



ANNUAL REVIEWS **Further**

Click [here](#) to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Measuring Public Opinion with Surveys

Adam J. Berinsky

Department of Political Science, Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139; email: berinsky@mit.edu

Annu. Rev. Polit. Sci. 2017. 20:309–29

First published online as a Review in Advance on
March 2, 2017

The *Annual Review of Political Science* is online at
polisci.annualreviews.org

<https://doi.org/10.1146/annurev-polisci-101513-113724>

Copyright © 2017 by Annual Reviews.
All rights reserved

Keywords

mass behavior, polls, survey research

Abstract

How can we best gauge the political opinions of the citizenry? Since their emergence in the 1930s, opinion polls—or surveys—have become the dominant way to assess the public will. But even given the long history of polling, there is no agreement among political scientists on how to best measure public opinion through polls. This article is a call for political scientists to be more self-conscious about the choices we make when we attempt to measure public opinion with surveys in two realms. I first take up the question of whom to interview, discussing the major challenges survey researchers face when sampling respondents from the population of interest. I then discuss the level of specificity with which we can properly collect information about the political preferences of individuals. I focus on the types of question wording and item aggregation strategies researchers can use to accurately measure public opinion.

INTRODUCTION

How can we best gauge the political opinions of the citizenry? The question is a critical one for any democratic society. To make the compromises and trade-offs essential to the functioning of a political system, we need information about both the direction and intensity of the public will. Throughout history a variety of techniques have been employed, ranging from monitoring mass protests and public speeches to evaluating letters published in newspapers and the partisan press (for a review, see Herbst 1993). But since their emergence in the 1930s, opinion polls—or surveys—have become the dominant way for politicians, media organizations, interest groups, and academic researchers to assess the public will.

In one of the most quoted statements in the field of public opinion, Key (1961, p. 8) noted that “to speak with precision of public opinion is a task not unlike coming to grips with the Holy Ghost.” But even if we have not come to grips with the larger meaning of public opinion, researchers and practitioners proceed as if we have. As Converse (1987) argued more than 25 years ago, in effect polls have become public opinion. The development of opinion polls over the last 75 years has changed how politicians relate to the mass public. Surveys have enabled politicians and the media to quickly and efficiently tap public sentiment. Indeed, Sidney Verba concluded in his presidential address to the 1995 American Political Science Association Meeting, “Surveys produce just what democracy is supposed to produce—equal representation of all citizens” (1996, p. 3; see also Geer 1996).

Polls are not without their flaws. Beginning with Blumer (1954), and continuing through Ginsberg (1986) and Herbst (1993), academics have raised a number of thoughtful critiques of the polling enterprise. However, these critics have not slowed the polling bandwagon. Nor should they. Surveys remain a critical method to gauge the political views of the mass public. They provide a more nuanced picture of the political cognition of individuals than do the blunt instruments of electoral returns. Moreover, even at a time when the “big data” revolution has opened up new ways to examine mass politics, surveys allow us to examine in detail the preferences and motivations of the public, using tools that we as researchers design to directly address our theoretic concepts of interest.

Within such a context, it is important to cast a critical eye on the polling enterprise. The fact that the field has come to a consensus does not mean that measuring public opinion is a straightforward endeavor. Even given the long history of polling, there is no agreement among political scientists on how to best measure public opinion through polls. This article is a call for political scientists to be more self-conscious about the choices we make when we attempt to measure public opinion with surveys.

There are two basic choices we make when conducting a poll: which people to interview and what questions to ask them. These choices seem simple, but neither is straightforward. The strategies one chooses can greatly affect the answers one gets. Thus, the voice of the people, as reflected in polls, is profoundly shaped by the decisions we make as researchers.

In this article, I examine the process of measuring public opinion with surveys through these two paths. I first take up the question of whom to interview. This question relates to sampling our respondents from the population of interest. The way in which we draw our samples influences the nature of the “public” we can generalize to. I discuss trends in survey sampling methods, with a special focus on the difficulties currently faced by the polling industry.¹ These difficulties do not necessarily signal the end of polling as we have come to know it, but the technological developments

¹My discussion in this article largely focuses on polling in the United States. However, the issues and concerns raised here are equally important to consider when measuring public opinion throughout the world. I illustrate this importance with some examples throughout the article.

of the modern era have altered the process of polling and have important implications for how we assess the information contained in surveys. Survey researchers are working on new methods to sample public opinion, but the costs and benefits of these methods are still contested and unclear. The first half of this article therefore does not offer clear guidance for how best to move forward, but it does provide an overview of the current state of the field and poses questions for scholars and practitioners to consider.

The second question—and one that has received much less attention in a big-picture sense—relates to the level of specificity with which we can accurately collect information about the political preferences of individuals. I focus on the particulars of the survey itself, namely, the kinds of questions researchers and practitioners ask their respondents. A great deal has been written about question wording effects in surveys—how the questions one asks and the order in which they are asked affect the answers one gets. Space constraints preclude a detailed discussion of this literature here, and Groves et al. (2009) have provided an excellent review. What I have in mind is a different and more fundamental question: At what level of abstraction can surveys gather meaningful information about the public's politically relevant wants, needs, and desires? As a practical matter, there is often a mismatch between the level of specificity in the attitudes that we would like people to have and in the attitudes they actually possess. That is, as policy makers and researchers, we would like people to be as discerning as we are when expressing support or opposition to different policies and programs; but, in reality, the public is far less discriminating. In light of this fact, is opinion best represented on a specific issue-by-issue basis, or is it instead better exemplified by a more general quantity, measured either through broad survey questions or the respondents' central tendency on opinions toward a broad variety of issues? My answer to this question is the focus of the second half of this article.

WHOM TO INTERVIEW

I begin this section with a discussion of sampling. When conducting a poll, it is important to select a representative set of individuals to canvass. Conducting a census of the entire US population is both prohibitively expensive and unnecessary. Instead, survey researchers select a sample from the population of interest (here residents of the United States).

From a technical point of view, the correct way to draw such a sample is through a process of simple random sampling (SRS). Simple random samples have two important properties: Each individual is chosen for inclusion in the sample by chance, and each member of the population has an equal chance of being included in the sample. Under these conditions, every possible sample of a given size has the same chance of selection. Each resident of the United States therefore has an equal chance to be included in a poll. But this technically correct procedure is impossible to achieve in practice, or at least impossibly expensive. Researchers have therefore always fallen short of this ideal. Instead, practitioners have employed other methods that ensure representative samples.

For most of the twentieth century, these procedures included face-to-face multi-stage designs and telephone interviewing by random-digit dialing (for an overview, see Lohr 2010). These methods, though not SRS, can be conducted in ways that allow researchers to approximate SRS. Statistical methods have been developed to account for the design components of modern survey sampling, such as clustering and stratification (Kish 1965, Lohr 2010).

In addition, researchers and practitioners have developed methods to account for imbalances that arise between the processes of drawing a sample and collecting a sample. Not all respondents who are selected to be in a poll agree to be interviewed. I discuss the implications of this nonresponse later in this article, but for now what is important is that researchers have worked out

ways to correct for this problem. In most cases, we only have limited information about the population (relative to the sample) in the form of information taken from the US Census. In these cases, weighting adjustments are typically applied to reduce the bias that nonresponse can cause in survey estimates (Lohr 2010). For example, if there are more old respondents in a survey sample than in the population, old respondents are given less weight and young respondents are given more, so that the distribution of age in the adjusted sample matches the population distribution. One approach commonly used to weight samples is raking. Raking matches cell counts to the marginal distributions of the variables used in the weighting scheme. Other techniques, such as regression weighting and propensity score weighting, take advantage of more detailed information. Lohr (2010) provides a useful overview of these different weighting methods (see also Lumley 2010).

The Changing Landscape of Polling

If I were writing this article 20 years ago—or even 15—this section would end here. Face-to-face and telephone sampling techniques were well worked out, and survey researchers could be confident that the group of respondents to their polls adequately represented the full US population. The question of whom to interview was asked and answered. However, beginning in the 1990s and especially since the turn of the century, a number of developments have thrown the polling industry into disarray. The representativeness of opinion polls is predicated on the ability of researchers to actually communicate with the respondents they select through their sampling procedure. This has become more and more difficult over time. Potential respondents are becoming harder to contact and, once researchers are able to contact a subset of such individuals, those potential respondents are less likely to agree to participate in polls.

Telephone interviewing. By the mid-1970s, over 90% of US households had a telephone (Lavrakas et al. 2007). Thus, the sample frame of telephone numbers provided good coverage of the target population of US residents. Moreover, contact rates were high—survey researchers could make contact with almost all individuals who were in the sample of potential respondents. In turn, cooperation rates were also high—the vast majority of these respondents would agree to answer the pollster’s questions.

Beginning in the late 1990s, nonresponse to opinion polls began to rise at alarming rates. Both contact rates and cooperation rates fell precipitously. The best source of overtime data on this question is the Pew Center for the Study of the People and the Press, which has released detailed figures on its ability to reach potential respondents for almost 20 years. This information is presented in **Figure 1**. In 1997, Pew was able to contact 90% of its potential respondents, and 43% of those individuals agreed to answer the survey, yielding an overall response rate of 36%. By 2012, the contact rate dropped to 62% and the cooperation rate fell to 14%, yielding a response rate of only 9%. Today, more than nine-tenths of potential respondents fail to make it into the sample of one of the best-run survey operations in the United States. Surely the response rate of organizations less skilled than Pew is even lower.²

Other technological changes have disrupted the survey research industry as well. An increasing number of Americans have abandoned landline telephones in favor of cell phones. Since 2003, the Center for Disease Control and Prevention’s National Health Interview Survey (NHIS) has

²Of course, low response rates in and of themselves do not necessarily lead to biased survey results. If the people who participate in surveys are similar to those who do not, opinion polls can remain representative of the mass public. But if nonrespondents are systematically different from respondents, the resulting measures of public opinion could be severely biased.

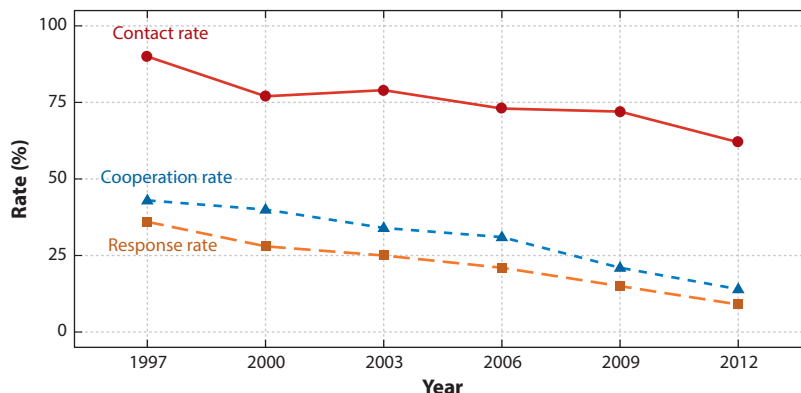


Figure 1

Decline in response rates to Pew surveys, 1997–2012 (Pew Res. Cent. 2012).

measured the prevalence of cell-only households in the United States. Over that time, the percentage of adults living in households with only wireless telephone service rose from 3% to 48% (Blumberg & Luke 2016). This sea change in communication technology has upended traditional phone polls. Cell-phone interviews are more expensive than landline interviews because surveyors are prohibited by law (the 1991 Telephone Consumer Protection Act, 47 U.S.C. 227) from using the autodialers that made telephone interviewing inexpensive. But there are more serious problems than the cost of doing business. Whereas early studies had indicated that the biases induced through the failure to contact cell-only households could be corrected by weighting a landline-only sample to reflect population characteristics on key variables such as age and race (Brick et al. 2006, Keeter 2006), by 2008 this solution appeared to have stopped working. Keeter et al. (2008) found that weighting landline samples to reflect the known characteristics of the population could not account for the different vote preferences of individuals who have only a cell phone. Individuals who had only cell phones were fundamentally different from individuals with landlines (conditioning on the kinds of demographic measures, such as age and race, that are available to survey researchers). Thus, probability samples of landline respondents excluded a distinct portion of the population from the sampling frame. Put another way, the opinions of cell-only individuals represent missing data for which we cannot easily compensate. Researchers in other disciplines have come to similar conclusions (see sidebar titled Cell Phone Coverage and Surveys).

CELL PHONE COVERAGE AND SURVEYS

The National Center for Health Statistics at the Centers for Disease Control and Prevention (CDC) compared National Health Interview Survey (NHIS) data from landline-only samples to those from “dual frame” samples that included both cell-only and landline households. They found that weighting a landline-only sample to match census demographics produced biased estimates of population characteristics and health behaviors (Blumberg & Luke 2009). The CDC has even raised concerns about its largest ongoing telephone interview health survey, the Behavioral Risk Factor Surveillance System. Researchers from the CDC conducted a parallel cell-only survey in 2008 and found large differences across a range of self-reported attitudes and behaviors between the combined sample and the landline-only sample, even after weighting (Hu et al. 2011).

As a result, major media and academic organizations have all moved to interviewing on both landlines and cell phones. However, to date, there is no widely accepted set of cell-phone surveying “best practices” (Lavrakas et al. 2010) and no consensus about how best to combine cell-phone and landline respondents into a single sample using “dual-frame” approaches (Lavrakas et al. 2010; for further discussion, see Brick et al. 2006, Kennedy 2007, Guterbock et al. 2011, Levine & Harter 2015). Cell phones are tied to individuals, whereas landlines are attached to households. Blending samples with these two types of points of contact is no easy feat. For instance, many individuals are reachable by both cell phone and landline. People who appear in both frames will have a higher probability of inclusion.

Perhaps, in time, cell-phone penetration will be broad enough to match the coverage of landline phones in the early 1990s. However, until that day, the path forward is unclear. The rise of cell phones clearly threatens the representativeness of surveys, but there are no obvious remedies for this problem.

Online interviewing. Polling through the internet has also been on the rise since the turn of the century. Increasing internet penetration among households has made it possible to reach a broad segment of the US population, at least in theory. And the ability to conduct self-administered surveys over the web has opened up a world of new possibilities for academics and practitioners alike. As Brick (2011) aptly notes, the low cost of administering surveys on the internet is attractive. However, there are reasons to be cautious about a jump to the internet. Unlike phone surveys and face-to-face interviews, the practice of internet surveying, so far, lacks the statistical theory and common understandings regarding data collection practices that would allow internet sampling to approximate SRS.

It is extremely difficult to define a target population that we would like to represent with our sample. Hillygus (2016) rightly points out that there exists no list of all internet users we can draw upon analogous to the list of telephone numbers. As a result, research over the internet is often conducted using nonprobability samples (although, of course, just as non-online polls are not inherently probability samples, online polls do not necessarily have to be nonprobability samples). For instance, some companies draw samples from online panels of volunteers. These samples are then adjusted to approximate the population in various ways, including quota sampling to match Census demographics, model-based propensity matching, and postsurvey adjustment via some weighting technique (Baker et al. 2010). But the key point remains that no matter what mechanism is employed, these samples from internet panels cannot be considered akin to random samples. A task force of the American Association for Public Opinion Research (AAPOR)—the leading organization for survey research scholars and practitioners—concluded that “there is currently no generally accepted theoretic basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population” (Baker et al. 2010, p. 758). Despite some promising work in this area (Gelman et al. 2016), in the six years since the AAPOR report was published, no consensus on this point has been reached.³

Address-based sampling. Another form of sampling that has been more widely used in recent years is address-based sampling (ABS). ABS is the process of sampling from a frame of address lists obtained from the US Postal Service. These computerized delivery sequence files ensure

³Some researchers have attempted to use social media platforms, such as Twitter, to measure public opinion. Although early efforts seem promising (O’Connor et al. 2010), a great many questions remain (Murphy et al. 2014).

high coverage of the universe of households—approaching 100% in some areas. Vendors have used these lists to construct frames and samples of potential respondents, updating these lists continuously to reflect changes in the delivery addresses.⁴ These methods clearly hold great promise for survey researchers. Surveyors can make initial contact with a probability sample through the mail and then direct respondents to an internet survey, for example. That said, fulfilling the promise of ABS is an ongoing process. As the AAPOR report notes, although contemporaneous coverage of ABS frames has been researched extensively, coverage is not a static property. In addition, Brick (2011) makes the excellent point that the administrative records used to construct frames for ABS were not designed for that purpose and, as a result, are not of the quality needed for sampling purposes. Thus, although ABS is extremely promising, much work needs to be done to ensure it is an effective method of recruiting respondents.

Declining cooperation. All of this discussion does not account for the threat posed by decreasing cooperation rates. Once a surveyor reaches potential respondents, they need to agree to be interviewed. As the Pew data presented above convincingly demonstrate, securing respondents' cooperation has become harder over the last 20 years. Some explanations attribute the decline in cooperation rates to increasing concerns about privacy and confidentiality and increased distrust of pollsters; from 1998 to 2006, the percentage of Americans saying they "would generally trust pollsters to tell the truth" fell from 55% to 34% (Kim et al. 2011). Others attribute the decline to lifestyle changes, as people feel they are busier and have less time to participate in surveys (Tourangeau 2004, Galea & Tracy 2007). Regardless of the cause, it has made the job of surveying the public even harder. Some pollsters have increased the use of monetary incentives (Tourangeau 2004, Groves 2006) and advance contact letters (De Leeuw et al. 2007) to increase cooperation rates. But the problem of nonresponse bias created by decreasing cooperation rates is not easily solved.

What Does the Future Hold?

The message from this review of the literature is clear: The question of whom to interview might be simple, but the answer is decidedly not. The procedures to select respondents that have served the profession well for decades need to be adapted to serve the needs of a new world. At the same time, significant questions remain about the utility of new technologies and methods. The modern polling landscape is the Wild West, in need of a sheriff to bring order.

Although the way forward is unclear, polls are not going anywhere; they remain a critical part of the political system. In concluding his survey of the history of survey research for *Public Opinion Quarterly*'s seventy-fifth anniversary issue, Robert Groves (2011, p. 870) rightly notes, "Survey research is not dying; it is changing." As the industry struggles to figure out how to collect samples of respondents in today's world, some basic facts need to be recognized and carefully considered when determining how best to construct a sample to assess the will of the people. The following series of suggestions provides what I hope is a productive starting point in thinking about how best to accurately sample mass opinion.

First, it may be useful to think of the problems currently plaguing surveys as missing-data problems. The central challenge of selecting respondents to interview is to minimize the differences between the people we ask our questions and the people we miss. For instance, the fundamental

⁴In addition, researchers interested in surveying potential voters can use voter files to contact respondents (Barber et al. 2014).

problem with continuing to use landlines to sample survey respondents is that the vast majority of people who should be in our sample are not in our sample—either because we cannot reach them or because they refuse to answer our questions once we contact them. From this perspective, the proper way to fix the problem is to consider how the people we can reach differ from the people we cannot reach—and account for those differences through weighting. We can adjust our samples, but in order to arrive at an accurate picture of the public will, we must measure the differences between respondents and nonrespondents to particular surveys. If we do not, our estimates of the will of the mass public will be wrong—as the failure of weighting to correct for cell-only bias after 2004 demonstrates. This concern is especially important because the auxiliary variables used for weighting are often chosen arbitrarily on the basis of their availability in both surveys and data sets measuring the characteristics of the general population—such as age, and gender—regardless of whether they will actually reduce the bias in our estimates. The fact that the United States has a relatively sparse collection of publicly available population-level demographic data (Massey & Tourangeau 2013) indicates that the weighting problem will not go away anytime soon.

Second, probability and nonprobability samples are fundamentally different, and a variety of factors need to be considered when comparing methods of gathering responses. It is important to continue to recognize these differences, even in an era when probability response rates are collapsing to the low single digits. As Brick (2011, p. 876) notes, “[A] common refrain is that a probability sample with a low response rate or coverage rate is no ‘better’ than a nonprobability or volunteer sample.” But this point of view may not be accurate. That a poll has a low response rate—even a vanishingly low response rate—does not mean that it is fundamentally flawed. In fact, it may be better than a nonprobability sample that looks more like the population on quantities we can measure. Probability samples are built on a strong inferential foundation because no matter how low the response rate, all the sampled units are actively recruited and encouraged to participate in the survey. This is not true for nonprobability samples. Put another way, there is a difference between me coming to you as a researcher and asking you to participate in my survey and you coming to me as a respondent and asking to be a part of my study. The fact that online panels may contain “professional” respondents who seek to take multiple surveys for the cash and other incentives offered (Hillygus et al. 2014) underscores this problem. Unless we can measure exactly the differences that lead to the self-selection behavior in nonprobability samples, we cannot account for these differences. That said, declining response rates are clearly troubling, and the relative merits of probability and nonprobability sampling should be subject to continuing debate—but it must be a debate informed by the limitations of both traditional forms of sampling and newer forms of sampling.

Finally, we need to think hard about the question of whom to interview—and we must not stop thinking about it simply because we have found what we take to be an acceptable solution in the present day. What is true today might not be true tomorrow, and it is always important to be vigilant when evaluating methods of sampling. The lessons of the 1936 Literary Digest debacle are illustrative here. In that case, the mail-back methodology that had accurately predicted every presidential election from 1920 through 1932 suddenly and spectacularly failed (see Squire 1988 for details). It is unclear if we are entering a similar predicament today. For instance, the polls leading up to the 2015 UK general election predicted a dead heat between the Conservative Party and the Labour Party. In the end, however, the Conservatives routed Labour by more than six percentage points. The British Election Study, which occurred right after the election and employed a random sample, correctly estimated the vote outcome, leading some to suspect that the failed prediction was driven by errors stemming from the nonrandom subject pools used by the majority of election pollsters (Clark 2015). Similar experiences in other large-scale, high-salience electoral contexts will lead to further questions about the utility of particular polling methods.

WHAT TO ASK

Once we have decided whom to interview in a poll, we must then decide what questions to ask them. This too is a difficult puzzle. As mentioned, a great deal has been written about question wording and ordering effects in opinion surveys (for a review, see Tourangeau et al. 2000). This literature has contributed to the development and growth of the survey enterprise in critical ways. However, in the remainder of this article, I consider a more fundamental question that touches on broader concerns: What kind of information can we meaningfully gather with political survey questions? The answer to this question turns on our assumptions and theories about the meaning of survey responses.

As a starting point, let us acknowledge two things we as political scientists know to be true about the mass public in the American political system. First, most of the people, most of the time, do not pay attention to politics.⁵ As Converse (1990) aptly notes, the mean level of political information among the mass public is vanishingly low. But, second, once in a survey interview, citizens are compliant. Ask them a question and they will give you an answer—perhaps not if the question is incomprehensible or obscure, but if the question seems reasonable, most people will try to give a reasonable answer.⁶

These two simple observations together have important implications for making sense of the information we collect in opinion polls. When choosing the questions we ask, we need to be cognizant of the fact that almost anyone will answer that question, even if they have little basis for their answer. Under such circumstances, how should we think of survey responses? For some scholars, the tendency of individuals to answer any survey question relegates public opinion polls to the trash bin—a string of meaningless nonattitudes (Moore 2008). But the fact that survey questions do not necessarily elicit a stable and deliberate stance on a given issue does not mean that answers to survey questions are worthless. Considering more fully the survey response process yields a more charitable view of opinion polls.

To do so, I begin at the level of the individual. What does an answer to a survey question reveal about a person's political views? The notion that surveys reveal preferences over political choices and policies generated through some sort of "file drawer" model—a fixed stance on a given issue that people call from memory when prompted by the interviewer—has long been discarded by psychologists and political scientists. Instead, survey responses are best seen as constructed preferences (Zaller 1992, Tourangeau et al. 2000; for a review, see Berinsky 2004).

The best description of this view in political science is Zaller's "Receive-Accept-Sample" (RAS) model, which builds on survey response models in psychology (see, e.g., Tourangeau & Rasinski 1988). Zaller (1992, p. 49) argues that individuals answer survey questions off the top of their head by "averaging across the considerations that are immediately salient to them" due to their personal characteristics and political experiences at the time of the survey interview. Survey responses therefore are a summary judgment over the mass of considerations—reasons for favoring one side of a controversy rather than another—that happen to be on their mind when they answer a particular question. The flow of information encountered about politics in daily life, and even the wording of particular poll questions, can bring about systematic changes in the kinds of considerations that are on people's minds. Because different considerations may be salient at different points in time, the response obtained from the same person may change from interview to interview. From Zaller's point of view, answers to survey questions are "opinion statements"

⁵This is a pessimists' reformulation of an astute observation, made most recently by Stimson (2015 [2004], p. 14), that "some people some of the time pay attention to government."

⁶For the logic behind this decision, see Schwarz (1996). For my interpretation, see Berinsky (2004).

that reflect a sample of respondents' underlying politically relevant considerations—the types of concerns and perspectives people bring to bear when considering issues in the realm of politics. These opinion statements may not meet the standards to which politicians would hold themselves with regard to their own opinions, but that does not mean such statements are meaningless. To use Bartels' (2003) terminology, people have attitudes, not preferences.

But even if the politically relevant thoughts of the average citizen rarely rise to the standard of preferences, these statements can have great meaning. They can provide a valuable window into the political thinking of ordinary Americans because even if they are somewhat inchoate, they can represent a sample of the considerations that resonate with the individual—the politically relevant wants, needs, and desires of those citizens.

There is, however, another side to this coin, and a word of caution is in order. Just because we can ask a survey question does not mean that we should take each and every survey response seriously. In the judgment of some scholars, asking questions about highly specific issues and policies can be a recipe for disaster. Bishop (2008) contends that responses to survey questions do not represent actual opinions about the specific policy issues being probed by pollsters, but are an “illusion” based on public ignorance of politics and aided by vague polling questions and variations in question form, wording, and context. Thus, for Bishop, measuring and aggregating survey responses is a meaningless exercise. He concludes that “public opinion polling today ‘creates’ public opinion that does not exist in the way intended by the pollsters’ questions and as interpreted by the journalistic community and the policymaking powers that be” (Bishop 2008, p. 150).

Bishop's words ring true. But taken to their logical extreme, they paint an inaccurate view of public opinion. Recognizing that particular responses to specific questions may be somewhat disconnected from their intended meaning does not mean abandoning the polling enterprise. Even Bishop (2008, p. 157) concedes that “respondents certainly do draw upon their general attitudes, interests and values”—the same kinds of cognitive concepts that Zaller terms predispositions—when answering survey questions. Survey questions may, as Bishop argues, force most people to “make up” answers on the spot, but these made-up answers are still meaningful because they reflect individuals' politically relevant considerations—their underlying distribution of preferences over the policies of government. Under most circumstances, Zaller's “attitude statements” do have meaning and political relevance.

Researchers and practitioners, then, should not necessarily take answers to specific questions at face value—heeding Bishop's warnings—but nor should they altogether reject measures of the public will as expressed in surveys. Instead, they should measure opinion at a medium level of specificity. The key is to find a balance where survey responses are general enough to reflect the respondent's distribution of considerations but are not so general as to lack relation to political controversies. Collected together, these individual survey responses can provide a window into public opinion. Such an approach fits well with both the needs of politicians and the realities of public engagement with the political world. As Stimson puts it (1999 [1991], p. 13):

Politicians must deal with issues at a high level of specificity Public opinion is by necessity general, unfocused. . . . [A] rational economizing collector of opinions would use that information to gain leverage on the evolution of public views at a very high level of generality, the level at which it is meaningful for a public mainly inattentive to public affairs.

In the next section, I discuss how to collect such information with specific examples, but for now, what is important is delineating the purpose of polls in a democratic system—to facilitate communication between masses and elites. Properly posed survey questions can do just this.

How to Target the Proper Level of Specificity

The argument in the previous section is that, at the individual level, survey responses are meaningful when the measurement instrument is calibrated at a moderate level of specificity: not so general as to be empty of content and not so specific that they risk falling into the trap described by Bishop, creating opinions where none exist. When such responses are collected together and aggregated across individuals, they can provide a clear measure of public opinion writ large. Thus, by my view, public opinion properly exists at a mezzo level, but where to locate that mezzo level is an open question. There are, after all, various ways to arrive at this end—and various means have been advanced by different scholars.

One approach is to ask individuals questions at a general level, for example, by measuring preferences over broad trade-offs between increasing spending and increasing the scope of government involvement in the public sphere. This approach focuses on the particular properties of individual questions. A second approach is to employ a series of survey questions about specific issues—for instance, to ask detailed policy questions about specific government programs—and construct a composite measure from that list of items. This approach considers the relationship of individual questions to each other. These two strategies are, I argue, complementary approaches that consider differential degrees of abstraction in survey questions. The first approach considers the specificity of questions. The second approach is akin to what Druckman & Jacobs (2006) term “lumping” together survey items (rather than “splitting” question responses) and considers aggregation across survey questions. I discuss both in turn.

Single-question approaches. When considering single-issue-based survey items, some questions may be more useful than others for measuring public opinion. Consider the following question, asked by the American National Elections Study (ANES) for more than 40 years and used widely in studies of public opinion (Brians & Wattenberg 1996, Jacoby 2000, Newman 2003, Berinsky 2004, Enns & Kellstedt 2008, Prior 2014):

Some people think the government should provide fewer services even in areas such as health and education in order to reduce spending. Other people feel it is important for the government to provide many more services even if it means an increase in spending. Where would you place yourself on this scale, or haven't you thought much about this?

Such a question is phrased at a relatively general level of assessment and taps general support for government involvement in society. Even someone who pays only the most casual attention to politics could answer this question and—in doing so—meaningfully draw upon a base of politically relevant considerations.

Moving down the ladder of generality in a related realm, consider the following item, used by Hacker & Pierson (2005) to assess attitudes toward tax policy in a more specific way:

Which of these would you prefer—a large tax cut plan that provides an across-the-board tax cut for everyone, or a smaller tax cut plan that provides targeted tax cuts mainly for lower and middle-income people?

This question taps a more specific concern about the role of government, relating to the particulars of tax policy, but it still proceeds at a somewhat general level. Citizens who pay only moderately close attention to tax policy might be able to formulate a reasonable answer to the survey item.

Moving further down (and from the realm of academic research to the larger political world), in the wake of the implementation of the 2010 Affordable Care Act, the Washington Post/Kaiser poll asked an even more specific question in the summer of 2012:

Which of these two descriptions comes closer to your view of what Medicare should look like in the future? OPTION A: Medicare should continue as it is today, with the government guaranteeing all seniors the same set of health insurance benefits. OR, OPTION B: Medicare should be changed to a system in which the government guarantees each senior a fixed amount of money to help them purchase coverage either from traditional Medicare or from a list of private health plans.

This question, though seemingly reasonable to someone steeped in politics, almost certainly has little relevance to the casual observer of that world, especially someone who has no direct experience with Medicare. It is full of jargon, but that is not the central problem. The question references particular government programs. Moreover, it contrasts those programs with private health plans on various cost dimensions, with no discussion of the types of benefits under either plan. Such questions tap into the specifics of government programs and might flummox the most devout political junkie. Thus, even some respondents who tried to answer the question might construct an answer that would not reflect what the question seems designed to measure—namely their considerations regarding government management of seniors' health care. Such an answer could be little more than statistical noise.

These three questions, chosen to represent different points on the continuum of specificity concerning the direct involvement of government in the lives of ordinary citizens, serve as a window into the possibility of measuring mezzo-level opinion with a single item. As the items become more specific, referencing the particulars of policy options and demanding greater expertise on the part of the respondents, they move into a realm in which they may not properly tap public sentiment. But if a researcher desires to measure public opinion with a single question, framing questions at a broader level of generality may make it possible to accurately gauge public sentiment.

Such an approach has been adopted by several scholars. In my own work on social welfare policy (Berinsky 2002, 2004), I examined items that measured these concerns at a broad level—including the ANES measure on the trade-off between services and spending detailed above. Jacoby (2000) adopted an analogous strategy in his study of framing and social welfare policy. Similarly, in the domain of war, Mueller (1973, p. 43) makes the case for the primacy of general questions, arguing that a question that asks if it was a “mistake” to go to war “asks for the respondent’s general opinion on the wisdom of the war venture itself, and thus seems to be a sound measure of a sort of general support for the war.”

Aggregation approach. Single-item measurement approaches tap only one dimension of the process of measuring public opinion. A second way to gauge general public sentiment is to be a “lumper” rather than a “splitter” (Druckman & Jacobs 2006) and consider the answers that respondents give across a number of survey questions. Such a strategy involves aggregating multiple items into a single measure at the individual level and then aggregating across individuals to get a measure of public opinion. This approach is more common, and it has become especially so over the last 20 years.

There are several ways to justify combining the answers individuals give to multiple survey questions into a single measure of preference. The most common justification is a measurement-error approach. Achen (1975) argued that any survey question measures the respondent’s political preferences with measurement error (see also Erikson 1978, 1979). Subsequently, scholars have used this insight to argue that any single survey question can be considered an indicator of the

AGGREGATING OPINIONS

A “test” mindset drives the measurement strategy employed in much of the current wave of representation literature (e.g., Clinton et al. 2004). Bafumi & Herron (2010) make the strongest case for such an approach, arguing that the best way to study the congruence between legislators and their constituencies is to examine congruence in the ideal points of representatives and the mass public—ideal points constructed using item-response techniques. This same logic drives Stimson’s path-breaking work on aggregate policy mood. His discussion of the movement over time of preferences toward spending across a variety of policy areas in various treatments (Stimson 1999 [1991], 2015 [2004]) provides the clearest incidence of the logic of aggregation. For Stimson, each question asking respondents to evaluate a policy is a question “testing” an underlying preference for spending. By examining the results of these tests in the aggregate, we can learn about the underlying shape of mass preferences toward government involvement in society more generally.

underlying general attitudes, interests, and values of the respondents. But each single question is imperfect and measures the respondents’ attitudes with error. Thus, aggregating across several survey items reduces the noise in the measures of public opinion and produces better estimates of respondents’ underlying issue preferences. Ansolabehere et al. (2008) employ this approach by asking respondents multiple survey questions on each issue and creating an average, or score, on that issue area. They show that this averaging approach creates individual responses that are highly stable over time (see sidebar titled *Aggregating Opinions*). To return to Zaller’s (1992) language, each answer reveals a snapshot of the considerations at the top of a respondent’s head at the time they answered the question. By aggregating across several questions, researchers can combine several of these snapshots together, separating signal from noise.

This measurement-error approach to measuring public opinion has spawned a great deal of work in recent years. But like the single-item approach to gauging public sentiment, this strategy comes in several varieties, proceeding in different ways. Paralleling my discussion in the last section, the most important distinction is the level of abstraction in the final measures of public opinion advanced by different scholars. **Table 1** provides an overview of three levels.

The first level considers public opinion as a broad measure, aggregated across a wide spectrum of issues—a “mood,” to use Stimson’s (1999 [1991]) term. Stimson argues that to properly measure opinion, we need to aggregate survey responses across both individuals and across issues. Setting aside the issue of aggregating across individuals to form constituencies (a question taken up indirectly in the first part of this article with the discussion of sampling), Stimson makes a

Table 1 Levels of abstraction in the measurement of public opinion

Level of abstraction	Dimensions of opinion	Examples
High	Single dimension—“policy mood”	Durr (1993), Ellis & Faricy (2013), Enns & Koch (2013), Jessee (2009), Luttig (2013), Stimson (1999 [1991], 2015 [2004]), Erikson et al. (2002), Tausanovitch & Warshaw (2013), Ura (2014), Wlezien (1995)
Medium	Two dimensions—economic and social	Ansolabehere et al. (2008), Carmines et al. (2012), Ellis & Stimson (2012), Jost et al. (2009), Kelly & Chambliss (1966), Klar (2014), Knoke (1979), Layman & Carsey (2002), Moskowitz & Jenkins (2004), Treier & Hillygus (2009)
Low	Multiple dimensions—policy-specific measures	Atkinson et al. (2011), Brulle et al. (2012), Enns (2014), Kellstedt (2000), Mulligan et al. (2013), Nicholson-Crotty et al. (2009), Shapiro & Bloch-Elkon (2006), Wilcox & Norranger (2001)

critical contribution in arguing that aggregating across issue questions allows us to gauge public opinion at its proper level of measurement—a high level of generality. Stimson is not interested in the responses citizens give to particular issue questions in and of themselves. Instead, Stimson views these items as mere indicators of an underlying construct that captures people’s general political predilections—akin to Zaller’s considerations (as discussed above).

Consider, for instance, preferences over spending on specific programs, such as education or job creation. We can ask respondents if the government should spend more, less, or the same on a series of particular initiatives. Each of these questions relates to a different program, but what matters for Stimson—what motivates mood—is the high-level preference for what government should do generally: Should government be bigger or smaller? Each spending question is a single answer to this question. Taken together, the average response to a series of questions provides a broader and more well-grounded answer to that larger question. Just as individual survey respondents are, to use Stimson’s language, “spokespeople for the herd” (1999 [1991], p. 2), individual survey questions are mere indicators of underlying general sentiment toward government action.

The “public policy mood” that we extract from these items therefore taps the broad “welfare state/size of government controversy that divides the parties” (Stimson 2015 [2004], p. 81). By asking many specific questions, Stimson casts a wide net, extracting general sentiment about policy through a series of specific tests. This far-reaching view gives us a broader window into the distribution of the underlying predispositions of the American public.

Empirically, this notion of mood has a great deal of power. Individual issue questions may at first glance appear to measure distinct indicators of policy sentiment, but in a series of updated analyses of mood over the last 25 years, Stimson argues that almost all specific issues have been subsumed by this single dimension. Compare, for example, Stimson’s discussion of abortion in *Public Opinion in America* (1999 [1991]), which first laid out the concept of mood, to its treatment in *Tides of Consent* (2015 [2004]). Whereas data through 1989 suggested that the abortion issue was “a domain by itself,” once Stimson incorporated polling data collected through 2012, he concluded that “abortion attitudes are beginning to align with the standard left-right debate that I have called the welfare state” (2015 [2004], p. 50).

Scholars have long employed Stimson’s measure of mood to measure public opinion, either whole cloth (Durr 1993, Erikson et al. 2002) or through a similar measure (Wlezien 1995, Enns & Koch 2013), and it is still widely used in political science (Ellis & Faricy 2013, Luttig 2013, Ura 2014). What is important for present purposes, though, is Stimson’s insight that public opinion can be best captured by a measure that aggregates broadly across a variety of seemingly distinct policies. Such an approach has been adopted widely in the field, even among those scholars who do not trace their intellectual lineage to Stimson’s work. In this vein, consider Jessee’s (2009) study of spatial voting. Jessee asks respondents whether they would support or oppose 31 proposals that were put to a vote in the US Senate. The topics range from minimum-wage increases to a declaration “that marriage in the United States shall consist only of the union of a man and a woman.” These bills therefore encompass both issues of economic policy and social policies. Jessee finds that the preferences are best explained by a single-dimensional model and proceeds to treat mass preferences as a one-dimensional construct, akin to Stimson’s mood measure (though Jessee himself does not make such a connection). More generally, Jessee’s approach here is quite similar to that of other scholars of representation who adopt item response theory (IRT) models to measure public sentiment, such as Tausanovitch & Warshaw (2013), who use an IRT model to estimate single-dimension state- and local-level public opinion, which they label “policy preferences.”⁷

⁷This approach extends beyond the United States. A similar philosophical approach has even been extended to a dimensional measurement of the public’s preferences in China (Pan & Xu 2015).

Moving down levels of abstraction in **Table 1**, other scholars advocate aggregating public opinion across different survey items but argue that combining items across different policy domains—as Stimson does—goes too far. A common approach is to consider two broad bases of opinion, rather than just one. Specifically, scholars distinguish between preferences over economic issues and those over social issues (reviewed by Carmines & d’Amico 2015). Even Stimson—the primary proponent of the single-dimension approach described above—has moved to considering an explicitly multidimensional approach. Ellis & Stimson (2012) advance a two-dimensional model of public opinion, which allows for both economic and cultural domains.⁸

A third approach moves even further down the scale of generality, generating issue-specific “moods” in the vein of Stimson but at a lower level of aggregation [see, e.g., Kellstedt (2000) on race; Nicholson-Crotty et al. (2009) and Enns (2014) on criminal justice policies; Wilcox & Norranger (2001) on moral issues; Mulligan et al. (2013) on cultural policies; Shapiro & Bloch-Elkon (2006) on foreign policy; and Brulle et al. (2012) on climate change policies]. Notably, Stimson himself has begun with his collaborators (Atkinson et al. 2011) to create policy-specific mood estimates. All of these authors still aggregate across questions to create a measure of public opinion, but they do so in a more limited issue-area scope. Essentially, they argue that even two policy domains—economic and social—do not parse public opinion finely enough to accurately characterize the policy space of public opinion.⁹

A Way Forward

Even if we consider the goal to measure public opinion at a mezzo level—as I have argued we should—there is no obvious prescription for how best to proceed. Each of the strategies described here has its drawbacks. First, consider the single-question strategy. Although the use of general questions should in theory force respondents to evaluate politics at a broader frame of reference, in practice, such questions might induce noisy survey responses. Research in psychology has shown that people sometimes answer general survey questions with a specific frame of reference in mind—one that can differ both across people and across situations (Tourangeau et al. 2000). For instance, when answering the most general “services and spending” question described above, some individuals might consider their preferences over healthcare spending, others might think of education reform, and still others might have something altogether different in mind. These shifting frames could alter the types of considerations that come to mind when people answer survey questions, thereby leading to fuzzier measures of public opinion.

The strategy of achieving generality by aggregating across questions for each individual respondent comes with potential problems as well. Broockman (2016) criticizes the use of ideological scales that compute the opinions of individuals by taking their average position on a variety of items—whether a straight average or the weighted average implied by factor analysis or IRT models. He argues that these scales are flawed measures of individual preferences because they inappropriately aggregate items across policy domains. In essence, he rejects the Stimson single-dimension approach in favor of the more specific approaches referenced in **Table 1**. Broockman argues that these measures do not capture the true positions of voters but instead measure their ideological consistency. Consider, for instance, a voter who takes a strong liberal position on one issue and a strong conservative position on another. The average of these positions would classify

⁸However, by Ellis & Stimson’s (2012) account, these two dimensions are highly correlated. Thus, the change from the original Stimson formulation is not as great as would appear at first blush.

⁹Of course, still other authors consider single-item measures of opinion on their own terms, even within a single domain; see, e.g., Lax & Philips (2009) and, on a larger scale, Page & Shapiro (1992).

that individual as an ideological moderate, but according to Broockman, that individual is best seen as ideologically mixed. After all, on neither issue can he or she be properly classified as moderate. As Broockman himself notes, this general idea is not new—psychologists and public opinion scholars have long struggled with the conflation of opinion “extremity” and “intensity” (Krosnick et al. 1993, Baron et al. 1996; for an early take, see Allport & Hartman 1925). But Broockman’s work is important for calling this longstanding concern to the attention of a new generation of scholars.

This is not, however, the final word on the matter. Broockman (2016) identifies an important limitation of some existing work, but if we take his argument to its logical conclusion, we may attempt to measure opinion at a level at which it does not exist. Broockman adopts a particular measurement strategy—respondents choose among very specific responses to very detailed questions. For example, on health care, he asks respondents which of seven detailed policy options they most prefer. One of these policy options references Britain’s government-sponsored health-care system, and another mentions tort reform among other specific changes. Respondents may provide answers to these kinds of questions, but only because we, as researchers, have asked them to do so in the context of the survey interview. At a theoretical level, asking people to answer the kinds of questions Broockman poses—and taking those answers seriously in a world in which we know people pay little attention to politics but are happy to answer our survey questions—goes too far down the levels of abstraction at which we can meaningfully measure individual opinions. Put another way, adopting Broockman’s approach may introduce its own pathologies; we may attribute genuine preferences where only thin whims exist, ending up with the illusion that Bishop (2008) rightly criticizes.

With these warnings in mind, there is a principled way to go about measuring opinion at the mezzo level. At the individual level, adopting general survey questions gives us a window into what individuals think about the political world. But we must be careful to pay attention to the larger context in which they are answering those questions. Similarly, moving to the multi-question measurement strategy, the use of issue scales can provide valuable information about the distribution of considerations of individuals—the politically relevant needs and desires of the mass public. But we would do well to heed Broockman’s (2016) warning to tread lightly when we construct such scales. In particular, when we approach the scaling of attitudes, our theory of what goes with what should be well tied to the particular empirical problems that we want to address. Broockman is correct that aggregating measures of opinions on issues that do not go together—items that tap empirically distinct issue domains—will lead us astray. The use of any issue-specific measure needs to come with an explicit argument about why we should believe that most respondents have at least partially thought-out attitudes, given the two observations I introduced at the beginning of this discussion regarding how individuals answer survey questions. Only then can any measure report attitudes that are well informed by respondents’ interactions with the political world.

My advice is simple. We should ask general survey questions that demand little specific expertise on the part of survey respondents, and we should combine these questions into measures of opinion in theoretically meaningful ways. Of course, this is not a hard and fast rule. On some issues—those in the public eye—it is possible to gather meaningful data on the public’s specific preferences. But as a general guiding principle, measurement should always go hand in hand with an argument about the validity of that measurement strategy.

Relying on theory is especially important because arguments about the meaningfulness of measures of public opinion cannot and will not be settled by hypothesis tests. Different tests of dimensionality may provide different interpretations of the same data (see Tabachnick & Fidell 1996, Kieffer 1999). For example, Stimson began in the 1980s by assuming that any different

dimensions of opinion would be orthogonal to one another. By this method, the first dimension was strong and clear, but the second dimension was a residual grouping of sorts, “a mishmash of stray issues that appear to have nothing in common with one another” (Stimson 2012, p. 26). However, by relaxing the assumption of orthogonality, a coherent second dimension emerges—albeit one highly correlated with the first dimension. This result underscores the fact that statistical analysis alone cannot uncover the proper way to measure opinion. Rather, adjudicating this kind of dimensional analysis is a politically informed way to approach the measurement of opinion. And that approach depends in large part on the decisions made by the researcher.

CONCLUSION

Polls are powerful tools for measuring popular sentiment. But we cannot simply ask any question of any citizen and expect to get meaningful information back. In this review, I have discussed how two critical decisions—whom to ask our questions, and what kinds of questions to ask—can profoundly shape our reading of the public will. Researchers must strive to develop sampling methods that can enable us to accurately represent the mass public at a time when changing communication technologies have undermined the traditional methods of survey sampling. The way forward is unclear, but it is a critical problem to address. At the same time, we cannot let this search for a technical solution to our problems allow us to forget that our survey responses are only as good as the questions we ask our respondents. There needs to be more systematic thinking about the level at which it is appropriate to ask survey questions. If we are careful with our tools, polls can serve democracy well.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

For helpful discussions regarding this article, I thank Michael Barber, David Broockman, Jamie Druckman, Jeremy Pope, and Chris Warshaw. I am grateful to Justin de Benedictis-Kessner, Nicolas Dumas, Meg Goldberg, Tom O’Grady, and Robert Pressel for research assistance. I also thank Jeff Lax and Melissa Schwartzberg for inviting me to speak at a conference on representation they sponsored. My remarks there contained the seeds of this article.

LITERATURE CITED

- Achen CH. 1975. Mass political attitudes and the survey response. *Am. Polit. Sci. Rev.* 69(4):1218–31
- Allport FH, Hartman DA. 1925. The measurement and motivation of atypical opinion in a certain group. *Am. Polit. Sci. Rev.* 19(4):735–60
- Ansolahehere S, Rodden J, Snyder JM Jr. 2008. The strength of issues: using multiple measures to gauge preference stability, ideological constraint, and issue voting. *Am. Polit. Sci. Rev.* 102(2):215–32
- Atkinson M, Baumgartner F, Coggins K, Stimson J. 2011. *Mood and agendas: developing policy-specific conceptions of mood*. Presented at Annu. Meet. Midwest Polit. Sci. Assoc., 69th, Chicago, Mar. 31–Apr. 3
- Bafumi J, Herron MC. 2010. Leapfrog representation and extremism: a study of American voters and their members in Congress. *Am. Polit. Sci. Rev.* 104(3):519–42
- Baker R, Blumberg SJ, Brick SM, Couper MP, Courtright M, et al. 2010. AAPOR report on online panels. *Public Opin. Q.* 74(4):711–81

- Barber MJ, Mann CB, Monson JQ, Patterson KD. 2014. Online polls and registration-based sampling: a new method for pre-election polling. *Polit. Anal.* 22(3):1–15
- Baron RS, Hoppe SI, Kao CF, Brunzman B, Linneweh B, Rogers D. 1996. Social corroboration and opinion extremity. *J. Exp. Soc. Psychol.* 32(6):537–60
- Bartels LM. 2003. Democracy with attitudes. In *Electoral Democracy*, ed. G Rabinowitz, M MacKuen, pp. 48–82. Ann Arbor: Univ. Mich. Press
- Berinsky AJ. 2002. Silent voices: social welfare policy opinions and political equality in America. *Am. J. Polit. Sci.* 46(2):276–87
- Berinsky AJ. 2004. *Silent Voices: Public Opinion and Political Participation in America*. Princeton, NJ: Princeton Univ. Press
- Bishop GF. 2008. Rational public opinion or its manufacture? Reply to Page. *Crit. Rev.* 20(1–2):141–57
- Blumberg SJ, Luke JV. 2009. Reevaluating the need for concern regarding noncoverage bias in landline surveys. *Am. J. Public Health* 99(10):1806–10
- Blumberg SJ, Luke JV. 2016. *Wireless substitution: early release of estimates from the National Health Interview Survey, July–December 2015*. Natl. Health Interview Surv. Early Release Program 1–13, Natl. Cent. Health Stat., Hyattsville, MD. <https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201605.pdf>
- Blumer H. 1954. What is wrong with social theory? *Am. Soc. Rev.* 19(1):3–10
- Brians CL, Wattenberg MP. 1996. Campaign issue knowledge and salience: comparing reception from TV commercials, TV news and newspapers. *Am. J. Polit. Sci.* 40(1):172–93
- Brick JM. 2011. The future of survey sampling. *Public Opin. Q.* 75(5):872–88
- Brick JM, Dipko S, Presser S, Tucker C, Yuan Y. 2006. Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opin. Q.* 70(5):780–93
- Broockman DE. 2016. Approaches to studying policy representation. *Legis. Stud. Q.* 41(1):181–215
- Brulle RJ, Carmichael J, Jenkins JC. 2012. Shifting public opinion on climate change: an empirical assessment of factors influencing concern over climate change in the US, 2002–2010. *Clim. Change* 114(2):169–88
- Carmines EG, D’Amico NJ. 2015. The new look in political ideology research. *Annu. Rev. Polit. Sci.* 18:205–16
- Carmines EG, Ensley MJ, Wagner MW. 2012. Political ideology in American politics: one, two, or none? *Forum* 10(3):4
- Clark T. 2015. New research suggests why general election polls were so inaccurate. *Guardian* 12 Nov. <http://www.theguardian.com/politics/2015/nov/13/new-research-general-election-polls-inaccurate>
- Clinton JD, Jackman S, Rivers D. 2004. The statistical analysis of roll call data. *Am. Polit. Sci. Rev.* 98(2):355–70
- Converse PE. 1987. Changing conceptions of public opinion in the political process. *Public Opin. Q.* 51:S12–24
- Converse PE. 1990. Popular representation and the distribution of information. In *Information and Democratic Processes*, ed. J Ferejohn, J Kuklinski, pp. 369–88. Champaign: Univ. Illinois Press
- De Leeuw E, Callegaro M, Hox J, Korendijk E, Lensvelt-Mulders G. 2007. The influence of advance letters on response in telephone surveys: a meta-analysis. *Public Opin. Q.* 71(3):413–43
- Druckman JN, Jacobs LR. 2006. Lumpers and splitters the public opinion information that politicians collect and use. *Public Opin. Q.* 70(4):453–76
- Durr RH. 1993. What moves policy sentiment? *Am. Polit. Sci. Rev.* 87(1):158–70
- Ellis C, Faricy C. 2013. Social policy and public opinion: how the ideological direction of spending influences public mood. *J. Polit.* 73(4):1095–110
- Ellis C, Stimson JA. 2012. *Ideology in America*. New York: Cambridge Univ. Press
- Enns PK. 2014. The public’s increasing punitiveness and its influence on mass incarceration in the United States. *Am. J. Polit. Sci.* 58(4):857–72
- Enns PK, Kellstedt PM. 2008. Policy mood and political sophistication: why everybody moves mood. *Br. J. Polit. Sci.* 38(3):433–54
- Enns PK, Koch J. 2013. Public opinion in the U.S. states: 1956 to 2010. *State Polit. Policy Q.* 13(3):349–72
- Erikson RS. 1978. Analyzing one-variable three-wave panel data: a comparison of two models. *Polit. Methodol.* 5(2):151–67
- Erikson RS. 1979. The SRC panel data and mass political attitudes. *Br. J. Polit. Sci.* 9(1):89–114

- Erikson RS, MacKuen MB, Stimson JA. 2002. *The Macro Polity*. Cambridge, UK: Cambridge Univ. Press
- Galea S, Tracy M. 2007. Participation rates in epidemiologic studies. *Ann. Epidemiol.* 17(9):643–53
- Geer JG. 1996. *From Tea Leaves to Opinion Polls: A Theory of Democratic Leadership*. New York: Columbia Univ. Press
- Gelman A, Goel S, Rivers D, Rothschild D. 2016. The mythical swing voter. *Q. J. Polit. Sci.* 11(1):103–30
- Ginsberg B. 1986. *The Captive Public: How Mass Opinion Promotes State Power*. New York: Basic Books
- Groves RM. 2006. Nonresponse rates and nonresponse bias in household surveys. *Public Opin. Q.* 70(5):646–75
- Groves RM. 2011. Three eras of survey research. *Public Opin. Q.* 75(5):861–71
- Groves RM, Fowler FJ Jr., Couper MP, Lepkowski JM, Singer E, Tourangeau R. 2009. *Survey Methodology*. Hoboken, NJ: Wiley
- Guterbock TM, Lavrakas PJ, Tompson TN, ZuWallack R. 2011. Cost and productivity ratios in dual-frame RDD telephone surveys. *Surv. Pract.* 4(2):1–7
- Hacker JS, Pierson P. 2005. Abandoning the middle: the Bush tax cuts and the limits of Democratic control. *Perspect. Polit.* 3(01):33–53
- Herbst S. 1993. *Numbered Voices: How Opinion Polling Has Shaped American Politics*. Chicago: Univ. Chicago Press
- Hillygus DS. 2016. The practice of survey research: changes and challenges. In *New Directions in Public Opinion*, ed. A Berinsky, pp. 34–53. New York: Routledge. 2nd ed.
- Hillygus DS, Jackson N, Young M. 2014. Professional respondents in non-probability online panels. In *Online Panel Research: A Data Quality Perspective*, ed. M Callegaro, R Baker, J Bethlehem, AS Göritz, JA Krosnick, PJ Lavrakas, pp. 219–37. New York: Wiley
- Hu SS, Balluz L, Battaglia MP, Frankel MR. 2011. Improving public health surveillance using a dual-frame survey of landline and cell phone numbers. *Am. J. Epidemiol.* 173(6):703–11
- Jacoby WG. 2000. Issue framing and public opinion on government spending. *Am. J. Polit. Sci.* 44(4):750–67
- Jessee SA. 2009. Spatial voting in the 2004 presidential election. *Am. Polit. Sci. Rev.* 103(1):59–81
- Jost JT, Federico CM, Napier JL. 2009. Political ideology: its structure, functions, and elective affinities. *Annu. Rev. Psychol.* 60:307–37
- Keeter S. 2006. The impact of cell phone noncoverage bias on polling in the 2004 presidential election. *Public Opin. Q.* 70(1):88–98
- Keeter S, Dimock M, Christian L. 2008. Calling cell phones in '08 pre-election polls. News release, Pew Res. Cent. People Press, Washington, DC. <http://www.people-press.org/files/legacy-pdf/cell-phone-commentary.pdf>
- Kellstedt PM. 2000. Media framing and the dynamics of racial policy preferences. *Am. J. Polit. Sci.* 44(2):245–60
- Kelly KD, Chambliss WJ. 1966. Status consistency and political attitudes. *Am. Sociol. Rev.* 31(3):375–82
- Kennedy C. 2007. Evaluating the effects of screening for telephone service in dual frame RDD surveys. *Public Opin. Q.* 71(5):750–71
- Key VO. 1961. *Public Opinion and American Democracy*. New York: Knopf
- Kieffer KM. 1999. An introductory primer on the appropriate use of exploratory and confirmatory factor analysis. *Res. Schools* 6(2):75–92
- Kim J, Gershenson C, Glaser P, Smith TW. 2011. The polls—trends: trends in surveys on surveys. *Public Opin. Q.* 71(5):165–91
- Kish L. 1965. *Survey Sampling*. New York: Wiley
- Klar S. 2014. A multidimensional study of ideological preferences and priorities among the American public. *Public Opin. Q.* 78(S):344–59
- Knoke D. 1979. Stratification and the dimensions of American political orientations. *Am. J. Polit. Sci.* 23(4):772–91
- Krosnick JA, Boninger DS, Chuang YC, Berent MK, Carnot CG. 1993. Attitude strength: one construct or many related constructs? *J. Pers. Soc. Psychol.* 65(6):1132–51
- Lavrakas PJ, Shuttles CD, Steeh C, Fienberg H. 2007. The state of surveying cell phone numbers in the United States 2007 and beyond. *Public Opin. Q.* 71(5):840–54
- Lavrakas PJ, Tompson TN, Benford R, Fleury C. 2010. *Investigating data quality in cell phone surveying*. Presented at Annu. Conf. Am. Assoc. Public Opin. Res., 65th, Chicago, May 13–16

- Lax JR, Phillips JH. 2009. Gay rights in the states: public opinion and policy responsiveness. *Am. Polit. Sci. Rev.* 103(3):367–86
- Layman GC, Carsey TM. 2002. Party polarization and “conflict extension” in the American electorate. *Am. J. Polit. Sci.* 46(4):786–802
- Levine B, Harter R. 2015. Optimal allocation of cell-phone and landline respondents in dual-frame surveys. *Public Opin. Q.* 79(1):91–104
- Lohr SL. 2010. *Dual frame surveys: recent developments and challenges*. Presented at Meet. Italian Stat. Soc., 45th, Padua, Jun. 29–Jul. 1
- Lumley T. 2010. *Complex Surveys: A Guide to Analysis Using R*. Hoboken, NJ: Wiley
- Luttig M. 2013. The structure of inequality and Americans’ attitudes towards redistribution. *Public Opin. Q.* 77(3):811–21
- Massey DS, Tourangeau R. 2013. Where do we go from here? Nonresponse and social measurement. *Ann. Am. Acad. Polit. Soc. Sci.* 645(1):222–36
- Moore DW. 2008. *The Opinion Makers: An Insider Exposes the Truth Behind the Polls*. Boston: Beacon
- Moskowitz AN, Jenkins JC. 2004. Structuring political opinions. *Sociol. Q.* 45(3):395–419
- Mueller J. 1973. *Wars, Presidents and Public Opinion*. New York: Wiley
- Mulligan K, Grant T, Bennett D. 2013. The dynamics of public opinion on cultural policy issues in the US, 1972–2010. *Polit. Behav.* 35(4):807–29
- Murphy J, Link MW, Childs JH, Tesfaye CL, Dean E, et al. 2014. Social media in public opinion research: executive summary of the AAPOR Task Force on emerging technologies in public opinion research. *Public Opin. Q.* 78(4):788–94
- Newman B. 2003. Integrity and presidential approval, 1980–2000. *Public Opin. Q.* 67(3):335–67
- Nicholson-Crotty S, Peterson DA, Ramirez MD. 2009. Dynamic representation(s): federal criminal justice policy and an alternative dimension of public mood. *Polit. Behav.* 31(4):629–55
- O’Connor B, Balasubramanyan R, Routledge BR, Smith NA. 2010. *From tweets to polls: linking text sentiment to public opinion time series*. Presented at Int. Conf. Weblogs and Social Media, 4th, Washington, May 23–26
- Page BI, Shapiro RY. 1992. *The Rational Public: Fifty Years of Trends in Americans’ Policy Preferences*. Chicago: Univ. Chicago Press
- Pan J, Xu Y. 2015. *China’s ideological spectrum*. Res. pap., Dep. Polit. Sci., Mass. Inst. Technol.
- Pew Res. Cent. 2012. *Assessing the representativeness of public opinion surveys*. Washington, DC: Pew Res. Cent. <http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/>
- Prior M. 2014. Visual political knowledge: a different road to competence? *J. Polit.* 76(01):41–57
- Schwarz N. 1996. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*. Mahwah, NJ: Erlbaum
- Shapiro RY, Bloch-Elkon Y. 2006. *Political polarization and the rational public*. Presented at Annu. Conf. Am. Assoc. Public Opin. Res., Montreal, 61st, May 18–21
- Squire P. 1988. Why the 1936 literary digest poll failed. *Public Opin. Q.* 52(1):125–33
- Stimson JA. 1999 (1991). *Public Opinion in America: Moods, Cycles, and Swings*. Boulder, CO: Westview. 2nd ed.
- Stimson JA. 2012. On the meaning and measurement of mood. *Daedalus* 141(4):23–34
- Stimson JA. 2015 (2004). *Tides of Consent: How Public Opinion Shapes American Politics*. Chapel Hill: UNC Press. 2nd ed.
- Tabachnick BG, Fidell LS. 1996. *Using Multiple Statistics*. New York: HarperCollins
- Tausanovitch C, Warshaw C. 2013. Measuring constituent policy preferences in Congress, state legislatures and cities. *J. Polit.* 75(2):330–42
- Tourangeau R. 2004. Survey research and societal change. *Annu. Rev. Psychol.* 55:775–801
- Tourangeau R, Rasinski KA. 1988. Cognitive processes underlying context effects in attitude measurement. *Psychol. Bull.* 103(3):299–314
- Tourangeau R, Rips LJ, Rasinski K. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge Univ. Press
- Treier SD, Hillygus DS. 2009. The nature of political ideology in the contemporary electorate. *Public Opin. Q.* 73(4):679–703

- Ura JD. 2014. Backlash and legitimation: macro political responses to Supreme Court decisions. *Am. J. Polit. Sci.* 58(1):110–26
- Verba S. 1996. The citizen as respondent: sample surveys and American democracy. *Am. Polit. Sci. Rev.* 90(1):1–7
- Wilcox C, Norrander B. 2001. Of mood and morals: the dynamics of opinion on abortion and gay rights. In *Understanding Public Opinion*, ed. B Norrander, C Wilcox, pp. 121–48. Washington, DC: CQ Press
- Wlezien C. 1995. The public as thermostat: dynamics of preferences for spending. *Am. J. Polit. Sci.* 39(4):981–1000
- Zaller J. 1992. *The Nature and Origins of Mass Opinion*. Cambridge, UK: Cambridge Univ. Press

Contents

Politics, Academics, and Africa <i>Robert H. Bates</i>	1
Qualitative Methods <i>John Gerring</i>	15
Just War Theory: Revisionists Versus Traditionalists <i>Seth Lazar</i>	37
International Courts: A Theoretical Assessment <i>Clifford J. Carrubba and Matthew Gabel</i>	55
Political Economy of Taxation <i>Edgar Kiser and Steven M. Karceski</i>	75
Comparing Political Values in China and the West: What Can Be Learned and Why It Matters <i>Daniel A. Bell</i>	93
Culture, Politics, and Economic Development <i>Paul Collier</i>	111
Progovernment Militias <i>Sabine C. Carey and Neil J. Mitchell</i>	127
Voter Identification Laws and Turnout in the United States <i>Benjamin Highton</i>	149
Climate Change and International Relations (After Kyoto) <i>Arild Underdal</i>	169
Social Movement Theory and the Prospects for Climate Change Activism in the United States <i>Doug McAdam</i>	189
Climate Change: US Public Opinion <i>Patrick J. Egan and Megan Mullin</i>	209
The Political Economy of Regional Integration <i>Christina J. Schneider</i>	229

Bureaucracy and Service Delivery <i>Thomas B. Pepinsky, Jan H. Pierskalla, and Audrey Sacks</i>	249
Feminist Theory Today <i>Kathy E. Ferguson</i>	269
When Does Globalization Help the Poor? <i>Nita Rudra and Jennifer Tobin</i>	287
Measuring Public Opinion with Surveys <i>Adam J. Berinsky</i>	309
Conflict and Cooperation on Nuclear Nonproliferation <i>Alexandre Debs and Nuno P. Monteiro</i>	331
From a Deficit of Democracy to a Technocratic Order: The Postcrisis Debate on Europe <i>Ignacio Sánchez-Cuenca</i>	351
Understanding the Political Economy of the Eurozone Crisis <i>Jeffrey Frieden and Stefanie Walter</i>	371
The Electoral Consequences of Corruption <i>Catherine E. De Vries and Hector Solaz</i>	391
Labor Unions, Political Representation, and Economic Inequality <i>John S. Ahlquist</i>	409
Coding the Ideological Direction and Content of Policies <i>Joshua D. Clinton</i>	433
Wealth Inequality and Democracy <i>Kenneth Scheve and David Stasavage</i>	451
The New New Civil Wars <i>Barbara F. Walter</i>	469
State Building in the Middle East <i>Lisa Blaydes</i>	487
Information, Uncertainty, and War <i>Kristopher W. Ramsay</i>	505
Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges <i>John Wilkerson and Andreu Casas</i>	529
Trading in the Twenty-First Century: Is There a Role for the World Trade Organization? <i>Judith Goldstein</i>	545

Police Are Our Government: Politics, Political Science, and the
Policing of Race–Class Subjugated Communities
Joe Soss and Vesla Weaver 565

Errata

An online log of corrections to *Annual Review of Political Science* articles may be found
at <http://www.annualreviews.org/errata/polisci>