Advance Access publication November 21, 2012

Political Analysis (2013) 21:48–69
doi:10.1093/pan/mps031

Asking About Numbers: Why and How

Stephen Ansolabehere
Department of Government, Harvard University, Cambridge, MA
e-mail: sda@gov.harvard.edu

Marc Meredith
Department of Political Science, University of Pennsylvania, Philadelphia, PA
e-mail: marcmere@sas.upenn.edu

Erik Snowberg
Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, and National Bureau of Economic Research
e-mail: snowberg@caltech.edu (corresponding author)

Edited by Jasjeet Sekhon

Survey questions about quantities offer a number of advantages over more common qualitative questions. However, concerns about survey respondents' abilities to accurately report numbers have limited the use of quantitative questions. This article shows quantitative questions are feasible and useful for the study of economic voting. First, survey respondents are capable of accurately assessing familiar economic quantities, such as the price of gas. Second, careful question design—in particular providing respondents with benchmark quantities—can reduce measurement error due to respondents not understanding the scale on which more complex quantities, such as the unemployment rate, are measured. Third, combining quantitative and qualitative questions sheds light on where partisan bias enters economic assessments: in perceiving, judging, or reporting economic quantities.

1 Introduction

Knowledge of the economy is important to a broad range of decisions: from private decisions, like investment and educational choices, to public decisions, such as voting for bond issues or holding the President accountable for his economic policy choices. Economic information is often expressed in numbers, for example, prices, rates, and even consumer confidence. Yet, there is very little public opinion research on what citizens know about economic quantities, and how they use that knowledge to make decisions.[1]

This lack of research is perplexing, as theories of economic voting are fundamentally rooted in numbers. In particular, the political economy literature focuses on how vote shares change with economic quantities, such as gross domestic product (GDP) growth, the unemployment rate, or inflation (e.g., Kramer 1971; Fair 1978; Alesina, Londregan, and Rosenthal 1993). Parallel investigations by survey researchers, however, have not yielded a set of findings that can be readily linked to statistical models used in the analyses of such aggregate data.[2] A fundamental cause of

Authors' note: We thank Mike Alvarez, Conor Dowling, Ray Duch, Jon Eguia, Ken Scheve, Emily Thorson, and Chris Wlezien for encouragement and suggestions, and seminar audiences at Columbia, MIT, MPSA, NYU, Temple, Wharton, and Yale for useful feedback and comments. Replication data may be found in Ansolabehere, Meredith, and Snowberg (2012). Supplementary Materials for this article are available on the *Political Analysis* Web site.

[1]Quantities are relevant for a wide range of policy areas: for example, average test scores, incarceration rates, abortion rates, and health care expenditures.

[2]For example, whether voting is based on aggregate economic outcomes or a voter's personal economic outcomes (whether voters are sociotropic or egotropic; see Kinder and Kiewiet 1979, 1981) does not affect patterns in aggregate data because, on average, personal outcomes improve when aggregate outcomes improve.

this disconnect is that survey researchers tend to rely on respondents' qualitative evaluations of the economy, rather than ask about quantities directly. Thus, developing standard questions to measure perceptions of economic quantities is essential for testing theories of economic voting. Our goal is to take a first step toward developing such standard questions by evaluating survey questions that directly ask about quantities.[3]

This article shows that survey respondents provide accurate responses to questions about familiar quantities, such as the price of gas. Moreover, we show that careful question design can reduce problems associated with quantitative questions when asking about more complex quantities, such as the unemployment rate. Finally, we show that quantitative questions seem to exhibit less partisan bias than qualitative questions. That is, this article summarizes the advantages of, and concerns about, quantitative questions, and makes some progress in alleviating those concerns.

For the past 30 years, the standard survey question used to measure economic evaluations and study economic voting has been some variant of the retrospective economic evaluation:

> Now, thinking about the economy in the country as a whole, would you say that over the past year, the nation's economy has gotten better, stayed about the same, or gotten worse?

There are a number of advantages to such *qualitative questions*. No sophisticated knowledge of economics is required to understand or respond to this question. It is also broad, allowing researchers to measure economic evaluations with a single question. Finally, this question is easily portable across surveys, facilitating comparisons across both time and space about the relationship between economic assessments and political behavior.

These advantages, however, come with some costs. The retrospective economic evaluation does not relate to a specific dimension of the economy—for example, GDP growth, levels of employment, tax policy, or changes in prices (inflation). Such vagueness invites projection of attitudes other than economic understanding, such as political beliefs. It also reduces complex economic assessments to one of three possible answers.[4] Finally, this question cannot separate a respondent's perception of economic conditions on several dimensions from his or her judgment of whether those perceived conditions represent an improvement.

While *quantitative questions* are not perfect, they have a distinct conceptual advantage over qualitative questions. Namely, they can be used to ask respondents about the basic building blocks of many theories: quantities. However, quantitative questions remain largely unused in the study of economic voting due to concerns about whether respondents are capable of understanding and reporting economic quantities. Previous work generally finds that a substantial number of people report wildly inaccurate views about quantities, such as the unemployment rate or inflation (Conover, Feldman, and Knight 1986; Holbrook and Garand 1996; Blendon et al. 1997; Curtin 2007). However, if a majority of voters are indeed unable to understand economic quantities, this calls into question the assumptions on which most political economic models are built.

In order to show respondents can handle numbers, we examine their assessments of the price of gas. Gasoline expenditures are a sizable portion of most households' budgets, and the price of gas is frequently encountered in respondents' everyday environment. Thus, inaccurate reports of the price of gas would suggest that there are few economic quantities respondents can accurately report. However, in three different election surveys run since 2006, individuals accurately report the average price of the gas. Moreover, there are few observable characteristics that consistently relate to either the bias or accuracy of reported gas prices. One characteristic that does make individuals more accurate is exposure to gas prices—for example, through driving—which is consistent with the idea that survey respondents are capable of reporting quantities with which they are familiar.

[3]In other domains, quantitative questions may be useful for measuring exposure or knowledge rather than perceptions. As our findings focus on the measurement of perceptions, more research is needed to understand how quantitative questions function in measuring these other dimensions.

[4]Many election surveys also let respondents choose between "somewhat" and "substantially" better and worse.

Asking about more complex quantities requires more care. In particular, we advocate the use of benchmarks to describe the scale on which the quantity of interest is measured. For example, most respondents are infrequently exposed to the unemployment rate, if at all, and thus provide very inaccurate assessments. A potential reason that is discussed in the literature for this inaccuracy is that not all respondents share a common understanding of what is meant by the unemployment rate, and hence that they are not reporting unemployment rates on the same scale. We show that providing respondents with the historical range and average level of the unemployment rate reduces differences in reported rates between higher and lower socioeconomic status (SES) respondents. This is consistent with benchmarks reducing inaccurate responses caused by differential understandings of the scale on which economic quantities are measured.

In contrast to the reported gas prices, many observable characteristics systematically associate with respondents' reports of the unemployment rate. Consistent with the mecroeconomic voting theory of Ansolabehere, Meredith, and Snowberg (2011a), we show in four election surveys in three different years that observable characteristics associated with a higher risk of unemployment increase reported unemployment rates. Moreover, reported unemployment rates rose in lockstep with the actual unemployment rate between 2008 and 2009.

Finally, we demonstrate that quantitative questions can help us better understand the sources of partisan differences in responses to qualitative questions (Wlezien, Franklin, and Twiggs 1997; Anderson, Mendes, and Tverdova 2004; Evans and Andersen 2006; Evans and Pickup 2010). Combining information about economic perceptions with qualitative evaluations allows us to infer the criteria respondents use to judge economic performance. Our results suggest that partisan differences result from biases in the reporting of economic assessments, whether quantitative or qualitative. Moreover, quantitative questions evince less partisan bias than corresponding qualitative questions.

2 Why Ask About Quantities?

It is perhaps surprising that public opinion research focuses on qualitative—rather than quantitative—survey questions, given that many theories are rooted in numbers. For example, the theory of economic voting, and aggregate-level studies about it, primarily focus on how political support for the incumbent party changes with various economic quantities, such as GDP growth, the unemployment rate, and inflation (Kramer 1971; MacKuen, Erikson, and Stimson 1992; Erikson, MacKuen, and Stimson 2002; van der Brug, van der Eijk, and Franklin 2007). However, public opinion research into the individual-level mechanisms of economic voting has focused on the qualitative, retrospective economic evaluation discussed in the introduction. Supporting this focus are concerns that quantitative questions are too cognitively difficult for survey respondents.

This section details the potential advantages of quantitative questions. We then explore the concerns that have prevented a wider adoption of quantitative questions, and briefly preview how our research addresses these concerns through the careful design of quantitative questions.

2.1 *Benefits of Quantitative Questions*

Quantitative questions have four main advantages over qualitative questions. Quantitative questions allow researchers to separate respondents' perceptions of conditions from their judgments of those perceptions; they allow open responses, allow perceptions to be compared with actual conditions, and place respondents' responses on a standard scale.

Theories of economic voting posit that voters form *perceptions* of economic performance, and then *judge* these perceptions relative to some benchmark.[5] Quantitative questions allow researchers to separately elicit information about these two steps. Without good measures of economic perceptions, researchers risk incorrectly both calculating the amount of economic voting and a poor understanding of the mechanisms underlying it (Kramer 1983; Wlezien, Franklin, and Twiggs 1997;

[5]The benchmark may be a fixed standard in retrospective voting models, or depend on perceptions of opposition candidates in prospective models.

Anderson, Mendes, and Tverdova 2004; Evans and Andersen 2006; Evans and Pickup 2010). Moreover, evaluations of complex issues, like the economy, are often multidimensional (Zaller and Feldman 1992). As economic quantities are very specific, quantitative questions allow researchers to measure perceptions of the economy on multiple dimensions, and, in combination with a respondent's overall, qualitative evaluation, understand how these multiple perceptions are judged.

A second advantage of quantitative questions is that response is *open*—that is, response options can be relatively unconstrained. In contrast, qualitative questions are often *closed*—that is, constrained—to keep responses relevant and to reduce the cost of coding answers (Schuman and Presser 1981).[6] Open questions offer two advantages over closed questions. First, responses to closed questions are affected by the presentation and choice of response categories (Schwarz et al. 1985; Rockwood, Sangster, and Dillman 1997). Second, closed questions limit the amount of variation in responses (Groves et al. 2004). For example, individuals who perceive the unemployment rate to be 7% and 10%, respectively, may both report that the economy has "gotten worse," even though those who think it is 10% believe that there is almost 50% more unemployment. In practice, this may result in a large loss of variation: on the 2008 American National Election Survey (ANES), 90% of the respondents reported that the economy had "gotten worse." Increasing the amount of variation captured by survey responses is particularly useful in understanding how economic information affects economic perceptions (Manski 2004).[7]

Third, as quantitative questions often have an objectively correct answer, researchers can compare respondents' answers with actual conditions. For example, Hetherington (1996) shows that more media consumption is associated with worse economic evaluations. However, this study could not determine which group is more or less accurate in its perceptions.[8] This contrasts with Ansolabehere, Snowberg, and Snyder (2005), who use quantitative questions to show that low media consumption is associated with more accurate perceptions of the amount of money spent in congressional campaigns.

A fourth advantage of quantitative questions is that they facilitate comparisons across respondents. Survey respondents are likely to differ in their interpretations of what is meant by verbal phrases (Beyth-Marom 1982; Wallsten 1986; Wright, Gaskell, and O'Muircheartaigh 1994). For example, research shows wide variation in the numerical interpretations of various verbal expressions of uncertainty: the probability assigned to the term "likely" varies from 42% at the 10th percentile to 81% at the 90th percentile (Beyth-Marom 1982). Such interpretation differences may cause two individuals with identical underlying perceptions to give different answers to qualitative questions. In contrast, quantities like gas prices and the unemployment rate are measured on a well-defined scale.

2.2 *Can Survey Respondents Handle Numbers?*

Despite these potential advantages, quantitative questions are infrequently used on political surveys.[9] The focus on qualitative questions comes largely from concerns that quantitative questions are too cognitively complex for survey respondents. This concern manifests itself in a number of ways: from a general perception that survey respondents cannot "handle" or do not like quantitative questions, to more specific mechanisms about how these cognitive costs create biases in

[6]It is possible to construct hybrid questions that put verbal labels on specific quantities or ranges of quantities. See Juster (1966) for an example.

[7]This potential variation, however, may not be fully utilized as respondents tend to choose focal numbers, like multiples of five, when reporting quantities (Tourangeau and Smith 1996; Tourangeau et al. 1997).

[8]Qualitative questions are occasionally used to identify inaccurate perceptions. For example, Bartels (2002) shows that a substantial percentage of the respondents on the 1988 ANES report that inflation got worse over the course of the Reagan administration when, objectively, inflation was significantly lower at the end of Reagan's term than at the beginning.

[9]Quantitative questions are more frequently used on economic surveys. For example, the Michigan Survey of Consumer Behavior asks both quantitative and qualitative questions about inflation expectations.

responses. Additionally, the fact that quantitative questions put responses on a common scale can turn from asset to hindrance if respondents have different understandings of that scale.[10]

General concerns that quantitative questions are too cognitively demanding are found in research showing that survey respondents report wildly inaccurate assessments of quantities, such as the population of minority groups, or the percentage of the federal budget spent on welfare and foreign aid (Nadeau, Niemi, and Levine 1993; Kuklinski et al. 2000; Gilens 2001). Moreover, respondents indicate a preference for communicating numerical values qualitatively rather than quantitatively, although there is some heterogeneity (Erev and Cohen 1990; Wallsten et al. 1993). Together, these findings question individuals' abilities to accurately report numbers (see Moxey and Sanford 2000 for a summary).

Even if respondents are capable of providing numerical assessments, there are concerns that they may not have the motivation to do so. While motivation is a general problem in survey research, two factors make the problem more acute when asking about quantities. First, *survey satisficing*—the behavioral pattern of providing acceptable, rather than optimal, answers—increases when assessments are difficult to perform and report (Krosnick 1991). Moreover, some respondents feel compelled to provide answers to factual questions, rather than stating they "don't know" (Nadeau and Niemi 1995). If these respondents lack the motivation to produce responses that accurately reflect their perceptions, then these additional answers may introduce unwanted noise.[11] For example, Bruine de Bruin et al. (2000) find that survey respondents reporting a probability of 0.5 are often expressing general uncertainty, rather than a belief of a 50% chance.

Despite these concerns, we find that in Section 3, when respondents are asked about quantities they have direct experience with, such as gas prices, their responses are extremely accurate. This contradicts the hypothesis that survey respondents are generally unable to report numbers, either because they are incapable or because they lack motivation. Yet, there still may be concerns that certain types of individuals, like those with less education, may find it difficult to accurately report quantities. However, few covariates systematically associate with the accuracy of gas prices assessments. Thus, there appears to be widespread ability to accurately report familiar economic quantities.

2.3 *Do Respondents Really Use a Common Scale?*

An advantage of quantitative questions is that they may put survey responses on a common scale. A common scale is useful as it allows researchers to interpret higher reported unemployment rates as higher perceptions of unemployment.[12] However, this interpretation may not be valid if respondents are unfamiliar with the scale on which a quantity is measured. Indeed, 23%, 34%, and 40% of a nationally representative sample report that they have never heard of the unemployment rate, Consumer Price Index, and Gross Domestic Product, respectively (Curtin 2007). Moreover, the percentage of low SES respondents who do not know these quantities is substantially larger.[13] This leads to concerns that heterogeneity in respondents' reports of technical economic quantities may be due to different levels of knowledge about the definition of those quantities, rather than different perceptions of the conditions that underly those quantities.

Although Section 4 shows this concern is valid, the same section also shows that careful question design produces patterns consistent with respondents using a common scale. In particular, we show that providing benchmarks in questions about quantities—in this case, information on the

[10]We advocate the use of benchmarks to put respondents on a common scale; see Section 4.

[11]On the other hand, if quantitative questions motivate respondents to produce accurate responses because they know there is a wrong answer, this may be an added benefit of quantitative questions. Bullock, Gerber, and Huber (2010) find that providing financial incentives reduces, but does not eliminate, partisan differences in reported perceptions of facts.

[12]Moreover, if subjects perceive a scale linearly, then those that report 10% unemployment can be interpreted as perceiving twice as much unemployment as those that report 5% unemployment. However, as none of our discussion requires a linear scale, we discuss our results in terms of the weaker condition of a monotonic, but not necessarily linear, common scale.

[13]For example, the percentage of people at the bottom, middle, and top tercile of the income distribution who have not heard of the unemployment rate is 38%, 21%, and 11%, respectively.

historical tendencies of the unemployment rate—reduces both the number of outlying responses, and differences between the reports of higher and lower SES respondents. This is consistent with benchmarks producing a common scale for all respondents.

A final concern is whether there is useful information in respondents' assessments of quantities. While this is a tricky question to answer, we present a variety of evidence demonstrating that reported economic quantities do reflect perceptions of the economy. First, changes in gas prices and the unemployment rate are reflected in changes in quantitative assessments over time. Second, the strong relationship between qualitative and quantitative economic assessments suggests responses to both, which are driven by common underlying factors. Third and finally, consistent with the mecroeconomic voting theory of Ansolabehere, Meredith, and Snowberg (2011a), cross-sectional variation in reported unemployment rates reflects differences in respondents' probability of being unemployed.

3 Survey Respondents Can Handle Numbers

The previous section highlights some skepticism about respondents' abilities to report perceptions of *any* numerical quantity. To address this skepticism, we demonstrate that respondents can accurately report the average price of gas in their state.[14] We focus on the price of gas because it is one of the most frequently encountered numbers in a respondent's everyday environment. Gasoline is also an important component of U.S. household consumption, and gas prices have an independent effect on presidential approval (Fox 2009).[15] Finally, actual gas prices are easy to observe, which facilitates comparison between respondents' assessments and actual prices.

As gas prices are salient and well understood by most people, we expect that perceptions will be quite accurate. Observing inaccurate assessments of gas prices would therefore suggest that there are few or no economic quantities that respondents can accurately report. In contrast, if individuals can accurately report the price of gas, then inaccurate assessments of other economic quantities are not simply a reflection of general innumeracy.

We asked respondents the following question on the 2006 and 2008 ANES, and on the 2008 Cooperative Congressional Election Survey (CCES):

> What is your best guess of the average price of a gallon of regular unleaded gasoline across all of [YOUR STATE] today?

Respondents accurately report the average price of gas in their state of residence. Figure 1 plots a histogram of the reported price of gas in each of the three surveys. We observe that the modal response in each survey tracks the actual average price of gas. It is particularly impressive that the large drop in gas prices between October and November 2008 is quickly reflected in survey responses.

Figure 2 plots the difference between respondents' reported price of gas and the actual average price of gas in the respondents' state in the month of the survey. This difference is distributed relatively evenly around a mean of about zero. On all three surveys, the average *bias*—the difference between the reported and actual price of gas—is slightly positive.[16] The mean response overestimates the price of gas by approximately ten cents on all the three surveys, whereas the median response overestimates the price by between six and ten cents.

Overall, the accuracy of assessments of gas prices shows survey respondents are capable of accurately answering quantitative questions about a familiar quantity. The *accuracy* of responses—the negative of the absolute difference between the reported and actual price of gas—is –20 to –30 cents across the three surveys, with the median deviation being slightly lower.[17] Given

[14]All of the data files to replicate our analysis can be found in Ansolabehere, Meredith, and Snowberg (2012).

[15]U.S. Bureau of Labor Statistics news release USDL-10-1390 indicates that gasoline and motor oil account for about 5% of American household expenditures between 2007 and 2009.

[16]We bottom- and top-code the bias at −\$1.00 and \$1.00, respectively.

[17]Note that as accuracy is defined as the negative of the absolute difference, accuracy increases as the absolute difference between a respondent's report and the actual price decreases. Gasoline price data are from the U.S. Department of Transportation.

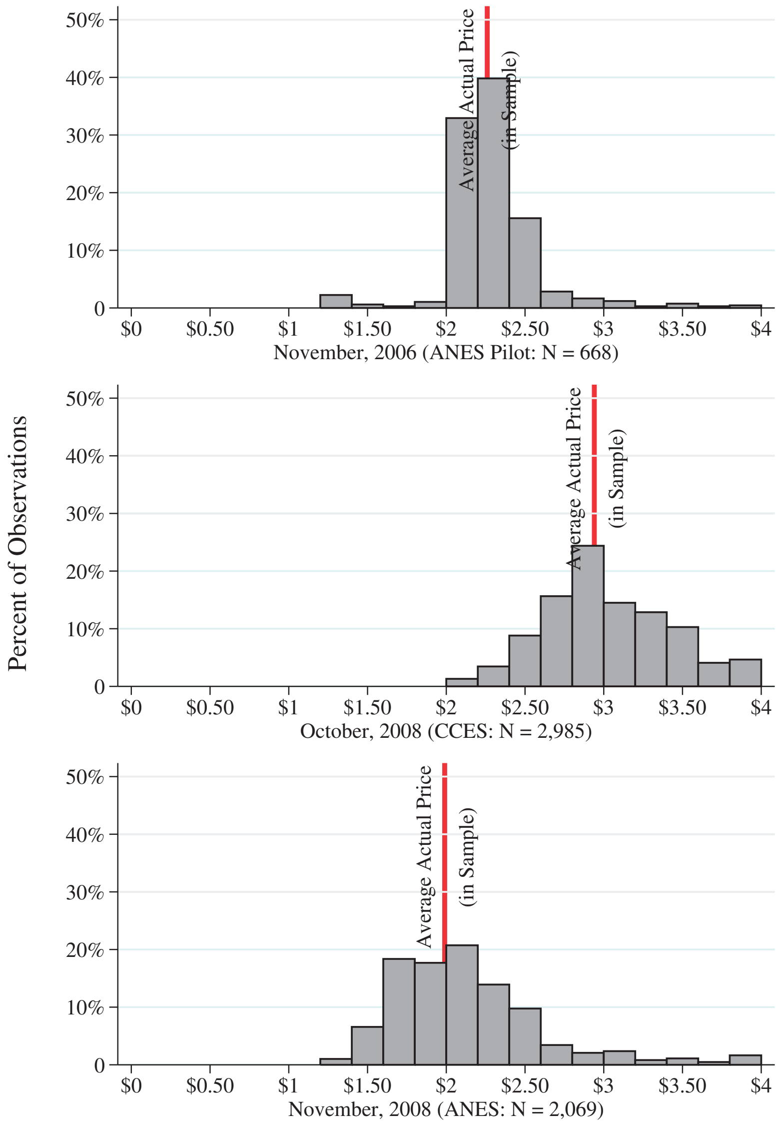

Fig. 1 Distribution of reports of average state gas price. Top and bottom 1% of the responses are set to the ninety-ninth- and first-percentile answer, respectively.

that the average price of gas in the sample ranges from two to three dollars, this suggests a 10%–15% average difference between the reported and the actual price of gas.

The above analysis shows respondents are accurate, on average. However, there may be considerable heterogeneity in respondents' abilities to perceive and report numerical quantities. For example, more education may help respondents report more accurate numeric assessments. Table 1 explores this possibility by using a multivariate regression to determine the correlates of bias and accuracy of reported gas prices in all three surveys.

The table shows that different types of people do not systematically differ in their ability to accurately answer questions about a familiar quantity. There are few covariates that are consistent predictors of either the bias or accuracy of assessments of the price of gas. Moreover, most of the coefficients are both small in magnitude and insignificant. There are two exceptions to this general pattern: Black and Hispanic respondents are significantly less accurate than non-Black, non-Hispanic respondents. Additionally, in contrast to the findings in the next section about reported unemployment rates, there is little partisan difference in assessments of the price of gas.

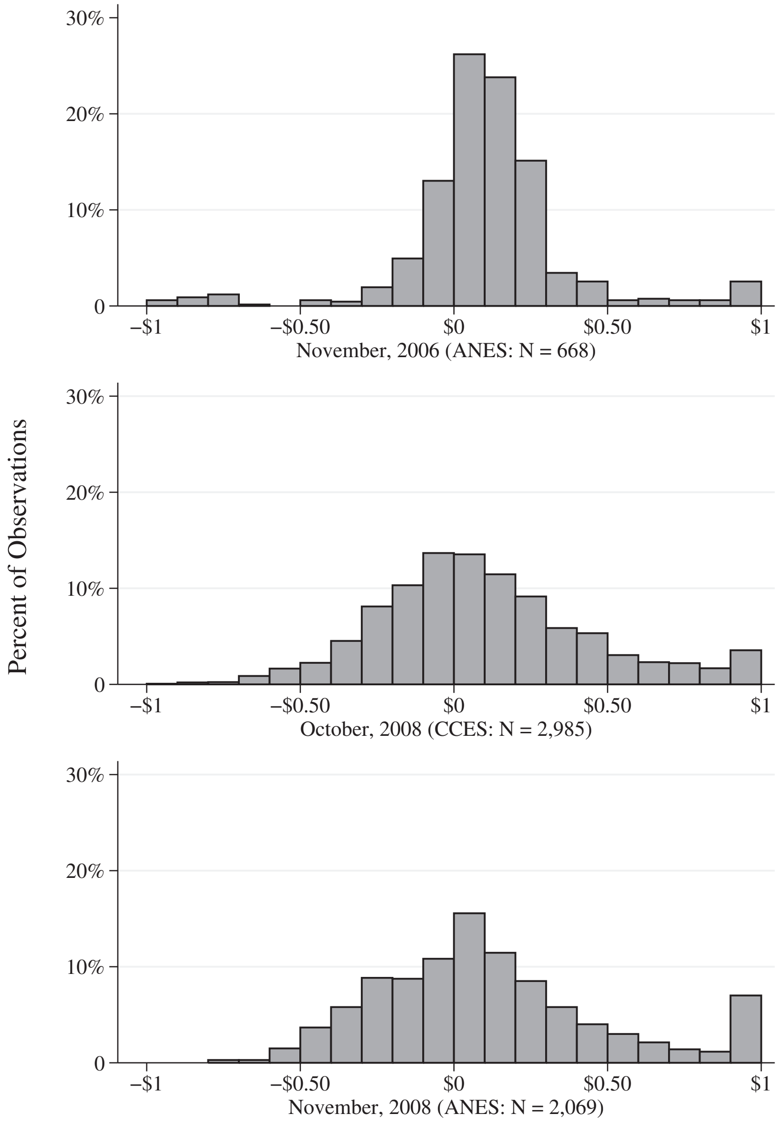

Fig. 2 Distribution of bias in reports of average state gas price. Errors are trimmed to be between –$1 and $1.

Finally, although education is not significantly correlated with the accuracy of gas price assessments, other activities that increase a respondent's familiarity with a numerical quantity are correlated with responses. In particular, the 2006 ANES asked the respondents both how many times a week they drove a car, and how often they noticed gas prices. Each additional day that a respondent reports driving makes them 0.8 cents (SE = 0.4) more accurate, while each additional time that a respondent reports noticing gas prices makes them 1.6 cents (SE = 0.4) more accurate (Ansolabehere, Meredith, and Snowberg 2011b). These point estimates suggest that every day that a respondent drives and notices gas prices is associated with a 12.5% improvement in his or her accuracy. This examination would not be possible with qualitative questions.

4 The Benefits of Benchmarking

Understanding individuals' perceptions of economic quantities is potentially quite useful for understanding the mechanisms underlying economic voting. However, as outlined in Section 2.3, many respondents are unfamiliar with complicated quantities like the unemployment rate and the scale on

Table 1 Correlates of reported average gas price

	2006 ANES (n = *668*)		*2008 CCES* (n = *2985*)		*2008 ANES* (n = *2069*)	
Dependent variable	*Bias*	*Accuracy*	*Bias*	*Accuracy*	*Bias*	*Accuracy*
Mean value	0.101	−0.193	0.102	−0.274	0.117	−0.292
Median value	0.101	−0.131	0.058	−0.202	0.078	−0.215
Democrat	−0.020 (0.027)	−0.007 (0.021)	0.069 (0.017)***	−0.038 (0.012)***	0.016 (0.025)	0.006 (0.018)
Independent	−0.015 (0.028)	0.012 (0.022)	0.069 (0.017)***	−0.039 (0.012)***	0.024 (0.025)	−0.013 (0.018)
Age (years)						
18–44	0.070 (0.031)**	−0.065 (0.024)***	−0.045 (0.019)**	−0.002 (0.013)	−0.053 (0.024)**	0.032 (0.017)*
45–64	0.059 (0.029)**	−0.037 (0.022)*	−0.019 (0.018)	0.015 (0.012)	−0.041 (0.024)*	0.024 (0.017)
Female	−0.011 (0.022)	−0.017 (0.017)	0.026 (0.013)**	−0.039 (0.009)***	0.010 (0.017)	−0.019 (0.012)
Married	0.012 (0.024)	0.004 (0.019)	−0.057 (0.014)***	0.035 (0.010)***	−0.070 (0.018)***	0.037 (0.013)***
Black	−0.035 (0.037)	−0.110 (0.029)***	0.037 (0.023)	−0.035 (0.016)**	0.025 (0.023)	−0.035 (0.016)**
Hispanic	0.122 (0.055)**	−0.190 (0.042)***	0.055 (0.023)**	−0.046 (0.016)***	0.037 (0.022)*	−0.067 (0.015)***
Some college	0.004 (0.028)	0.019 (0.021)	0.077 (0.016)***	−0.017 (0.011)	−0.011 (0.020)	0.017 (0.014)
Bachelor's degree	−0.016 (0.028)	0.019 (0.022)	0.107 (0.017)***	−0.032 (0.012)***	0.029 (0.023)	−0.007 (0.016)
Income less than $20,000	0.047 (0.048)	0.016 (0.037)	0.021 (0.030)	−0.026 (0.021)	0.041 (0.040)	−0.026 (0.028)
Income between $20,000 and $40,000	0.000 (0.043)	−0.021 (0.033)	−0.057 (0.025)**	0.024 (0.018)	−0.030 (0.038)	0.012 (0.027)
Income between $40,000 and $80,000	−0.016 (0.037)	−0.006 (0.028)	−0.069 (0.022)***	0.027 (0.016)*	−0.059 (0.037)	0.033 (0.026)
Income between $80,000 and $120,000	−0.032 (0.041)	−0.028 (0.031)	−0.060 (0.025)**	0.040 (0.017)**	−0.026 (0.039)	−0.004 (0.028)
Unemployed	−0.069 (0.061)	−0.057 (0.047)	0.092 (0.026)***	−0.041 (0.018)**	0.074 (0.032)**	−0.025 (0.022)
Days driving per week	−0.018 (0.006)***	0.008 (0.004)*				
Notice gas prices per week	−0.009 (0.005)**	0.016 (0.004)***				
Constant	0.187 (0.061)***	−0.232 (0.047)***	0.070 (0.032)**	−0.244 (0.023)***	0.172 (0.046)***	−0.309 (0.033)***

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% levels two-tailed, respectively. Regressions also include minor and missing party, church attendance, union membership, and missing income indicators. The omitted categories are Republicans, age ≥65 years, male, not married, White, ≤12 years of education, $120,000 or more for income, employed, not in union, and do not attend church. All regression estimated using ordinary least-squares (OLS).

which those quantities are measured. As a result, individuals may report their perceptions of the percentage of people who are not working when asked about their perceptions of the unemployment rate. Thus, heterogeneity in responses may be driven by heterogeneity in individuals' understanding of what the unemployment rate is meant to measure. While such heterogeneity may be useful for measuring familiarity or knowledge of the unemployment rate, it is problematic to use these assessments to examine how perceptions of labor market conditions affect voter behavior.

This section advocates the use of benchmarks when asking about complex economic quantities. Benchmarks refer to reference values of the quantity in question that are embedded in the question wording, and ideally provide a sense of scale of the quantity to those who do not know it, without changing the perceptions of those who do.

We investigate the effect of including benchmarks when asking the following question, asked on the 2006 and 2008 ANES and the 2008 and 2009 CCES:

> As far as you know, what is the current rate of unemployment? That is, of the adults in the United States who wanted to work during the second week of October, what percent of them would you guess were unemployed and looking for a job?

The following benchmarks were sometimes added to the above question on the 2008 and 2009 CCES:

> The unemployment rate in the United States has varied between 2.5% and 10.8% from 1948 till today. The average unemployment rate during that time was 5.8%.

Comparing the responses of those who were exposed and not exposed to the benchmarks shows that including benchmarks reduces the heterogeneity in reported unemployment rates, especially among low SES respondents. Yet, adding these benchmarks does not reduce meaningful variation in reported unemployment rates. We show these results in a randomized experiment with approximately 4000 respondents, administered as part of the 2008 CCES. Three-quarters of these respondents were asked about the unemployment rate with a question that included the benchmarks, and the remaining one-quarter were asked the same question without them. We refer to these as the benchmarked and non-benchmarked samples, respectively.[18]

4.1 *A Simple Model of Survey Response*

The lack of familiarity with complicated economic quantities, like the unemployment rate, raises questions about whether or not assessments of these quantities are on a common scale. For example, respondents who are not familiar with the technical definition of the unemployment rate may report their perception of the percent of people out of work, rather than those that are out of work *and* looking for a job. That is, they may report their perception of the labor force nonparticipation rate, rather than their perception of the unemployment rate.[19] Thus, heterogeneity in assessments of the unemployment rate may primarily reflect heterogeneity in definitions and scaling, rather than heterogeneity in perceptions of unemployment.

A well-designed question would substantially reduce heterogeneity in responses due to different understandings of the quantity of interest, without reducing meaningful variation.[20] To ascertain whether including our benchmarks satisfies these goals requires a theory of what exactly constitutes meaningful variation, and which responses are due to lack of a common scale. Meaningful

[18]Due to an error administrating the survey, the benchmarked sample allowed respondents to enter unemployment rates to the tenth of a percent, whereas the unframed sample was only able to enter them to the nearest percent. Truncating or rounding the responses in the benchmarked sample has almost no effect on the results presented here, although we cannot rule out that this biased responses by serving as a cue about the expected precision of response.

[19]A large literature also shows that individuals struggle to estimate low probabilities, like the probability that people are unemployed (Kahneman and Tversky 1979; Snowberg and Wolfers 2010).

[20]In a model of on-the-spot survey response (Converse 1964; Zaller 1992), a well-designed question would provide information that would not influence a respondent's perception of the economy, but instead help them translate that perception into a common numeric scale. To put this another way, when asking a respondent to rate the President's foreign policy on a scale of 1–10, the respondent should not draw any information about foreign policy from the fact that a 1–10 scale is used.

variation is easy to define: it is variation due to differences in perceptions of the economic concept underlying the quantity. As this cannot be measured directly, the rest of this section builds a simple model of survey response that makes predictions about what would occur if a question's design reduces heterogeneity due to non-common scaling while maintaining meaningful variation.

We assume there are three types of survey respondents: those who know the scale on which the unemployment rate is measured, those who do not know the scale, and those who are not attempting to correctly answer the question. Benchmarks would ideally provide a sense of scale to those who do not know it, without reducing the meaningful variation among those who do.

This simple structure is enough to deduce two patterns that we should observe if our question is well designed. The first pattern is that more information about the definition and the tendencies of the unemployment rate should reduce the proportion of respondents reporting very large unemployment rates (say, >15%). In particular, including the precise definition of the unemployment rate should produce an initial reduction and the inclusion of benchmarks a further reduction.

Second, there should be a similar relationship between reported unemployment and qualitative economic evaluations among those who report an unemployment rate within the historical limits in *both* the benchmarked and non-benchmarked samples. Moreover, as only those who are not attempting to correctly answer the question provide responses greater than the historical maximum in the benchmarked sample, there should not be a significant relationship between reported unemployment rates and qualitative economic evaluations above the historical maximum.

To deduce a third pattern, we maintain an additional two hypotheses that refine the definition of what does—and does not—constitute meaningful variation. The argument in Section 2.3 suggests that low SES respondents will be particularly likely to have a different understanding of the scale on which unemployment is measured. Thus, we assume low SES respondents will be more likely to report a number closer to the labor force nonparticipation rate. Maintaining this hypothesis, reductions in the amount of low SES respondents, who report very high levels of unemployment, should be interpreted as evidence that the benchmarks are reducing heterogeneity due to different understandings of the unemployment rate.

We further maintain that meaningful variation follows the mecroeconomic voting hypothesis of Ansolabehere, Meredith, and Snowberg (2011a). Under this hypothesis, individuals gather local economic information to inform them about their future risks of unemployment. Thus, those who face a greater risk of unemployment should perceive the unemployment rate to be higher.

Taking these two maintained hypotheses together, we deduce a third pattern that should be observed if our benchmarks are well designed: there should be substantial heterogeneity in the non-benchmarked sample, with those at a greater risk of unemployment and those with low SES providing answers that are, on average, much higher. The benchmarked sample should provide responses consistent with their relative risk of unemployment, but the differences between high and low SES respondents should be substantially reduced.

The next three subsections evaluate the extent to which the above patterns are observed in data from a randomized experiment implemented as part of the 2008 CCES.

4.2 *Reduction of Very High Responses*

While we believe it is important to precisely define the quantity of interest, it is unlikely that this has much of an effect on survey responses. Compared to the previous work, we observe fewer very high responses: only 30% of the respondents in the unbenchmarked sample report unemployment rates above the historical high of 10.8%, compared with 50% in the previous studies by Conover, Feldman, and Knight (1986) and Holbrook and Garand (1996). However, this is likely due to differences in the population being surveyed or the mode of survey, or the changes across time in the salience of unemployment.[21] Assessments above the historical high of 10.8% are particularly concentrated among low SES respondents.

[21]In particular, a simple experiment, run on Mechanical Turk, found that the distribution of responses was the same whether or not we provided a definition of the unemployment rate. We do not provide an opt-out prompt when asking

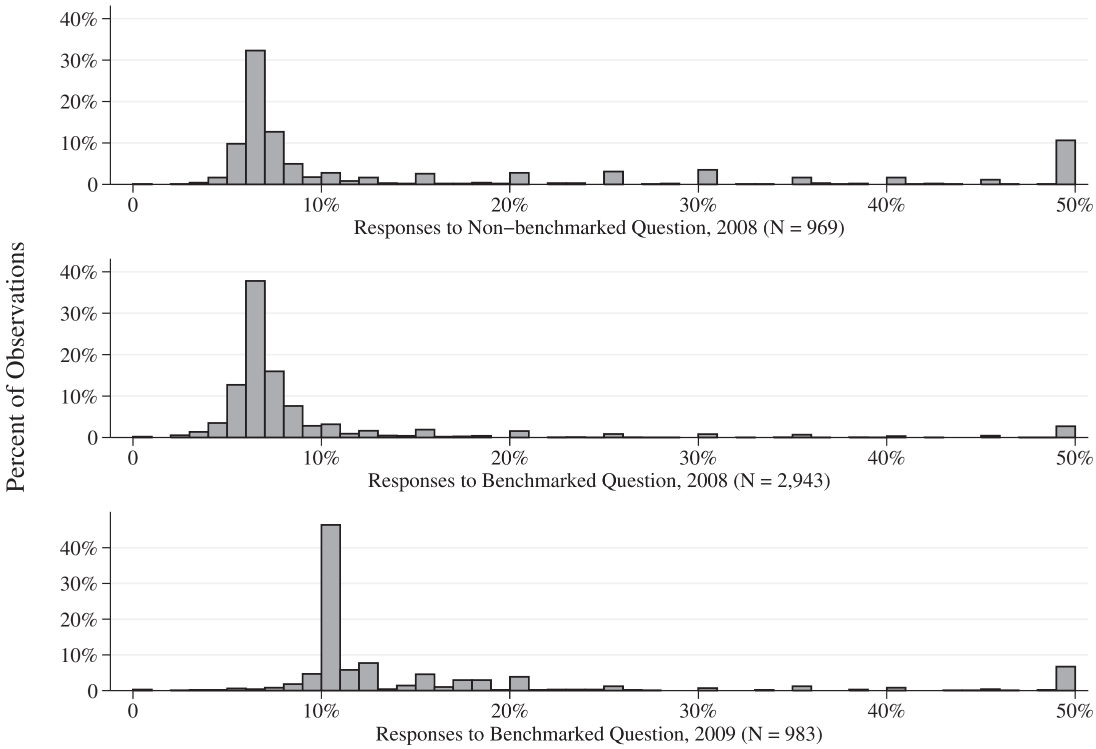

Fig. 3 Distribution of reports of the national unemployment rate.

The top panel of Fig. 3 suggests that a substantial proportion of the population is roughly aware that 6.5% was the true unemployment rate in October 2008, with just under one-half of the sample reporting unemployment rates of between 6% and 7%. However, another one-third of the responses are higher than 10.8%, the highest unemployment rate since the Bureau of Labor Statistics unemployment series began in 1948. This implies either that there are many respondents who believe the unemployment rate is at least twice as high as it actually is, or that many of the extreme answers come from a poor understanding of the scale of the unemployment rate.

Next, we add benchmarks to the question. The benchmarks we construct provide the respondent with a sense of the smallest and largest feasible unemployment rates by telling them the historical highs and lows of unemployment during the previous 60 years. It also gives guidance on the average level of unemployment across time.

As shown in the middle panel of Fig. 3, providing respondents with benchmarks substantially reduces, but does not eliminate, the heterogeneity in assessments of the unemployment rate. Just as in the non-benchmarked sample, slightly fewer than 50% of the respondents report unemployment rates of 6% and 7%. However, the percentage of answers above the largest historical unemployment rate is reduced from over 30% to ∼15% of the sample. Thus, the first pattern conjectured above holds: more information about the definition and tendencies of the unemployment rate reduces the proportion of respondents reporting very large unemployment rates.

4.3 *Comparison with Qualitative Evaluations*

If respondents view higher unemployment as a sign the economy is getting worse, then there should be a relationship between perceptions of unemployment and qualitative economic evaluations, like

these questions, and request that respondents report a number if they try to skip the question. Respondents can opt out by attempting to skip the question a second time. Our resulting response rates are 99% in the benchmarked and 98% in the non-benchmarked samples. This protocol is consistent with Curtin's (2007) finding that providing an opt-out significantly reduced responses, without significantly increasing the accuracy of responses.

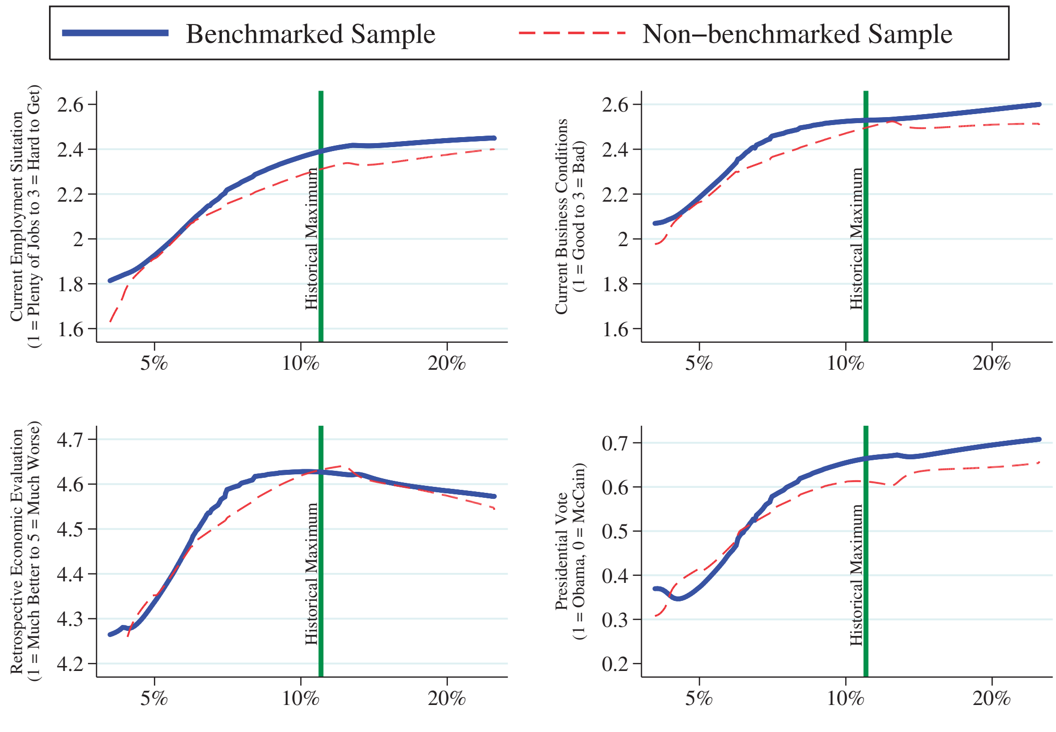

Fig. 4 Relationship between qualitative and quantitative evaluations. Each figure is Loess smoothed using a bandwidth of 0.8 (Cleveland, Devlin, and Grosse 1988).

the standard retrospective question discussed in the introduction. That is, if reported unemployment rates are capturing variation in unemployment perceptions, we expect that respondents who report higher unemployment rates will report more negative evaluations of the aggregate economy.[22]

This subsection examines how including benchmarks affects the estimated relationship between the reported unemployment rate and retrospective economic evaluations. As noted above, if our simple model of survey response is correct, and the benchmarks are well designed, then there should be a similar relationship between reported unemployment and qualitative economic evaluations among those who report an unemployment rate within the historical limits in both the benchmarked and non-benchmarked samples. Moreover, there should be no relationship between reported unemployment rates and qualitative economic evaluations above the historical maximum in the benchmarked sample.

These predicted patterns are consistent with the data shown in Fig. 4. In particular, the top panels show the (Loess smoothed) relationship between respondents' reported unemployment rate assessments and their qualitative assessment of business conditions and the employment situation, elicited by asking:

> How would you rate the present general business conditions in your area? Good, normal, or bad?
> What would you say about available jobs in your area right now? There are plenty of available jobs in my area, there are not so many available jobs in my area, or jobs in my area are hard to get?[23]

[22]Of course, unemployment perceptions and the threshold at which a respondent will say the economy has gotten worse are not necessarily independent. The answers to these questions could be uncorrelated or negatively correlated, if respondents who have a higher threshold also systematically perceive higher levels of unemployment. We view this as an unlikely possibility, and one that is contradicted by the data in the next section.

[23]These are the standard current business condition and employment questions from the Conference Board survey. Note that these questions ask about local economic conditions, as opposed to the retrospective economic evaluation, which asks about national economic conditions.

We code the extremes of "good" and "plenty of jobs" as 1, and "bad" and "hard to get" as 3. All data come from the 2008 CCES.[24] The bottom two panels show the relationship between respondents' reported unemployment perceptions and their retrospective economic evaluation, and vote choice, respectively.[25]

All four panels are plotted on a log scale to focus on the variation within the historical boundaries. For all three qualitative assessments, the relationship between the qualitative assessment and unemployment rate assessment within that range is quite similar in the benchmarked and non-benchmarked sample. The same pattern holds for vote choice. Statistical tests, contained in the Supplementary Appendix, confirm this appearance. Moreover, in all four panels, there appears to be no relationship between qualitative assessments or vote choice and unemployment rate assessments above the historical maximum in the benchmarked sample.[26] Once again, statistical tests contained in the Supplementary Appendix confirm this appearance.

Thus, the second pattern conjectured above holds: including benchmarks does not affect the relationship between unemployment rate assessments and qualitative assessments in the range defined by historical limits. Moreover, there is no relationship between unemployment rate assessments and qualitative assessments above the historical maximum.

4.4 *Mecro-patterns of Unemployment Perceptions*

Finally, we examine how introducing the benchmark changes differences in assessments of unemployment between groups. We show that although including benchmarks reduces the heterogeneity between groups, it does not affect the overall pattern of responses: namely, respondents in groups with a higher risk of unemployment report higher average assessments of the unemployment rate. These patterns are thus consistent with the third set of patterns deduced in Section 4.1, and hence, with the question being well designed under our maintained hypothesis. Moreover, these patterns are substantially different from patterns found in qualitative evaluations, leading to the conclusion that quantitative questions capture additional, meaningful variation.

Data from the benchmarked and non-benchmarked sample are examined in Table 2. Columns 1 and 2 present coefficients from least absolute deviation (LAD) regressions of the reported unemployment rate, in the benchmarked and non-benchmarked sample, respectively, on a host of covariates. LAD regressions, sometimes referred to as median regressions, minimize the impact of outliers on estimated coefficients. The coefficients in these regressions report how the median value of reported unemployment associates with a one-unit change in the independent variable, holding all else equal. For example, the median respondent making under $20,000 a year reports unemployment rates that are 0.77 percentage points (SE = 0.21) higher in the benchmarked sample, as compared to 13.5 percentage points (SE = 5.8) higher in the non-benchmarked sample. This is consistent with the general pattern: the same variables significantly relate to unemployment assessments in both the benchmarked and non-benchmarked samples. However, the magnitudes are sometimes dramatically different.

Adding benchmarks generally reduces differences between groups in their assessments of unemployment. To perform the comparison, we must first recode each respondent's unemployment assessment as the percentile of the distribution that his or her unemployment assessment falls

[24]While it would be interesting to compare quantitative and qualitative employment questions in 2009 as well, in that year, 90% of the respondents reported jobs were "hard to get." This shows one of the benefits of using quantitative questions: the fact that responses are open means that they measure variation even in extreme circumstances.

[25]The questions were asked in the following sequence: retrospective economic evaluation, business conditions, qualitative employment, vote choice, and unemployment rate. As the unemployment question was asked last, the benchmark could not have affected answers to these other questions. Additional questions were asked before, after, and in between these questions.

[26]One additional result of note is that there is no relationship between qualitative assessments or vote choice and unemployment rate assessments above the historical maximum in the non-benchmarked sample either. Our model implies that responses above the historical maximum in the non-benchmarked sample come from a combination of people who do not know the scale on which unemployment is measured and people who are not attempting to answer the question. This pattern thus suggests that those who do not know the scale on which the unemployment rate is measured do not share a common misunderstanding of the scale.

https://doi.org/10.1093/pan/mps031 Published online by Cambridge University Press

Table 2 Correlates of reported unemployment rate in 2008 CCES

Dependent variable	*Level (LAD)*			*Percentile (OLS)*		
Sample	*Benchmarked*	*Non-benchmarked*	*Difference*	*Benchmarked*	*Non-benchmarked*	*Difference*
Democrat	0.55 (0.07)***	1.50 (0.53)***	−0.95 (0.54)*	12.09 (1.23)***	9.07 (1.57)***	3.02 (1.99)
Independent	0.31 (0.06)***	0.50 (0.38)	−0.19 (0.39)	8.12 (1.00)***	3.74 (1.75)**	4.38 (2.01)**
Age (years)						
18–24	0.69 (0.48)	6.50 (6.03)	−5.81 (6.04)	12.84 (2.62)***	15.95 (4.25)***	−3.11 (4.99)
25–44	0.49 (0.08)***	1.50 (0.77)*	−1.01 (0.78)	10.44 (1.61)***	10.10 (2.37)***	0.34 (2.87)
45–64	0.20 (0.05)***	0.50 (0.51)	−0.30 (0.51)	4.95 (1.32)***	5.74 (1.79)***	−0.79 (2.22)
Married male	0.17 (0.07)**	1.00 (0.60)*	−0.83 (0.61)	0.01 (1.48)	1.62 (2.83)	−1.61 (3.20)
Unmarried female	0.66 (0.10)***	3.00 (1.24)**	−2.34 (1.24)*	11.45 (1.50)***	12.16 (3.21)***	−0.71 (3.54)
Married female	0.62 (0.11)***	3.00 (1.06)***	−2.38 (1.07)**	10.31 (1.34)***	17.13 (2.49)***	−6.83 (2.82)**
Black	0.61 (0.23)***	2.00 (2.98)	−1.39 (2.99)	6.90 (1.41)***	6.79 (2.92)**	0.11 (3.24)
Hispanic	0.10 (0.13)	1.50 (1.54)	−1.40 (1.54)	0.39 (2.15)	4.37 (2.46)*	−3.98 (3.27)
Some college	−0.30 (0.07)***	−1.50 (0.89)*	1.20 (0.89)	−5.21 (1.14)***	−8.25 (2.62)***	3.04 (2.86)
Bachelor's degree	−0.38 (0.07)***	−2.00 (0.93)**	1.62 (0.94)*	−7.35 (1.28)***	−12.99 (2.43)***	5.64 (2.75)**
Income less than \$20,000	0.77 (0.21)***	13.50 (5.79)**	−12.73 (5.80)**	10.46 (2.39)***	19.95 (5.80)***	−9.50 (6.27)
Income between \$20,000 and \$40,000	0.41 (0.10)***	3.00 (1.62)*	−2.59 (1.62)	6.55 (1.69)***	14.45 (3.51)***	−7.90 (3.90)**
Income between \$40,000 and \$80,000	0.05 (0.06)	0.50 (0.58)	−0.45 (0.58)	2.88 (1.53)*	6.17 (2.53)**	−3.29 (2.96)
Income between \$80,000 and \$120,000	0.02 (0.06)	0.50 (0.56)	−0.48 (0.56)	0.72 (1.71)	2.98 (2.78)	−2.25 (3.27)
Unemployed	0.20 (0.18)	1.00 (2.16)	−0.80 (2.17)	3.22 (1.80)*	1.94 (2.99)	1.28 (3.49)
State unemployment	0.12 (0.02)***	0.00 (0.13)	0.12 (0.13)	2.31 (0.36)***	0.52 (0.57)	1.79 (0.68)***
Constant	5.05 (0.16)***	4.50 (1.37)***		16.82 (3.62)***	24.04 (5.60)***	

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% levels two-tailed, respectively, with SEs block bootstraped by state in LAD regressions and clustered by state in OLS regressions. Benchmarked: $n = 2943$; non-benchmarked: $n = 969$. Regressions also include minor and missing party, church attendance, union membership, and missing income indicators. The omitted categories are Republicans, ≥age 65 years, male, not married, White, ≤12 years of education, \$120,000 or more for income, employed, not in union, and do not attend church.

within his or her sample. This is necessary because, if our theory is correct, respondents in the two samples are implicitly using different scales. Thus, in this coding, hundred indicates that a respondent had the highest report in his or her sample and zero the lowest.

Columns 4 and 5 examine how changes in the percentile of respondents' reports (within their sample) are correlated with the same covariates as before. For example, in the non-benchmarked sample, holding all else equal, the average respondent making under $20,000 a year reports a 20.0 percentile (SE = 5.8) higher unemployment rate than the average respondent making more than $120,000 a year. In comparison, this shrinks to 10.5 percentiles (SE = 2.4) higher in the benchmarked sample. Column 6 reports that the difference between these estimates is 9.5 percentiles (SE = 6.3), although this difference is not statistically significant. Including benchmarks causes statistically significant reductions in the difference in reports between those who never attended college and those with a bachelor's degree, as well as those with low and high incomes. In addition, we estimate a statistically significant reduction in the difference between married women and unmarried men, who are the least and most likely gender–marriage combinations, respectively, to participate in the labor force. As we conjecture that those with less education, lower incomes, and those outside the labor force are less likely to be aware of the scale on which the unemployment rate is measured, this suggests that our benchmarks help put individuals on a common scale.[27]

A final question of interest here is whether we could have observed the same mecroeconomic patterns using only qualitative assessments. To determine this, in Table 3, we regress the same three qualitative assessments used in Fig. 4 on the same set of demographic controls in Table 2. The results are striking: the only macroeconomic pattern in Table 3 is found in different income groups' qualitative evaluation of the employment situation.[28] This shows that quantitative questions are able to capture additional meaningful variation that is difficult to observe in qualitative evaluations.

The only variables that consistently correlate with all three qualitative economic evaluations examined in Table 3 are partisan indicators. The next section uses both quantitative and qualitative questions to understand the sources of this partisan bias.

Thus, all three sets of patterns deduced above are consistent with the data. While this implies that our question is well designed under the maintained hypotheses, it can by no means prove that it is well designed. Our maintained hypothesis could be wrong, or there may be other implications contradicted by the data.

Finally, we note that including benchmarks has some potential drawbacks. The goal of benchmarks is to reduce the difference between respondents' perceptions of the economic conditions measured by the quantity of interest and their reported perceptions of the quantity—that is, to reduce measurement error. However, benchmarks may exacerbate, rather than reduce, measurement error. Some survey respondents might use information contained in the benchmark to form, rather than scale, their perceptions. This could occur if respondents know the current unemployment rate, but not the scale on which the unemployment rate is measured. Then, providing the scale would cause respondents to update on the state of the economy, and perhaps alter responses to subsequent questions (Blinder and Krueger 2004).

Benchmarks also could affect a respondent's ability or desire to express their true perception of the quantity. For example, our benchmarks may inhibit some respondents who know the scale on which the unemployment rate is measured from expressing their perception that the unemployment rate is above its post-1948 peak.

While our benchmarks certainly are not perfect, we do not find much evidence that they generate the forms of measurement error discussed in the previous paragraph. While many forms of

[27]There are concerns about the reliability of Internet surveys, such as the CCES (see Malhotra and Krosnick 2007; Yeager et al. 2009; Stephenson and Crête 2011 for an opposing view). To alleviate such concerns, we replicate the regressions in Table 2 on responses to a reported unemployment rate question on the 2006 and 2008 ANES in the Supplementary Appendix. On both surveys, the question did not contain any benchmarks, and the 2006 question differed slightly as it asked respondents their perceptions of the unemployment rate in their state of residence. It is reassuring that the results reported in Table A.2 of the Supplementary Appendix show very similar patterns in the ANES data to those observed in the CCES data. For more information on response rates and data reliability in the 2008 CCES, see Ansolabehere (2011), especially the guide.

[28]Conducting the analyses in Table 3 using an ordered probit produces qualitatively similar results.

Table 3 Correlates of qualitative economic assessments in 2008 benchmarked CCES

Dependent variable	*Retrospective economic evaluation*	*Business conditions*	*Employment situation*
Democrat	0.53 (0.03)***	0.43 (0.02)***	0.40 (0.03)***
Independent	0.31 (0.04)***	0.19 (0.03)***	0.22 (0.03)***
Age (years)			
18–24	−0.09 (0.06)	−0.20 (0.06)***	−0.02 (0.06)
25–44	−0.04 (0.04)	0.02 (0.04)	0.03 (0.04)
45–64	−0.00 (0.03)	0.07 (0.04)*	0.07 (0.03)**
Married male	−0.09 (0.03)***	0.01 (0.04)	0.01 (0.04)
Unmarried female	0.00 (0.04)	0.06 (0.03)*	0.02 (0.05)
Married female	−0.11 (0.03)***	−0.01 (0.04)	−0.01 (0.04)
Black	−0.02 (0.03)	−0.05 (0.04)	0.04 (0.04)
Hispanic	−0.11 (0.05)**	−0.07 (0.03)**	−0.02 (0.06)
Some college	0.03 (0.03)	0.01 (0.03)	−0.00 (0.04)
Bachelor's degree	0.03 (0.03)	0.00 (0.03)	−0.02 (0.03)
Income less than $20,000	−0.09 (0.06)	0.02 (0.06)	0.30 (0.08)***
Income between $20,000 and $40,000	−0.10 (0.05)**	0.02 (0.04)	0.24 (0.04)***
Income between $40,000 and $80,000	−0.03 (0.04)	−0.05 (0.04)	0.11 (0.04)***
Income between $80,000 and $120,000	−0.04 (0.04)	−0.09 (0.04)**	0.03 (0.04)
Unemployed	0.03 (0.05)	0.15 (0.05)***	0.18 (0.06)***
State unemployment	0.01 (0.01)	0.09 (0.01)***	0.09 (0.01)***
Constant	4.34 (0.08)***	1.67 (0.11)***	1.27 (0.12)***

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% levels two-tailed, respectively, with robust SEs clustered at the state level, $n = 2943$. Regressions also include minor and missing party, church attendance, union membership, and missing income indicators. The omitted categories are Republicans, age ≥65 years, male, not married, White, ≤12 years of education, $120,000 or more for income, employed, not in union, and do not attend church.

measurement error would attenuate the relationship between the reported unemployment rate and the qualitative economic evaluations, statistical tests contained in the Supplementary Appendix show the relationship is slightly stronger when using the benchmarked responses. We also do not find many cases of respondents survey satisficing by responding with the historical minimum, maximum, or average when they receive the benchmark. Between 2008 and 2009, the median reported unemployment rate in the benchmarked samples increased from 6.5% to 10.2%, which almost perfectly matched the actual change. Moreover, ~45% of the responses in 2009 were above the historical maximum, as compared to 15% in 2008, suggesting that many respondents were willing to report perceptions above the historical maximum. Finally, providing respondents with benchmarks did not change responses to questions that followed in the survey.

5 Understanding Partisan Bias

Quantitative questions can be used to better understand the variation in responses to qualitative questions. Qualitative economic questions often require respondents to make judgments about the quality of economic conditions, for example, whether jobs are "easy" or "hard" to find. Quantitative assessments do not involve such judgment. Thus, comparing quantitative and qualitative assessments can help determine the extent to which variation in responses to qualitative questions results from respondents using different criteria to judge economic conditions. In particular, we examine quantitative and qualitative assessments of unemployment to understand whether Democrats and Republicans judge economic conditions differently. We find preliminary evidence that partisanship either affects the reporting of both perceptions and evaluations, or affects economic judgments in a particularly odd way: opponent partisans would have to be more lenient on the incumbent to rationalize the observed patterns.

Consistent with previous results, Table 4 shows that supporters of the incumbent party (Republicans in 2008) report more positive assessments of employment in both quantitative and

Table 4 Unemployment assessments by party in 2008 benchmarked CCES

	Partisan identification		
	Republican (%)	*Independent (%)*	*Democratic (%)*
Qualitative			
Unemployment evaluation			
Positive ($n = 451$)	28.1	16.4	5.8
Neutral ($n = 1436$)	52.1	48.3	51.0
Negative ($n = 960$)	19.8	35.3	43.2
Unemployment rate			
Less than 5.6 ($n = 445$)	24.7	13.9	10.2
Between 5.6 and 7.0 ($n = 1497$)	53.5	55.6	49.3
Greater than 7.0 ($n = 905$)	21.8	30.5	40.6

Note: Numbers are percent of the column total.

qualitative questions (Wlezien, Franklin, and Twiggs 1997; Anderson, Mendes, and Tverdova 2004; Evans and Andersen 2006; Evans and Pickup 2010). However, Table 4 indicates that the qualitative reports are more related to partisan identification than the quantitative reports. In particular, nearly an identical number of respondents report that the unemployment rate is below 5.6% as report a positive evaluation of the employment situation. Yet, five times as many Republicans as Democrats report a positive qualitative evaluation of the employment situation (28.1% versus 5.8%) compared with two-and-a-half times as many Republicans as Democrats reporting an unemployment rate under 5.6% (24.7% versus 10.2%).

The literature has identified three potential sources of partisan differences in responses to qualitative economic questions. First, perceptions of the economy may relate with partisan affiliations. This could occur either because partisanship directly affects economic perceptions (Gerber and Huber 2010), or because partisanship is related to unmeasured determinants of economic perceptions, like personal experience with economy (Ansolabehere, Meredith, and Snowberg 2011a). Second, partisan affiliation may affect respondents' reports conditional on perceptions. Specifically, "partisan cheerleading" may cause supporters of the incumbent political party to report economic assessments that are more favorable than their actual economic perceptions. Third and finally, partisanship may affect the criterion used to judge the economy. For example, a Democrat may judge that a 2% growth rate is "acceptable" when a Democrat is in power, but "unacceptable" when a Republican is in power. It is difficult to separate these three sources of partisan difference in cross-sectional data, as all three have the same effect: supporters of the incumbent report more positive evaluations of economic performance than opponents.[29]

Preliminary evidence about the source of partisan bias in quantitative questions comes from comparing the correlates of reported unemployment rates on the 2008 and 2009 CCES. Over this time period, Barack Obama (a Democrat) replaced George W. Bush (a Republican) in the White House. In contrast to 2008, Table A.1 in the Supplementary Appendix shows that, in 2009, Republicans reported slightly higher rates of unemployment than Democrats. Moreover, other patterns of response remained largely the same as in 2008.[30] This suggests that differential economic experiences cannot be the only explanation for the partisan differences we observe in reported unemployment rates in Table 4. However, because quantitative assessments do not involve judgment, the partisan differences in reported unemployment rates observed in Section 4 must result from differences either in perceptions or in reporting.

[29]Previous research uses experimental or quasi-experimental variation in survey design to isolate survey effects (Wilcox and Wlezien 1993; Palmer and Duch 2001; Sturgis, Choo, and Smith 2009), or eschews survey data altogether and considers consumption data (Gerber and Huber 2009). Unfortunately, these techniques do not separate whether partisanship affects perceptions or judgments of economic conditions.

[30]As shown in Table A.1 in the Supplementary Appendix, differences in assessments between age groups were attenuated in 2009, whereas differences in assessments between race/ethnicity and education groups were enhanced.

Table 5 Conditional distribution of reported unemployment rates in 2008 benchmarked CCES

	25th percentile (%)	*50th percentile (%)*	*75th percentile (%)*
Qualitative			
Unemployment evaluation			
Positive			
Republican	5.1	6.0	6.3
Independent	5.6	6.1	7.0
Democratic	6.0	6.5	8.0
Neutral			
Republican	5.8	6.0	7.0
Independent	6.0	6.4	7.5
Democratic	6.0	7.0	8.5
Negative			
Republican	6.0	7.0	8.9
Independent	6.1	7.0	10.0
Democratic	6.1	7.0	10.0

Moreover, quantitative assessments can also be used to indirectly test for partisan differences in criteria used to judge the employment situation. To see this, consider a model where respondents randomly draw perceptions of the unemployment rate from a distribution that may vary by political party, and judge that perception according to whether it is higher or lower than a threshold. Respondents who perceive that the unemployment rate is below their evaluative threshold report a positive evaluation of unemployment, otherwise they report a negative evaluation. If supporters of the incumbent party apply a less stringent threshold when evaluating the economy, then supporters who report a positive evaluation will have a higher perception of the unemployment rate than opponents who report a positive evaluation. To put this another way: among those reporting a positive qualitative economic evaluation, the highest reported unemployment rates should come from members of the incumbent party. Likewise, among those reporting a negative qualitative economic evaluation, the lowest reported unemployment rates should come from members of the opposition party.

We do not find evidence consistent with criteria used to judge the employment situation being affected by partisanship. In contrast, Table 5 shows that in the 2008 CCES, Democrats generally report higher unemployment rates than Republicans, conditional on their qualitative assessments of unemployment.[31] For example, the interquartile range of reported unemployment rates among Republicans with a neutral qualitative evaluation is 5.8%–7.0%. In comparison, the interquartile range of Independents and Democrats with a neutral qualitative evaluation is 6.0%–7.5% and 6.0%–8.5%, respectively. Thus, if partisans are using different criteria to judge the employment situation, it would have to be the case that Republicans are using a stricter criteria than Democrats. As this is counter to theory, we conclude that partisan differences mainly enter in reporting of economic assessments (whether elicited using qualitative or quantitative questions). Moreover, as the distribution of unemployment reports by party are less skewed than qualitative unemployment assessments, we conclude that these differences are less pronounced in quantitative questions.

6 Conclusion

Many theories in political science, such as theories of economic voting, are fundamentally rooted in numbers. In particular, the focus in the economic voting literature is on how vote shares change with changes in economic quantities, such as GDP, inflation, or the unemployment rate. While survey questions that ask about numbers would form a tighter link between theory and survey data, numerous concerns have limited their use.

[31]We cannot perform a similar analysis in 2009 because roughly 90% of the sample reports negative evaluations of the employment situation; see Footnote 4.3.

We have shown that survey respondents can handle quantitative questions, especially about familiar quantities, such as the price of gas. Moreover, respondents' accuracy is affected by little else other than the regularity with which they are exposed to information about the price of gas. This finding stands in contrast to the recent public opinion work demonstrating the inaccuracy of responses, particularly among certain types of individuals, to open-ended questions about quantities.[32]

Asking questions about more complex and unfamiliar quantities is a greater challenge. However, we are confident that these challenges can be overcome in many situations through careful question design. In particular, we show that providing information about the historical tendencies of a quantity can be quite useful in giving respondents a sense of the scale of the quantity in question, without obscuring meaningful variation in responses.

Quantities are important to voters' evaluations of many policies. In particular, budget and trade deficits, the cost of social programs, the number of people affected by a policy, and the number of war dead are all naturally expressed as numbers. The results here suggest that many of the barriers to using quantitative questions are surmountable, and provides some guidance on how they may be overcome.

That said, the value of quantitative questions goes beyond the ability to reduce measurement error or discover new, interesting patterns in the data. Many political economy theories are about specific quantities—GDP growth, levels of employment, tax policy, or changes in prices (inflation)—they are not about "feelings about the economy." Qualitative survey questions seem to conflate the underlying variables of interest—perceptions of economic quantities—with outcomes—evaluations of those perceptions and resulting political behavior. Thus, qualitative questions make, at best, an indirect statement about consumers' or voters' utilities. Those statements are not irrelevant, but it has long been known that direct comparisons of utilities are difficult. And, more importantly, they are not the primitives of theoretical models.

References

Alesina, Alberto, John Londregan, and Howard Rosenthal. 1993. A model of the political economy of the United States. *American Political Science Review* 87(1):12–33.

Anderson, Christopher J., Silvia M. Mendes, and Yuliya V. Tverdova. 2004. Endogenous economic voting: Evidence from the 1997 British election. *Electoral Studies* 23(4):683–708.

Ansolabehere, Stephen. 2011. *CCES Common Content, 2008*. Version 4, http://hdl.handle.net/1902.1/14003 (accessed June 12, 2012).

Ansolabehere, Stephen, Marc Meredith, and Erik Snowberg. 2011a. Mecro-economic voting: Local information and micro-perceptions of the macro-economy. Presented at the annual meeting of the Midwest Political Science Association, Chicago.

———. 2011b. Sociotropic voting and the media. In *The ANES book of ideas*, eds. John H. Aldrich and Kathleen McGraw. Princeton, NJ, Princeton University Press.

Ansolabehere, Stephen, Erik Snowberg, and James M. Snyder. 2005. Unrepresentative information the case of newspaper reporting on campaign finance. *Public Opinion Quarterly* 69(2):213–31.

Ansolabehere, Stephen, Marc Meredith, and Erik Snowberg. 2012. *Replication data for: Asking about numbers: Why and how*. http://hdl.handle.net/1902.1/18787 IQSS Dataverse Network [Distributor] V1 [Version].

Bartels, Larry M. 2002. Beyond the running tally: Partisan bias in political perceptions. *Political Behavior* 24(2):117–50.

Beyth-Marom, Ruth. 1982. How probable is probable? A numerical translation of verbal probability expressions. *Journal of Forecasting* 1(3):257–69.

Blendon, Robert J., John M. Benson, Mollyann Brodie, Richard Morin, Drew E. Altman, Daniel Gitterman, Mario Brossard, and Matt James. 1997. Bridging the gap between the public's and economists' views of the economy. *Journal of Economic Perspectives* 11(3):105–18.

Blinder, Alan S., and Alan B. Krueger. 2004. What does the public know about economic policy, and how does it know it? *Brookings Papers on Economic Activity* 2004(1):327–87.

Bruine de Bruin, Wandi, Baruch Fischhoff, Susan G. Millstein, and Bonnie L. Halpern-Felsher. 2000. Verbal and numerical expressions of probability: It is a fifty-fifty change. *Organizational Behavior and Human Decision Processes* 81(1):115–31.

[32]For a few examples, see Conover, Feldman, and Knight (1986); Nadeau, Niemi, and Levine (1993); Holbrook and Garand (1996); Kuklinski et al. (2000); Gilens (2001); Sigelman and Niemi (2001); Kaplowitz, Fisher, and Broman (2003); Ansolabehere, Snowberg, and Snyder (2005); Martinez, Wald, and Craig (2008) and Herda (2010).

Bullock, John G., Alan S. Gerber, and Gregory A. Huber. 2010. Partisan bias in factual beliefs about politics. Mimeo, Yale University.

Cleveland, William S., Susan J. Devlin, and Eric Grosse. 1988. Regression by local fitting: Methods, properties, and computational algorithms. *Journal of Econometrics* 37(1):87–114.

Conover, Pamela Johnstone, Stanley Feldman, and Kathleen Knight. 1986. Judging inflation and unemployment: The origins of retrospective evaluations. *Journal of Politics* 48(3):565–88.

Converse, Philip E. 1964. The nature of beliefs systems in mass publics. In *Ideology and discontent*, ed. David E. Apter. Glencoe, IL: Free Press of Glencoe.

Curtin, Richard. 2007. What U.S. consumers know about economic conditions. Memeo, University of Michigan.

Erev, Ido, and Brent L. Cohen. 1990. Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organization Behavior and Human Decision Processes* 45(1):1–18.

Erikson, Robert S., Michael B. MacKuen, and James A. Stimson. 2002. *The macro polity*. Cambridge, UK, Cambridge University Press.

Evans, Geoffrey, and Mark Pickup. 2010. Reversing the causal arrow: The political conditioning of economic perceptions in the 2000–2004 U.S. presidential election cycle. *Journal of Politics* 72(4):1236–51.

Evans, Geoffrey, and Robert Andersen. 2006. The political conditioning of economic perceptions. *Journal of Politics* 68(1):194–207.

Fair, Ray C. 1978. The effect of economic events on votes for president. *Review of Economics and Statistics* 60(2):159–73.

Fox, Gerald T. 2009. Partisan divide on war and the economy: Presidential approval of G. W. Bush. *Journal of Conflict Resolution* 53(6):905–33.

Gerber, Alan S., and Gregory K. Huber. 2009. Partisanship and economic behavior: Do partisan differences in economic forecasts predict real economic behavior?. *American Political Science Review* 103(3):407–26.

———. 2010. Partisanship, political control, and economic assessments. *American Journal of Political Science* 54(1):153–73.

Gilens, Martin. 2001. Political ignorance and collective policy preferences. *American Political Science Review* 95(2):379–95.

Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey methodology*. Hoboken, NJ, Wiley.

Herda, Daniel. 2010. How many immigrants? Foreign-born population innumeracy in Europe. *Public Opinion Quarterly* 74(4):674–95.

Hetherington, Marc J. 1996. The media's role in forming voters' national economic evaluations in 1992. *American Journal of Political Science* 40(2):372–95.

Holbrook, Thomas, and James C. Garand. 1996. Homo economus? Economic information and economic voting. *Political Research Quarterly* 49(2):351–75.

Imbens, Guido W., and Thomas Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2):615–35.

Juster, F. Thomas. 1966. Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association* 61(315):658–96.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–92.

Kaplowitz, Stan A., Bradley J. Fisher, and Clifford L. Broman. 2003. How accurate are perceptions of social statistics about blacks and whites? Effects of race and education. *Public Opinion Quarterly* 67(2):237–43.

Kinder, Donald R., and D. Roderick Kiewiet. 1979. Economic discontent and political behavior: The role of personal grievances and collective economic judgments in congressional voting. *American Journal of Political Science* 23(3):495–527.

———. 1981. Sociotropic politics: The American case. *British Journal of Political Science* 11(2):129–61.

Kramer, Gerald H. 1971. Short-term fluctuations in U.S. voting behavior, 1896–1964. *American Political Science Review* 65(1):131–43.

———. 1983. The ecological fallacy revisited: Aggregate-versus individual-level findings on economics and elections, and sociotropic voting. *American Political Science Review* 77(1):92–111.

Krosnick, Jon A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5(3):213–36.

Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, and Robert F. Rich David Schwieder. 2000. Misinformation and the currency of democratic citizenship. *Journal of Politics* 62(3):790–816.

MacKuen, Michael B., Robert S. Erikson, and James A. Stimson. 1992. Peasants or bankers? The American electorate and the U.S. economy. *American Political Science Review* 86(3):597–611.

Malhotra, Neil, and Jon A. Krosnick. 2007. The effect of survey mode and sampling on inferences about political attitudes and behavior: Comparing the 2000 and 2004 ANES to Internet surveys with nonprobability samples. *Political Analysis* 15(3):286–323.

Manski, Charles F. 2004. Measuring expectations. *Econometrica* 72(5):1329–76.

Martinez, Michael D., Kenneth D. Wald, and Stephen C. Craig. 2008. Homophobic innumeracy? Estimating the size of the gay and lesbian population. *Public Opinion Quarterly* 72(4):753–67.

Moxey, Linda M., and Anthony J. Sanford. 2000. Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology* 14(3):237–55.

Nadeau, Richard, and Richard G. Niemi. 1995. Educated guesses: The process of answering factual knowledge questions in surveys. *Public Opinion Quarterly* 59(3):323–46.

Nadeau, Richard, Richard G. Niemi, and Jeffrey Levine. 1993. Innumeracy about minority populations. *Public Opinion Quarterly* 57(3):332–47.

Palmer, Harvey D., and Raymond M. Duch. 2001. Do surveys provide representative or whimsical assessments of the economy? *Political Analysis* 9(1):58–77.

Rockwood, Todd H., Roberta L. Sangster, and Don A. Dillman. 1997. The effect of response categories on questionnaire answers: Context and mode effects. *Sociological Methods & Research* 26(1):118–40.

Schuman, Howard, and Stanley Presser. 1981. *Questions and answers in attitude surveys*. New York, Academic Press.

Schwarz, Norbert, Hans-J. Hippler, Brigitte Deutsch, and Fritz Strack. 1985. Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly* 49(3):388–95.

Sigelman, Lee, and Richard G. Niemi. 2001. Innumeracy about minority populations: African Americans and Whites compared. *Public Opinion Quarterly* 65(1):86–94.

Snowberg, Erik, and Justin Wolfers. 2010. Explaining the favorite-longshot bias: Is it risk-love or misperceptions? *Journal of Political Economy* 118(4):723–46.

Stephenson, Laura B., and Jean Crête. 2011. Studying political behavior: A comparison of Internet and telephone surveys. *International Journal of Public Opinion Research* 23(1):24–55.

Sturgis, Patrick, Martin Choo, and Patten Smith. 2009. Response order, party choice, and evaluations of the national economy: A survey experiment. *Survey Research Methods* 3(1):7–12.

Tourangeau, Roger, Kenneth Rasinski, Jared B. Jobe, Tom W. Smith, and William F. Pratt. 1997. Sources of error in a survey on sexual behavior. *Journal of Official Statistics* 13(4):341–65.

Tourangeau, Roger, and Tom W. Smith. 1996. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60(2):275–304.

van der Brug, Wouter, Cees van der Eijk, and Mark Franklin. 2007. *The economy and the vote: economic conditions and elections in fifteen countries*. Cambridge, UK, Cambridge University Press.

Wallsten, Thomas S. 1986. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General* 115(4):348–65.

Wallsten, Thomas S., David V. Budescu, Rami Zwick, and Steven M. Kemp. 1993. Preference and reasons for communicating probabilistic information in numerical or verbal terms. *Bulletin of the Psychonomic Society* 31:135–38.

Wilcox, Nathaniel, and Christopher Wlezien. 1993. The contamination of responses to survey items: Economic perceptions and political judgments. *Political Analysis* 5(1):181–213.

Wlezien, Christopher, Mark Franklin, and Daniel Twiggs. 1997. Economic perceptions and vote choice: Disentangling the endogeneity. *Political Behavior* 19(1):7–17.

Wright, Daniel B., George D. Gaskell, and Colm A. O'Muircheartaigh. 1994. How much is quite a bit? Mapping between numerical values and vague quantifiers. *Applied Cognitive Psychology* 8:479–96.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levindusky, Alberto Simpser, and Rui Wang. 2009. Comparing the accuracy of RDD telephone surveys and Internet surveys conducted with probability and non-probability samples. Mimeo, Stanford University.

Zaller, John. 1992. *The nature and origins of mass opinion*. Cambridge, UK, Cambridge University Press.

Zaller, John, and Stanley Feldman. 1992. A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science* 36(3):579–616.