explore the relationship of vote strength and action. Recall that at the end of each session, respondents were asked a question which called for a commitment with apparent consequence outside the test. They were asked which of the issues they would be interested in getting more information for how they might take action or influence public policy and allowed to check all that apply. While these checks are not a perfect measure of all actions voters would take at the polls, on social media, or in grassroots campaigning, they do provide a window into a world beyond voters’ stated preferences.

We note at the outset that the use of the QV methodology does not cause voters to check boxes with any greater frequency. Indeed, the below table shows the percent of box-checking (that is, request for more information on how to take action) is nearly identical for negative “strong” votes, “moderate” votes, and positive “strong” votes:

<b>Vote Intensity</b>	<b>Strong negative</b>	<b>Moderate</b>	<b>Strong Positive</b>
<b>% boxes checked under Likert</b>	30.3	23.5	42.3
<b>% boxes checked under QV</b>	30.4	23.4	42.8

Figure 12. Percent asking for more information by intensity of Likert and QV voting in the Likert-only and QV-only conditions.

While these percentages are essentially identical, they hide a stark difference between the Likert and QV methodologies. Said simply, the “strong” vote in Likert aggregates a wide range of belief – from the voter who espouses strong support but caves under the pressure of scarcity, to the voter who redoubles his/her vote intensity when given a credit 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 below, the “strongest” voters in Likert are likely to behave with great interest: 42% want more information. In contrast, QV disaggregates these voters, finding those that are less likely to act (the blue, QV, vote of 3) despite giving a “strong” vote, and those whose votes speak quite powerfully to their probability of taking action (the blue, QV, votes of 4, 5, and 6+). In essence, QV breaks the red Likert line into multiple pieces, sorting voters into groups that better reveal their true action-based preferences (both weaker and strong).

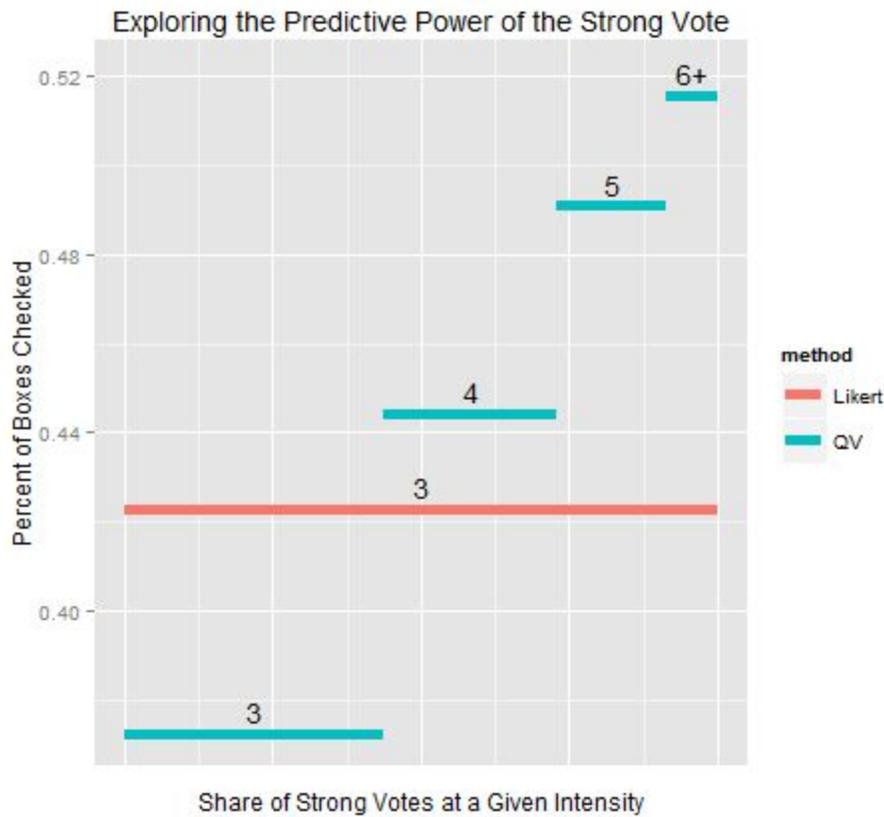


Figure 13. The probability a voter will ask for information on an issue when giving a certain “strong” positive vote in Likert and QV. The width of each segment shows what fraction of “strong” votes are at the stated intensity. QV levels 6-10 have been grouped to ensure a sample size to produce a stable estimate. Similar results occur with negative “strong” votes.

## DISCUSSION

This study shows that QV offers a promising alternative to Likert. The results demonstrate that QV achieves a standard data distribution, which Likert has never done; engages a higher level of thoughtfulness; attenuates extremism to expose deeper insight; and better predicts behavior than Likert at high levels of vote strength.

Perhaps the most powerful macro effect of QV is one of moderation. Because voters can no longer use the extreme ends of the scale without consequence, 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 down to just one or two key issues, while QV allows for a greater diversity of expression (Louviere & Islam 2008).

In terms of engagement, the “total revisions made” (TRM) results and the revisions on proposals suggest that QV engenders greater thoughtfulness among its users. Indeed, in qualitative UX testing and in the comments section available at the end of the survey, QV voters often cite two main reasons for making revisions. First, they note that this 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 isolated intensity levels, and not something the voter has considered before. Second, while voters are tempering and ordering their preferences, they still want to show these preferences as strongly as possible. Thus, many voters work to use every last credit in expressing their opinions (hence, the wealth of revision triangles in the upper right corner of the 80<sup>th</sup> percentile QV plot of Figure 4). These two forces – preference realignment and voting profile maximization – are examples of what Mullainathan & Shafir (2013) refer to as the “focus dividend”: “the positive outcome of scarcity capturing the mind” (Mullainathan & Shafir 2013, p. 24).

Taken together, the results on revision tendencies and the accumulating vote graphs suggest some of the changes that are going on at the individual voter levels. 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 these changes. Furthermore, the accumulating vote graphs suggest that the 80<sup>th</sup> percentile of Likert voting is just beginning to look like the 20<sup>th</sup> percentile of QV voting, a possible consequence of the focus dividend inducing greater thoughtfulness in the presence of credit scarcity.

This aspect of 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 through phone interviews or face-to-face interviews, the only person setting the speed of the interview is the respondent. Thus QV’s demonstrated ability to engage respondents in such a way that the respondent is willing to return to questions already answered and change their responses seems a strong step in addressing this aspect of satisficing.

In the study, greater engagement and reduced satisficing show great promise for understanding the populations survey samples are taken from. 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 issues the seven separate 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 draw out their opinions. Strong ideological divides in one methodology may dissolve in a different methodology.

Not only does QV disincentivize extreme ratings/voting, it tempers ratings such that they are more meaningful and predictive of relevant and even costly behavior. QV imbues the notion of a “strong” vote with greater granularity, disaggregating those who claim to care about an issue but are unlikely to act, from those who claim to care and are likely to act. Furthermore, these findings are not simply the result of arbitrary scale size differences between these instances of Likert and QV. 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 between 5–7 points were more reliable than scales with fewer or more points. Seven points were found to be optimal for bipolar scales such as those used here. As such, QV (with little optimization) outperforms Likert in its near-optimal configuration.

## **FUTURE CONSIDERATIONS**

Despite this demonstrated added value of QV, there are clear challenges ahead. For example, the histograms of QV data show an unexpected dip at vote intensity zero. An initial hypothesis to account for this dip at zero is that given the quadratic pricing mechanism we implemented, respondents are left with unused credits and thus seek to redistribute their remaining credits in such a way that allows them to get to zero credits left: namely, purchasing one vote for issues on which they had originally wanted to remain neutral (the zero vote level). It is unclear if purchasing this vote reflects 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 they hold a belief on an issue that they truly do not have — an example of what Mullianathan and Shafir (2013) call “tunneling”. Future testing should address this. One path would be to move from discrete to continuous vote allocation. This would allow respondents to purchase tenths or even hundredths of votes. Another would be to “pay” 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 to understand and optimize the parameters available within the QV framework. For example, experimentation should take place with the number of items on the survey, types of issues included, and more extensive qualitative inquiry into respondents’ interactions with QV content.

While the above analysis of QV in understanding public opinion empirically validates its potential, fully realizing such potential will likely require at least two avenues of additional experimentation:

- 1) Demonstrating or evaluating the 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 in the future. As part of the survey, we garnered user ratings and requested feedback on the survey experience. Respondents rated the Likert-only experience 8.6 out of 10 on average (10 being the most positive rating) and the QV-only experience a 7.9. Though the difference is small, it is non-trivial. A basic sentiment analysis of the comments left by the portion of the sample who left 10 words or more in feedback (260 total respondents) showed that “highly positive” comments outrated “highly negative” comments 3:1. Among those “highly negative” comments, some respondents seemed to struggle with the limitation imposed by the quadratic pricing mechanism and felt they were unable to express how they truly felt. This 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 affordances and constraints 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.
- Krosnick, J. A. (1988). Attitude importance and attitude change. *Journal of Experimental Social Psychology*, 24(3), 240–255.
- Krosnick, J. A., Visser, P. S., & Harder, J. (2009). The Psychological Underpinnings of Political Behavior. *Handbook of Social Psychology*.
- Krosnick, J. A., & Fabrigar, L. R. (2012). Designing Rating Scales for Effective Measurement in Surveys. *Lyberg/Survey Survey Measurement and Process Quality*, 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](https://www.nsf.gov/sbe/ac_materials/the_future_of_survey_research.pdf). Accessed 21 March 2016
- Lalley, S. P., & Weyl, E. G. (2015). Quadratic Voting. *SSRN Electronic Journal SSRN Journal*.
- 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*.
- Sides, J. (2014, January 8). 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
- 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. (2013). Quadratic Vote Buying. *SSRN Electronic Journal SSRN Journal*.