Pew Research Center

FOR RELEASE February 18, 2020

Assessing the Risks to Online Polls From Bogus Respondents

Approval ratings can be influenced several percentage points by bogus takers of opt-in polls

BY *Courtney Kennedy, Nick Hatley, Arnold Lau, Andrew Mercer, Scott Keeter, Joshua Ferno, and Dorene Asare-Marfo*

FOR MEDIA OR OTHER INQUIRIES:

Courtney Kennedy, Director, Survey Research
Rachel Weisel, Senior Communications Manager
Calvin Jordan, Communications Associate

202.419.4372

www.pewresearch.org

RECOMMENDED CITATION

Pew Research Center, February 18, 2020
"Assessing the Risks to Online Polls from Bogus Respondents"

About Pew Research Center

Pew Research Center is a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping the world. It does not take policy positions. It conducts public opinion polling, demographic research, content analysis and other data-driven social science research. The Center studies U.S. politics and policy; journalism and media; internet, science and technology; religion and public life; Hispanic trends; global attitudes and trends; and U.S. social and demographic trends. All of the Center's reports are available at www.pewresearch.org. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

[ACCORDIAN BOX] **How we did this**

We compared data from six online sources used for public polling, including three prominent sources of opt-in survey samples, one crowdsourcing platform, and two survey panels that are recruited offline using national random samples of residential addresses and surveyed online. One of the address-recruited samples comes from the Pew Research Center's American Trends Panel. The study included more than 60,000 interviews with at least 10,000 interviews coming from each of the six online sources. All samples were designed to survey U.S. adults ages 18 and over.

Assessing the Risks to Online Polls from Bogus Respondents

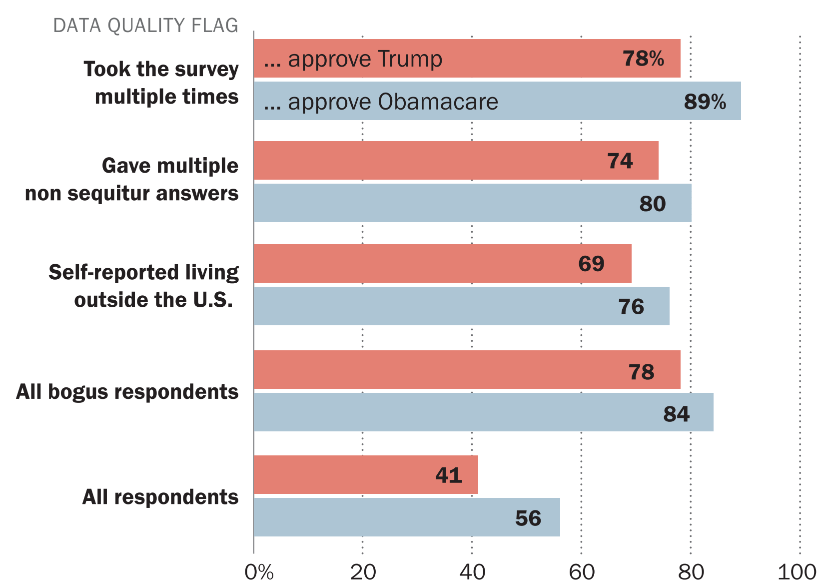
Approval ratings can be influenced several percentage points by bogus takers of opt-in polls

More than 80% of the public polls used to track key indicators of U.S. public opinion, such as the President's approval rating or support for Democratic presidential candidates, are conducted using online opt-in polling.[1] A new study by Pew Research Center finds that online samples fielded with widely-used opt-in sources contain small but measurable shares of bogus respondents (about 4% to 7%, depending on the source). Critically, these bogus respondents are not just answering at random, but rather they tend to select positive answer choices – introducing a small, systematic bias into estimates like presidential approval.

This pattern is not partisan. While 78% of bogus respondents reported approving of President Donald Trump's job performance, their approval rating of the 2010 health care law, also known as Obamacare, was even higher, at 84%. Open-ended answers show that some respondents answer as though they are taking a market research survey (e.g., saying "Great product" regardless of the question).

Bogus respondents tend to approve of everything

% of respondents flagged who say they…



Notes: Figures are unweighted.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

While some challenges to polls are ever-present (e.g., respondents not answering carefully or giving socially desirable answers), the risk that bad actors could

[1] This finding is based on an analysis of the 6,872 polls conducted between Jan. 1 and Sept. 19, 2019 and used by FiveThirtyEight.com to track presidential approval.

compromise a public opinion poll is, in some respects, a new one. It is a consequence of the field's migration toward online convenience samples of people who sign themselves up to get money or other rewards by taking surveys. This introduces the risk that some people will answer not with their own views but instead with answers they believe are likely to please the poll's sponsor. It also raises the possibility that people who do not belong in a U.S. poll (e.g., people in another country) will try to misrepresent themselves to complete surveys and accrue money or other rewards.

With that backdrop, this study was launched to measure whether behavior of this sort is present in widely-used online platforms and, if so, to what extent and consequence. This study defines a bogus respondent as someone who met one or more of the following criteria: took the survey multiple times; reported living outside the United States (the target population is U.S. adults); gave multiple non sequitur open-ended answers; or always said they approve/favor regardless of what was asked.

The study finds that not all online polls suffer from this problem. Online polls that recruit participants offline through random sampling of residential addresses have only trace levels of bogus respondents (1% in each of two address-recruited panels tested). In address-recruited panels, there are too few bogus cases to have a perceptible effect on the estimates.

Sources, sample sizes and incidences of bogus cases

	All interviews	Bogus interviews	% Bogus
Opt-in crowdsourced			
Workers who get paid for tasks posted to a crowdsourcing website			
Opt-in crowdsourced	10,879	756	7%
Opt-in survey panels			
Online convenience samples of people sourced from various places			
Opt-in panel 1	10,002	613	6%
Opt-in panel 2	10,000	375	4%
Opt-in panel 3	11,054	624	6%
ABS survey panels			
People recruited offline via random sampling of residential addresses but surveyed online			
ABS panel 1	10,178	97	1%
ABS panel 2	10,526	74	1%

Note: Company names are masked in this report because the purpose is to study how polls are conducted (i.e., the methodology), not which company was used.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

The study compares data from six online sources used for public polling: three prominent sources of opt-in survey samples (two marketplaces and one panel),[2] one crowdsourcing platform,[3] and two survey panels that are recruited offline using national random

[2] Opt-in survey marketplaces source respondents from various third-party companies (see Appendix A), most of which are themselves opt-in panels. For simplicity, this report uses the term "opt-in panel" to describe both these marketplaces and a sample drawn from a single panel.
[3] A crowdsourcing platform is a website on which employers can hire remotely located workers to perform discrete tasks (e.g., transcription, image coding).

samples of residential addresses but surveyed online. One of the address-recruited samples comes from the Center's American Trends Panel. The study included more than 60,000 interviews with at least 10,000 interviews coming from each of the six online sources. All samples were designed to survey U.S. adults ages 18 and over. Analyses are unweighted as this is an examination of the credibility of respondents' answers.[4]

This is not the first study to find untrustworthy interviews in online surveys. This study is the first, however, to compare data quality from multiple opt-in and address-recruited survey panels, as well as a crowdsourcing platform. This study is also the first to employ sample sizes large enough to reliably estimate the incidence of bogus respondents, as well as the demographics and political attitudes reported by bogus respondents in each source.

Some poll questions are more affected by bogus respondents than others. Questions that allow the respondent to give a positively valanced answer show larger effects than those that do not. For example, a classic poll question designed to get a high-level read on public sentiment asks whether things in the country are "generally headed in the right direction" or "off on the wrong track." The share saying "generally headed in the right direction" drops two percentage points in the opt-in survey panel polls when bogus respondents are removed. In the crowdsourced poll, the figure drops four points when removing bogus cases. However, other questions – such as political party affiliation or views on new gun laws – do not appear to map onto this behavior and show little to no influence from bogus cases on topline results.[5]

Bogus respondents can have small but measurable effect on opt-in poll results

% saying things in the country are "generally headed in the right direction" when bogus respondents are…

	...Included	...Excluded	Diff
Opt-in crowdsourced	37	33	-4
Opt-in panel 1	39	37	-2
Opt-in panel 2	37	35	-2
Opt-in panel 3	38	36	-2
ABS panel 1	30	30	0
ABS panel 2	31	31	0

Notes: Figures are unweighted. All six polls were conducted online. ABS refers to polls that are recruited offline through residential address-based sampling. Significant differences **in bold**.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

Part of the explanation is that a segment of opt-in respondents express positive views about everything – even when that means giving seemingly contradictory answers. This study includes seven questions in which respondents can answer that they "approve" or have a "favorable" view of

[4] When the analysis is run with weighted data, the main findings are basically unaffected. A summary is provided in Appendix A.
[5] Ahler and colleagues found that bogus cases can bias estimates of relationships between survey variables. While this topic is not addressed in this report, it could be explored with the publicly available micro-dataset.

something. About half of the question topics (Vladimir Putin, Theresa May, Donald Trump) tend to draw support from conservative audiences, while the others are more popular with left-leaning audiences (Emmanuel Macron, Angela Merkel and the 2010 health care law).[6] If respondents are answering carefully, it would be unusual for someone to express genuine, favorable views of all seven.

The study found 4% of crowdsourced respondents gave a favorable response to all seven questions, followed by 1% to 3% of the opt-in survey panel polls. There were a few such respondents in the address-recruited polls, but they constitute less than a half of one percent. The upshot is that small but nontrivial shares of online opt-in respondents seek out positive answer choices and uniformly select them (e.g., on the assumption that it is a market research survey and/or that doing so would please the researcher). In a follow-up experiment in which the order of responses was randomized, researchers confirmed that this approve-of-everything response style was purposeful (not simply a primacy effect) (see Chapter 8).

But these uniformly positive respondents are not alone in nudging approval ratings upward. Respondents exhibiting other suspect behaviors answer in a similar way. For example, if always-approving cases are set aside, the study finds that 71% of those giving multiple non sequiturs to open-ended questions approve of the 2010 health care law, as do 80% of those found taking the survey more than once.[7,8] Similarly, when always-approving cases are set aside, 42% of those taking the survey multiple times express a favorable view of Vladimir Putin, as do 32% of those giving multiple non sequitur answers. These rates are roughly three times higher than Putin's actual favorability rating among Americans (about 9%), according to high-quality polling. These patterns matter because they are suggestive of untrustworthy data that may bias poll estimates and not merely add noise.

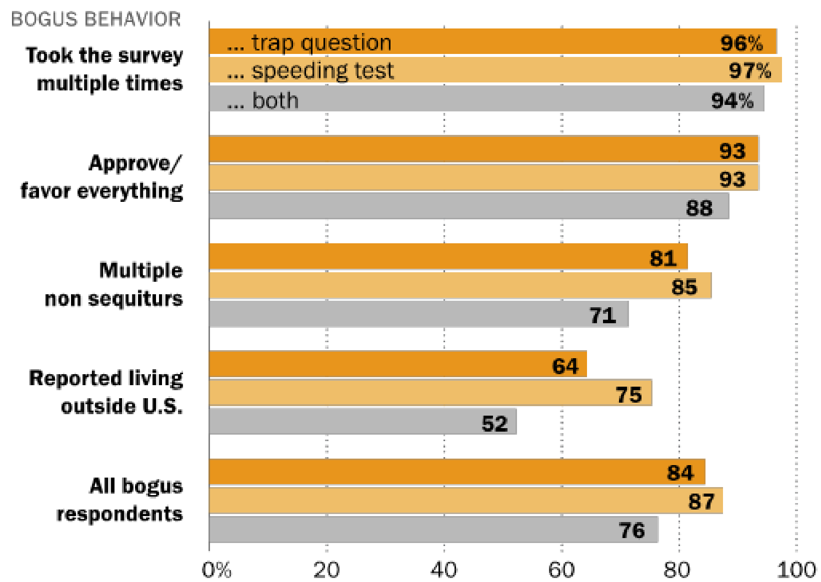
The study also finds that two of the most common checks to detect low quality online interviews – looking for respondents who answer too fast or fail an attention check (or "trap") question — are not very effective. The attention check question read, "Paying attention and reading the instructions carefully is critical. If you are paying attention, please choose Silver below." Some 84% of bogus respondents pass the trap question and 87% pass a check for responding too quickly.

[6] When asked about world leaders, respondents were allowed to answer "Never heard of." Donald Trump was the subject of two questions (presidential job approval and an overall favorability rating). Also, this survey was conducted while Theresa May was Prime Minister of the United Kingdom.
[7] These figures are computed slightly differently from those in the graphic on page 2. The page 2 estimates are based on all respondents who received the data quality flag mentioned. The estimates here are based on respondents who were flagged for the behavior mentioned but did not uniformly answer approve/favor to all seven such questions.
[8] For reference, a Kaiser Family Foundation poll conducted in April 2019 (around the time of this data collection) measured public support for the health care law at 50%.

After using those checks to remove cases, the opt-in recruited polls examined here still had 3% to 7% of interviewing coming from bogus respondents, compared to 1% in address-recruited online samples.

Checks for speeding and attention fail to catch most bogus respondents

% of bogus respondents passing…



Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

One of the more notable implications of the study is evidence suggesting people in other countries might be able to participate in polls intended to measure American public opinion. Other researchers have documented foreign respondents in India and Venezuela participating in American social science research using crowdsourcing platforms. This study confirms those findings. Some 5% of crowdsourced respondents were using an IP address based outside of the U.S., and the most common host countries for the foreign IP addresses were the Seychelles and India. In the address-recruited online samples, by contrast, the rate of foreign IP addresses was 1%, and the most common host countries for the foreign IP addresses were Canada and Mexico.[9] Virtually no respondents in the opt-in survey panel samples had IP addresses from outside the U.S., suggesting that the survey panels have controls in place guarding against that. Other key findings from the study include:

Bogus interviews were prone to self-reporting as Hispanic or Latino. Overall, 10% of study respondents identified as Hispanic, but the rate was three times higher (30%) among cases flagged for bogus behavior. According to the Census Bureau's American Community Survey, Hispanics make up 16% of the U.S. adult population. While some of the bogus respondents could very well be Hispanic, this rate is likely inflated for several reasons. In particular, Hispanic

[9] Differences between an internet user's location and their IP address location can arise from several factors. Those include where the owner of the IP has it registered, where the agency that controls the IP is located, and proxies. For example, it is common for users on the Verizon network who live in the northern U.S. to show a Canadian IP because that is where the controlling agency of the IP is located.

ethnicity was measured with a stand-alone "yes/no" question, so people answering at random would be expected to report "yes" about half the time. As a consequence of this greater propensity for bogus respondents to identify as Hispanic, substantive survey estimates for Hispanics (such as presidential approval) are at risk of much greater bias than for the sample as a whole (see Chapter 6).

Open-ended questions elicited plagiarized answers and product reviews from some opt-in and crowdsourced respondents. Responses to open-ended questions show that in all six sources, most respondents appear to be giving genuine answers that are responsive to the question asked. That said, 2% to 4% of opt-in poll respondents repeatedly gave answers that did not match the question asked, compared to 0% of address-recruited panel respondents. Further examination of the 6,670 non sequitur answers in the study revealed several different types: unsolicited product reviews, plagiarized text from other websites found when entering the question in a search engine, conversational text, common words and other, miscellaneous non sequitur answers. Plagiarized responses were found almost exclusively in the crowdsourced sample, while answers sounding like product reviews as well as text sounding like snippets from a personal conversation were more common in the opt-in survey panels.

One open-ended question was particularly effective for detecting bogus respondents. The question, "What would you like to see elected leaders in Washington get done during the next few years? Please give as much detail as you can," elicited twice as many plagiarized answers as the question eliciting the second most (176 versus 78). Two thirds (66%) of the plagiarized answers were snippets from various biographies of George Washington. These respondents (nearly all of whom were from the crowdsourced sample) had apparently put the question into a search engine, and the first two search results happen to be online biographies of the first U.S. President.

Do changes of 2 or 3 percentage points really matter?

Findings in this study suggest that– with multiple, widely used opt-in survey panels – estimates of how much the public approves or favors something are likely biased upward, unless the pollster performs data cleaning beyond the common checks explored here. The bias comes from the roughly 4% to 7% of respondents who are either not giving genuine answers or are not actually Americans. Online polls recruited offline using samples of addresses do not share this problem because the incidence of low-quality respondents is so low. In absolute terms, the biases documented in this report are small and their consequences can be viewed several ways:

- It almost certainly does not matter if, in a single poll, the President's approval rating is 43% versus 41%. Such a difference is typically within the margin of error and does not change what poll says about the overall balance of public sentiment.

- It is more debatable whether it matters if numerous national polls are overestimating public support for a policy or a president by a few percentage points. For policies like the Affordable Care Act, where public support has been somewhat below or somewhat above 50%, a small, systematic bias across polls could conceivably have consequences.

- It's also important to consider what happens if policy makers and the public lose more trust in polls due to data coming from people who give insincere answers or who should not be in the survey in the first place. The problems uncovered in this study are minor in any given survey, but they point to the potential for much more serious problems in the near future, as reliance on opt-in samples increases and the barriers for entry into the public polling field continue to fall. For researchers using random national sampling or even well-designed opt-in samples, one risk is that a highly public scandal involving a low-quality opt-in sample has the potential for damaging the reputation of everyone in the field. This research suggests that there is considerable work to be done to reduce this risk to an acceptable level.

Bots or people answering carelessly

Fraudulent data generated by survey bots is an emergent threat to many opt-in polls. Survey bots are computer algorithms designed to complete online surveys automatically. At least one such product is commercially available and touts an "undetectable mode" with humanlike artificial intelligence. Bots are not a serious concern for address-recruited online panels because only individuals selected by the researcher can participate. They are, however, a potential concern for any opt-in poll where people can self-enroll or visit websites or apps where recruitment efforts are common.

There are numerous anecdotal accounts of bots in online opt-in surveys. Rigorous research on this issue, by contrast, is scarce. One major difficulty in such research is distinguishing between bots and human respondents who are simply answering carelessly. For example, logically inconsistent answers or nonsensical open-ended answers could be generated by either a person or a bot. This report details the response patterns observed and, where possible, discusses whether the pattern is more indicative of a human or an algorithm. Categorizing cases as definitively bot or not a bot is avoided because typically the level of uncertainty is too high. On the whole, data from this study suggest that the more consequential distinction is between interviews that are credible versus those that are not credible (or bogus), regardless of the specific process generating the data.

Implications for polling

The study finds that no method of online polling is perfect, but there are notable differences across approaches with respect to the risks posed by bogus interviews. The crowdsourced poll stands out as having a unique set of issues. Nearly all of the plagiarized answers were found in that sample, and about one-in-twenty respondents had a foreign IP address. For experimental research, these problems may be mitigated by imposing additional controls and restricting participation to workers with a task completion or approval rate of at least 95%. But requiring a 95% worker rating is a dubious criterion for polls purporting to represent Americans of all abilities, education levels and employment situations. Furthermore, the presence of foreign respondents was just one of several data quality issues in the crowdsourced sample. If all the interviews with a foreign IP address are removed from the crowdsourced sample, the rate of bogus respondents (4%) is still significantly higher than that found in samples recruited through random sampling.

For online opt-in survey panels and marketplaces, concerns about data quality are longstanding. Perhaps the most noteworthy finding here is that bogus respondents can have a small, systematic effect on questions offering a positively valanced answer choice. This should perhaps not come as a surprise given that many if not most surveys conducted on these platforms are market research assessments of how much people approve or disapprove of various products, advertisements, etc. It is difficult to find any other explanation for out-of-the-blue answers like, “m I love this has good functions meets the promise and is agreed to the money that is paid for it. ans.” This study suggests that some quality checks may help detect and remove some of these cases. But it is unclear which public pollsters have routine, robust checks in place and how effective they are. This study shows that if no quality checks are done, one should expect approval-type estimates to be impacted.

To be sure, opt-in polls do not have a monopoly on poor respondent behavior. A number of address-recruited respondents failed various data quality checks in this study. That said, the incidences were so low that poll estimates were not affected in a systematic way.

Does this study mean that polls are wrong?

No. While some of the findings are concerning, they do not signal that polling writ large is broken, wrong or untrustworthy. As the 2018 midterm (and even national-level polling from the 2016 election) demonstrated, well-designed polls still provide accurate, useful information. While not included in this study, other methods of polling – such as live telephone interviewing or one-off surveys in which people are recruited though the mail to take an online survey – can perform well when executed carefully.

As for online polls, the study finds that survey panels recruited offline using random sampling of mailing addresses performed very well, showing only a trace level of bogus respondents. Panels and marketplaces that use opt-in sourcing showed higher levels of untrustworthy data, but the levels were quite low. Rather than indicating some polls are wrong, this study documents a number of data quality problems – all of which are currently low level but that have the potential to grow worse in the near future.

Overview of research design

This study was designed to measure the incidence of untrustworthy interviews in online platforms routinely used for public polls. Center researchers developed a questionnaire (Appendix E) containing six open-ended questions and 37 closed-ended questions. The beginning of the questionnaire is designed to look and feel like a routine political poll. In fact, the opening questions are modeled after those used by several of the most prolific public polls conducted online.

As other researchers have noted, open-ended questions can be an effective tool for identifying problematic respondents. Open-ended questions (e.g., "What would you like to see elected leaders in Washington get done during the next few years?") require survey-takers to formulate answers in their own words. Researchers leveraged this to categorize open-ended answers for several suspicious characteristics (see Appendix B). Similarly, a number of closed-ended questions were also designed to detect problematic responding (see Chapter 7). Other questions probed commonly polled topics such as evaluations of presidential job performance and views of the Affordable Health Care Act.

In total, six online platforms used for public polling were included. Three of the sources are widely used opt-in survey panels. One is an opt-in crowdsourcing platform. Two of the sources are survey panels that interview online but recruit offline. For both panels recruited offline, most panelists were recruited using address-based sampling (ABS), and so "address-recruited" is used throughout the report as a shorthand. Before using ABS, both panels recruited offline using random samples of telephone numbers (random digit dialing). For the purposes of this study, the important property is that everyone in these two panels was recruited offline by randomly sampling from a frame that covers virtually all Americans.[10]

Each sample was designed to achieve at least 10,000 interviews with U.S. adults age 18 and older in all 50 states and the District of Columbia. Data collection took place in March and April 2019. The exact field dates for each sample and additional methodological details are provided in Appendix A. The micro-dataset is available for download on the Pew Research Center website.

Limitations and caveats

Generalizability is challenging in studies examining the quality of online opt-in surveys because such surveys are not monolithic. Sample vendors and public pollsters vary widely both in their quality control procedures and the extent to which those procedures are communicated publicly. While some organizations publicize the steps they take to identify and remove bogus respondents, the practice is far from universal, and a review of methods statements from opt-in polls used to track presidential approval, for instance, turned up no mention of data quality checks whatsoever. This makes it difficult for even savvy consumers of polling data to determine what kind of checking, if any, has been performed for a given poll.

Broadly speaking, this study speaks to online polls where the pollster performs little to no data quality checking of their own. To the extent that public pollsters routinely use sophisticated data quality checks – beyond the speeding and trap questions addressed in this report – the results from this study may be overly pessimistic.

While it is not reasonable to expect such pollsters to detail exactly how they try to detect bogus cases (as that may tip off bad actors), some discussion of procedures in place would be useful to polling consumers trying to ascertain whether this issue is addressed at all. A plausible scenario is that at least some pollsters rely on the panels/marketplaces selling the interviews to be responsible for data quality and security. The data in this study were collected under just that premise, and the results demonstrate that reliance on opt-in panels can lead to non-trivial shares of bogus cases.

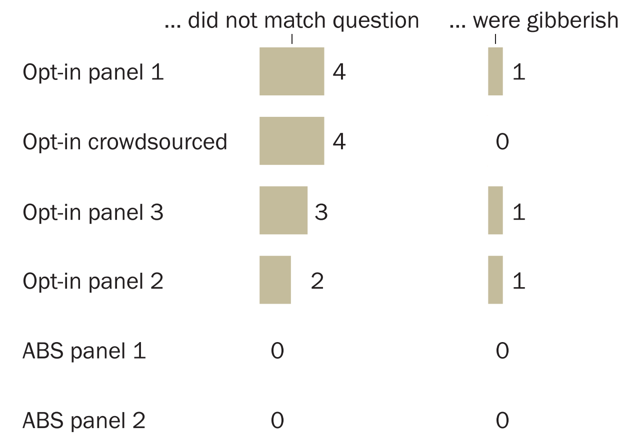
[10] The panels recruited offline could both be described as probability-based panels because the probability that virtually all U.S. adults have of being selected for inclusion into the panel is known. The three opt-in survey panels and the crowdsourced sample, by contrast, are all nonprobability sources because the chances of selection are not known. This report uses the more descriptive terms ("address-recruited," "opt-in recruited," and "crowdsourced") to focus on the processes used.

1. Answers that did not match the question were concentrated in opt-in polls

By their very nature, open-ended questions pose a greater cognitive challenge to respondents than closed-ended questions. As a result, answers to them also offer a more sensitive indicator of whether a respondent is sincere or not.

About 2% to 4% of opt-in respondents repeatedly give answers that don't follow from the question

% respondents giving two or more answers that …



Notes: All six polls were conducted online. ABS refers to polls that are recruited offline through residential address-based sampling. Figures are unweighted and based on the six open-ended questions asked in the survey.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

This study included six open-ended questions:

- How would you say you are feeling today?
- When you were growing up, what was the big city nearest where you lived?
- When you visit a new city, what kinds of activities do you like to do?
- How do you decide when your computer is too old and it's time to purchase a new one?
- In retirement what skill would you most like to learn?
- What would you like to see elected leaders in Washington get done during the next few years? Please give as much detail as you can.

Researchers manually coded each of the 375,834 open-ended answers into one of four categories: responsive to the question; does not match the question; gibberish; or did not answer (respondent left it blank or gave a "don't know" or refusal type answer).[11] In all six sources most respondents appear to be giving genuine answers that are responsive to the question asked. That said, the study found that 2% to 4% of opt-in poll respondents repeatedly gave answers that did not match the question asked. Throughout the report we refer to such answers as non sequiturs. There were a

[11] Appendix B provides the protocol for the open-ended coding. Appendix C provides the inter-coder reliability analysis. Researchers initially coded blank and don't know/refusal type answers separately. In this report, however, those categories are combined.

few such respondents in the address-recruited panel samples, but as share of the total their incidence rounds to 0%.

Some non sequitur answers are more suspicious than others. For example, when asked “When you were growing up, what was the big city nearest where you lived?,” some respondents gave geographic answers like “Ohio” or “TN.” But others gave answers such as “ALL SERVICES SOUNDS VERY GOOD” or “content://media/external/file/738023”.[12] Answering with a state when the question asked for a city is qualitatively different from answering with text that has absolutely no bearing on the question. This is among the reasons this report focuses on cases that gave multiple non sequiturs, as opposed to just one. Multiple instances of giving answers that do not match the question is a more reliable signal of a bogus interview than a single instance.

The study found that gibberish answers to open-ended questions were rare in both opt-in recruited and address-recruited online polls. The share of respondents giving multiple gibberish responses ranged from 0% to 1% depending on the sample. Examples of gibberish include “hgwyvbufhffbbibzbkjgfbgjjkbvjhj” and “Dffvdgggugcfhdggv Jr ffv. .” Certain letters, especially F, G, H, and J, were prominent in many gibberish answers. These letters are in the middle of QWERTY keyboards, making them particularly convenient for respondents haphazardly typing away. For a bot, by contrast, no letter requires more or less effort than another letter. This suggests that gibberish answers were probably given by humans who were satisficing (answering in a lazy manner) as opposed to bots.

There were differences across sources in the rates of not answering the open-ends, but that difference stems from administrative factors. The share giving a blank, “don’t know,” or refusal-type answer to two or more open-ended questions ranged from 13% to 19% for the address-recruited samples versus 1% to 2% for the opt-in samples. One reason for this is that people in the address-based panels are told explicitly that they do not have to answer every question. Each Pew Research Center survey begins with a screen that says, “You are not required to answer any question you do not wish to answer.” This is done to respect respondents’ sensibilities about content they may find off-putting for one reason or another. With public polls using opt-in panels, by contrast, it is common for answering questions to be required, though some pollsters permit not answering. In this study, opt-in survey panel respondents were not shown any special instructions and were required to answer each question. A follow-up data collection discussed in Chapter 8 allowed opt-in respondents to skip any question. When allowed to skip questions, 5% of

[12] Open-ended answers presented in this report are unedited. This means that all characters, symbols, spelling and grammatical errors are original to the answers given. Notably, this includes ellipses (…), which all came from respondents and not from the research team. The only exception to this policy is on page 21 where an expletive was redacted.

opt-in survey panel respondents gave a blank, don't know, or refusal type answer to two or more open-ended questions.

Non sequitur answers came in several forms

Examination of the 6,670 non sequitur answers revealed several different types: apparent product reviews, plagiarized text from other websites, conversational text, common words and miscellaneous non sequiturs.

Bogus answers took several forms, including product reviews and plagiarism

Product reviews *(two examples out of 1,017 total non sequiturs of this type)*

Great product

it is fun and easy to use

Plagiarized answers *(two examples out of 444 total non sequiturs of this type)*

Hack Reactor teaches you to think like a software engineer. Our grads are prepared for a world where next year's most important tech hasn't been invented yet.

Washington served as a general and commander-in-chief of the colonial armies during the American Revolution, and later became the first president of the United States, serving from 1789 to 1797. He died on December 14, 1799, in Mount Vernon, Virginia.

Conversational answers *(two examples out of 412 total non sequiturs of this type)*

Thank God for you and your family are doing well and that you are

Y'all need a panda tail to go to bed and go get food or drinks sugar or drinks and then I eat a chicken nuggets

Common words *(two examples out of 1,753 total non sequiturs of this type)*

Ok

YES

Miscellaneous *(two examples out of 3,044 total non sequiturs of this type)*

Bedroom Pop

content://media/external/file/738023

Notes: All text, punctuation and symbols, including ellipses (...) are original to the answer given, not inserted by the researcher. This analysis combines data from all six open ended questions and all six samples.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls from Bogus Respondents"

PEW RESEARCH CENTER

Some respondents, particularly in the crowdsourced poll, gave plagiarized text

Some of the answers not only made no sense but were clearly lifted from elsewhere on the internet. Often these answers came from websites that are top hits when one enters the survey question into a search engine. For example, a Google search for "What would you like to see elected leaders in Washington get done during the next few years?," returns a webpage on MountVernon.org, two articles on WashingtonPost.com, and one webpage for the Washington State Legislature. Opt-in respondents gave answers plagiarizing text from all four.

The question "What would you like to see elected leaders in Washington get done during the next few years?" elicited 176 plagiarized answers

Examples of plagiarized answers

It seemed as if everyone rejoiced at the election of our first chief executive except the man ... On February 4, 1789, the 69 members of the Electoral College made George

Crowdsourced respondent lifting from https://www.smithsonianmag.com/history/george-washington-the-reluctant-president-49492/

Washington officially the State of Washington, is a state in the Pacific Northwest region of the .... It includes large areas of semiarid steppe and a few truly arid deserts in the rain shadow of the ...... The North Cascades Highway, State Route 20, closes every year due to snowfall and avalanches in the area of Washington Pass.

Crowdsourced respondent lifting from https://en.wikipedia.org/wiki/Washington_(state)

A whole new season of competition, cooperation, screaming, and storming off the set start here with the chilling tale of Camp Grizzly! It's like Friday the 13th but with more cross-dressing and fewer mommy issues.

Opt-in panel 1 respondent lifting from https://vrv.co/watch/G6J0KXM0R/Board-as-Hell:Board-as-Hell-Halloween-Special-Camp-Grizzly

By signing up you agree to receive email newsletters or alerts from ... We now have the answer from special counsel Robert Mueller to the ... want to see the report, and hear from him why he made the decision he ... is one of the central issues for Congress for the next several months. ... The 2020 Election.

Crowdsourced respondent lifting from https://www.politico.com/newsletters/playbook/2019/03/25/mueller-is-done-whats-next-for-washington-414520

The nine elected executives serve a four-year term. All run independently. The Superintendent of Public Instruction may not declare a partisan affiliation. They are listed in their order of ascension to the office of Governor

Opt-in survey panel 1 respondent lifting from http://leg.wa.gov/legislature/Pages/ElectedOfficials.aspx

Notes: The full question text was, "What would you like to see elected leaders in Washington get done during the next few years? Please give as much detail as you can." All text and punctuation, including ellipses, are original to the answer, not inserted by the researcher.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

The study found 201 respondents plagiarizing from 125 different websites. Appendix D lists the source website for each plagiarized answer that was detected. While there were a few instances of this occurring in one of the opt-in survey panels, almost all of the respondents who gave plagiarized answers came from the crowdsourced sample (97%). No plagiarizing respondents were found in the address-recruited online polls. Some 194 respondents in the crowdsourced poll (2%) were found giving at least one plagiarized answer.

One possible explanation for the plagiarism is that some respondents were not interested in politics and searched for help crafting an acceptable answer. Even in this best case scenario it is very questionable that the respondents genuinely held the views expressed in their answers. Many other respondents answered openly that they do not follow or care about politics, without plagiarizing.

The charitable explanation of respondents just needing a little help crafting an answer falls apart when considering responses to “How would you say you are feeling today?” This is a simple question for which it is almost unimaginable that someone would bother to plagiarize an answer. In this study 35 respondents did so. They gave answers such as:

> The word feeling implies that the person is able to change from feeling to feeling say in physical wellbeing after ...
>
> Crowdsourced respondent lifting from https://preply.com/en/question/how-are-you-feeling-today-and-how-do-you-feel-today-41333

> It could be interpreted as "I am understanding and feeling the same emotions as ... I don't recall ever hearing someone say "I feel you" to mean .
>
> Crowdsourced respondent lifting from https://ell.stackexchange.com/questions/81023/saying-i-feel-you-in-a-conversation

> Eventually, the cold or flu will go away and you will get over it or get better. When you are feeling 80 % better, you might describe yourself as over the worst. If you are completely better, you may say you have recovered.
>
> Crowdsourced respondent lifting from https://dictionaryblog.cambridge.org/2014/11/12/are-you-feeling-any-better-talking-about-colds-and-flu/

Many plagiarized responses were remarkably similar. Not only did they tend to pull from some of the same websites; they often lifted the exact same text. For example, when asked “When you visit a new city, what kinds of activities do you like to do?,” 15 respondents gave the same passage of text from an article in the travel section stltoday.com.

When researchers tracked plagiarized responses back to their source website, they found variance in where text was lifted. In some cases, the plagiarized text was at or very near the top of the webpage. In other cases, portions of text were lifted from the middle or bottom of a page. Among those plagiarizing, it was common for them to join different passages with an ellipsis (…). It seems likely that at least some of the ellipses come from the respondent copying directly from the search engine results page.

In numerous instances, different respondents plagiarized the same text

Researcher screenshot of answers to "When you visit a new city, what kinds of activities do you like to do?"

```
==========================================================================================================
CITYVISIT
==========================================================================================================
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot...
```

Notes: Each row corresponds to one interview.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

Those who plagiarized did not do so consistently. Of the 201 plagiarizers, only seven were found to have plagiarized answers for all six open-ends. On average, respondents detected doing this at least once gave plagiarized answers to two of the six open-ends. When the plagiarizers appeared to answer on their own, the results were not good. Their other answers were often non sequiturs as well, just not plagiarized. For example, it was common for plagiarizers to answer "When you were growing up, what was the big city nearest where you lived?" by giving the name of a state. This is suggestive of perhaps a vague but not precise understanding of the question. The inconsistent reliance on plagiarism is more suggestive of a human answering the survey than a bot.

And while plagiarized answers were found for each open-end, the question asking respondents what they would like to see elected leaders in Washington get done was by far the most likely to elicit this behavior. Researchers detected 176 plagiarized answers to that question. The second highest was 78 plagiarized answers to "When you visit a new city, what kinds of activities do you like to do?"

Question about elected leaders in Washington elicited most plagiarized responses

Number of plagiarized answers found for each open-ended question

Notes: Each open-ended question was administered to all 62,639 study respondents.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

The question about elected leaders in Washington appears to have been the Waterloo for untrustworthy respondents for at least two reasons. One is the homonym "Washington," and the other is conceptual difficulty. It was serendipitous that the top search result for "What would you like to see elected leaders in Washington get done during the next few years?," was https://www.mountvernon.org/george-washington/the-first-president/election/10-facts-about-washingtons-election/, which is clearly the wrong meaning of "Washington." In total, 117 respondents answered the question with text plagiarized from this or another online biography of the first U.S. president.

Apart from the homonym, it seems likely that some respondents found this question more difficult to answer. Questions like "How are you feeling today?" require no special knowledge and can be answered in a word or two. By contrast, this question about elected officials achieving unspecified goals almost certainly posed a greater cognitive challenge.

Additionally, the instruction to "Please give as much detail as you can" alerted respondents to the fact that longer answers were desirable. It seems plausible that some people felt comfortable answering the simpler questions on their own but looked for a crutch to come up with an answer to the elected leaders question. These two characteristics of having a homonym with multiple popular meanings and probing a relatively challenging topic may prove useful to researchers writing future questions designed to identify untrustworthy respondents.

Some respondents answered as though they were reviewing a product

About 15% of all the non sequitur answers sounded like a product review. For example, when asked, "When you were growing up, what was the big city nearest where you lived?," over 150 respondents said "excellent," "great," "good," or some variation thereof.[13] More pointed answers included "awesome stocking stuffers" and "ALL SERVICES SOUNDS VERY GOOD." None of these answers are responsive to the question posed. Researchers coded whether these evaluations were positive or negative and found that almost all of them (98%) were positively valanced.

While rare in all sources, product review and plagiarized non sequitur responses were found almost exclusively in the opt-in samples

% of respondents giving each kind of non sequitur answer at least once

	Plagiarized	Product review	Conversation	Common words	Other
Opt-in crowdsourced	2	1	0	2	5
Opt-in panel 1	0	1	1	3	5
Opt-in panel 2	0	1	0	2	3
Opt-in panel 3	0	1	0	3	4
ABS panel 1	0	0	0	1	1
ABS panel 2	0	0	0	0	1

Notes: All six polls were conducted online. ABS refers to polls that are recruited offline through residential address-based sampling. Figures are unweighted and based on the six open-ended questions asked in the survey.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

While rare in all the sources tested, product review sounding answers were found almost exclusively in the opt-in samples. Researchers found that 1% of the respondents in the crowdsourced sample and each of the opt-in samples gave a product-review sounding answer to at least one of the open-ends, compared to 0% of the address-recruited respondents.

There are two plausible explanations that stand out as to why some respondents offered these bizarre answers that had nothing to do with the question. Opt-in surveys are routinely (if not mostly) used for market research – that is, testing to determine how best to design or market a product like insurance, automobiles, cosmetics, etc. If someone were looking to complete a high

[13] For five of the six open ended questions, answers such as "good," "terrible," or "excellent" were clearly non sequitur and coded as such. However, for "How would you say you are feeling today?," such answers are credible and were coded as responsive.

volume of opt-in surveys with minimal effort, they might simply give answers assuming each survey is market research. If one assumes that the question is asking about a product, then rote answers such as "I love it" and "awesome" would seem on target.

A second explanation stems from the fact that many online surveys end with a generic open-ended question. For example, questions like "Do you have any feedback on this survey?" or "Do you have any additional comments?" are common. Even if a respondent did not necessarily assume the survey was market research, they may have assumed that any open-ended question they encountered was of this general nature. In this scenario, answers like "good" or "I like so much" offer rote evaluations of the survey itself rather than a product asked about in the survey.

A few of the non sequitur answers explicitly talked about a product ("Great product," "product is a good," "ITS IS EXCELLENT BRAND I AM SERVICE."), presumably placing them under the first explanation. Other answers explicitly talked about the survey ("this was a great survey," "I love this survey I want more"), presumably placing them under the second explanation. Far more non sequitur answers, though, simply offered comments like "good" or "great," making it impossible to know what they were intended to reference. The commonality with all of these cases is that they offered an evaluation that may have made sense for a different (perhaps market research) survey but was unrelated to the question asked.

Some answers sounded like snippets of a conversation between two people

The study also found respondents giving non sequitur answers that sounded like snippets from a personal conversation. For example, when asked what was the big city nearest where they grew up, one respondent said, "Yes we can have dinner at the" and another said, "I love love joe and ahhh so much fun and joe and joe joe." The repetition of words exhibited by the latter answer is common among these conversational sounding non sequiturs.

Over 400 answers sound like snippets of a personal conversation

Examples of conversational sounding non sequiturs to "When you visit a new city, what kinds of activities do you like to do?"

Y'all need a panda tail to go to bed and go get food or drinks sugar or drinks and then I eat a chicken nuggets

– Respondent in opt-in panel 1

You can please just hang in there and I look forward to hearing from you soon

– Respondent in opt-in panel 3

Thank God for you and your family are doing well and that you are

– Respondent in opt-in panel 1

Thanks for the update and for the update and for

– Respondent in opt-in panel 2

C.J please don't have yo [expletive] touches my bad for a while now that don't make no same thing to me because they screen shot it from Instagram and then share it from Instagram and then share it from Instagram and then share it from Instagram and then share it from

– Respondent in opt-in panel 1

Notes: All text, punctuation and symbols, including ellipses (…) are original to the answer given, not inserted by the researcher.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

A likely explanation for these types of responses is the use of predictive text features on mobile devices. When a smartphone user begins typing, most smartphones attempt to predict the desired word and offer suggestions for words that are likely to come next based on that person's prior texts and emails. These features are intended to reduce the amount of typing that users need to do on their devices. However, repeatedly tapping the predictive text results in sequences where each consecutive word would plausibly follow the one that came before it in a conversation but are otherwise meaningless.

Indeed, virtually all (99%) of respondents giving multiple conversational non sequiturs completed the survey on a mobile device (smartphone or tablet). Overall, 53% respondents in the study answered on a mobile device (this does not include the crowdsourced sample, for which device type was not captured).

In general, these conversational sounding answers are exceedingly rare. Less than 1% of all respondents in the study gave an answer of this type. When they did occur, these answers were concentrated in two of the opt-in samples. Of the 134 total respondents giving at least one conversational non sequitur answer, 46% were in opt-in panel 1 and 37% were in opt-in panel 3. The crowdsourced poll and opt-in panel 2 accounted for an additional 8% and 7% of these respondents, respectively.

Many non sequitur answers were just common words

In total, researchers identified five types of non sequitur answers: those that looked like a product evaluation, plagiarism, snippets of conversation, common words, and other/miscellaneous. Most non sequitur answers fell into those last two groups. For example, when asked "How would you say you are feeling today?," 40 respondents said "yes" or "si" and an additional 42 respondents said "no." Such answers were coded as non sequiturs using common words.

The bucket of other/miscellaneous non sequiturs contained any non sequitur that did not fall into the other four buckets. To give some examples, when asked "How would you say you are feeling today?," other/miscellaneous non sequitur answers included "Bedroom Pop," "Bot," and "content://media/external/file/738023". These answers were relatively heterogeneous with no clear explanation as to how they were generated.

Giving the same answer to six different open-ended questions

Giving the exact same answer to each of the six open-ended questions is another suspicious pattern. Typically, these were single word answers such as "Yes," "like," or "good". Of all the patterns in this study that could be indicative of a bot as opposed to a human respondent, this is perhaps the most compelling. That said, it is far from dispositive. A human respondent who is satisficing could generate the same pattern.

Some respondents gave the exact same answer to all open-ended questions

Verbatim answers (10 examples out of 160 respondents giving the same answer to every open-ended question)

How would you say you are feeling today?	When you were growing up, what was the big city nearest where you lived?	When you visit a new city, what kinds of activities do you like to do?	In retirement what skill would you most like to learn?	What would you like to see elected leaders in Washington get done during the next few years?
Yes	Yes	Yes	Yes	Yes
Si	Si	Si	Si	Si
No	No	No	No	No
lol	lol	lol	lol	lol
None	None	None	None	None
I like it	I like it	I like it	I like it	I like it
Is good	Is good	Is good	Is good	Is good
I agree	I agree	I agree	I agree	I agree
initials	initials	initials	initials	initials
too much well	too much well	too much well	too much well	too much well

Note: Each row corresponds to one interview. Due to space constraints, five of the six open-ended question are shown. Answers to the sixth open end are identical to those displayed.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

Giving the exact same answer to all six open-ended questions is exceedingly rare. The overall study incidence is less than one percent (0.3%). Due to the study's very strong statistical power, it is possible to detect variation across the sample sources. This behavior was almost completely absent (0.0%) from the crowdsourced sample as well as the two address-recruited panel samples. Among opt-in panels, however, roughly 1-in-200 respondents did this (0.5%). The incidences were similar across the opt-in panels, ranging from 0.4% to 0.6%.

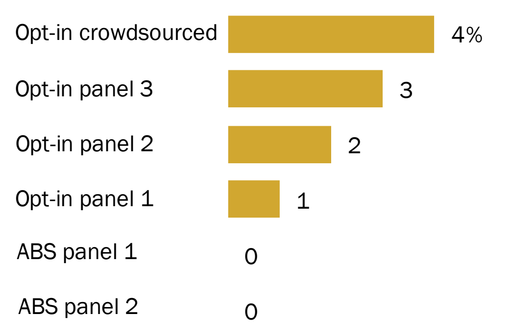
2. Respondents who approve of everything

The study found a segment of respondents who expressed positive views about everything – even when that meant giving seemingly contradictory answers. This suggests untrustworthy data that stands to bias poll estimates. If a nontrivial share of respondents seek out positive answer choices and always selected them (e.g., on the assumption that it is a market research survey and/or that doing so would please the researcher), that could systematically bias approval ratings upward. The study included seven questions in which respondents could answer that they approve or favor something. Specifically, the survey asked:

- Do you approve or disapprove of the job Donald Trump is doing as President?
- What is your overall opinion of U.S. President Donald Trump?[14]
- What is your overall opinion of British Prime Minister Theresa May?
- What is your overall opinion of Russian President Vladimir Putin?
- What is your overall opinion of German Chancellor Angela Merkel?
- What is your overall opinion of French President Emmanuel Macron?
- Do you approve or disapprove of the health care law passed by Barack Obama and Congress in 2010?

About 4% of crowdsourced respondents say they approve of everything

% answering approve/favorable for all seven questions



Notes: All six polls were conducted online. ABS refers to polls that are recruited offline through residential address-based sampling.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"
PEW RESEARCH CENTER

If respondents are answering carefully, it would be unusual to express genuine, favorable views of Emmanuel Macron, Angela Merkel, the Affordable Care Act (ACA), Theresa May, Donald Trump and Vladimir Putin. The first half of the list tend to draw support from left-leaning audiences while the latter are more popular with conservative audiences.

The study found 2% of respondents gave an approve or favorable response to each of these seven questions. The rate was highest in the crowdsourced poll (4%) followed by all three opt-in panels (ranging from 1% to 3%). There were a few such respondents in the address-recruited polls, but as share of the total their incidence rounds to 0%. Researchers confirmed that

[14] This battery asking about opinions of world leaders offered respondents an explicit "Never heard of option" for each leader.

this behavior was purposeful – not simply a primacy effect – in a follow-up experiment in which the order of responses was randomized (see Chapter 8).

While approving of everything might seem benign, it was strongly associated with bad data quality. About one-in-seven (15%) respondents who approved of everything had an IP address from outside the U.S. About 7% of always-approving respondents took the survey multiple times, and a sizable share (40%) gave multiple non sequitur answers to the open-ended question. The rates of all these behaviors are significantly higher than among all the study respondents.

"Approving of everything" behavior is strongly associated with bad data quality

% flagged for data quality issue

Data Quality Flag	Respondents who approve of everything	All study respondents
Gave multiple non sequitur answers	40%	2%
Used foreign IP address	15%	1%
Self-reported living outside the U.S.	12%	1%
Took the survey more than once	7%	0%
Unweighted n	1,195	62,639

Notes: Figures are unweighted.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

This always-approve behavior is related to giving unsolicited positive product-type evaluations in the open-ended questions. Among the 413 respondents who answered an open-end with a positive product evaluation-sounding answer, half (50%) answered "approve"/"favorable" all seven times on the closed-ended questions.[15]

While some of these respondents may have been answering honestly, a more plausible explanation is that this pattern represents error. Critically, this error is not mere "noise" but rather has the potential to systematically change the poll results.

[15] Researchers also examined the possibility that some respondents gave uniformly negative answers on those seven questions. That behavior was much less common (1% of all respondents) and did not correlate with other signals of problematic data (e.g., giving non sequitur answers or taking the survey multiple times), so there was not a compelling justification to label respondents bogus based on that pattern.

3. Imperfect metrics of whether respondents live in the U.S.

Some respondents simply answer that they live outside the U.S.

One way to measure how many people taking a poll do not actually live in America is simply to ask. A question in the study did just that, asking respondents whether they currently live "outside the U.S." or "inside the U.S." While informative, such a measure has notable limitations. If a foreign respondent is aware that they do not belong in a U.S. survey, they are unlikely to answer truthfully. Several research teams have documented this scenario of foreign respondents taking steps to conceal themselves in U.S. research using an online crowdsourced platform.

Another possibility is that some people answering "outside the U.S." are Americans temporarily living or traveling abroad. This situation is common, for example, with members of the U.S. military. Such people might vote in U.S. elections and in other respects still be part of "American public opinion." Given that only a small fraction of Americans live abroad, however, we would expect the share accurately reporting living abroad in a secure, representative survey to be very low.

Finally, some answers to this question may reflect measurement error from respondents who are either trolling or answering haphazardly. This form of error stands to bias estimates up while foreign respondents lying about their location would bias estimates down.

While rare in all samples, the share of respondents self-reporting that they currently live outside the U.S. was higher for the opt-in samples than for the address-recruited samples. Among the opt-in sources, the incidence ranged from 1% (crowdsourced sample and opt-in panels 2 and 3) to 2% (opt-in panel 1). There were some respondents giving that answer in the address-recruited panel samples, but their incidence rounds to 0%.

One of the two ABS panels is Pew Research Center's own American Trends Panel, allowing us to examine whether any of those panelists reporting that they live abroad actually do based on the mailing address that we have on file for them. This check showed that none of the panelists had a mailing address outside the U.S. This suggests that for the address-recruited panels the trace-level reports of living abroad probably reflect measurement error (e.g., from satisficing) or panelists who are traveling.

A self-report of living outside the country was strongly associated with other signs of bad response behavior. For example, giving multiple non sequitur open-ended answers was much more common among those saying they live outside the U.S. than inside the U.S. (42% versus 2%, respectively). Also, 12% of the respondents self-reporting that they live outside the U.S. completed the survey using a foreign IP address. By comparison, 1% of the respondents self-reporting that they live inside the U.S. completed the survey using a foreign IP address.

On a related note, some opt-in panel polls included a few respondents appearing to answer open-ended questions in a foreign language. The survey was administered in English and Spanish only. But when asked "How would you say you are feeling today?," an opt-in panel 2 respondent answered in Pashto ("هحختهح نتهخ"), and an opt-in panel 3 respondent answered in Portuguese ("Sim e muito bom"). Both reported that they currently live outside the country. No foreign language (non-English and non-Spanish) responses were detected for the crowdsourced sample or either of the address-recruited panels, though there are many instances of low English proficiency in the crowdsourced sample. Use of these foreign languages is not in itself proof that the respondents do not belong in a U.S. opinion poll, but it is suspicious – particularly alongside a self-report of living outside the U.S.

About one-in-20 crowdsourced respondents have a foreign IP address

IP address is another useful, though imperfect, piece of information about where online poll respondents live. In general, the geolocation of an IP address is a useful indicator of an internet user's approximate location. But differences between the user's location and their IP address location can and do arise from several factors. Those include: where the owner of the IP has it registered, where the agency that controls the IP is located, and proxies. For example, it is common for users on the Verizon network who live in the northern U.S. to show a Canadian IP because that is where the controlling agency of the IP is located. Due to such discrepancies, the geolocation of an IP location alone cannot be considered conclusive evidence about where individuals live.

Among the address-based respondents, 1% had a non-U.S. IP address, which compared to 5% of the crowdsourced respondents.[16] The latter result is almost identical to the non-U.S. rate found by Ahler and colleagues (6%) when examining IP addresses from a crowdsourced sample. In the opt-in survey panels, there were a few respondents with a non-U.S. IP address, but their share of all interviews round to 0%. This suggests that the opt-in panels may have controls in place to guard against this. In fact, several prominent online opt-in survey panels mention in their online information that they use ReleventID, which uses IP geolocation as one of the criteria for identifying fraudulent respondents.

[16] In this study a U.S. IP address refers to those assigned to the 50 states or the District of Columbia. IP addresses associated with U.S. territories are coded as non-U.S. because the territories are rarely if ever included in the target population for national public opinion polls.

Among the foreign IP addresses in the address-recruited panels, Canada and Mexico were the most common host countries. The other foreign IP addresses came from a very dispersed set of countries.

For crowdsourced samples, prior research found that participants with a foreign IP address were particularly likely to come from India or Venezuela. The Center study found a somewhat different pattern. The most common source country for IP addresses outside the U.S. was the Seychelles (125 cases). India was the second most common (99 cases). Then there was a sizable gap before the third most common, Canada (39 cases).

The Seychelles is the modal country of foreign IP addresses used by crowdsourced respondents

Number of respondents with IP address in the country, by sample source

	Crowd-sourced	Opt-in panel 1	Opt-in panel 2	Opt-in panel 3	ABS panel 1	ABS panel 2
United States	10,289	10,001	9,996	11,046	10,074	10,427
Seychelles	125	0	0	0	0	1
India	99	0	0	0	5	3
Canada	39	1	0	0	9	7
Romania	31	0	0	3	5	5
Uruguay	24	0	0	0	0	0
Venezuela	22	0	1	0	1	0
Other country	248	0	3	5	75	74
Unassigned	2	0	0	0	9	9
TOTAL	10,879	10,002	10,000	11,054	10,178	10,526

Note: Figures are unweighted. ABS refers to polls that are recruited offline through residential address-based sampling. Unassigned refers to cases where the IP address was listed as belonging to a private network, the IP address was missing or the IP address had no match in the country look-up.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

The Seychelles result is particularly curious. With a total population of about 95,000, this archipelago off the coast of Africa is more likely to be home to servers or networks masking foreign respondents' location as opposed to the home of the actual participants. Notably, this particular data center is known to be used by software companies whose products are aimed at masking an internet user's identity and location.

While we cannot know for certain where the users of these services are physically located, an earlier study by TurkPrime (2018) found that 89% of participants with IP addresses associated with these kinds of data centers were located in India.[17]

[17] The vegetable known in America as an eggplant is known in India as a brinjal. When shown a picture of this vegetable in the TurkPrime study, 89% of participants operating through a server farm identified it as a brinjal while 96% of those not using a server farm identified it as an eggplant.

The fact that a crowdsourced interview was traced back to a data center does not necessarily imply that the response is bogus as some portion of the U.S. adult population use these kinds of online privacy services legitimately. However, while IP addresses originating from data centers made up 2% of completes among the address-recruited samples, they comprised 8% of the crowdsourced interviews. For all three opt-in panels, addresses originating from data centers made up less than one percent of completed interviews, a pattern suggestive of screening on the part of the opt-in sample providers.

Duplicate IP addresses more common in crowdsourced poll

IP address data can also help detect instances where a given person may have answered the survey more than once. As with foreign geolocation, however, duplicate IP addresses sometimes have a benign explanation (e.g., internet service provider assigned multiple customers the same IP address). So while a duplicate IP is suspicious and may signal a fraudulent interview, it is not dispositive. Some very low-level duplicate rate could be expected just based on benign factors.

In total, 2% of the study interviews came from an IP address that appears in the dataset more than once. The rate among crowdsourced respondents was 5%, which is identical to the rate found by Ahler and colleagues (2019) for crowdsourced respondents. Among the opt-in recruited survey panels in this study, the rate of duplicate IP addresses ranged from 0% (opt-in panel 3) to 3% (opt-in panel 1). In both address-recruited panels, the rate was 1%. Duplicates were more common among foreign IP addresses than domestic ones (12% and 2%, respectively), but most duplicates (92%) were domestic IP addresses.

On balance these data show that the incidence of suspicious IP attributes – the IP address being a duplicate or based in another country – was much greater in an online crowdsourced sample than opt-in recruited or address-recruited panel samples. The opt-in survey panels appear to have controls in place to preclude responses from foreign IP addresses or duplicate IP addresses.

4. Two common checks fail to catch most bogus cases

A number of data quality checks have been developed for online surveys. Examples include flagging respondents who fail an attention check (or trap) question, complete the survey too quickly (speeders), give rounded numeric answers, or give the same or nearly the same answer to each question in a battery of questions (straight-lining). Perhaps the two most common of these are the flags for failing an attention check and for speeding.[18]

A key question is whether these common checks are sufficient for helping pollsters identify and remove bogus respondents before they bias public poll results. This analysis defines a bogus respondent as someone who did any of four things: reported living outside the country, gave multiple non sequitur answers, took the survey multiple times, or always said they approve/favor regardless of what was asked.[19] The rate of bogus respondents was 7% in the crowdsourced poll, 5% on average in the three opt-in panel polls, and 1% on average in the two address-recruited panel polls.

Most bogus respondents pass checks for speeding and attention

% of bogus respondents passing…

Bogus behavior	…Trap question	…Speeding test	…Both
Took survey multiple times	96	97	94
Approve/favor everything	93	93	88
Multiple non sequiturs	81	85	71
Reported living outside U.S.	64	75	52
All bogus respondents	84	87	76

Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

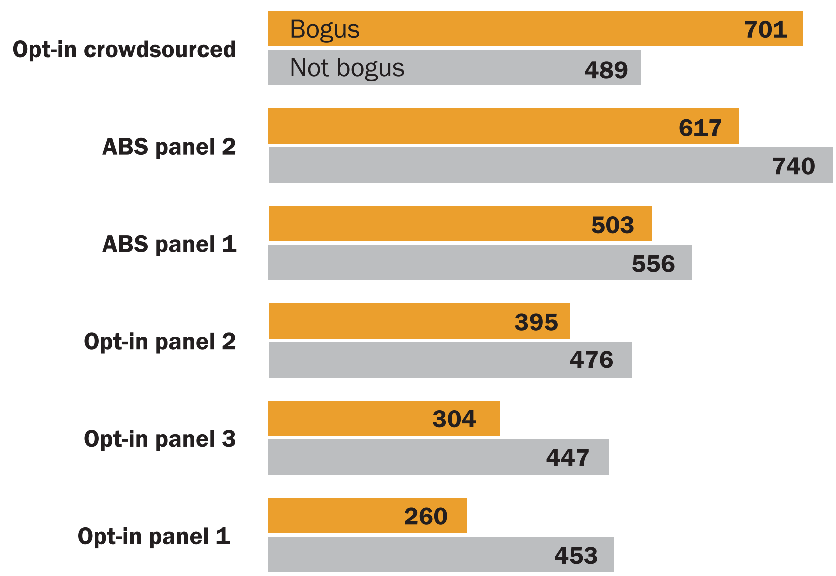
The attention check question in this study read, "Paying attention and reading the instructions carefully is critical. If you are paying attention, please choose Silver below." Overall, 1.4% of the 62,639 respondents in the study failed the attention check by selecting an answer other than "Silver." Among the bogus cases, most of them passed the attention check (84%). In other words, a standard attention check does not work for detecting the large majority of cases found to be giving the type of low quality, biasing data bogus respondents engage in. This result suggests that respondents giving bogus data do not answer at random and without reading the question – the behavior attention checks are designed to catch. Instead, this result corroborates the finding from the open-ended data that some bogus respondents, especially those from the crowdsourcing platform, are often trying very hard to give answers they think will be acceptable.

[18] Some have recommended against attention check questions as they have been found to harm data quality in questions asked later in the survey. That said, attention checks are still fairly common practice among researchers using opt-in sources.
[19] This definition was selected because the behaviors are fairly egregious. Other behaviors (such as claiming to follow a very obscure news story) could conceivably be considered bogus. But to the extent that less egregious behaviors are included in the definition, the risk of mischaracterizing mostly genuine interviews increases.

Results for speeding were similar.[20] Overall, 1.5% of the 62,639 study respondents were flagged for speeding. Speeding was defined as completing the survey in under three minutes when the median response time was seven minutes. Among the bogus cases, about nine-in-ten (87%) were not speeders.[21] This suggests that a check for too-fast interviews is largely ineffective for detecting cases that are either giving bogus answers or should not be in the survey at all. In the crowdsourced sample, the bogus respondents had a longer median completion time than other respondents (701 versus 489 seconds, respectively).

In crowdsourced poll, bogus respondents took over three minutes longer to complete the survey than others

Median time to complete the survey (in seconds)



Notes: All six polls were conducted online. ABS refers to polls that are recruited offline through residential address-based sampling. Figures are unweighted.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

These results are consistent with the findings from other research teams. Both Ahler and colleagues (2019) and TurkPrime (2018) found that fraudulent crowdsourced respondents were unlikely to speed through the questionnaire. Ahler and colleagues found that "potential trolls and potentially fraudulent IP addresses take significantly longer on the survey on average." The TurkPrime study found that crowdsourced workers operating through server farms to hide their true location took nearly twice as long to complete the questionnaire as those not using a server farm. They note that their result is consistent with the idea that respondents using server farms "a)

[20] For five of the six samples, speeding was defined using screen-level response time data. For the crowdsourced sample, however, time spent on each screen was not available and so speeding is defined using the time it took to complete the entire survey, which includes time spent on the introduction and closing screens, as well as questions that were not administered to all samples (see Appendix E). The proportion of the crowdsourced respondents flagged as speeding is, thus, lower than it otherwise would have been if timings at the level of the individual screens were available.

[21] Sensitivity analysis shows that if speeding is defined as answering in under four minutes (instead of under three minutes) the share of all study respondents coded as speeding would increase from 1.5% to 5.6%. Under this more expansive definition of speeding, 75% of bogus respondents would still pass (i.e., not be flagged for speeding).

have a hard time reading and understanding English and so they spend longer on questions" and "b) are taking multiple HITs at once."

Using the union of the two flags is also only partially effective as a means of identifying bogus respondents. About three-quarters (76%) of bogus cases pass both the attention check and the fast response check. Purging based on speeding and a trap question appears to be somewhat more effective for opt-in and address-recruited panels than the crowdsourced sample. On average, those flags removed 29% of the cases identified as bogus in the opt-in and address-recruited panels but just 7% of the bogus cases in the crowdsourced sample. In sum, these two common data quality checks seem to help but appear to be far from sufficient in terms of removing most bogus interviews.

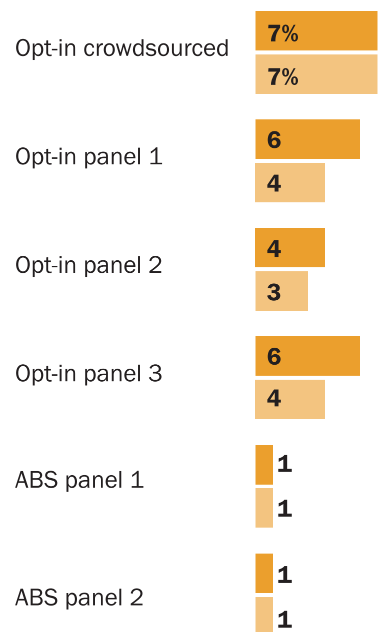
Respondents taking the survey multiple times was rare and limited to opt-in sources

Another possible quality check is to look for instances where two or more respondents have highly similar answers across the board. Similar to looking at duplicate IP addresses, having similar sets of answers could be an indicator of the same person taking the survey more than once.

Whether a pair of interviews having the same answers on a large proportion of closed-ended questions indicates duplication is [exceedingly tricky] to figure out, because various survey features such as the number of questions, the number of response options, the number of respondents, and the homogeneity within the surveyed population affect how natural it is for any two respondents to have very similar answers. However, because the questionnaire in this study also included six open-ended questions, it becomes possible to identify potential duplicate respondents with much higher confidence.

After removing speeders and attention check failures, most bogus cases remain

% of interviews that are identified as bogus before (dark orange) and after (light orange) purging speeders and trap question failures



Notes: ABS refers to polls that are recruited offline through residential address-based sampling. Figures are unweighted.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

For each open-ended question, researchers compared each respondent's answer to all the other respondents' answers using a metric for measuring the similarity between two strings of text.[22] This was done separately for each of the six samples. If, for a particular pair of respondents, three or more of their answers to the six open-ended questions exceeded a certain threshold, that pair was flagged for manual review. A researcher then reviewed each pair to assess whether they were a probable duplicate based on word choice and phrasing across multiple open-ended questions. When similar answers consisted entirely of short, common words (e.g., "good" or "not sure"), researchers did not consider that sufficiently strong evidence of a duplicate, as there is not enough lexical content to make a confident determination.

At the end of this process, researchers found that duplicates represented 0.3% of all interviews. The incidence of duplicates was highest in the crowdsourced sample (1.1%), while in the opt-in samples, the incidence ranged from 0.1 to 0.3%. No duplicate interviews were identified in the address-recruited samples.

Researchers examined whether the having an IP address flagged as a duplicate (as described in Chapter 3) was related to the interview being flagged as a duplicate based on this analysis of open-end answers. While there was a relationship, relying on IP addresses alone to detect people answering the survey multiple times is insufficient. Out of the 172 respondents flagged as duplicates based on their open-ended answers, there were 150 unique IP addresses.

[22] It is also possible that the same respondent might end up in more than one sample and thus take the survey more than once that way, but the computational cost of comparing open-ended responses between samples was judged to be too high.

Open-ended questions helped to identify instances of people taking the survey multiple times

Example of instance in the crowdsourced sample where the same person appears to have taken the survey four times, always using a different IP address

Question	Respondent 1	Respondent 2	Respondent 3	Respondent 4
How would you say you are feeling today?	Am feeling relaxed and more positive thoughts today, and also a excited day for me.	Am feeling relaxed and perfect minded today. Excited day for me	Am feeling relaxed and clean minded and also a excited day for me	Am feeling relaxed and clean minded today, also an excited day for me.
When you were growing up, what was the big city nearest where you lived?	Los Angeles	Los Angeles	Los Angeles	Los Angeles
When you visit a new city, what kinds of activities do you like to do?	I like to found the famous place of that city for visiting those scenarios .	I like to travel through that city and find the famous place of that city for visiting there.	I like to find out the famous place of the city and like to enjoy those beautiful scenario on that day	I like to find out the famous place of that city for visiting those scenarios.
How do you decide when your computer is too old and it's time to purchase a new one?	when computer starts to works slower and often hanging in the desktop leads to purchase a new one	It becomes slower in working and often hanging in the system leads to purchase a new one.	When computer started to work slower than its actual speed and also a often hanging leads to purchase a new one	When the computer starts to work slower than its actual speed and also a often hanging in the desktop leads to purchase me a new one
In retirement what skill would you most like to learn?	Renovating skills	Renovating skills	Renovating skills	Renovating skills
What would you like to see elected leaders in Washington get done during the next few years? Please give as much detail as you can.	Elected leaders must do their duties correctly and they must introduce the new laws for gun owning to prevent the cruelty from that guns and also they should increase the security for the public.	Elected leaders must do their duties correctly and they must re-correct the laws for owning guns to reduce the cruelness with the guns, and also they must increase the security for the people.	Elected leaders must do their duties correctly and they must enact new laws for the gun owning to reduce those cruelties made for the public and also they should increase the security for the people.	Elected leaders must do their duties correctly and enact new laws for gun owning to reduce the cruelty made to the people in this society through the gun and also they must increase the security of all the people in that country.

Notes: The proportion of identical closed-ended answers between the six distinct pairs that can be formed from these four respondents ranges from 78% to 91%, or 25 to 29 out of 32 closed-ended questions.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"
PEW RESEARCH CENTER

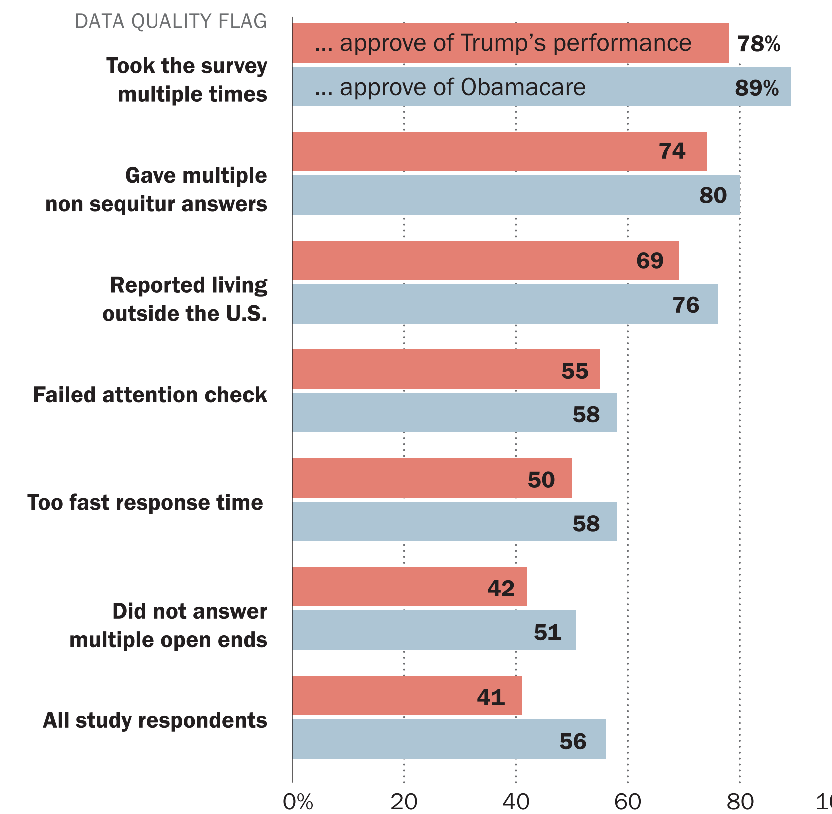
5. Bogus respondents bias poll results, not merely add noise

Respondents who consistently say they approve or favor whatever is asked are not the only ones introducing bias. Those flagged for other suspicious behaviors also answered political questions in ways that differ from other adults. In particular, those saying they currently live outside the U.S. or who give multiple non sequitur answers expressed much higher levels of support for both Donald Trump's job performance and the 2010 health care law (also known as Obamacare), relative to other respondents. Nearly three quarters (74%) of respondents giving multiple non sequiturs said they approve of Trump's job performance, compared with 41% of the study respondents overall. Similarly, 80% of those giving multiple non sequiturs said they approve of the 2010 health care law, compared with 56% of the study respondents overall.

The combination of these two views is relatively rare in the public. According to the address-recruited samples, 12% of those who approve of the president's job performance say they approve of the 2010 health care law. Among those giving multiple non sequiturs, however, 86% of those approving of Trump's job performance approve of the ACA. Given that this subgroup expresses a highly unusual viewpoint and is known to have members providing bogus data, this combination of attitudes should probably not be taken at face value.

Bogus respondents are unusually likely to say they approve of both Obamacare and Trump

% of respondents flagged who say they…



Note: Figures unweighted. "Did not answer" refers to leaving the question blank or giving a don't know or refusal type answer.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

Notably, traditional quality checks tend to flag respondents who express more common political views. Respondents flagged for answering too quickly, not answering questions, or failing an attention check question are not very different from the study participants as a whole on these attitudes. For example, approval of the Affordable Care Act ranges from 51% to 58% among respondents receiving those various flags – similar to the overall approval rating for the ACA in the study (56%).

Based on these findings, it is understandable how prior research teams looking at those traditional flags could have concluded that such respondents were not that different and are perhaps best kept in the survey analysis. But the data quality flags highlighted in this study (e.g., taking the survey multiple times, giving non sequitur answers) tell a very different story. Those flags show suspect respondents giving systematically different answers for key questions. Because the answers are systematic (e.g., largely favorable to the 2010 health care law or to Trump), they stand to move topline survey figures rather than merely adding noise.

To quantify the consequences for poll results, researchers computed estimates for key political questions with and without bogus respondents. This analysis uses the same definition of bogus respondents introduced above (a respondent who reported living outside the country, gave multiple non sequitur answers, took the survey multiple times, or always said they approve/favor regardless of what was asked).

Bogus respondents have small but consistent effect on some opt-in poll figures

Poll estimate when bogus respondents are included versus excluded (unweighted)

	Trump job approval			% Country is on right track			% Favorable of V. Putin		
	All interviews	**Bogus cases excluded**	***Diff.***	**All interviews**	**Bogus cases excluded**	***Diff.***	**All interviews**	**Bogus cases excluded**	***Diff.***
Opt-in crowdsourced	35	31	*-4*	37	33	*-4*	16	11	*-5*
Opt-in panel 3	42	40	*-2*	38	36	*-2*	15	12	*-3*
Opt-in panel 2	42	41	*-1*	37	35	*-2*	13	10	*-3*
Opt-in panel 1	43	41	*-2*	39	37	*-2*	13	12	*-1*
ABS panel 1	43	43	*0*	30	30	*0*	10	10	*0*
ABS panel 2	39	39	*0*	31	31	*0*	9	9	*0*

Note: Figures unweighted. A bogus respondent is defined as someone who reported living outside the country, gave multiple non sequitur answers, reported using a non-existent website, or always said they approve/favor regardless of what was asked. ABS refers to polls that are recruited offline through residential address-based sampling.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

As noted above, the rate of bogus respondents was 7% in the crowdsourced poll, 5% on average in the three opt-in panel polls, and 1% on average in the two address-recruited panel polls. With so few address-recruited panelists affected, it is not surprising that removing them has essentially no effect on estimates. For the opt-in polls, by contrast, the rate of bogus respondents is high enough for them to have a measurable, if small, impact. In opt-in panel 3, for example, the president's job approval rating drops two percentage points (from 42% to 40%) when bogus cases are excluded. In the crowdsourced sample, Trump's job approval drops by four percentage points when bogus cases are removed (35% to 31%).[23] Similarly, Vladimir Putin's favorability rating drops by three percentage points in two of the opt-in panel polls when bogus cases are removed.

Not all survey estimates, however, are affected by bogus respondents. For example, estimates of the political party people trust more on the economy do not change at all for two of the opt-in panel polls when bogus respondents are dropped. Similarly, there is no change in some of the opt-in estimates for the share saying that protecting the right to own guns is a higher priority than enacting new laws to try to reduce gun violence, when bogus cases are dropped.

On other estimates there is no discernable, consistent effect from bogus cases

Change in poll estimates when bogus respondents are excluded (unweighted)

	% Trust Republicans in Congress (over Dems) on the economy			*% Use Facebook at least once a month*			*%Higher priority is protecting gun rights*		
	All interviews	**Bogus cases excluded**	***Diff.***	**All interviews**	**Bogus cases excluded**	***Diff.***	**All interviews**	**Bogus cases excluded**	***Diff.***
Opt-in crowdsourced	45	44	*-1*	81	81	*0*	30	31	*+1*
Opt-in panel 3	49	49	*0*	80	81	*+1*	37	37	*0*
Opt-in panel 2	51	50	*-1*	80	80	*0*	35	35	*0*
Opt-in panel 1	50	50	*0*	77	78	*+1*	40	39	*-1*
ABS panel 1	50	50	*0*	68	68	*0*	36	36	*0*
ABS panel 2	47	47	*0*	71	71	*0*	34	34	*0*

Note: Figures unweighted. A bogus respondent is defined as someone who did any of four things: reported living outside the country, gave multiple non sequitur answers, reported using a non-existent website, or always said they approve/favor regardless of what was asked. ABS refers to polls that are recruited offline through residential address-based sampling.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

[23] Virtually all national polls fielded around the time of this study showed the president's approval rating closer to 40% than 30%. At first glance it may seem then that bogus respondents were making the crowdsourced poll more accurate (i.e., closer to 40%). That conclusion is not correct, however, because these estimates are not weighted. Crowdsourced samples are well documented to have a young, educated, liberal bias. The fact that Trump's unweighted approval rating based on the crowdsourced sample is in the mid to low 30s reflects the inherent biases of the platform and unweighted nature of the analysis.

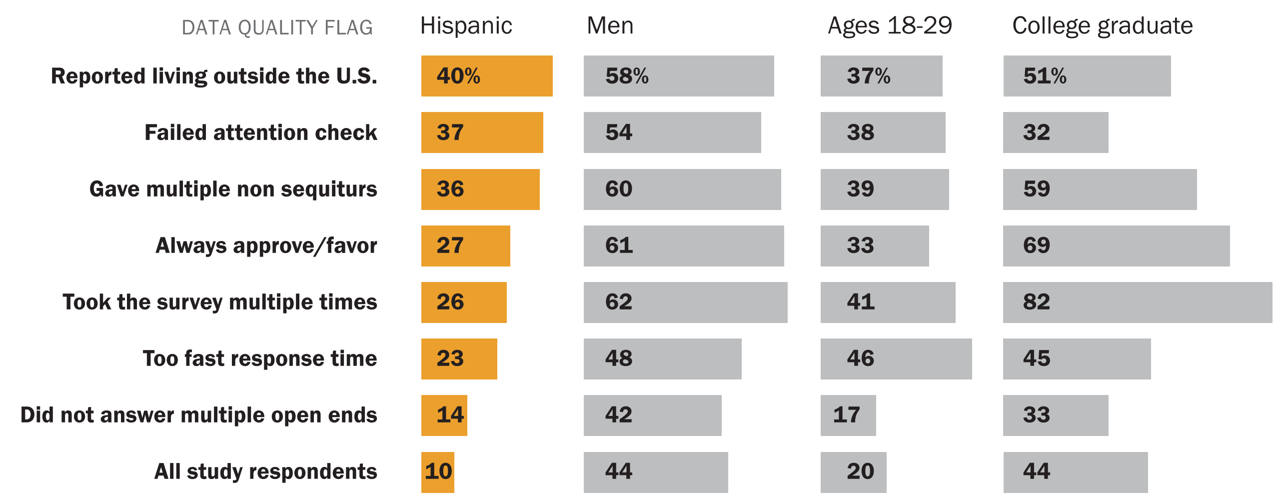
One characteristic stands out when looking at which questions are influenced by bogus cases versus which ones are not. Questions that allow the respondent to give a positively valanced answer appear to be most affected. For example, the question “Would you say things in this country today are generally headed in the right direction (or) off on the wrong track?” allows the respondent to say something is going right rather than wrong. As discussed above, respondents giving bogus data are very prone to giving positive answers. By contrast, the question battery “Who do you trust more to handle each of the following issues… Democrats in Congress or Republicans in Congress?” was basically unaffected by bogus respondents. The choice of Democrats versus Republicans apparently does not map onto this behavior of giving uniformly positive answers. Put simply, the bias from bogus data documented in this study is politically agnostic – neither pro-Republican nor pro-Democrat.

6. Cases tripping flags for bogus data disproportionately say they are Hispanic

Cases flagged for suspicious survey behavior have a very different demographic profile than nonsuspicious cases, based on respondents' self-reported characteristics. In particular, respondents flagged for certain suspicious behaviors were quite likely to say they are Hispanic. The baseline rate of respondents in the study self-reporting as Hispanic is 10% (the actual population rate[24] is 16%). However, 30% of those giving at least two non sequitur answers, taking the survey multiple times, always saying they approve/favor regardless of what was asked, or saying they currently live outside the U.S. said they were Hispanic.

Respondents flagged for suspicious survey behavior often say they are Hispanic

% of cases flagged who say they are …



Notes: Figures are unweighted. "Did not answer" refers to leaving the question blank or giving a don't know or refusal type answer.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

While some share of these suspicious respondents could very well be Hispanic, this rate is likely inflated. Hispanic ethnicity was measured with a stand-alone yes/no question. Therefore, respondents answering at random would be expected to report "yes" at a higher rate than the true incidence.

[24] Figure based on the 2018 American Community Survey.

Relative to the study respondents (and the population) as a whole, cases flagged for suspicious behavior skew male, young and educated. There is no obvious explanation why respondents would misreport those characteristics, but it is difficult to know if they should be taken at face value.

Respondents who decline to answer two or more of the open-ended questions exhibit none of those skews and in general look quite like the overall sample. This result underscores the fact that declining to answer some questions is very different from these other behaviors.

Bogus cases have a particularly large effect on estimates for Hispanic Americans

The fact that bogus cases are disproportionately likely to report being Hispanic means that the damage from bogus cases is particularly large for Hispanic estimates. For example, when looking at the opt-in panel polls, Trump's job approval among non-Hispanic whites changes by one percentage point, on average, when bogus cases are removed. By contrast, his approval rating among Hispanics drops five percentage points on average when bogus cases are removed. The damage to the Hispanic estimate from the crowdsourced sample is far worse – an 18 percentage point change. Among the address-recruited online panel polls, none of the subgroup estimates changed by more than a percentage point.

Bogus interviews are particularly damaging for Hispanic survey estimates

% of approving of Trump's job performance when bogus respondents are included versus excluded (unweighted)

	Trump approval among Hispanics			*Trump approval among non-Hispanic blacks*			*Trump approval among non-Hispanic whites*		
	All interviews	**Bogus cases excluded**	***Diff.***	**All interviews**	**Bogus cases excluded**	***Diff.***	**All interviews**	**Bogus cases excluded**	***Diff.***
Opt-in crowdsourced	44	26	*-18*	21	17	*-4*	37	34	*-3*
Opt-in panel 3	33	27	*-6*	21	17	*-4*	47	46	*-1*
Opt-in panel 2	28	25	*-3*	13	11	*-2*	48	46	*-2*
Opt-in panel 1	36	29	*-7*	16	12	*-4*	49	49	*0*
ABS panel 1	30	29	*-1*	10	9	*-1*	50	50	*0*
ABS panel 2	26	26	*0*	12	11	*-1*	45	45	*0*

Note: Figures unweighted. A bogus respondent is defined as someone who reported living outside the country, gave multiple non sequitur answers, reported using a non-existent website, or always said they approve/favor regardless of what was asked. ABS refers to polls that are recruited offline through residential address-based sampling.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

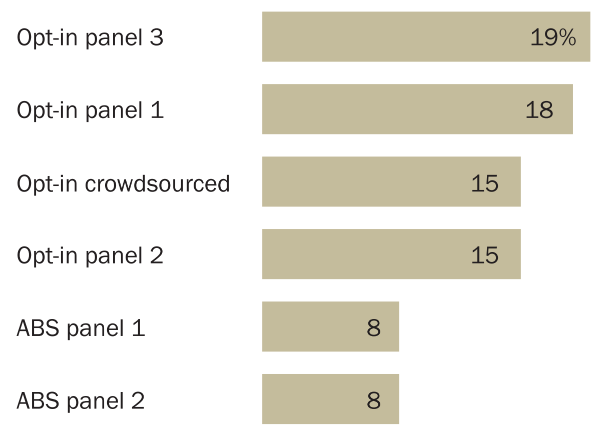
7. Other tests for attentiveness show mixed results

The study included other data quality checks based on those used in prior studies of response quality. Several of these tests found further evidence that opt-in respondents are more prone to giving low-quality answers than address-recruited respondents. Other tests found no meaningful differences across the sources.

One test developed decades ago involves asking respondents about a fictitious behavior. Reports of doing such a behavior means that the respondent gave an erroneous answer. This study operationalized that idea by asking respondents to select which of five websites they use at least once a month. The list included three real sites (YouTube, Instagram, Facebook) and two fictitious ones (FizzyPress and Doromojo). The share selecting at least one of the made-up websites was 2% in crowdsourced sample and ranged from 1% to 2% in the opt-in survey panel samples. There were some such cases in the address-recruited panels, but in both samples they rounded to 0%.

Opt-in respondents about twice as likely as address-recruited ones to report following an obscure news story

% saying they have been following an obscure news story "very" or "fairly" closely



Notes: All six polls were conducted online. ABS refers to polls that are recruited offline through residential address-based sampling. Figures are unweighted.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

A similar check tested the credibility of answers to questions about following the news. Polls are often used to measure the extent to which the public is paying attention to various news stories. The survey asked, "How closely, if at all, have you been following news about China's revised criminal procedure law?" and provided a four-point scale ranging from "Very closely" to "Not at all closely." This is a real news story, but it received virtually no coverage from American news outlets. [25] The revision became law about five months before the survey (Oct. 26, 2018) and does not appear to have been reported by any national U.S. news broadcast or newspaper. It was covered in the English language by Reuters, China's state-run press agency (Xinhua) and the Library of Congress's Global

[25] To avoid respondents feeling as though this question was posed without context, it was preceded by a question about attention to another, more high profile, international news story – the armed conflict in Syria. All four answers choices to the Syria question are plausible (i.e., following "very," "fairly," "not too," or "not at all" closely), and so answers to that question are not of interest in this study.

Legal Monitor. Only a very small proportion of U.S. adults can therefore credibly claim to have been following this news closely. Asking about an obscure news story was preferable to asking about a fictitious one in light of discussions around fake news in today's media environment.

Between 15% and 19% of respondents in the opt-in recruited polls said they have been following news about this highly obscure story either "very" or "fairly" closely. Among the address-recruited online polls, the rate was about half that (8%).

Unlikely answers to other questions were very rare in all samples

The study offered respondents several other opportunities to report suspicious answers. In general, these other questions turned up very little.

Several checks for unlikely answers show no meaningful differences by method

% who say they …

	… Live in Antarctica	… Are under age 18	… Are over age 110	… Voted in 2018 midterm and currently are not registered
Opt-in crowdsourced	0%	0%	0%	0%
Opt-in panel 1	0%	1%	0%	1%
Opt-in panel 2	0%	0%	0%	0%
Opt-in panel 3	0%	0%	0%	1%
ABS panel 1	0%	n/a	n/a	0%
ABS panel 2	0%	0%	0%	0%

Note: Figures unweighted. Age could not be measured on ABS panel 1 due to restrictions on re-asking common demographics. Living in Antarctica was measured in a question asked only of those who said they currently live outside the U.S. ABS refers to polls that are recruited offline through residential address-based sampling.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

The respondent's age was asked[26] without any value restrictions, but very few respondents gave an out-of-range answer. The share of all study respondents reporting an age under 18 and the share giving an age over 110 both round to 0%. The study also asked whether respondents are currently registered to vote and whether they voted in the 2018 midterm election. The share of respondents reporting that they did vote in the 2018 midterms but are not currently registered to vote should be very low, considering that the election was held just four or five months before the survey

[26] Age could not be measured in one of the address-recruited panels due to restrictions on re-asking common demographics questions.

depending on the sample. The incidence of that response pattern was very low (1% or less) in all sources.

As discussed above, another question asked whether respondents currently live outside or in the U.S. Those who reported living outside the country were administered a follow up question asking in which region of the world they live. Antarctica was listed as the first answer choice, but very few respondents selected it. Less than one half of one percent of all respondents in the study reported living outside the U.S. *and* living in Antarctica.

Little difference across sources in tendency to select answers listed first

For several questions, a random half of respondents from each source were shown the answer options in one order while the other half were shown the reverse order. For example, when asked "During the last twelve months, how often did you talk with any of your neighbors?," half the respondents were shown options ordered from highest frequency to lowest ("Basically every day," "A few times a week," "A few times a month," "Once a month," "Less than once a month," and "Not at all"), while the other half were shown the same list ordered from lowest frequency to highest. If respondents are answering carefully, they should give the same answer regardless of which order they received. However, this is not always the case. Because most people read from top to bottom, they tend to think about and answer with answer choices near the top slightly more than answers near the bottom. This is called a primacy effect.

Any online survey may be susceptible to primacy effects, particularly for questions where the answer may be difficult to recall precisely. In this study, opt-in recruited panelists were no more likely to show primacy effects compared to address-recruited respondents. The primacy effects themselves were meager, averaging less than one percentage point across all study respondents.

Researchers also tested for primacy effects in answers to a question asking about the respondent's state of residence using a drop-down box. States were listed in alphabetical order, raising the possibility that careless respondents would over-report living in states like Alabama, Alaska, Arizona or Arkansas. Researchers compared the share of respondents reporting that they live in each state to the actual share of U.S. adults living in those states (according to the American Community Survey). There was no evidence in any of the samples that respondents were carelessly selecting a state listed at the top of the list.

Sources show similar levels of attention to question wording in exclusion statement test

The study included a between subjects experiment to test attentiveness to question wording. A random half of respondents within each sample was asked, "How many hours do you usually spend online each week?" The other half was asked the same question with an instruction at the end: "How many hours do you usually spend online each week? Do NOT include time spent checking email." On average, people receiving the version with the instruction to exclude email should report fewer hours than those receiving the shorter version. The difference between the average number of hours given by those shown the short version and the average number of hours given by people shown the long version (with the exclusion instruction) is an aggregate measure of attentiveness to the question wording. In this case, larger differences between those averages are better, as that suggests more attentiveness.

Wording experiment suggests that the likelihood of respondents reading carefully is similar across samples

Average number of hours respondents report spending online per week

	Instructed to exclude email?		
	No	Yes	Difference
Opt-in crowd-sourced	31	27	4
Opt-in panel 1	23	21	2
Opt-in panel 2	24	21	3
Opt-in panel 3	25	22	3
ABS panel 1	18	15	3
ABS panel 2	20	16	4

Note: 151 respondents in the crowdsourced sample are excluded from this analysis because they gave a text (not numeric) answer. ABS refers to polls that are recruited offline through residential address-based sampling. Figures are unweighted.
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

The difference between the means in two versions were similar across the sources. There was a statistically significant difference between one of the address-recruited panels and one of the opt-in panels, but this did not translate to a consistent difference between address-recruited and opt-in panels as a whole.

8. Results from a follow-up data collection

In analyzing the data, researchers identified two issues that had the potential to affect the study's conclusions. First, the survey was designed to be administered the same way for each of the six online sources. But after interviewing was completed, researchers discovered that there was a discrepancy with respect to whether respondents were allowed to skip questions. Respondents in the two address-recruited and the one crowdsourced sample were not required to answer each question, but those in the opt-in samples were. This presented a potential problem, as forcing respondents to answer each question could conceivably affect their behavior and, in particular, their likelihood of giving answers that flagged them as a bogus respondent. Researchers needed to know if the higher incidence of bogus respondents in the opt-in samples was attributable to this difference. To find the answer, it was necessary to field the survey again on the opt-in sources, this time without forcing respondents to answer each question.

The second issue concerned the approve-of-everything response pattern. As discussed in Chapter 2, a small share of respondents answered with "approve" or "favorable" each time such a question was asked. This behavior was concentrated in the opt-in samples. As this report explains, the most likely explanation is that opt-in polls are primarily used for market research, and so offering rote "approve" answers is logical on the assumption that such answers would please the sponsor. This is a key finding because it demonstrates that bogus respondents, rather than just adding noise, stand to bias certain estimates.

An alternative explanation for the approve-of-everything response style is what is known in polling as a primacy effect. A primacy effect is the tendency for some respondents to select answers shown at or near the top of the answer list. For example, in the question asking about the President's job performance the first answer choice was "Strongly approve" and the last was "Strongly disapprove." Conceivably, the approve-of-everything respondents could have simply been selecting answers near the top, which in this study happen to be positively-valanced. To test this, it was necessary to field the survey again, this time presenting the negative answer choices first. If the approve-of-everything behavior was observed, even when such answers were shown near the bottom, this would show that the behavior is purposeful and that rotating the answer choices does not help.

Researchers addressed both potential concerns by fielding a follow-up data collection. The survey was fielded again from Dec. 2 – 7, 2019 with 10,122 interviews from opt-in panel 1 and 10,165 interviews from opt-in panel 3. Respondents to the first survey were ineligible for the follow-up study. Opt-in panel 2 was not used because it was not needed to answer the two questions raised above. The rates of bogus responding and approve-of-everything response style were similar across the three opt-in sources. If we learned that permitting respondents to skip questions or rotating the approve/disapprove options increased data quality in panels 1 and 3, it would be very reasonable to assume that that would also hold for panel 2. All three opt-in panels generally performed about the same.

The important difference between the main study and the follow-up study was two-fold. First, respondents were allowed to skip questions. Second, a split-form experiment was administered. A random 50% of respondents received the same response ordering as the main study with positive (approve/favorable) answers shown first, and the other random 50% of respondents received the reverse ordering with negative (disapprove/unfavorable) answers shown first. The follow-up study asked the same questions as the main study, with two minor exceptions. Because a new British Prime Minister took office between the first and second data collection, the name was updated in the question (Theresa May to Boris Johnson). Also, a language preference question was added to better assign English versus Spanish.

If the approve-of-everything behavior was merely a primacy effect (not purposeful), the follow-up study would have found a lower rate of the behavior when negative answers were shown first, as opposed to second. But that did not happen.

The incidence of respondents giving uniformly "approve"/"favorable" answers was basically the same regardless of the ordering of the answer choice. In opt-in panel 3, 3% of respondents approved of everything when positive answers were shown first, and the same amount did this when negative answers were shown first. The pattern was the same for opt-in panel 1, though with both rates being lower.

Approve-of-everything responding is not simply a primacy effect

% answering approve/favor for all seven questions

	Opt-in panel 1	Opt-in panel 3
Main study		
Positive answers shown first	1%	3%
Follow-up study		
Positive answers shown first	1%	3%
Negative answers shown first	1%	3%

Notes: Figures are unweighted.
Source: Main study conducted March 13-22, 2019. Follow-up study conducted December 2-7, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

This result indicates that the small but measurable share of opt-in respondents who apparently approve no matter what is asked about do so intentionally. They sought out the positive answers even when they had to look for them. They were not lazily selecting the first answer shown. This suggests that randomizing the response options would not eliminate this source of apparent bias. Interestingly, the overall incidence of this behavior was the same in the follow-up study as it was in the main study. This bolsters confidence in the generalizability of the main study findings.

There are several other data points worth noting that discredit the notion that the approve-of-everything pattern is merely a primacy effect. One might expect that those answering approve/favor regardless of the question are always selecting the *first* answer choice. For example, on a four-point scale (e.g., "very favorable," "mostly favorable," "mostly unfavorable," and "very unfavorable"), perhaps the always approving cases consistently select the most positive answer available ("very favorable"). That is not the case. For example, when the main study asked for an overall opinion of Vladimir Putin, 45% of the always approving respondents say "very favorable" while 55% say "mostly favorable." Most approve-of-everything respondents selected the second choice, not the first. The same pattern was observed for the questions asking about Merkel, Macron and May.

In addition, if approve-of-everything respondents were simply picking answers near the top of every question, most would have answered the attention check (or trap question) incorrectly. In fact, 93% of the always approve cases answered this attention check correctly in the main study, and a nearly identical 94% of the always approve cases did so in the follow-up. In sum, a good deal of randomized and non-randomized data indicated that the approve-of-everything behavior is largely purposeful. It may be exacerbated when positive choices are offered first, but the follow-up study showed that even when positive choices are not offered first this small segment of opt-in respondents will seek them out.

The follow-up study also tested whether allowing opt-in respondents to skip questions would reduce the bogus incidence. Researchers created a flag for bogus cases in the follow-up study using the same definition as the main study. In one opt-in panel, the bogus rate was *higher* when respondents could skip, while in the other panel it was lower. For opt-in panel 3, the incidence of bogus cases was 6% in the main study that prohibited skipping for opt-in respondents, and it was 8% in the follow-up study that allowed respondents to skip. For opt-in panel 1, the incidence of bogus cases was 6% in the main study that prohibited skipping for opt-in respondents, and it was 3% in the follow-up study that allowed respondents to skip. In neither case was the rate of bogus respondents as low as it was for the address-recruited panels (1%).

Opt-in polls still have higher rates of bogus data when respondents can skip items

% of respondents flagged for poor data quality

	Opt-in panel 1		Opt-in panel 3		ABS panel 1	ABS panel 2
	Main (no skipping)	Follow-up (skipping OK)	Main (no skipping)	Follow-up (skipping OK)	Main (skipping OK)	Main (skipping OK)
Flagged as bogus	**6%**	**3%**	**6%**	**8%**	**1%**	**1%**
Gave multiple non sequitur answers	4%	2%	3%	5%	0%	0%
Approve of everything	1%	1%	3%	3%	0%	0%
Self-report living outside the U.S.	2%	0%	1%	2%	0%	0%
Took survey multiple times	0%	0%	0%	0%	0%	0%

Notes: Figures are unweighted. ABS refers to polls that are recruited offline through residential address-based sampling. A respondent was flagged as bogus if they took the survey multiple times, reported living outside the U.S., gave multiple non sequitur open-ended answers, or always said approve/favorable regardless of what was asked.
Source: Main study conducted March 13-22, 2019. Follow-up study conducted December 2-7, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

In general, the follow-up study sample from opt-in panel 1 showed better data quality than the main study sample. The incidence of non sequitur open-ends and self-reports of living outside the U.S. were lower in the follow-up. In opt-in panel 3, by contrast, the follow-up sample had poorer data quality than the main study sample. The incidence of non sequitur answers and self-reports of living outside the U.S. were both higher in the follow-up. Interestingly, while none of the opt-in panel 3 respondents plagiarized an open-ended answer in the main study, 15 respondents from that panel did so in the follow-up study (see Appendix D). They pulled from several of the sources tracked in the main study, including websites for Mount Vernon and the Washington State Legislature, as well as a website helping non-English speakers answer "How are you feeling today?"

If allowing opt-in respondents to skip questions was the key to achieving good data quality then we would have seen the bogus rates in both opt-in panels decline in the follow-up study, perhaps to the low level observed for the address-recruited samples. But that is not what happened. Opt-in panel 1 did perform better when answering was not required, but the incidence of bogus cases was still significantly higher than the levels observed in the address-recruited samples. Meanwhile, opt-in panel 3 got worse, with the incidence of bogus cases climbing to a striking 8% in the follow-up.

Given that one opt-in panel did worse when skipping was allowed but another panel did better, it is not clear that requiring respondents to answer questions has a strong, systematic effect on the incidence of bogus cases. It is worth noting that opt-in panels 1 and 3 source respondents from many of the same third party companies. In this study alone, sources used by both panels 1 and 3 include CashCrate, A&K International, DISQO, Market Cube, MySoapBox, Persona.ly, Tellwut and TheoremReach. The variance in data quality may have more to do with the relative shares of respondents coming from such sources than it necessarily does with the forced response setting. This is a topic worthy of future investigation.

Notably, all of the key findings from the main study were replicated in the follow up. For example, most bogus respondents (76%) in the main study passed both an attention check and a check for speeding. The share of bogus cases passing those same two checks in the follow-up was similar (70%). Similarly, a suspiciously high share of bogus cases in the main study reported being Hispanic (30%). In the follow-up this rate was 31%. The follow up study also replicated the finding that bogus respondents can have a small systematic effect on approval-type questions. For example, the estimated share expressing a favorable view of Vladimir Putin dropped four percentage points (from 20% to 16%) in the follow up-study when bogus respondents were removed from the opt-in panel 3 sample, and this estimate dropped one percentage point when bogus respondents were removed from the opt-in panel 1 sample (from 14% to 13%).

9. Conclusions

While the growth of online interviewing is a prominent trend in polling, there is variation within that trend in how researchers recruit respondents. This study evaluated three respondent sourcing approaches (using workers from a crowdsourcing website, opt-in survey panels, and address-recruited survey panels) and found that sourcing affects data quality. Specifically, crowdsourced and opt-in survey panel respondents were more likely than those recruited via random sampling of addresses to give bogus data. Bogus data came in several forms including duplicate interviews, answers that had no bearing on the question, answers that were uniformly positive regardless of what was asked and interviews filled out by people who are probably not Americans. Two common data quality checks (one for speeding and another for attention) failed to detect most respondents flagged as bogus.

The study results raise concerns about how secure some public opinion polls are from fraudulent interviews. Consistent with other research, the data from this study suggests that fraudulent respondents in the crowdsourced sample are often foreign residents posing as Americans. For opt-in survey panels, there is little if any evidence of that, and it appears that widely used opt-in panels manage to keep out internet users with foreign IP addresses. With the opt-in survey panels and the crowdsourced sample, however, the study found a small but measurable segment of respondents who seem to operate on the assumption that it is a market research survey and therefore give pleasing (not genuine) answers.

Researchers using crowdsourced marketplaces for social science experiments say it is best practice to restrict participation to workers with a task completion or approval rate of at least 95%. While that may be sound advice for conducting randomized experiments with crowdsourced subjects, it is a dubious constraint when the goal is to obtain a representative sample of Americans for purposes of estimating public opinion. This is tantamount to a pollster paying one company (albeit a very decentralized one) to interview some of their better-performing employees and then describing the results as information about American public opinion. Such a process can be expected to systematically exclude those experiencing hardships affecting their work or those with lower cognitive abilities, not to mention the 99% of the public that does not work on the platform.

The data in this study were collected on the premise that the panels/marketplaces selling online interviews are responsible for ensuring data quality and that additional checking by the researcher is unnecessary. Given that many public polls are described simply as being “conducted online,” it seems likely that at least some researchers operate on this assumption. Other researchers using opt-in data presumably have their own checks in place to try to address the issues raised in this report. To help the public better differentiate trustworthy and untrustworthy polls, it would be helpful if poll methodology statements mentioned what checking, if any, was performed.

Acknowledgements

This report was made possible by The Pew Charitable Trusts. Pew Research Center is a subsidiary of The Pew Charitable Trusts, its primary funder.

This report is a collaborative effort based on the input and analysis of the following individuals:

Research team

Courtney Kennedy, *Director, Survey Research*
Nick Hatley, *Research Analyst*
Scott Keeter, *Senior Survey Advisor*
Andrew Mercer, *Senior Research Methodologist*
Arnold Lau, *Research Analyst*
Joshua Ferno, *Assistant Panel Manager*
Dorene Asare-Marfo, *Survey Methodologist*
Skye Toor, *former Data Science Assistant*
Patrick van Kessel, *Senior Data Scientist*
Adam Hughes, *Associate Director, Data Labs*

Communications and editorial

Rachel Weisel, *Senior Communications Manager*
Hannah Klein, *Communications Manager*
Calvin Jordan, *Communications Associate*
Travis Mitchell, *Copy Editor*

Graphic design and web publishing

Bill Webster, *Information Graphics Designer*
Travis Mitchell, *Digital Producer*

Colleagues both within and outside of Pew Research Center contributed to the development and execution of this study. In particular, we appreciate insights from Jake Soffronoff about approaches for detecting suspicious cases.

Appendix A: Survey methodology

This report included six online platforms used for public polling. Three are opt-in survey panels or "marketplaces." One is an opt-in crowdsourcing labor website. Two are survey panels that interview online but recruit offline, one of which is Pew Research Center's American Trends Panel. For both panels recruited offline, most panelists were recruited using address-based sampling (ABS). Before using ABS, both panels recruited offline using random samples of telephone numbers (random digit dialing).

The Center's American Trends Panel sample was interviewed using normal procedures. The crowdsourced sample was fielded in-house at the Center using a prominent crowdsourcing labor market website. The other four samples were conducted for the Center by a survey data collection firm that served as a coordinating vendor. The Center contracted with the vendor to conduct a "national polling study." The research aims of the study were not discussed with the coordinating vendor. The questionnaire was the only study document provided to the coordinating vendor and the panels to which they subcontracted.

The vendor was instructed to use formatting, style, and respondent sourcing that is normal for political polls conducted on each of the panels. In theory, Center researchers could have required a set of elaborate quotas for each opt-in source. That was purposefully avoided because it would have damaged the generalizability of the results. Had we required our own custom set of quotas, the study results would only be generalizable to opt-in surveys sampled the way Pew Research Center would do it. But that was not the study goal. Instead, the inferential goal was to field a national public opinion poll from each source using the specifications used by the panel provider for such polls. While some researchers using opt-in sources do use elaborate quotas, many do not. This study is designed to speak to the quality of data public pollsters received when they rely on the opt-in provider to ensure that the sample is sound.

The coordinating vendor was instructed to conduct at least 10,000 interviews with U.S. adults age 18 and older in all 50 states and the District of Columbia from each source. English and Spanish administration was available and used for four of the six sources. For the crowdsourced sample and opt-in survey panel 2, it was only feasible to conduct interviews in English.

The crowdsourced portion of the study was fielded in 11 waves beginning March 19 and ending April 4, 2019. The average wave size was approximately 1,250 respondents. The survey was available to any U.S. adult on the crowdsourced platform. No other filters available through the crowdsourced platform were used. Crowdsourced respondents were paid $2.50 to complete the study.

Source	Sampled	Completes	Break-offs	Field dates
Crowdsourced	11,009	10,879	130	March 19-April 4, 2019
Opt-in panel 1	25,527	10,002	757	March 13-18, 2019
Opt-in panel 2	23,773	10,000	687	March 14-19, 2019
Opt-in panel 3	34,927	11,054	759	March 13-21, 2019
ABS panel 1	16,652	10,178	181	March 13-22, 2019
ABS panel 2	13,482	10,526	74	April 1-15, 2019

Notes: A break-off was defined as a case in which the respondent answered the first two survey questions but abandoned the interview at some point after that. For the crowdsourced portion of this study, “sampled” refers to the number of unique respondents that accepted and began the task listing from the crowdsource website.

Address-recruited panel 1 had a study-specific response rate of 61%. The cumulative response rate to the survey (accounting for nonresponse to recruitment, to the current survey and for panel attrition) was 5%.

Address-recruited panel 2 had a study-specific response rate of 78%. The cumulative response rate to the survey (accounting for nonresponse to recruitment, to the current survey and for panel attrition) was 4%.

Quotas used for opt-in panel samples

The following quotas were specified and used by the opt-in survey panel vendors to produce their national samples of U.S. adults.

Opt-in Panel 1: Internal quotas	
Male	49%
Female	51%
18-24	11%
25-34	22%
35-44	20%
45-54	19%
55-64	15%
65-99	13%

Opt-in Panel 2: Internal quotas

Overall	10,050
Male 18-24	625
Male 25-34	920
Male 35-44	814
Male 45-54	934
Male 55-64	810
Male 65+	902
Female 18-24	597
Female 25-34	890
Female 35-44	815
Female 45-54	855
Female 55-64	868
Female 65+	1,122

Opt-in Panel 3: Internal quotas

Male	50%
Female	50%
18-24	13%
24-44	41%
45-65	30%
65+	16%
Hispanic	11%
Not Hispanic	89%
Black	12%
White	70%
Other race/ethnicity	18%
Midwest	22%
Northeast	18%
South	37%
West	23%

Opt-in sample third party sources

With the address-recruited panels, respondents are recruited by sampling residential addresses from the U.S. Postal Service Computerized Delivery Sequence File, and in years prior, by sampling telephone numbers using random digit dial. With the crowdsourced sample poll, the sample consists of workers on the crowdsourcing site. Two of the three opt-in survey panel vendors used third party companies, listed below (combined and de-deduplicated to avoid panel identification). One of the opt-in survey panel vendors reported using no third party companies.

.Jobs - Surveys	e-RG/Elite - USA - English	Opinion Share	Triaba US
A Triaba USA (Spanish)	freeup_site	opinionnow.com - USA	Tribe Surveys U.S.
A&K International - USA	FS Surveys	P2Sample	Union Street Enterprises, Inc.
AdGate Rewards	FusionCash	Paid Research Poll	Usability Testing Panel - USA
Adscend US	Giftizma	Panel Champ - USA	USOpinionPoll
A-K International	Gilhaus Research	PanelOptimus USA	ViewPoint Panel (USA)
ANQPANEL - USA	GRL - API - All	Persona.ly - USA	Vindale Research
ArcaMax Research	iAngelic	PINCHme USA	Vivatic US
AskPolonia USA	iGain	PI-Opinion	WRM USA
Attapoll	iGlobal Surveys	PollBuzzer PLUS USA	YourSay USA
BAP USA	Immersive Camp	Pollfish1	Z - Test Supplier
Billaway USA	InboxDollars	Prodege	Zonaencuestas USA
Bitburst	InboxResearch	Qmee - USA	
Bizrate Rewards	inMarket	QuickRewards.net	
Bovitz, Inc.	Innovate	Reel Change	
Branded Research	InstaGC - US	Revenue Universe	
Branded Surveys	Insticator - USA	RevenueHut - USA	
Cashbackearners.com	iOpenUSA	Rewardia.com.au	
CashCrate	iRazoo	Samples Avenue	
Centiment	Loop Surveys - USA (English)	Saybucks - United States	
CFS Panel USA	Survey Pronto USA	Snap Surveys USA	
CinchDollarsUSA	Surveyeah US	Springboard America	
ClixSense	Market Cube USA	SurveyEveryOne - USA	
Crowdology USA	Marketagent - USA	Surveys4Rewards - USA	
Dale US	MindField Online	SurveyTouch - USA	
Dalia Research - US	MUSICLOVERS Panel USA	Swagbucks - USA	
DataDiggers USA	MyPoints	TapResearch	
Decision Analyst	MySoapBox	TapResearch US - ES	
DEFY Media Panel - USA	Neobux	TellMeSo	
DISQO USA	On Device Research US	Tellwut US	
EarningStation	Opeepl	The Opinion Room	
Ebuno USA	Opini USA	TheoremReach US	
EmbeePay Panel USA	Opinion Capital	ToastyEgg Surveys	

Pre-testing

A pre-test with 32 interviews was conducted March 6-7, 2019 using one of the address-based panels.

Questionnaire

All six samples were administered a set of common questions. Appendix E provides the study questionnaire. There were a few questions measured for some of the samples but not others. This was because up-to-date demographic information was already on file for the address-recruited panelists but not for the opt-in panelists or crowdsourced respondents. For example, answers for sex, Hispanic ethnicity and race were measured in the study questionnaire for the opt-in respondents and measured in the most recent panel profile survey for the address-recruited respondents. Also, when the study began on March 13, 2019, it included a question asking whether respondents approved or disapproved of special counsel Robert Mueller's investigation of Russian interference in the 2016 U.S. election. That question was dropped from the study (not asked in subsequent interviews) on March 22 when the resulting report was submitted to the Attorney General. This decision was made because after March 22, the wording of the question no longer worked; the nation's focus shifted from "the investigation" to "the report." All instances in which a question was administered to some respondents but not others are noted in Appendix E.

Data quality checks

One study goal was to assess data quality under several different data checking scenarios. Methods statements for public polls conducted via opt-in methods generally make no mention of data quality checks, and so it was essential that one of the scenarios feature no checking.

Standard protocol for the Center's American Trends Panel is to review each survey dataset and remove interviews where the respondent refused to answer a large share of questions, always selected the first answer choice shown, or always selected the last answer choice shown. This cleaning process kicks out approximately 0.05% of interviews each survey. But to make the address-recruited data as comparable as possible to the opt-in data, those checks were not performed for this study.

Similarly, the other address-recruited panel has a standard set of quality checks. Once interviewing had completed, the vendor for the other address-recruited panel recommended excluding 30 of their cases due to refusing to answer 75% or more of the substantive survey questions. But again, to make the address-recruited data as comparable as possible to the opt-in data, those checks were not performed for this study.

The coordinating vendor also recommended using some of the responses collected in the survey to identify and remove low quality cases. Specifically, they recommended dropping cases that failed the attention check (COLOR); said they lived outside the US; reported an out-of-range age; indicated they are Spanish survey takers but said they couldn't understand Spanish in the open ends; or gave nonsensical open-ended answers. These checks were not performed on the data for this study. The recommended cleaning used questions (such as "Do you currently live inside or outside the U.S.?") were asked only because Center researchers put them in the survey; they are not commonly asked in public polls. Cleaning the datasets with the custom checks that we designed would have undercut the purpose and generalizability of the study.

Had routine quality checks been applied to the address-recruited panels, results from those sources would presumably have come out slightly better than those in the report – especially regarding item nonresponse to the open-ended questions. Similarly, to the extent that some public pollsters using opt-in sources may be performing routine quality checks, the opt-in results reported would be overly pessimistic. Chapter 4 of the report aims to provide readers with information about how well some of the most common checks perform.

Analysis of IP addresses

IP address is not included in the public dataset, as that may be considered personally identifiable information. It was, however, used in the report analysis. To obtain the geographical location of IP addresses, each IP address was compared to Classless Inter-Domain Routing (CIDR) ranges that are provided by The Internet Registry System and managed and distributed by the five Regional Internet Registries. The association of CIDR ranges to countries was used to classify respondents' IP address country. This was done using an R package called 'iptools'.

Following the procedure described by Ahler and colleagues, each IP address was also used to collect information from the AbuseIPDB API. For each IP address, researchers queried the API to retrieve information on usage type, the associated domain name, the internet service provider, an indicator for if an address is found on a blacklist of malicious addresses, and geographical information associated.

Usage type refers to how the IP address is used and has 11 different classifications. Usage type is classified as either Commercial, Content Delivery Network, Data Center/Web Hosting/Transit, Fixed Line ISP, Government, Library, Military, Mobile ISP, Organization, Reserved, or University/College/School. According to Ahler et al., the blacklist of malicious IP addresses is generated by AbuseIPDB users. The two primary reasons an IP address would be flagged are a website associated with the IP is caught spreading malware or engaging in phishing, or bad

Internet traffic like a DDoS attack originates from the IP. Since the geographical information from the AbuseIPDB API is sourced from a different database than the source of CIDR ranges used with the iptools R package, it was used as a validation measure of placing IP addresses in countries.

Analysis with weighted data

While the report analysis is based on unweighted data, researchers created a weight for each sample to assess whether the main study findings hold up when weights are applied. Each of the six samples was weighted separately. The weighting used an iterative technique called raking to align the sample with population benchmarks for U.S. adults on the dimensions listed in the accompanying table. The same raking dimensions were used for all six samples.

No base weight was available for the crowdsourced and opt-in survey panel samples, so only raking was performed. For the two address-recruited panels, the panel base weights (adjusting for differential probabilities of selection) were applied prior to the raking step, per standard practice. The weights for all six samples were trimmed separately at the 1st and 99th percentiles to reduce the variance in estimates from weighting.

The analysis found that the main patterns in the report (based on unweighted data) hold up when these weights are used. For example, the overall incidence of bogus respondents in the study is 4% with and without the weights applied. The average incidence of bogus respondents in the opt-in survey panel samples is similar with or without weighting (6% and 5%, respectively). Weighting also has little effect on the average incidence of bogus respondents in the address-recruited panel samples (1% weighted and unweighted). The share of bogus respondents passing both the attention check and the fast response check drops somewhat (from 76% to 69%) but remains a clear majority when weighting is applied.

Weighting dimensions

Variable	Benchmark source
Gender Age Education Race/Hispanic origin Hispanic nativity Home internet access	2017 American Community Survey
Region x Metropolitan status	2018 CPS March Supplement
Volunteerism	2015 CPS Volunteering and Civic Life Supplement
Voter registration	2016 CPS Voting and Registration Supplement
Party affiliation	Average of the three most recent Pew Research Center telephone surveys.

Note: Estimates from the ACS are based on non-institutionalized adults. Voter registration is calculated using procedures from Hur, Achen (2013) and rescaled to include the total US adult population.

PEW RESEARCH CENTER

One of the largest differences between the weighted and unweighted estimates concerns the crowdsourced sample. The share of interviews coming from bogus cases is 7% unweighted versus 4% weighted. This difference seems to stem from the fact that a very high share of bogus crowdsourced respondents report being college graduates (88%) – substantially more than the college graduate rate among bogus cases in the other samples. The severe overrepresentation of college graduates requires that, on average, the crowdsourced bogus cases get weighted down (i.e., have less influence on estimates).

Follow-up data collection

As discussed in Chapter 8, a follow-up data collection was fielded with fresh samples from opt-in panel 1 and opt-in panel 3. The purpose of this data collection was to address two important questions raised in the main study: whether the approve-of-everything response behavior was purposeful or mostly just a primacy effect and whether the fact that the opt-in panel respondents were not allowed to skip questions affected their likelihood of giving bogus data. Opt-in panel 2 was not used because in the main study all three opt-in panels performed about the same, and so including all three opt-in survey panels in the follow-up was deemed unnecessarily costly.

The follow-up survey was fielded from Dec. 2 – 7, 2019 with 10,122 interviews from opt-in panel 1 and 10,165 interviews from opt-in panel 3. Respondents to the first survey were ineligible for the follow-up study. As with the main study, the Center used a coordinating vendor, which was instructed to conduct 10,000 interviews with U.S. adults in all 50 states and the District of Columbia from each source. The survey was available in English and Spanish. The vendor was instructed to use formatting, style, and respondent sourcing that is normal for political polls conducted on each of the panels. The quotas used were the same as in the main study (reported above).

Respondents could skip any of the questions asked. In each of the samples, respondents were randomly assigned to receive survey form 1 or form 2. Form 1 displayed questions as they had been displayed in the main study, where answer choices like “approve” or “favorable” were shown first and choices like “disapprove” or “unfavorable” were shown second. Form 2 displayed all the same questions but answer choices like “disapprove” or “unfavorable” were shown first and choices like “approve” or “favorable” were shown second.

Notifications for researchers downloading the study microdata

The micro-dataset for this study is available on the Pew Research Center website. Data users should be aware of the following:

1. Some open-ended answers contain emojis. In general, the emojis render in R and Microsoft Excel. They may not render properly in SPSS and possibly other survey software.

2. A small number of open-ended answers were redacted because the answer appeared to contain personally identifiable information. Such answers appear in the dataset as "REDACTED [PII]."

3. Some open-ended answers contain vulgar or offensive comments. As with all respondent data, inclusion by the Center does not constitute an endorsement or recommendation of any viewpoint, service or policy. Unless an answer appeared to contain personally identifiable information, the answers are left intact so that other researchers can see the raw data for themselves.

Appendix B: Protocol for coding open-ended answers

Overview

The questionnaire contained the following six open-ended questions. Analysis of answers to these questions is central to the study findings. This section details the procedures researchers used to code the data.

[FEELS]	How would you say you are feeling today?
[GREWUPCITY]	When you were growing up, what was the big city nearest where you lived?
[CITYVISIT]	When you visit a new city, what kinds of activities do you like to do?
[COMPUTER]	How do you decide when your computer is too old and it's time to purchase a new one?
[RETIRE]	In retirement what skill would you most like to learn?
[GETDONE]	What would you like to see elected leaders in Washington get done during the next few years? Please give as much detail as you can.

Coding blank, gibberish, don't know/refused, responsive or non sequitur

In the public dataset, the variables FEELS_CODE, … , GETDONE_CODE provide the codes used to categorize each answer as blank, gibberish, don't know/refused, responsive, or non sequitur. The codes are defined as follows:

0 = Blank	Response field is completely blank. No characters or emojis.
1 = Gibberish	Answers that are gibberish. This includes answers that are only punctuation or letters not forming real words. Examples: "ahhjfvadvasdv," "-----," ".," and "5ry6754etdvhuyji"

2 = Don't know/Refuse Answers expressing that the respondent does not know how to answer or is unwilling to answer. Examples: "Don't know," "Dunno," "idk," "Not sure," "No opinion," "No comments," "Not sure," and "?"

3 = Responsive Answers that are responsive to the question. Examples: "Feelin aaaight," "Just want to relax," and "Bring the BIBLE back into schools"

4 = Non sequitur Answers that do not follow from (are not responsive to) the question asked. Example: Q: *How would you say are you feeling today?* A: "I love this product!" or "Ceiling Pop"
Answers consisting of numbers or emojis that are not responsive to the question are assigned this code as are answers found to have been plagiarized from another website.

General principles

Coders were instructed to adhere to the following principles.

Principle i. Employ a generous, permissive definition of answers that are responsive to the question. Give the respondent the benefit of the doubt. For example, the following answers arguably don't make perfect sense but are close enough to be coded responsive:
Q: *How would you say are you feeling today?* A: "good live," "secure," "depends"

Principle ii. Disregard spelling and/or grammatical errors. Many answers contain spelling or grammar errors. As long as it is possible to discern roughly what the respondent was trying to say, code the answer as responsive if it is on topic. For example: "gr8," "grate," and "hapyp" all count as responsive answers to FEELS.

Principle iii. Curt or sarcastic responses should be coded as responsive provided they make sense based on the question asked. For example, the following answers should be coded 3: Q: *How would you say are you feeling today?* A: "about what?," "with my hands," and "meh"

Principle iv. Several answer characteristics may warrant coding as non sequitur. In particular:
● If the question asks about the respondent *personally* (see FEELS, CITYVISIT, RETIRE) but the response focuses on an object (e.g., "I like it," "it's fast"), a non sequitur code may be appropriate
● If the answer expresses satisfaction (e.g., "amazing," "Very satisfied," "good") but that does not align with the question, a non sequitur code may be appropriate. This applies to all open ends except FEELS.
● If the question asks about the respondent personally (see FEELS, CITYVISIT, RETIRE) we

would generally expect answers from the first-person perspective, though pronouns like "I" and "me" do not necessarily need to be used. If answers to such questions inexplicably use second person (e.g., "Develop your skills") or third person (e.g., People enjoy learning about it"), a non-sequitur code (4) may be appropriate. But in all cases, best judgment should be used.
● If the answer just repeats a few words from the question but does not elaborate or give a coherent answer, code as non sequitur.

Principle v. For answers containing both gibberish and words, focus on the words.
● If the answer contains responsive words and some gibberish, code as responsive.
● If the answer contains words unconnected to the question and some gibberish, code as non sequitur.

Considerations specific to individual open-ended questions

Coders were provided with the following guidelines that are specific to individual questions.

[FEELS] **How would you say you are feeling today?**
● This question was asked after several political items, leading many respondents to interpret it as "how are they feeling today about politics." To address this, all politics-related answers to FEELS should be coded as responsive even if they do not specifically refer to the respondent's feelings. Such answers are arguably responsive given the context in which FEELS was asked.
● For this question only, percentages (e.g., "75%") and letter grades (e.g., "B") should be counted as responsive answers since those are common ways for people to articulate how they are feeling.
● Answers like the following should be coded as refusal: "Don't know," "Not sure," "No comment," "None of your business," and "n/a"
● Answers like the following should be coded as non sequitur: "Yes," "No," "never," "none," and "nothing"

[GREWUPCITY] **When you were growing up, what was the big city nearest where you lived?**
● Answers indicating that the question is not applicable should be coded as responsive because there are legitimate reasons why some people might feel it does not apply to them. Answers that reference moving, being in a military family, not living near any big cities, living in a big city (as opposed to near one), or the ambiguity of "big city" should be counted as responsive. For example, responses like the following should all be coded as responsive: "Military brat," "What do you mean by big city," "Moved around," "Several different cities," "Boston or Chicago," "Multiple," "I lived in the city," "None" and "n/a"
● Common city abbreviations count as responsive. For example: "OKC," "NY," "NYC," "PHL," and "SLC"

● The question asks for a city, so answers that are a state or a country should be coded as non sequitur.
● Answers like the following should be coded as refusal: "No comment," "Unsure," "idk," "don't want to say"

[CITYVISIT] **When you visit a new city, what kinds of activities do you like to do?**
● Answers indicating that the question is not applicable should be coded as responsive because there are legitimate reasons why some people might feel it does not apply to them. For example, answers referencing not liking cities, not traveling, or a disability count as responsive. Similarly, count answers like "n/a" responsive even if there is no explanation.
● Curt or sarcastic responses should be coded as responsive if they are on topic. For example: "Anything," "Nothing," "Leave," "Everything," "Whatever," "Historic," "A lot," "All," "Cheap," "Depends," and "None" count as responsive (3).
● Answers like the following should be coded as refusal: "Don't know," "Not sure," "No comment," and "None of your business"

[COMPUTER] **How do you decide when your computer is too old and it's time to purchase a new one?**
● Answers indicating that the question is not applicable should be coded as responsive because there are legitimate reasons why some people might feel it does not apply to them. For example, answers referencing having no computer, just using a smartphone, having a new computer, not being the person who decides, or never having replaced a computer count as responsive. Similarly, count "n/a" or "none" type responses as responsive even if there is no explanation.

[RETIRE] **In retirement what skill would you most like to learn?**
● Answers indicating that the question is not applicable should be coded as responsive because there are legitimate reasons why some people might feel it does not apply to them. For example, answers referencing already being retired, will never be able to retire, have never thought retirement, don't want to retire, or retirement is too far into the future count as responsive.
● Many respondents gave some type of "I don't know" answer. Some made clear that they did not know because retirement is too far away or because they have never given it any thought. Others did not offer a reason. Because "don't know" is such a common, arguably legitimate reaction to this question, all answers along those lines should be coded as responsive. This means that no answers to RETIRE shall be coded as don't know/refused.
● Answers like the following should be coded as responsive: "none," "n/a," and "nothing"

[GETDONE] **What would you like to see elected leaders in Washington get done during the next few years? Please give as much detail as you can?**

● Answers that express feelings like frustration, lack of faith in the political system, or belief that politicians will just do what they want should be coded as responsive.
● Answers that start with something like "I don't know" but offers some answer like "cut taxes" should be counted as responsive.
● Answers that are just a name or list of names such as "Barack Obama" or "Bernie" should be coded as non sequitur.
● This question was asked after a question on volunteering, which seems to explain why some answers are just people describing the volunteer work that they do. Those should all be coded non sequitur. While this is a departure to the guideline for FEELS, any reasonable reading of this question makes clear that it is asking about elected officials not volunteer activities that the respondent may be doing.
● For this question "none" and "n/a" should be coded as refusal, while "nothing" should be coded as responsive.

Coding non sequitur answers

In the public dataset, the variables FEELS_NONSEQTYPE, … , GETDONE_NONSEQTYPE provide the codes used to categorize each non sequitur answer as positive product evaluation, negative product evaluation, common expression, conversational, plagiarized, or other non sequitur. The codes are defined as follows:

5 = Positive product evaluation	Answers that sound like a customer evaluation of a product, specifically a positive evaluation. This includes responses that are simply descriptors (e.g., "Great," "Awesome," "Good," "Cool," "Like," and "Nice") as well as more elaborate answers, such as ("love it," "is great," and "excellent brand")
6 = Negative product evaluation	Answers that sound like a customer evaluation of a product, specifically a negative evaluation. This includes responses that are simply descriptors (e.g., "Bad" or "Poor quality") as well as more elaborate answers.
7 = Common expression	Answers that are solely or primarily comprised of the following words or phrases: "Yes," "Yeah," "No," "OK," "Thank you," "I agree," "None," "Nothing," "Never," "Hi," "Hello," "Hey," or "What." Variations such as "Nope," and "not really" are included. Note that this is an exhaustive list, not a set of examples. Also note that common expressions that are positive or negative in tone should generally be coded 5 or 6.

8 = Conversational text	Answers that sound like part of a conversation between people. In general, these answer sound informal and they often, but not always, contain repeated phrases. Examples: "You can please just hang in there and I look forward to hearing from you soon," "5AM said he had no plans for the warehouse but," "Thank you for the update and for the update and for the update," and "Greg and I are are you?"
9 = Copied from a website	Answers that sound like they were copied and pasted from a source outside the survey. These answers have one or more of the following characteristics: (i) Sounds like marketing/advertising. Examples: "SEO provides the very bedrock and foundation for skyrocketing your sales," and "Explore by Foot. VirtualTourist members agree that the best way to see a new destination is to experience it by foot" (ii) Sounds like it came from a Wikipedia entry, a news story, or other information-based website. Examples: "George Washington was commander in chief of the Continental Army…," or "Administers education programs for children in state institutions" (iii) Sounds like it came from an online Q&A website, chatroom, etc. Examples: "Hello! Tell me please what is the difference between 'How are you feeling today' and 'How do you feel today'? As for me…" (iv) Unusually stilted or formal. Examples: "An illegal act by an officeholder constitutes political corruption"
10 = Other	Answers that do not fit any of the other categories. Examples, "California," "Can I just get my points," Bedroom Pop," "content://media/external/file/738023," "$25," and "new." This category also include answers that appeared to come from an English as second language respondent who may not have understood the question.

Appendix C: Reliability analysis for open-ended codes

A senior member of the research team manually coded more than 360,000 open-ended answers using the coding protocol in Appendix B. To evaluate the reliability of these codes, a sample of study interviews was selected and independently coded by a team of five researchers. This sample consisted of all 6,940 respondents for which one or more open-ended answers was initially coded as gibberish, item nonresponse or non sequitur. An additional 500 cases whose answers were all coded as either responsive or blank were randomly selected from each of the six samples. In total, 57,599 open-ended answers from 9,940 respondents were used to measure the reliability of the coding scheme.[27] The sample was randomly divided into four equal sized batches, and each batch was coded by two different researchers such that each open-ended answer was independently coded a total of three times including the initial code.

A coding scheme is said to be reliable when different people following the same set of instructions tend to agree on the proper classification of the answers to be coded. If coders frequently disagree about the classification, then the coding would be unreliable. Across all questions and answers in this study, the reliability coders reached the same conclusion as the primary coder 94% of the time. The agreement rate varied somewhat by question, ranging from 92% for GETDONE to 96% for GREWUPCITY and RETIRE.

Reliability scores for coding of open-ended questions

Intercoder reliability scores before adjudication

Question	Percent agreement (%)	Krippendorf's alpha
CITYVISIT	95	.81
COMPUTER	93	.84
FEELS	96	.77
GETDONE	92	.87
GREWUPCITY	96	.85
RETIRE	96	.81
Overall	**94**	**.85**

Notes: See Appendix B the full wording of the open-ended questions
Source: Surveys conducted March 13-22, 2019; March 19-April 4, 2019; April 1-15, 2019.
"Assessing the Risks to Online Polls From Bogus Respondents"

PEW RESEARCH CENTER

In addition to calculating the chance of agreement with the initial coding, Krippendorf's alpha was also computed to measure the chance-adjusted probability of agreement between all four coders.[28] Krippendorf's alpha was chosen as a reliability metric for this analysis because it accommodates multiple coders and allows for the possibility that not every answer will have been coded by the same set of people. For a given question, the

[27] Answers that were left blank by the respondent were coded automatically. Because these did not involve any individual discretion or judgment to code, they were excluded from the reliability analysis.
[28] Krippendorff, K. 2004. "Reliability in Content Analysis: Some Common Misconceptions and Recommendations." *Human Communication Research* 30(3):411–33.

value of alpha ranges from 0 to 1, where 0 means that coders always disagree or are assigning codes randomly and 1 means that coders always agree on the correct classification.

While there is no one-size-fits-all threshold, an alpha of 0.8 is generally considered to be desirable. Taking all of the questions and answers together, the codes have an alpha of 0.85. Individually, all but one of the questions had an alpha of 0.8 or higher. The exception was FEELS which had an alpha of .77. After the reliability coding was completed, answers where two or more of the reliability coders disagreed with the primary coder were reviewed and a final code was chosen.