Measure for Measure: An Experimental Test of Online Political Media Exposure

Andrew M. Guess

Department of Political Science, Columbia University, New York, NY, USA
e-mail: amg2232@columbia.edu

Edited by R. Michael Alvarez

Self-reported measures of media exposure are plagued with error and questions about validity. Since they are essential to studying media effects, a substantial literature has explored the shortcomings of these measures, tested proxies, and proposed refinements. But lacking an objective baseline, such investigations can only make relative comparisons. By focusing specifically on recent Internet activity stored by Web browsers, this article's methodology captures individuals' actual consumption of political media. Using experiments embedded within an online survey, I test three different measures of media exposure and compare them to the actual exposure. I find that open-ended survey prompts reduce overreporting and generate an accurate picture of the overall audience for online news. I also show that they predict news recall at least as well as general knowledge. Together, these results demonstrate that some ways of asking questions about media use are better than others. I conclude with a discussion of survey-based exposure measures for online political information and the applicability of this article's direct method of exposure measurement for future studies.

1 Introduction

Survey-based measures of media exposure are notoriously imperfect. Respondents tend to inflate their news consumption in self-reports, which can paint a misleading picture of the overall news audience and bias estimates of regression parameters in studies of media effects. Recent research has estimated that for network news, this measurement error can overstate aggregate exposure by at least 200% (Prior 2009a).

Such problems, and persistent findings since the 1940s of "minimal effects" in observational studies, once led some scholars to abandon the media-effects enterprise entirely (Lazarsfeld et al. 1944; Graber 1997). Nearly 50 years after the canonical studies of the Columbia School, Bartels (1993) described this state of affairs as "one of the most notable embarrassments of modern social science."

Yet research on media effects continues. Since the pioneering experiments of Iyengar and Kinder (1987), scholars have worked to identify how exposure to campaign content or news coverage in the lab affects opinion change, candidate evaluations, vote choice, and other outcomes. Others have since studied these effects in the field (e.g., Gerber et al. 2011). But even though researchers in such designs control the content and timing of treatments, measuring subjects' exposure can still be necessary for the analysis—for example, to estimate complier average causal effects (CACE). At the same time, survey-based measures, in overcoming problems of external validity, remain a vital tool for disentangling the role of the media in real-world studies of political behavior.

Thus far, efforts to validate existing survey-based measures have suffered from a persistent problem: lacking a measure of true exposure (e.g., Romantan et al. 2008; Prior 2009b). I address this shortcoming by demonstrating the usefulness of a simple yet powerful new method of recovering subjects' actual history of exposure. The approach necessitates a shift in focus from the more traditional domain of television and newspapers to the Internet. In doing so, this article provides the first evidence of the accuracy of measures designed specifically to capture online political news consumption.

Such measures will only increase in importance as more Americans use the Internet as their primary source for news and information about the political world. A recent Pew report found that 39% of respondents went online or used a mobile device to catch up on news "yesterday," and this proportion increased to 50% when social networks, e-mail, and podcasts were included (Pew Research Center for the People & the Press 2012). Given continuing generational gaps in news consumption habits, online sources will likely overtake more traditional media in the near future. But so far, the literature on media exposure has lagged behind.

The measure of actual exposure is constructed from subjects' browsing history. Web browsers, such as Firefox and Google Chrome, continually collect information on which pages users visit and when they do so. These are saved locally and remain secure. However, it is possible with users' permission to obtain data on which of a predetermined list of sites has been visited within a given period of time. This can be matched to specific question wordings in surveys. In this article, I demonstrate how to generate the data both indirectly and directly: first by taking advantage of how browsers display visited and unvisited links to users, and second by using a browser plug-in that collects the information automatically.

I find that among the three self-reported measures of media exposure tested, asking respondents to list the Web sites they have visited to learn about politics in an open-ended format produces the best results. In the aggregate, open-ended questions recover an accurate picture of the overall online audience for news, and at the individual level they cause substantially less overreporting. I also show in a validation exercise that exposure measures constructed from open-ended responses predict the reception of recent political news stories better than general political knowledge. However, there is still a worrying level of absolute misreporting across all question types, reinforcing the arguments of some scholars that direct measures—such as the one demonstrated in this article—should be used when possible for studying individuals' political behavior. However, the open-ended survey measure appears suitable for use in capturing aggregate exposure.

The article proceeds as follows. First, I briefly review the critiques of existing exposure measures and discuss the special challenges of studying online media. Second, I outline three distinct experimental designs intended to test and benchmark the survey-based measures of online media exposure. Third, I present basic results comparing the performance of the measures against each other and against the true measure on several different metrics. Fourth, I analyze characteristics of both individuals and sites that predict misreporting in surveys. Fifth, I test the survey-based exposure measures against a measure of general political knowledge in predicting news recall. A final section then concludes with a discussion of how both open-ended survey questions and the direct measures I introduce can improve research.

2   Improvements versus Proxies

Scholars have proposed a number of workarounds to improve existing exposure measures. One class of solutions aims to incentivize accuracy among respondents, either by offering them money or by giving them more time (Prior and Lupia 2008). Another attempts to anchor respondents' expectations by providing them with basic population-level summaries of behavioral frequencies, such as how often most Americans watch network news broadcasts (Burton and Blair 1991; Prior 2009b).

These proposed solutions may not be reliable, however, and reporting population-level data requires this information to be known in advance. Especially in the case of questions probing subjects' new-media consumption habits, this may not always be the case, or at least the population frequencies in question may not be stable over time.

Other scholars have concluded that substitutes should be used instead. Most prominently, Price and Zaller (1993) have proposed and validated general political knowledge questions as a proxy for news media exposure. However, as Althaus and Tewksbury (2007) note in their report on media use measures in the American National Election Studies (ANES), such knowledge questions must be periodically aligned with contemporary realities and thus may make longitudinal comparisons difficult.

The final approach is to forego exposure measures entirely in favor of *direct* behavioral measures. A notable recent attempt to do so relied on a mobile-phone-based ambient recording system deployed by a media research firm in Chicago and New York (LaCour 2013). By matching digital audio signatures with actual broadcasts in subjects' media markets, the innovative study design enabled the collection of accurate, continuously updated records of exposure to specific programs on television and radio. Unfortunately, the likelihood of another such study in the near future is low: Issues of cost aside, the company, Integrated Media Measurement Incorporated, has been acquired and its study discontinued.

In sum, while the problems of relying on self-reports are manifest and well known, the proposed alternatives have shortcomings of their own. Improving on existing measures should continue to be a priority, with special attention to the shifting media landscape in which studies are being conducted.

3 Survey-Based Measures of Online Media Exposure

One of the main recommendations of Althaus and Tewksbury's report for the ANES is to adapt existing measures to more closely match respondents' experiences with newer forms of media. Such efforts will have to address the challenges discussed by Mutz and Young (2011, 1023–1024) in their recent survey of public opinion and communication research:

> Indeed, asking the average citizen whether he or she watches, reads, or listens to "news" these days is the classic example of a bad survey question because the very definition of what constitutes "news" is in flux. Because scholars have yet to come to grips with all of these recent changes, we know little about where people are getting their exposure to political information and argument, and whether the source makes any difference.

An important feature of the online news environment that differs markedly from the traditional realms of television, radio, and newspapers is the sheer quantity of available sources. The Internet offers potentially thousands of avenues for learning about the political developments of the day, from mainstream news sites and aggregation portals to blogs and other online publications targeted to narrower audiences. A corollary to this proliferation of sources is that outlets often cater to a specific range of the ideological spectrum (Mullainathan and Shleifer 2005).

But if the number of potential sources of exposure presents an additional difficulty to the measurement task, the way people interact with political news online may actually make it easier. The bulk of Internet news consumers primarily visit a small handful of mainstream Web sites—many but not all of which are associated with established newspapers or cable channels (Gentzkow and Shapiro 2011; Hindman 2008). Moreover, while TV viewership may be considered at least in part a passive activity, accessing political information online is a more purposeful activity that might lead to more accurate recall (Mutz and Young 2011).

Given these attributes of the online media environment, the studies presented here allow for a wide range of specific online sources to be measured while focusing as much as possible on subjects' deliberate—as opposed to incidental—exposure to political content.

3.1 *Types of Survey Questions*

Based on a review of the survey response and media exposure literatures, I have identified three different types of questions that can be adapted for capturing online media habits for political news.

First, ***check-all questions*** allow respondents to check off any number of options from a given list of information sources, from none to all. This is a common format: For example, the National Annenberg Election Survey (NAES), in its 2008 online panel, allowed multiple selections from lists in its exposure questions on television. Dilliplane et al. (2013) validate this "program list technique" as a measure for exposure to political information on television. The ANES also adopted versions

of this question format in its 2012 Evaluations of Government and Society Study and pre-election questionnaire.

Second, ***open-ended questions*** allow respondents to give their answers without restrictions on length or the choice set (Sudman et al. 1996, 225). The Pew Research Center has used this format in its Internet & American Life Project surveys; its 2010 post-election survey asked respondents which Web sites they used for campaign and election news, coding up to three responses.

And third, ***forced-choice, yes/no questions*** (Smyth et al. 2006) require respondents to choose either "Yes" or "No" to each question in a sequence, in this case a list of online political news sources. "Don't Know" and not answering are not options with this type of question.

Given these questions, I derived and preregistered (Humphreys et al. 2013)[1] two hypotheses on their relative performance as benchmarked against the measure of true exposure, which is described further in the next section.

First, Sudman et al. (1996) argue that open-ended questions are useful for self-reporting of behavioral frequencies, an application closely related to the measurement of media exposure (especially frequent media exposure, as it is often framed in question wording). One advantage of this type of question is that since respondents are not presented with a preselected list of choices, it reduces errors in judgment caused by the familiarity of a given item: "Reliance on accessibility or familiarity, however, also leaves a respondent open to error. Although true frequency can increase familiarity, so can factors like ease of perceiving an item, expectation induced by context, and probably many other variables" (Tourangeau et al. 2000, 142–143). This leads to my expectation that open-ended prompts will counteract respondents' tendency to overreport their exposure to online political media as compared to the check-all format:

> *Hypothesis* 1
> **(Open-Ended versus Check-All)** *Respondents to open-ended prompts will be less likely to overreport their exposure to online political media than respondents asked to select from specific outlets in a list.*

Unlike open-ended questions, both the forced-choice and check-all question types present respondents with a list of choices. However, the task faced by respondents differs substantially. Check-all questions can encourage satisficing (Krosnick 1991): respondents can leave all items unchecked or select only the first option they can justify. But forced-choice questions require a response one way or another for each item, which should make satisficing harder (Sudman and Bradburn 1982). It should also result in more "Yes" responses, "both because respondents process throughout the list and because they more deeply process each individual response option, making them more likely to think of reasons the options apply" (Smyth et al. 2006).

> *Hypothesis* 2
> **(Forced-Choice versus Check-All)** *Respondents to a successive series of forced-choice, yes–no prompts will be more likely to overreport their exposure to online political media than respondents asked to select from specific outlets in a list.*

4   Research Design and Data

To test these hypotheses, and to more generally determine the accuracy of these three survey-based measures, I deployed a series of survey experiments on Amazon.com's Mechanical Turk online marketplace (MTurk).[2] In addition to a pilot study in 2012, the surveys were in the field in four separate batches from November 22, 2013, to January 11, 2014.

MTurk is an online marketplace for posting modular, low-cost "human intelligence tasks" (HITs). Among other applications, it enables quick recruitment of subjects comprising an opt-in online sample in many ways more representative of the general population than samples found in

---

[1] Before analysis began, I uploaded a pre-registration document to my public academic Web site: http://polisci.columbia. edu/files/polisci/u227/PreregistrationDesignDocumentGuess2013a.pdf.

[2] The MTurkR package by Leeper (2013) proved useful for this step.

published lab experiments. For example, Berinsky et al. (2012) found that while MTurk samples are younger than other adult convenience samples, the subjects are older than those found in convenience samples composed of students at college campuses.[3]

Data collected in this way can be susceptible to the weaknesses of crowdsourcing in general, namely habitual users and "spammers" who race through surveys or submit meaningless responses. For this article, I took two steps to reduce the possibility of data contamination. First, at the front end, participation was restricted to registered users who are residents of the United States, age 18 or older, and whose approval rate (as determined by task requesters) for completing HITs was 95% or greater. Second, I included a "screener" question at the beginning of the survey to induce attentiveness. Following one of the recommendations of Berinsky et al. (2013), respondents could only continue with the survey after successfully completing the screener question. Since the authors found that "shirkers" differ from other respondents in meaningful ways, I opted for this "training" approach and did not drop any respondents.[4]

In the survey experiments, respondents were randomly assigned to one of the three question types for measuring online media exposure. The questions were designed to hold as much as possible constant between conditions, including the available choices and the time period respondents were asked to think about (30 days). Question wordings were as follows.[5]

Check-all condition. This question, adapted from Annenberg, read: "Which of these Web sites have you visited or used in the past 30 days for news, if any? Select ALL answers that apply." Respondents were shown a list of 27 sites, including the main network and cable news homepages, news Web sites such as NYTimes.com, niche sites such as Politico, and a balanced selection of smaller partisan Web sites. Respondents were also given the option to check "Other (please specify)," which included a text box.

Open-ended condition. Subjects were given a large blank text box with the following instructions: "Please list any Web sites or blogs that you have visited in the past 30 days for news. **Take some time to ensure that you think of all the sites you have visited.**" Respondents were encouraged to "take some time" and "think" about all the online sources they have visited in order to coax a full response in the absence of preselected choices. (The word "blogs" was also added to ensure that subjects did not skew their responses toward large or mainstream news sites.)

Forced-choice condition. Respondents were shown an ordered list of Web sites—the same as in the check-all condition—with the choices of "Yes" and "No" given next to each and the question "Have you visited this site?" Above the list was this sentence: "Which of these Web sites have you visited or used in the past 30 days for news, if any?"

To determine the list of choices given to respondents in the first and third conditions, sites were first selected on the basis of total visits as determined by a pilot administered in the summer of 2012. This was then augmented to ensure that sites for the major cable and network news organizations, newspapers, and large online portals (AOL, Yahoo!) were covered. Finally, a balanced selection of sites for partisan and nonpartisan political enthusiasts was added (e.g., ThinkProgress, Politico, Drudge Report).[6]

---

[3]MTurk nonprobability samples are decidedly skewed along several dimensions. Using a nationally representative face-to-face survey as a benchmark (the 2008 ANES), Berinsky et al. (2012) reported that in their investigation, MTurk respondents were more likely to be female (60.1% versus 54.9%), better educated, have lower income, and be white (95% versus 86.8%). Subjects recruited by MTurk are also far less likely to be religious (41.8% said they have no religion, compared to 20.1% in the ANES) or married (39% versus 50.1%), somewhat more liberal or Democratic-leaning, and less interested in politics.

[4]The screener question asked respondents for their favorite color, but only those who read all of the text saw the following instructions: "To demonstrate that you've read this much, just go ahead and select both red and green among the alternatives below, no matter what your favorite color is. Yes, ignore the question below and select both of those options" (see Berinsky et al. 2013). An error message was displayed until respondents correctly answered the question, after which they were allowed to continue.

[5]See the Online Appendix for full wordings and available responses for all three questions.

[6]The social link aggregator Reddit was also included with the expectation that the site is popular among respondents recruited via MTurk.

4.1 *Study 1: The Link Classification Technique*

In Study 1, respondents were asked to complete a *Link Classification Task* after answering one of the randomly assigned survey questions. This task allowed for the creation of the measure of actual online media exposure.

To do this, I presented subjects with an extended list of "masked" URLs of political news sites and blogs. Each link was displayed with exactly the same generic text, indicating nothing about the identity or content of the corresponding URL. Web browsers display visited and unvisited links in different colors; the survey exploited this feature by asking all respondents to check off each line containing a link displayed as visited (i.e., purple in most cases) by their browsers. Subjects were told that the "linked Web sites are all related to political news." Respondents saw a long page filled with hyperlinks labeled solely with the word "LINK"—a sea of purple and blue.[7] To the left of each line on the page, there was a single check box. Following the instructions, any line containing a purple ("visited") link was to be checked. In this way, the survey collected a snapshot of subjects' recent browser history for those specific sites.

The list of sites included in the *Link Classification Task*—and thus the universe of sources with which the survey-based measures can be benchmarked—was assembled as follows.[8] First, adapting a strategy used by Gentzkow and Shapiro (2011), I obtained comScore data on total unique visitors per month to U.S. Web sites, averaged from July to August 2013 for users at home and at work. I first took all sites with at least 300 unique visitors per month on average in the "General News" category and all sites with at least 50 unique visitors per month on average in the "Politics" category (which was smaller). I then added any sites in the larger "News/Information" category with at least 1,500 unique visitors that were not already included. Next, I winnowed down this list by removing sites that were primarily local in focus, non-news-related, or not based in the United States (with the exception of sites like BBC and *The Guardian* that cover American politics extensively). I ensured that the sites of the top 10 newspapers by total average circulation (both print and digital as of March 2013, according to the Alliance for Audited Media) were included. Finally, I crawled several directories of partisan political blogs to complete the list.[9]

The final result was a list of 155 links, along with a limited number of variations to ensure that the sites were properly captured by the technique.[10] Each line of the *Link Classification Task* consists of one to three separate URLs pointing to the same site. URLs were coded to point to main pages in order to capture purposeful behavior rather than incidental visits to specific pages.

This method of constructing the true measure of online media exposure relies on the assumption that subjects, when presented with generically labeled hyperlinks to sites on the Internet, will truthfully distinguish between those marked as visited and those marked as unvisited by their browser software. This assumption is plausible since the task, divorced from any substantive context involving politics or media, is mechanical and straightforward. However, it does need to be tested.

A validity check using hard-coded color displays demonstrated that the method is generally accurate. In Study 1, a fourth treatment condition randomly presented any of the three survey questions, followed by a version of the *Link Classification Task* in which all the link colors were fixed in advance unbeknown to subjects. Thus, there was a predetermined "correct" distribution of responses. Figure 1 illustrates Study 1's design.

I randomly selected 5 of the 155 links to be hard-coded as visited.[11] The proportion of the 132 subjects in the validity condition who correctly marked a visited link ranges from 89.4% to 97%,

---

[7]See the Online Appendix for a screen shot of how this looked to respondents.

[8]See the Online Appendix for a full list of the sites included. The order of the links displayed was randomly shuffled for each respondent.

[9]I used the following listings as additional sources for sites to include: the top 25 in Alexa's Politics > News and Media and News > Weblogs categories; Evan Carmichael's Top 50 Political Blogs of 2009; Rightwing News' 100 Most Popular Conservative Websites for 2013; The Free Republic's Top 100 Conservative Political Websites of 2007; and the List of Political Blogs (http://politicalbloglistings.blogspot.com).

[10]For example, a user might visit Yahoo! by typing either "www.yahoo.com" or "yahoo.com"; both versions were included for sites where applicable.

[11]This sparseness was by design: In the 2012 pilot, the mean number of sites checked as "visited" was 3.4.
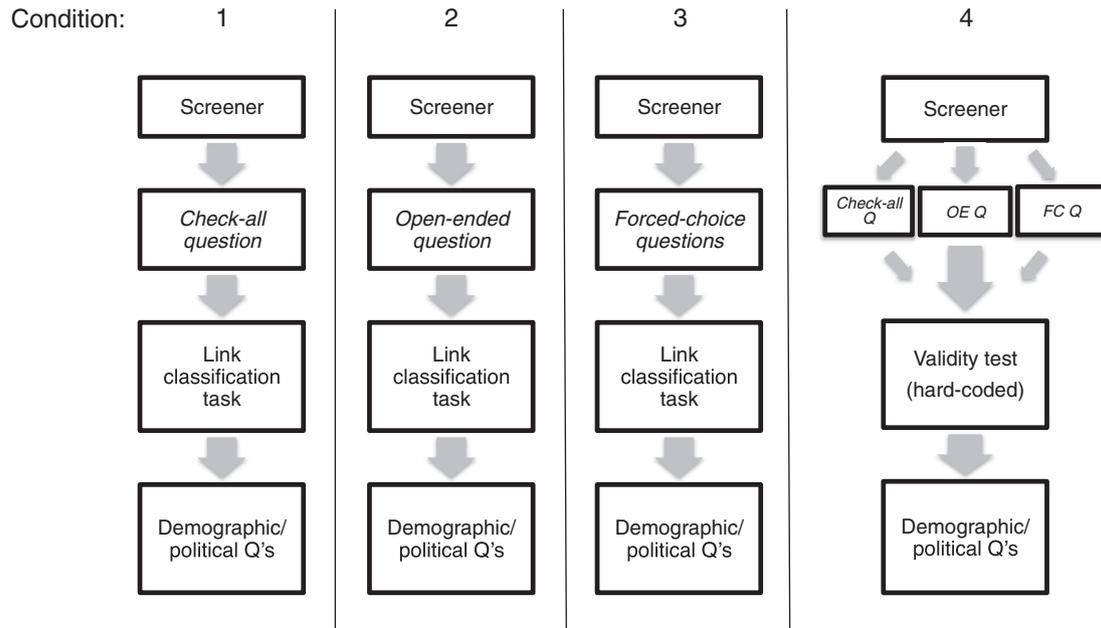
Fig. 1 Survey flow for Study 1. Subjects were randomly assigned to treatment conditions 1–4.

with an average of 95% accuracy. The low end of the range corresponds to the fact that I marked the second but not the first of two links in row 93 as visited. The high end of the range corresponds to the purple link in the first row, consistent with satisficing. The proportion of subjects who correctly left the unvisited links unmarked ranges from 99.2% (in only two cases) to 100% (in the other 148 cases). Overall, this test verifies that the method of asking subjects to manually check visited (i.e., purple) links provides an accurate measure of browsing history.

4.2 *Study 2: Browser Widget*

In Study 2, the *Link Classification Task* was replaced with an automated method of collecting the same information: a software plug-in for Google Chrome Web browsers. The other change from Study 1 was that there were only three conditions, one for each survey question type; the validity condition was not needed in this design.

Since Chrome has powerful capabilities for querying users' Web histories necessary for this kind of application, I only allowed subjects using that browser to enter the study.[12] Google Chrome is currently the second most widely used Web browser in the United States, behind Microsoft Internet Explorer.[13] But among the MTurk respondent pool, Chrome is by far the most popular: 62.9% of respondents in Study 1 used the browser to take the survey.

I implemented the design by asking respondents at the end of the survey to follow a link to install the browser extension. Since requiring workers to install software is a violation of MTurk policies, this step was left as optional for bonus credit. In the final step of the survey, respondents were given a unique identifier code and told to run the extension by clicking the button added to their browser's toolbar. When this was done, a small window appeared asking for the code. Once users submitted the code, the plug-in immediately scanned their Web history, encoded the data, and sent it to the survey database. Subjects were encouraged to uninstall the extension once it finished.

---

[12]The browser extension was written in JavaScript using the Google Chrome API. Munson et al. (2013) used a similar approach in a recent study of media slant in users' online reading habits.

[13]A total of 29.99% of users compared to 37.5% for IE, according to StatCounter Web analytics for January–December 2013. Source: http://gs.statcounter.com/#browser-US-monthly-201301-201312-bar.

The advantages of this approach are considerable: If the browser extension is installed and run properly, there will be no measurement error in retrieving a subject's exposure data. Perhaps even more attractive is the ability to precisely control the time period for which the browser history is queried. The *Link Classification Method* is susceptible to variations in the frequency with which subjects clear their browser histories, which could range from daily to never. With the browser extension, both the software and the survey questions were calibrated to the same time period—the past 30 days.

A final advantage of this approach is that the software can be adjusted for different search criteria. In the present design, the code was written to capture intentional behavior: For instance, a visit to a specific page on NYTimes.com but not the main page itself was *not* counted as exposure. This most closely matches the way exposure history was captured in the *Link Classification Task*. However, future designs could easily change this rule and incorporate any number of permissible URL variations via regular expressions.

From the respondent's perspective, the process is likely less tedious and time-consuming than scrolling through a list of 155 links. However, there are some downsides. First, some users may feel uncomfortable with installing third-party software that accesses their Web history. I addressed this issue by making clear that the plug-in only searches against a fixed list of sites (identical to the one used for the *Link Classification Task*), which respondents could view before continuing. Second, and related, this method necessarily cut down the subject pool. In addition to non-Chrome users, survey respondents who did not wish to install a browser extension were dropped from the analysis.

As a result, the effective response rate for this design was 14.6%, for a final sample size of 85 (check-all condition: $n = 30$; open-ended condition: $n = 21$; forced-choice condition: $n = 34$). Although this attrition raises the possibility of bias, I show in the Online Appendix that demographic and political characteristics of the respondents were comparable between the full sample and the subset that installed the browser extension.

4.3   *Study 3: Testing News Recall*

While the previous two designs focused on accuracy, Study 3 was intended to investigate how well the survey-based exposure measures predict news recall as compared to a measure of general political knowledge (Price and Zaller 1993), thus serving as an important test of validity.[14]

Like Study 2, Study 3 kept only the three main treatment conditions. The other difference was the inclusion of three questions testing the recall of recent political news stories. These were modeled on a series of questions included in the 1989 ANES Pilot Study and analyzed by Price and Zaller (1993) in their original validation of general knowledge questions as proxies for news reception.

The wordings were as follows, each beginning with a yes-or-no question and followed by an open-ended question which was then coded:

1. "Do you remember any recent stories about a new policy in Colorado taking effect on Jan. 1, 2014? *[If yes:]* Do you recall what that policy involves?"[15]
2. "Have you seen or heard any news stories about the *Duck Dynasty* star Phil Robertson? *[If yes:]* Do you know why he was recently in the news?"[16]
3. "Do you remember hearing about certain federal government benefits that expired at the end of 2013? *[If yes:]* Do you recall what kind of benefits they were?"[17]

---

[14]I thank an anonymous reviewer for suggesting this test.
[15]Correct answer: legalization of marijuana for recreational use.
[16]Correct answer: inflammatory comments about gays and African Americans.
[17]Correct answer: expiration of Emergency Unemployment Compensation program.

Table 1 Overview of the accuracy measures used in the analysis

| Exposure Data | | Measures | | |
Actual	Reported	Overreporting	Underreporting	Misreporting
0	0	0	0	0
0	1	1	0	1
1	0	0	1	1
1	1	0	0	0

For a given individual, each line represents one of the possible combinations of actual and reported exposure to a site.

The general knowledge questions were included in all versions of the design.[18]

4.4 Measures

In the analyses that follow, I make use of several measures of accuracy. I construct measures capturing individuals' tendency to overreport, underreport, and misreport true exposure in either direction. Table 1 summarizes how each measure is constructed.

To code responses to the open-ended question, I first examined the results, inductively generating a running list of mentioned Web sites (along with spelling variations) included in the *Link Classification Task*. I separately collected common shorthand designations for various sites in the list. Then, I coded exposure by iteratively running text searches for the resulting strings (disregarding case).

For each respondent, these measures are summed over all the possible Web sites they could have reported visiting given the treatment condition (27 for the check-all and forced-choice conditions, 155 for the open-ended prompt). Thus, in the table above, the measures of overreporting, underreporting, and misreporting would be created by summing down the rows. In the same way, I also generated raw counts of individuals' total number of actual and reported site visits.

Finally, at the site level, I created measures of total visits (both actual and reported) per site across all respondents in a given condition.

5 Results

For the main results using the *Link Classification Task*, I pool across Studies 1 and 3, keeping observations from treatment groups 1–3 ($n = 1,112$). For these observations the design is identical.[19]

The total number of visits per site across all conditions—according to the measure of actual exposure—tracks the findings of Gentzkow and Shapiro (2011), who showed that the bulk of traffic to online political news sites is taken up by a relative few mainstream outlets.[20] While this would never be true in a representative sample, Reddit, a social news aggregator, was by far the most visited site among the MTurk respondents, with 558 total visits. The next-most-visited site was CNN.com, with 293 visitors in the sample, followed by the Huffington Post with 146 and Buzzfeed with 131. The counts drop off fairly quickly, leaving a "long tail" of smaller, more narrowly targeted outlets such as Politico (37) and Talking Points Memo (14).

Figure 2 breaks down these total counts by treatment condition, showing both actual and reported exposure to online news sources summed across respondents. The left graph shows

---

[18]I asked four questions: (1) "Who is the Chief Justice of the U.S. Supreme Court?"; (2) "For how many years is a United States senator elected—that is, how many years are there in one full term of office for a U.S. senator?" (open-ended); (3) "Who is the Speaker of the House of Representatives?"; (4) "On which of the following does the U.S. federal government spend the least money?" with available choices of foreign aid (the correct answer), Medicare, national defense, and Social Security.
[19]Balance checks and summary statistics are reported in Table A1 of the Online Appendix.
[20]These descriptive findings are shown in Fig. A1 of the Online Appendix.

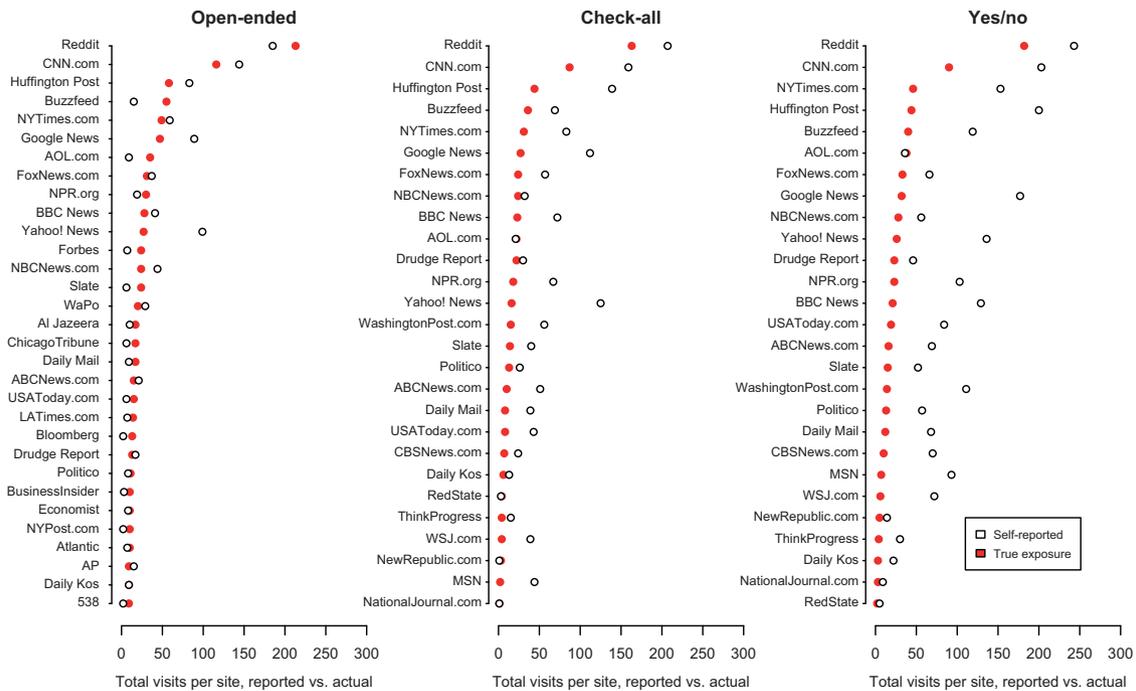## Self-reported and actual exposure, all 3 conditions

Fig. 2 Number of respondents reporting exposure to each site (actual and reported) for each treatment condition in Studies 1 and 3. Each graph is in reverse order by true exposure, with 27 each for the check-all and yes/no conditions and the 31 top sites in the open-ended condition.

reported versus actual exposure for those assigned to the open-ended prompt. There does not seem to be an obvious pattern of either over- or underreporting exposure, at least in the aggregate. The middle graph focuses on the check-all question, which seems to induce some amount of overreporting in general. The same is true of the forced-choice, yes/no question as shown on the right, which displays at least as much overreporting as the check-all question.

I plot the same data slightly differently in Fig. 3 to make direct comparisons easier and to show the amount of uncertainty in these effects. The graphs plot differences between the reported and actual exposure totals shown in Fig. 2 for each treatment condition, along with 95% confidence intervals.[21] Almost all of the differences in the open-ended condition are statistically indistinguishable from 0, while most of the differences in the check-all and yes/no conditions are positive and significant, illustrating a general tendency toward overreporting. Overall, the latter seems to induce somewhat more overreporting. One peculiarity of these graphs is that Yahoo! News, Google News, and the Huffington Post exhibit a fair degree of overreporting in all three conditions. This is possibly because some respondents visited individual article pages rather than the main page but still counted this as exposure, an artifact of the *Link Classification Task* design.[22]

Such idiosyncrasies raise the possibility that site-level characteristics systematically promote over- or underreporting. In the Online Appendix, I explore whether a site's overall audience or partisan leaning (as measured by comScore data) has an effect on self-reported exposure. I find some evidence that audience, but not partisan leaning, is associated with overreporting.

Next, I report the site-level totals from Study 2, which used the browser extension in place of the *Link Classification Task*, in Fig. 4. The results are striking: The same general pattern holds as in

---

[21]Figure A2 in the Online Appendix plots the differences for all three conditions in the same graph.

[22]Conversely, in a pilot version of this study, YouTube was the most visited site according to the *Link Classification Task* yet was rarely reported as a news source. Since this likely reflected respondents' conception of YouTube as an entertainment and not a news resource, it was kept out of Studies 1–3 altogether.
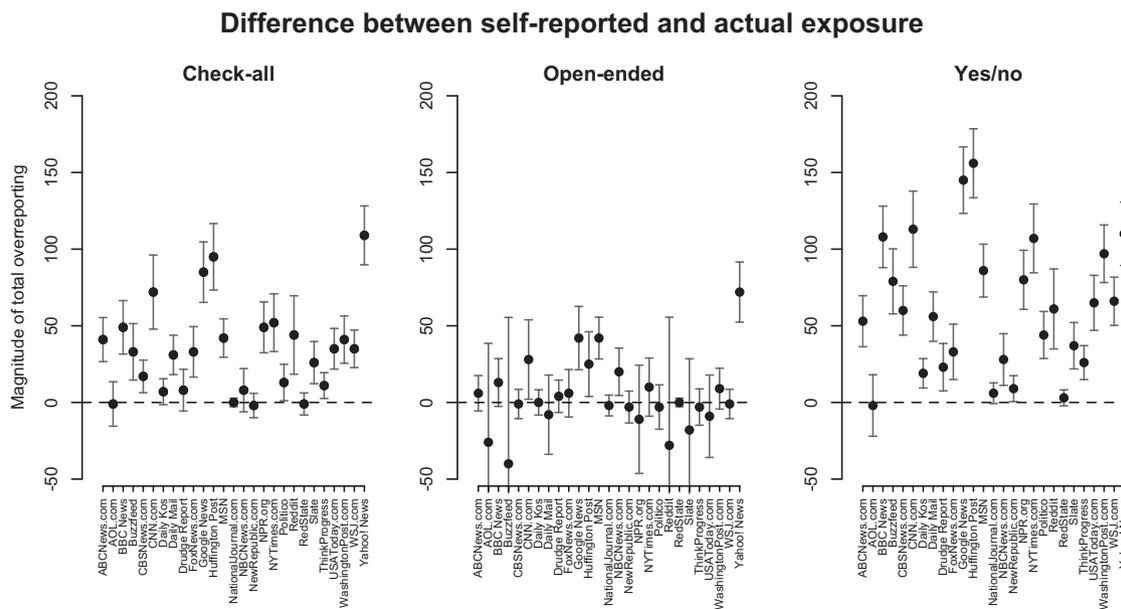
Difference between self-reported and actual exposure

Fig. 3 Plots, with 95% confidence intervals, of the difference between total reported and total actual visits per site for each survey question type. Confidence intervals were generated using *t*-tests comparing the means of the underlying self-reported and actual exposure data for each treatment group.

Fig. 2, but with less noise. These figures illustrate the apparent trade-off between measurement error and response rate with this type of survey instrument. The magnitudes on the *x*-axis are much lower than in Fig. 2 because of the sample size.

5.1 *Individual Determinants of Misreporting*

Despite the aggregate site-level results reported above, it is possible that they obscure important effects on the individual level. For example, does the open-ended prompt produce a more accurate overall picture of online news exposure because respondents are induced to report more accurately about all sources? Or is it perhaps because the question causes respondents to misreport in two different directions that cancel out?

To investigate, I report regressions in Table 2 modeling the determinants of individuals' propensity to overreport, underreport, and misreport exposure to online media. In all models, the check-all condition is the base case. To facilitate ease of interpretation, I use OLS with Huber–White robust standard errors to report the findings.[23]

I include dummies for the treatment conditions, in addition to a full range of demographic and political controls. *Knowledge* was constructed using the questions testing general political knowledge, and is coded from 0 to 4. *Attention* was an attempt to measure respondents' lack of attention devoted to political news online.[24] I also include several variables that attempt to correct for any technological issues that could have inadvertently caused measurement error with the *Link Classification Task*. *Computer* is a dummy variable for subjects' response to the question "Are you completing this survey from the computer/Web browser combination you typically use to read news online, including about American politics?" This was intended to isolate the effect of respondents who reported exposure to media that they used another computer or browsing device

---

[23]Table A2 in the Online Appendix shows the same results with both quasi-Poisson and negative binomial models as robustness checks since the dependent variable is overdispersed count data. All substantive effects discussed in the text also hold in both models, while additional significant findings are not consistent between the models.

[24]"When you watch or read news on the Internet, how much attention do you usually pay to news about politics?"
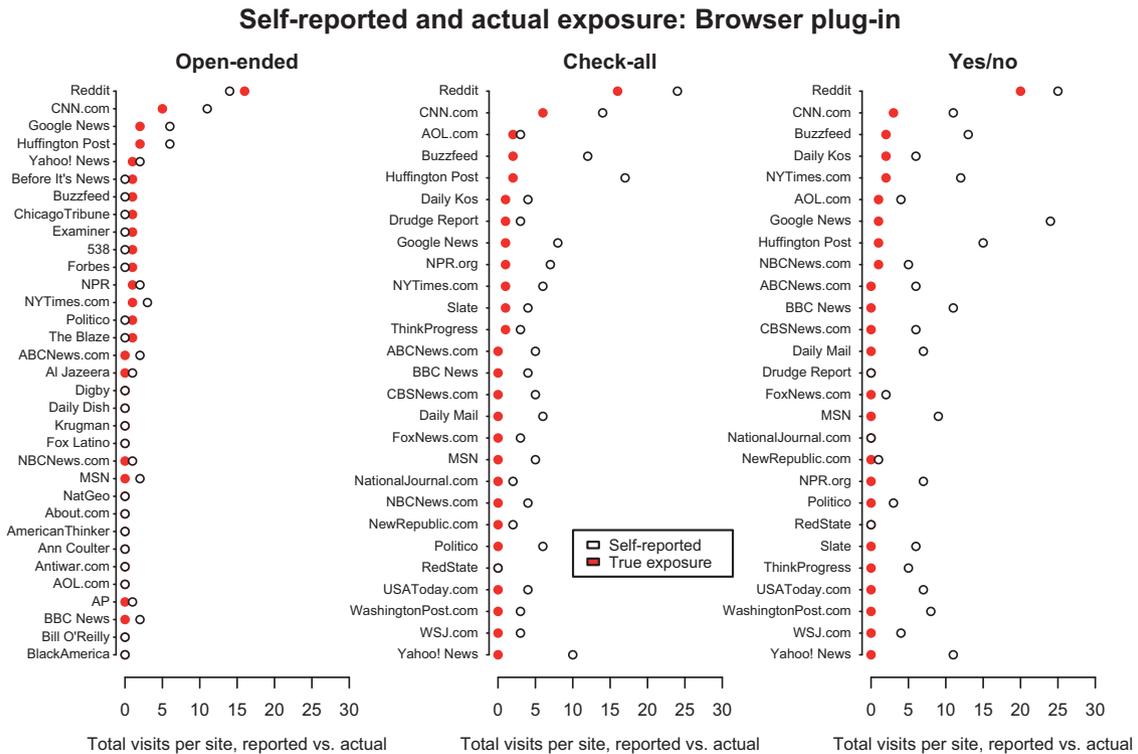
Fig. 4 Number of respondents reporting exposure to each site (actual and reported) for each treatment condition in Study 2. Each graph is in reverse order by true exposure, with 27 each for the check-all and yes/no conditions and the 33 top sites in the open-ended condition.

to access. *Cleared* is a variable coded from the question "Have you recently cleared the Web history of the browser you are currently using to complete this survey (i.e., within the last 30 days)?" This captures the possibility that some users will have nearly empty Web histories that could skew the actual exposure measure. And finally, *Purple* coded respondents' open-ended comments at the end of the survey, where they were asked if they encountered any problems with the survey, such as not seeing any purple links in the *Link Classification Task*. This was intended to capture any other possible technological issues that could have caused the *Link Classification Task* to fail.

Several results stand out. First, the only models with any real explanatory power are those that predict overreporting. This makes sense since overreporting is the most well-known problem with self-reported exposure measures. As expected, and confirming both preregistered hypotheses, the open-ended condition causes significantly less overreporting than the check-all question, and the yes/no question causes more. In terms of magnitudes, the coefficient on the open-ended condition, –1.53, implies that the absolute level of overreporting (2.33 additional sites on average for the check-all condition) is reduced by two-thirds. The yes/no condition, in contrast, increases overreporting by an even greater magnitude. Models 3 and 4 suggest that open-ended questions also encourage more underreporting, which also makes sense given the cost of recall. This is also a likely reason why the open-ended question, despite significantly less overreporting, does not perform any better than the check-all question when it comes to misreporting in general.

It appears that a number of respondents had either recently cleared their browser histories or experienced some trouble displaying the colored links. In particular, those with recently cleared histories were far more likely to "overreport," although that result is an artifact of measure construction—the actual exposure for these respondents was artificially low.

A few demographic characteristics stand out as significant explanatory factors. Echoing a well-known finding in self-reports of voting (Belli et al. 2001), respondents who were more knowledgeable about politics were more likely to overreport. In addition, respondents with more education

Table 2 OLS regressions with robust standard errors in parentheses ($n = 1,112$)

	Dependent variable:					
	Overreporting		Underreporting		Misreporting	
	(1)	(2)	(3)	(4)	(5)	(6)
Open-Ended	−1.61***	−1.53***	1.47***	1.41***	−0.14	−0.11
	(0.17)	(0.17)	(0.38)	(0.37)	(0.41)	(0.39)
Yes/no	1.82***	1.82***	0.002	0.002	1.82***	1.82***
	(0.27)	(0.26)	(0.13)	(0.15)	(0.28)	(0.28)
Age		−0.01		0.01		−0.001
		(0.01)		(0.01)		(0.01)
Hawaii/Pacific		1.34		0.84**		2.19
		(1.55)		(0.38)		(1.54)
Other Race		0.52		−1.58		−1.06
		(0.57)		(1.40)		(1.47)
Native		−2.54***		−2.42		−4.96**
		(0.76)		(1.98)		(2.08)
Black		0.66*		−0.01		0.65
		(0.34)		(0.33)		(0.44)
Asian		−0.28		0.12		−0.16
		(0.32)		(0.27)		(0.40)
Hispanic		−0.65*		2.22		1.57
		(0.34)		(2.03)		(2.03)
Female		−0.23		0.08		−0.15
		(0.19)		(0.25)		(0.30)
Income		−0.01		0.07		0.06
		(0.04)		(0.06)		(0.07)
Education		0.10		0.24*		0.34**
		(0.07)		(0.12)		(0.14)
Ideology		−0.07**		−0.01		−0.07**
		(0.03)		(0.02)		(0.04)
Party ID		0.02		0.10		0.11
		(0.07)		(0.15)		(0.16)
Knowledge		0.42***		0.10		0.51***
		(0.09)		(0.14)		(0.16)
Attention		−0.08		−0.001		−0.08
		(0.07)		(0.01)		(0.08)
Computer		−0.36		0.39*		0.03
		(0.24)		(0.23)		(0.31)
Cleared		0.99***		−0.37		0.63**
		(0.20)		(0.24)		(0.30)
Purple		0.63***		−0.51		0.12
		(0.23)		(0.59)		(0.62)
Constant	3.33***	2.33***	0.64***	−1.43	3.97***	0.89
	(0.15)	(0.58)	(0.08)	(1.17)	(0.16)	(1.29)
Adj. R^2	0.17	0.22	0.02	0.03	0.02	0.04

*Note: *$p < .1$; **$p < .05$; ***$p < .01$.

were more likely to misreport exposure, a finding that seems to be driven more by underreporting. Finally, more conservative respondents were less likely to overreport exposure than more liberal respondents.

6 Predicting News Reception

If a particular measure of media exposure is to be useful beyond a mere accounting of "who visited what Web site when," it should be able to predict respondents' ability to report political events

discussed in the news. In a classic treatment of news reception, Price and Zaller (1993) propose and validate general knowledge as a predictor of news recall, arguing that it performs better as an "indicator of a general propensity for learning about news events" (138). In Study 3 ($n = 700$), I incorporated questions about recent "news events," structured in the same fashion as the questions examined in the original paper. This allowed me to investigate whether any of the particular survey-based measures perform as well as, or better than, general knowledge.

The three events I asked about—Colorado's legalization of recreational marijuana, *Duck Dynasty* star Phil Robertson's controversial remarks about gays and African Americans, and the expiration of extended unemployment benefits—varied in terms of their overall resonance. The predictive power of general knowledge is clear: As knowledge increases (from 0 to 4), the percentage of respondents who recall the marijuana story increases from nearly 52% to nearly 75%, and the percentage who recall the *Duck Dynasty* story increases from 48% to 74%. The unemployment issue is less well known, but recall of that story increases as well, from under 19% to over half.

To investigate the performance of the exposure measures, I first created an index of overall news reception by summing up the indicators for each news story. Table 3 regresses this index on total reported exposure and general political knowledge separately within each treatment group. In the middle column, the coefficient on *Exposure* is statistically significant and larger than that for *Knowledge*: For respondents randomly assigned to the open-ended prompt, the total number of Web sites they reported being exposed to predicts news reception over and above the level of general political knowledge. The exposure measure in the check-all condition is also significant, but its effect is more modest. Also notable is that for those assigned to the open-ended group, the *Attention* measure behaves as we would expect, reducing overall news reception.

7   Discussion

The results presented here will be of considerable use to scholars seeking to improve research designs in both observational and experimental settings—especially given that data collection online is increasingly popular for both practical and substantive reasons (Iyengar 2011). Taken together, the findings of the three studies converge on the conclusion that among survey-based measures of online media exposure, open-ended questions perform the best. Evaluated against baseline measures of true exposure constructed in two different ways, they produce an accurate snapshot of aggregate online media use. They also encourage significantly less overreporting of exposure to specific Web sites, although this is partially balanced out by more underreporting. As expected, forced-choice, yes/no questions cause even more overreporting than check-all questions, and both result in an inflated picture of the overall audience for online political media.

Holding actual exposure constant, the kinds of sources that tend to get reported most are those with the greatest reach. This is likely both because those sources are more likely to be listed as a choice to be selected in closed-ended questions and because they are more likely to be easily retrieved from memory.

On the individual level, more knowledgeable respondents are more likely to overreport exposure to political media, and there is some evidence that more conservative respondents are less likely to overreport. Without further research on a representative sample, it is unclear whether these findings generalize to the population, but they raise the possibility that there are demographic and political correlates of misreporting behavior on surveys of media exposure.

In addition to performing better in terms of overreporting and aggregate exposure, open-ended questions predict reception of recent political news stories better than Price and Zaller's knowledge measure, at least in the sample analyzed here.

In general, however, it is important to remember that the absolute level of misreporting is high, relative to overall exposure, regardless of the survey-based method used. Fortunately, the methods demonstrated here point a way forward. In applications where direct behavioral measures of Internet-based activity are needed, online surveys can be equipped to present respondents with a version of the *Link Classification Task* tailored to a particular universe of sites. As demonstrated in the validity test, this procedure generates nearly error-free data on actual online exposure without

Table 3 OLS with Huber–White robust standard errors in parentheses

	DV = News reception (0-3)		
	Check-all (1)	Open-ended (2)	Yes/no (3)
Exposure	0.05**	0.14***	0.01
	(0.02)	(0.03)	(0.02)
Knowledge	0.21***	0.01	0.14*
	(0.06)	(0.05)	(0.07)
Attention	−0.11	−0.14***	−0.14*
	(0.07)	(0.05)	(0.07)
Age	0.001	−0.001	0.01*
	(0.01)	(0.01)	(0.01)
Pacific		−0.01	0.91***
		(0.15)	(0.30)
Other	0.73***	−0.08	−0.43
	(0.26)	(0.42)	(0.48)
Native	−0.88***		
	(0.22)		
Black	−0.52**	−0.12	−0.03
	(0.22)	(0.34)	(0.22)
Asian	0.15	−0.12	−0.18
	(0.24)	(0.19)	(0.23)
Hispanic	−0.31	−0.12	0.14
	(0.19)	(0.27)	(0.30)
Female	0.12	0.13	−0.10
	(0.14)	(0.12)	(0.12)
Income	−0.03	0.03	−0.04
	(0.02)	(0.02)	(0.03)
Education	0.02	0.05	0.12**
	(0.04)	(0.05)	(0.05)
Ideology	0.05	−0.09***	0.002
	(0.04)	(0.04)	(0.003)
Party ID	0.06	0.08*	−0.04
	(0.05)	(0.05)	(0.04)
Computer	0.30*	0.08	−0.47***
	(0.16)	(0.16)	(0.14)
Cleared	−0.003	0.09	0.21*
	(0.12)	(0.12)	(0.13)
Purple	−0.22	0.04	0.02
	(0.14)	(0.14)	(0.13)
Constant	0.67	1.45***	1.29***
	(0.51)	(0.37)	(0.43)
n	216	253	231

Note: *$p < .1$; **$p < .05$; ***$p < .01$.

the privacy concerns of a software extension. If deployed as part of a representative sample online, the method could shed light on exactly what types of people are more or less likely to report using a particular source.

Classifying links based on appearance is a relatively straightforward and unintrusive way of obtaining data on subjects' online browsing behavior for a given domain. Its main disadvantage is inconvenience to respondents (which increases linearly with the number of URLs included in the list). In research designs where using this procedure is infeasible, researchers have two options. First, they can opt for survey-based measures: open-ended responses that are coded, analyzed, then possibly adjusted for known biases documented in representative data. This is an especially

attractive option for studies not focusing entirely or at all on Internet-based media, since open-ended queries are endlessly adaptable.

Second, researchers can consider using the browser extension method. While Study 2 illustrated the trade-off between measurement error and response rate with this design in the survey context, it may be better suited for more controlled settings. In laboratory experiments or computer-administered online surveys, in which researchers have some degree of control over the equipment, this may actually be the least intrusive option. And most compellingly, this approach could be harnessed to tap into more refined measures of actual browsing behavior, such as logging not just whether and when users have visited a given URL, but how many times. This "dosage" measure would help separate substantively important habitual media habits from one-off visits, which the current method cannot distinguish between (Mutz and Young 2011).

References

Althaus, S. L., and D. H. Tewksbury. 2007. Toward a new generation of media use measures for the ANES. *Report to the Board of Overseers of the ANES*.

Bartels, L. M. 1993. Messages received: The political impact of media exposure. *American Political Science Review* 87(2):267–85.

Belli, R. F., M. W. Traugott, and M. N. Beckmann. 2001. What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics* 17(4):479–98.

Berinsky, A., G. Huber, and G. Lenz. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis* 20(3):351–68.

Berinsky, A. J., M. F. Margolis, and M. W. Sances. 2013. Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*.

Burton, S., and E. Blair. 1991. Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly* 55(1):50–79.

Dilliplane, S., S. K. Goldman, and D. C. Mutz. 2013. Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science* 57(1):236–48.

Gentzkow, M., and J. M. Shapiro. 2011. Ideological segregation online and offline. *Quarterly Journal of Economics* 126):1799–839.

Gerber, A. S., J. G. Gimpel, D. P. Green, and D. R. Shaw. 2011. How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *American Political Science Review* 105(1):135–50.

Graber, D. A. 1997. *Mass media and American politics*. Washington, DC: CQ Press.

Hindman, M. 2008. *The myth of digital democracy*. Princeton, NJ: Princeton University Press.

Humphreys, M., R. Sanchez de la Sierra, and P. van der Windt. 2013. Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis* 21(1):1–20.

Iyengar, S. 2011. Experimental designs for political communication research: Using new technology and online participant pools to overcome the problem of generalizability. In *Sourcebook for political communication research: Methods, measures, and analytical techniques*, eds. E. P. Bucy and R. L. Holbert. New York: Routledge.

Iyengar, S., and D. R. Kinder. 1987. *News that matters: Television and American opinion*. Chicago: University of Chicago Press.

Krosnick, J. A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5(3):213–36.

LaCour, M. J. 2013. A balanced news diet, not selective exposure: Results from Erie to Arbitron. Paper presented at the Annual Meeting of the American Political Science Association Annual, Chicago.

Lazarsfeld, P. F., B. Berelson, and H. Gaudet. 1944. *The people's choice: How the voter makes up his mind in a presidential campaign*. New York: Columbia University Press.

Leeper, T. 2013. Crowdsourcing with R and the MTurk API. *Political Methodologist* 20(2):2–7.

Mullainathan, S., and A. Shleifer. 2005. The market for news. *American Economic Review* 95(4):1031–53.

Munson, S., S. Y. Lee, and P. Resnick. 2013. Encouraging reading of diverse political viewpoints with a browser widget. *International Conference on Weblogs and Social Media*.

Mutz, D., and L. Young. 2011. Communication and public opinion. *Public Opinion Quarterly* 75(5):1018–44.

Pew Research Center for the People & the Press. 2012. In Changing News Landscape, Even Television is Vulnerable. http://www.people-press.org/2012/09/27/in-changing-news-landscape-even-television-is-vulnerable/ (accessed November 24, 2013).

Price, V., and J. Zaller. 1993. Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly* 57(2):133–64.

Prior, M. 2009a. The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly* 73(1):130–43.

———. 2009b. Improving media effects research through better measurement of news exposure. *Journal of Politics* 71(3):893–908.

Prior, M., and A. Lupia. 2008. Money, time, and political knowledge: Distinguishing quick recall and political learning skills. *American Journal of Political Science* 52(1):169–83.

Romantan, A., R. Hornik, V. Price, J. Cappella, and K. Viswanath. 2008. A comparative analysis of the performance of alternative measures of exposure. *Communication Methods and Measures* 2(1–2):80–99.

Smyth, J., D. Dillman, L. Christian, and M. Stern. 2006. Comparing check-all and forced-choice question formats in Web surveys. *Public Opinion Quarterly* 70(1):66–77.

Sudman, S., and N. M. Bradburn. 1982. *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Sudman, S., N. M. Bradburn, and N. Schwarz. 1996. *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The psychology of survey response*. Cambridge, UK: Cambridge University Press.