

FOR RELEASE MAY 9, 2019

When Online Survey Respondents Only ‘Select Some That Apply’

Forced-choice questions yield more accurate data than select-all-that-apply lists

BY Arnold Lau, Courtney Kennedy

FOR MEDIA OR OTHER INQUIRIES:

Arnold Lau, Research Analyst
Courtney Kennedy, Director of Survey Research
Rachel Weisel, Communications Manager
202.419.4372
www.pewresearch.org

RECOMMENDED CITATION

Pew Research Center, May 2019, “When Online Survey Respondents Only Select Some That Apply”

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.

© Pew Research Center 2019

When Online Survey Respondents Only ‘Select Some That Apply’

Forced-choice questions yield more accurate data than select-all-that-apply lists

Anyone who has taken a survey has likely been given the option to “check all that apply” when answering a question. The instruction is widely used in data collection because of its ease and efficiency. But when designing an online survey questionnaire, there is more than one way to ask a respondent to select which options in a series applies to them.

A pollster can show respondents a list and ask them to select all that apply, *or* the pollster can ask separately about each option. The two methods do not always yield identical results. Studies have repeatedly indicated that more survey respondents endorse (answer “yes” to) each option under the latter approach, known as “forced-choice,” sometimes resulting in very different estimates. However, previous studies have not been clear on which question format produces the more accurate estimates.¹

Using its national online [American Trends Panel](#) (ATP), Pew Research Center conducted a large, randomized experiment comparing these two question formats. The experiment was part of a survey conducted July 30-Aug. 12, 2018, among 4,581 U.S. adults. Respondents were asked whether they or someone in their immediate family had experienced various undesirable events (e.g., treated for addiction to drugs or alcohol, or lost a home to foreclosure). The study randomly assigned half the respondents to answer using a select-all-that-apply list while the other half answered a series of forced-choice “yes/no” questions.

The challenge of this sort of research is in determining which measurement gets it more “right.” The [research](#) that undergirds this experiment shows that, when it comes to undesirable but relatively common events, respondents are more likely to underreport their actual experiences than to overreport something they haven’t experienced. For example, it’s more plausible that someone treated for addiction would decline to tell a pollster about their experience than it is that someone who has never been treated for addiction would report that they had been. For the

¹ Smyth, Jolene, Don Dillman, Leah Christian and Michael Stern. 2006. [“Comparing Check-All and Forced-Choice Question Formats in Web Surveys.”](#) Public Opinion Quarterly; Callegaro, Mario, Michael Murakami, Ziv Tepman and Vani Henderson. 2015. [“Yes-no answers versus check-all in self-administered modes: A systematic review and analysis.”](#) International Journal of Market Research.

purposes of this study, then, higher victimization estimates are thus presumed to reflect more accurate responses.²

The new study found compelling evidence that forced-choice questions yield more accurate results than select-all-that-apply lists. In the randomized experiment, panelists were first asked, “Which of the following, if any, have happened to you, personally?” and were shown six items. Later in the questionnaire, they were then asked, “Which of the following, if any, has happened to you OR someone in your family?” and were shown a different set of six items. Each panelist was randomly assigned to receive both questions either as select-all-that-apply lists or as forced-choice batteries. The study found that victimization estimates were consistently higher using the forced-choice format versus selecting all that apply. For example, the estimated share of U.S. adults saying that they had personally been denied coverage by a health insurance company was 19% using the forced-choice format versus 13% using select-all-that-apply. The estimated share of U.S. adults saying that they or someone in their immediate family had lost a job and struggled to find another one was 63% using the forced-choice format versus 51% under select-all-that-apply. On average, victimization estimates were 8 percentage points higher with individual forced-choice questions than with a checklist.

While question format clearly affected endorsement rates, it had little effect on which items were most or least endorsed. The arrangement of most to least endorsed items is identical or highly similar between forced-choice and select-all-that-apply formats, with differences being almost entirely within the margins of sampling error. This implies that if only the rank order is of interest, either format will do.

A companion analysis using the Center’s ATP archive examined whether primacy effects, a type of measurement error, were larger for select-all-that-apply questions than forced-choice ones. Primacy effects refer to the tendency for some respondents to endorse whichever item occurs at the top of the list more often than items below. An example of a primacy effect in archival ATP data is when respondents were shown a list of personal traits and asked to select the ones that described them well. The order in which the traits were listed varied across respondents. A total of 57% of respondents endorsed whichever trait was listed first, while only 42% of respondents endorsed the trait listed last – an order effect of 15 percentage points. Other researchers have examined whether such order effects are larger for checklists than for forced-choice formats, and the findings have been mixed.³

² Tourangeau, Roger, Lance Rips and Kenneth Rasinski. 2000. “The Psychology of Survey Response.”

³ Callegaro, Mario, Michael Murakami, Ziv Tepman and Vani Henderson. 2015. “[Yes-no answers versus check-all in self-administered modes: A systematic review and analysis.](#)” International Journal of Market Research.

Across 21 ATP surveys conducted between November 2015 and August 2018, the average order effect observed for select-all-that-apply questions was 3 percentage points, compared to 0 percentage points for forced-choice questions. While this suggests that primacy effects may be slightly larger for checklists than forced-choice questions, the overall relationship is weak. The archival analysis is also much more tentative than the aforementioned experimental analysis because in the archival data the question formats were not tested in a head-to-head fashion. There are several other potential explanations for differences observed, such as certain question topics being both more prone to primacy effects and being more likely to be fielded as select-all-that-apply questions. Another factor could be the tendency for select-all-that-apply questions to have more answer choices due to having more space. The study's findings regarding differences in items that were endorsed most or least and primacy effects are consistent with previous studies.

Based on the results of these analyses, Pew Research Center has adopted a policy of using a forced-choice format instead of a select-all-that-apply list in its online surveys whenever possible.

Pros and cons of different question formats

Select-all-that-apply questions are formatted succinctly and can thus take up only a single screen. By contrast, forced-choice batteries require a separate question for each item, which is more likely to require respondents to either scroll down or click through several screens. This increases the amount of time it takes for respondents to get through these questions, which can be burdensome.

However, select-all-that-apply questions ask respondents to consider all the options at the same time. Responding accurately with respect to each item in the list can require a fair amount of discipline from respondents. Those lacking enough motivation may “satisfice,” or respond only in a satisfactory manner rather than in the most accurate way possible. When presented with a sizeable array of options, respondents may select only *some* that apply rather than all.

Forced-choice questions, as their name implies, force the respondents to provide a separate answer for each item, one by one. This format encourages respondents to more deeply consider each option, especially as they are not simultaneously juggling all the other options. Prior studies have shown that respondents are more likely to endorse options when presented to them as a forced-choice compared to when the same options are presented as a select-all-that-apply.

Proponents of forced-choice questions argue that this increased rate of endorsement is a sign that the format encourages deeper cognitive processing, or salience.⁴ However, it is also possible that

Forced-choice formats require more space and take more time than select-all-that-apply

Screenshots from Pew Research Center randomized experiment

The image shows two overlapping screenshots of survey questions. The left screenshot shows a forced-choice question: "Which of the following, if any, have happened to you, personally?" with three items: "Had a home break-in", "Had your wallet or purse stolen", and "Denied coverage by a health insurance company". Each item has two radio button options: "Yes, has happened to me" and "No, has not happened to me". The right screenshot shows a select-all-that-apply question: "Which of the following, if any, have happened to you, personally?" with the instruction "[Check all that apply]". It lists the same three items as the left screenshot, but each has a checkbox option. At the bottom of the right screenshot is a checkbox for "None of these".

Source: Survey of U.S. adults conducted July 30-Aug. 12, 2018.
 “When Online Survey Respondents Only ‘Select Some That Apply’”

PEW RESEARCH CENTER

⁴ Smyth, Jolene, Don Dillman, Leah Christian and Michael Stern. 2006. “[Comparing Check-All and Forced-Choice Question Formats in Web Surveys.](#)” *Public Opinion Quarterly*; Thomas, Randall K., Barlas, Frances M., Buttermore, Nicole R., and Jolene D. Smyth. 2017. “Acquiescence

this format is much more susceptible to acquiescence response bias, which is when respondents tend to endorse any assertion made in a question, regardless of its content.⁵ If satisficing is the larger problem, then estimates derived from select-all-that-apply formats could be lower than they should be (and thus more inaccurate). However, if acquiescence bias dominates, then estimates derived from forced-choice formats may be artificially inflated.

One proven way of dealing with acquiescence bias is to use more direct, construct-specific wording, where the response categories are tailored to be directly relevant to the question being asked.⁶ For example, a question about whether respondents have or have not done a range of activities in the past year can spell out the response options as “Yes, have done this” and “No, have not done this” rather than merely “Yes” and “No.”⁷

Bias in Yes-No Grids? The Survey Says... No.” Poster presentation. American Association for Public Opinion Research; Haynes, Harper, et al. 2018. “Yes, No, and Maybe So: Assessing the Data Quality of Check-All and Forced-Choice Question Formats.” Poster presentation. American Association for Public Opinion Research.

⁵ Krosnick, Jon. 1999. “[Survey Research.](#)” Annual Review of Psychology; Callegaro, Mario, Michael Murakami, Ziv Tepman and Vani Henderson. 2015. “[Yes-no answers versus check-all in self-administered modes: A systematic review and analysis.](#)” International Journal of Market Research.

⁶ Höhne, Jan and Timo Lenzner. 2018. “[New Insights on the Cognitive Processing of Agree/Disagree and Item-Specific Questions.](#)” Journal of Survey Statistics and Methodology.

⁷ Krosnick, Jon, Paul Lavrakas and Nuri Kim. 2014. “Survey Research”. In Reis, Harry and Charles Judd, eds. “Handbook of Research Methods in Social and Personality Psychology.” 2nd ed.

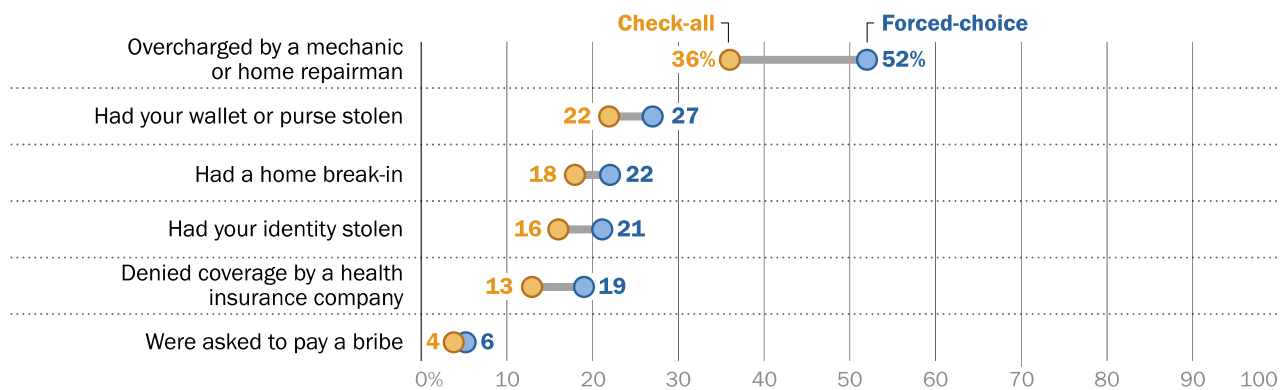
Reporting of undesirable experiences is higher with forced-choice questions than select-all-that-apply

Endorsement rates – the share of survey respondents who answer “yes” to an item – tend to be higher with a forced-choice format than with select-all-that-apply. Are the higher endorsement rates for forced-choice questions caused by respondents answering the items with greater care, or with greater acquiescence?

To answer this question, Pew Research Center conducted an experiment on the American Trends Panel. Panelists were first asked, “Which of the following, if any, have happened to you, personally?” and were shown six items. Later in the questionnaire, they were then asked, “Which of the following, if any, has happened to you OR someone in your family?” and were shown a different set of six items. Each panelist was randomly assigned to receive both questions either as select-all-that-apply lists or as forced-choice batteries. Panelists who were shown the select-all-that-apply list were also given the option to check a box at the bottom of the list stating that none

In questions about personal experiences, victimization estimates are consistently higher for forced-choice than for select-all-that-apply

Which of the following, if any, have happened to you, personally?



Note: “None of these” responses for select-all-that-apply condition not shown.

Source: Survey of U.S. adults conducted July 30-Aug. 12, 2018.

“When Online Survey Respondents Only ‘Select Some That Apply’”

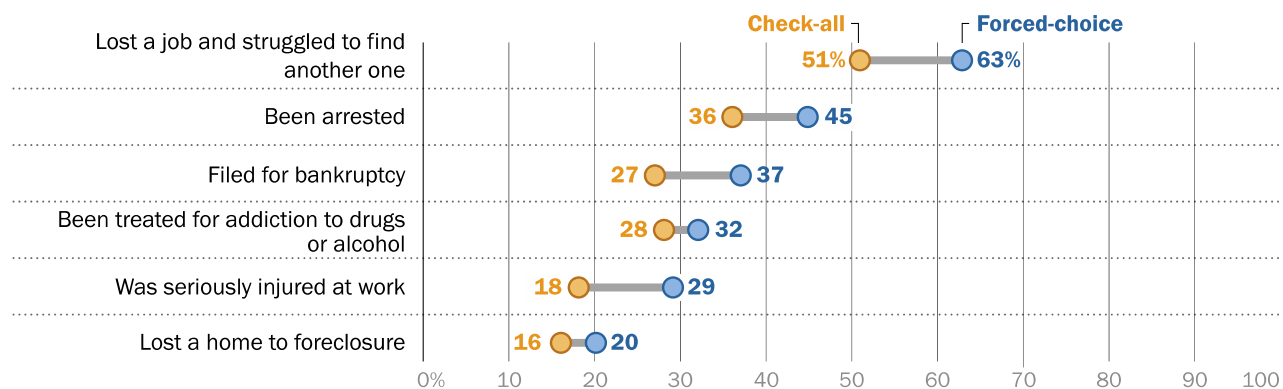
PEW RESEARCH CENTER

of the options applied to them.⁸ Panelists who were given forced-choice batteries were also given construct-specific wording for their answer choices; that is, rather than “Yes” and “No,” panelists chose between “Yes, has happened to me [or my family]” and “No, has not happened to me [or my family].”

Across all twelve items, endorsement rates were higher for the forced-choice format than for select-all-that-apply. For example, 19% of forced-choice respondents said that they had personally been denied coverage by a health insurance company, compared with 13% of select-all-that-apply respondents. Similarly, 63% of forced-choice respondents said that they or someone in their family had lost a job and struggled to find another one, compared with only 51% of select-all-that-apply respondents. The largest difference came from asking if respondents had personally been overcharged by a mechanic or home repairman, with 52% in the forced-choice condition reporting the experience versus 36% in the select-all-that-apply condition – a 16 percentage point difference.

Victimization estimates are consistently higher for forced-choice in questions about family experiences

Which of the following, if any, has happened to you OR someone in your immediate family?



Note: “None of these” responses for select-all-that-apply condition not shown.

Source: Survey of U.S. adults conducted July 30-Aug. 12, 2018.

“When Online Survey Respondents Only ‘Select Some That Apply’”

PEW RESEARCH CENTER

Acquiescence response bias occurs when it is easier or more socially desirable to endorse a response option rather than do the opposite. Respondents in this experiment were not expected to be susceptible to carelessly reporting potentially embarrassing or disheartening events, nor was there any social incentive to do so.⁹ Factors that usually drive acquiescence should in this case

⁸ For both questions, 1% of respondents in the select-all-that-apply condition opted to select “none of these”.

⁹ Tourangeau, Roger, Lance Rips and Kenneth Rasinski. 2000. “The Psychology of Survey Response.”

discourage endorsement. Regardless, forced-choice endorsement rates were higher than those for select-all-that-apply by an average of 8 percentage points across all twelve items. This tells us that the estimates from the forced-choice format are more accurate, whereas the checklist format is more prone to respondents only selecting *some* that apply.

Forced-choice and select-all-that-apply questions yielded similar rank orders of endorsement

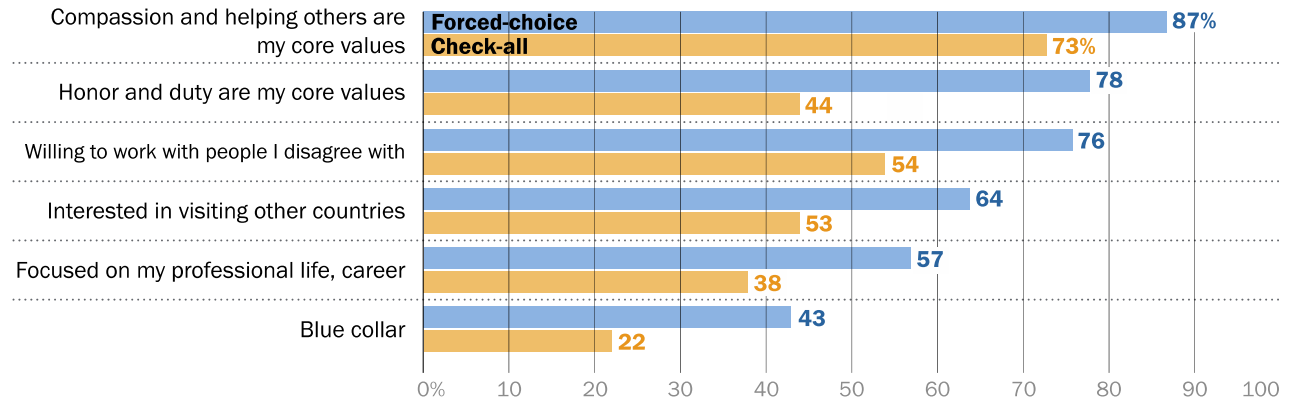
For some applications, having accurate estimates for each individual item in a battery or list is less important than having an accurate ranking of which items were the most to least endorsed. When asked which of a series of events had happened to them personally, the question format made no difference in the rank order. Being overcharged by a mechanic or home repairman was the most reported event across both forced-choice and select-all-that-apply conditions, while being asked to pay a bribe was the least reported event, with the rank order of all other items in between agreeing regardless of question format.

For events that had happened to respondents or someone in their families, filing for bankruptcy and being treated for alcohol or drug addiction switch positions between the two experimental conditions. However, the ranking switch is arguably of minor consequence at best, with 27% reporting filing for bankruptcy and 28% reporting addiction treatment under select-all-that-apply, compared to 37% and 32% respectively under forced-choice.

A similar experiment conducted earlier with the American Trends Panel appears to challenge this at first glance. In an ATP survey of 4,867 U.S. adults conducted Sept. 14-28, 2017, respondents were randomly assigned to answer a battery of questions using a forced-choice or select-all-that-apply format. Respondents were asked whether a series of personal traits described them well. In that experiment, the statement “honor and duty are my core values” ranked second using the forced-choice format and fourth using the select-all-that-apply format. There was a gaping 34 percentage point difference for that item between the two conditions (78% with forced-choice, 44% with select-all-that-apply). The precise reason for this large difference is unclear, though the survey literature indicates that vague, attitudinal concepts (like those in the traits experiment) are more susceptible to question formatting effects than concrete, behavioral concepts (like those in the victimization experiment presented earlier).

Vague, attitudinal concepts could be more susceptible to question wording effects

Which of these describes you well?



Note: "None of these" responses for select-all-that-apply condition not shown.

Source: Survey of U.S. adults conducted Sept. 14-28, 2017.

"When Online Survey Respondents Only 'Select Some That Apply'"

PEW RESEARCH CENTER

Order effects are generally similar between forced-choice and check-all-that-apply questions

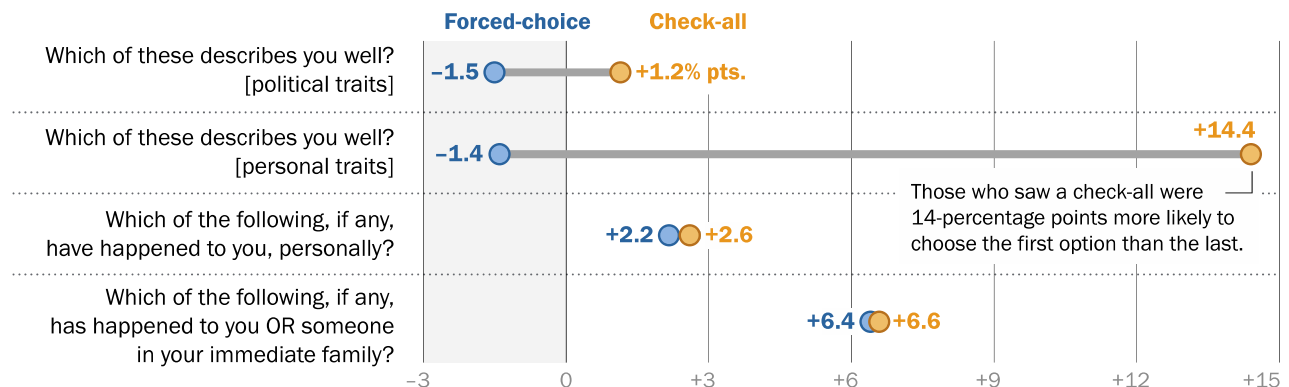
This study also considered whether the size of primacy effects – the tendency to select the first option(s) listed – differs for forced-choice and select-all-that-apply questions. The study specifically considered primacy effects rather than other types of order effects because the surveys considered were all conducted online with respondents reading the questions. (Telephone survey pollsters, by contrast, tend to worry about *recency* effects because the most recent option spoken by the interviewer is the easiest to recall).

Primacy effects are another indication of people responding in a manner that is merely “good enough.” Respondents being more likely to select whatever item appears closer to the top of a list could indicate that they merely skimmed or did not read the item all the way through. It is standard practice to randomize the order in which respondents see items in order to minimize the impact of primacy effects. However, in some applications, items must be presented in a specific order. The concern is that select-all-that-apply lists, in particular, encourage primacy effects.

The experimental data provides only inconclusive evidence that this is the case. Researchers calculated primacy effects as the signed difference in the endorsement rate between whatever item respondents saw at the first position and whatever item they saw last. For the family victimization battery discussed above, the primacy effects were virtually the same for the forced-choice and

Experiments on the relationship between question format and primacy effects were mixed

Signed difference in endorsement rate between item respondents saw at first position and item respondents saw at last position, in percentage points



Source: Surveys of U.S. adults conducted Sept. 14-28, 2017, and July 30-Aug. 12, 2018.
 “When Online Survey Respondents Only ‘Select Some That Apply’”

PEW RESEARCH CENTER

select-all-that-apply formats (6.4 and 6.6 percentage points, respectively). The same was true for the personal victimization battery, with the forced-choice format displaying a primacy effect of 2.2 percentage points and the select-all format displaying an effect of 2.6 percentage points.

However, for the personal trait battery, the select-all-that-apply format yielded a large, 14.4-point primacy effect, while the forced-choice format showed no primacy effect (-1.4 percentage points). From the same survey, a battery that asked respondents which of a series of political traits described them well – such as being an environmentalist or being a supporter of rights for lesbian, gay, bisexual and transgender (LGBT) people – yielded primacy effects close to zero for both question formats.

Given that the order effects observed in this handful of tests were mixed, researchers sought a larger source of data, albeit a non-experimental one. Researchers scoured 21 surveys conducted on the ATP between November 2015 and August 2018 and identified 40 select-all-that-apply and 25 forced-choice questions that had previously been asked.¹⁰

Unlike with the experiments reported above, respondents in these surveys all received the same question format. This means that comparisons between the forced-choice results and select-all-that-apply results could be confounded by factors like differing topics or differing numbers of items. This limitation makes the results from the archival analysis much more tentative than results from the experiments reported above.

In the archival data, the average primacy effect was 0 percentage points among the forced-choice questions examined, while it was 3 percentage points for the select-all-that-apply questions. Taken together, the experimental and observational archival data suggest that forced-choice formatting is, if anything, less susceptible to primacy effects than select-all-that-apply questions, but that effect is usually small and not always consistent.

¹⁰ This count includes the survey experiments, with each condition being treated as a separate question. Questions with 3 or less items and questions asked only of a miniscule subset of respondents were excluded from this analysis.

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

Arnold Lau, *Research Analyst*

Courtney Kennedy, *Director, Survey Research*

Andrew Mercer, *Senior Research Methodologist*

Nick Bertoni, *Panel Manager*

Nick Hatley, *Research Analyst*

Scott Keeter, *Senior Survey Advisor*

Claudia Deane, *Vice President of Research*

Communications and editorial

Rachel Weisel, *Communications Manager*

Hannah Klein, *Communications Associate*

Calvin Jordan, *Communications Associate*

Travis Mitchell, *Digital Producer, Copy Editor*

Graphic design and web publishing

Bill Webster, *Information Graphics Designer*

Methodology

The American Trends Panel survey methodology

The American Trends Panel (ATP), created by Pew Research Center, is a nationally representative panel of randomly selected U.S. adults. Panelists participate via self-administered web surveys. Panelists who do not have internet access at home are provided with a tablet and wireless internet connection. The panel is being managed by Ipsos.

Data in this report is primarily drawn from the panel waves conducted Sept. 14-28, 2017 among 4,867 respondents, with a margin of sampling error for the full sample of 2.3 percentage points, and July 30-Aug. 12, 2018 among 4,581 respondents, with a margin of sampling error for the full sample of 2.4 percentage points. All in all, this report references a total of 21 panel waves conducted between November 2015 and August 2018.

As of August 2018, members of the American Trends Panel were recruited from several large, national landline and cellphone random digit dial (RDD) surveys conducted in English and Spanish. At the end of each survey, respondents were invited to join the panel. The first group of panelists was recruited from a Pew Research Center survey conducted Jan. 23 to March 16, 2014. Of the 10,013 adults interviewed, 9,809 were invited to take part in the panel and a total of 5,338 agreed to participate.¹¹ The second group of panelists was recruited from a Pew Research Center survey conducted Aug. 27 to Oct. 4, 2015. Of the 6,004 adults interviewed, all were invited to join the panel, and 2,976 agreed to participate.¹² The third group of panelists was recruited from a survey conducted April 25 to June 4, 2017. Of the 5,012 adults interviewed in the survey or pretest, 3,905 were invited to take part in the panel and a total of 1,628 agreed to participate¹³.

The ATP data was weighted in a multi-step process that began with a base weight incorporating the respondents' original survey selection probability and the fact that in 2014 some panelists were subsampled for invitation to the panel. Next, an adjustment was made for the fact that the propensity to join the panel and remain an active panelist varied across different groups in the sample. The final step in the weighting uses an iterative technique that aligns the sample to population benchmarks on a number of dimensions. Gender, age, education, race, Hispanic origin

¹¹ When data collection for the 2014 Political Polarization and Typology Survey began, non-internet users were subsampled at a rate of 25%, but a decision was made shortly thereafter to invite all non-internet users to join. In total, 83% of non-internet users were invited to join the panel.

¹² Respondents to the 2014 Political Polarization and Typology Survey who indicated that they are internet users but refused to provide an email address were initially permitted to participate in the American Trends Panel by mail, but were no longer permitted to join the panel after February 6, 2014. Internet users from the 2015 Pew Research Center Survey on Government who refused to provide an email address were not permitted to join the panel.

¹³ White, non-Hispanic college graduates were subsampled at a rate of 50%.

and region parameters come from the U.S. Census Bureau's 2016 American Community Survey. The county-level population density parameter (deciles) comes from the 2010 U.S. Decennial Census. The telephone service benchmark comes from the July-December 2016 National Health Interview Survey and is projected to 2017. The volunteerism benchmark comes from the 2015 Current Population Survey Volunteer Supplement. The party affiliation benchmark is the average of the three most recent Pew Research Center general public telephone surveys. The Internet access benchmark comes from the 2017 ATP Panel Refresh Survey. Respondents who did not previously have internet access are treated as not having internet access for weighting purposes. Sampling errors and statistical tests of significance take into account the effect of weighting. Interviews are conducted in both English and Spanish, but the Hispanic sample in the American Trends Panel is predominantly native born and English speaking.

The following table shows the unweighted sample sizes and the error attributable to sampling that would be expected at the 95% level of confidence for different groups in the survey:

Group	Unweighted sample size	Plus or minus ...
Total sample (Wave 37)	4,581	2.4 percentage points
Total sample (Wave 29)	4,867	2.3 percentage points

Sample sizes and sampling errors for other subgroups are available upon request.

In addition to sampling error, one should bear in mind that question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of opinion polls.