Mode Effects in Online Knowledge Surveys

This white paper reports the results of five surveys designed to guide the Annenberg Constitution Day Civics Knowledge survey's transition from random digit dial (RDD) phone samples to an online panel survey mode. Because the Constitution Day survey focuses on knowledge, the move from phone to an online mode increases the opportunities for respondents to misrepresent their command of civics knowledge by reporting answers obtained through an online search. Here we report the differences observed by mode, describe the effects of our efforts to depress online searching, and explain the protective processes that we have implemented.

The Annenberg Public Policy Center. 2023. "Mode Effects in Knowledge Surveys."
August 30, 2023.

Table of Contents

Acknowledgments

This report is a collaborative effort based on the input and analyses of the following individuals. Find related reports online at www.annenbergpublicpolicycenter.org/.

Primary Researchers

Kathleen Hall Jamieson, Director, APPC
Matt Levendusky, Director AIOD Survey
Shawn Patterson Jr., Research Analyst
Ken Winneg, Managing Director of Survey Research

Editorial and Graphic Design Team

Michael Rozansky, Director of Communications

Acronyms and Abbreviations

APPC

The Annenberg Public Policy Center at the University of Pennsylvania

ASAPH

The Annenberg Science and Public Health Multi-Wave Panel Survey

AIOD

The Annenberg Institutions of Democracy Multi-Wave Panel Survey

Overview

The Annenberg Civics Knowledge Survey, first conducted in 2006, focuses on the public's understanding of the Constitution of the United States. Since 2013, the survey results have been released annually for Constitution Day (September 17) and known as the Annenberg Constitution Day Civics Survey. These nationally representative phone surveys have been widely cited as measures of the U.S. public's civics knowledge.

High-quality random digit dialing (RDD) phone surveys, however, have become increasingly difficult to conduct reliably and prohibitively costly for most researchers. In October 2022, *The New York Times* reported that only 0.4% of dials yielded a completed interview (Cohn 2022). As a result, acquiring a random sample of 1,000 U.S. adults would require 250,000 phone calls. When respondents are compensated, as they are in surveys conducted by the Annenberg Public Policy Center, response rates generally improve, as the 3% response rate from the 2022 Civics Knowledge Survey demonstrates. But these response rates are still low enough to raise questions about the representativeness of the survey. While weighting can adjust for observable differences with the target population, other unobserved characteristics that drive these response rates can threaten generalizability.

As a result, survey firms are increasingly moving toward self-administered online surveys, which are completed by respondents without an interviewer's assistance. Major news organizations such as CNN and polling operations such as Pew and Gallup have either added online panels to their survey strategies or replaced RDD surveys altogether. For example, Pew's American Trends Panel (ATP), a nationally representative address-based online panel of over 10,000 adults, serves as the workhorse sample for Pew's research.

Self-administered online surveys create a possible concern for survey research in the form of mode effects – that is, systematically different responses from otherwise equivalent samples caused by the method of survey administration. For example, Pew (2015) found that "[q]uestions about societal discrimination against several different groups…produced large differences, with telephone respondents more apt than Web respondents to say that gays and lesbians, Hispanics and blacks face a lot of discrimination." They also found that web respondents were more likely to give a "very unfavorable" rating to political figures compared to phone respondents. Whether driven by social desirability bias, differential processing, or print vs. aural item presentation, changes in mode can systematically affect survey responses.

Online surveys assessing objective knowledge – such as the Annenberg Civics Knowledge Survey – present one further complication. In this context respondents are able to search for the answer to the question online during the survey. By some estimates, between 15-25% of respondents do so when presented with knowledge questions (see Graham 2023).

The purpose of the APPC Mode Effects project, which was conducted in August 2022, was to [1] assess the degree of information search (sometimes referred to as "cheating") in a Civics Knowledge Survey administered online; [2] explore the costs and benefits of various mitigation strategies for information search; and [3] report the effects of implementing some of these strategies on the 2023 Civics Knowledge Survey.

The Random Digit Dial national probability telephone survey sample was drawn by SSRS, an independent research company. Interviews were conducted among 1,113 U.S. adults in all 50 states from August 2-August 13, 2022. The online survey sample was drawn from a nationally representative probability sample from SSRS's Opinion Panel. A total of 1,584 respondents completed a survey from August 2-August 8, 2022. Both samples were weighted to represent the target U.S. adult population. Excluding demographic questions, 37 identical questions were asked of each sample.

Overall, we find few differences in the **attitudes** respondents express toward the U.S. Supreme Court by mode. Online panelists were no more likely to view the Supreme Court as too powerful and were no less likely to trust the Court to operate in the best interest of the American people. Nor were they more likely to support adding justices to the U.S. Supreme Court or stripping their jurisdiction over controversial topics.

On issues of **knowledge**, however, we find considerable differences by mode. Compared to the 47% of phone respondents who could name all three branches of government, 72% of online panelists were able to do the same. Consistent differences by mode were seen for listing the three branches of government, enumerating protections within the First Amendment, and the effect of a 5-4 Supreme Court decision. Given the stark differences in findings between attitudes and knowledge, we hypothesized that information search constitutes a considerable portion of the overall mode effect.

To test this hypothesis, we conducted a survey experiment on the SSRS national probability online omnibus panel among 1,004 U.S. adults from May 31-June 7, 2023, to explore possible mitigating interventions. Half of respondents were randomly presented with a **pledge** not to use outside sources in answering their questions. We then tracked the **time** respondents spent on each item and embedded **paradata flags**

indicating whether respondents left the webpage during the survey's administration. Finally, we ended the knowledge battery with a **catch** question[1] – an item so difficult we assume that no respondent could respond correctly without searching for the answer.

Overall, we find considerable evidence that the pledge treatment suppressed online searching. Seven in 10 (72%) of the control respondents reported knowing all three branches, compared to 6 in 10 (63%) among those in the pledge condition – a 9% reduction.

We find both costs and benefits associated with the remaining interventions in this experiment. Timers disproportionately affect older respondents. Catch questions underestimate the prevalence of overall information search on other items. Although paradata flags are effective at showing that respondents left the page, they do not indicate why they left – and disqualifying those who fail this test reduces the sample by a quarter, with only small impacts on the overall distribution of responses.

With this knowledge in hand, we added the pledge to the August 2023 wave of the Annenberg Science and Public Health (ASAPH) survey prior to the conventional civics knowledge battery. This approach has the advantage of exposing roughly the same panelists to the same questions a year later. But where they did not receive the pledge in 2022, they were exposed to it in 2023. <u>We find strong evidence suggesting the efficacy of the pledge.</u> Whereas 72% of ASAPH (untreated) panelists could name all three branches in 2022, only 66% could in 2023.

We find similar post-pledge results for the protections of the First Amendment, but less consistent results for the more difficult, conceptual questions pertaining to 5-4 decisions, the final arbitration of constitutionality, and First Amendment protections on social media. These results suggest that the interventions are indeed reducing information search as the items most affected by the treatment are those easiest to search for online.

While this project was motivated by our desire to minimize information search in our online survey, our last substantive analyses explore other mode effects observed in these surveys, particularly the importance of explicit "Don't know" response options in online surveys and the possible effects of various probing protocols for multiple-response items.

---

[1] In what year did the U.S. Supreme Court decide the case *Geer v. Connecticut*? Answer: 1896.

This report concludes by discussing conceptual concerns with measuring civics knowledge in 2023 using self-administered surveys.

2022 Civics Knowledge Survey Mode Effects

We first report the results of three surveys conducted in fall 2022 on the Annenberg Science and Public Health (ASAPH) panel, an SSRS Opinion Panel omnibus, and a random digit dial (RDD) sample. The purpose of these surveys was to assess potential mode effects in online surveys.

All surveys were conducted by SSRS. The first survey was conducted from August 2-8, 2022, among 1,584 panelists from the ASAPH panel. This is a proprietary panel of 1,834 U.S. adults in wave 7.1 owned by the Annenberg Public Policy Center. The second survey occurred from August 2-13, 2022, among a representative sample of 1,113 adults contacted through random digit dialing (RDD, the phone sample). The third data collection occurred from September 9-12, 2022, among a sample of 1,032 panelists from the SSRS Opinion Panel Omnibus (the omnibus sample). All surveys were conducted within a narrow time period, reducing the risk of underlying change in attitudes or knowledge. They were also all conducted by SSRS, minimizing possible house effects.

Demographic Composition

Figure 1 presents the weighted demographic distributions of the samples. **Figure 2** presents the model coefficients predicting an online panelist using both weighted and unweighted data. These results suggest small differences in racial composition – with fewer Asian respondents in phone surveys – and larger imbalances in age, internet usage, and partisanship – with the online sample having older, more partisans, and considerably more frequent internet users. To account for these possibly confounding imbalances, all analyses for differences in attitudes and knowledge will use the weighted data and control for age, education, gender, party identification, racial and ethnic identity, and frequency of internet usage. However, the key conclusions are robust to unweighted specifications.

Figure 1 – Demographic Composition of Samples
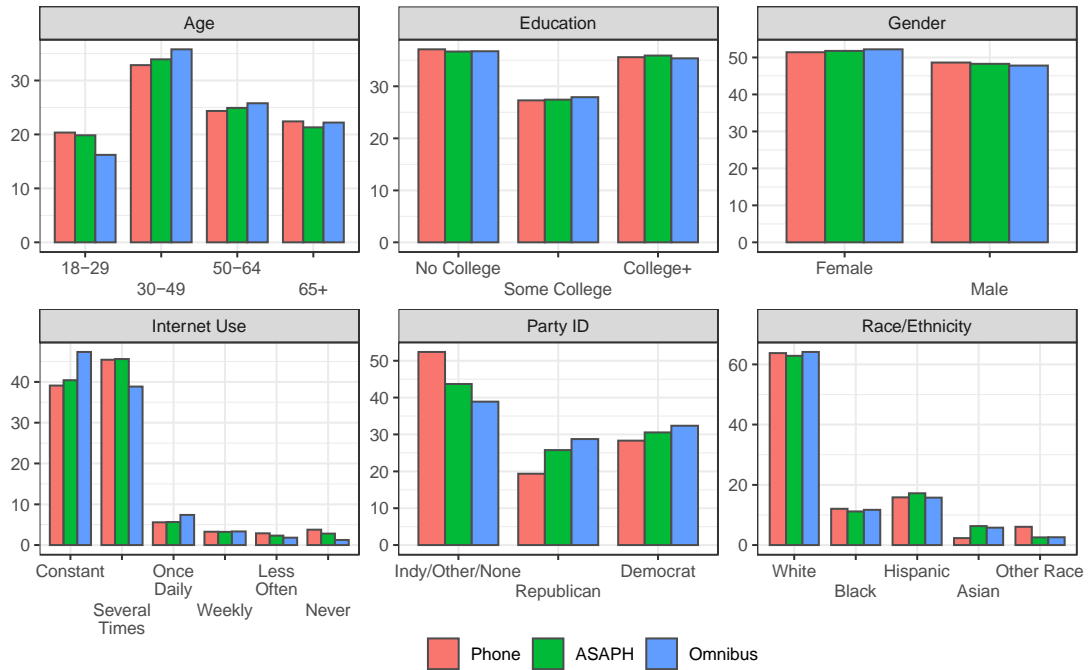
Figure 2 – Predictors of Online Sample Panelist
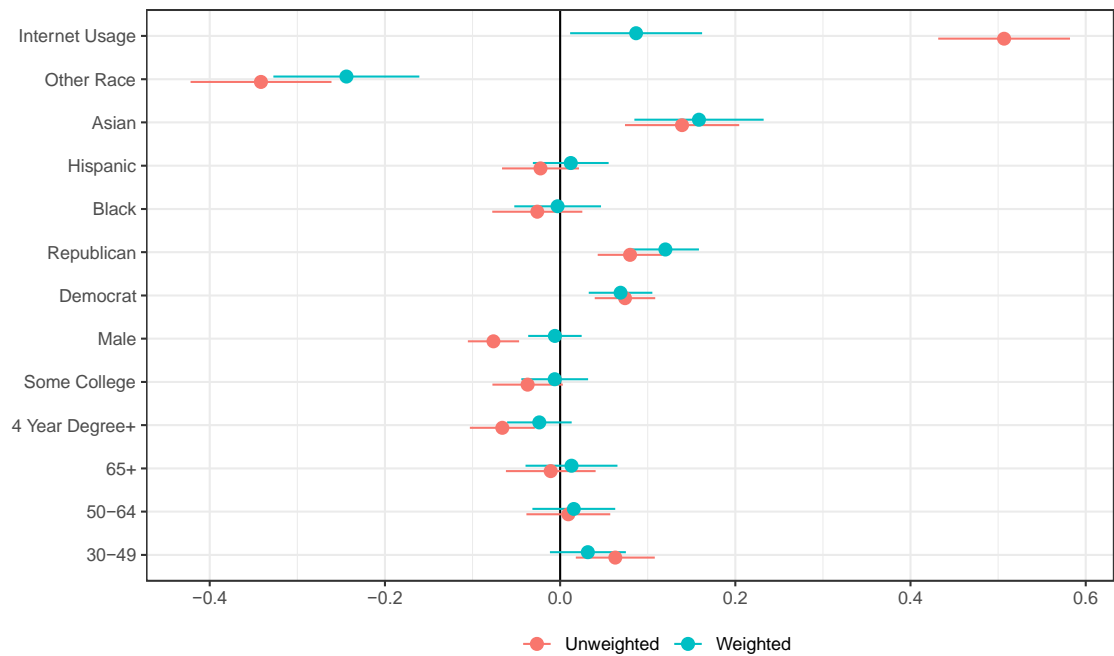
Attitudinal Differences

In both the ASAPH and phone samples we posed a series of questions assessing perceptions of legitimacy of the U.S. Supreme Court. Respondents were asked whether the U.S. Supreme Court had too little, the right amount, or too much power (**Too Much Power**)**,** how much trust they had in the Supreme Court to operate in the best interests of the American people (**Best Interest**), and how strongly they supported increasing the number of justices on the Supreme Court (**Packing**). We also asked how strongly they agreed or disagreed with the following statements:

   a. **Abolish:** If the Supreme Court started making a lot of rulings that most Americans disagreed with, it might be better to do away with the Court altogether.
   b. **Politicians:** Supreme Court Justices are just like any other politicians; we cannot trust them to decide court cases in a way that is in the best interests of our country.
   c. **Mixed Up:** The U.S. Supreme Court gets too mixed up in politics.
   d. **Less Independent:** The U.S. Supreme Court ought to be made less independent so that it listens a lot more to what the people want.
   e. **Strip Jurisdiction 1:** The right of the U.S. Supreme Court to decide certain types of controversial issues should be reduced.
   f. **Strip Jurisdiction 2:** When Congress disagrees with the Supreme Court's decisions, Congress should pass legislation saying the Supreme Court can no longer rule on that issue or topic.

These nine items are coded such that higher values correspond with greater hostility toward the institutional legitimacy of the Supreme Court and are scaled to form an illegitimacy index (α=0.87). **Table 1** predicts this index, as well as its individual components, as a function of survey mode and the above demographics.

We find few systematic differences in the responses of individuals by mode. While phone panelists were slightly more likely to agree with stripping the jurisdiction and making the Supreme Court less independent, they were less likely to support abolishing the Court and viewing the justices as politicians. Moreover, as we will see below, the magnitudes of the effects are much smaller than with knowledge. There are no differences on the other items, nor on the index. Importantly, these models control for demographics, which accounts for the imbalances noted above.

Overall, we find that perceptions of legitimacy for the U.S. Supreme Court are not systematically affected by the survey mode.

Table 1 – Attitudinal Differences by Survey Mode

	Legitimacy Scale	Too Much Power	Best Interest	Pack Court	Abolish Abolish	Politicians Politicians	Mixed Up	Less Independent	Strip Jurisdiction 1	Strip Jurisdiction 2
Mode										
Phone	−0.01 (0.01)	−0.03 (0.02)	0.00 (0.01)	0.01 (0.02)	−0.06*** (0.01)	−0.03* (0.01)	−0.01 (0.01)	0.04** (0.01)	−0.02 (0.01)	0.03* (0.01)
Age										
30-49	−0.05*** (0.01)	−0.09*** (0.03)	−0.04** (0.02)	0.00 (0.02)	−0.09*** (0.02)	−0.02 (0.02)	0.02 (0.02)	−0.07*** (0.02)	−0.06** (0.02)	−0.05** (0.02)
50-64	−0.11*** (0.01)	−0.21*** (0.03)	−0.11*** (0.02)	−0.07** (0.03)	−0.17*** (0.02)	−0.07*** (0.02)	−0.02 (0.02)	−0.15*** (0.02)	−0.09*** (0.02)	−0.09*** (0.02)
65+	−0.10*** (0.02)	−0.18*** (0.03)	−0.13*** (0.02)	−0.09*** (0.03)	−0.15*** (0.02)	−0.05* (0.02)	0.01 (0.02)	−0.12*** (0.02)	−0.10*** (0.02)	−0.05* (0.02)
Education										
4 Year Degree+	−0.09*** (0.01)	−0.13*** (0.02)	0.01 (0.02)	−0.03 (0.02)	−0.11*** (0.02)	−0.09*** (0.02)	−0.04* (0.02)	−0.19*** (0.02)	−0.09*** (0.02)	−0.10*** (0.02)
Some College	−0.05*** (0.01)	−0.07** (0.02)	0.00 (0.02)	−0.05* (0.02)	−0.03* (0.02)	−0.04** (0.02)	−0.03 (0.02)	−0.08*** (0.02)	−0.05** (0.02)	−0.06*** (0.02)
Gender										
Male	−0.07*** (0.01)	−0.15*** (0.02)	−0.04*** (0.01)	−0.06*** (0.02)	−0.06*** (0.01)	−0.03** (0.01)	−0.03* (0.01)	−0.11*** (0.01)	−0.07*** (0.01)	−0.04** (0.01)
Party Identification										
Democrat	0.10*** (0.01)	0.13*** (0.02)	0.11*** (0.01)	0.29*** (0.02)	0.03* (0.02)	0.08*** (0.02)	0.10*** (0.01)	0.07*** (0.02)	0.08*** (0.02)	0.11*** (0.02)
Republican	−0.12*** (0.01)	−0.14*** (0.02)	−0.16*** (0.02)	−0.17*** (0.02)	−0.07*** (0.02)	−0.16*** (0.02)	−0.13*** (0.02)	−0.08*** (0.02)	−0.12*** (0.02)	−0.09*** (0.02)
Race/Ethnicity										
Black	0.06*** (0.02)	0.03 (0.03)	−0.01 (0.02)	0.05 (0.03)	0.10*** (0.02)	0.06** (0.02)	0.00 (0.02)	0.12*** (0.02)	0.08*** (0.02)	0.06** (0.02)
Hispanic	0.04** (0.01)	0.06* (0.03)	−0.08*** (0.02)	0.10*** (0.02)	0.06*** (0.02)	0.02 (0.02)	−0.04* (0.02)	0.09*** (0.02)	0.04* (0.02)	0.10*** (0.02)
Asian	0.01 (0.02)	−0.13** (0.03)	−0.11*** (0.03)	0.12** (0.04)	0.08* (0.03)	0.04 (0.03)	−0.02 (0.03)	0.06 (0.04)	0.03 (0.03)	0.10** (0.03)
Other Race	0.04 (0.02)	0.11* (0.05)	0.00 (0.03)	−0.01 (0.04)	0.06 (0.03)	0.03 (0.03)	0.02 (0.03)	0.06 (0.04)	0.01 (0.03)	0.03 (0.03)
Internet Usage										
Internet Usage	−0.03 (0.02)	0.02 (0.05)	0.03 (0.03)	0.08 (0.04)	−0.08** (0.03)	−0.01 (0.03)	0.01 (0.03)	−0.07* (0.03)	−0.11*** (0.03)	−0.12*** (0.03)
Num.Obs.	2549	2535	2534	1898	2534	2527	2528	2524	2522	2512
R2	0.194	0.107	0.128	0.265	0.114	0.110	0.073	0.146	0.106	0.113
R2 Adj.	0.190	0.102	0.123	0.259	0.109	0.105	0.068	0.141	0.101	0.108

* p < 0.05, ** p < 0.01, *** p < 0.001

Knowledge Differences

In all three surveys from 2022, we asked respondents a series of civics knowledge questions pertaining to the structure and function of U.S. government. Respondents were asked to recall: [1] the individual branches of government (legislative, executive, judicial); [2] five protections enumerated in the First Amendment (speech, religion, press, assembly, and petition); and [3] the meaning of a 5-4 Supreme Court decision. Each response was rescaled (0-1) such that 0 was incorrect/no recall and 1 was correct/complete recall. These three items scale roughly into a single index (α=0.6), however the patterns are present for each of the three items individually.

Table 2 predicts the overall knowledge index as well as its individual components as a function of mode and demographics. Across all models, online panelists perform better than phone samples. With the index, First Amendment, and three branches items, the ASAPH panel also performs better than the omnibus panel. The difference between the ASAPH panelists and the omnibus panelists (both online), however, is much smaller than the differences by mode (online vs. phone). The civics knowledge mode effects are significantly and consistently larger than the largest mode difference found in the attitudinal questions – three to four times larger.

For example, 72% of ASAPH panelists, 66% of omnibus respondents, and 48% of phone panelists could name all three branches of the federal government – an 18-24% difference by mode. Similarly, while only 1% of phone panelists could name all five protections in the First Amendment, 16% of the ASAPH panel and 14% of the omnibus sample could. Interestingly, 15% of ASAPH panelists and 17% of omnibus panelists incorrectly named "the right to bear arms" as a first amendment protection compared to only 9% of phone panelists.

These results suggest two key points: First, there are considerable mode effects for political knowledge items in these surveys – differences that are statistically significant and robust to demographic controls. Second, however, the differences are primarily but not exclusively such that online panelists appear more informed. However, as the misattribution of the Second Amendment's protections to the First Amendment highlights, this is not always the case (for more on this point, see Section 3.3 – Other Mode Effects).

Table 2 – Knowledge Differences by Survey Mode

	Index	Index	3 Branch	3 Branch	1st Amend	1st Amend	5-4	5-4	2nd Amend	2nd Amend
Mode										
Omnibus	0.14*** (0.01)	−0.04*** (0.01)	0.11*** (0.02)	−0.05** (0.01)	0.16*** (0.01)	−0.04*** (0.01)	0.15*** (0.02)	−0.01 (0.02)	0.08*** (0.02)	0.02 (0.01)
Phone		−0.18*** (0.01)		−0.16*** (0.01)		−0.21*** (0.01)		−0.16*** (0.02)		−0.06*** (0.01)
ASAPH	0.18*** (0.01)		0.16*** (0.01)		0.21*** (0.01)		0.16*** (0.02)		0.06*** (0.01)	
Age										
30-49	0.00 (0.01)	0.00 (0.01)	−0.05** (0.02)	−0.05** (0.02)	−0.03* (0.02)	−0.03* (0.02)	0.08*** (0.02)	0.08*** (0.02)	0.03 (0.02)	0.03 (0.02)
50-64	0.05*** (0.01)	0.05*** (0.01)	−0.02 (0.02)	−0.02 (0.02)	0.00 (0.02)	0.00 (0.02)	0.16*** (0.02)	0.16*** (0.02)	0.02 (0.02)	0.02 (0.02)
65+	0.06*** (0.02)	0.06*** (0.02)	−0.01 (0.02)	−0.01 (0.02)	0.00 (0.02)	0.00 (0.02)	0.20*** (0.03)	0.20*** (0.03)	0.02 (0.02)	0.02 (0.02)
Education										
4 Year Degree+	0.20*** (0.01)	0.20*** (0.01)	0.27*** (0.02)	0.27*** (0.02)	0.14*** (0.01)	0.14*** (0.01)	0.18*** (0.02)	0.18*** (0.02)	−0.05*** (0.01)	−0.05*** (0.01)
Some College	0.13*** (0.01)	0.13*** (0.01)	0.19*** (0.02)	0.19*** (0.02)	0.12*** (0.01)	0.12*** (0.01)	0.09*** (0.02)	0.09*** (0.02)	−0.03 (0.02)	−0.03 (0.02)
Gender										
Male	0.09*** (0.01)	0.09*** (0.01)	0.08*** (0.01)	0.08*** (0.01)	0.06*** (0.01)	0.06*** (0.01)	0.14*** (0.01)	0.14*** (0.01)	−0.01 (0.01)	−0.01 (0.01)
Party Identification										
Democrat	0.03** (0.01)	0.03** (0.01)	0.03 (0.01)	0.03 (0.01)	0.01 (0.01)	0.01 (0.01)	0.04* (0.02)	0.04* (0.02)	0.00 (0.01)	0.00 (0.01)
Republican	0.01 (0.01)	0.01 (0.01)	−0.01 (0.02)	−0.01 (0.02)	0.01 (0.01)	0.01 (0.01)	0.02 (0.02)	0.02 (0.02)	0.01 (0.02)	0.01 (0.02)
Race/Ethnicity										
Black	−0.15*** (0.01)	−0.15*** (0.01)	−0.12*** (0.02)	−0.12*** (0.02)	−0.11*** (0.02)	−0.11*** (0.02)	−0.22*** (0.02)	−0.22*** (0.02)	0.01 (0.02)	0.01 (0.02)
Hispanic	−0.08*** (0.01)	−0.08*** (0.01)	−0.09*** (0.02)	−0.09*** (0.02)	−0.07*** (0.02)	−0.07*** (0.02)	−0.10*** (0.02)	−0.10*** (0.02)	−0.02 (0.02)	−0.02 (0.02)
Asian	−0.16*** (0.02)	−0.16*** (0.02)	−0.21*** (0.03)	−0.21*** (0.03)	−0.10*** (0.03)	−0.10*** (0.03)	−0.17*** (0.04)	−0.17*** (0.04)	0.01 (0.03)	0.01 (0.03)
Other Race	−0.02 (0.02)	−0.02 (0.02)	0.00 (0.03)	0.00 (0.03)	0.04 (0.03)	0.04 (0.03)	−0.11** (0.04)	−0.11** (0.04)	−0.04 (0.03)	−0.04 (0.03)
Internet Usage										
Internet Usage	0.25*** (0.02)	0.25*** (0.02)	0.32*** (0.03)	0.32*** (0.03)	0.19*** (0.03)	0.19*** (0.03)	0.23*** (0.04)	0.23*** (0.04)	−0.01 (0.03)	−0.01 (0.03)
Num.Obs.	3518	3518	3505	3505	3518	3518	3512	3512	3518	3518
R2	0.258	0.258	0.189	0.189	0.156	0.156	0.149	0.149	0.015	0.015
R2 Adj.	0.255	0.255	0.185	0.185	0.152	0.152	0.146	0.146	0.011	0.011

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Mitigating Mode Effects in Online Surveys

We assume that the large knowledge gaps and the small, inconsistent differences in attitudes that we found are mode effects primarily driven by internet-search. As Graham (2022) notes: "The shift to online surveys has made it easier for respondents to look up the answers to questions designed to test factual knowledge (Clifford and Jerit 2014; Liu and Wang 2014; Strabac and Aalberg 2011; Shulman and Boster 2014)."

To address this phenomenon, researchers have developed several methods to **detect** and **deter** information search. Search can be **detected** using *self-reports* (Jensen and Thomsen 2014), *catch questions* – questions too difficult to answer correctly without looking them up (Bullock et al. 2015; Motta et al. 2016) – or *paradata flags* that can detect and track when respondents leave the survey page (Diedenhofen and Musch 2017). In a series of online experiments, Graham (2022) finds that combining detection and deterrence methods (pledge, paradata, and catch questions) made it possible to "deter or detect more than 90% of search, leaving search to affect about 0.5% of the remaining observations."

Deterrence Methods

In Clifford and Jerit's (2016) survey experiments, the most effective deterrent of outside search was a commitment pledge to not use outside resources. Prior to asking respondents the knowledge questions, they provided the following prompt:

- **Commitment**: "It is important to us that you do NOT use outside sources like the Internet to search for the correct answer. Will you answer the following questions without help from outside sources?" (Yes, No)

These authors found this to be the most effective intervention, and critically it did not reduce either the respondents' satisfaction with the survey or their enthusiasm for completing future surveys.

In another condition, they provided a timer condition:

- **Timer:** "Please do NOT use outside sources like the Internet to search for the correct answer. You will have 30 seconds to answer each question."

However, a handful of respondents in the timer condition mentioned feeling frustrated with the timers in their post-survey reactions. It was not as effective at reducing information search as the pledge and it was more intrusive.

An additional detection strategy gathers the length of time respondents spend per item as paradata in order to flag outliers. This move assumes that those who take longer on an item are more likely to have searched for the answers to the questions elsewhere. However, item-response times may be correlated with important demographics, which could bias any results. Moreover, it could be that thinking deeply about an item takes longer than searching, making timers a poor indicator of information search.

Detection Methods

There are many ways to detect information search in online surveys. Researchers can embed **paradata flags** that record when respondents leave the web page. Another common method is use of **catch questions**, e.g., In what year was *Geer v. Connecticut* decided? Those individuals who can correctly answer this question can be safely assumed to have sought outside sources.

It's important to note that these methods do not detect information search directly. A respondent's paradata would be flagged if they searched for the names of the three branches of government or if they checked their e-mail. And while the number of individuals fascinated with the politics of 19th century interstate wild fowl transportation (the issue at play in *Geer*) is likely small, it isn't zero. Moreover, searching for the answer to an impossible question may not be a perfect indicator of search behavior on other, less difficult items.

These methods also raise the question of what to do with those flagged respondents. As Smith, Clifford, and Jerit (2020) note, "[d]ropping respondents from the analysis (e.g., those who engage in outside search) may harm the representativeness of the sample and should be discouraged, especially if knowledge is measured post-treatment in an experimental design (Montgomery, Nyhan, and Torres 2016)." Another solution would be to control for information search in statistical models. But Smith, Clifford, and Jerit also show that "controlling for information search did little to reduce the biases." The optimal strategy, of course, is to discourage information search before it happens, but this is more difficult to guarantee.

2023 Information Search Experiment

To test the effectiveness of these various interventions, APPC contracted with SSRS to conduct a stand-alone survey of 1,004 respondents on the SSRS Opinion Panel in May 2023. This survey included several civics knowledge items as well as measures assessing attitudes toward the Supreme Court. Half of respondents were assigned to a pledge condition:

- **PLEDGE:** We want to see how much information about government gets out to the public from television, newspapers and the like. It is important to us that you do **NOT** use outside sources like the Internet to search for the correct answer. Will you answer the following questions without help from outside sources?
- 99% of treated respondents chose "Yes."

All respondents were then asked a series of civics knowledge questions. For these knowledge items, we embedded item-level timers and paradata flags – a count of the number of times respondents left the web browsing tab containing the survey.

Finally, we provided an open-ended textbox catch question:

- In what year did the U.S. Supreme Court decide *Geer v. Connecticut*?

We assumed that no one – beyond the few truly fascinated with 19th century interstate transportation of wild fowl – would know this case without googling. We apply these mitigation strategies to assess the level of search present in our online samples.

Figure 3 summarizes the share of respondents flagged by each method. Since roughly 20% of respondents were flagged in either the three branches or First Amendment items, we set the threshold for outliers on time at the top 20th percentile, to flag a roughly equivalent portion of the total sample.

Consistent with previous research, the pledge has a significant impact on the prevalence of behaviors associated with search. Only 3% of respondents in the pledge condition correctly answered the catch question, compared to 8% in the control. The number of respondents flagged by the paradata trackers was nearly double across multiple items in the control condition. This evidence is consistent with previous findings showing that a pledge can reduce information search by half. However, there is less evidence to suggest that the pledge affected the amount of time respondents spent on each item.

Figure 3 – Share of Respondents Flagged by Paradata & Catch Questions

Share of Respondents Flagged by Mitigation Strategy
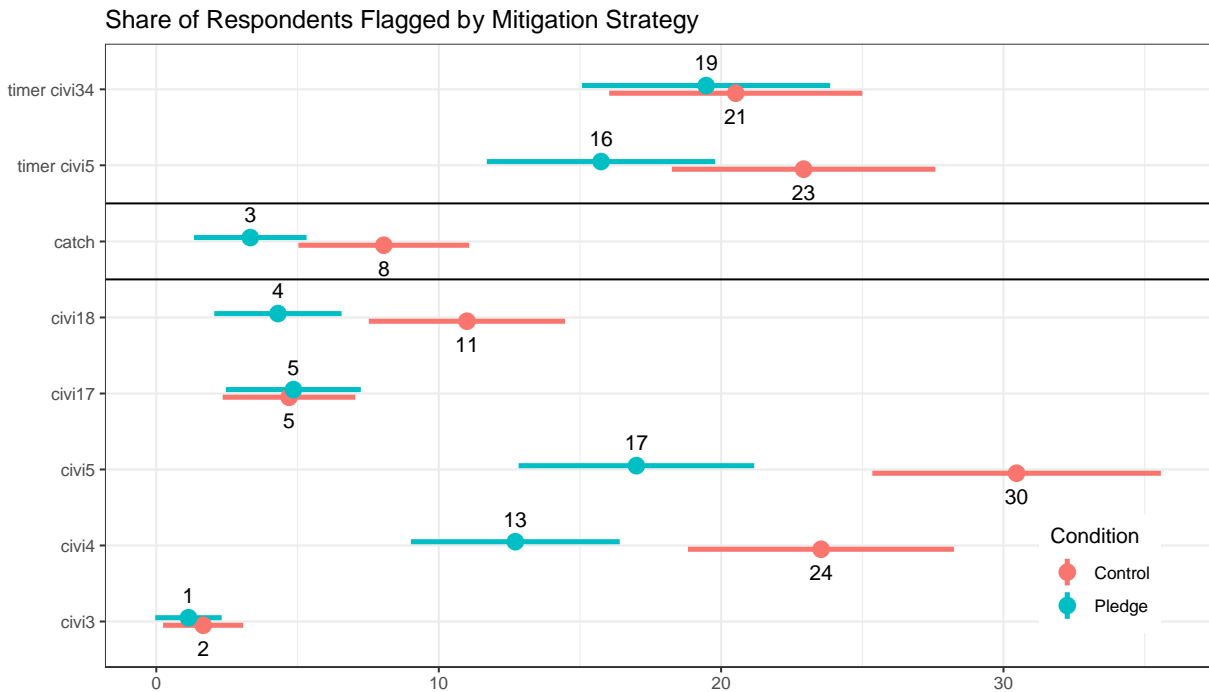
How much are our estimates improved by these strategies? **Figure 4** provides a visual example of these mitigation strategies in action. The first panel shows the share of respondents correctly identifying the three branches of government: legislative, executive, and judicial. We can see that among the full sample, the pledge had a 9% effect – 72% of those untreated correctly identified all three compared to 63% in the treated condition. Each subsequent pair of columns subtracts respondents who were flagged by a particular mitigation strategy. First, a very small number of respondents did not agree to the **Pledge**. Second, we filtered out those respondents who correctly responded to the **Catch** question. Third, we removed those who left the page multiple times when answering a question (**Para (2+)**). And finally, as our most stringent filter, we remove anyone who left the browser at any point during either the initial probe or recall request (civi3 and civi4, **Para (1)**).

A few important findings can be seen in **Figure 4**. First, we omitted the timers from the analysis. Timers proved too biased a measure of search. Counter to expectations, those who took *longest* on the survey's three branches item were not most likely to recall three branches (see **Table 3**). In the treated condition, the average response time for an individual identifying all three branches was 57 seconds, compared to 75 seconds for

those who could recall two branches. In the untreated condition, the average response time for an individual identifying all three branches was 51 seconds, compared to 61 seconds for those who could recall two branches. Similarly, older respondents tend to take longer than younger respondents in both conditions. This may suggest that thinking about an item takes longer than searching for the answer online. Due to uncertainties around age and findings counter to expectations, we caution against using timers to correct for information search.

Figure 4 – Effects of Mitigation Strategies on Knowledge of Three Branches
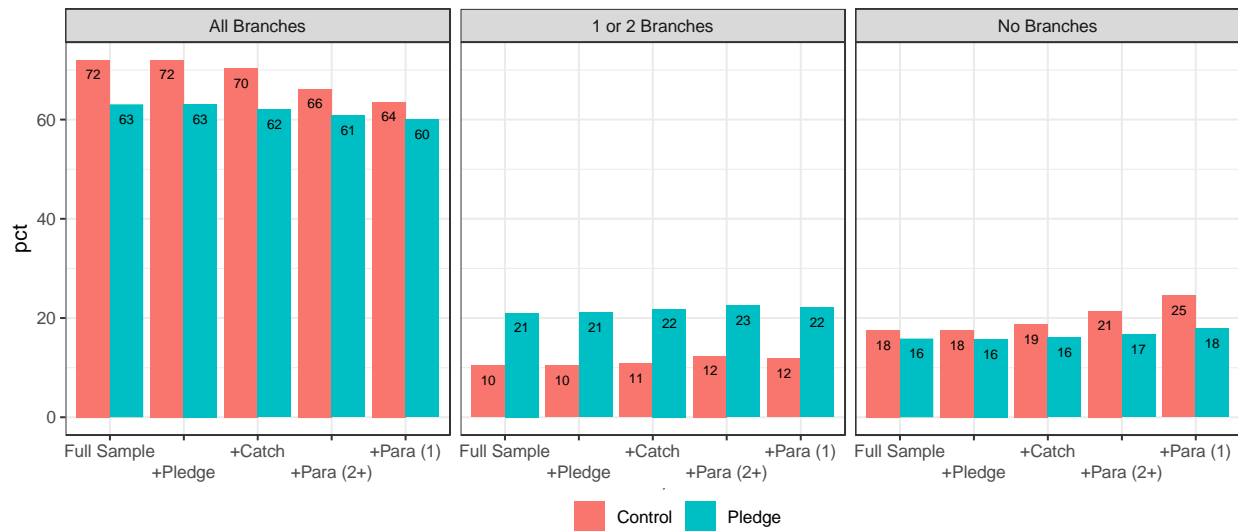
Table 3 – Average Response Times to CIVI4 by Knowledge and Age

	0 Branches	1 Branch	2 Branches	3 Branches
Pledge Condition	24.6	35.0	74.9	56.7
Control Condition	16.5	33.2	61.0	51.0
	18-29	**30-49**	**50-64**	**65+**
Pledge Condition	51.2	39.1	65.7	68.6
Control Condition	32.1	46.1	55.9	54.9

Note: Time measured in seconds.

Second, the mitigation strategies have significantly less impact on the treated condition than on the control condition: 72% of the control condition reported knowing all three branches, compared to 64% (**-8%**) after excluding all flagged respondents. In the control

condition, however, only 63% reported knowing all three, and 60% (**-3%**) after exclusions.

Third, across all interventions, there are significantly fewer respondents in the control condition naming some (1 or 2) of the three branches. This suggests that those who are searching for information are more likely to be those who could, upon reflection, name some, but not all, of the branches.

Fourth, the pledge condition alone has a comparable substantive effect on the topline result to all mitigation efforts combined. 63% of the full treated sample could recall all three branches, compared to 64% among the control sample with the catch question and paradata. But this omits the cost – applying these three filters reduces the sample size of our survey by 244 respondents, or 24% of the sample.

Which respondents are being caught by these particular filters? **Figure 5** provides two alluvial plots displaying the changing composition of the sample after each filter. Here we can again see the effect of the treatment pledge. Only 17% of respondents are flagged after applying all filters, compared to 32% in the control condition (**-15%**), effectively cutting in half the prevalence of search-suggestive behavior. We can also see that at each step, the vast majority of those flagged for possible information search are those who had accurately named all three branches of government.

Table 4 shows the effect of the pledge on the knowledge index described earlier after each successive mitigation strategy – highlighting the fact that the results are robust to demographic controls. In the model diagnostics at the bottom of the table, we also report the size of the treated and control groups, and the proportion of those respondents remaining after surviving each subsequent mitigation strategy is employed. Also of note is the improved model fit across each specification – demographics explain a greater proportion of the variation in civics knowledge among the filtered sample, despite the 24% reduction in sample size.

Are there demographic commonalities among these flagged respondents? Not among the usual suspects. Neither age, education, gender, party identification, internet usage, nor racial/ethnic identity are significant predictors of being flagged by catch question or paradata trackers. The results of these models can be found in the **Appendix**.

While the catch questions and paradata flags help detect potential search behavior, are they worth the cost to sample size? To answer this question, we now apply these results to the 12th wave of the ASAPH panel.

Figure 5 – Effects of Mitigation Strategies on Sample Composition

Effect of Mitigation Strategies on Composition of Control Group

Effect of Mitigation Strategies on Composition of Treatment Group
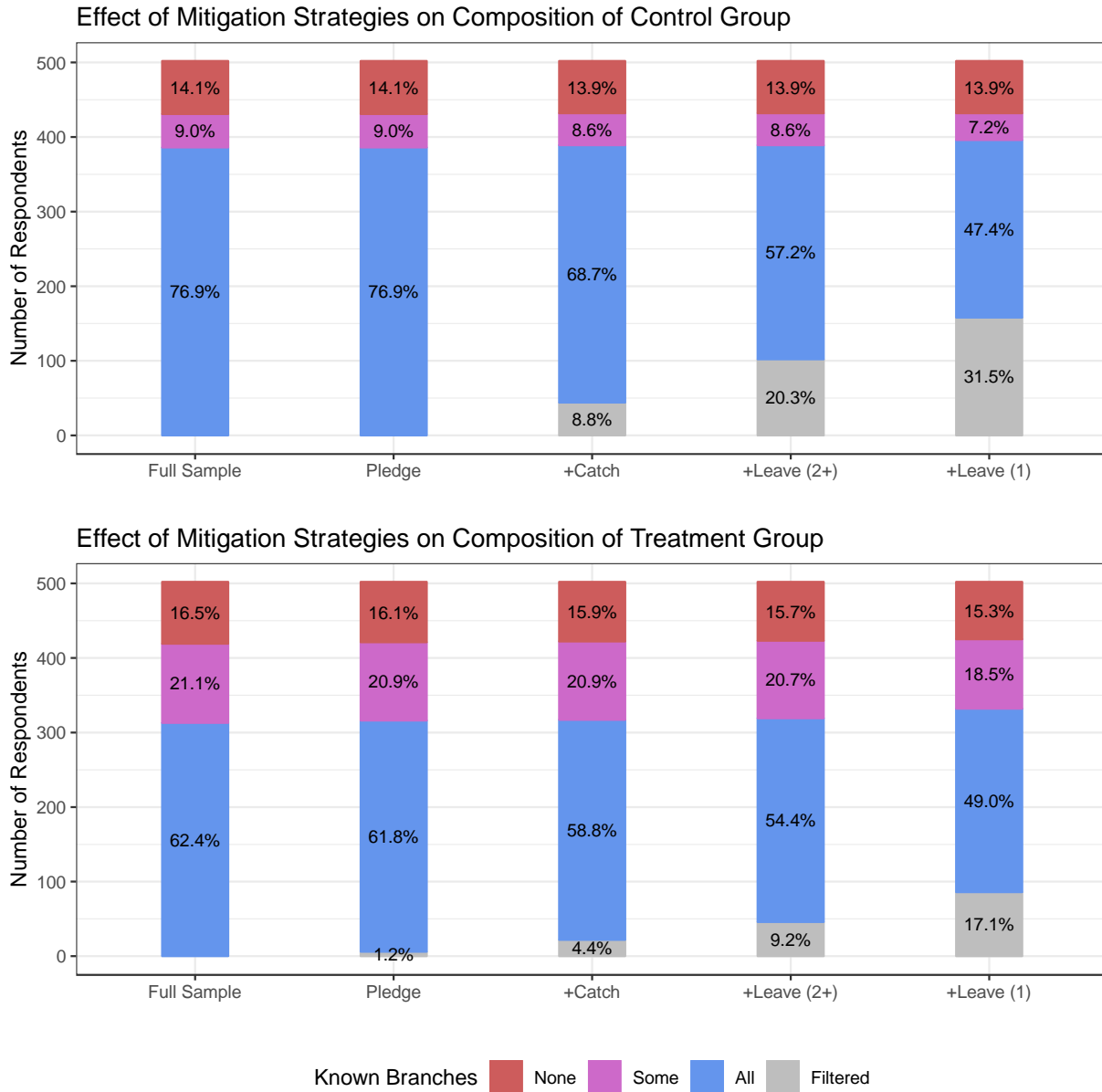
Table 4 – Effects of Mitigation Strategies and Pledge Treatment on Knowledge

	Full Sample	Pledge	+Catch	+Leave (2+)	+Leave (1)
Pledge	-0.05**	-0.05**	-0.04*	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
30-49	-0.01	-0.01	-0.01	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
50-64	0.10***	0.10***	0.08**	0.08**	0.11***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
65+	0.09**	0.09**	0.08**	0.09**	0.10**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
4 Year Degree+	0.23***	0.23***	0.25***	0.26***	0.29***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Some College	0.16***	0.16***	0.18***	0.18***	0.18***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Male	0.10***	0.09***	0.10***	0.10***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Democrat	-0.03	-0.03	-0.02	-0.01	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Republican	-0.07***	-0.07***	-0.07**	-0.07**	-0.07**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Black	-0.16***	-0.16***	-0.19***	-0.19***	-0.17***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Hispanic	-0.10***	-0.10***	-0.10***	-0.11***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Asian	-0.12**	-0.12**	-0.12***	-0.12**	-0.11**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Other Race	-0.07	-0.07	-0.14*	-0.13*	-0.12
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)
Internet Usage	0.10*	0.10*	0.06	0.06	0.10
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Num.Obs.	945	940	884	806	715
Cntrl—Tr	502 — 502	100% — 99%	91% — 96%	80% — 91%	69% — 83%
R2	0.251	0.251	0.270	0.282	0.304
R2 Adj.	0.240	0.240	0.258	0.270	0.290

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The Cntrl-Tr numbers show the size of the control and treated groups in column 1, and then the share of those respondents remaining after each filter is applied.

2023 Civics Knowledge Survey

In August 2023, we included the standard civics knowledge battery at the beginning of the twelfth wave of ASAPH – a one-year follow-up from their exposure to the original ASAPH civics knowledge panel. Theoretically, a year gap between surveys should eliminate any potential sensitization for these particular questions. Notably, while many online panels (like the SSRS Opinion Panel) participate in a range of surveys, ASAPH panelists are proprietary and have not participated in any online panels pertaining to civics, government, or politics.
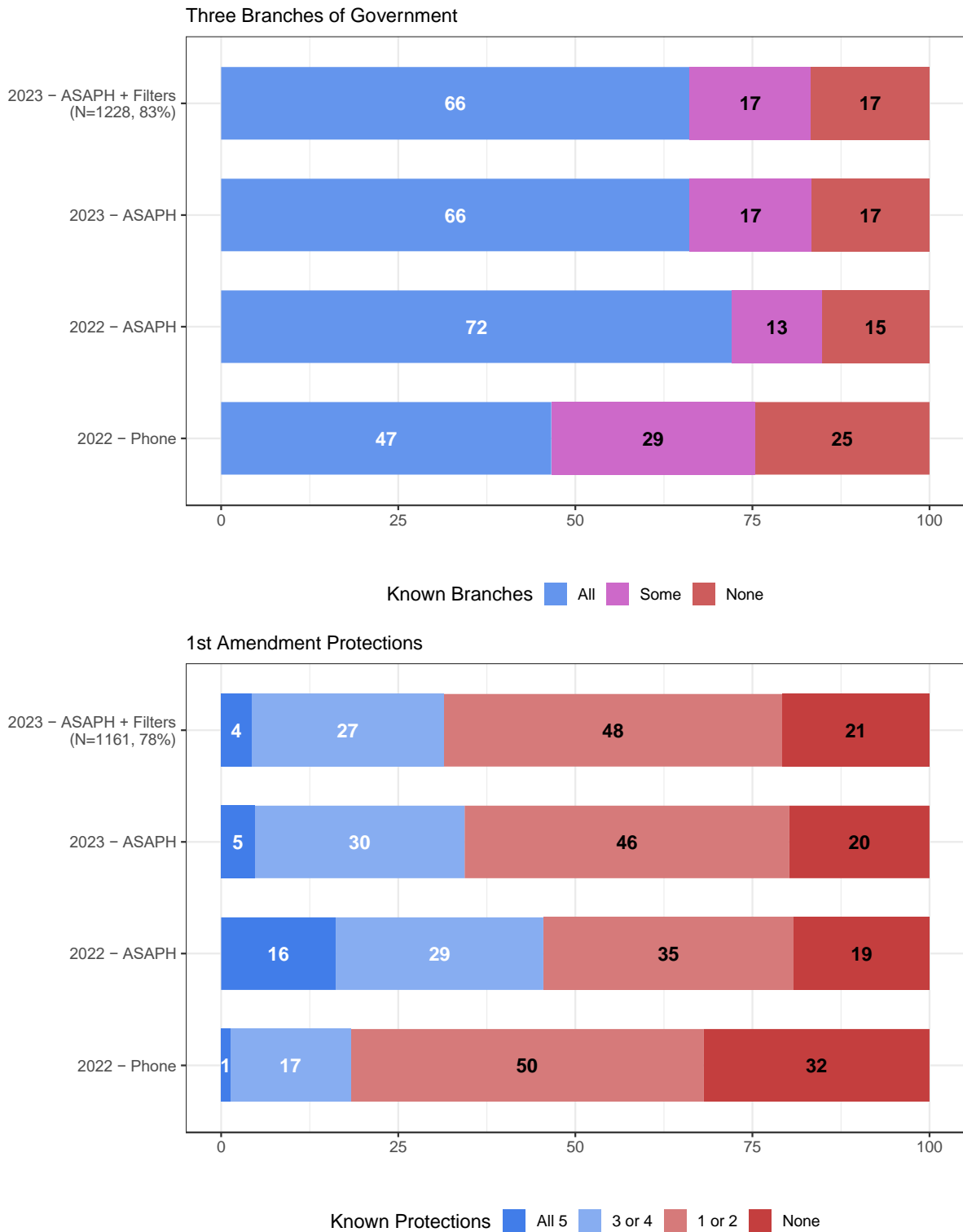
All respondents were provided the same pledge as in the experimental condition above. All civics knowledge questions were embedded with paradata trackers. We did not include a catch question due to both space constraints and worry that a difficult question may affect respondents' likelihood of information search on subsequent public health knowledge items in the ASAPH survey.

Main Results

Figure 6 presents a comparison between First Amendment and three branches civics knowledge items asked on both the 2022 and 2023 ASAPH panel. We again embedded paradata trackers in these two items as they had the highest flag prevalence in the summer 2023 experiment. The first row of each panel shows the topline results for the 2023 ASAPH panel once those who rejected the pledge or left the page are excluded, whereas the second row is the entire 2023 ASAPH sample. As comparisons, row three provides the untreated/unpledged 2022 ASAPH and row four the 2022 RDD samples.

Even after treatment and mitigation filters are applied, there are still notable mode differences for the three branches and First Amendment items. Where only 48% could name all three branches on the phone in 2022, 66% of respondents could online in 2023. Where only 1% could name all five First Amendment protections over the phone, 4% and 5% could in the filtered and unfiltered 2023 sample, respectively.

Figure 6 – Three Branches and First Amendment Protections by Survey Mode

Three Branches of Government

Known Branches   ■ All   ■ Some   ■ None

1st Amendment Protections

Known Protections   ■ All 5   ■ 3 or 4   ■ 1 or 2   ■ None

Composition of Flagged Respondents

When deciding whether we should include or exclude the flagged panelists, it's worth considering who would be excluded. Unlike the omnibus experiment there are some demographic predictors of being flagged among the ASAPH panelists. **Figure 7** presents the model coefficients predicting whether a respondent was flagged either through failure to agree to the pledge or tripping the paradata alert for leaving the page. Recall that the catch question was not included in ASAPH 2023 to avoid affecting subsequent questions (see Graham n.d.).

Figure 7 – Predicting Flagged Respondents in ASAPH 2023
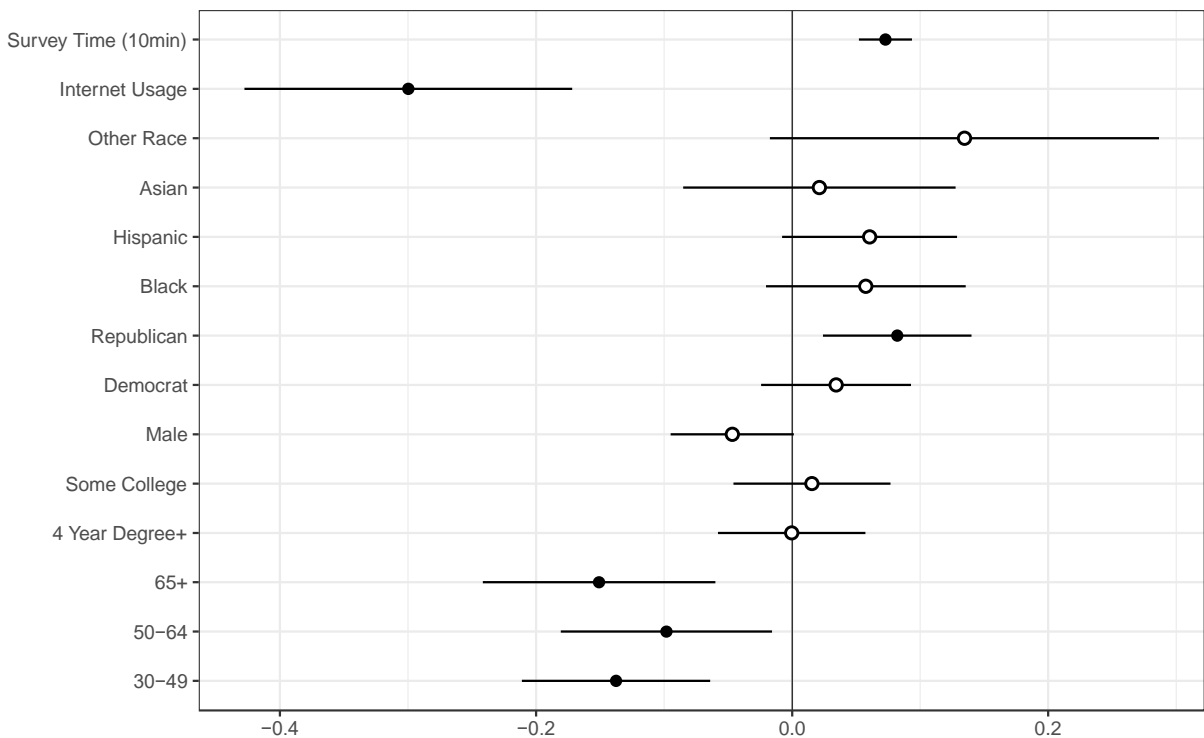
What we can see from this figure is that those who took longer to complete the entire ASAPH wave (not just the civics module) and Republicans were more likely, overall, to be flagged, whereas older respondents and those with the highest degree of internet usage were less likely. This raises possible concerns for bias in paradata tracking. Perhaps younger respondents more commonly multitask while using the internet and innocuously jump between tasks more frequently. Perhaps those with greater internet savvy are better able to avoid these trackers or are more likely to be using the internet on two devices simultaneously.

Other Mode Effects

Figure 8 highlights another form of mode difference to consider going forward. Consider the two items:

CIVI17. If the president and Supreme Court differ on whether an action by the president is constitutional, who has the final responsibility for determining if the action is constitutional- (the president), (Congress), or (the Supreme Court), or are you not sure?

1   President
2   Congress
3   Supreme Court **[CORRECT]**
8   Not sure/Don't know
999   **[PN: IF CATI:]** (DO NOT READ) Refused
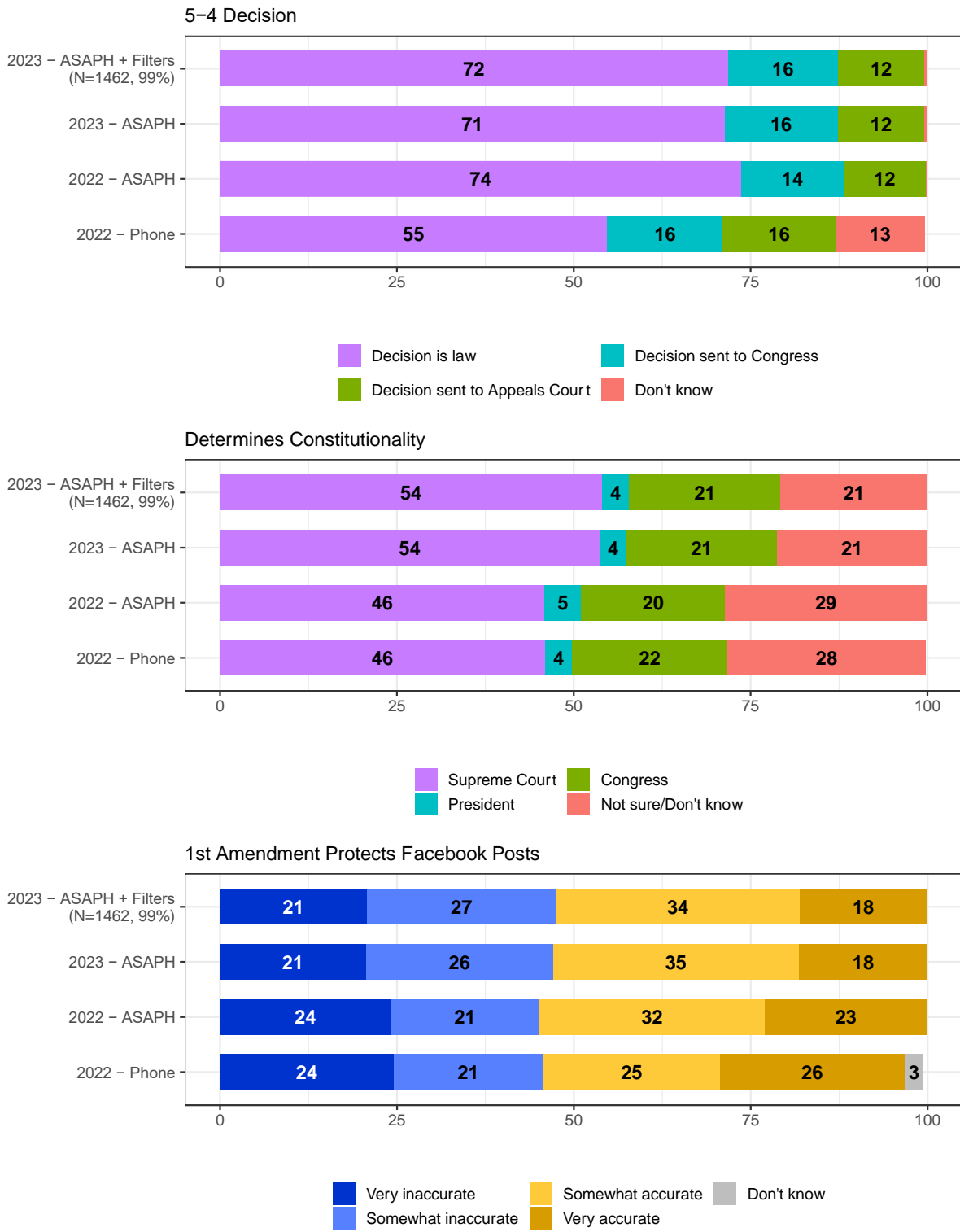999   **[PN: IF WEB:]** Decline/Web blank

CIVI18. If the U.S. Supreme Court rules on a case 5 to 4, does this mean...

[PN: IF CATI: [READ LIST]**]**

1   The decision is the law and needs to be followed **[CORRECT]**
2   The decision is sent back to Congress for reconsideration
3   The decision is sent back to the federal court of appeals to be decided there
998   **[PN: IF CATI:]** (DO NOT READ) Don't know
999   **[PN: IF WEB:]** Decline/Web blank

For determining constitutionality, an explicit "don't know" option is provided. For 5-4 decisions, the option is not explicitly provided. Now consider the distribution of "don't know" responses across the four samples in **Figure 8**. The numbers are comparable by year in the explicit condition, but very different in the omitted condition. Where a negligible portion of respondents skip a forced choice item online, 13% of respondents *volunteered* "don't know" over the phone. While the context of the research should determine whether to include an explicit "don't know" response, it is important to consider that without one, comparisons across mode are even further from apples-to-apples.

Figure 8 – Civics Knowledge by Survey Mode

5−4 Decision

Survey Mode	Decision is law	Decision sent to Congress	Decision sent to Appeals Court	Don't know
2023 − ASAPH + Filters (N=1462, 99%)	72	16	12	
2023 − ASAPH	71	16	12	
2022 − ASAPH	74	14	12	
2022 − Phone	55	16	16	13

- Decision is law
- Decision sent to Congress
- Decision sent to Appeals Court
- Don't know

Determines Constitutionality

Survey Mode	Supreme Court	President	Congress	Not sure/Don't know
2023 − ASAPH + Filters (N=1462, 99%)	54	4	21	21
2023 − ASAPH	54	4	21	21
2022 − ASAPH	46	5	20	29
2022 − Phone	46	4	22	28

- Supreme Court
- President
- Congress
- Not sure/Don't know

1st Amendment Protects Facebook Posts

Survey Mode	Very inaccurate	Somewhat inaccurate	Somewhat accurate	Very accurate	Don't know
2023 − ASAPH + Filters (N=1462, 99%)	21	27	34	18	
2023 − ASAPH	21	26	35	18	
2022 − ASAPH	24	21	32	23	
2022 − Phone	24	21	25	26	3

- Very inaccurate
- Somewhat inaccurate
- Somewhat accurate
- Very accurate
- Don't know

Responses to the First Amendment item suggest another possible mode effect. In 2022, 15% of panelists believed the right to "bear arms" or "own a firearm" was found in the First Amendment (see **Figure 9**). When those same panelists were asked not to search for answers, that figure rose to 22%, consistent with the conclusion that the pledge reduced outside search and therefore decreased overall recorded knowledge. However, in both the treated and untreated ASAPH wave conditions, and in both the treated and untreated omnibus respondents conditions (23% and 16%, respectively), far more respondents provided this incorrect response compared to those surveyed on the phone (9%). These differences are robust to demographic controls (see **Table 2**).

Figure 9 – Individual First Amendment Protections

Knowledge of 1st Amendment Protections
(% of respondents who named each protection)
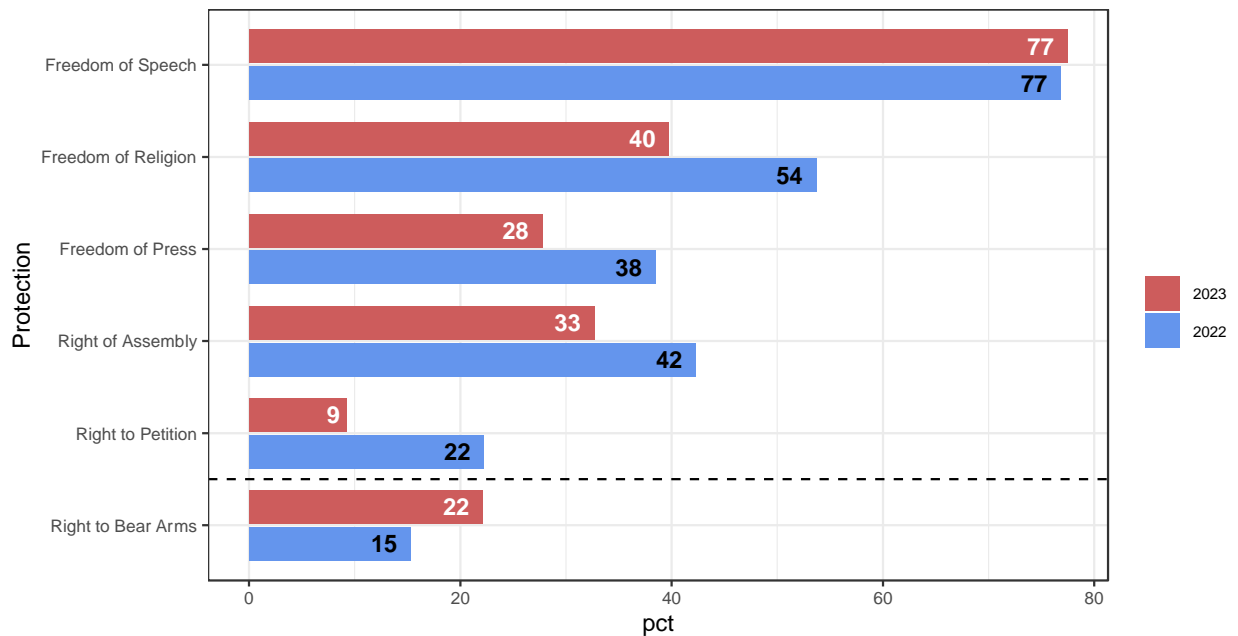
Source: ASAPH Survey, Aug. 2023
N=1482, MOE=+/−3.5%
©2023 Annenberg Public Policy Center

Why are respondents online more likely to give this incorrect response? One hypothesis is the mode. Consider how this item is administered:

CIVI5. As you may know, the First Amendment is part of the U.S. Constitution. Can you name any of the specific rights that are guaranteed by the First Amendment of the Constitution, or not?

> **[PN: IF CATI:** IF YES: Would you mind naming them?**]**

> **[PN: IF CATI:** PROBE TO THE NEGATIVE: Any others?**]**

> **[PN: IF CATI:** (DO NOT READ LIST)**]**

Phone administrators (CATI) probe until individuals say they don't know any others (up to five), but respondents are never provided with the number 5. They don't know how many rights they are "supposed" to know. In contrast, web panelists see five empty text boxes and receive:

> **[PN: IF WEB:** *(Please type each First Amendment right you know of in the text boxes below, up to 5)*

In other words, online panelists have reason to assume that the correct number of protections is five. This could encourage increased guessing, and when individuals guess, they pull protections from elsewhere in the Bill of Rights. Subsequent analyses will test these hypotheses systematically, but for the purposes of this report, the finding reiterates the important differences subtle changes by mode can have on response outcomes.

Discussion

Recommendations

The key recommendation implemented as a result of this analysis is inclusion of the pledge against outside information search. In both the literature (see Graham 2023) and this experiment, the pledge alone can reduce information search by nearly half. We find no evidence of a backlash effect, e.g., for no items did search-suggestive behavior increase in the face of the pledge. In our experiment, all respondents exposed to the pledge continued through to answer the final item on the survey, suggesting that the pledge did not affect the respondents' willingness to participate in the survey, a finding consistent with past research (Clifford and Jerrit 2016).

As for the individual mitigation efforts, the evidence is decidedly more mixed and warrants consideration of the costs and benefits of each method. Timers appear to be noisy and possibly biased against particular demographic groups. Catch questions under-identify search overall and only roughly predict particular search-suggestive behaviors. Other scholars have also found a backlash effect to catch questions – that is, by asking respondents a question too hard to be reasonably known, they are encouraged to search the answer to that question AND subsequent questions (Graham n.d.). Paradata filters appear effective at reducing the differences between treatment and control in our experiments, but do so at the expense of an unacceptably large portion of our sample. And when applied to our ASAPH panel, these flags were biased against demographic groups in a manner that could reflect non-search behavior.

Beyond Detection and Deterrence

Another solution to the problem of information search is to design survey questions that are more difficult to answer online. For example, Kleinberg (2022) compares the mode effects for survey questions based on the ease of finding the answers in a search engine (i.e., those whose answers appear as the first search and in the title vs. those that would require multiple links to answer, e.g., "Who is the Chief Justice of the Supreme Court?" vs. "On which of the following policy areas does the U.S. spend the least?") and finds a smaller mode effect on more difficult to search questions. Similarly, DeBell (2022) demonstrates that asking people to identify pictures of political figures produces smaller differences by mode than recalling names of officeholders.

Evidence for this approach can be seen in **Figure 8**. Consider CIVI17 – who determines the constitutionality of an action when the President and Supreme Court disagree? This item is more conceptual (e.g., more difficult to Google), and has the smallest mode effects in 2022, has the smallest differences between treated and untreated conditions between waves, and is only minimally affected by the filter.

Also consider the following item:

CIVI9A. How accurate is it to say that the First Amendment's protection of freedom of speech means that Facebook must permit all Americans to freely express themselves on Facebook pages?

[PN: IF CATI: Is that... [READ LIST]**]**

1	Very accurate
2	Somewhat accurate
3	Somewhat INaccurate **[CORRECT]**
4	Very INaccurate **[CORRECT]**
998	**[PN: IF CATI:]** (DO NOT READ) Don't know
999	**[PN: IF CATI:]** (DO NOT READ) Refused
999	**[PN: IF WEB:]** Decline/Web blank

As shown in the last panel of **Figure 8**, here too we see relative stability year and mode. We again see the effect of non-explicit "don't know" options (albeit much smaller than in the case of 5-4 decisions).

These solutions are effective and worth pursuing but would change the outcomes of interest and limit our ability to compare results with those generated in the past decades of the survey's existence. Yet as we consider designing new items it is worth considering the difficulty of the question and how difficult it would be to Google the answer.

Conclusion

The differences we find between modes invite the question, what sorts of recall of foundational constitutional knowledge are of greatest use to the public. One might make the case that aural requests to recall do not approximate at least some

circumstances in which such knowledge is needed or useful. Where such recall might be useful in a jury room, it is less so when making sense of a printed ballot.

These results also raise the question – what does it mean to be an informed citizen in 2023? Relying on the works of Berelson, Lazersfeld, and McPhee (1954), Barber (1969), and Neuman (1986), Delli Carpini and Keeter (1993, 1996), we have answered this question by tracking knowledge of "what government is and does" – "is" being the structure of government, e.g., federalism, separation of powers, civil liberties, the party system, and "does" focusing on the history, facts, and alternatives of ongoing policy debates. Few would argue that these are not the foundational tenets.

However, operationalizing these concepts into questions may require changes with both time and mode. For example, in their classic work, Delli Carpini and Keeter (1993) include knowledge of the veto override threshold in their knowledge index. On the heels of Reagan's nine veto overrides and over three decades of Democratic majorities and supermajorities in Congress, this could indeed be a relevant piece of information. However, there have only been nine overrides total since this index was published. In today's evenly divided, deeply polarized Congress, veto overrides are less relevant to the workings of government.

Missing from this initial index is any inquiry into the mechanics of the filibuster. This is understandable: the 1991-1992 Congress had fewer than 50 cloture votes. But no major piece of legislation in today's Congress can ignore the threat of filibuster, with 298 cloture votes in 2019-2020 and 289 in 2021-2022. In the first 2024 Republican primary debate, former Governor Nikki Haley (R-SC) criticized former Vice President Mike Pence's position on abortion, stating: "Don't make women feel like they have to decide on this issue when you know we don't have 60 Senate votes." Understanding the cloture threshold is arguably far more relevant to informed citizenship and representational accountability today than the veto override.

This is a small, perhaps unrepresentative, example. Yet, as we transition to a new mode for assessing the public's civics knowledge, it is important to consider not just how we should adapt our items to these new modes, but what types of knowledge would best inform good, discerning citizenship.

References

Barber, David. 1969. *Citizen Politics*. Markham Press.

Berelson, Bernard, Paul Lazarsfeld, and William McPhee. 1954. *Voting*. University of Chicago Press.

Bullock, John , Alan Gerber, Seth Hill and Gregory Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10:1–60.

Clifford, Scott and Jennifer Jerit. 2014. "Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies." *Journal of Experimental Political Science* 1: 120–131.

Clifford, Scott, and Jennifer Jerit. 2016. "Cheating on political knowledge questions in online surveys: An assessment of the problem and solutions." *Public Opinion Quarterly* 80: 858-887.
> https://doi.org/10.1093/poq/nfw030

Cohn, Nate. 2022. "Who in the World is Still Answering Pollsters' Phone Calls?" *The New York Times*. October 12, 2022.
> https://www.nytimes.com/2022/10/12/upshot/midterms-polling-phone-calls.html

DeBell, Matthew. 2022. "Measuring Political Knowledge and Not Search Proficiency in Online Surveys." *International Journal of Public Opinion Research* 34: edac002.
> https://doi.org/10.1093/ijpor/edac002

Delli Carpini, Michael and Scott Keeter. 1993. "Measuring Political Knowledge: Putting First Things First." *American Journal of Political Science* 37: 1179-1206.
> https://doi.org/10.2307/2111549

------. 1996. *What Americans Know About Politics and Why It Matters*. Yale University Press.

Diedenhofen, Birk and Jochen Musch. 2017. "PageFocus: Using paradata to detect and prevent cheating on online achievement tests." *Behavior Research Methods* 49:1444– 1459.

Graham, Matthew. 2023. "Detecting and Deterring Information Search in Online Surveys." *American Political Science Review*. Early View.
> https://doi.org/10.1111/ajps.12786

Graham, Matthew. Working Paper. "Catch Questions Backfire."
> https://m-graham.com/papers/Graham_CatchQuestionsBackfire.pdf

Jensen, Carsten and Jens Peter Frølund Thomsen. 2014. "Self-reported cheating in web surveys on political knowledge." *Quality and Quantity* 48: 3343–3354.

Kleinberg, Mona. 2022. "Measuring Political Knowledge in Online Surveys: How Question Design can Improve Measures." *Public Opinion Quarterly* 86.3: 736-747.
> https://doi.org/10.1093/poq/nfac024

Liu, Mingnan and Yichen Wang. 2014. "Data Collection Mode Effects On Political Knowledge." *Survey Methods: Insights from the Field*: 1–12.

Montgomery, Jacob, Brendan Nyhan and Michelle Torres. 2018. "How Post-Treatment Bias Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62: 760–775.

Motta, Matthew, Timothy Callaghan and Brianna Smith. 2016. "Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys." *International Journal of Public Opinion Research* 29: edw027.

Munzert, Simon, et al. 2022. "Who's cheating on your survey? A detection approach with digital trace data." *Political Science Research and Methods*. First View.
https://doi.org/10.1017/psrm.2022.42

Neuman, Russell. 1981. "Differentiation and Integration: Two Dimensions of Political Thinking." *American Journal of Sociology* 86: 1236-68.

Shulman, Hillary and Franklin Boster. 2014. "Effect of Test-Taking Venue and Response Format on Political Knowledge Tests." *Communication Methods and Measures* 8: 177– 189.

Smith, Brianna, Scott Clifford, and Jennifer Jerit. 2020. "TRENDS: How internet search undermines the validity of political knowledge measures." *Political Research Quarterly* 73.1: 141-155.
https://doi.org/10.1177/1065912919882101

Strabac, Zan and Toril Aalberg. 2011. "Measuring political knowledge in telephone and web surveys: A cross-national comparison." *Social Science Computer Review* 29:175–192.

Vezzoni, Cristiano and Riccardo Ladini. 2017. "Thou shalt not cheat: How to reduce internet use in web surveys on political knowledge." *Rivista Italiana di Scienza Politica* 47:251– 265.

Appendix

Table A1 – (Lack of) Demographic Differences in Experimental Flagging

	(1)	(2)
Pledge		-0.15***
		(0.03)
30-49	0.03	0.02
	(0.04)	(0.04)
50-64	0.01	0.00
	(0.04)	(0.04)
65+	0.01	0.00
	(0.05)	(0.05)
4 Year Degree+	-0.05	-0.05
	(0.03)	(0.03)
Some College	0.02	0.03
	(0.04)	(0.03)
Male	-0.02	-0.03
	(0.03)	(0.03)
Democrat	-0.02	-0.02
	(0.03)	(0.03)
Republican	0.00	0.00
	(0.03)	(0.03)
Black	0.00	-0.01
	(0.05)	(0.05)
Hispanic	0.03	0.03
	(0.04)	(0.04)
Asian	0.04	0.02
	(0.06)	(0.06)
Other Race	-0.07	-0.04
	(0.10)	(0.10)
Internet Usage	0.13	0.11
	(0.09)	(0.08)
Num.Obs.	945	945
R2	0.010	0.039
R2 Adj.	-0.004	0.025

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

About the Annenberg Public Policy Center

The Annenberg Public Policy Center of the University of Pennsylvania was founded in 1993 and created FactCheck.org in 2003. By conducting and disseminating research, staging conferences and hosting policy discussions, its scholars have addressed the role of communication in politics, science, adolescent behavior, child development, health care, suicide prevention, civics, and mental health, among other important areas. The center's researchers have drafted materials that have helped policy makers, journalists, scholars, constituent groups and the general public better understand the role that media play in their lives and the life of the nation.

APPC's motto is "Research and Engagement That Matter," and its work has informed the policy debates around campaign finance, children's television, internet privacy, tobacco advertising, the tone of discourse in Washington, and disinformation. Scholars at the policy center have offered guidance to journalists covering difficult stories, including terrorist threats, suicide, mental health, Covid-19, the Zika virus, and vaccination hesitancy. The center's discussions of key public policy issues have brought together industry representatives, advocates, government officials, media platforms, and the scholarly community. Its research has examined what messages work best to reduce the spread of HIV and drug use, how to improve candidate discourse and fact checking, how to best communicate about vaccination, and specific strategies for parents to use to monitor their children's media exposure.

Through Annenberg Classroom and the Civics Renewal Network, which it organized and manages, APPC has developed award-winning materials to help educators and schools do a better job of teaching youth about citizen rights, civic responsibility, democracy and the Constitution.

APPC's ongoing funding comes from an endowment established for it by the Annenberg Foundation in 1993.

If there are any questions about these materials, please contact our office at (215) 898-9400 or send us an email.