

Using Experimental Vignettes in Telephone Surveys to Study how Survey Methods and Findings Affect the Public's Evaluation of Public Opinion Polls. Considering a Dual Process Approach

Allyson L. Holbrook¹ · Paul J. Lavrakas^{1,2} · Timothy P. Johnson¹ ·
Andrew Crosby³ · Polina Polskaia⁴ · Xiaoheng Wang⁵ · Xiaoyan Hu⁸ ·
Evgenia Kapousouz¹ · Young Ik Cho⁶ · Henning Silber⁷

¹University of Illinois at Chicago, Department of Public Policy, Management, and Analytics

²Independent Consultant

³University of California, Riverside, School of Public Policy

⁴Pace University, Department of Public Administration, Dyson College

⁵Wichita State University, Hugo Wall School of Public Affairs

⁶University of Wisconsin - Milwaukee, Joseph J. Zilber College of Public Health

⁷University of Michigan, Survey Methodology Program, Survey Research Center at the Institute for Social Research

⁸Oakland University, Department of Political Science, College of Arts and Sciences

Understanding how the public thinks about and understands survey findings is an important part of understanding the role of surveys in policymaking and democracy more broadly. In this paper, we examine the results of three vignette experiments conducted within representative sample surveys in which the methodology and results of a hypothetical public opinion survey assessing support for specific public policy proposals was manipulated. Respondents' beliefs about the accuracy of the survey described in the vignette experiment, and their beliefs about whether or not it should be considered by policymakers were measured. We used both manipulated (e.g., the methodological rigor of the survey described in the vignette) and measured (e.g., respondents' opinions on the proposed policy about which public opinion was measured in the vignette) independent variables to test four different theoretical models that could be used to explain public evaluations of public opinion surveys: (1) the rational actor model, which suggests that people will evaluate more methodologically rigorous surveys more positively; (2) the science literacy model, which suggests that people high in science literacy will evaluate more methodologically rigorous surveys more positively than will people low in science literacy; (3) the motivated reasoning model, which suggests that people will evaluate surveys more positively when the survey results are consistent with their prior opinions than when they are inconsistent; and (4) a dual process model approach, which suggests that people will evaluate more methodologically rigorous surveys more positively only when they are both able and motivated to do so. We found some support for the scientific literacy and motivated reasoning models, but these findings were qualified by an interaction between factors associated with respondent motivation, ability, and survey methodology rigor that strongly supports the dual process model perspective.

Keywords: survey climate; motivated reasoning; dual processing models; vignette research; public opinion

Supplementary Information The online version of this article (<https://doi.org/10.18148/srm/2025.v19i2.8261>) contains supplementary information.

Corresponding author: Allyson L. Holbrook, Department of Public Policy, Management, and Analytics, University of Illinois at Chicago, 400 S. Peoria St., Chicago, IL 2100 AEH (MC278), USA (Email: allyson@uic.edu)

Science and the results of scientific research affect and inform almost every aspect of our lives (Durant, Evans, and Thomas, 1989). Researchers have studied a variety of dimensions related to public perceptions of science, including knowledge, beliefs, and attitudes (Pardo and Calvo, 2004). Research on knowledge of science has focused on public understanding of scientific findings, facts, or evidence, but it has also examined public understanding of the methods by which science is conducted (e.g., random assignment to experimental groups or the role of a placebo condition; e.g., National Science Board, National Science Foundation, 2020).

Although much of this research has focused on how the public thinks about physical science research, there is a growing acknowledgement that the public's understanding of social science research findings and methods are also important (e.g., Schäfer, 2016). Our research focuses on examining public understanding of the methods and findings of the most ubiquitous social research methodology: public opinion surveys and polls. Specifically, we are interested in the factors that influence public evaluations of surveys (specifically their perceived accuracy and usefulness to policymakers) and the conditions under which people use a survey's methodological quality and its results to evaluate it.

We report findings from three vignette experiments conducted with general population samples in the US state of Ohio to investigate and compare the rational actor, science literacy, motivated reasoning, and dual process perspectives for understanding public beliefs regarding the credibility of survey results. We tested these hypotheses in data aggregated across three vignette survey experiments about three different current affairs issues, conducted as part of three different probability-based sample telephone surveys. To our knowledge, this is the first research directly examining all four of these possible models and the first direct application of a dual process model approach to understanding how the public perceives surveys and their methodologies.¹ We note that our paper presents information about the results and methodology of a survey as part of the vignettes in the telephone surveys. By that, information is presented aurally to respondents. Although media coverage of survey results

and methodology is frequently presented aurally, most previous research on the topic has examined processing of aurally and visually presented information.

1 Public Understanding of Public Opinion Surveys

Extensive research has investigated the degree to which public opinion polls influence attitudes and behaviors, including voter turnout (Vannette and Westwood, 2013), voting behavior (Lavrakas et al., 1991; Sinclair and Plott, 2012), public opinion (Rothschild and Malhotra, 2014; Toff, 2018), opinion expression (Noelle-Neumann, 1993), public policy (Jacobs and Shapiro, 2000; Page, 1994; Shapiro, 2011) and other topics (Moy and Rinke, 2012). As survey response rates within the general public have continued to decrease, researchers have also begun to investigate variables and conditions associated with public evaluations of opinion surveys, which some evidence suggests are also worsening (Kim et al., 2011). These concerns have taken on added importance as negative opinions about surveys have in fact been found to be associated with both unit nonresponse (Loosveldt and Storms, 2008; Stocké and Langfeldt, 2004) and item nonresponse (Rogelberg et al., 2001; Stocké, 2006).²

2 Survey Quality and Evaluations: Rational Actor Approach

It seems reasonable to expect that objective features of individual survey studies may influence respondent evaluations of their quality and trustworthiness (Salwen, 1987). For example, research suggests that survey sponsorship can be associated with public perceptions of opinion polls and their credibility. In a nationwide U.S. survey, for example, Presser et al. (1998) found that mentioning a survey sponsor who holds a directional position on the policy issue of interest reduced the perceived poll credibility in four of the six vignettes examined. Consistent with the rational actor

¹ Previous literature (e.g., Kuru et al., 2020; Stadtmüller et al., 2022; Stefkovics and Kmetty, 2024) has explicitly examined motivated reasoning and rational actor models, finding greater support for the former than the latter. In addition, a number of studies have tested the science literacy model (e.g., Kuru et al., 2020; Stadtmüller et al., 2022; Stefkovics and Kmetty, 2024) by demonstrating that people are more likely to use survey data quality to evaluate the survey when they have greater knowledge or cognitive ability, although these studies did not typically frame these results directly in terms of this theoretical perspective. No studies that we are aware of have examined the hypothesis suggested by dual process models – that people will be more likely to evaluate carefully and use information such as survey methodology quality when they are both motivated and able to do so.

² Several respondent characteristics are also known to be associated with evaluations of public opinion polls. Older persons, for example, have generally been shown to be less trusting of polls (De Vreese and Semetko, 2002; Dran and Hildreth, 1995; Lavrakas et al. 1991; Price and Stroud, 2005; Stefkovics and Kmetty, 2024), although some recent studies have reported the opposite effect, with Stadtmüller et al., 2022 reporting greater survey trust among older respondents and Johnson et al. (2024) finding a borderline positive association between these measure. More educated and more politically involved individuals are also known to have lower trust in opinion polls (Lavrakas et al., 1991; Stadtmüller et al., 2022; Stefkovics and Kmetty, 2024; Traugott, 1991; Tsfati, 2001), perhaps because they may have greater exposure to polls and are more aware of the general methodological limitations of survey research.

perspective (Downs, 1957) this suggests that respondents may be less willing to accept findings from studies conducted by sources perceived by some to be non-credible, such as foreign or partisan entities.

Several studies have also examined associations between survey source and public perceptions of opinion surveys. Presumably, rational actors would be expected to have more favorable perceptions of survey findings produced by organizations viewed as being more trustworthy. Kim et al. (2000) reported an experiment in which they found surveys conducted by “traditional media” were believed to be more credible than those conducted by online survey firms in the United States. At the time the Kim et al. (2000) study was conducted in 1999, online surveys were still in their infancy and few were conducted using probability sampling methods, perhaps accounting for public beliefs that they were less credible. Contrary to rational actor expectations, though, other available research has not shown survey source to be associated with perceptions of poll credibility (Kuru et al., 2017; Presser et al., 1998) or trust (Salwen, 1987; Stadtmüller et al., 2022).

Methodological rigor has also been reasoned to be a potential correlate of public perceptions of survey results. Experiments reported by Kuru et al., (2020) have confirmed that the public is able to recognize better quality opinion polls when confronted with polls that vary in quality. As part of vignette experiments conducted in Germany (Stadtmüller et al., 2022), Hungary and the U.S. (Stadtmüller et al., 2022), respondents were found to express greater trust when surveys were described as having a larger sample size and when reported as being “representative” or probability-based. Perhaps as a consequence of their educational experiences, Salwen (1987) reported that undergraduate students considered polls more trustworthy when probability sampling methods were indicated. Contrary evidence, however, has been reported by Johnson et al. (2024). Using a national probability sample of U.S. adults, they found that presenting multiple elements of a survey’s methodology as part of a vignette significantly decreased trust in the survey’s findings, which the authors speculated may have been a consequence of providing respondents with cues regarding study limitations, provoking less trust in poll results.

These results in part suggest that people may use factors associated with scientific (i.e., methodological) rigor and objectivity to evaluate the quality of surveys and the trustworthiness of their results in a manner consistent with the rational actor model. Such factors could include sampling (probability versus nonprobability), the types of organizations involved in the research (including the nonpartisan vs. partisan nature of the survey sponsor and the organization conducting the survey), survey participation rate, and sample size. This leads to the first hypothesis that we tested

in the current study about the predictors of people’s evaluations of surveys:

H1: Methodological rigor will be positively associated with evaluations of a survey, such that people evaluate more rigorous surveys more positively and less rigorous surveys more negatively. (*Rational Actor Hypothesis*)

3 Survey Quality and Education: Science Literacy Approach

An elaboration of the Rational Actor perspective might consider the possibility that rational choice as it pertains to evaluations of public opinion polling can only operate when respondents have sufficient scientific literacy or experience to make rational decisions (cf. Li and Guo, 2021; Miller, 1983). Recent investigations evaluating various measures of scientific literacy have presented some evidence consistent with this idea. Weisberg and colleagues (2021), for example, reported that a general understanding of science facts and how science is conducted is associated with acceptance of scientific theories such as climate change, vaccine safety and evolution. Other investigators have also found positive correlations between measures of science knowledge and acceptance of scientific theories (Miller et al., 2006; McPhetres et al., 2019; Weisberg et al., 2018). Given this research, it seems reasonable to expect that those individuals with a greater understanding of science would be better able to recognize linkages between how surveys are conducted and the credibility of their findings.

More directly, several studies have found education (a proxy for scientific literacy) to moderate the relationship between provision of methodologic details and perceived trustworthiness of survey findings. Stadtmüller et al. (2022), and a replication study by Stefkovics and Kmetty (2024), each found that persons with higher education were more likely to identify polls for which methodologic details were provided as being more trustworthy³. Kuru et al., (2020) also found that more educated respondents were more likely to assign more credibility to higher quality polls when asked to compare them with polls of lower quality. We refer to this as the Science Literacy hypothesis:

H2: Methodological rigor will predict survey evaluations as described in H1, but only for respondents with the ability to

³ A footnote in Johnson et al. (2024), however, reported an opposite finding, with more educated respondents reporting less trust in a survey when a block of methodologic information was provided, compared to when no methodologic information was presented. They speculated that informing respondents of a survey’s methods may provide cues about study limitations, resulting in less expressions of trust in findings.

understand the information (we used education as a proxy for science literacy). (*Science Literacy Hypothesis*)

4 Survey Results and Evaluations: Motivated Reasoning Approach

A different theoretical perspective suggests that evaluations of surveys are primarily driven by the extent to which the survey results confirm or disconfirm an individual's own pre-existing opinions. Contrary to Rational Actor models of human behavior, this research provides accumulating evidence that people do not consistently process information in an objective and unbiased manner (Epley and Gilovich, 2016; Lord et al., 1979). Rather, they evaluate evidence via motivated reasoning processes⁴ and are more likely to accept evidence that is consistent with their pre-existing beliefs or opinions than evidence that is not (Donovan et al., 2020; Redlawsk et al., 2010).

For example, in a survey experiment reported by Madison and Hillygus (2020), respondents were more likely to believe that opinion polls were credible when findings were consistent with their pre-existing opinions. Similarly, Ts-fati (2001) found that left-leaning Israeli respondents were more likely, and right-leaning respondents were less likely, to trust survey findings predicting a victory for leftist Labor Party candidate Shimon Peres in the 1996 Prime Minister election; that is, a finding that Ts-fati interpreted as evidence that "people are more likely to trust polls when the polls report what they want to hear. (p. 439)" A study by Presser et al. (1998) reported that, in four out of six policy issues examined, respondents assigned more credibility to those poll findings that were consistent with their prior beliefs on the respective issue. Experiments by Kuru et al. (2017) showed that respondents who held issue positions that contradicted with the results of polls perceived those polls to be less credible. They also found this decreased credibility to be strongest when individuals had high levels of political knowledge. In a follow-up study, Kuru et al. (2020) observed similar patterns for candidate horse race questions, with respondents indicating that polls finding that their favored candidate was leading to be more credible. Also, during the month before the 1988 US presidential election, Democrats were less likely than Republicans to believe poll results showing their candidate (Michael Dukakis) to be behind George Bush in the race (Lavrakas et al. 1991). This literature is consistent in suggesting the importance of motivated reasoning processes when individuals evaluate the

quality and/or legitimacy of findings from public opinion surveys.

Other research also supports motivated reasoning in that perceptions of the findings of opinion polls appear to influence relevant opinions. For example, research in Taiwan demonstrated that survey respondents were more likely to perceive media bias when confronted with poll findings that did not support their partisan beliefs and candidates (Chia and Chang, 2017). In Denmark, respondents who had voted for the losing side in a 2000 referendum on the introduction of the Euro were more likely to support policies that placed greater restrictions on the publication of public opinion polls (de Vreese and Semetko, 2002).

Therefore, in contrast to the rational actor or scientific literacy perspectives, the Motivated Reasoning hypothesis suggest that survey evaluations are dependent on consistency with pre-existing respondent beliefs:

H3: The consistency of Attitude-poll results with one's own attitudes will predict one's evaluations of surveys, such that people perceive surveys with findings that are consistent with their prior attitudes more positively than surveys with findings that are inconsistent with their prior attitudes. (*Motivated Reasoning Hypothesis*)

In summary, past studies have directly or indirectly tested hypotheses derived from the rational actor, science literacy, and motivated reasoning perspectives (e.g., Kuru et al., 2017; 2020; Stadtmüller et al., 2022; Stefkovics and Kmetty, 2024). There is clear evidence from these studies for the motivated reasoning perspective and science literacy perspectives and mixed evidence for the rational actor perspective. However, the previous literature has not considered dual process models which posit that people are most likely to carefully process (and use) information when they are *both* able *and* motivated to do so. This perspective suggests that ability and motivation factors interact positively rather than competing with one another. Below, we briefly review dual process models and the hypothesis derived from this perspective.

5 Who Might Consider Survey Quality?: A Dual Process Approach

Dual process models suggest that information can be processed in two different ways (or in ways that fall along a continuum; Chaiken & Trope, 1999; Claypool, O'Mally, & DeCoster, 2012) by different individuals. In some instances, people process information quickly and automatically and this processing tends to rely on heuristics and other cues (cf. Kahneman, 2013). In other cases, they process information more deeply and intentionally, paying

⁴ Motivated reasoning is a form of confirmation bias, a cognitive process by which individuals tend to seek out and accept information consistent with their pre-existing beliefs while rejecting information that may contradict those beliefs (Nickerson, 1998; Plous, 1993).

more attention to the content of the information and evaluating it more stringently. These dual process models have been widely used for many decades to understand a variety of cognitive processes, including attitude formation and change (Chaiken, 1980; Petty & Cacioppo, 1981).

Dual process models suggest that two types of factors determine whether people will process information carefully or whether they will process it more superficially—(1) factors related to whether people are motivated to do so and (2) those related to whether they are able to do so. Specifically, people will carefully process information more thoughtfully when they are *both* motivated *and* able to do so (e.g., Petty & Cacioppo, 1981). This leads to our final hypothesis:

H4: People will act like rational actors (as described in H1) only when they are both motivated and able to process the information about methodological rigor. (*Dual Process Hypothesis*)

6 Methods

Our study used three survey datasets that were gathered by the Center of Survey Research (CSR) at The Ohio State University (OSU) two decades ago.⁵ The data came from three vignette experiments (one experiment in each survey) that were administered by telephone interviewers. In each experiment, and using random assignment to conditions, different respondents were read different vignettes about a current event issue describing a “hypothetical” poll and its findings. This was done in a way that randomly varied different aspects of the poll, including its findings and methodology (e.g., sample size, participation rate, data collection mode, organization conducting the survey, and sponsor). After a respondent heard the poll described, s/he was asked two questions about the poll’s results. Preliminary findings related to only one of those questions were presented by Lavrakas et al. (2000).

6.1 The Buckeye State Poll

The data for the three probability-based random digit dialing telephone surveys were gathered as part of the Buckeye State Poll (BSP), which was conducted monthly in Ohio by the Center for Survey Research at The Ohio State Univer-

sity. Recruitment of sampled respondents and data collection were carried out by part-time professional telephone interviewers. The response rates (AAPOR RR 3) for each survey ranged between 40–50% (AAPOR, 2023). The cooperation rates for AAPOR COOP3 were in the 75–80% range (AAPOR, 2023). The first experiment was conducted in a Franklin County, Ohio survey in September 1997 ($n = 719$). The second experiment was part of a statewide Ohio survey conducted in March 2000 ($n = 582$). And the third experiment was part of a statewide Ohio survey conducted in April 2000 ($n = 797$). (See Supplementary Material Section A for more details about these BSPs.)

6.2 Survey Questionnaires and Variables

Survey questionnaires. Each of the three survey questionnaires began with a series of economic-indicator items related to consumer confidence. Then came a series of items about a current event topic, which differed for each survey. The current event topics used for the three datasets were (a) funding of public education in Ohio, (b) gun control in Ohio, and (c) the Ohio state lottery. Prior to the vignettes, respondents were asked an attitudinal item about their own beliefs/views toward the current event topic that was focused upon in that month’s BSP. The experimental vignette⁶ designs were part of this middle section within the current event topic sequence of questions and focused on the current event topic. Each questionnaire finished with a series of demographic questions. The survey interviews took approximately 15–20 minutes to complete.

Experimental Vignette Designs. The exact wording of the vignette to which each respondent was exposed varied randomly according to a multiple factorial design. An example vignette from the survey on handgun control is shown below with the information that was varied randomly shown in brackets [] and in italics:

Suppose you heard some details about a public opinion poll on what Ohioans think about a ban on the sale of all handguns, except those that are issued to law enforcement officers and other authorized persons. The poll found that [65/35] percent of Ohioans favored this ban on the sale of all handguns.

There were [1000/2000] adult Ohioans surveyed in this poll and they were sampled by [*interviewers in*

⁵ At that time, some of the data were used for preliminary analyses and those findings were presented in an American Association for Public Opinion Research (AAPOR) paper (Lavrakas, Diaz-Castillo, and Monson, 2000). The first author of this AAPOR paper was the faculty director of the OSU-CSR at the time and the second and third authors were graduate research assistants with the OSU-CSR.

⁶ Within each survey questionnaire, the vignette design was one in which the wording of a question was randomly altered and randomly assigned to a respondent, so that different respondents heard different vignette versions comprised of different conditions of the researcher-controlled experimental treatment(s) used for the questions (cf. Vargas, 2008; Eifler and Petzold, 2019).

Ohio shopping malls asking every 10th person who walked past them to fill out a questionnaire/randomly selecting Ohioans with e-mail addresses and asking them to fill out a questionnaire on an Internet site].

About [70/30] percent of the Ohioans who were sampled participated in the survey.

The survey was paid for by [a major newspaper in Ohio/the National Rifle Association] and conducted by [a market research firm in Ohio/the Gallup Organization].

Using this approach, the original researchers (i.e., Lavrakas, et al. 2000) conducted the three studies in an iterative fashion. Each included a vignette experiment that systematically manipulated information in the description of the hypothetical survey. They also randomly varied whether the respondents heard the poll result before the methodological information or in the opposite order, and whether the first dependent variable (see below) was measured before or after the second dependent variable.

Each vignette experiment was basically constructed using a 2 (poll results) \times 2 (sample size) \times 2 (participation rate) \times (sampling mode) \times 2 (poll sponsor) \times 2 (polling organization) factorial experimental design with some slight variations across the three studies.⁷ Table 1 shows the vignette information that was randomized in the three studies.⁸

Dependent variables. Immediately after the vignette was heard, respondents were asked two questions about the survey described in the vignette to assess the perceived ac-

curacy of the survey and whether or not they believed its results should be considered by elected officials:

How accurate do you believe this poll is? Would you say it is ...

<1> extremely accurate,

<2> quite accurate,

<3> fairly accurate,

<4> not too accurate, or

<5> not at all accurate?

<9> UNCERTAIN

When our elected officials are considering legislation about [CURRENT EVENT ISSUE], do you think they should consider the results of this poll in making their decisions or not?

<1> YES, SHOULD CONSIDER

<2> NO, SHOULD NOT CONSIDER

<9> UNCERTAIN

Responses to the first question were recoded into an Accuracy variable that ranged from 0 (not at all accurate) to 1 (extremely accurate). Responses to the second question were recoded into a Consideration variable coded 0 for “should not consider” and 1 for “should consider.”⁹ The

by Auspurg, Hinz, and colleagues examined vignette data conducted in telephone surveys). However, Andernach and Schunk (2014) directly investigated the feasibility of factorial surveys using CATI and concluded that the evidence that factorial vignette studies could be implemented in CATI surveys was encouraging. Our review also suggests that factorial surveys using CATI methods are in fact commonly reported in the peer-reviewed literature. These include studies in several prominent journals, including the *American Journal of Political Science* (Berinsky & Mendelberg, 2005), the *American Sociological Review* (Pager & Quillian 2005), *BMJ Open* (Whiddett et al., 2016) the *Journal of Ethnic and Migration Studies* (Canan & Foroutan, 2016), *Political Research Quarterly* (Djupe & Calfano, 2012), and *Public Opinion Quarterly* (Hopkins & King, 2010).

We would argue that the telephone administration of our vignette experiments was likely to produce valid data because the vignette experiments that we used manipulated 5-6 factors with 2-3 levels each, representing the low end of recommended complexity. Null findings also seem the most likely result if respondents are unable to process vignette information carefully, so the concern about telephone surveys would not undermine the validity of significant findings. Nonetheless, we acknowledge this as both a limitation and a strength of our research. It is a weakness because our findings are limited to telephone surveys (or at least aurally presented information). However, it is a strength because information about survey methods and results may be presented aurally by the media, so assessing processing in this mode reflects a realistic information presentation situation that is unique to our studies.

⁹ The second dependent variable in each of the three surveys was asked as an open-ended question, with the respondents' verbatim responses being coded by interviewers into one of the three answers that were used for this item. That is, interviewers were not asked to code the answers they heard into a more varied response scale of options, due to concerns about the interviewers' abilities to do that reliably. One limitation of this question is that it included only two substantive responses, so it has limited variance. To the extent this is likely to impact results, it is likely to make it more difficult to find the predicted effects.

⁷ In all, there were 576 different randomly-generated vignette versions used across the three surveys. This takes into account the nature of the particular factorial designs in each of the three surveys, including the randomization of the orders in which the two dependent variables were presented, and the order in which the hypothetical poll results vs. the order of the hypothetical poll methods within each vignette were presented. Based on our calculation, the first survey included $2^7 = 128$ possible vignette versions, the second survey included $2^8 = 256$ possible vignette versions, and the third survey included $2^6 \times 3 = 192$ possible vignette versions.

⁸ Some readers might be concerned that these vignettes were read to respondents in telephone surveys, so respondents needed to process the information as it was read to them and cannot control the pace of the interview. Some authors (e.g., Auspurg and Hinz, 2014; Auspurg, Hinz, and Walzenbach, 2019) have recommended that researchers use self-administered modes for vignette studies, particularly those that manipulate numerous dimensions. These authors also identify 6-8 variable dimensions with 2-3 levels as a “mid-level of complexity” and recommend that researchers limit the number of vignettes per respondent to no more than 10 vignettes to prevent fatigue (Auspurg, Hinz, and Walzenbach, 2019). Regarding the latter, our research exposed respondents to only one vignette.

However, no direct evidence regarding the unsuitability of the telephone mode is presented by these authors (indeed, none of the research

Table 1*Information Randomly Assigned in the Vignette*

Survey	Issue	Poll results Favor	Sample size	Participation rate	Poll Sponsor	Data Collection Mode and Sampling Strategy	Polling organization
August 1997	School vouchers	40 or 60%	200 or 1000	80 or 20%	Daily newspaper or Religious denomination	Mall intercept or RDD	Not mentioned
March 2000	Handgun control	65 or 35%	1000 or 2000	70 or 30%	Major Ohio newspaper or NRA	Enhanced mall intercept or Web survey	Gallup or Market research firm in Ohio
April/May 2000	Eliminate Ohio Lottery	60 or 40%	100, 2000 or 10,000	45 or 55%	Major Ohio newspaper or Ohioans Against the Lottery	Held constant	Group of volunteers or Gallup

“uncertain” responses for both questions were coded as missing.

Manipulated indicators of survey quality. A total of five of the vignette factors were associated with methodological rigor (sample size, participation rate, survey sponsor, organization that conducted the survey, and the data collection mode/sampling approach):

- For the sample size variable, each vignette presented sample sizes that ranged from 100 to 10,000 across the three surveys. In order to rescale this variable and also account for the greater importance of increased sample size at lower levels, we took the natural log of this value to account for diminishing returns of larger sample sizes (e.g., for a sample size of 100, $\ln(100) = 4.61$; for a sample size of 10,000, $\ln(10,000) = 9.21$) and then rescaled these values so that the variable ranged from 0 to 1 where higher values indicated larger sample sizes.
- The participation rate reported in the vignette ranged from 20 to 80%. This was recoded so that participation rates could range from a possible value of 0 (for 20%) to 1 (for 80%).
- The poll sponsor variable indicated the type of organization that sponsored the hypothetical poll. It was recoded as a dichotomous variable, with 0 representing an advocacy (i.e., potentially biased) organization¹⁰ and 1 representing a well-known daily newspaper.
- The polling organization variable indicated the type of organization that carried out the recruitment and data collection. That was recoded into 0, 0.5, and 1, with 0 representing groups of volunteers, 0.5 representing market research firms, and 1 representing the Gallup organization.
- The last survey quality indicator was about the sampling approach and respondent recruitment mode. It was recoded to 0.0, 0.33, 0.67, and 1.0, with 0 representing mall intercept surveys, 0.33 representing enhanced mall intercept surveys, 0.67 representing Web surveys, and 1 representing RDD surveys¹¹. The larger value implies survey modes that are more likely to yield a representative unweighted final sample.

¹⁰ At the time in the U.S. that the School Voucher survey was carried out, it was well known within the general population that private religious-based schools were strong advocates of implementing school voucher laws. Thus, in that experimental vignette, the poll sponsor being cited as a “religious denomination” (as opposed to a “Major Newspaper”) would be perceived as a biased sponsoring organization by many Americans.

¹¹ At the time of the Handgun survey, it was reasoned that a *random-selected* survey of emails addresses would more likely be perceived by the general public as a more representative survey sampling method than sending interviewers to shopping malls (without specifying how the malls were selected). However, at the time of this survey, it was also reasoned that a Web survey would be perceived as less likely to yield a representative sample than an RDD survey.

Table 2*Descriptive Statistics for Categorical Variables*

Variable	Label	Value	Frequency	Percent
Categorical Variables				
Reported poll sponsor (manipulated)	Newspaper	1	1076	51
	Advocacy group	0	1022	49
Reported polling organization (manipulated)	Gallup	1	682	49
	Market research firm	0.5	306	22
	Group of volunteers	0	391	28
Reported data collection mode/sampling strategy (manipulated)	RDD survey	1	1142	54
	Web survey	0.67	287	14
	Enhanced mall intercept (randomly selected every nth person to ask to participate)	0.33	295	14
	Mall intercept	0	374	18
Order of dependent measures (manipulated)	Consideration/accuracy	1	1021	49
	Accuracy/consideration	0	1077	51
Order of information within vignette (manipulated)	Methods/results	1	1087	52
	Results/methods	0	1011	48
Opinion Consistency (calculated)	Consistent	1	948	49
	Inconsistent	0	995	51
Consideration (measured)	Yes	1	1181	59
	No	0	820	41

As previously mentioned, not all the five variables were manipulated in each of the surveys. For example, the polling organization was not included in the School Voucher survey vignette. In addition, the data collection mode was held constant (the telephone mode) in the Lottery survey; thus, the values for this variable in the Lottery survey were all recoded to 1.

In addition to these five variables, we created three indices from these variables. The first was an objective survey quality index (OSQI) which were averages of the sample size, participation rate, and mode/sampling strategy quality variables. The second was a subjective survey quality index (SSQI) which was an average of the survey sponsor and data collection organization variables. Finally, we created an total survey quality index (TSQI), which was calculated as an average of all five of the manipulated quality vari-

ables.¹² (see Tables 2 and 3). These indices all ranged from 0 to 1 with higher values indicating greater quality.

These indices were used to test whether respondents used survey quality in evaluating surveys (H1), whether only highly educated respondents did so (H2), and whether only respondents who were motivated and able to carefully process this information did so (H4).

Opinion consistency. For each survey, respondents were asked their opinion about the issue addressed in the vignette *before* they were exposed to the vignette experiment:

- In the August 1997 survey, respondents were asked: *Do you favor or oppose the use of tax-supported vouchers to help finance private education?*
- In the March 2000 survey, respondents were asked: *Would you favor or oppose a ban on the sale of all handguns, except those that are issued to law enforcement officers and other authorized persons?*
- In the April 2000 survey, respondents were asked: *Do you believe the Ohio Lottery should continue as it is, be eliminated altogether, or be expanded to include multi-state games such as Powerball?* Respondents who said

¹² Creating indices of the manipulated variables makes assumptions that are not necessarily warranted, including that these factors are equally important indicators of methodological quality. We recoded each of these variables to range from 0 to 1 for the lowest and highest quality, but in some cases these coding choices were arbitrary. The goal of creating these indices was to assess whether respondents' accuracy and use ratings were responsive to an overall aggregate measure of survey methods quality, albeit a rough and imperfect one. Because analyses using OSQI and SSQI showed similar patterns for the two

separate quality indices, we present results using the TSQI in the main text for parsimony and analyses examining OSQI and SSQI separately are shown in Section C of the Supplementary Materials.

Table 3*Descriptive Statistics for Continuous Variables*

Variable	Mean	Standard deviation	Minimum	Maximum	Sample Size
Continuous Variables					
Reported survey participation rate (manipulated)	0.49	0.34	0	1	2098
Reported survey sample size (manipulated)	0.48	0.30	0	1	2098
Objective Survey Quality Index (OSQI)	0.55	0.21	0.05	0.86	2098
Subjective Survey Quality Index (SSQI)	0.54	0.39	0	1	2098
Total Survey Quality Index (TSQI)	0.55	0.19	0.03	0.92	2098
Accuracy	0.44	0.22	0	1	2053

they believed the Ohio Lottery should be eliminated altogether supported the proposed policy of eliminating the lottery described in the vignette, while those who said the Ohio Lottery should either continue as is or be expanded opposed the proposed policy.

These policy opinion questions were used along with the survey result factor that indicated the proportion of the public that endorsed the particular policy addressed in the vignette to construct an Opinion Consistency variable. This measured the consistency of the poll results in the vignette to which a given respondent was randomly exposed with the respondent's own previously reported opinion towards the current event issue that the poll measured. This variable was coded 1 if the survey result was consistent with the respondent's prior opinion and 0 if it was not.

This variable was used to test whether respondents evaluated surveys whose results they agreed with more favorably than those whose results they disagreed with (H3) and whether people might be more motivated to scrutinize (and therefore use) methodological information to evaluate surveys when the survey result is inconsistent with their own prior opinions (H4).

Vignette order. An additional independent variable was used in our analyses, and this came from the original surveys. This variable (Vignette Order) indicated the order of the information about the poll result and methodology of the hypothetical poll which was randomized in the vignette that a respondent heard. This occurred in two orders: one presented the poll results *before* the methodology of the poll was explained, and the other order had the poll results presented *after* the methodological details were given. Vignette Order was coded "0" if results were first presented and methods were second, and to "1" if methods were first presented and results were second. This variable was used to test whether respondents would be more motivated

to scrutinize (and therefore use) survey methodological information to evaluate surveys when they knew the survey result was inconsistent with their own prior opinions prior to hearing about the methodology (H4).

DV order. Another independent variable was used and represented the order in which the two dependent variables (i.e., Accuracy and Consideration) were measured was also manipulated experimentally as part of the vignette experiment. The variable, DV Order, was coded 0 if Accuracy was measured before Consideration and 1 if Consideration was measured before Accuracy. This variable was included as a control variable in all analyses but was not directly relevant to our hypotheses.

Education. A final independent variable from the original surveys that was used in the analyses was the respondents' self-reported educational attainment level. Respondents were asked: "What is the highest grade or year of school you have completed?"¹³ Answers were coded into three categories: (1) high school degree or less (baseline),¹⁴ (2) some college, and (3) four-year college degree or more. That variable then was used to create two binary dummy variables for the analyses, each coded 0/1. One dummy va-

¹³ As is often used in U.S. general population surveys, the educational attainment item was asked as an open-ended question, which the interviewer coded into a provided list of educational levels that the respondent did not see or hear. This list of attainment levels, which the interviewer used to code the respondent's verbatim response into ranged from 1st grade to a Doctorate or an Advanced professional degree. Thus, this scale was an ordinal (not an interval) variable. As such, the coded answers were recoded into a smaller set of categories for analytic purposes, as is routinely done in most general population surveys in the U.S.

¹⁴ Respondents with less than a high school education were combined with those with a high school education due to a relatively small number of respondents reporting less than a high school education.

Regression Models Predicting Accuracy and Consideration Beliefs: Unstandardized Coefficients (Standard errors)

Predictor	Accuracy Beliefs (OLS)						Consideration Beliefs (Logistic)																	
	1. Main effects			2. Main effects + two-way interactions			3. Main effects + two- and three-way interactions			4. Main effects + two-, three-, and four-way interactions			5. Main effects			6. Main effects + two-way interactions			7. Main effects + two- and three-way interactions			8. Main effects + two-, three-, and four-way interactions		
	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p
Total Survey Quality Index (TSQI)	0.01	0.04	0.58	0.04	0.05	0.48	-0.05	0.07	0.53	-0.08	0.08	0.30	0.20	0.24	0.42	0.04	0.51	0.94	0.35	0.68	0.61	-0.24	0.74	0.73
Opinion Consistency (Consistency)	0.04	0.02	0.01	0.03	0.03	0.41	0.06	0.06	0.31	0.02	0.07	0.79	0.40	0.09	<0.01	0.50	0.33	0.12	0.93	0.56	0.09	0.21	0.63	0.74
Order of information within vignette (vignette order: 0 = results/methods; 1 = methods/results)	-0.001	0.02	0.94	0.03	0.03	0.44	-0.09	0.06	0.11	-0.13	0.07	0.04	-0.04	0.09	0.68	-0.13	0.32	0.68	0.11	0.54	0.84	-0.60	0.63	0.34
Order of dependent variables DV order: 0 = accuracy/consideration; 1 = consideration/accuracy)	0.01	0.02	0.49	0.01	0.01	0.56	0.01	0.01	0.63	0.004	0.01	0.66	0.35	0.10	<0.01	0.35	0.10	<0.01	0.36	0.10	<0.01	0.35	0.10	<0.01
Education (high school degree or less baseline group)																								
Some college	-0.02	0.02	0.13	-0.03	0.04	0.42	-0.16	0.06	0.01	-0.16	0.07	0.03	-0.11	0.11	0.33	-0.38	0.38	0.31	-0.12	0.58	0.83	-0.24	0.64	0.70
Four-year college degree or more college degree)	-0.05	0.02	<0.01	-0.18	0.04	<0.01	-0.24	0.07	<0.01	-0.32	0.07	<0.01	-0.16	0.11	0.17	-10.08	0.39	0.01	-10.10	0.62	0.08	-2.36	0.75	0.01
TSQI*Consistency	-	-	-	-0.04	0.05	0.45	-0.10	0.10	0.31	-0.02	0.11	0.83	-	-	-	-0.60	0.50	0.22	-1.31	0.91	0.15	-0.02	1.07	0.98
TSQI*Vignette order	-	-	-	-0.09	0.05	0.07	0.12	0.10	0.23	0.19	0.11	0.09	-	-	-	0.01	0.49	0.99	-0.38	0.91	0.67	0.91	1.08	0.40
Consistency*Vignette order	-	-	-	0.02	0.02	0.35	-0.01	0.07	0.93	0.08	0.10	0.41	-	-	-	0.09	0.19	0.66	-0.47	0.63	0.45	1.02	0.91	0.26
TSQI*Some college	-	-	-	-0.001	0.06	0.99	0.22	0.11	0.04	0.22	0.12	0.08	-	-	-	0.08	0.59	0.89	-0.31	0.99	0.75	-0.09	1.12	0.93
TSQI*College degree	-	-	-	0.14	0.06	0.02	0.26	0.11	0.02	0.40	0.13	0.01	-	-	-	1.43	0.60	0.02	1.45	1.05	0.17	3.78	1.30	0.01
Consistency*Some college	-	-	-	0.03	0.02	0.30	0.06	0.08	0.48	0.05	0.10	0.62	-	-	-	0.48	0.23	0.04	0.14	0.74	0.85	0.39	0.95	0.68
Consistency*College degree	-	-	-	0.06	0.02	0.01	-0.01	0.08	0.88	0.14	0.10	0.18	-	-	-	0.14	0.23	0.53	-0.16	0.74	0.83	2.18	1.02	0.03
Vignette Order*Some college	-	-	-	0.01	0.02	0.81	0.25	0.08	0.01	0.25	0.10	0.01	-	-	-	0.01	0.23	0.95	-0.34	0.72	0.63	-0.05	0.92	0.96
Vignette Order*College degree	-	-	-	0.04	0.02	0.07	0.25	0.08	0.01	0.41	0.11	<0.01	-	-	-	0.16	0.23	0.50	0.45	0.73	0.53	2.87	1.02	0.01
TSQI*Consistency*Vignette order	-	-	-	-	-	-	0.06	0.11	0.60	-0.10	0.16	0.54	-	-	-	-	-	-	0.88	1.00	0.38	-1.83	1.56	0.24
TSQI*Consistency*Some college	-	-	-	-	-	-	-0.04	0.13	0.76	-0.03	0.17	0.88	-	-	-	-	-	-	0.39	1.20	0.74	-0.02	1.64	0.99
TSQI*Consistency*College degree	-	-	-	-	-	-	0.13	0.13	0.31	-0.12	0.17	0.47	-	-	-	-	-	-	0.54	1.21	0.66	-3.74	1.75	0.03
TSQI*Vignette order*Some college	-	-	-	-	-	-	-0.43	0.12	0.01	-0.42	0.17	0.01	-	-	-	-	-	-	0.47	1.19	0.69	-0.06	1.59	0.97
TSQI*Vignette order*College degree	-	-	-	-	-	-	-0.38	0.13	0.01	-0.66	0.18	<0.01	-	-	-	-	-	-	-0.56	1.21	0.64	-4.97	1.78	0.01
Consistency*Vignette order*Some college	-	-	-	-	-	-	-0.02	0.05	0.68	-0.009	0.15	0.95	-	-	-	-	-	-	0.22	0.46	0.64	-0.25	1.39	0.86

Table 4

(Continued)

Predictor	Accuracy Beliefs (OLS)						Consideration Beliefs (Logistic)																	
	1. Main effects			2. Main effects + two-way interactions			3. Main effects + two- and three-way interactions			4. Main effects + two-, three-, and four-way interactions			5. Main effects			6. Main effects + two-way interactions			7. Main effects + two- and three-way interactions			8. Main effects + two-, three-, and four-way interactions		
	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p
Consistency*Vignette order* College degree	-	-	-	-	-	-	0.001	0.05	0.98	-0.30	0.15	0.04	-	-	-	-	-	-	0.02	0.46	0.96	-4.64	1.42	0.001
TSQI*Consistency*Vignette order*Some college	-	-	-	-	-	-	-	-	-	-0.02	0.25	0.93	-	-	-	-	-	-	-	-	-	0.84	2.40	0.73
TSQI*Consistency*Vignette order*College degree	-	-	-	-	-	-	-	-	-	0.56	0.24	0.03	-	-	-	-	-	-	-	-	-	8.57	2.47	0.001
(Pseudo-)R ²	0.02			0.03			0.03					0.04	0.01			0.02			0.02					0.03
N	1905			1905			1905					1905	1870			1870			1870					1870

riable was for the Some College category and one for the Four-Year College Degree category.

Education was used as a proxy for scientific literacy to test whether respondents with greater science literacy (i.e., higher education respondents) would be more likely to use a survey's methodological rigor to evaluate it than respondents with lower science literacy. Education was also as a proxy for respondents' ability to process methodological information in order to test H4.

6.3 Analytic Approach

We conducted all analyses using Stata. We began by assessing whether there were important differences in the dependent variables across the three surveys to determine whether we needed to control for Survey when combining the data across surveys. We found no associations between Survey and Accuracy ($F(2,2050) = 1.07, p = 0.34, N = 2052$) or Consideration ($\chi^2(2) = 0.95, p = 0.62, N = 2001$). Therefore, we combined data across the three surveys and conducted our analyses using simple OLS and logistic regression without controlling for survey as a clustering variable.

We began by estimating models that regressed each dependent variable on each of the manipulated survey methodology factors separately for each survey and combined across surveys (H1) along with Opinion Consistency (H3), Vignette Order, DV Order, and Education. We then conducted these analyses using SSQI and OSQI instead of the individual manipulated quality variables. Finally, we estimated models using a single TSQI index. These analyses tested H1 and H3 using different approaches to operationalizing survey quality. We coded education into three categories for these analyses to help ensure that we had a sufficient sample size in each education category. Next, we estimated models for each dependent variable with all two-way interactions between the OSQI, Opinion-Consistency, Vignette Order, and Education (controlling for SSQI); models for each dependent variable with all two-way interactions between the SSQI, Opinion-Consistency, Vignette Order, and Education (controlling for OSQI); and models for each dependent variable with all two-way interactions between the TSQI, Opinion-Consistency, Vignette Order, and Education. This allowed us to test H2 by assessing whether the impact of each of the three survey quality indices varied for respondents with different levels of Education (as a proxy for science literacy). We next estimated models for each dependent variable with all two- and three-way interactions (we did so for completeness, although these did not provide direct tests of any of our hypotheses).

Table 5*Regression Models Predicting Accuracy and Consideration Beliefs split by Education: Unstandardized Coefficients (Standard errors)*

Predictor	Accuracy Beliefs (OLS)						Consideration Beliefs (Logistic)					
	1. High school degree or less			2. Some college			3. Four-year college degree or more			4. High school degree or less		
	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p
Total Survey Quality Index (TSQI)	-0.08	0.08	0.31	0.13	0.09	0.14	0.32	0.10	0.01	-0.27	0.74	0.72
Opinion Consistency (Consistency)	0.02	0.07	0.79	0.07	0.08	0.37	0.16	0.08	0.05	0.22	0.64	0.73
Order of information within vignette (Vignette Order: 0 = results/methods; 1 = methods/results)	-0.14	0.07	0.05	0.11	0.07	0.12	0.27	0.08	0.01	-0.62	0.63	0.33
DV Order (DV order: 0 = accuracy/consideration; 1 = consideration/accuracy)	0.01	0.02	0.57	0.0002	0.02	0.99	0.002	0.02	0.91	0.48	0.15	0.01
TSQI*Consistency	-0.02	0.11	0.83	-0.05	0.13	0.70	-0.17	0.14	0.21	-0.02	1.07	0.99
TSQI*Vignette order	0.19	0.12	0.10	-0.23	0.13	0.07	-0.47	0.14	0.01	0.94	1.09	0.39
Consistency*Vignette order	0.08	0.10	0.41	0.07	0.11	0.52	-0.22	0.11	0.05	1.04	0.91	0.26
TSQI*Consistency*Vignette order	-0.10	0.17	0.55	-0.12	0.19	0.52	0.46	0.19	0.02	-1.84	1.57	0.24
(Pseudo-)R ²	0.01			0.03			0.07			0.02		
N	766			569			570			740		
				568			562			562		

Table 6*Effect of Quality split by Consistency and Vignette order for Respondents with a Four-year Degree or More: Unstandardized Coefficients (Standard errors)*

Predictor	1. Consistency = Consistent Vignette order = Methods/Results			2. Consistency = Consistent Vignette order = Results/Methods			3. Consistency = Inconsistent Vignette order = Methods/Results			4. Consistency = Inconsistent Vignette order = Results/Methods		
	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p	Coef.	S.E.	p
Predicting Accuracy												
Total Survey Quality Index (TSQI)	0.14	0.10	0.17	0.16	0.10	0.11	-0.14	0.10	0.14	0.33	0.09	< 0.01
DV Order (DV order: 0 = accuracy/ consideration; 1 = consideration/accuracy)	0.05	0.04	0.21	-0.04	0.04	0.31	0.03	0.04	0.35	-0.04	0.03	0.23
R ²		0.02			0.02			0.02			0.10	
N		154			132			151			133	
Predicting Consideration												
Total Survey Quality Index (TSQI)	2.44	0.96	0.01	-0.23	0.87	0.79	-0.51	0.92	0.58	3.60	1.07	0.01
Order of dependent variables (DV order: 0 = accuracy/consideration; 1 = consideration/ accuracy)	0.40	0.35	0.24	0.19	0.36	0.59	0.64	0.34	0.06	-0.32	0.37	0.39
Pseudo-R ²		0.04			0.002			0.02			0.07	
N		152			133			145			132	

To test H4, we estimated models for each dependent variable with all two-, three-, and four-way interactions between these variables using each survey quality index separately. Specifically, H4 predicts a four-way interaction between survey quality, Opinion Consistency, Vignette Order and Education, such that survey quality is hypothesized to predict Accuracy and Consideration among high education respondents who were told about survey findings that contradicted their prior opinions before the survey methodology was described to them.

Finally, for subgroups/conditions in which quality predicted both dependent variables, we tested whether perceptions of Accuracy mediated the impact of the survey quality on Consideration using the *sem* command in Stata.

7 Results

7.1 Descriptive Statistics

Table 2 shows descriptive statistics for all the variables in our analyses aggregated across the three surveys. All manipulated variables used random assignment, but because some of the variables were not manipulated in all studies, sample sizes for each level of some of the manipulated variables vary.

7.2 Main Effects: Rational Actor vs. Motivated Reasoner

Models 1 and 5 in Table 4 show the results of models where each dependent variable is regressed on the main effects of the TSQI, Opinion Consistency, Vignette Order, DV Order, and Education. There was no support for H1(Rational Actor)—the main effect of the TSQI was not significant in either analysis. Providing support for H3, Opinion Consistency was a significant predictor of Accuracy ratings, such that respondents rated polls that were consistent with their pre-existing attitudes as being more accurate than those that did not (coefficient = 0.04, SE = 0.02, $p = 0.001$). In addition, respondents with a four-year college degree or more education rated the surveys as less accurate than did those with a high school degree or less education (coefficient = -0.05 , SE = 0.02, $p < 0.001$).

Model 5 in Table 4 shows that the main effects findings for Consideration overlapped somewhat. Again, supporting H3, Opinion Consistency was a strong predictor of reporting that the survey should be considered by policymakers (logistic coefficient = 0.40, SE = 0.09, $p < 0.001$). However, Education was not a significant predictor of Consideration.

The order of the dependent variable (DV Order) was a significant predictor of consideration (logistic coefficient = 0.35, SE = 0.10, $p < 0.001$). Respondents were more likely to report that the poll results should be considered by policymakers if they were asked Consideration before Accuracy than if they were asked Accuracy first. This was not an effect we predicted, but it makes sense if the act of answering about Accuracy leads respondents to be more aware that they should be concerned about the accuracy of the survey when making their Consideration judgments.¹⁵

The main effects model provides no evidence for the Rational Actor Model (H1) and strong evidence for the Motivated Reasoning Model (H3). The strongest direct predictor of evaluations of the survey (both Accuracy and Consideration) was whether the survey finding regarding public opinion on a proposed policy was consistent or inconsistent with a person's pre-existing opinion about the policy.

7.3 Interaction Effects: Scientific Literacy and Dual Process

The main effects reported above were qualified by a number of significant interactions. Models with main effects and all two-way interactions between the TSQI, Opinion Consistency, Vignette Order and Education are shown in Models 2 and 6 in Table 4.¹⁶ These analyses provide some support for H2, the Science Literacy hypothesis. The interaction between the TSQI and college degree was positive and significant for both the Accuracy variable (coefficient = 0.14, SE = 0.06, $p = 0.02$) and the Consideration variable (coefficient = 1.43, SE = 0.60, $p = 0.02$), providing evidence that respondents who received a college degree were

¹⁵ Table C1 (see Section C of the Supplementary Materials) shows bivariate associations between each of manipulated survey quality variables, the other independent variables and the two dependent variables separately for each of the three surveys and combined across surveys. Table C2 (see Section C of the Supplementary Material) shows the results of multivariate models predicting Accuracy and use ratings with separate manipulated survey quality factors and the other independent variables separately for each survey and combined across surveys. Finally, Models 1 and 5 in Tables C3a and C3b show multivariate analyses predicting each of the dependent variables with the OSQI, SSQI, and other independent variables. Overall, these results suggest that the individual manipulated survey quality factors are not consistently or strongly associated with Accuracy or Consideration across studies, providing little support for H1 (Rational Actor). Similar to the TSQI variable, OSQI and SSQI also do not consistently predict either Accuracy or Consideration.

¹⁶ Parallel analyses for OSQI and SSQI are shown in Section C, Tables C4a and C4b. These results suggest that similar patterns are shown for the index of objective survey quality factors (OSQI) and the index of subjective survey quality factors (SSQI) providing further support for combining these into TSQI.

more likely to consider survey quality than were those without a college degree when considering survey accuracy and whether a survey's results should be considered by policymakers.

For completeness, Models 3 and 7 in Table 4 include main effects and all two- and three-way interactions between these variables. Models 4 and 8 in Table 4 include all main effects, and two-, three-, and four-way interactions between these variables. Most notably, although some of the two- and three-way interactions are significant, all these effects are qualified by the interactions shown in the last two rows of Table 4. For Accuracy ratings, the interaction between the TSQI, Opinion Consistency, Vignette Order, and the dummy variable for a four-year college degree or more was significant (coefficient = 0.56, SE = 0.24, $p = 0.03$) and this same interaction was highly significant for the Consideration dependent variable (logistic coefficient = 8.57, SE = 2.47, $p = 0.001$). Parallel analyses for OSQI and SSQI are shown in Tables C4a and C4b of the Supplementary Material. These results show similar patterns of interactions for OSQI and SSQI.

In order to illustrate the exact nature of this four-way interaction, we conducted analyses looking at the TSQI by Opinion Consistency by Vignette Order interaction by Education level (see Table 5; parallel analyses for OSQI and SSQI are shown in Tables C5a and C5b). For both dependent variables, the main effect of the TSQI was significant only for respondents with at least a college degree (Accuracy: coefficient = 0.32, SE = 0.10, $p = 0.001$; Consideration: coefficient = 3.52, SE = 1.07, $p = 0.001$), providing evidence for the science literacy perspective (H2). However, these main effects were qualified for respondents with at least a four-year college degree by a three-way interaction between the TSQI, Opinion Consistency, and Vignette Order (Accuracy: coefficient = 0.46, SE = 0.19, $p = 0.02$; Consideration: logistic coefficient = 6.76, SE = 1.91, $p < 0.001$). Only among respondents with at least a four-year college degree, Accuracy and Consideration were a joint function of the TSQI, Opinion Consistency, and Vignette Order.

Table 6 further illustrates the nature of this interaction by showing the effect of the TSQI (controlling for DV Order) for respondent subgroups split by Opinion Consistency and Vignette Order for respondents with a four-year college degree or more (parallel results are shown for OSQI and SSQI in Table C6 in Section C of the Supplementary Material). These results show that methodological quality is the strongest predictor of both Accuracy and Consideration among respondents with at least a four-year college degree, when the survey result was *inconsistent* with the respondent's prior attitude on the current event issue on which the survey focused and when respondents were told the survey result *before* they were told the survey's method-

ological information (see Model 4 in Table 6), providing strong support for H4.

Mediation Analysis. For those with at least a four-year college degree who had prior attitudes that were inconsistent with the survey result they learned about, and who were told the survey result before the survey methodology in the vignette that was read to them, we also tested whether Accuracy perceptions mediated the effect of the TSQI on Consideration. Among these respondents and consistent with the results described thus far, the TSQI significantly predicted Accuracy beliefs (coefficient = 0.25, SE = 0.08, $p = 0.002$). In addition, Accuracy significantly predicted Consideration beliefs (coefficient = 1.39, SE = 0.17, $p < 0.001$).¹⁷ Tests of direct and indirect effects (using the Stata *estat teffects* command) showed that the TSQI had both a significant direct effect on Consideration beliefs (coefficient = 1.39, SE = 0.17, $z = 8.00$, $p < 0.001$) and a significant indirect effect (coefficient = 0.35, SE = 0.12, $z = 2.91$, $p = 0.004$). These results suggest that the impact of the TSQI on Consideration beliefs was partly mediated by perceptions of Accuracy for this subgroup of respondents.

Effect of the Order of Dependent Variables. As described above, the variable indicating the order of the dependent variables was a significant predictor of Consideration beliefs but not of perceived Accuracy (see row 4 in Models 1 and 5 in Table 4), such that respondents were less likely to indicate that they thought the survey result should be considered by policymakers when they were asked about survey Accuracy before being asked about Consideration by policymakers (relative to respondents who were asked about Consideration by policymakers before being asked about Accuracy). We did not have a hypothesis about this variable and its possible effects, but our post hoc interpretation of this is that asking respondents about Accuracy first reminded respondents that a survey might not be accurate, thereby tending to reduce their belief that the survey should be considered in policymaking.

We also tested whether the manipulation of the ordering of the dependent variables moderated the effect of quality or varied by respondent education (by examining interactions between this order variable, education, and the TSQI). None of these interactions were significant, suggesting that this effect did not moderate the effect of the TSQI on the dependent variables and its effect did not differ across education levels (or levels of other variables). This is also illustrated

¹⁷ We also tested these relationships using the Stata *gsem* command, which allowed us to specify that Consideration was dichotomous while Accuracy was continuous, but the Stata *estat teffects* command is not available for the *gsem* procedure. The results for the key relationships showed the same substantive results regardless of whether *sem* or *gsem* was used.

in row 4 of Models 4, 5, and 6 in Table 5, which shows that the effect of the order of the dependent variable (while larger and conventionally significant only among respondents with a high school degree or less), was consistently positive and of comparable size. Row 2 of the bottom panel of Table 6 shows more variability in the effect of dependent variable order on Consideration, but these interactions were not significant.

8 Discussion and Limitations

To our knowledge, our research is the first comprehensive comparison of models testing the factors that influence respondents' evaluations of surveys and the first to consider whether dual process models explain the conditions under which respondents use information about the quality of a survey's methodology.

We found no support for the rational actor model (H1) that suggests that respondents will evaluate surveys based on their quality. However, weak support was found for the science literacy model—there was some albeit limited evidence that the evaluations by more educated respondents were more likely to be affected by quality than the evaluations by less educated respondents (H2). That is, more educated respondents were more likely to evaluate survey quality in a rational manner.

When examining main effects of the predictors, the motivated reasoning model was strongly supported (H3)—respondents evaluated surveys that had findings consistent with their prior attitudes more positively than surveys whose results were at odds with those attitudes. However, the motivated reasoning evidence finding was qualified by a set of complex, yet statistically significant, interactions that suggested that the consistency of a survey result with one's opinion may also play a role as a motivator to process information about the survey's methodology, but only among respondents who are able to do so. We found strong support for the dual process model (H4) hypothesis suggesting that survey quality would influence evaluations when respondents were motivated *and* able to carefully consider survey quality.

Our research has several strengths. It uses data from three independent representative samples drawn using probability sampling thereby providing results that are generalizable to the population being studied and stronger external validity than nonprobability samples (used in most of the research on this topic),¹⁸ and it is strong in internal validity because randomized experiments were used to isolate the influence

of the key independent variables (cf. Lavrakas et al., 2019). We also combined data across surveys collected over multiple years that asked conceptually similar questions about three different proposed policies—suggesting that our results will generalize across issues.

Although these experiments were conducted some time ago, we believe that our reanalysis of the data originally gathered in these three vignette experiments remains timely in 2025, in part, because of the nature of the findings which we argue would very likely be even more compelling if the data had been gathered nowadays because of the much more intense political polarization that has arisen among Americans in the past two decades. Furthermore, since the original data were gathered, the credibility of U.S. election polls (and polls in several other countries) has become a controversial issue in itself. This change also plays into the belief that the experiments would have yielded stronger findings had they been conducted more recently. Finally, we believe that the findings remain relevant in 2025 because, although the way polls are conducted has changed dramatically over time, the basic psychological processes by which we posit people process information that is generated by election polls is not likely to have changed.

Furthermore, our data and analyses also suggest a number of potential directions for future research. One would be to conduct more current vignette experiments, as these data were gathered some time ago. However, we argue that we are measuring and testing general cognitive processes that are unlikely to change over time. We also acknowledge that our *R*-squared statistics are low, although they do identify statistically reliable findings. These values may, in part, be due to the single items used for our dependent variables. It also may be due to the cognitive complexity of the vignette sequence leading up to the dependent variables. That is, the complexity may have increased respondent-related measurement error more than would have occurred with a less complex questionnaire sequence. However, it also raises suspicions that our models were under-specified due to limited funding that prohibited a longer set of questions, and that there may be other important predictors that were not measured in these surveys. Finally, in each of the three surveys, the second dependent variable, consideration, was measured through an open-ended question format. Respondents' responses were coded into one of three predetermined categories (yes, no, uncertain). While this approach is understandable for a phone survey, it still poses a limitation in terms of capturing a wider range of variance in responses. Assigning responses to a broader range of op-

¹⁸ One of the advantages of conducting experiments in representative sample surveys is the high external validity (cf. Lavrakas et al. 2019). We can argue strongly that these results are generalizable to adult Ohio-

ans, but we acknowledge that we cannot know for sure whether these results would generalize to very different populations (e.g., other countries or cultures) or to experiments conducted in different contexts (e.g., a Web survey instead of a telephone survey).

tions for the dependent variable could be considered in the future.

Another potential direction for future research would be to examine the possible role of data collection mode on our findings. Because the vignettes were read to respondents, they could not go back to reconsider the methodology when they were told the results after the methodology or when they were reminded that survey accuracy might be important before being asked about the usage of the findings by policymakers. This logic suggests that one might find different results if the vignettes were presented in a self-administered questionnaire. There, the order of information within the vignette and the order of the dependent variables might matter less because respondents may have been able to go back and review the methodology if the survey result or the question about accuracy motivated them to do so.

9 Conclusion

We found support for several models that could be used to explain how members of the public think about and evaluate public opinion surveys. Like previous research, we found highly reliable evidence for a motivated reasoning perspective. However, our research represents the first application of a dual process model to this topic and we found highly reliable evidence that people use the quality of a survey's methodology to evaluate the survey when they are *both* motivated *and* able.

Acknowledgements The data used in this project were gathered by the Center for Survey Research at Ohio State University using funding from *The Columbus Dispatch* newspaper and the OSU College of Behavioral and Social Sciences. We also would like to thank Lillian Diaz-Castillo Hoffman and Quin Monson for helping to plan and conduct the original surveys and creating the original datasets.

References

- AAPOR (2023). *Standard definitions: final dispositions of case codes and outcome rates for surveys* (10th edn.). American Association for Public Opinion Research.
- Andernach, B., & Schunck, R. (2014). *Investigating the Feasibility of a Factorial Survey in a CATI*. SFB 882 Working Paper Series No. 28. Investigating the Feasibility of a Factorial Survey in a CATI (d-nb.info)
- Auspurg, K., & Hinz, T. (2014). *Factorial survey experiments*. New York: SAGE.
- Auspurg, K., Hinz, T., & Walzenbach, S. (2019). Are factorial survey experiments prone to survey mode effects? In P.J. Lavrakas, et al. (Ed.), *Experimental methods in survey research: techniques that combine random sampling with random assignment* (pp. 371–392). Hoboken: John Wiley.
- Berinsky, A.J., & Mendelberg, T. (2005). The indirect effects of discredited stereotypes in judgments of Jewish leaders. *American Journal of Political Science*, 49(4), 845–864.
- Canan, C., & Foroutan, N. (2016). Changing perceptions? Effects of multiple social categorisation on German population's perceptions of muslims. *Journal of Ethnic and Migration Studies*, 42(12), 1905–1924.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.
- Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual process theories in social psychology*. New York: Guilford.
- Chia S. C., Chang T. 2017. Not my Horse: Voter Preferences, Media Sources, and Hostile Poll Reports in Election Campaigns. *International Journal of Public Opinion Research* 29(1):23–45.
- Claypool, H. M., O'Mally, J., DeCoster, J. (2012). Dual-process models of information processing. In *Encyclopedia of the science of learning* (pp. 1046–1048). Boston: Springer.
- De Vreese, C.H., & Semetko, H.A. (2002). Public perception of polls and support for restrictions on the publication of polls: Denmark's 2000 euro referendum. *International Journal of Public Opinion Research*, 14(4), 367–390.
- Djupe, P.A., & Calfano, B.R. (2012). The deliberative pulpit? The democratic norms and practices of the PCUSA. *Journal for the Scientific Study of Religion*, 51(1), 90–109.
- Donovan, K., Kellstedt, P.M., Key, E.M., & Lebo, M.J. (2020). Motivated reasoning, public opinion, and presidential approval. *Political Behavior*, 42, 1201–1221.
- Downs, A. (1957). *An economic theory of democracy*. New York: Harper & Row.
- Dran, E.M., & Hildreth, A. (1995). What the public thinks about how we know what it is thinking. *International Journal of Public Opinion Research*, 7(2), 128–144.
- Durant, J. R., Evans, G. A., Thomas, G. P. (1989). Public understanding of science. *Nature*, 340, 11–14.
- Eifler, S., & Petzold, K. (2019). Validity aspects of vignette experiments: expected “what-if” differences between reports of behavioral intentions and actual behavior. In P.J. Lavrakas, et al. (Eds.), *Experimental methods in survey research: techniques that combine random sampling with random assignment* (pp. 393–413). Hoboken: John Wiley.

- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3), 133–140.
- Hopkins, D.J., & King, G. (2010). Improving anchoring vignettes: designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*, 74(2), 201–222.
- Jacobs, L.R., & Shapiro, R.Y. (2000). *Politicians don't pander: political manipulation and the loss of democratic responsiveness*. Chicago: University of Chicago Press.
- Johnson, T.P., Silber, H., & Darling, J.E. (2024). Public perceptions of pollsters in the United States: experimental evidence. *Social Science Quarterly*, 105, 114–127.
- Kahneman, D. (2013). *Thinking, fast and slow*. New York: Random House.
- Kim, J., Gershenson, C., Glaser, P., & Smith, T.W. (2011). Trends in surveys on surveys. *Public Opinion Quarterly*, 75(1), 165–191.
- Kim, S.T., Weaver, D., & Willnat, L. (2000). Media reporting and perceived credibility of online polls. *J&MC Quarterly*, 77(4), 846–864.
- Kuru, O., Pasek, J., & Traugott, M.W. (2017). Motivated reasoning in the perceived credibility of public opinion polls. *Public Opinion Quarterly*, 81(2), 422–446.
- Kuru, O., Pasek, J., & Traugott, M.W. (2020). When polls disagree: how competitive results and methodological quality shape partisan perceptions of polls and electoral predictions. *International Journal of Public Opinion Research*, 32(3), 586–603.
- Lavrakas, P.J., Holley, J.K., & Miller, P.V. (1991). Public reactions to polling news during the 1988 presidential election campaign. In P.J. Lavrakas & J.K. Holley (Eds.), *Polling and Presidential Election Coverage* (pp. 151–183). Newbury Park: SAGE.
- Lavrakas, P.J., Diaz-Castillo, L., & Monson, Q. (2000). *Experimental investigations of the cognitive processes which underlie judgments of poll accuracy*. Presented at the 2000 AAPOR Conference, Portland.
- Lavrakas, P.J., Kennedy, C., de Leeuw, E.D., West, B.T., Holbrook, A.L., & Traugott, M.T. (2019). Probability survey-based experimentation and the balancing of internal validity and external validity concerns. In P.J. Lavrakas, M.W. Traugott, C. Kennedy, A.L. Holbrook, E.D. de Leeuw & B.T. West (Eds.), *Experimental methods in survey research: techniques that combine random sampling with random assignment* (pp. 1–17). Hoboken: Wiley.
- Li, Y., & Guo, M. (2021). Scientific literacy in communicating science and socio-scientific issues: prospects and challenges. *Frontiers in Psychology*, 12, 758000.
- Loosveldt, G., & Storms, V. (2008). Measuring public opinions about surveys. *International Journal of Public Opinion Research*, 20(1), 74–89.
- Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
- Madison, G.J., & Hillygus, D.S. (2020). All the best polls agree with me: bias in evaluations of political polling. *Political Behavior*, 42, 1055–1072.
- McPhetres, J., Rutjens, B.T., Weinstein, N., & Brisson, J.A. (2019). Modifying attitudes about modified foods: Increased knowledge leads to more positive attitudes. *Journal of Environmental Psychology*, 64, 21–29.
- Miller, J.D. (1983). Scientific literacy: a conceptual and empirical review. *Scientific Literacy*, 112(2), 29–48.
- Miller, J.D., Scott, E.C., & Okamoto, S. (2006). Public acceptance of evolution. *Science*, 313, 765–766.
- Moy, P., & Rinke, E.M. (2012). Attitudinal and behavioral consequences of published opinion polls. In C. Holtz-Bacha & J. Strömbäck (Eds.), *Opinion polls and the media: reflecting and shaping public opinion* (pp. 225–245). Basingstoke: Palgrave Macmillan.
- National Science Board (NSB), National Science Foundation. 2022. Science and Technology: Public Perceptions, Awareness, and Information Sources. Science and Engineering Indicators 2022. NSB-2022-7. Alexandria, VA. Available at <https://nces.nsf.gov/pubs/nsb20227>.
- Nickerson, R.S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220.
- Noelle-Neumann, E. (1993). *The spiral of silence* (2nd edn.). Chicago: University of Chicago Press.
- Page, B.I. (1994). Democratic responsiveness? Untangling the links between public opinion and policy. *Political Science & Politics*, 27(1), 25–29.
- Pager, D., & Quillian, L. (2005). Walking the talk? What employers say versus what they do. *American Sociological Review*, 70(3), 355–380.
- Pardo, R. & Calvo, F. (2004). The cognitive dimension of public perceptions of science: Methodological issues. *Public Understanding of Science*, 13(3), 203–227.
- Petty, R.E., & Cacioppo, J.T. (1981). *Attitudes and persuasion: classic and contemporary approaches*. Dubuque: Brown.
- Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.

- Presser, S., Lavrakas, P., Price, V., & Traugott, M. (1998). *Public opinion about polls: how people decide whether to believe survey results*. St. Louis. Paper presented at the annual meeting of the American Association for Public Opinion Research.
- Price, V., & Stroud, N.J. (2005). Public attitudes toward polls: evidence from the 2000 U.S. presidential election. *International Journal of Public Opinion Research*, 18(4), 393–421.
- Redlawsk, D.P., Civettini, A.J.W., & Emmerson, K.M. (2010). The affective tipping point: Do motivated reasoners ever “get it”? *Political Psychology*, 31(4), 563–593.
- Rogelbert, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D. & Horvath, M. (2001). Organizational Research Methods, 4(1), 3-25.
- Rothschild, D., & Malhotra, N. (2014). Are public opinion polls self-fulfilling prophecies? *Research and Politics*. <https://doi.org/10.1177/2053168014547667>.
- Salwen, M.B. (1987). Credibility of newspaper opinion polls: source, source intent and precision. *Journalism Quarterly*, 64(4), 813–819.
- Schäfer, M. S. (2016). Social science in society. Public understanding of science, 25(3), 394-396.
- Shapiro, R. Y. (2011). Public opinion and American democracy. *Public Opinion Quarterly*, 75(5), 982–1017.
- Sinclair, B., & Plott, C.R. (2012). From uninformed to informed choices: voters, pre-election polls and updating. *Electoral Studies*, 31(1), 83–95.
- Stadtmüller, S., Silber, H., & Beuthner, C. (2022). What influences trust in survey results? Evidence from a vignette experiment. *International Journal of Public Opinion Research*. <https://doi.org/10.1093/ijpor/edac012>.
- Stefkovics, Á., & Kmetty, Z. (2024). Trust in survey results. A cross-country replication experiment. <https://osf.io/preprints/osf/crwq7>.
- Stocké, V. (2006). Attitudes towards surveys, attitude accessibility and the effects on respondents' susceptibility to nonresponse. *Quality & Quantity*, 40, 259–288.
- Stocké, V., & Langfeldt, B. (2004). Effects of survey experience on respondents' attitudes towards surveys. *Bulletin de Methodologie Sociologique*, 81, 5–32.
- Toff, B. (2018). Exploring the effects of polls on public opinion: how and when media reports of policy preferences can become self-fulfilling prophecies. *Research and Politics*. <https://doi.org/10.1177/2053168018812215>.
- Traugott, M. W. (1991). Public attitudes about news organizations, campaign coverage, and polls. Polling and presidential election coverage, 134-150.
- Tsfati, Y. (2001). Why do people trust media pre-election polls? Evidence from the Israeli 1996 elections. *International Journal of Public Opinion Research*, 13(4), 433–441.
- Vanette, D. & Westwood, S. (2013). Voter mobilization effects of poll reports during the 2012 presidential campaign. Presented at the 2013 Annual AAPOR Conference, Boston.
- Vargas, P. (2008). Vignette question. In P.J. Lavrakas (Ed.), *The encyclopedia of survey research methods* (pp. 947–949). Thousand Oaks: Sage.
- Weisberg, D.S., Landrum, A.R., Metz, S.E., & Weisberg, M. (2018). No missing link: knowledge predicts acceptance of evolution in the United States. *BioScience*, 68, 212–222.
- Weisberg, D.S., Landrum, A.R., Hamilton, J., & Weisberg, M. (2021). Knowledge about the nature of science increases public acceptance of science regardless of identity factors. *Public Understanding of Science*, 30, 120–138.
- Whiddett, D., Hunter, I., McDonald, B., Norris, T., & Waldon, J. (2016). Consent and widespread access to personal health information for the delivery of care: a large scale telephone survey of consumers' attitudes using vignettes in new zealand. *BMJ Open*, 6, e11640. <https://doi.org/10.1136/bmjopen-2016-011640>.