

Who Cares? Measuring Preference Intensity in a Polarized Environment

October 2, 2020

Among all the policy issues voters have preferences over, some are more likely than others to matter for people's thoughts, intentions and behavior. Such differences in preference intensity are rarely measured in surveys. We test a new survey tool — Quadratic Voting for Survey Research (QVSR) — designed to address this methodological gap. QVSR gives respondents a fixed budget to 'buy' votes in favor of (against) a set of policy proposals. Because the price for each vote is quadratic, expressing support for (opposition to) a given proposal becomes increasingly costly. In a polarized political context, we expect QVSR to better measure preference intensity relative to existing ordinal survey tools (e.g. Likert and issue importance items). To test this expectation, we ask participants to take the same survey, varying only the survey technology. We find that QVSR better differentiates between intense and weak preferences, as proxied by respondents' behavior on simplified real-world tasks.

Word Count: 10,581

Keywords: Survey research, Public opinion, Preference intensity, Attitude strength, Likert item, Issue importance, Quadratic voting for survey research.

Many research endeavors in political science require knowing not only what people want, i.e. their preference¹ orientation, but also how much they want it, i.e. their preference intensity. For example, political theorists ponder over the weight to give to a policy intensely preferred by a minority and only weakly opposed by a majority. Students of political change point out that while voters hold preferences on more than one policy, only intense policy preferences, referred to as ‘salient’ in this literature,² will decisively influence political behavior.

Despite its relevance, preference intensity is poorly measured in surveys. Most studies rely on a version of the Likert item, a survey technology developed in the 1930s by the psychologist Rensis Likert. Survey respondents are asked to evaluate a policy statement by picking one of 5 (or 7) ordered responses ranging from ‘strongly agree (favor)’ to ‘strongly disagree (oppose).’ Answers collected using this survey tool capture a mix of preference orientation and preference intensity without a way of disentangling one from the other. Likert items raise two additional concerns. First, respondents who answer a battery of these items have only limited incentives to consider trade-offs across policy issues (the abundance problem).³ In other words, the data collected provides limited information on the issues people might prioritize when faced with a forced choice (e.g. in the case of Brexit, immigration control at the expense of access to the single market). A second concern, most common in studies of contemporary American politics, is the distortion introduced by partisan motives, by which survey answers cluster on both extremities of the ordinal response scale (the bunching problem). The same response categories (e.g. *strongly* agree-favor) combine respondents who care about the issue and respondents who do not care as intensely and are merely “paying lip service to the party norm” (Zaller 2012). In rare instances, a Likert item will be followed by a question asking how important a given issue is to the respondent. Responses are collected using a similar categorical scale ranging from ‘not at all important’ (1) to ‘very important’ (5) (Miller and

¹ In this paper, we define ‘preferences’ as evaluations of statements, in this case policy-relevant statements, ranging from positive to negative. We also distinguish between ‘preference orientation’ and ‘preference intensity.’ Our definitions are in broad agreement with similar concepts in social psychology such as ‘attitude orientation’ and ‘attitude strength.’

² Through the remainder of the paper, we avoid the concept of salience as its definition in political science is somewhat vague, describing in some instances something akin to preference intensity as defined in this paper, and in other instances, something that is “on top of voters’ minds” because of the issues elites decide to politicize over others. A salient issue according to the latter, need not be salient according to the former.

³ Such trade-offs can be due to the incompatibility of options (one cannot support more government spending and tax cuts), resource scarcity (politicians cannot implement all policies), or the specific features of the choice space (one’s favorite policy bundle might not be offered by any existing candidate or party).

Peterson 2004; Howe and Krosnick 2017). This survey technique improves measures of preference intensity by explicitly asking respondents how much they care. Yet, it does not explicitly address the abundance and bunching problems. Our goal in this paper is to improve the way we routinely measure preference intensity. Solutions to the abundance problem, such as issue-ranking, have proven difficult to implement⁴ and, to the best of our knowledge, no solutions to the bunching problems have been suggested.

As with any measurement exercise, successfully measuring differences in preference intensity requires answering three questions. First, what is preference intensity conceptually? Second, which survey technology should best measure preference intensity in theory? Third, which survey technology best measures preference intensity in practice? In response to the first question, we follow common practice in political economy and conceptualize preference intensity as the willingness to incur the costs of aligning one's political behavior with one's preference orientation. In other words, the more intense the preference, the more an individual is likely to act to move the status quo closer to her preferred policy outcome.

To answer the second question, we propose a simple model of survey answer and use it to compare different survey technologies. In particular, we study a new way of measuring preference intensity called Quadratic Voting for Survey Research (QVSR). This method mimics real world trade-offs by asking respondents to vote on a bundle of issues under conditions of scarcity: respondents are constrained by a fixed budget with which to 'buy' the votes. Because the price for each vote is quadratic, it becomes increasingly costly to acquire additional votes to express more intense support for (or opposition to) the same issue. The budget constraint and quadratic pricing compel respondents to arbitrate between the issues in the bundle and make it costly to express intense preferences by voting repeatedly for the same issue (Lalley and Weyl 2018). This set up contrasts sharply with Likert items' world of abundance in which respondents face no trade-offs and can pick end-of-scale responses (e.g. strongly favor/oppose) at no cost. With QVSR, respondents faced with scarcity are forced to 'de-bunch.' Under the assumption that the psychological cost of misreporting one's preference depends positively on preference intensity, we show that respondents are expected

⁴ In a recent contribution, Hanretty, Lauderdale and Vivyan (2020) use a modified version of conjoint analysis to identify the issues voters care the most about. While very helpful, this technique was not designed to provide an individual-level measure of preference intensity.

to prioritize issues they care about the most at the expense of issues they care about the least. We therefore expect the additional variance (i.e. higher entropy) obtained using QVSR in place of Likert-style items to be more informative of differences in preference intensity.

Finally, to answer the third question (whether or not QVSR does indeed better measure preference intensity), we randomly assign individuals to take the same survey varying only the technology used to measure policy preferences. Alongside Likert items and QVSR, we test a third survey tool which combines a Likert item with an issue importance item designed to more directly measure preference intensity by asking respondents whether an issue is personally important to them ('Likert+' for short). Likert+, because it also places respondents in a world of abundance and does not penalize partisan motives, should theoretically face some of the same limits as Likert item.

To examine which survey tool best measures preference intensity, we also need an independent and reliable measure of preference intensity. By our definition of preference intensity, engaging in a costly behavior related to a given policy issue indicates more intense preferences on this same issue. We consequently ask respondents to perform three behavioral tasks related to policy issues included in the survey (e.g. a donation to a non-profit advocating for gun control). Higher values on this behavioral outcome imply a more costly behavior (e.g. amount donated) and consequently more intense preferences. We then compare each tool's discrimination ability, i.e. its ability to distinguish between respondents whose behavior suggests intense preferences and those whose behavior suggests weak preferences. In addition, building on workhorse assumptions in political economy, we assume that people directly affected by a policy hold more intense preferences over this policy than individuals not affected by it. We capture the likelihood that one will benefit or be harmed by a given policy by collecting information on socio-economic factors that are correlated with exposure to a reform. We examine each tool's ability to distinguish between respondents most likely to be affected and those least likely to be affected.⁵

When compared to Likert, QVSR does better on all tests. When compared to Likert+, QVSR does better on a large majority of tests. More specifically in the latter case, while Likert+ sometimes

⁵ As described in more details below, we regress the behavioral outcomes on survey answers and compare effect sizes across survey methods, using normalized survey responses. A larger effect size indicates that a given method better discriminates between the highest and lowest intensity respondents (and anyone in between).

matches QVSR's ability to discriminate between high and low intensity respondents, it never outperforms QVSR. In light of our findings, and absent evidence that QVSR significantly increases survey costs (Quarfoot et al. 2017), we see relatively few downsides of switching to QVSR. Answers collected using this method have higher variance with better discriminating abilities, properties especially valuable for experimental studies seeking to unpack the micro-foundations of models of democratic politics (e.g. Kuziemko et al. (2015), Cruces, Perez-Truglia and Tetaz (2013)).

Measuring Preference Intensity

We propose a model that formalizes two related concerns: What exactly does a given survey question measure? And to what extent are responses to a survey question truthful? We build on this model to examine how answers to these two questions vary across survey tools. We start with Likert items, the standard tool used to measure policy preferences. We also examine two alternative survey tools. One is a Likert item followed by an issue importance item (Likert+). The other is the above-mentioned QVSR, which we present in more detail below. In the main text, we explain the model's main ingredients, intuitions, and results. The formal model is presented in detail in the appendix (see Appendix A.1). For expository purposes, we discuss each concern sequentially, but both are addressed simultaneously in the model.

Defining Key Concepts: Preference Orientation and Preference Intensity

Conceptually, when one considers a number of proposed policy changes, e.g. building a wall on the border between the U.S. and Mexico or legislating to give same sex couples the right to adopt a child, one may assume that an individual's policy preferences on these reforms are characterized by (at least) two main parameters: preference orientation and preference intensity.

Preference orientation describes whether the individual agrees or not with the policy proposal as described. Formally, one's preference orientation on each issue $k = 1, \dots, K$ can be described by a real number α_{ik} in the interval $[-1, 1]$, where $\alpha_{ik} = 1$ means full agreement with the policy, and $\alpha_{ik} = -1$ full disagreement with this policy. Intermediate values correspond to intermediate opinions,

something, for instance, that might be due to ambivalence (e.g. support for the policy's core principle but not specifically as stated in the survey question) or neutrality towards the issue.

Preference intensity describes how important the proposed policy change is to the individual. Formally, how much an individual cares about an issue, i.e. preference intensity, can be captured by a positive number β_{ik} in the interval $[0, 1]$, where $\beta_{ik} = 0$ means that this individual does not care about this issue, and $\beta_{ik} = 1$ means that the question is of the utmost importance to her. What does it mean for an individual to *care* about a given issue? In line with common practice in economics, our definition of preference intensity is inherently tied to observable behavior. Following Arrow (1951), we define preference intensity as the willingness to incur the costs of aligning one's political behavior with one's political preference orientation. By that we mean any behavior that moves the status quo closer to one's preference orientation on a given issue. When deciding whether or not to act on one's preferences, individuals face at least two types of costs, the direct personal cost of engaging in a given behavior (e.g. cost of donating money to a charity supporting a desired change, cost of turning out to vote, etc.) and the cost of acting on a given policy preference, which can be at the expense of one's preferences on a competing policy issue (see fn 3). In practice, β_{ik} incorporates both direct costs and trade-off costs.⁶

With these definitions in mind, we can examine the relationship between tool-specific survey answers, and α_{ik} and β_{ik} . For expository purposes, we focus on the variants of Likert, Likert+ and QVSR used in our empirical study.

Measuring Preference Orientation and Preference Intensity with Surveys

Likert Item Following best practices in survey design (Malhotra, Krosnick and Thomas 2009), we asked respondents assigned to the Likert tool about their opinion using a sequence of two branching questions: “Do you favor, oppose, or neither favor nor oppose: [Example] Giving same sex couples

⁶ This behavioral definition of preference intensity is closely related to what social psychologists Krosnick and Abelson (1992) discuss under the umbrella term of ‘attitude strength,’ which they define as the extent to which a given attitude “affects one’s cognition or behavior.” Krosnick and Abelson identify five sub-attributes of attitudes, all contributing to attitude strength. Our objective in this paper is not to challenge existing typologies. Instead, we simply aim for parsimony, and focus on two arguably central concepts in political economy: a preference for or against a status-quo-changing proposal, and the extent to which this preference affects relevant behavior.

the legal right to adopt a child?” Respondents who pick the favor or oppose option then see the following prompt: “Do you favor [oppose] that a great deal, moderately, or a little?” Respondents who initially select “neither” are not asked a follow-up question. Likert items’ wording suggests that answers collected using this survey tool capture a mix of preference orientation and intensity. Indeed, when an individual reports that she opposes giving same sex couples the legal right to adopt a child *a great deal*, she conveys not only that she opposes the proposed policy, but also that she feels strongly about it. Formally, we denote by $x_{ik} \in [-1, 1]$ the answer that individual i would give to this question under Likert if fully *sincere*. Later in this section, we examine what happens when respondents have the incentive to systematically deviate from their sincere views.

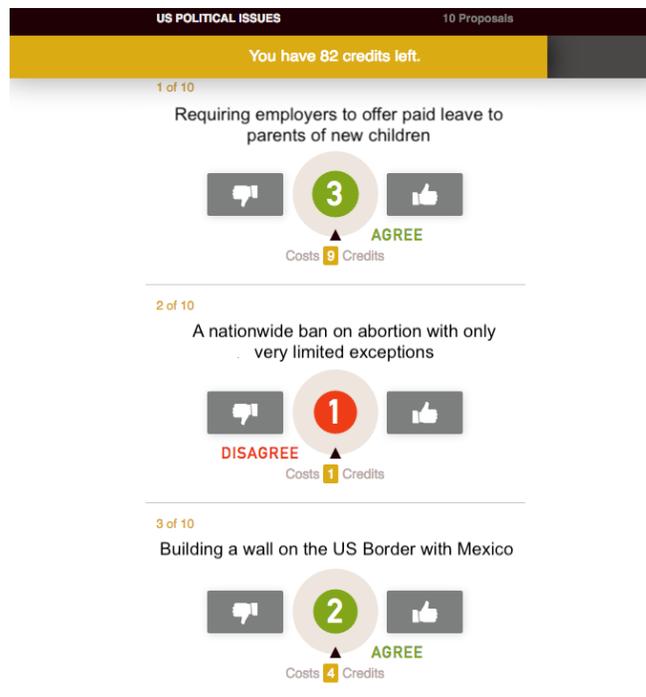
Likert Item Followed by Issue Importance Item With Likert+, there is a follow-up question: “How important is this issue to you personally?” The response options are: “extremely important; very important; somewhat important; not too important; not important at all.” Assuming people are fully *sincere*, Likert+ collects two pieces of information: x_{ik} , as with a standard Likert item, and β_{ik} , directly measured using the follow-up question. The expected improvement with Likert+ compared to Likert is straightforward: it provides a direct measure of preference intensity.

Quadratic Voting for Survey Research QVSR draws on research on quadratic voting (Posner and Weyl 2018).⁷ Each person starts with an artificial budget of credits, which she can use to ‘buy’ votes in favor of (against) a fixed set of proposals. A respondent can cast more than one vote per proposal. A distinctive feature of QVSR is that the cost of buying votes is quadratic: buying one vote in support of (against) one proposal costs one credit; buying two votes for the same proposal costs four credits; buying three votes costs nine credits; and so on. In our own survey, respondents assigned to the QVSR tool are given a budget of 100 credits to spend across ten different survey questions. Figure 1 shows what the survey looks like to the respondent. The first three issues involve positions on the border wall, abortion and parental leave. Respondents can scroll down to vote on the other seven issues examined in the survey. The order in which the 10 proposals are

⁷ In this original formulation, QV was intended as a means of arriving at efficient social decisions when voting on policies that have a high probability of being implemented. Lalley and Weyl (2018) primarily assume that influencing policy is the main motivation of citizens. In contrast, we examine the implications of QV in a very different context, namely survey research, where influencing policy is probably not the main motivation of citizens.

presented is fully randomized. Notice that respondents can use credits in favor of or against each proposal. The cost of each vote increases according to the quadratic form and is displayed below each question. Remaining credits are displayed on the top of the smart phone screen. Respondents can go back to revise their answers until they are satisfied with how they have allocated their credits. The maximum that a respondent can spend in favor of (against) any question is 10 votes (100 credits) though doing so would mean not being able to cast a vote on any of the other 9 issues.

Figure 1: Screenshot of the QVSR Version of the Survey



With QVSR, the respondent faces a budget constraint that prevents her from opposing or favoring *a great deal* too many options. When the budget constraint is binding, an individual cannot report her policy preferences as she would have had under Likert or Likert+. Instead she has to deviate from her unconstrained response on each issue. In our model, we assume that the reluctance to deviate from one's unconstrained answer on one issue is positively correlated with preference intensity. We formally show that, compared to standard Likert items, an individual will allocate more credits to issues over which she has more intense preferences. Answers collected using QVSR

are still expected to be a mix of α_{ik} and β_{ik} , as with Likert items. Yet these answers should be more informative of preference intensity than answers measured using Likert.

With Likert+, the respondent is asked about β_{ik} directly. By design, Likert+ relies on two questions instead of one, and in that sense gathers more information than QVSR. This can be seen as a positive or a negative depending on how much one values parsimony. Another major difference is that QVSR compels individuals to explicitly compare across issues. In Likert+, this comparison is only implicit: asked about the importance of a given issue, respondents could have their answers to the other nine issue importance items in mind but there is no way for the researcher to ensure that this is indeed the case. Under QVSR, the procedure is baked into the hard budget constraint.⁸ However, in our view, the main advantage of QVSR over Likert+ follows from a different set of considerations pertaining to the fact that respondents are only partially sincere. We now turn to this issue.

Non-Sincere Responses: Measuring Policy Preferences in a Polarized Environment

As with any survey instrument, another potential issue with Likert is that respondents' answers to surveys might be only partially sincere. In particular, strategic or partisan motives may induce respondents to inflate their opposition to (or support for) a given reform. The same applies when it comes to expressing how important an issue is to them personally. Consider the example of a Republican voter being asked whether she favors or opposes building a wall between the U.S. and Mexico and whether this issue is important to her personally. Assume further that, in reality, this respondent has some doubts about the adequacy of this policy to address immigration issues, or thinks that this issue is only of secondary importance compared to more pressing economic issues. Still, since this issue is high on the agenda of the Republican party, in a highly polarized political environment, and out of loyalty to her preferred party or elected official, she may feel a psychological pressure to answer that she strongly supports this reform, and that this issue is of the utmost importance (Bullock and Lenz 2019).

⁸ A previous study of how people engage with QVSR finds that respondents proceed by trial and error, moving back and forth across items to best distribute their votes and credits, and thus being forced to compare (Quarfoot et al. 2017).

We model respondents as being intrinsically motivated to report their true preferences, which we call the *sincerity motive*, as well as simultaneously pursuing other objectives when answering surveys. For example, they might also want to strategically influence policy or signal a partisan identity, something we call the *signaling motive*. We assume these conflicting motives to be present whatever the survey technology.

Likert Item With standard Likert items, we show that answers are typically a weighted average of the sincere answer (x_{ik}) and some signaling target. In particular, in a polarized environment where parties are quite extreme, we expect the partisan motive to push respondents to systematically inflate their views, and to report stronger support or opposition than what they actually feel. In the model, we denote by \hat{x}_{ik}^L the answers under standard Likert assuming mixed motives. Our expectation in a polarized environment is that $|\hat{x}_{ik}^L| > |x_{ik}|$.

As a result, responses will bunch at the extremities of the response scale (strongly favor/oppose) making it difficult to distinguish between 1) respondents who strongly favor (or oppose) a given policy change and are ready to act on this preference and 2) respondents who, while expressing the same position, do not care as much and are merely “paying lip service to the party norm” (Zaller 2012), or expressing their dislike for the opposite party (Mason 2015). In other words, if political polarization is high, voters polled using a Likert item will appear to strongly disagree on key issues, when in practice, only a subset of voters really care. When the latter group is unequally distributed across parties (e.g. members of one party strongly support a policy while members of the other only weakly oppose it), variants of Likert items can provide an inadequate picture of public opinion, one that overlooks important asymmetries in preference intensity across parties.

In a polarized two-party system where political elites provide coherent ideological bundles, such bunching at the extremities of a scale occurs on more than one key issue: a strong signaling motive implies ‘strong’ support for a bundle of policies associated with their party. In this case, most of the variance in the data is due to *between* party differences. The available information on differences *within* each political party is, by contrast, much smaller. Relatedly, this “bunching to the extremes” on more than one issue results in limited individual-level information regarding which policies a respondent cares about the most. Instead, she appears to care about many things, with little

information on how she might address policy trade-offs central to democratic politics.

Likert Item Followed by Issue Importance Item Likert+ combines a Likert item and an issue importance item. With regards to the Likert item, the logic is the same as the one stated above ($|\hat{x}_{ik}^{L+}| > |x_{ik}|$). Regarding the issue importance item, a similar partisan motive may push individuals away from reporting their ‘true’ β_{ik} (preference intensity). For example, if an issue is perceived as being very controversial and hotly debated in the political arena, with one’s preferred party indicating that this issue is of the highest priority for the party, loyal partisans may have the incentive to inflate the importance of the issue. Denoting by $\hat{\beta}_{ik}^{L+}$ the answers to the issue importance question under Likert+, one might expect that $\hat{\beta}_{ik}^{L+} > \beta_{ik}$. Assuming a strong signalling motive, the issue importance question potentially suffers from the same bunching issues as the standard Likert items.

Quadratic Voting for Survey Research With QVSR, as explained above, if the budget constraint is binding, an individual cannot report her policy preferences as she would have had under Likert scales (the \hat{x}_{ik}^L), and she will comparatively allocate more credits to issues over which she has intense preferences. A second advantage of QVSR, which is more subtle, is that QVSR not only better captures β_{ik} , it may also provide a more accurate measure of x_{ik} , i.e. the “sincere” component of the unconstrained Likert answer. To understand why, remember that when a signaling motive is present, answers under Likert (the \hat{x}_{ik}^L) are typically a weighted average between the sincere answer (the x_{ik} parameter) and some signaling target. The lower the strength of the sincerity motive compared to the signaling/partisan motive, the further away the Likert answers will be from the sincere answers. Now, remember that under QVSR the respondent comparatively allocates more votes to issues over which she has intense preferences. If one assumes that there is a positive correlation between preference intensity and the strength of the sincerity motive, the model then predicts that under QVSR, the individual will allocate more credits to issues on which the unconstrained answer (\hat{x}_{ik}^L) is closer to the sincere answer (x_{ik}), thus better capturing the true x_{ik} parameter.

Lastly, we can qualify the predictions made earlier regarding the comparison between QVSR and Likert+. When one explicitly takes into account the fact that some respondents may be insincere and have the incentive to inflate the importance of polarizing issues, then answers to the issue

importance question will be a noisy measure of the true intensity parameter. In such a context, and based on the reasoning laid above, QVSR's 'revealed preference' approach could potentially provide a comparatively better measure of preference intensity. In the end, the comparison between Likert+ and QVSR depends on the weight of the sincerity motive. Given the impossibility to directly assess the latter, which tool performs best in the current U.S. context is ultimately an empirical question.

Empirical Predictions

Based on the argument made in the previous sections, we expect both QVSR and Likert+ to better measure preference intensity than a standard Likert item. Assuming respondents are sincere, the advantages of using QVSR over Likert+ are mostly practical, namely parsimony and improved inter-item comparability resulting from forced trade-offs. However, in the current American context, where strong partisan motives compete with the sincerity motive (Bullock and Lenz 2019), we can reasonably expect QVSR to outperform Likert+. Conversely, the finding that, despite the context, Likert+ outperforms QVSR would provide strong evidence against the use of QVSR to measure preference intensity.

To examine empirically which tool best measures preference intensity, we first need to define what are the latter's main empirical manifestations. As a reminder, preference intensity first and foremost affects individuals' likelihood of engaging in costly behavior related to their preferred policy outcome. The relationship between preference intensity on the one hand, and costly issue-relevant behavior on the other, constitutes our first identifying criterion, the behavior criterion.

Beyond behavior, what other observable features of the world can be tied to preference intensity? A common assumption and well-documented finding in political economy is that individuals significantly affected by a policy are more likely to mobilize around this policy (see for example Bouton et al. (2018)). We thus posit that preferences over policies that "directly affect (a voter's) rights, privileges, or lifestyle in some concrete manner" (Howe and Krosnick 2017: 328) are more intense than preferences over policies that have no direct personal implications. This relationship between personal exposure and preference intensity is our second identifying criterion, the exposure criterion.

Let's now imagine a survey question asking respondents their opinion of policy proposal P . Preferences over P are measured using either Likert, Likert+ or QVSR. Each of the three survey technologies can be thought of as a classification tool that distributes a surveyed population across a fixed number of response categories (e.g. 7 categories in the standard Likert scale, 11 in Likert +, an average of 15 in QVSR). The survey tool that best measures preference intensity is the one that best classifies respondents according to their behavior and exposure. In other words:

Prediction 1: Behavior. Relative to Likert answers, both Likert+ and QVSR answers better classify respondents based on their likelihood of engaging in a costly action that aligns with their orientation toward P .

Prediction 2: Exposure. Relative to Likert answers, both Likert+ and QVSR answers better classify respondents based on their likelihood of being directly affected by policy P .

As previously stated, the comparison between QVSR and Likert+ depends on the relative strength of the sincerity motive, something we do not observe. As a result, whether QVSR outperforms Likert + in our target population (American citizens in 2018-2019) remains an empirical question.

Empirical Design and Strategy

To test predictions 1 (behavior) and 2 (exposure), we rely on a survey experiment.⁹ The main component of our survey took place from October 5 to October 9, 2016. The survey was administered to a general population of English-speaking US citizens over the age of 18 who reside in the United States. The survey company, GfK-Ipsos, uses a probability-based web panel designed to be representative of the U.S. population. Respondents (N=3666) were randomly assigned to one of the three survey tools and asked to provide their opinion on 10 policy issues (listed below). Respondents assigned to the QVSR tool were asked to watch a video explaining how QVSR works

⁹ For an overview of the survey design, see Appendix A.2 and A.7.

before they answered the policy questions.¹⁰ All respondents were asked to perform a task with “real world” consequences. Specifically, they were given the opportunity to donate lottery money to single-issue advocacy groups (e.g. immigration control, gun control). Four months later (between January 31 and February 18, 2017), we recontacted a random subset of these respondents and asked them to answer the same 10 survey questions using the survey tool they were assigned to in the first wave (number of responses, N= 1569).¹¹ In wave 2, we collected information on two additional behavioral tasks.¹² First, respondents were asked to play dictator games with an individual who shares or opposes their views on two policy issues (see details below). Second, they were given the opportunity to write to their Senators about real bills that were moving through Congress at the time of the survey. Note that throughout the paper, when we examine the relationship between survey answers and behavior, we only use answers collected in wave 1, though using data collected in wave 2 does not change the results. Readers should nevertheless keep in mind that, in at least two cases (dictator game and letter writing tasks), the attitudinal data is analyzed alongside behavioral data collected four months apart.

Table 1 summarizes the sample sizes for each wave and outcome of interest. Note that the dropout rate is roughly twice as large in the QVSR treatment (14 percent) than in the Likert and Likert+ conditions. Most of the dropout occurred when respondents were asked to watch the video explaining how QVSR works. We found no evidence that this attrition was correlated with key variables of interests (see Appendix A.2 for more details). The difference in sample sizes between the behavior and exposure predictions in wave 1 is due to a subset of individuals who skipped the donation task. These individuals are distributed equally across the three treatment conditions. Below we list the ten survey questions asked in both wave 1 and wave 2:

Do you Favor or Oppose:

¹⁰ The video can be found here: https://www.youtube.com/watch?v=GrY_RzDsQLY.

¹¹ Participation in wave 2 is not predicted by treatment condition and policy preferences in wave 1, nor by partisanship. See Appendix A.2 for more on balance across treatment conditions and wave 2 participation.

¹² Wave 1 took on average 9 min to complete. The second wave enabled us to collect additional behavioral data without burdening respondents in wave 1.

Table 1: Sample Sizes: Overview

Behavior (Wave 1)	Likert	Likert +	QVSR
Valid observations	1215	1217	1116
Consented	1306	1305	1289
Exposure (Wave 1)	Likert	Likert +	QVSR
Valid observations	1291	1298	1191
Consented	1306	1305	1289
Behavior (Wave 2)	Likert	Likert +	QVSR
Valid observations	523	539	470

- Giving same sex couples the legal right to adopt a child
- Laws making it more difficult for people to buy a gun
- Building a wall on the U.S. Border with Mexico
- Requiring employers to offer paid leave to parents of new children
- Preferential hiring and promotion of blacks to address past discrimination
- Requiring employers to pay women and men the same amount for the same work
- Raising the minimum wage to \$15 an hour over the next 6 years
- A nationwide ban on abortion with only very limited exceptions
- A spending cap that prevents the federal government from spending more than it takes
- The government regulating business to protect the environment

Measuring Behavior and Exposure

The first behavioral task (wave 1) elicits the willingness to incur a monetary cost to promote a political cause one agrees with. At the beginning of the task, respondents were told that, as participants to the survey, they had been automatically entered into a lottery with “a prize of \$100 for 40 randomly selected respondents (among 4000 or so).” They were then prompted to imagine that they were among the winners and asked whether they wanted to donate part of their lottery

money to an advocacy group. They had a choice between four advocacy groups working on two issue areas: immigration and gun control. For each issue area, we chose organizations that fall on different sides of the political divide: for and against immigration, as well as for and against gun control. Respondents could choose not to donate or to donate to one of the four advocacy groups. Whatever they did not donate, they could keep.

The second task (wave 2) approximates a situation in which respondents are cross-pressured. The task involved a dictator game in which respondents were faced with a clear trade-off: reward or punish another participant –identified as an independent– who agrees with them on one issue but disagrees with them on another. To start, we asked respondents how they would behave in three dictator games involving a Republican, a Democrat and an Independent respectively (the order was randomized). Respondents had the option to donate anywhere between \$0 and \$100 of some lottery money (following a similar set up as the one in wave 1). After they made their decisions, we informed them of the Independent respondent’s preferences on two issues, against the border wall and against gun control. We asked respondents if they wanted to change the amount they had previously decided to donate to this individual.¹³ We chose this specific mix of preferences (anti-gun control and pro-immigration) because they rarely coexist in practice, and can be, as a result, expected to generate ambivalence among the majority of respondents. For example, for most Democrats the task involved a clear trade-off: punishing the Independent for donating to a pro-gun organization meant also punishing her for donating to a pro-immigration organization.

The third behavioral task (wave 2) captures the willingness to spend time and effort promoting a political cause one agrees with. Respondents were given the option to write a short text about one of two policy proposals being discussed in Congress at the time. One proposal sought to restrict abortion rights and was sponsored by Sen. Roger Wicker (R-Miss.). The other sought to increase the minimum wage and was sponsored by Sen. Bernie Sanders (I-Vermont). We did not mention who the bill sponsors were, only the content of the bills. The texts provided by the respondents were then integrated into a letter, which was ultimately sent to the Senate committees in charge of reviewing the policy proposals. Comments were anonymous. Section A.7 in the Appendix provides

¹³ The Independent’s preferences were conveyed using information on donation patterns in wave 1. In practice, respondents could only donate to one non-profit. After implementing the lottery, we picked an independent to receive the funds based on her survey answers, not her donation behavior.

screen shots of all three tasks. Table 2 provides an overview of the outcome variables derived from these tasks and used in the remainder of the analysis.

With two exceptions, measures of exposure (prediction 2) were directly purchased from GfK-Ipsos, meaning that respondents were not primed to think about their own situation. The two exceptions are information on whether one would benefit from a minimum wage increase and whether one would benefit from paid parental leave. Questions used to construct these two exposure variables were asked at the beginning of the survey. Table 3 lists all exposure variables alongside the relevant policy issues.

Table 2: Behavioral Outcomes and Relevant Survey Question

Variable	Description	Mean (Stand. dev.)	Survey question expected to correlate with behavior
Donation to gun-related advocacy group	Equal to the \$ amount donated multiplied by 1 if donated to pro gun control and -1 if donated to anti gun control advocacy group.	9.8 (33.5)	Laws making it more difficult for people to buy a gun
Donation to immigration-related advocacy group	Equal to the \$ amount donated multiplied by 1 if donated to anti immigration and -1 if donated to pro immigration advocacy group.	1.6 (28.7)	Wall on the border with Mexico
Punishment of Independent respondent (1)	Equal to the \$ amount <i>taken off</i> the amount previously donated to the Independent. If respondent <i>gave</i> additional \$ then amount multiplied by -1 . Note that individuals who donated in the first round but did not change their answers in the second, and individuals who did not donate in the two rounds are both coded as 0.	4.9 (13.5)	Laws making it more difficult for people to buy a gun/ Wall on the U.S. border with Mexico
Punishment of Independent respondent (2)	Equal to the \$ amount <i>taken off</i> the amount previously donated to the Independent as a <i>proportion</i> of the amount originally donated. If respondent <i>gave</i> additional \$ then amount multiplied by -1 . Note that individuals (N = 24) who did not donate in the first round but nevertheless asked to take \$ away from the Independent are coded as missing (N = 24).	0.17 (0.38)	Laws making it more difficult for people to buy a gun/ Wall on the U.S. border with Mexico
Letter writing on the minimum wage bill	Equal to the length of text written (number of characters).	76 (139)	Raising the minimum wage to \$15/h over the next 3 years (absolute values)
Letter writing on the abortion bill	Equal to the length of text written (number of characters).	59 (90)	A nationwide ban on abortion with only very limited exceptions (absolute values)

Table 3: Proxies of Personal Exposure and Relevant Policy Proposal

Variable	Measurement (Origin)	Survey question expected to correlate with exposure	Coding (Mean)
Sexual orientation	Which of the following best describes how you think of yourself? Gay/Lesbian or Straight (GfK-Ipsos)	Giving same sex couples the legal right to adopt a child	Equal to 0 if heterosexual, 1 otherwise (0.07)
Gun ownership	Are there any guns or revolvers that are kept in your home or garage? (Y/N) Do any of these guns personally belong to you? (Y/N) (GfK-Ipsos)	Laws making it more difficult for people to buy a gun	Equal to 1 if does not own gun or none in the house, 0 otherwise (0.63)
Proximity to immigration event	Where were you / your parents / grand parents born?	Wall on the U.S. border with Mexico	Equal to 1 if one or both parents born abroad, 0 otherwise (0.16)
Gender	What is your gender? (GfK-Ipsos)	Require employers to pay women and men the same amount for the same work?	Equal to 1 if female, 0 otherwise (0.51)
Race	What is your race/ethnicity (GfK-Ipsos)	Preferential hiring and promotion of blacks to address past discrimination	Equal to 1 if black, = 0 otherwise (0.13)
Proximity to birth of child	Do you have children? How old is your youngest? Do you plan on having children (another child) one day?	Require employers to offer paid leave to parents of new children	Equal to 1 if about to have a child, plans to have a child soon or already has a toddler, 0 otherwise (0.19)
Wage level relative to current/proposed minimum wage	Are you paid at/below/above the minimum wage? Are you paid at/below/above \$15/h ?	Raising the minimum wage to \$15/h over the next 3 years	Equal to 1 if would benefit from increase to \$15/h, 0 otherwise (0.29)
Religiosity / born-again	Would you describe yourself as a born-again or Evangelical Christian? How often do you attend religious service? (GfK-Ipsos)	A nationwide ban on abortion with only very limited exceptions	Equal to 1 if *not* a born-again, = 0 otherwise (0.63)

Estimation Strategy

Each of the three survey technologies can be thought of as an ordinal classification tool that differs from the other two in terms of the total number of ordinal categories and the distribution of observations across these categories. We discuss each in turn.

Number of Response Categories Likert has 7 response categories: from -3 (strongly oppose) to 3 (strongly favor), centered around 0 (neither/nor). QVSR in theory ranges from -10 (up to 10 votes against) to 10 (up to 10 votes in favor), centered around 0 (no votes). In practice, the highest number of votes cast is 7 ¹⁴ not 10 , meaning that for a given issue QVSR sorts respondents across 15 response categories in total. In the case of Likert+, we combine the Likert and issue importance items into one single scale by multiplying values of the first Likert branching question (favor = 1 , opposes = -1 , neither = 0) by the values of the issue importance scale. The result is an ordinal scale with 11 response categories ranging from -5 (oppose, extremely important) to 5 (favor, extremely important) and centered around 0 (neither/nor).¹⁵ In the remainder of the analysis, we use normalized answers such that the lowest possible answer corresponds to zero ($-3 / -5 / -7$) and the highest possible to 1 ($3/5/7$).

Distribution of Responses As shown in the bottom panel of Figure 2 (see emphasis in red), when measured using a Likert item, the distribution of answers to a politicized issue such as gun control is uni-modal: most answers bunch on one extreme of the scale. This pattern is somewhat less pronounced in Likert+, implying that, while most respondents strongly support gun control, not everyone believes this issue to be personally important to them. Responses in QVSR, by design, exhibit no such bunching patterns.

More response categories and less bunching imply higher information entropy for Likert+ and QVSR on the one hand relative to Likert on the other. For 8 out of 10 survey items, QVSR also has higher entropy relative to Likert+.¹⁶ The aim of our empirical analysis is to establish whether

¹⁴ There are a few respondents casting 8 or 9 votes, we merge them with the '7 votes' response categories.

¹⁵ We discuss alternative combinations of the two items in Appendix A.3.

¹⁶ Shannon entropy for each of the 10 issues broken down by survey tool are provided in Appendix A.6.

this higher entropy is more than just noise. If more response categories and more de-bunching in QVSR is informative of preference intensity, then individuals assigned through their answers to the lowest available score provided by QVSR should be very different from those assigned to the highest available score. The difference in this case is in terms of their behavior and their exposure. Individuals assigned to intermediary scores, should not exhibit a behavior (exposure) that signals more intense (less intense) preferences than individuals with the highest (lowest) score. In practice, this implies a positive and monotonic relationship between a) response categories and b) the mean of the outcome of interest (behavior or exposure) conditional on the response category.

Behavioral Tasks (P1) To assess the extent to which a given survey tool successfully sorts respondents into increasingly discriminating categories, we first match a given outcome Y to the relevant policy issue in the survey (e.g. donation to a gun-related organization and attitudes toward gun control). Table 2 provides more detail on the three continuous behavioral outcomes (donation in dollars, punishment in dollars, writing in number of characters). Thanks to randomization, the distribution of outcome variables is the same across the three treatment conditions (see Appendix A.2). We then regress the behavioral outcomes over the normalized survey responses (X),¹⁷ interacted with a categorical variable identifying the treatment conditions with the Likert condition as the reference category:

$$Y_i = \beta_0 + \alpha_1 D_{i,Likert+} + \alpha_2 D_{i,QVSR} + \beta_1 X_i + \beta_2 X_i D_{i,Likert+} + \beta_3 X_i D_{i,QVSR} + \beta_4 J_4 + \dots + \beta_j J_j + \varepsilon_i \quad (1)$$

where J_4, \dots, J_j are dummy variables that indicate membership in a block used for black randomization (see Appendix A.2). Regression coefficients β_1 , $\beta_1 + \beta_2$ and $\beta_1 + \beta_3$ can be interpreted as the difference between $E(Y/X = 1)$ and $E(Y/X = 0)$ for Likert, Likert+ and QVSR respectively. The better discriminating tool is the one with a larger difference between the two quantities of interest, i.e. with a larger regression coefficient.¹⁸

¹⁷ Note that when predicting the number of characters written, we use the absolute value of the preference variables, i.e. 0-3 for Likert, 0-5 for Likert+ and 0-7 for QVSR.

¹⁸ We checked for a non-parametric relationship and, for all methods, the standard errors do not rule out a monotonic

To understand the intuition behind this estimation strategy we turn to Figure 2, which plots average donations to the gun control charities by (normalized) survey response on the gun control question, further broken down by survey methodology. The lines captures the three regression coefficients mentioned above (see Figure 3 for the actual estimates). As shown on this figure, the regression slope is larger for QVSR than for Likert. Individuals who chose the end-of-scale response categories in Likert end up de-bunching under QVSR. This de-bunching aligns with their behavior on the donation task: people who donate less chose smaller values in QVSR than people who donated more. This is captured by the magnitude of the regression slope: individuals who do not donate are no longer pulling the regression slope down by ‘sharing’ the extreme response categories with people who care enough to donate. Using the regression coefficients as a summary statistics, we can also see that the sorting achieved with QVSR better aligns with preference intensity than the sorting achieved with Likert+.¹⁹

Exposure proxies (P2) The exposure variables are re-coded as binary indicators equal to 1 if one is likely to be affected by a given policy proposal, and equal to 0 otherwise (see Table 3 for more details). We follow the same procedure as the one described for prediction 1 with two differences: we use a logistic regression and plot differences in predicted probabilities instead of regression coefficients. In Appendix A.3, we also present results from an OLS regression, the results remain substantively unchanged.

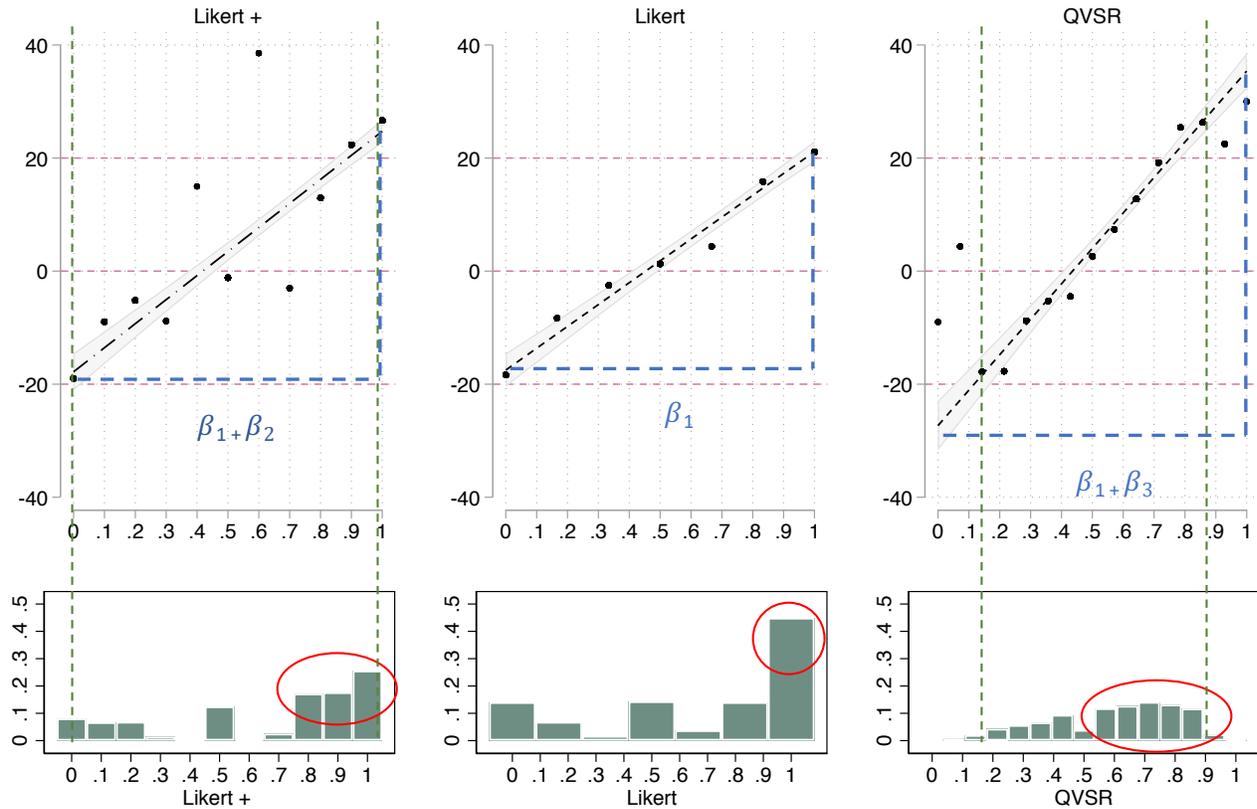
Results

Behavioral Tasks (P1) Figure 3 plots regression coefficients obtained using equation 1. The higher the regression coefficient, the better a given tool is at distinguishing between respondents

relationship. In Appendix A.6, for each survey tool, we plot \bar{Y} for all values of X and overlay a linear fit line based on the underlying data. For three outcomes, a quadratic function provides a better fit. As we discuss, this does not affect the interpretation of our results.

¹⁹ Note that if increasing the number of response categories was enough to generate new information, then Likert items would not be limited to 7 categories. As shown by Revilla, Saris and Krosnick (2014), with Likert, the quality of the data collected deteriorates as the number of categories increases.

Figure 2: Estimation Strategy



Likert, in the center of the figure provides the benchmark. A visual comparison indicates that the coefficient in Likert+ is only marginally larger than that for Likert. Notice the difference in slope between Likert and QVSR. The full estimates are available in Figure 3. Compare (red circles) the bunching in Likert and the variance gained under Likert+ and QVSR.

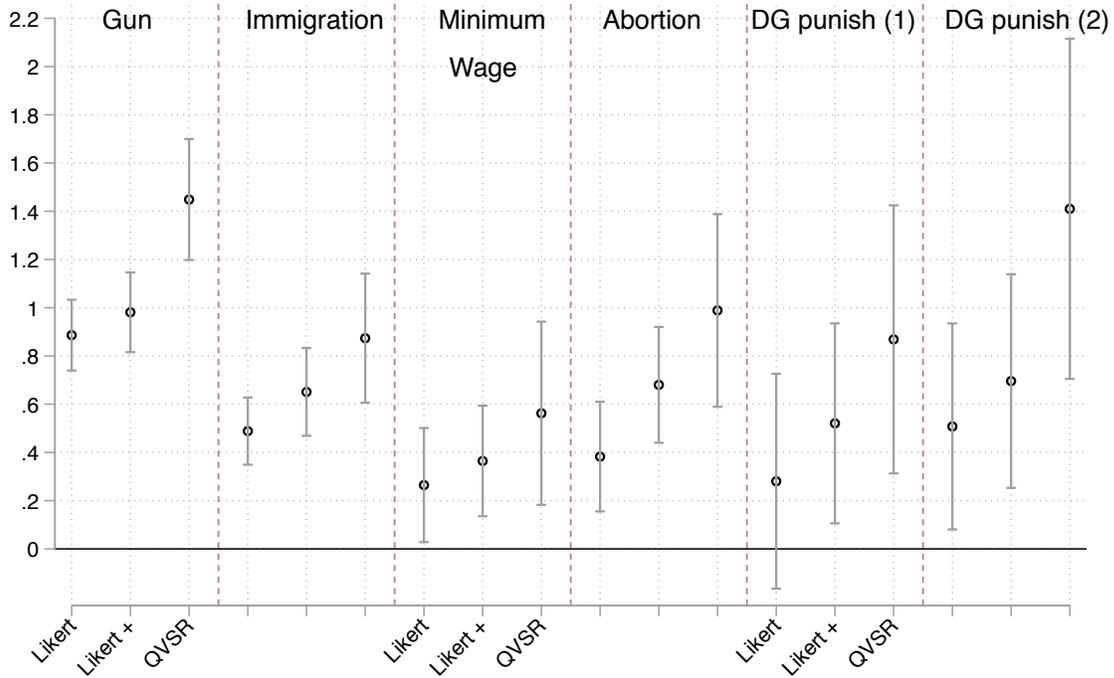
Y-axis: Donation to gun-related advocacy group. X-axis: answers (normalized) for the following item “Laws making it more difficult for people to buy a gun.”

with high and low preference intensity (as proxied by behavior).²⁰ In the case of the first donation task, QVSR outperforms Likert on the two outcomes of interest (‘gun’ and ‘immigration’). QVSR also outperforms Likert+ on one of the two outcomes (‘gun’). In contrast, there is no evidence that Likert+ does systematically better than Likert or QVSR at discriminating between respondents based on their donation patterns. Due to smaller sample sizes, results for the other two tasks (4 outcomes in total) are only indicative. Still, larger regression coefficients also suggest that QVSR

²⁰ Note that the predictor for the punishment task is a combination of answers on the gun control and the border wall items. Specifically, we took the difference between the two answers: a positive value indicates someone who is more to the ‘left’ on gun control than on the border wall, whether in preference orientation or preference intensity. This difference was then normalized to enable comparison across survey methods. The higher the value, the more someone should want to punish the Independent.

has higher discrimination abilities than the other two methods.

Figure 3: Regression Coefficients for Behavioral Outcomes



For details on each task, see text.

Interpretation: a switch from the smallest response category (0) to the largest (1) is associated with a β increase in Y . The increase is equal to β times the standard deviation of Y .

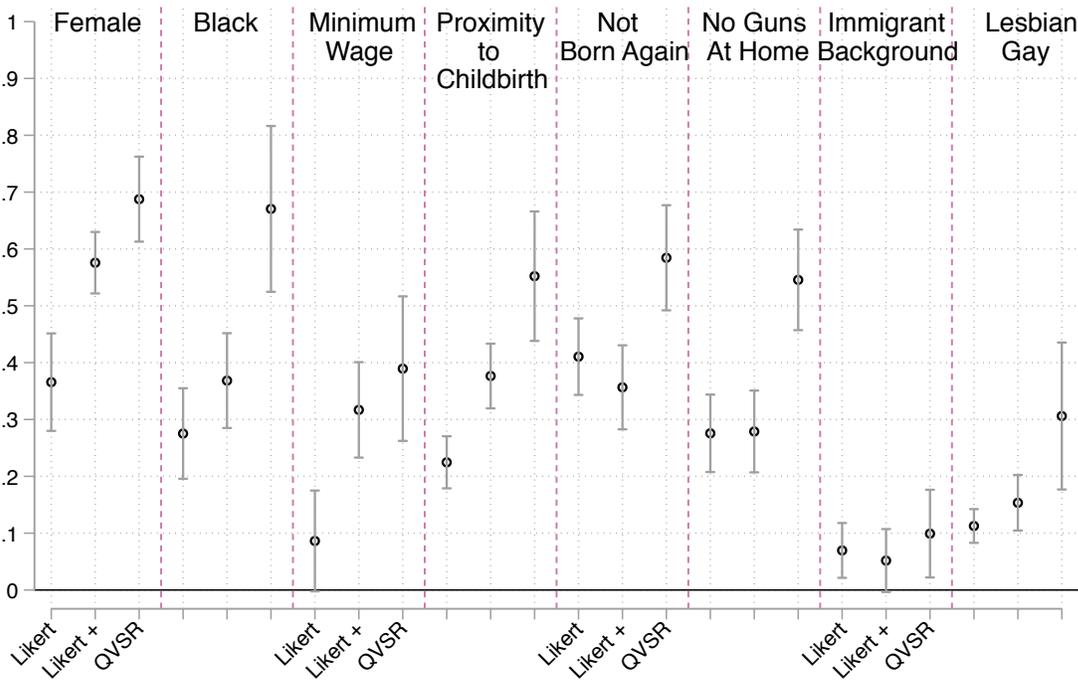
Note that for the punishment task, the predictor is normalized difference in answers on the gun control item minus answers on the border wall item (see fnt 20).

Exposure proxies (P2) Figure 4 plots differences in predicted probabilities obtained using equation 1 modified by adding a logit link function. Higher differences indicate higher discrimination abilities. For all exposure proxies but one (immigrant background), QVSR outperforms Likert items. Likert+ outperforms Likert for only 3 out of 8 exposure proxies. QVSR outperforms Likert+ for 5 out of 8 exposure proxies, while Likert+ never outperforms QVSR.

Figure 5 provides a visual overview of the relationship between the estimates provided in Figure 4 and the underlying data, focusing on the case of gender. Notice how in Likert and Likert+, there is very little variance in survey answers: most people appear to *strongly* support gender equality in the work place or find it *very* or *extremely* important. The additional variance gained by switching from Likert to Likert+ is informative of respondents' gender: women are more likely than men to

be in the ‘extremely important’ response category. This is especially true when switching to QVSR. As shown on the right-hand side of Figure 5 there is a striking linear and monotonic relationship between the number of votes in QVSR and the percentage of women as a share of individuals who cast the same number of votes ($P(Y = 1/X = x)$). Individuals who cast a higher number of votes are much more likely to be women (for all other outcomes see Appendix A.5.2).

Figure 4: Predictions for Proxies of Exposure



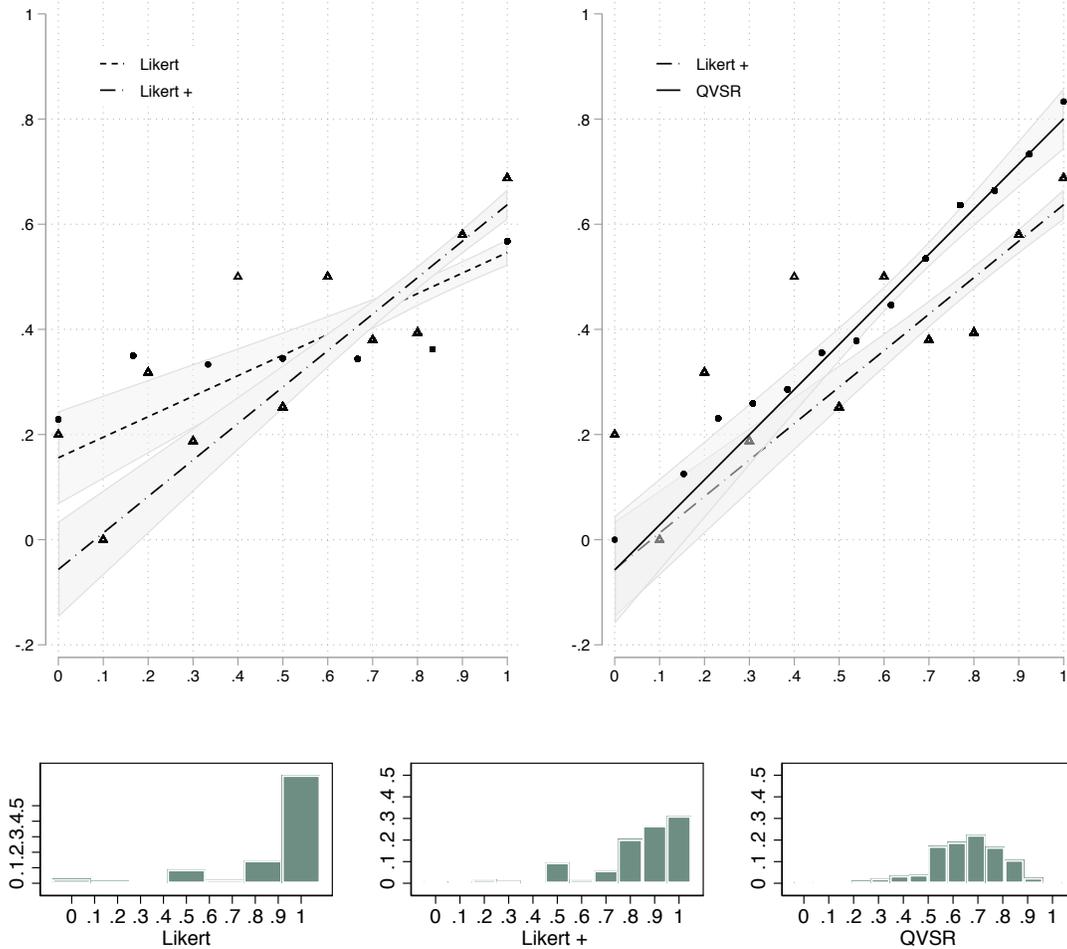
For details on each proxy, see text.

Interpretation: a switch from the smallest response category (0) to the largest (1) is associated with a β increase (in % points) in the probability of exhibiting the mentioned feature Y . Probabilities computed using a logit regression (see text). OLS estimates are available in Appendix A.3.

Summary and Next Steps

QVSR generates more information thanks to a novel technique that increases the cost of choosing the extremities of the scale. This information is meaningful and correlates with preference intensity. The value of the empirical tests presented above comes from the diversity of outcomes considered in the analysis. Jointly, our results indicate that QVSR is never the worst performer and is, most of the time, more informative of preference intensity than the second best performing survey tool. In

Figure 5: Respondent's Gender and Response to Pay Equity Item



Y-axis: gender of respondent (female = 1), scatter plot represents the share of all respondents who are female conditional on having the same value on X, i.e. $E(Y=1/X=x)$. X-axis: survey answers by survey method, normalized to vary from 0 to 1.

Each line plots the relationship between Y and X derived from a linear regression (not a logit as in Figure 4. Lines are overlaid on binned scatter plots of share who are women. Each bin is a survey response category.

line with our argument in section 1, Likert+ also appears to carry more information on preference intensity than Likert, but its relative performance is less consistent than that of QVSR.

The behavioral tasks used to test prediction 1, while more costly than simply answering a survey question, are nevertheless still relatively low-stakes. There are two advantages to this design. First, it captures the reality of everyday politics from the point of view of the individual voter, namely a relatively low-stakes (at least materially) yet highly symbolic activity. Second, it provides a conservative test of QVSR. Evidence that QVSR outperforms Likert and Likert+ items for such

low-cost outcomes implies that QVSR will outperform them in predicting more costly behaviors.

Additional methodological issues remain to be investigated. For example, we have yet to examine the impact of changing the menu of options, as well as the consequences of switching to other forms of pricing (e.g. linear instead of quadratic). Most importantly, we have set aside the issue of anchors that allow comparability across menus and surveys, though, empirically, the latter concern does not appear to hinder the interpretation of our results. Possible anchors include substituting one of the questions with a non-political item. A revealed preference anchor might also include a small monetary lottery, with respondents buying the right with their credits for tickets in the lottery. To facilitate follow-up studies, we are making available a web application enabling researchers to design a survey using different variants of QVSR. We hope this tool will help generate additional evidence regarding the pros and cons of QVSR and spur new research on theoretically-informed solutions to the abundance and the bunching problems highlighted in this paper.

Discussion

Despite its theoretical importance, there currently exists no routine way of measuring preference intensity and the concept is often overlooked in empirical research on political behavior. Studies that explicitly emphasize preference intensity often rely on the issue importance item without much insight on its ability to measure preference intensity. As discussed in our model, in a polarized environment with highly salient partisan identities, such ordinal scales risk imperfectly capturing the share of respondents who “really” care: in order to pay lip-service to the party line, many respondents might endorse extreme statements and inflate the importance of issues that are high on their preferred party’s agenda.

Our arguments and results advance survey methodology in three ways. First the model presented in this paper (and in Appendix A.1) lays the foundations for evaluating potential solutions to the abundance and bunching problems identified earlier. Second, to the best of our knowledge, no existing study uses real-world behavior and policy exposure proxies as benchmarks to evaluate the performance of different methods for measuring preference intensity. Against recent claims that Likert+ is an unhelpful measure of preference intensity (e.g. Hanretty, Lauderdale and Vivyan

(2020)), we show that the additional variance generated by this survey tool does carry information on preference intensity.

Third, and most importantly, with QVSR, we test a new “revealed preference” approach to the measurement of preference strength. Instead of asking explicitly and openly how important an issue is, QVSR uses the fact that it forces trade-offs across issues to measure how much people care. Our results point to QVSR’s comparatively higher discrimination ability, making this survey tool particularly valuable for work on the micro-foundations of models of democratic politics. For example, Kuziemko et al. (2015) ran a survey experiment to understand how support for redistribution is affected by information on inequality. They find no effect of their informational treatment on support for redistributive social policies, which were measured using a Likert item. QVSR, because it yields more discriminatory variance and a better ‘signal’ would provide more solid grounds to conclude that information on inequality does not affect preference for redistribution (whether preference orientation or preference intensity).

QVSR could also benefit descriptive work, helping, for example identify what Converse calls issue publics, i.e. small groups of voters with intense preferences on a few issues and outsized influence at key steps in the election process. In our survey, QVSR appears most informative on issues such as affirmative action or tackling budget deficits. Republicans have comparatively much more intense preferences on these issues than Democrats. On issues of gun control or abortion, results point to a possible “hold-up” of the Republican party by a small minority of voters with intense conservative preferences on these issues. Noticeably, there is a similarly sized group of Republican voters with equally strong liberal preferences (i.e. in favor of gun control and against an abortion ban). More descriptive statistics for other issues are provided in Appendix A.5.

Having better measures of preference intensity could also contribute to research on the role of policy preferences. For example, there is a core disagreement between behavioral studies of public opinion and work in political economy. To simplify, political economists tend to assume that individuals hold political preferences and that these preferences directly, and consequently predict politically-relevant behavior. Students of public opinion have challenged the accuracy and adequacy of these behavioral assumptions. Achen and Bartels (2016), for instance, argue that the causal arrow between preferences and behavior should be reversed: citizens first pick a politician

for reasons that have little to do with policy preferences and then adopt that politician's policy views (see also Lenz (2013); Freeder, Lenz and Turney (2018)). "Even the more attentive citizens," they write, "mostly adopt the policy positions of their parties as their own: they are mirrors of the parties, not masters" (p 299). One way to rescue political economy's core assumptions while also accounting for the empirical patterns documented by those critical of this framework is to emphasize that preferences vary in intensity.²¹ Intense "attitudes have the characteristics that" political economists "assume are possessed by all (preferences)" (Howe and Krosnick 2017: 328). While intensely held attitudes might result in party switching, weakly held ones are more likely to be affected by partisan elite rhetoric (Zaller 1992, 2012). In other words, other preferences are easier to change or manipulate, but the result of this change is unlikely to have meaningful behavioral consequences (Lenz 2009; Zaller 2012).²² A tool like QVSR will help revisit this important debate.

²¹ See Krosnick (1999); Howe and Krosnick (2017); Miller and Peterson (2004) for more on this theme.

²² Zaller, in a recent review of the research published since his 1992 opus, similarly distinguishes between strong and weak attitudes.

References

- Achen, Christopher and Larry Bartels. 2016. *Democracy for Realists: Why Elections Do Not produce Responsive Government*. Princeton, Oxford: Princeton University Press.
- Arrow, Kenneth. 1951. *Social Choice and Individual Values*. New York: Wiley.
- Bouton, Laurent, Paola Conconi, Francisco Pino and Maurizio Zanardi. 2018. Guns, Environment, and Abortion: How Single-Minded Voters Shape Politicians' Decisions. Technical report CEPR Discussion Paper No. DP12801.
- Bullock, John and Gabriel Lenz. 2019. "Partisan Bias in Surveys." *Annual Review of Political Science* 22:325–342.
- Cruces, Guillermo, Ricardo Perez-Truglia and Martin Tetaz. 2013. "Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment." *Journal of Public Economics* 98:100–112.
- Freder, Sean, Gabriel S Lenz and Shad Turney. 2018. "The Importance of Knowing 'What Goes With What'." *Journal of Politics* 81(1):274–290.
- Hanretty, Chris, Benjamin Lauderdale and Nick Vivyan. 2020. "A Choice-Based Measure of Issue Importance in the Electorate." *American Journal of Political Science* 64(3):519–535.
- Howe, Lauren and Jon Krosnick. 2017. "Attitude Strength." *Annual Review of Psychology* 68:327–351.
- Krosnick, Jon A. 1999. "Survey research." *Annual review of psychology* 50(1):537–567.
- Krosnick, Jon and Robert Abelson. 1992. The Case for Measuring Attitude Strength in Surveys. In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, ed. J.M. Tamur. Russell Sage Foundation pp. 177–203.
- Kuziemko, Ilyana, Michael Norton, Emmanuel Saez and Stefanie Stantcheva. 2015. "How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments." *American Economic Review* 105(4):1478–1508.
- Lalley, Steven and E Glen Weyl. 2018. "Quadratic Voting: How Mechanism Design Can Radicalize Democracy." *American Economic Association Papers and Proceedings* 108:33–37.

- Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Priming Hypothesis." *American Journal of Political Science* 53(4):821–837.
- Lenz, Gabriel S. 2013. *Follow the Leader?: How Voters Respond to Politicians' Policies and Performance*. Chicago: University of Chicago Press.
- Malhotra, Neil, Jon A Krosnick and Randall K Thomas. 2009. "Optimal Design of Branching Questions to Measure Bipolar Constructs." *Public Opinion Quarterly* 73(2):304–324.
- Mason, Lilliana. 2015. "'I Disrespectfully Agree': The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* 59(1):128–145.
- Miller, Joanne and David Peterson. 2004. "Theoretical and empirical implications of attitude strength." *The Journal of Politics* 66(3):847–867.
- Posner, Eric and E Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton: Princeton University Press.
- Quarfoot, David, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein and Ellen Konar. 2017. "Quadratic Voting in the Wild: Real People, Real Votes." *Public Choice* 172(1-2):283–303.
- Revilla, Melanie, Willem Saris and Jon Krosnick. 2014. "Choosing the Number of Categories in Agree–Disagree Scales." *Sociological Methods & Research* 43(1):73–97.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.
- Zaller, John. 2012. "What Nature and Origins Leaves Out." *Critical Review* 24(4):569–642.