Survey Professionalism: New Evidence from Web Browsing Data

Bernhard Clemm von Hohenberg*[1], Tiago Ventura[2], Jonathan Nagler[3], Ericka Menchen-Trevino[4], and Magdalena Wojcieszak[5]

[1]Institute of Political Science, University of the Bundeswehr Munich
[2]McCourt School of Public Policy, Georgetown University
[3]Department of Politics, New York University
[4]Independent researcher
[5]Department of Communication, UC Davis

September 6, 2025

Abstract

Online panels have become an important resource for research in political science, but the compensation offered to panelists incentivizes them to become "survey professionals," raising concerns about data quality. We provide evidence on survey professionalism exploring three U.S. samples of subjects who donated their browsing data, recruited via Lucid, YouGov, and Facebook (total $n = 3,886$). Survey professionalism is common, but varies across samples: by our most conservative estimate, we find 1.7% of respondents on Facebook, 7.6% on YouGov, and 34.7% on Lucid to be professionals (under the assumption that professionals are as likely as non-professionals to donate data after conditioning on observable demographics available from all online survey takers). However, evidence that professionals lower data quality is limited: they do not systematically differ demographically or politically from non-professionals and do not exhibit more response instability. They are, however, somewhat more likely to speed, straightline, and attempt to take questionnaires repeatedly. To address potential selection issues in donating of browsing data, we present sensitivity analyses with lower bounds for survey professionalism. While concerns about professionalism are warranted, we conclude that survey professionals do not, by and large, distort inferences of research based on online panels.

Keywords: survey professionalism, external validity, representativeness, data quality, web browsing data

*Corresponding author. Email: bernhard.clemm@unibw.de

1 Introduction

Online panels have become an essential resource for low-cost survey research in political science and other fields. Yet, the recruiting methods of many commercial providers raise urgent questions about data quality and representativeness (Krupnikov, Nam, and Style 2021; Cornesse and Blom 2023; Jerit and Barabas 2023; Hopkins and Gorton 2023). A complicated web of panel companies, who often recruit via third-parties and even from one another, obscures the origin of respondents (Enns and Rothschild 2022).

From the panelist's perspective, the sheer number of platforms offering payment makes it attractive to become a "survey professional," namely someone who does many online surveys and spends substantial time answering them. Survey taking is actively advertised as a full-time job (see Supplementary Materials (SM) Section A) and professionals are likely to maximize their revenue by taking as many surveys as possible. Little is known about the extent and the consequences of survey professionalism. Research into the topic has relied entirely on self-reports (e.g., Zhang et al. 2020; Matthijsse, De Leeuw, and Hox 2015), which may not be the best way to measure this phenomenon. Knowing that professionalism is undesirable for researchers, respondents are likely to under-report survey taking.

Our article provides a novel empirical strategy to address the phenomenon, using *behavioral* measures constructed from web browsing data. This allows us to estimate the prevalence of survey professionalism for our population of interest—that is, online survey takers—and put it in perspective to other behaviors such as using social media. We consider three potential downstream consequences for data quality. First, professionals may differ from non-professionals sociodemographically and politically (Krupnikov, Nam, and Style 2021). Second, survey professionalism may entail speeding, satisficing, or insufficient effort in responding (Zhang et al. 2020). Third, it is also possible that survey professionals take the same questionnaire multiple times to increase their earnings (Ahler, Roush, and Sood 2019).

We analyze three different U.S. samples with subjects who participated in our surveys and were willing to donate browsing data. We also have basic sociodemographics on participants unwilling to provide browsing data, which allows us to construct target weights based on both groups combined and weight our analyses towards the population of online survey takers. Our identification assumption is that, after conditioning on observable demographics (through weighting), professionals are equally common among those respondents who donate browsing data and those who do not donate. We provide sensitivity analysis results relaxing this assumption, allowing the share of professionals between donors and non-donors to vary. The samples were recruited using different platforms that constitute a range of recruitment approaches common in political science, i.e., through Facebook ads (n = 707), Lucid (n = 2,222), and YouGov (n = 957), with their browsing data totaling over 96 million web visits. We identify survey taking in three ways: based on existing lists of questionnaire websites, by using regular expressions applied to URLs, and by manually coding the most frequent websites in our data.

We first report estimates of the extent of survey professionalism across the three samples, weighted towards the population of online survey takers. Our results indicate that survey professionalism is most prevalent in the Lucid sample (54.6% of all visits, or 43.0% of time spent browsing), followed by YouGov (23.9% of all visits, or 18.2% of time spent browsing), and Facebook (10.2% of all visits, or 8.7% of time spent browsing). Out of four dichotomous measures of survey professionalism we offer, our lowest estimate puts the percentage of survey professionals at 1.7% of Facebook respondents, 7.9% of YouGov respondents, and 34.3% of Lucid respondents. In a sensitivity analysis, in which we drop the assumption that survey professionals are as likely to donate data as non-professionals, the lower bounds for this measure are at 1.1% of Facebook, 4.8% of YouGov, and 19.8% of Lucid respondents.

Second, we compare survey professionals and non-professionals on a range of sociodemographics and political characteristics, finding few consistent differences. Third, we compare the two groups on several indicators of response quality. Speeding through surveys is more common among survey professionals. However, we find only weak evidence for satisficing (measured by straightlining through grid questions), and no evidence that survey professionals show lower response stability over time. Lastly, by analyzing visits to well-known questionnaire platforms, we explore how common it is that participants attempt to take the same questionnaire multiple times. We find that 26.7% of Facebook, 71.5% of Lucid, and 15.3% of YouGov respondents attempted to take *at least one* questionnaire more than once. This behavior is much more common among professionals than non-professionals.

In sum, although we uncover the widespread presence of survey professionalism across three types of opt-in online samples, evidence that this affects data quality negatively is limited. Professionals may be more prone to speeding and straightlining, but these behaviors can be measured and contained by excluding respondents. The prevalence of repeated attempted questionnaire taking is arguably our most worrying finding.

2 Extent and consequences of survey professionalism

Previous literature has mostly understood survey professionalism in terms of the number of panels someone is a member of, or the number of surveys done in a certain time period (e.g., Gittelman and Trimarchi 2010; Callegaro et al. 2014; Zhang et al. 2020). This conceptualization is geared towards measurement via survey self-reports such as "How many Internet surveys have you completed before this one?" (cf. Zhang et al. 2020). Such questions do not easily translate into measurements using fine-grained web browsing data, since concepts such as "a survey" cannot be neatly counted in the vast browsing data a subject produces.

We offer a slightly different definition of survey professionalism, which follows the core ideas of previous work, while lending itself to measurement with web browsing data. Our unit of analysis is the visit to a survey site (henceforth "survey visit"), as well as its duration. We use a broad definition of survey visits, including filling out a questionnaire and the steps necessary before and after: signing up on platforms offering survey jobs, selecting surveys on such platforms, and obtaining the rewards. These various activities cannot be cleanly separated from one another, as many online platforms provide several or all of these functionalities. We define and measure a subject's survey professionalism as the number (and duration) of their survey visits. We offer details on measurement in Section 3.2.

We use our measure of survey professionalism in two ways. First, we examine it as a continuous variable, i.e., someone can have more or fewer survey visits or spend more or less time on surveys. We use this continuous variable to measure the overall extent of survey professionalism. Beyond this continuous understanding, the literature has also categorized subjects into "professionals" and "non-professionals." We follow this dichotomy and build three distinct binary measures, defining a professional as someone with (a) more than 100 survey visits a day; (b) more than 50% of their visits to survey sites; (c) more than 50% of their browsing time spent on survey sites. Additionally, we create a fourth measure that counts as professional (d) a subject meeting any of the three criteria above. We present the first approach in the main paper and report the others in the SM.

2.1 The extent of survey professionalism

This conceptualization and measurement allows us to quantify the extent of survey professionalism, a goal pursued by researchers for two decades. An early study reported substantial numbers of respondents to be members of twenty or more panels (Ipsos Insight, NPD Group, and TNS 2006). Later studies in the U.S., the UK (Gittelman and Trimarchi 2010), and the Netherlands (Willems, Van Ossenbruggen, and Vonk 2006) confirmed that membership in multiple panels was common. The issue with these estimates is that they rely on self-reports. As respondents may surmise that heavy survey participation is not desirable from the researcher's perspective, it is plausible that they under-report survey taking. In general, respondents tend to misreport media-related behaviors (Parry et al. 2021). Estimating the number of surveys taken in a certain period may be especially challenging to remember. As the internet is now the most important mode for survey recruitment in political science, we specifically target the population of *online* survey takers. Relying on online behavioral data, we ask: *What is the degree of survey professionalism among online survey takers? (RQ1)*

2.2 Sociodemographic peculiarities of survey professionals

A common worry about online survey research is that samples may systematically differ from the general population (Krupnikov, Nam, and Style 2021). Most studies focus on demographic and political characteristics of *convenience samples* compared to the population (Malhotra and Krosnick 2007; Chang and Krosnick 2009; Valentino et al. 2020; Huff and Tingley 2015; Berinsky, Huber, and Lenz 2012). There is less evidence on the peculiarities of *survey professionals*, who are conceptually distinct from members of convenience samples. Reviewing a number of U.S. studies, Whitsett (2013) concluded that frequent survey respondents were more likely to have lower education and no full-time job, although no differences emerged in terms of gender and age. More recent studies reached mixed conclusions on variables such as age, education, gender, ethnicity, political interest and participation (Zhang et al. 2020; Matthijsse, De Leeuw, and Hox 2015). We combine survey questions on demographics and political variables with our behavioral data and ask: *Do survey professionals differ from non-professionals sociodemographically and politically? (RQ2)*

2.3 Low response quality among survey professionals (RQ3)

Concerns that online respondents might deliver lower quality of responses—manifesting in speedy responding, low attention, or a lack of careful thinking to produce meaningful responses (satisficing)—have been raised early on (Cape 2008). Although the literature evaluates the quality of responses obtained from convenience samples in general (Chang and Krosnick 2009; Zhang and Gearhart 2020; Ternovski and Orr 2022), evidence on professional respondents is scarce. Zhang et al. (2020) and Matthijsse, De Leeuw, and Hox (2015) find that professionals in the U.S. and the Netherlands were more likely to speed through the questionnaire, but not more likely to engage in straightlining—that is, respondents choosing the same scale position on all items of a battery even if some are reverse-coded, a common way to measure satisficing (Chang and Krosnick 2009). Building on this research, we ask: *Are survey professionals more likely to engage in straightlining (RQ3a) and speeding (RQ3b) than non-professionals?*

A related concern about survey data quality regards insufficient effort responding (IER), which means that survey participants do not pay attention to the questions prior to responding (McGonagle, Huang, and Walsh 2016; Toich, Schutt, and Fisher 2022). As IER results in similar response patterns as random responses (Huang et al. 2015), it should increase over-time response instability. Previous studies have explored this issue for convenience samples in general (Shapiro, Chandler, and Mueller 2013; Holden, Dennie, and Hicks 2013), but not for survey professionals. Taking advantage of the multi-wave structure of our three datasets, we ask: *Do survey professionals exhibit higher over-time response instability than non-professionals? (RQ3c)*

2.4 Survey professionals and repeated participation (RQ4)

Lastly, we examine a more problematic possibility, namely that highly active survey professionals may try to complete the same questionnaire[1] multiple times to maximize their revenue. Only a handful of studies have contemplated measuring this type of behavior, partly due to the methodological challenges of ascertaining multiple survey taking attempts by a single individual (Berinsky, Huber, and Lenz 2012; Ahler, Roush, and Sood 2019). Most of these findings are restricted to the "closed system" of MTurk, in which the researcher directly invites participants. Other sample providers often recruit respondents from third parties so that a professional respondent who is member of both panel A and B might be invited to the same survey through both panels (Ternovski and Orr 2022). In contrast to studies that only check whether participation in *one* questionnaire may be repeated, we explore whether our participants may have attempted to repeatedly participate in *any* questionnaire we can identify. We ask: *What is the extent to which participants attempt to take the same questionnaire more than once, and do survey professionals engage in more repeated participation than non-professionals? (RQ4)*

1. Throughout the paper, we distinguish between "survey sites" / "survey taking," which is the wider concept including the signing up to surveys and getting paid for them, and "questionnaires" / "questionnaire taking," which refers to the narrower activity of answering questions.

3 Data and Methods

We rely on data collected from three distinct U.S. samples. For our main analyses, we only examine subjects who took the survey and shared their browsing data ("donors"). We use subjects not willing to donate ("non-donors") to construct demographic-based weights to estimate survey professionalism among the entire samples. Our first sample was recruited in 2018 through ads on Facebook for a two-wave survey. All survey participants were paid, and those who uploaded browsing data could win one of five $100 Amazon gift cards. Our second sample was acquired in 2019 for a three-wave survey from Lucid (now Cint), a provider that aggregates respondents from third-party sources. Respondents were paid up to $25 for providing their browsing data. The third sample was recruited in 2018 from the YouGov Pulse panel for a two-wave survey, compensated according to YouGov standards. All three studies received ethical approval.

The three data sets combine survey responses with individual-level records of browsing behavior for two months for the YouGov sample and for 90 days before each wave for the Lucid and Facebook samples. In the Facebook and Lucid samples, after informed consent and some initial survey questions, participants were directed to an open-source tool, Web Historian, where they were asked to submit their browsing history stored in their browsers. In this method, web visits are collected retrospectively and only for desktop devices. In the YouGov sample, browsing data were collected with a tracking tool that participants had consented to install when signing up as panelists. YouGov browsing data were collected on-the-go, i.e., each web visit instantly produces a record, and include both desktop and mobile data. SM B.2.1 provides details on how browsing data are recorded.

Across samples, some donating subjects participated in the surveys but provided little browsing data. As such subjects would distort some of the proportional metrics we calculate—for example, someone who submitted five visits in total, all of which are to a survey site, would be treated as doing surveys 100 percent of the time—we exclude subjects who submitted data from less than seven days. For details on sample recruitment and data exclusions, see SM

B.1. Our final Facebook sample has $n_{subjects} = 707$ and $n_{webvisits} = 16.4$ million in the first wave (up to 90 days), the Lucid sample $n_{subjects} = 2{,}222$ and $n_{webvisits} = 73.8$ million in the first wave (up to 90 days), and the YouGov sample $n_{subjects} = 957$ and $n_{webvisits} = 6.4$ million (up to 60 days).[2]

3.1 External validity

A potential weakness of our research design relates to systematic differences between donors and the population of interest, i.e., all online survey takers (donors plus non-donors). All three studies were—at least in part—open to both donors and non-donors, although the recruitment process differed slightly across the studies. On Facebook and Lucid, participants landing on the survey were first asked a few sociodemographic questions and then invited to donate their browsing data. If they declined, they could not proceed with the survey. On Facebook, out of 2,775 people who participated in our study, there were 820 (29.5%) donors (of which 707 had more than seven days of browsing data and were thus included in the main analysis) and 1,955 non-donors. On Lucid, out of 15,589 initial participants, there were 2,462 (15.8%) donors (of which 2,222 had more than seven days of browsing) and 13,127 non-donors. On YouGov, we analyze two independent sets of respondents. One set consisted of panelists who had installed YouGov's tracking software. The other set consisted of panelists from YouGov's "main" panel of respondents. We refer to the second set as the "non-donor" set (although it may include some respondents who have donated data to other researchers or to YouGov). Both sets of respondents received identical surveys during an identical time frame but we do not know how many panelists YouGov incentivized to provide tracking data to produce the respondents we have in our analysis. On Yougov, we then have 1,179 donors (of which 957 had more than seven days of browsing) and 4,543 non-donors.

As we have basic socio-demographic information about donors and non-donors, we can

2. For the continuous measures of professionalism, we focus on wave-1 data in the Lucid and Facebook samples to avoid attrition bias. In the YouGov sample, browsing data was collected for the whole period even if subjects did not return to wave 2. For the binary professionalism measures, we use all waves across all samples. For the analyses based on survey responses from both waves (RQ3c) we only use subjects participating in both. For analyses of repeated participation, we use data from all waves.

explore potential differences between these groups. As detailed in SM Section B.1.2, we find that donors in the Facebook sample are not significantly different from non-donors in terms of age, gender, education, and ethnicity. In the Lucid sample, donors have a similar gender and ethnicity profile but are somewhat younger, more educated, more liberal, and more likely to identify as Democrats. The starkest differences emerge for the Yougov sample, where donors are younger, less educated, less likely to be white, more positive towards the out-party, more likely to follow politics, more politically knowledgeable and interested, but not significantly different in terms of gender and partisanship.

Although all of these differences are small in size, they raise the question of representativeness. As discussed, we target the population of online survey takers—for which there is no census. However, we assume that donors and non-donors taken together are a fair approximation of the population of online survey takers.[3] To maximize external validity, we construct target weights based on the sets of donors and non-donors combined, and weight the samples used for our main analyses with raking (Deville, Särndal, and Sautory 1993). For the Facebook sample, we base our weights on age, gender, education, and ethnicity. For Lucid and Yougov, we add ideology and partisanship. See SM Section B.4 for details on the weighting procedure. SM Section C.1 reports results for $RQ1$ without weights.

Our weighting approach should alleviate concerns of representativeness under the identification assumption that survey professionalism is as common among donor as among non-donors (after adjusting for the weighting variables). However, survey professionalism might be positively related to the willingness to donate browsing data, as both are presumably motivated by monetary rewards. Thus, we may overestimate the extent of professionalism. We run two analyses to address this concern. First, if survey professionals are more common among donors, we should see that donors have lower attrition rates than non-donors. For the sample where we can test this implication (Yougov), we do *not* find support for differential attrition (see SM Figure C.14). Second, we perform a sensitivity analysis for our estimates

3. We recognize that this statement needs to be qualified to the extent that the recruitment processes of the three platforms are a black box, and it is impossible to definitively understand why a given individual was invited to a survey or not by the platforms' mechanisms.

of the percentage of professionals varying the assumption that the share of professionals is the same among donors as among non-donors, finding that the share of professionals remains substantial (see SM Section C.6). This sensitivity analysis provides the necessary caveats when interpreting the prevalence of survey professionalism.

3.2 Measuring survey professionalism

In all three browsing data sets, each row represents a web visit by a subject and has three main variables: a subject identifier, the URL, and a timestamp of the visit. From the URL, we first extract the *host*, defined as the part between the *scheme* (e.g., "https://") and the *path* (beginning at the first "/"). The host always includes the *domain*, but sometimes it conveys more information. For example, the URL "https://survey123.panel456.com/r/abcd" would have the host "survey123.panel456.com" and the domain "panel456.com". For details on the complexities of URL parsing, see Clemm von Hohenberg et al. (2024).

Identifying survey visits. We identify survey visits at the level of URL hosts. As exemplified above, a URL host sometimes provides more information than its domain, allowing us to minimize false negatives. As defined earlier, survey visits comprise both filling out questionnaires and the steps before and after that, such as accepting study invitations and obtaining rewards. To capture the variety of websites linked to survey taking, we follow three different approaches.

- First, we rely on a report of questionnaire software published by Bevec and Vehovar (2021). From their comprehensive list (see their Table 12 on p.7), we filtered out irrelevant types (e.g., "UX tool") and manually verified the web addresses connected to each questionnaire software, such as "qualtrics.com" or "surveymonkey.com".
- Second, assuming that the content of URLs provides clues about their content, we classified all hosts that contained the word "survey" as survey taking.
- Third, an approach based purely on the identification of the word "survey" might

miss important sites in our data. Two trained research assistants manually coded the 500 most frequently visited hosts from each of our three datasets for whether people are taking surveys on them, that is either responding to questionnaires or being recruited/rewarded for such activities. Detailed coding instructions can be found in SM B.2.2. Each of the three lists was coded by two coders with an overlap of at least 10% of cases. Inter-coder reliability for the overlaps was high (Facebook: 92% agreement; Lucid: 88%; YouGov: 94%).

The final list of hosts based on these three approaches can be found in the reproduction materials. We present results broken down by the three approaches in SM Figure C.8 and highlight the most dominant survey sites in SM Figure C.9.

Measuring survey professionalism. Based on this list of hosts, we classify each visit by each participant in the three datasets as a survey visit or not. We then compute several variables: the subject's count of survey visits; the subject's time spent on survey sites; the proportion of survey visits out of all visits, the proportion of time spent on survey sites out of the overall browsing time of each participant. To measure time spent, we use timestamps to order all visits for a person chronologically and take the time until the subsequent visit as the duration—unless this time exceeds a threshold of five minutes. In this case, the participant likely paused their browsing activity, and we code the visit duration as missing (cf. Clemm von Hohenberg et al. 2024). We use these continuous count and duration variables to measure the degree of survey professionalism ($RQ1$). To address $RQ2$, $RQ3$ and $RQ4$, we dichotomize these continuous metrics to create four alternative classifications of survey professionals. For our primary analyses we define a survey professional (a) as someone who has on average more than 100 survey visits per active day (i.e., a day on which they were online). In the SM, we report results based on three alternative definitions, namely (b) a respondent with more than 50 percent of all visits to survey sites; (c) a respondent who spends more than 50 percent of all browsing time on survey sites, and (d) a respondent who meets *any* of the three conditions.

Measuring repeated participation. To measure how often individuals attempt to take the same questionnaire multiple times ($RQ4$), we take advantage of the fact that questionnaire platforms assign unique URLs to the same questionnaire. For example, "https://www.surveymonkey.com/r/MT2B22T" will permanently point to the same questionnaire. The methodological challenge here is not to count visits to generic URLs from the same platform, e.g., "https://www.surveymonkey.com/login," as a repeated attempt. We identify patterns in URLs that reliably designate unique questionnaires for eleven common questionnaire platforms (e.g., Qualtrics, SurveyMonkey, QuestionPro). See SM B.2.3 for details.

Whenever a questionnaire URL (e.g., "https://www.surveymonkey.com/r/MT2B22T") appears more than once in a participant's browsing data, the participant potentially took this questionnaire repeatedly. However, some visits to the the same questionnaire can be legitimate, for example, a participant accidentally clicking twice on the questionnaire link when invited, or coming back to a questionnaire to finish it. Hence, we only count visits to the same questionnaire URL as repeated if done at least one hour later (we vary this parameter in the SM) and if not happening directly after each other. We emphasize again that we measure *attempts* to take questionnaires repeatedly because we do not know if the participant completed the questionnaire multiple times, as questionnaire owners might have guardrails in place to prevent such behavior.

3.3 Survey-based measures

Sociodemographics. To compare professionals and non-professionals in terms of sociodemographics ($RQ2$), we rely on measures of age, gender, education, and ethnicity. We also compare our samples to 2020 U.S. census statistics on these four variables.

Political outcomes. To explore whether professionals differ from non-professionals on various political characteristics ($RQ2$), we analyze partisanship, ideology, an out-party feeling thermometer (dropping Independents), political interest, political knowledge, and following politics in the news. All measurements are reported in detail in SM Table B.3. For better comparability across data sets, we recode ideology, political interest, political knowledge, and following politics to a scale from 0 to 1. We benchmark our sample statistics against ANES 2020 survey data.

Speeding and straightlining. Across all samples, we use individual duration (in seconds) of taking the first-wave survey to assess speeding (*RQ3b*). To provide comparability with previous work, we also analyze the percentage of participants who were 30%, 40%, or 50% faster than the median duration (Greszki, Meyer, and Schoen 2015). To detect straightlining (*RQ3a*), we use multi-item batteries in which, for some items, higher values represent a higher value on the construct, while the reverse is the case for other items (Facebook: five-item battery about gun control attitudes; Lucid: five-item battery measuring attribution of malevolence to the out-party; YouGov: six-item battery on abortion attitudes). We flag subjects if they chose the same scale value on all items except if that value was the midpoint.

Over-time response instability. To identify if survey professionals change their responses to the same item at a higher rate than non-professionals (*RQ3c*), we use all variables that were measured across waves with precisely the same question wording and a numeric response scale. We identified 59 such variables in the Facebook data, 151 in the Lucid data, and 31 in the YouGov data. These are primarily standard political science questions, such as issue attitudes, media use, or political preferences.

4 Results

4.1 The extent of survey professionalism (RQ1)

How much do the participants in our three samples engage in survey taking? Figure 1 shows the percent of visits to survey sites out of all visits (bars in dark tones). These are aggregate-level percentages showing how much out of the sample's web visits go to survey sites. For comparison, the figure also plots the percent of visits to *facebook.com*, *google.com*, *amazon.com* and *youtube.com*. Survey-taking constitutes a substantial part of participants' browsing: 54.6% of all visits in the Lucid sample and 23.9% in the YouGov sample are visits to survey-taking sites. This is much more than visits to, for example, *google.com* (6.0% and 11.3%, respectively). Survey professionalism is least prevalent in the Facebook sample, i.e., 10.2% of all visits are to survey-taking sites, which is less than visits to *google.com* (25.1%).

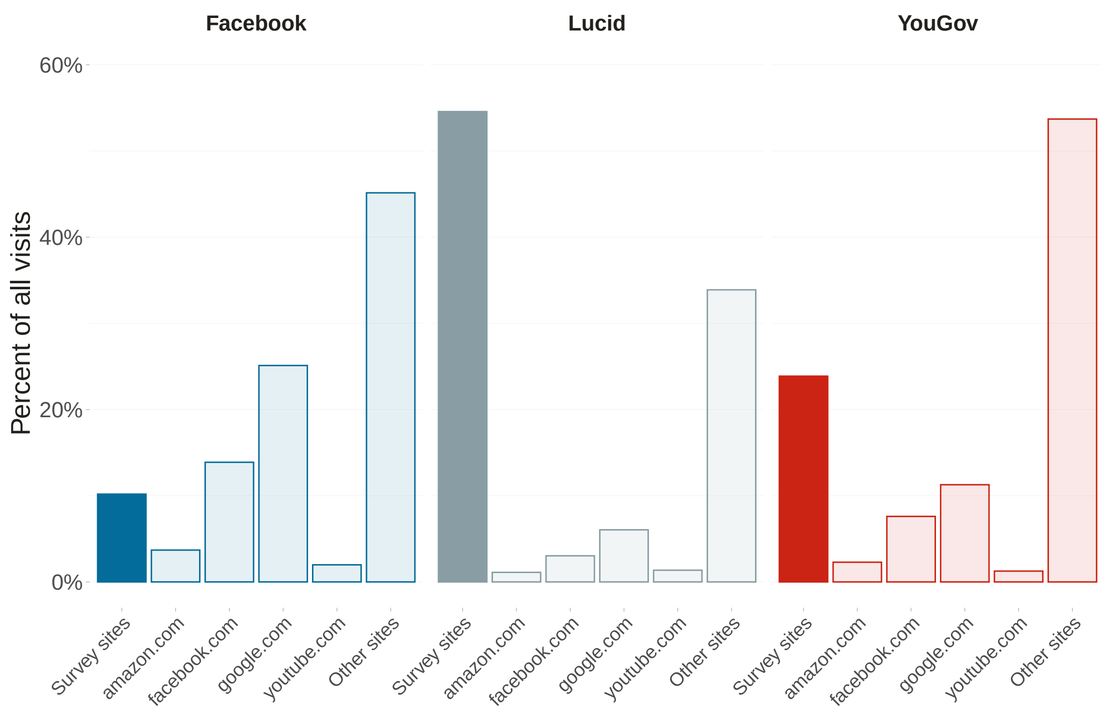

Figure 1: Percent of survey visits (out of all visits) compared to visits to popular web domains.

Figure 2 plots the distribution of survey visits per respondent. Although extreme values exist across the three samples, the shapes of the distributions differ drastically. The Facebook sample has the most skewed distribution—a few people account for most visits—followed by the YouGov sample. The Lucid sample is the least skewed, indicating that it is more common to engage in a lot of survey taking among Lucid participants. The average percentages of survey visits based on individual-level variation, depicted in SM Figure C.10, are slightly smaller than the aggregate percentages(Facebook: 8.4%; Lucid: 49.7%; YouGov: 16.3%).

In SM C.5, we report statistics in terms of visit duration. A substantial share of our samples' browsing time is spent on survey sites, whether measured as an aggregate percentage

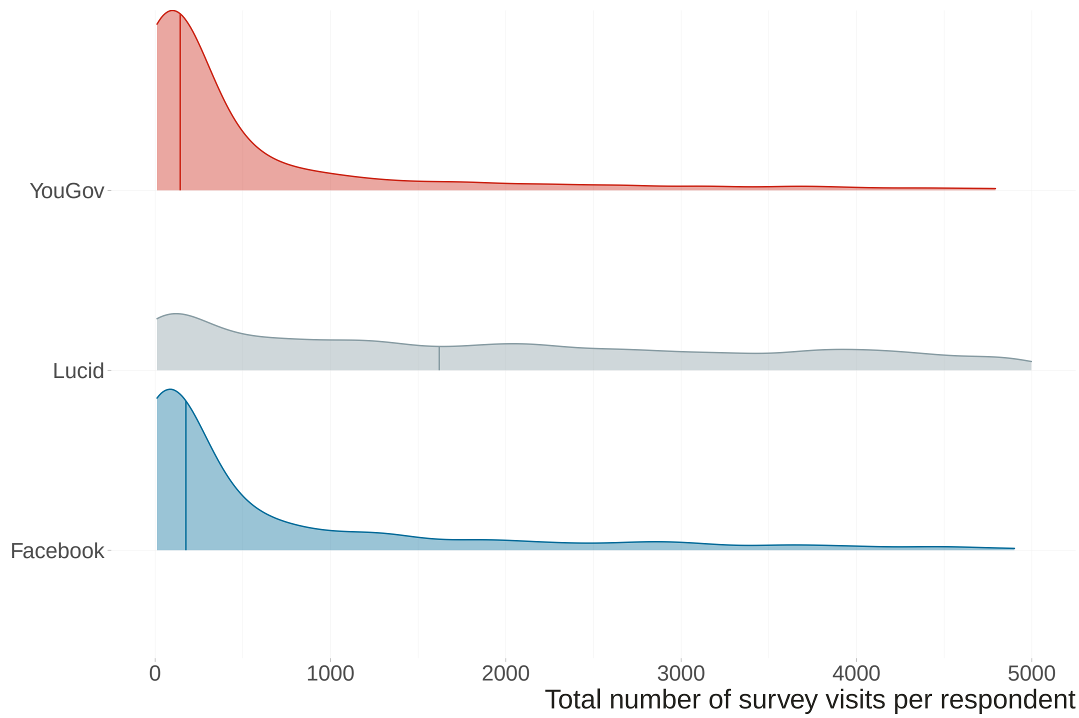

Figure 2: Individual level distribution of survey visits per respondent across the three samples. Lines represent the median value for the distribution.

(Facebook: 8.7%; Lucid: 43%; YouGov: 18.2%) or as an average individual-level percentage (Facebook: 6.8%; Lucid: 41%; YouGov: 15.2%). In SM C.3, we report which of our three approaches to identify survey sites, and which specific sites, account for most survey visits. Overall, the distribution of the three methods is similar across datasets. SM Figure C.9 presents the ten most prevalent survey sites across the samples. We see that *swagbucks.com*, *mturk*, *samplicio.us*, *decipherinc.com* represent some of the largest survey sites. This shows that participants active on one panel, e.g., YouGov, take surveys across different (and somewhat competing) platforms.

To address the next set of research questions ($RQ2$, $RQ3a, b, c$, and $RQ4$), we classify subjects as "professionals" or "non-professionals" using the four different categorizations described earlier. As shown in Figure 3, in the Lucid sample, the estimates of survey professionals vary between 34.7% and 66.2%, and go up to 71.4% when including subjects categorized as professionals by any of the measures; in the YouGov sample, between 7.6% and 11.6%, with 16.6% when using any of the categories; and in the Facebook sample, between 1.6% and 10.9%, with 11.9% when using any of the categories. For the remainder of the paper, we use the first definition of survey professionals (> 100 visits per active day).

As aforementioned, if we assume that survey professionalism is as common among donors as among non-donors (after weighting the sample of donors), these estimates should be unbiased. However, this assumption may be unrealistic and we relax it in a sensitivity analysis (details in SM Section C.6). Using the indicator of professionalism yielding the most conservative estimates, the share of professionals in the Facebook and YouGov samples does not change substantially, reducing from 1.7% to 1.1%, and from 7.6% to 4.6%, respectively. Bounds are wider in Lucid, and the estimation for the percentage of professionals would decrease from 34.7% to 20.3%.

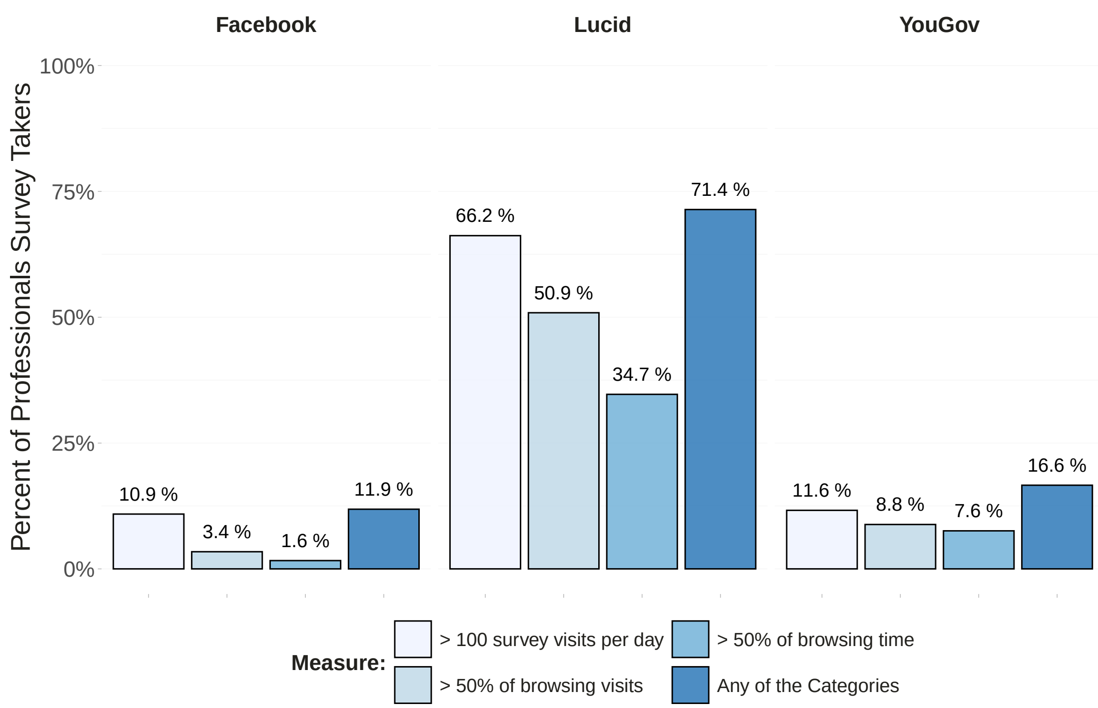

Figure 3: Percent of survey professionals for different definitions of survey professionalism.

In sum, survey taking is prevalent, but also varies substantively across samples: Lucid is the panel with the highest proportion of survey professionals; the Facebook sample has the lowest proportion. The YouGov sample sits between these extremes, indicating that even on a high-quality panel from one of the most reputable survey companies, the presence of survey professionals is not negligible. As the YouGov sample was collected from both desktop and mobile devices, we explore potential differences between these two modes (SM Section C.2): professionalism tends to be slightly more prevalent among those who primarily use desktop.

4.2 Sociodemographic and political differences (RQ2)

Do survey professionals differ from non-professionals in terms of demographic and political characteristics? Table 1 reports the central tendencies of the two groups, as well as of the U.S. population, on a number of key variables. The first rows, which compare professionals and non-professionals demographically, show the most pronounced differences in the Lucid sample, in which professionals are older, more highly educated, and more ethnically white than non-professionals at conventional levels of statistical significance. We see similar patterns in the Facebook sample, but no statistically significant difference in terms of age in the YouGov sample. Note that in the Facebook and YouGov samples, the number of professional respondents is small, which affects statistical power. SM Tables D.4, D.5 and D.6 report the same statistics for our alternative categorizations, showing that the differences are not very robust: Only the finding that professionals are older in the Lucid sample remains significant across approaches.

The second part of Table 1 reports political differences. Across the samples, survey professionals tend to be more conservative than non-professionals—although this difference only remains statistically significant across categorizations in the Lucid sample (see SM Tables D.4, D.5 and D.6). Professionals also tend to feel more positive towards out-partisans: In the YouGov sample, this difference is statistically significant no matter the categorization of professionals. In the Lucid sample, professionals are further more politically interested. In the YouGov sample, survey professionals are less politically knowledgeable and less likely to follow politics than the non-professionals. Taken together, we do not find consistent sociodemographic or political differences between professionals and non-professionals.

Table 1: Survey professionals vs. non-professionals vs. population (professionals = more than 100 survey visits / day)

		Facebook			Lucid			Yougov		
	U.S. population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	40-44	⋆⋆	35-39	46	⋆⋆⋆	37	56		55
Gender (% female)	50.8	74.8 (5.3)		74.9 (1.8)	54.4 (1.4)		54.0 (2.0)	58.4 (5.5)		55.2 (2.0)
Education (% Bachelor or more)	30.4	56.0 (6.0)		56.0 (2.0)	47.1 (1.4)	⋆	41.8 (1.9)	43.9 (5.7)		40.7 (2.0)
Ethnicity (% white)	62.6	89.1 (3.7)		84.3 (1.5)	81.6 (1.0)	⋆	77.1 (1.6)	78.5 (4.1)		77.4 (1.5)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.7 (0.3)	○	3.1 (0.1)	3.8 (0.1)		3.7 (0.1)	3.6 (0.3)		3.6 (0.1)
Ideology (0-1)	0.54 (0.006)	0.49 (0.03)	⋆⋆	0.40 (0.01)	0.54 (0.01)	⋆⋆	0.50 (0.01)	0.56 (0.03)	○	0.50 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	28.9 (3.3)		25.0 (0.9)	28.2 (0.8)		28.8 (1.1)	19.1 (2.9)	⋆	12.6 (0.9)
Political interest (0-1)	0.4 (0.006)	0.65 (0.04)		0.66 (0.01)	0.69 (0.01)	⋆⋆	0.63 (0.02)	0.70 (0.04)	⋆⋆	0.82 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.65 (0.01)		0.62 (0.02)	0.62 (0.04)		0.68 (0.01)
Following politics (0-1)		0.64 (0.04)		0.61 (0.01)				0.57 (0.04)	⋆⋆	0.69 (0.01)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for age, chi-squared tests for gender, education and race, and t-tests for all other variables (○ $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$). Sociodemographic population data from the U.S. Census; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

Table 2: Response quality of survey professionals vs. non-professionals (professionals = more than 100 survey visits / day)

	Facebook			Lucid		Yougov		
	Professionals		Non-professionals	Professionals	Non-professionals	Professionals		Non-professionals
Straightliner (%)	2.6 (1.8)		0.9 (0.4)	1.9 (0.4)	1.4 (0.5)	7.8 (2.8)	⋆	2.9 (0.5)
Duration (median seconds)	733 (77)	○	846 (25)	1167 (20)	1253 (42)	1466 (139)		1804 (60)
Duration (% 30% faster than median)	33.3 (5.6)	⋆	20.7 (1.6)	22.9 (1.1)	21.4 (1.5)	33.0 (5.2)	⋆	21.0 (1.6)
Duration (% 40% faster than median)	26.7 (5.2)	⋆⋆	13.6 (1.4)	15.6 (0.9)	15.2 (1.3)	22.9 (4.6)	⋆⋆	12.7 (1.2)
Duration (% 50% faster than median)	13.1 (3.9)		8.1 (1.1)	10.1 (0.8)	10.7 (1.2)	8.9 (2.7)		6.5 (0.9)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for survey duration, and chi-squared tests for the proportion of those 30/40/50 percent faster than the median duration and for proportion of straightliners (○ $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$).

4.3 Response-quality differences (RQ3a,b,c)

To examine whether response quality varies between professional and non-professional participants, we again report results using our first categorization of professionals (see SM Section E for the other categorizations). The first row in Table 2 reports straightlining behavior on grid questions among survey professionals and non-professionals (*RQ3a*). Overall, across the samples, professionals show a higher incidence of straightlining. For the YouGov sample, this difference is statistically significant at the 95% level—however, it is insignificant for the alternative professionalism measures.

The more pronounced differences appear in terms of speeding (*RQ3b*), reported in rows 2-5. First, consider the absolute measure of duration of taking our surveys. For the Lucid sample, the median professional is 6.9% faster compared to the median non-professional, representing a difference of one minute and twenty-six seconds. This difference is yet more pronounced for the YouGov and Facebook samples: professionals are 18.7% and 13.4% faster respectively. When we compare speeding rates among professionals and non-professionals using the different criteria for finishing faster than the median (rows 3-5), we again see that professionals speed more than non-professionals (though the difference is not statistically significant throughout).

Next, *RQ3c* asks if professionals show more response instability across waves than non-professionals. Our quantity of interest is the variability of wave-two responses after taking into account responses in wave one. To estimate this parameter, we fit a Bayesian heteroscedastic linear model that allows us to estimate the difference in the standard deviation between professionals and non-professionals explicitly. Formally, the model can be described as:

$$Y_{iw_2} \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

$$\mu_i \sim \alpha + \beta_1 Y_{iw_1} + \beta_2 SP_i + \beta_3 Y_{jw_1} \cdot SP_i + \epsilon_i$$

$$\sigma_i = SP_i \cdot \sigma_{\text{pro}} + (1 - SP_i) \cdot \sigma_{\text{nonpro}}$$

We regress the wave-two response for outcome (Y_{iw_2}) on the wave-one response for the same outcome (Y_{iw_1}), plus the interaction of the wave-one response with a dummy for survey professionalism (SP_i). We allow the standard deviation of the residual (Y_{iw_2}) to vary conditional on SP_i. Our quantity of interest is the difference in the posterior distributions between σ_{pro} and σ_{nonpro}, which represents the difference in response stability between professional and non-professionals. The full model specification is discussed in SM Section E.2. Similar modeling approaches have been used to investigate the stability of policy attitudes (Clark, Nordstrom, and Reed 2008; Alvarez and Brehm 1995; Garner and Palmer 2011).

Figure 4 presents the distribution of z-scores for differences in the standard deviation between professionals and non-professionals for all survey outcomes measured in both waves. Using this distribution, we calculate exact p-values (Imbens and Rubin 2015, Ch. 5) for the null hypothesis that the average z-score of the interactive terms is equal to zero. Across all three samples, only a few coefficients for the interactive term achieve statistical significance at conventional levels and all three distributions are centered around zero, showing no consistent patterns of lower response stability among professionals. Thirteen outcomes (out of 59) yield statistically significant parameters in the Facebook sample, nineteen (out of 149) in the Lucid sample, and six (out of 31) in the YouGov sample. For some of these outcomes, professionals actually show *lower* standard deviation than non-professionals. The exact p-values are higher than 0.78 in all three cases, which does not allow us to reject the null hypothesis. SM Section E.2.1 reports the point estimates with confidence intervals for all models. To ensure robustness of these findings, we also run an alternative model in which we regress the absolute difference between wave-two and wave-one outcome on the professionalism dummy using a simple OLS model. If professionals showed less response stability, these absolute differences should be higher, which is not the case.

To summarize, despite the high prevalence of survey professionals across the three samples, we do not find substantive differences in the quality of responses between professionals and non-professionals. Professionals do speed through surveys more, but they are not more likely to straightline through grid questions or to show more unstable responses to questions asked across waves. As an additional piece of evidence about potential quality differences, we also

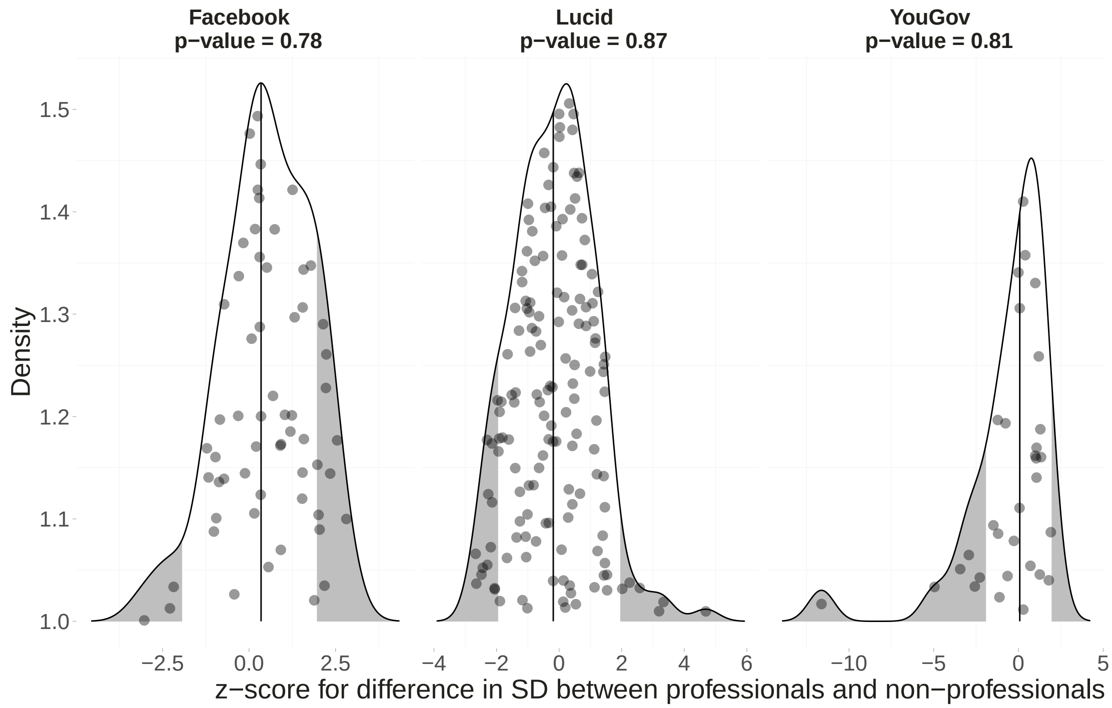

Figure 4: Z-scores for the difference in between-wave standard deviation between professionals and non-professionals. Gray areas contain z-scores larger than 1.96.

analyze whether the effects of two treatments administered in the Lucid sample vary between professionals and non-professionals, finding no differences (SM Section E.3).

4.4 Repeated questionnaire participation (RQ4)

To examine if respondents attempt to take the same questionnaires repeatedly, we track if an individual's data show two or more visits to the same questionnaire URL with a time difference of at least one hour and not happening directly after each other, as detailed earlier.[4] For the eleven questionnaire platforms we analyze, our data sets contain 70,220 unique questionnaires, which have been visited by respondents in our samples 273,840 times. It is not surprising that questionnaires are taken more than once *across individuals*. However, here we ask whether

4. We emphasize that we can only measure *attempts* to take the same questionnaire. We know when the respondent went to the same questionnaire URL more than once, but we do not observe whether the survey provider may have turned away the respondent.

the same individual attempts to take the same questionnaire more than once.

Table 3 reports several statistics capturing attempts to take the same questionnaire repeatedly. First, the percent of respondents attempting to take *at least one* questionnaire repeatedly is 26.7% for Facebook, 71.5% for Lucid and 15.3% for YouGov. This statistic does not necessarily imply that participants in this set attempt to take *many* questionnaires repeatedly. The second row speaks to the absolute number of repeated questionnaires per individual, of which we take the average across all participants. For example, if one individual attempted to take five questionnaires repeatedly and the other had fifteen attempts, then their average would be ten (no matter how many questionnaires they did in total). The table shows an average of 1.1 and 0.5 for the Facebook and YouGov samples, but 8.3 for the Lucid sample. In other words, on average, a Lucid participant attempted to take more than eight questionnaires more than once. The third row summarizes, as an average, the share of questionnaires participants take repeatedly out of all questionnaires they take. For example, if one participant took fifty unique questionnaires in total and of those ten multiple times, her percentage would be 20%; if another took one hundred unique questionnaires and ten more than once, his percentage would be 10%; their average would be 15%. The table shows that the average percentages of repeated participation range from 2.2% (YouGov) to 5.0% (Facebook) and 6.9% (Lucid).

Table 3: Repeated questionnaire participation

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)	26.7 (1.7)	71.5 (1.0)	15.3 (1.2)
Number of repeated questionnaires per participant (mean)	1.1 (0.2)	8.3 (0.4)	0.5 (0.1)
Percent of repeated questionnaires per participants (mean)	5.0 (0.5)	6.8 (0.2)	2.2 (0.3)

Note: Standard errors in parentheses.

We emphasize that our findings do not imply that YouGov participants take *YouGov* questionnaires multiple times, as the company operates its own (closed) questionnaire platform, which we could not include in our identification of unique questionnaire URLs. However, our YouGov panelists *do* attempt to repeat questionnaires on the eleven other platforms we identified, even if they are not sent there by YouGov.

Table 4 shows that among professionals, the share of those who have taken at least one questionnaire repeatedly is much higher than among non-professionals. For example, in the Facebook sample, 74.4% of professionals took at least one questionnaire more than once, whereas this is the case for only 21.1% of non-professionals. For Lucid, these numbers are 85.1% versus 44.6%; for YouGov, 69.2% versus 8.2%. In all three samples, professionals also show higher absolute and relative numbers of repeated participation.

Table 4: Repeated questionnaire participation, professionals vs. non-professionals (professionals = more than 50 of browsing time to survey sites)

	Facebook		Lucid		Yougov	
	Non-Professionals	Professionals	Non-Professionals	Professionals	Non-Professionals	Professionals
Subjects taking at least one questionnaire repeatedly (%)	21.1 (1.6)	74.4 (5.0)	44.6 (1.9)	85.2 (1.0)	8.2 (1.0)	69.2 (4.7)
Number of repeated questionnaires per participant (mean)	0.6 (0.1)	5.2 (1.0)	1.6 (0.2)	11.6 (0.6)	0.1 (0.0)	3.5 (0.6)
Percent of repeated questionnaires per participants (mean)	4.8 (0.5)	6.2 (0.9)	5.5 (0.4)	7.4 (0.2)	1.6 (0.3)	6.7 (0.7)

Note: Standard errors in parentheses.

These results are based on our cutoff of one-hour. We present results with alternative cutoffs (six hours and one day) in SM Section F.1. Although the prevalence of repeated participation slightly decreases, it is still high: For example, when using the 24-hour cutoff, the share of subjects taking at least one questionnaire more than once is still 22.5%, 68.2% and 13.1% for Facebook, Lucid and YouGov respectively. We also disaggregate results by the questionnaire platforms we identified in SM Section F.3.

We also consider an alternative explanation of repeated participation: Several visits to the same questionnaire could be the result of participants taking breaks, which would manifest in long response times for individual questions. However, we do not find that it is very common that respondents take as long as one hour for individual questions (see SM Section F.5).

Lastly, we consider whether our results could be driven by survey providers over-targeting respondents from hard-to-reach groups, who can be reached comparatively well with online surveys (Hopkins and Gorton 2023; Rosenzweig et al. 2020). SM Section F.4 zooms in on respondents over 65, from non-white racial groups, and Republicans. All three groups are somewhat more likely to attempt taking surveys repeatedly. We advise researchers to be mindful of these dynamics when using online surveys to recruit hard-to-reach participants.

5 Discussion

Our article uses a novel measurement approach based on web browsing data to offer three key findings about the extent and effects of survey professionalism in online surveys. First, professional survey-taking represents a substantial portion of the online activity of the samples. Lucid (now Cint) shows the highest prevalence of survey professionalism, followed by YouGov and Facebook. Visits to survey sites are more than half of all visits in the Lucid sample, around a quarter in the YouGov sample, and one tenth in the Facebook sample.

Second, although survey professionals constitute a non-trivial part of our samples, we do not find that they introduce significant inferential problems for online research. Although there are some demographic and political differences between professionals and nonprofessionals, these largely depend on the sample and the categorization of professionals. More importantly, survey professionalism does not seem to have pronounced implications for data quality. Professionals do show a greater tendency to speed through surveys. YouGov professional panelists also engage in more straightlining. But to the extent that these behaviors are observable, data providers and researchers themselves can decide whether to remove speeders and straightliners from their analyses. Crucially, survey professionals do not show greater instability of responses to the same questions over time. We take this as evidence that they answer questions at least as attentively as non-professionals.

Our third core finding hints at one problematic consequence of survey professionalism. 26.7% of subjects in the Facebook sample, 15.3% in the YouGov sample, and 71.5% in the

Lucid sample attempted to take at least one questionnaire multiple times. Across all three samples, repeated participation is substantially higher among survey professionals. These findings suggest that bad actors are present in the survey ecosystem and underscore the importance of building systems to detect repeated participation.

Some limitations of our study are worth highlighting. Our study design, based on behavioral measurement of survey taking, arguably offers higher internal validity than previous studies based on self-reports. Yet we acknowledge that our behavioral measure requires subjects' willingness to share their web-browsing data. As we show in SM Table B.1, donors vary in some key demographics compared to non-donors. Even after adjusting for these differences by weighting, professionalism might be more prevalent among donors. To address this concern, we provide a sensitivity analysis bounding our results. Second, it is unclear to what extent our results generalize to other vendors. We provide a range of recruitment methods commonly used by political scientists. Nevertheless, the prevalence and consequences of survey professionalism may differ among other common vendors such as MTurk or Prolific. Future work should examine our questions across different panels and time periods, potentially based on strategies and measures presented here.

Despite these limitations, we offer previously unavailable evidence on the extent and implications of survey professionalism in U.S. panels. We encourage researchers to continue examining these phenomena and their consequences for what the public knows from research. Decades ago, lab experimentalists were criticized for their reliance on college students (mostly white, more affluent, and better educated than the general population). Online panels were hoped to offer an antidote through access to a more diverse set of individuals partaking in surveys or experiments. Although online samples are also unusual as they attract survey professionals, our finding that these do not, by and large, distort inferences makes us cautiously optimistic.

Data availability statement: Data and code are available on the Harvard Dataverse.

Authors Contributions: BCvH, TV, JN and MW designed the research. BCvH and TV did all the data analysis and wrote the first version of the paper. BCvH, TV, JN and MW authors contributed to revising the manuscript. JN, MW, EMT and BCvH collected the data.

Acknowledgments: We thank Nejla Asimovic, Michael Bailey, Raphael Callel, Taylor Boas, James Druckman, Ernesto Calvo, and seminar participants at the Summer Institute of Computational Social Science (Rutgers) 2024, at the American Political Science Annual Meeting 2024, at the European Political Science Association 2024 conference, at NYU's Center for Social Media and Politics, and at the Social Data Science and AI Lab at Ludwig-Maximilians-Universitaet Munich for helpful feedback.

Funding Statement: JN gratefully acknowledges funding for data collection for this work from Democracy Fund, and that the Center for Social Media and Politics at New York University is supported by funding from the John S. and James L. Knight Foundation, the Charles Koch Foundation, Craig Newmark Philanthropies, the William and Flora Hewlett Foundation, and the Siegel Family Endowment. This research was also supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. MW gratefully acknowledges the support of the European Research Council (ERC Starting Grant, EXPO- 756301, PI Magdalena Wojcieszak).

Conflict of Interest: JN was listed as an investigator on a research grant from Facebook for "Misinformation on Social Media: Investigating the Role of Digital Literacy" that was awarded to Princeton University. MW was the principal investigator on two research grants from Facebook and is part of the 2020 Facebook and Instagram Election Study.

References

Ahler, Douglas J., Carolyn E. Roush, and Gaurav Sood. 2019. "The micro-task market for lemons: data quality on Amazon's Mechanical Turk." *Political Science Research and Methods,* 1–20. doi:10.1017/psrm.2021.57.

Alvarez, R Michael, and John Brehm. 1995. "American ambivalence towards abortion policy: Development of a heteroskedastic probit model of competing values." *American Journal of Political Science,* 1055–1082.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–368. doi:10.1093/pan/mpr057.

Bevec, Domen, and Vasja Vehovar. 2021. *A WebSM Study: Web Survey Software 2021.* Ljubljana: Faculty of Social Sciences, Centre for Social Informatics. https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=eng&id=133467.

Callegaro, Mario, Ana Villar, David S. Yeager, and Jon A. Krosnick. 2014. "A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples." In *Online Panel Research: A Data Quality Perspective,* 23–53. John Wiley & Sons. http://www.wiley.com/WileyCDA/WileyTitle/productCd-1119941776.html.

Cape, Pete. 2008. "ASC Conference - Multiple panel members: saints or sinners?" *International Journal of Market Research* 50 (5): 702–704. doi:10.2501/S1470785308200122.

Chang, Linchiat, and Jon A. Krosnick. 2009. "National Surveys Via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 73 (4): 641–678. doi:10.1093/poq/nfp075.

Clark, David H, Timothy Nordstrom, and William Reed. 2008. “Substitution is in the variance: Resources and foreign policy choice.” *American Journal of Political Science* 52 (4): 763–773.

Clemm von Hohenberg, Bernhard. 2023. “Truth and bias, left and right: testing ideological asymmetries with a realistic news supply.” *Public Opinion Quarterly* 87 (2): 267–292.

Clemm von Hohenberg, Bernhard, Sebastian Stier, Ana S. Cardenal, Andrew Guess, Ericka Menchen-Trevino, and Magdalena Wojcieszak. 2024. “Analysis of Web Browsing Data: A Guide.” *Social Science Computer Review* 0 (0). doi:10.1177/08944393241227868.

Cornesse, Carina, and Annelies G. Blom. 2023. “Response Quality in Nonprobability and Probability-based Online Panels.” *Sociological Methods & Research* 52 (2): 879–908. doi: 10.1177/0049124120914940.

DeBell, Matthew, and Jon A Krosnick. 2009. “Computing weights for American national election study survey data.” *nes012427. Ann Arbor, MI, Palo Alto, CA: ANES Technical Report Series.*

Deville, Jean-Claude, Carl-Erik Särndal, and Olivier Sautory. 1993. “Generalized raking procedures in survey sampling.” *Journal of the American statistical Association* 88 (423): 1013–1020.

Enns, Peter K., and Jake Rothschild. 2022. *Do you know where your survey data come from?* Medium. https://medium.com/3streams/surveys-3ec95995dde2.

Garner, Andrew, and Harvey Palmer. 2011. “Polarization and issue consistency over time.” *Political Behavior* 33:225–246.

Gittelman, Steve, and Elaine Trimarchi. 2010. *Online research … and all that jazz!: The practical adaption of old tunes to make new music.*

Greszki, Robert, Marco Meyer, and Harald Schoen. 2015. "Exploring the effects of removing "too fast" responses and respondents from web surveys." *Public Opinion Quarterly* 79 (2): 471–503.

Holden, Christopher J., Trevor Dennie, and Adam D. Hicks. 2013. "Assessing the reliability of the M5-120 on Amazon's Mechanical Turk." *Computers in Human Behavior* 29 (4): 1749–1754. https://doi.org/10.1016/j.chb.2013.02.020.

Hopkins, Daniel J, and Tori Gorton. 2023. "On the Internet, No One Knows You're an Activist: Patterns of Participation and Response in an Online, Opt-in Survey Panel." *Political Research Quarterly,* 10659129241268818.

Huang, Jason L., Nathan A. Bowling, Mengqiao Liu, and Yuhui Li. 2015. "Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions." *Journal of Business and Psychology* 30:299–311. doi:10.1007/s10869-014-9357-6.

Huff, Connor, and Dustin Tingley. 2015. "'Who are these people?' Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2 (3). doi:10.1177/2053168015604648.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Ipsos Insight, NPD Group, and TNS. 2006. *Heavier Responders in Online Survey Research.* https://www.ipsos.com/sites/default/files/news_and_polls/2006-10/mr061023-1POV.pdf.

Jerit, Jennifer, and Jason Barabas. 2023. "Are Nonprobability Surveys Fit for Purpose?" *Public Opinion Quarterly* 3:816–840. doi:10.1093/poq/nfad037.

Kane, John V, Yamil R Velez, and Jason Barabas. 2023. "Analyze the attentive and bypass bias: Mock vignette checks in survey experiments." *Political Science Research and Methods* 11 (2): 293–310.

Krupnikov, Yanna, H. Hannah Nam, and Hillary Style. 2021. "Convenience Samples in Political Science Experiments." In *Advances in Experimental Political Science,* edited by James N. Druckman and Donald P. Green, 165–183. Cambridge University Press.

Malhotra, Neil, and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 15 (3): 286–323. doi.org/10.1093/pan/mpm003..

Matthijsse, Suzette M., Edith D. De Leeuw, and Joop J. Hox. 2015. "Internet panels, professional respondents, and data quality." *Methodology* 11 (3): 81. doi:10.1027/1614-2241/a000094.

McGonagle, Alyssa K., Jason L. Huang, and Benjamin M. Walsh. 2016. "Insufficient Effort Survey Responding: An Under-Appreciated Problem in Work and Organisational Health Psychology Research." *Applied Psychology* 65 (2): 287–321. doi:10.1111/apps.12058.

Parry, Douglas A., Brittany I. Davidson, Craig J.R. Sewall, Jacob T. Fisher, Hannah Mieczkowski, and Daniel S. Quintana. 2021. "A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use." *Nature Human Behaviour* 5 (11): 1535–1547. doi:10.1038/s41562-021-01117-5.

Rosenzweig, Leah, Parrish Bergquist, Katherine Hoffmann Pham, Francesco Rampazzo, and Matto Mildenberger. 2020. "Survey sampling in the Global South using Facebook advertisements." *SocArXiv Papers.*

Shapiro, Danielle N., Jesse Chandler, and Pam A. Mueller. 2013. "Using Mechanical Turk to Study Clinical Populations." *Clinical Psychological Science* 1 (2): 213–220. doi:10.1177/2167702612469015.

Ternovski, John, and Lilla Orr. 2022. "A Note on Increases in Inattentive Online Survey-Takers Since 2020." *Journal of Quantitative Description: Digital Media* 2. doi:10.51685/jqd.2022.002.

Toich, Margaret J., Elizabeth Schutt, and David M. Fisher. 2022. "Do you get what you pay for? Preventing insufficient effort responding in MTurk and student samples." *Applied Psychology* 71 (2): 640–661. doi:10.1111/apps.12344.

Valentino, Nicholas A., Kirill Zhirkov, D. Sunshine Hillygus, and Brian Guay. 2020. "The Consequences of Personality Biases in Online Panels for Measuring Public Opinion." *Public Opinion Quarterly* 84 (2): 446–468. doi:10.1093/poq/nfaa026.

Whitsett, Healey C. 2013. "Understanding 'Frequent Survey Responders' on Online Panels." *Nera Economic Consulting,* https://www.nera.com/insights/publications/2013/understanding-frequent-survey-responders-on-online-panels.html.

Willems, Pieter, Robert Van Ossenbruggen, and Ted Vonk. 2006. "The effects of panel recruitment and management on research results: A study across 19 online panels." Presented at ESOMAR Panel Research, http://www.websm.org/db/12/12228/Web%20Survey%20Bibliography/The_effects_of_panel_recruitment_and_management_on_research_results_A_study_across_19_online_panels/.

Zhang, Bingbing, and Sherice Gearhart. 2020. "Collecting Online Survey Data: A Comparison of Data Quality among a Commercial Panel & MTurk." *Survey Practice* 13 (1). doi:10.29115/SP-2020-0015.

Zhang, Chan, Christopher Antoun, H Yanna Yan, and Frederick G Conrad. 2020. "Professional Respondents in Opt-in Online Panels: What Do We Really Know?" *Social Science Computer Review* 38 (6): 703–719. doi:10.1177/0894439319845102.

SUPPLEMENTARY MATERIALS (SM)

Survey Professionalism: New Evidence from Web Browsing Data

Table of Contents

A Background **3**

B Data collection and methods **4**

B.1 Details on sample recruitment and exclusions 4

B.1.1 Time frame of data collection 4

B.1.2 Comparing data donors and non-donors 4

B.1.3 Facebook sample 5

B.2 Behavioral measures 5

B.2.1 The generation of browsing data 5

B.2.2 Identifying survey sites with manual coding 6

B.2.3 Identifying repeated questionnaire participation 7

B.3 Self-reported measures 8

B.4 Weighting procedure 11

C Prevalence of survey professionalism (RQ1): additional results **12**

C.1 Unweighted results 12

C.2 Disaggregation by desktop/mobile users 14

C.3 Disaggregation by identification approaches and survey sites 14

C.4 Average individual-level percentage of survey visits 15

C.5 Analogous results based on duration 18

C.6 Sensitivity Analysis for the Prevalence of Professionalism 20

D Sociodemographic and political differences (RQ2): Additional results **22**

D.1 Alternative professionalism indicators 22

E Response-quality differences (RQ3): additional results **24**

E.1 Speeding and straightlining (RQ3a,b) 24

E.1.1 Alternative professionalism indicators 24

E.2 Stability of Survey Responses (RQ3c) 26

E.2.1 Bayesian heteroscedastic linear model 26

E.2.2 Absolute-difference OLS model 27

E.3 Treatment effect differences 30

F Repeated questionnaire participation (RQ4): additional results **32**

F.1 Alternative time cutoffs 32

F.2 Alternative professionalism indicators 33

F.3 Disaggregation by questionnaire platforms 35

F.4 Disaggregation by hard-to-reach groups 36

F.5 Break patterns as an alternative explanation 37

A Background

The screenshot shown in Figure A.1 illustrates how survey taking is advertised as a "professional" activity.

Finding a job > 20 Companies That Will Pay You To Take Surveys Online

20 Companies That Will Pay You To Take Surveys Online

By Indeed Editorial Team
Published February 8, 2021

In the age of technology, you can make money from the comfort of your own home by taking surveys online. Companies who pay people for taking surveys value market research and can use your replies to improve their business operations. You might enjoy a job taking surveys if you prefer working from home to commuting to an office every day. In this article, we explore 20 companies who offer payment for taking surveys online.

Related: How to Make Money Online

Figure A.1: Example of an article promoting survey taking as a paid occupation (retrievable at https://archive.ph/pLcII)

B Data collection and methods

B.1 Details on sample recruitment and exclusions

B.1.1 Time frame of data collection

All three data sets combine survey responses with individual-level records of browsing behavior of two months for the YouGov sample and of 90 days before each wave for the Lucid and Facebook samples. Facebook surveys started on March 29, 2018 (Wave 1) and July 16, 2018 (Wave 2). Lucid surveys started on March 18, 2019 (Wave 1), August 23, 2019 (Wave 2) and November 9, 2019 (Wave 3). YouGov surveys were carried out starting on October 4, 2018 for Wave 1 and December 3, 2018 for Wave 2. On YouGov, the browsing data collection covers the period around the Wave 1 survey, starting in September 4 and going up to November 8, 2024.

Across samples, some subjects participated in the surveys but provided little browsing data. We exclude subjects who submitted data from less than seven days. In addition, we use data from only the first wave for the Lucid and Facebook samples (i.e., up to 90 days per sample), but for both waves for the YouGov sample (i.e., 1 month before and 1 month after Wave 1). In the Facebook and Lucid samples, anyone who did not participate in later waves also did not submit any browsing data for those waves. Hence, including all waves would overweight participants who returned to later waves. This is not the case in the YouGov sample, where browsing data is collected for the later waves irrespective of whether the participant does the later surveys.

B.1.2 Comparing data donors and non-donors

A potential weakness of our research design relates to systematic differences between online survey takers who donate their browsing data and the population of interest, i.e., all online survey takers. All three studies were at least partly open to both donors and non-donors, and we have collected socio-demographics on both groups. Table B.1 compares demographic and political attitudes between participants who donated their web browsing data and those who did not. On Facebook, we have 820 donors and 1,955 non-donors; on Lucid, 2,462 donors and 13,127 non-donors, and on Yougov, 1179 donors and 4,543 non-donors. The table shows that in the Facebook sample, only the age difference is significant; In the Lucid sample, significant differences emerge for age, education, partisanship, and ideology; in the Yougov sample, significant differences emerge from age, education, ethnicity, out-party feeling, political interest, and political knowledge.

Table B.1: Data donors vs. non-donors

	Facebook			Lucid			Yougov		
	Donors		Non-donors	Donors		Non-donors	Donors		Non-donors
Sociodemographics									
Age (median years)	35-39	⋆	35-39	40	⋆⋆⋆	44	50	⋆⋆⋆	57
Gender (% female)	74.8 (1.5)		74.8 (1.2)	53.6 (1.0)		54.1 (0.5)	54.4 (1.6)		56.0 (0.7)
Education (% Bachelor or more)	52.6 (1.8)		52.7 (1.3)	48.9 (1.0)	⋆⋆⋆	44.5 (0.5)	36.0 (1.5)	⋆⋆⋆	42.6 (0.7)
Ethnicity (% white)	84.3 (1.3)		84.2 (1.0)	79.0 (0.8)		80.2 (0.4)	71.9 (1.4)	⋆⋆⋆	78.4 (0.6)
Political outcomes									
Partisanship (1-7)				3.6 (0.0)	⋆⋆⋆	3.8 (0.0)	3.6 (0.1)		3.7 (0.0)
Ideology (0-1)				0.49 (0.01)	⋆⋆⋆	0.54 (0.00)	0.49 (0.01)	◦	0.51 (0.00)
Thermometer out-party (1-100)							15.3 (0.8)	⋆	13.0 (0.4)
Political interest (0-1)							0.77 (0.01)	⋆⋆⋆	0.81 (0.00)
Following politics (0-1)							0.65 (0.01)	⋆⋆	0.68 (0.00)
Political knowledge (0-1)							0.62 (0.01)	⋆⋆⋆	0.67 (0.00)

Note: Standard errors for means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for age, chi-squared tests for gender, education and race, and t-tests for all other variables (◦ $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$). Sociodemographic population data from the US Census; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

B.1.3 Facebook sample

As participant recruitment via Facebook involves more degrees of freedom than through panel providers, we here detail the procedure. We recruited participants using Facebook advertisements targeting adults in the US. The advertisements appeared on the pages of 266,827 Facebook users, and 3,735 clicked on the link (1.4% click-through rate), which directed them to the landing survey, which inviting them to the the full Wave 1 survey and provided a link to upload their browsing histories. After informed consent, participants could complete the landing survey with or without uploading their online browsing data. All survey participants received $2 and those who uploaded their browsing data had a chance to win one of five $100 Amazon gift cards. Three months later, we asked the same participants to complete Wave 2 and again upload their browsing data for $10. As a quality check, we analysed participants' browsing histories to verify that they were indeed in the US and different individuals.

B.2 Behavioral measures

B.2.1 The generation of browsing data

In the Lucid and Facebook samples, browsing data was collected with Web Historian, a tool that collects participants' browsing history. Common browsers such as Chrome or Firefox record a visit in the browsing history when a user navigates to a webpage. Insofar as testable, browsers record an entry in the history only when the webpage is initially loaded. When a user opens a URL in one browser tab, then switches to a different tab for a while, and then returns to the original tab, the browser's history does not record a new visit. When tabs refresh in the background, whether because of inactivity, system resource management, or browser/computer restart, it generally does not count as a new visit in the browsing history, either. Some browsers use features such as "tab discarding" to manage memory, which can cause tabs to refresh when revisited, but this does not create a new entry. In summary, browsers are designed to not clutter a user's history with multiple entries for the same URL unless the user actively navigates to it again. However, they also lack a precise measurement of time spent on each visit.

For the Yougov sample, browsing data was recorded with a commercial tracking solution. These solutions—such as Wakoopa—generally attempt to deliver a more fine-grained measurement of how much time a user spends on each URL, even when switching back and forth between tabs. However, what exactly counts as a visit ultimately remains a black box. In our own past research, we have attempted in vain to get precise documentation of the data generation process from commercial vendors (Clemm von Hohenberg et al. 2024).

B.2.2 Identifying survey sites with manual coding

Beyond (1) the list of questionnaire sites by Bevec and Vehovar (2021, Table 12) and (2) categorizing all URL hosts containing the word "survey" as survey sites, we (3) manually coded the most frequented 500 URL hosts in each of the three data sets. The coding instructions for this last approach read as follows:

Goal: Code each host in hosts_500_LUCID, hosts_500_FB and hosts_500_YOUGOV to identify sites where people are (1) taking surveys or (2) engage in other activity for rewards/coupons (e.g., playing games or watching videos). "Taking surveys" includes both the actual responding to questions, and the recruitment and payment process before and after the survey. The same goes for "other activity", by which we mean both the actual compensated activity and the recruitment/payment. Steps how to identify such sites:

- Look at the URL itself (e.g. you would already know that wikipedia.org is not a survey / other rewarded activities site)
- Visit the URL. Some hosts share a second-level domain, e.g., c.opinionetwork.com & ps.opinionetwork.com and you may not able to visit these hosts. In that case, try without the subdomain (e.g. opinionetwork.com). Often the fact that there are several subdomains indicates that it is a platform on which companies have their own subdomains to distribute surveys.
- Google the name of the URL. If that does not give you a clue, google "[URL] surveys rewards" to find any discussion about this site as a place to take surveys/engage in other rewarded activity.

Codebook:

- "code": 1 = survey site; 2 = other activity compensated by rewards/coupons; 0 = not related to surveys/other rewards; 99 = unsure / unclear - try to use infrequently and always add notes if you use this to explain why it is unclear. "OVERLAP" signifies that this host already appears in one of the other two sheets
- "coder": your initials
- "code_old": code from previous, slightly different method You can take this as a guideline, but do double-check.

B.2.3 Identifying repeated questionnaire participation

Table B.2 below lists the questionnaire platforms for which we could identify URL patterns reliably pointing to unique and permanent questionnaires. Column "Regex in URL path" shows the regular expressions applied to the paths of URLs with the respective host.

Table B.2: Survey software platforms

Platform	Host(s)	Regex in URL path
Confirmit (now Forsta)	`confirmit.com`	`^/wix/[a-zA-Z0-9]`
Surveygizmo	`surveygizmo.com` `surveygizmo.eu`	`^/s3/[a-zA-Z0-9]`
Surveymonkey	`surveymonkey.com`	`^/r/[a-zA-Z0-9]`
Qualtrics	`qualtrics.com`	`^/jfe/form/[a-zA-Z0-9]`
Dynata	`survey.cmix.com`	`^/[A-Z0-9]`
Questionpro	`questionpro.com`	`/t/[a-zA-Z0-9]` `/a/TakeSurvey\?tt=[a-zA-Z0-9]`
Formsite	`formsite.com`	`[a-zA-Z0-9]/index\.html$`
Unipark	`unipark.com`	`^/uc/[a-zA-Z0-9]`
Typeform	`typeform.com`	`^/to/[a-zA-Z0-9]`
Formstack	`formstack.com`	`^/forms/(?![a-zA-Z0-9]*index\.php)/[A-Z0-9]`
Zoho	`survey.zohopublic.com` `survey.zohopublic.eu`	`^/zs/[a-zA-Z0-9]`

B.3 Self-reported measures

Table B.3 below describes the political characteristics used to address part of RQ3a, in terms of their wording, their response scale and any recoding.

Table B.3: Wording of political survey variables

Dataset	Question wording	Response scale	Recoded scale
Partisanship			
Facebook	Please select the option that best describes your political party affiliation	A strong Democrat (1) … A strong Republican (7)	1-7
Lucid	Please select the option that best describes your political party affiliation.	A strong Democrat (1) … A strong Republican (7)	1-7
Yougov	PROVIDED AS META VARIABLE	A strong Democrat (1) … A strong Republican (7)	1-7
ANES	Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what? DEPENDING ON REPONSE: Would you call yourself a strong [DEMOCRAT / REPUBLICAN] or a not very strong [DEMOCRAT / REPUBLICAN]? OR: Do you think of yourself as closer to the Republican Party or to the Democratic Party?	A strong Democrat (1) … A strong Republican (7)	1-7
Ideology			
Facebook	In politics, people sometimes talk of the political “left” and “right”. Where would you place yourself on this scale, where 0 means extreme left and 10 means extreme right?	Extreme left (0) … Extreme right (1)	0-1
Lucid	In politics, people also sometimes talk of the political ”liberal” and ”conservative”. Where would you place yourself on this scale, where 0 means liberal and 10 means conservative?	Liberal (0) … Conservative (10)	0-1
Yougov	As shown on the scale below, some people in the U.S. tend to identify more with the political left, while others tend to identify more with the political right. […] Please place yourself on this scale.	Far left (0) … Far right (100)	0-1
ANES	We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale, or haven’t you thought much about this?	Extremely liberal (0) … Extremely conservative (7)	0-1
Out-party feeling			
Facebook	We’d like you to rate several different groups using something called a ”feeling thermometer”. The higher the number, the warmer or more favorable you feel toward the group; the lower the number, the colder or less favorable. Please rate how you feel about the following groups. [DEMOCRATS / REPUBLICANS]	0 … 100	0-100
Lucid	We’d like you to rate several different groups using something called a ”feeling thermometer”. The higher the number, the warmer or more favorable you feel toward the group; the lower the number, the colder or less favorable. Please rate how you feel about the following groups. [DEMOCRATS / REPUBLICANS]	0 … 100	0-100
Yougov	Please rate each of the following political figures on a scale from 1 to 100. If you are not familiar with the person or group listed, leave it blank and move onto the next item. [THE DEMOCRATIC PARTY / THE REPUBLICAN PARTY]	0 … 100	0-100

Table B.3: Wording of political survey variables *(continued)*

Dataset	Question wording	Response scale	Recoded scale
ANES	I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. [THE DEMOCRATIC PARTY / THE REPUBLICAN PARTY]	0 … 100	0-100
Political interest			
Facebook	How interested would you say you are in politics?	Not at all interested (1) … Very interested (7)	0-1
Lucid	How interested are you in the following topics? [POLITICS]	Not at all interested (1) … Very interested (7)	0-1
Yougov	Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs…	Most of the time (1) … Hardly at all (4)	0-1
ANES	How interested would you say you are in politics? Are you very interested, somewhat interested, not very interested, or not at all interested?	Very interested (1) … Not at all interested (4)	0-1
Political knowledge			
Lucid	(1) Do you happen to know how many times an individual can be elected President of the United States under current laws? Please indicate the number of times in the box below. (2) Do you happen to know which party currently has the most members in the U.S. House of Representatives in Washington? (3) 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? Please indicate the number of years in the box below. (4) On which of the following does the U.S. federal government currently spend the least?	OPEN-ENDED OR MULTIPLE CHOICE	0-1
Yougov	(1) China has imposed additional tarrifs on some agricultural imports from the United States. Which of the following best describes United States' actions on trade with China since December 2016: (2) Which of the following best describes the U.S. stock market since December 2016. (3) Which of the following best describes the official unemployment rate, as reported by the United States Government, since December 2016.	OPEN-ENDED OR MULTIPLE CHOICE	0-1
ANES	(1) 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? (2) On which of the following does the U.S. federal government currently spend the least? (3) Do you happen to know which party currently has the most members in the U.S. House of Representatives in Washington?	OPEN-ENDED OR MULTIPLE CHOICE	0-1
Following politics in the media			
Facebook	How closely do you follow politics on TV, radio, newspapers, or the Internet?	Not at all (1) … Very closely (7)	0-1
Yougov	How often do you read news about politics online?	At least 10 times a day (7) … Never (1)	0-1

B.4 Weighting procedure

As described in the paper, to account for potential differences between donors and non-donors in our samples, we present results using weights for all our analysis. To build the weights, we use raking methods (Deville, Särndal, and Sautory 1993) implemented with the R package *anesrake*. Our target population consists of all online survey takers, approximated by the combined set of donors and non-donors. For the Facebook sample, we weight on age, gender, education, and ethnicity. For Lucid and YouGov, we additionally incorporate ideology and partisanship. To ensure robust weighting, we require that each category within our weighting variables contains at least 5% of the sample (DeBell and Krosnick 2009). For a small set of participants with missing demographic information, we assigned a weight equal to 1. These missing participants are less than 10% across the three samples.

C Prevalence of survey professionalism (RQ1): additional results

C.1 Unweighted results

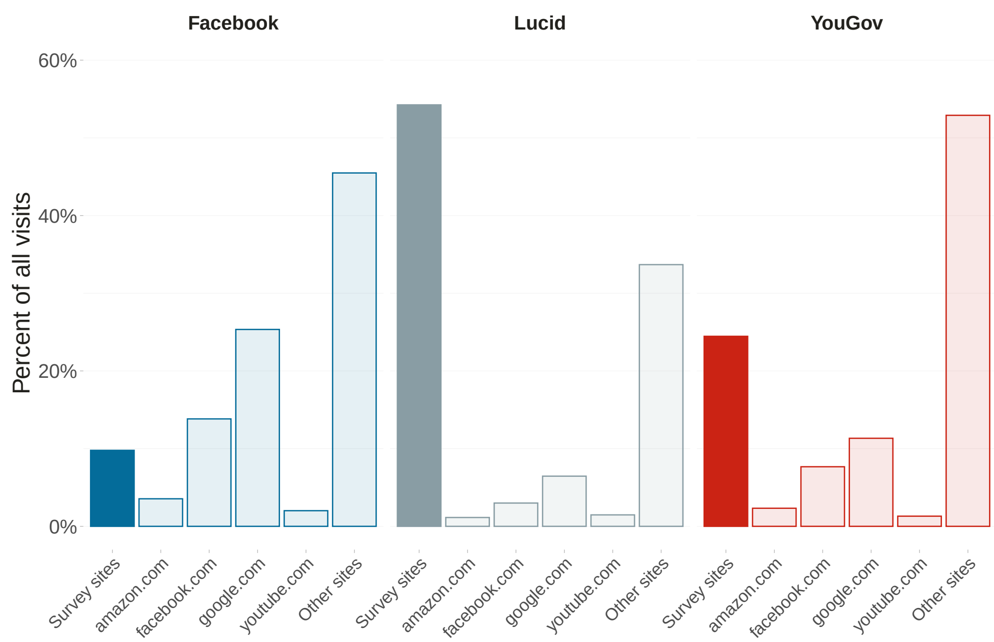

Figure C.2: Percent of visits to survey sites (out of all visits) compared to visits to popular web domains, unweighted.

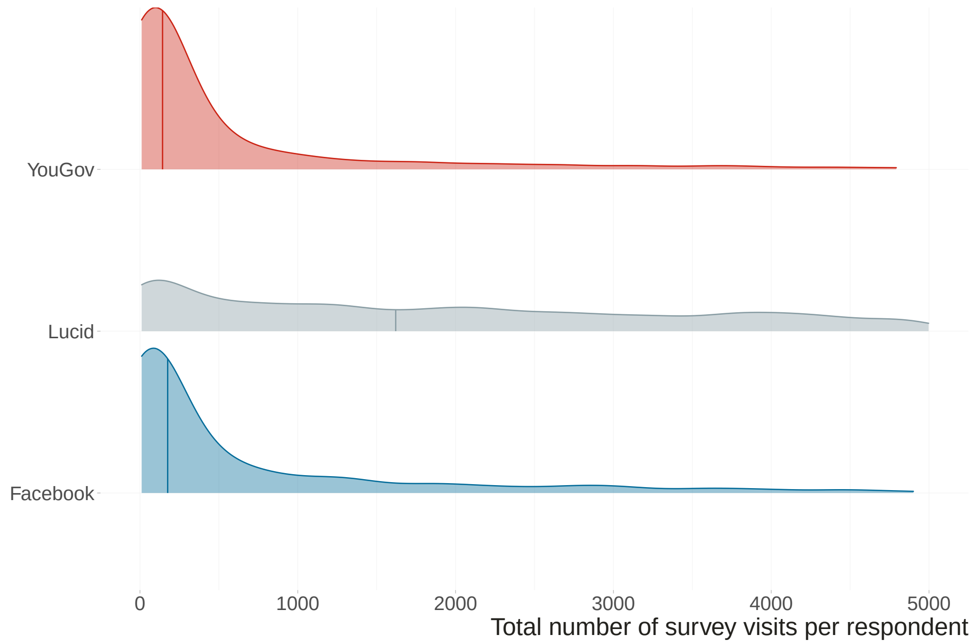

Figure C.3: Individual level distribution of survey visits, unweighted.

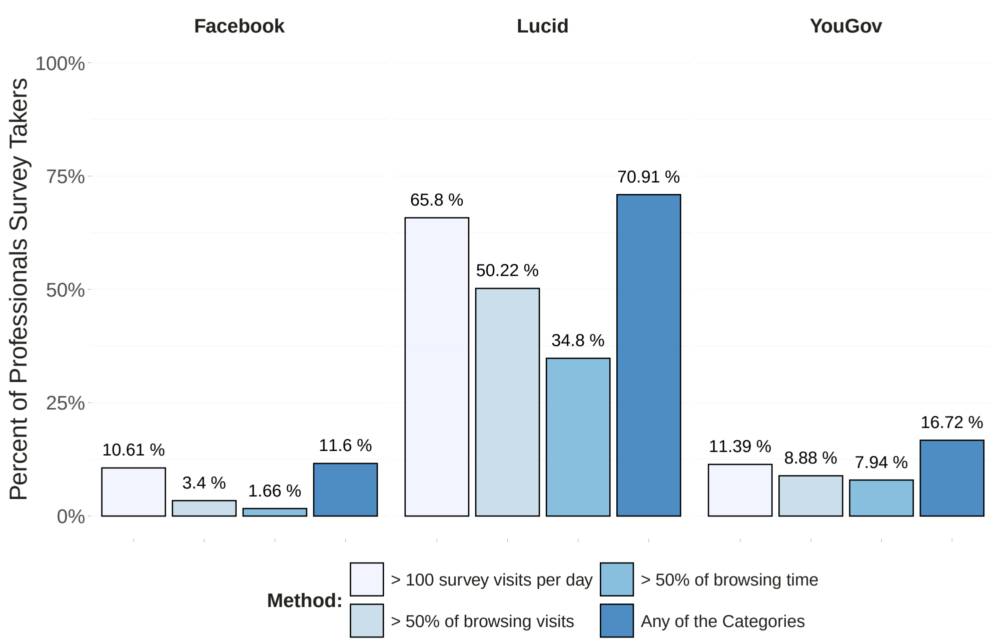

Figure C.4: Percent of survey professionals for different definitions of survey professionalism, unweighted.

C.2 Disaggregation by desktop/mobile users

Analogously to Figures 1, 2, and 3 in the main paper, Figures C.5, C.6 and C.7 show the prevalence of survey professionalism for Yougov, disaggregated into desktop and mobile users. We classify respondents as desktop users if more than 90% of their browsing originates from a laptop or computer, and mobile users if more than 90% of their browsing comes from a mobile device. We use this definition (rather than 100%) because it identifies the primary mode in which participants donate data. Using this definition, we identify 530 (55.4%) of donors as primarily desktop users, 398 (41.6%) as mobile users, and 29 as mixed users. Since the number of mixed users is small, we refrain from drawing inferences about this group.

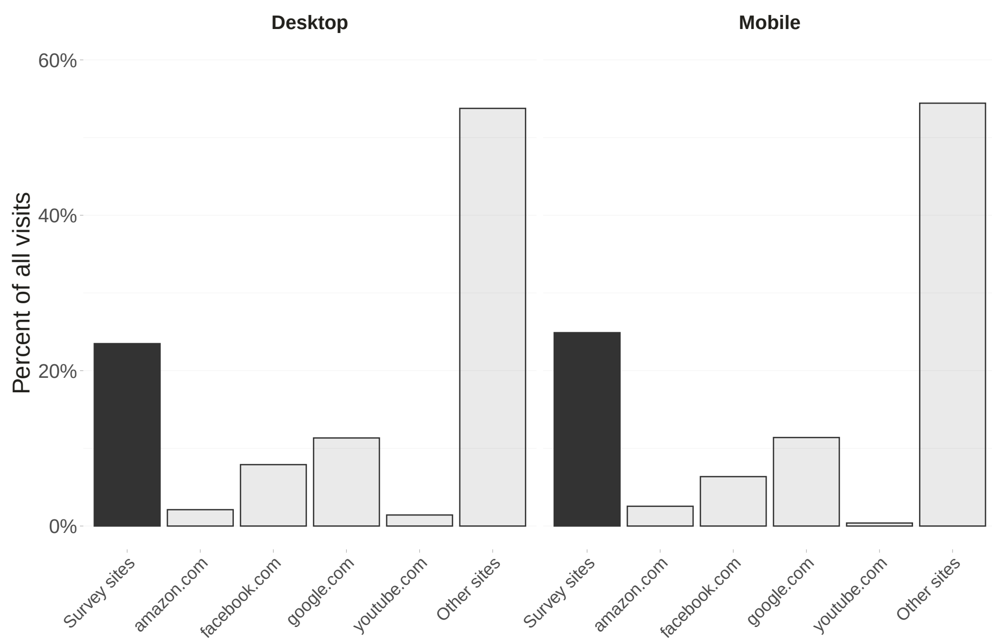

Figure C.5: Percent of visits to survey sites (out of all visits) compared to visits to popular web domains, for Yougov sample and disaggregating desktop and mobile users.

C.3 Disaggregation by identification approaches and survey sites

Figure C.8 shows the same measures as Figure 1 in the main paper, but splits survey visits according to identification method. Figure C.9 shows the ten most frequent URL hosts among the survey visits and reports how much of the total each accounts for.

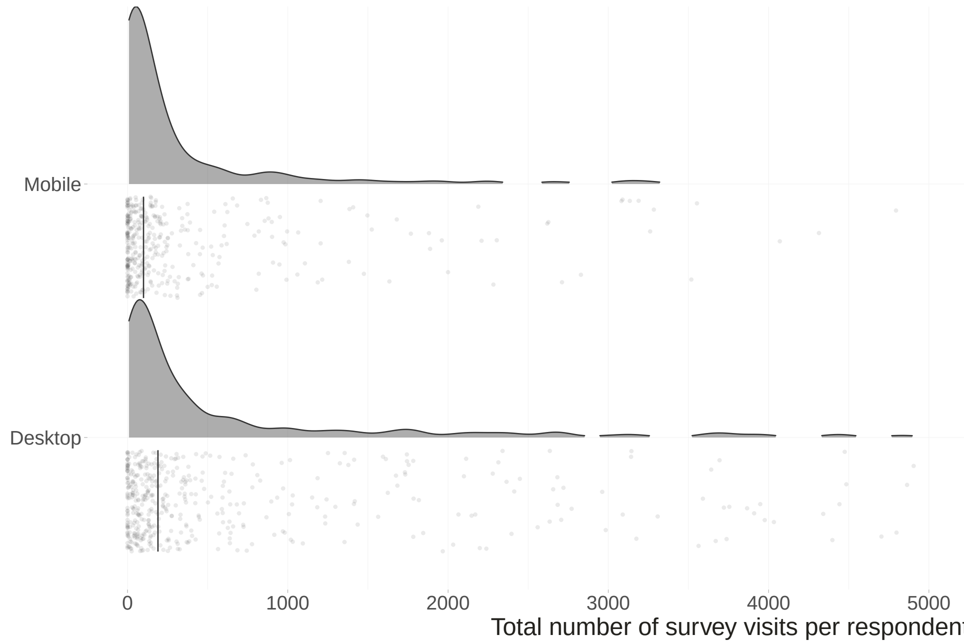

Figure C.6: Individual level distribution of survey visits, for Yougov sample and disaggregating desktop and mobile users.

C.4 Average individual-level percentage of survey visits

Figures C.10 takes a slightly different perspective on the data, reporting individual-level survey taking. Figure C.10 shows the average individual-level percentage of survey visits. This statistic is calculated by dividing each individual's number of survey visits by his or her total number of visits and take the average of these proportions.

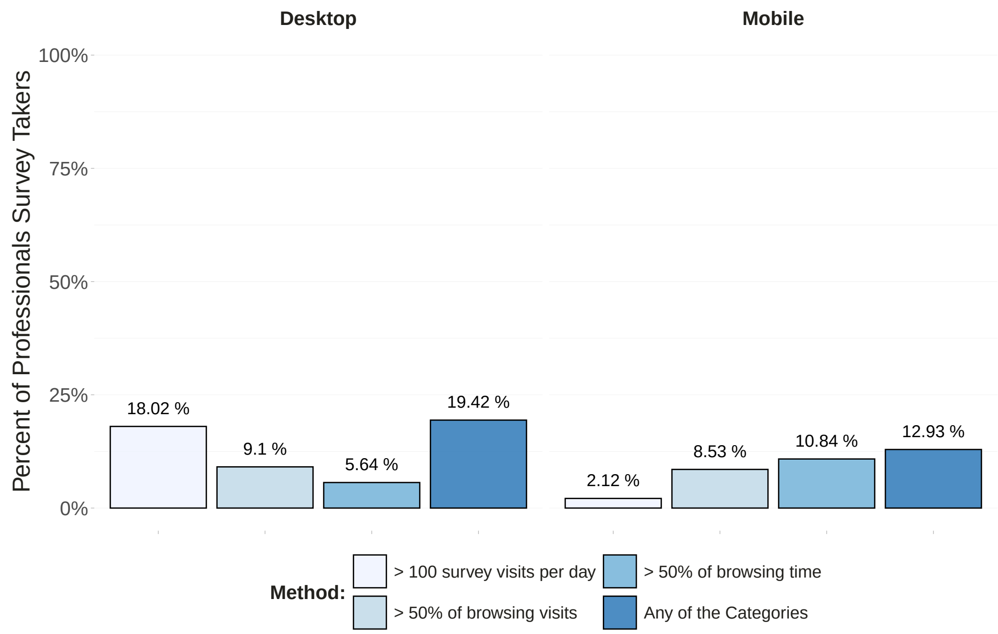

Figure C.7: Percent of survey professionals for different definitions of survey professionalism, for Yougov sample and disaggregating desktop and mobile users.

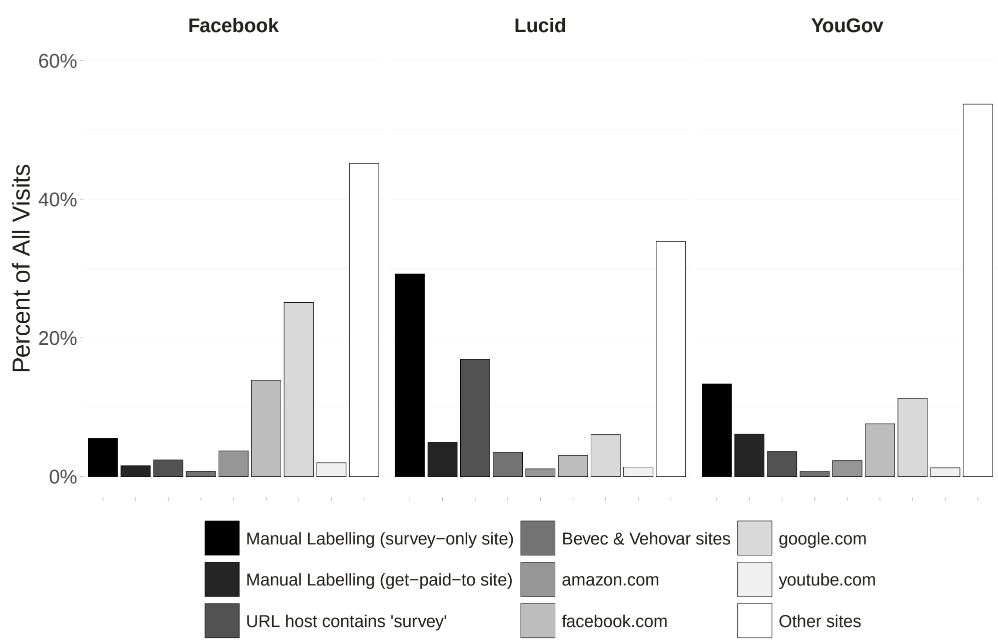

Figure C.8: Percent of visits to survey sites (out of all visits), split by method, and compared to popular web domains.

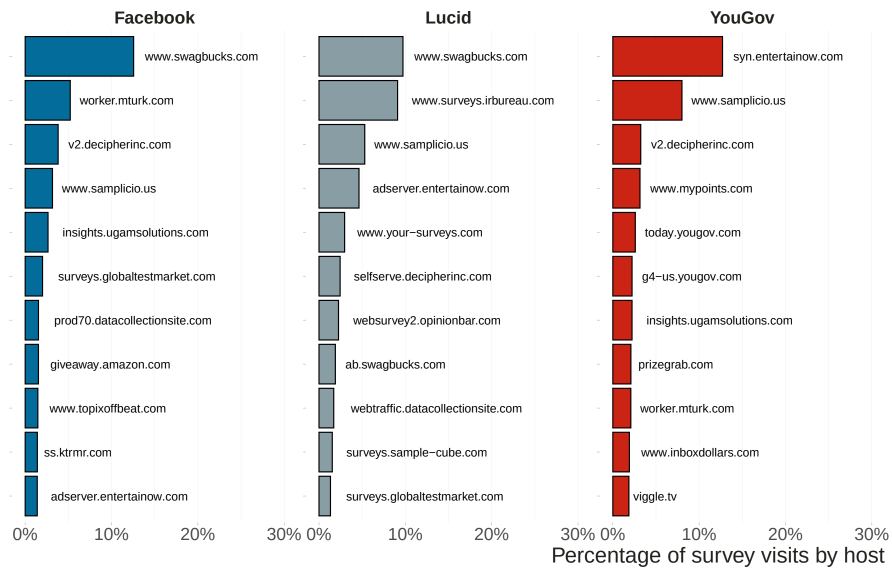

Figure C.9: Ten most frequented survey sites, by sample.

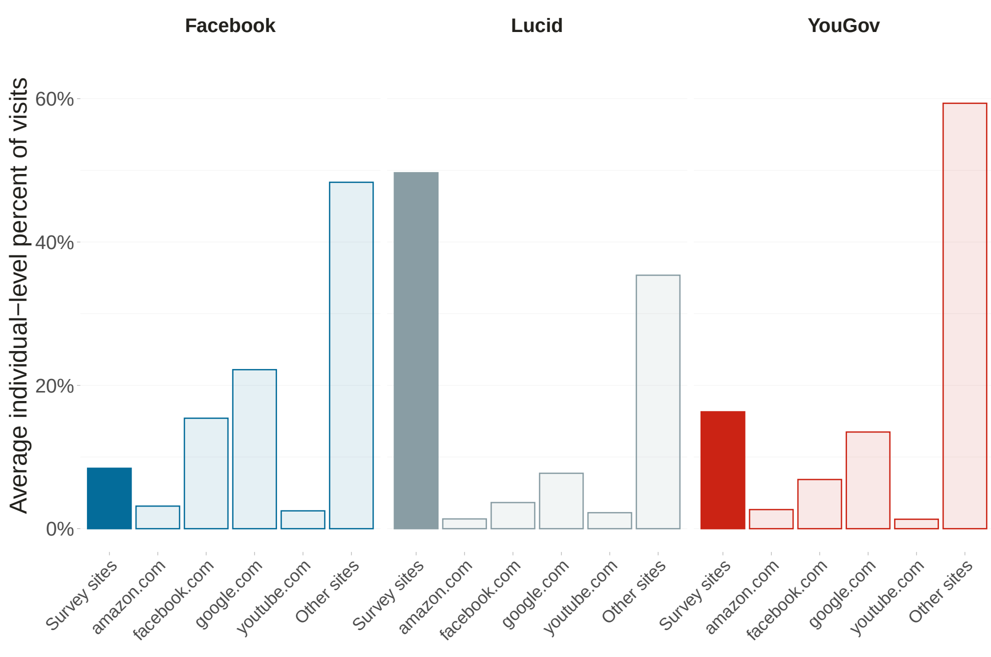

Figure C.10: Average individual-level percent of visits to survey sites out of all visits, compared to popular web domains.

C.5 Analogous results based on duration

Figure C.11 shows, analogously to Figure 1 in the main paper, the aggregate proportion of survey-taking time out of all browsing time. Figure C.12 shows, again, the average individual-level percentage of survey-taking time out of all browsing, and Figure C.13 the distribution of this individual-level percentage.

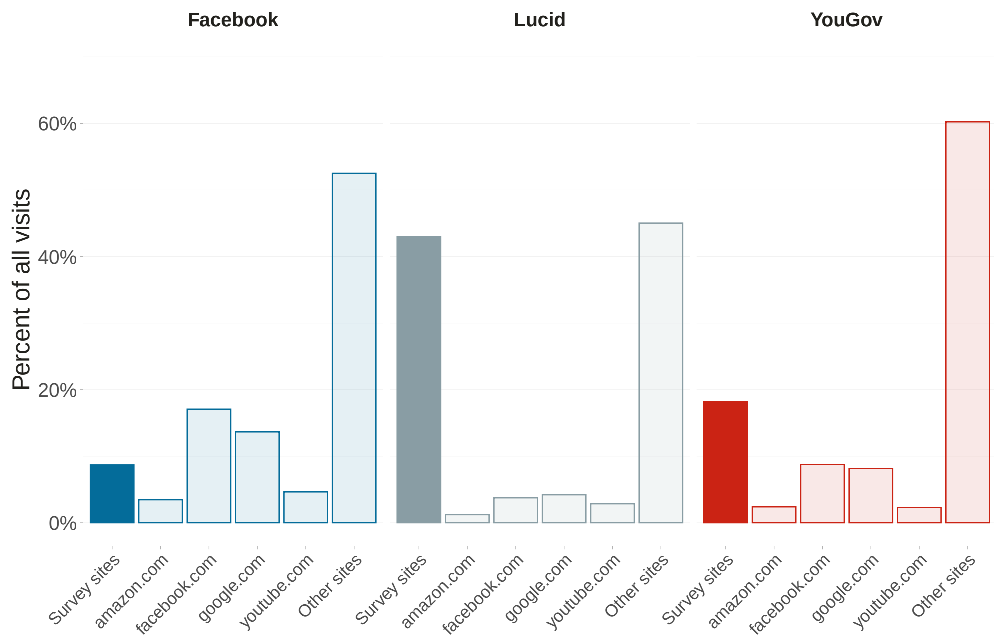

Figure C.11: Percent of visit time to survey sites (out of all browsing time).

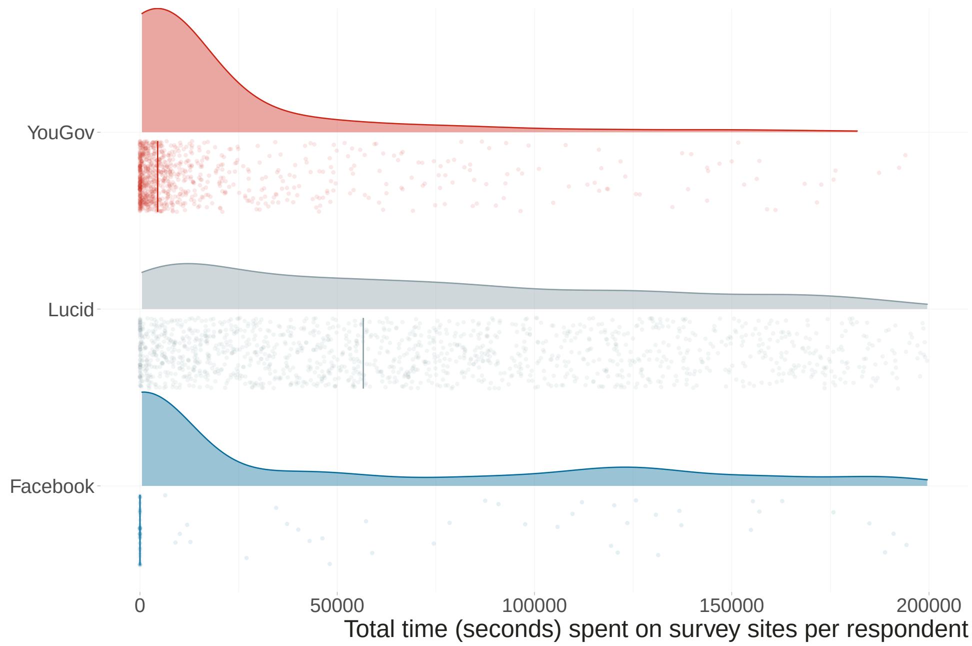

Figure C.12: Distribution of individual-level percent of visit time to survey sites out of all browsing time

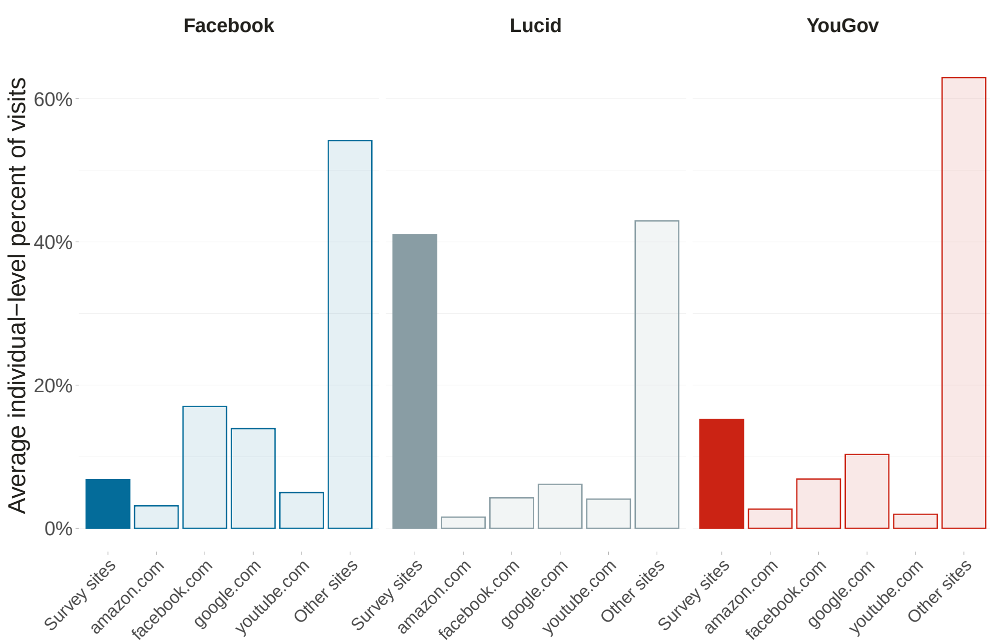

Figure C.13: Distribution of individual-level percent of visit time to survey sites out of total browsing time.

C.6 Sensitivity Analysis for the Prevalence of Professionalism

In the manuscript, we present all results for the prevalence of professionalism using weighted results to approximate our target population of online survey takers. We acknowledge that weighting only addresses concerns about representativeness insofar as observables are concerned. One might believe that survey professionalism is strongly related to participants' willingness to donate their web-browsing data, which – if true – would imply that we overestimate the extent of survey professionalism.

In this section, we provide two additional pieces of information to address this concern. First, we look at the levels of attrition between waves for donors and non-donors in our sample. As the Facebook and Lucid studies only invited donors to do the full survey (and the later waves), we can only perform this test for the Yougov sample. The logic behind the test is simple: if survey professionals are more willing to do things in surveys for money, then we would expect them to be more likely than non-professionals to accept the offer to do a follow-up survey (i.e., 2nd wave) in a panel. If survey professionals are more common among donors, we should see that donors have lower attrition rates (i.e., higher completion rates) than non-donors. When we compare donors and non-donors in the Yougov sample, we do not see any meaningful differences (less than 2 percentage points) in completion rates, as figure C.14 shows. Although by no means conclusive, we take this as suggestive evidence that professionals are not more likely to share their browsing data than non-professionals, since they are not more likely than non-donors to accept the monetary reward of doing the follow-up survey.

Second, we perform a sensitivity analysis to add bounds to the percentage of survey professionals across our samples. The critical assumption we make in our analysis in the paper is that given the set of observables we weight on, the prevalence of professionals is the same across donors and non-donors. Since unobserved cofounders may affect both the willingness to donate browsing data and survey professionalism, this assumption may be incorrect. Our sensitivity analysis relaxes this assumption, decomposing the prevalence of professionals across the two types of users (donors and non-donors), and allowing the share of professionalism to vary between these types (which our weighted results assume to be constant) [5].

For example, what would our estimates likely be if among donors, there were 50%

5. This sensitivity analysis requires two additional assumptions. First, it requires us to assume that the share of donors and non-donors we observe in our samples is representative of these types in the population of online survey takers on each platform. Second, we need to assume that no donor type exists in the non-donor data. With these two assumptions, we can hold constant the share of donors and non-donors in the sensitivity analysis and vary only one parameter in the decomposition

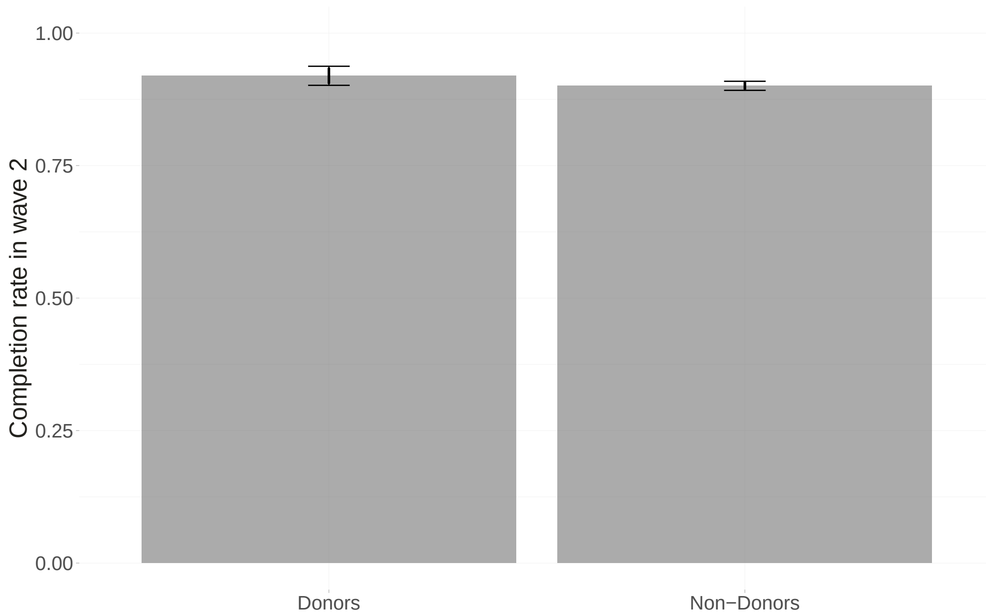

Figure C.14: Attrition rate of donors on non-donors on YouGov

more or 50% fewer professionals than among non-donors? Figure C.15 shows the sensitivity of our estimates following this logic. We examine each of the survey platforms (listed on the right side of the panel), and present results based on each of the measures of survey professionalism (listed on the left side of the panel). We vary the share of professional among non-donors from being equal to the number of professionals among donors (0% change) to up to 50% more or 50% fewer survey professionals. For example, on Lucid, 84.2% of respondents are non-donors and 15.8% are donors. Among donors, we find that 66% (based on our measure of greater than 100 visits/day) are professionals. Now if we assume, per our sensitivity analysis, that among non-donors the share of professionals changes in 50%, the lower bound in the target population would be 38.5%, and the upper bound would be 94.5% of survey takers are professionals. The bounds are wide on Lucid, but the share of professions is still substantial.

And even with these broad bounds on the relative prevalence of professionals in the donor and non-donor samples, the share of professionals for Facebook and Yougov does not changed much. For instance, again using the criteria of greater than 100 visits per day, given our parameters we bound the estimate for percentage of respondents being survey professionals between 7.1% and 14.8% on Facebook, and between 7.6% and 17.6% on YouGov. This sensitivity analysis provides the reader with the necessary caveats to interpret the prevalence measures discussed in the paper.

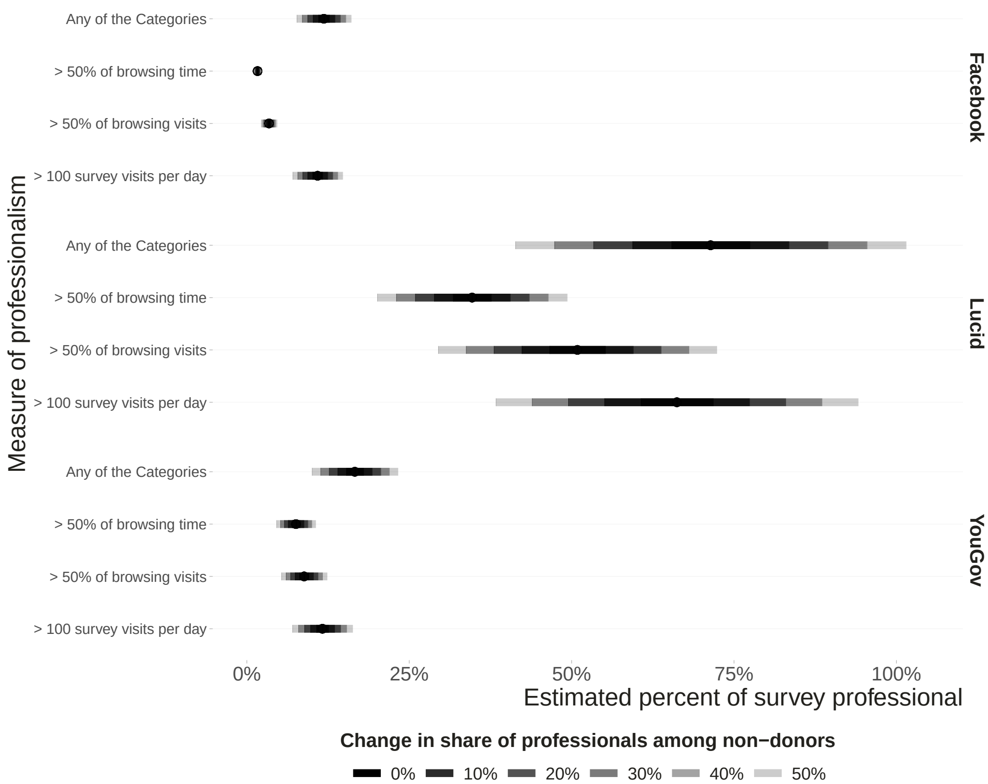

Figure C.15: Sensitivity analysis for the prevalence of survey professionals

D Sociodemographic and political differences (RQ2): Additional results

D.1 Alternative professionalism indicators

Below, we report sociodemographic and political differences between professionals and non-professionals when professionals are defined as those with more 50 percent visits to survey sites (Table D.4), when defined as anyone with more than 50 percent of browsing time to survey sites (Table D.5), and when meeting any of the three criteria (Table D.6).

Table D.4: Survey professionals vs. non-professionals vs. population (professional = more than 50 percent of visits to survey sites)

		Facebook			Lucid			Yougov		
	U.S. population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	45-49		35-39	46	***	41	56		55
Gender (% female)	50.8	61.1 (10.8)		75.3 (1.7)	53.4 (1.6)		55.2 (1.6)	61.6 (5.8)		54.9 (1.9)
Education (% Bachelor or more)	30.4	38.5 (10.8)		56.6 (1.9)	45.5 (1.6)		45.2 (1.6)	27.9 (5.7)	*	42.4 (2.0)
Ethnicity (% white)	62.6	95.3 (4.6)		84.5 (1.4)	80.9 (1.2)		79.3 (1.3)	73.9 (5.0)		77.9 (1.5)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.7 (0.6)		3.2 (0.1)	3.8 (0.1)		3.7 (0.1)	3.7 (0.3)		3.6 (0.1)
Ideology (0-1)	0.54 (0.006)	0.47 (0.06)		0.41 (0.01)	0.55 (0.01)	***	0.50 (0.01)	0.55 (0.03)		0.50 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	28.2 (5.0)		25.3 (0.9)	29.7 (0.9)	*	26.8 (0.8)	20.4 (3.0)	*	12.7 (0.9)
Political interest (0-1)	0.4 (0.006)	0.64 (0.06)		0.66 (0.01)	0.70 (0.01)	**	0.65 (0.01)	0.69 (0.04)	**	0.82 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.64 (0.01)		0.64 (0.01)	0.59 (0.04)		0.68 (0.01)
Following politics (0-1)		0.62 (0.06)		0.61 (0.01)				0.54 (0.04)	***	0.69 (0.01)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for age, chi-squared tests for gender, education and race, and t-tests for all other variables ($\circ$ $p < 0.1$; $\star$ $p < 0.05$; $\star\star$ $p < 0.01$; $\star\star\star$ $p < 0.001$). Sociodemographic population data from the US Census; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

Table D.5: Survey professionals vs. non-professionals vs. population (professional = more than 50 percent of browsing time to survey sites)

		Facebook			Lucid			Yougov		
	U.S. population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	5 (1)		5 (0)	43	**	43	54		56
Gender (% female)	50.8	48.6 (14.5)	*	75.2 (1.7)	53.7 (1.9)		54.6 (1.4)	64.9 (6.0)		54.8 (1.9)
Education (% Bachelor or more)	30.4	48.2 (14.5)		56.5 (1.9)	46.7 (1.9)		44.6 (1.4)	27.3 (5.8)	*	42.2 (2.0)
Ethnicity (% white)	62.6	91.6 (8.0)		84.4 (1.4)	79.9 (1.5)		80.2 (1.1)	72.8 (5.3)		77.9 (1.5)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.0 (0.6)		3.2 (0.1)	3.8 (0.1)		3.7 (0.1)	4.0 (0.3)		3.6 (0.1)
Ideology (0-1)	0.54 (0.006)	0.42 (0.06)		0.41 (0.01)	0.55 (0.01)	**	0.51 (0.01)	0.55 (0.04)		0.51 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	35.0 (6.3)		25.4 (0.9)	30.9 (1.1)	**	26.8 (0.7)	21.7 (3.3)	**	12.7 (0.9)
Political interest (0-1)	0.4 (0.006)	0.74 (0.09)		0.66 (0.01)	0.69 (0.02)		0.66 (0.01)	0.64 (0.04)	***	0.82 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.63 (0.02)		0.65 (0.01)	0.53 (0.05)	**	0.68 (0.01)
Following politics (0-1)		0.68 (0.09)		0.61 (0.01)				0.49 (0.04)	***	0.69 (0.01)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for age, chi-squared tests for gender, education and race, and t-tests for all other variables ($\circ$ $p < 0.1$; $\star$ $p < 0.05$; $\star\star$ $p < 0.01$; $\star\star\star$ $p < 0.001$). Sociodemographic population data from the US Census; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

Table D.6: Survey professionals vs. non-professionals vs. population (professional = any of the categories)

		Facebook			Lucid			Yougov		
	U.S. population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	40-44	**	35-39	46	***	36	56		55
Gender (% female)	50.8	72.9 (5.2)		75.0 (1.8)	53.9 (1.3)		55.1 (2.1)	58.6 (4.4)		54.9 (2.0)
Education (% Bachelor or more)	30.4	55.9 (5.8)		56.4 (2.0)	46.4 (1.3)		42.6 (2.1)	36.1 (4.5)		42.0 (2.1)
Ethnicity (% white)	62.6	89.9 (3.4)		83.9 (1.5)	81.2 (1.0)	*	77.1 (1.8)	74.7 (3.6)		78.1 (1.6)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.6 (0.3)		3.1 (0.1)	3.8 (0.1)		3.7 (0.1)	3.7 (0.2)		3.6 (0.1)
Ideology (0-1)	0.54 (0.006)	0.48 (0.03)	*	0.40 (0.01)	0.54 (0.01)	***	0.49 (0.01)	0.55 (0.02)		0.50 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	28.4 (3.2)		25.1 (1.0)	28.3 (0.7)		28.6 (1.1)	20.2 (2.4)	**	12.0 (0.9)
Political interest (0-1)	0.4 (0.006)	0.65 (0.03)		0.66 (0.01)	0.69 (0.01)	*	0.63 (0.02)	0.70 (0.03)	***	0.83 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.65 (0.01)		0.62 (0.02)	0.61 (0.03)	*	0.68 (0.01)
Following politics (0-1)		0.64 (0.03)		0.61 (0.01)				0.56 (0.03)	***	0.70 (0.01)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for age, chi-squared tests for gender, education and race, and t-tests for all other variables ($\circ$ $p < 0.1$; $\star$ $p < 0.05$; $\star\star$ $p < 0.01$; $\star\star\star$ $p < 0.001$). Sociodemographic population data from the US Census; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

E Response-quality differences (RQ3): additional results

E.1 Speeding and straightlining (RQ3a,b)

E.1.1 Alternative professionalism indicators

Below, we report differences in speeding and straightlining when professionals are defined as those with more 50 percent visits to survey sites (Table E.7), when defined as anyone with more than 50 percent of browsing time to survey sites (Table E.8), and when meeting any of the three criteria (Table E.9).

Table E.7: Response quality of survey professionals vs. non-professionals (professionals = more than 50 percent visits to survey sites)

	Facebook		Lucid			Yougov		
	Professionals	Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Straightliner (%)	4.1 (4.1)	1.0 (0.4)	2.2 (0.5)	○	1.2 (0.4)	3.2 (2.0)		3.4 (0.6)
Duration (median seconds)	833 (191)	833 (25)	1179 (23)		1203 (34)	1604 (241)		1787 (54)
Duration (% 30% faster than median)	26.0 (9.4)	21.8 (1.6)	22.1 (1.2)		22.7 (1.3)	32.4 (5.7)	⋆	21.3 (1.5)
Duration (% 40% faster than median)	26.0 (9.4)	14.6 (1.4)	15.2 (1.1)		15.7 (1.1)	20.7 (4.9)	○	13.1 (1.2)
Duration (% 50% faster than median)	17.7 (8.1)	8.3 (1.1)	9.9 (0.9)		10.8 (0.9)	7.6 (3.2)		6.6 (0.9)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for survey duration, and chi-squared tests for the proportion of those 30/40/50 percent faster than the median duration and for proportion of straightliners (○ $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$).

Table E.8: Response quality of survey professionals vs. non-professionals (professionals = more than 50 of browsing time to survey sites)

	Facebook		Lucid			Yougov		
	Professionals	Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Straightliner (%)	0.0 (0.0)	1.2 (0.4)	2.8 (0.6)	⋆	1.2 (0.3)	2.6 (1.9)		3.5 (0.6)
Duration (median seconds)	909 (250)	830 (24)	1166 (27)		1210 (26)	1847 (364)		1767 (47)
Duration (% 30% faster than median)	17.0 (11.0)	22.4 (1.6)	23.5 (1.5)		21.7 (1.1)	31.0 (6.1)	○	21.6 (1.5)
Duration (% 40% faster than median)	8.6 (8.3)	15.3 (1.4)	16.5 (1.3)		14.6 (0.9)	19.1 (5.2)		13.4 (1.2)
Duration (% 50% faster than median)	0.0 (0.0)	8.9 (1.1)	10.6 (1.1)		9.9 (0.8)	8.4 (3.5)		6.6 (0.9)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for survey duration, and chi-squared tests for the proportion of those 30/40/50 percent faster than the median duration and for proportion of straightliners (○ $p < 0.1$; ⋆ $p < 0.05$; ⋆⋆ $p < 0.01$; ⋆⋆⋆ $p < 0.001$).

Table E.9: Response quality of survey professionals vs. non-professionals (professionals = any of the measures)

	Facebook			Lucid		Yougov		
	Professionals		Non-professionals	Professionals	Non-professionals	Professionals		Non-professionals
Straightliner (%)	3.6 (2.0)	$\star$	0.8 (0.4)	1.9 (0.4)	1.1 (0.5)	5.7 (1.9)	$\circ$	2.9 (0.6)
Duration (median seconds)	725 (70)	$\circ$	843 (25)	1164 (19)	1291 (43)	1709 (100)		1787 (61)
Duration (% 30% faster than median)	34.5 (5.4)	$\star\star$	20.8 (1.6)	22.9 (1.1)	20.8 (1.6)	28.7 (4.1)	$\circ$	21.0 (1.6)
Duration (% 40% faster than median)	28.4 (5.1)	$\star\star\star$	13.5 (1.4)	15.5 (0.9)	15.0 (1.4)	18.5 (3.5)	$\circ$	12.9 (1.3)
Duration (% 50% faster than median)	13.5 (3.8)		8.1 (1.1)	10.1 (0.8)	10.3 (1.2)	8.4 (2.3)		6.4 (0.9)

Note: Standard errors of means in parentheses. Significance of differences between professionals and non-professionals were tested with a Kolgomorov-Smirnoff test for survey duration, and chi-squared tests for the proportion of those 30/40/50 percent faster than the median duration and for proportion of straightliners ($\circ$ $p < 0.1$; $\star$ $p < 0.05$; $\star\star$ $p < 0.01$; $\star\star\star$ $p < 0.001$).

E.2 Stability of Survey Responses (RQ3c)

E.2.1 Bayesian heteroscedastic linear model

Here, we more fully describe the Bayesian heteroscedastic linear model to identify the stability of survey responses between professionals and non-professionals. Recall the mathematical model introduced in the main paper:

$$Y_{iw_2} \sim \mathcal{N}(\mu_i, \sigma_i^2)$$
$$\mu_i \sim \alpha + \beta_1 Y_{iw_1} + \beta_2 SP_i + \beta_3 Y_{jw_1} \cdot SP_i + \epsilon_i$$
$$\sigma_i = SP_i \cdot \sigma_{\text{pro}} + (1 - SP_i) \cdot \sigma_{\text{nonpro}}$$

We assume non-informative priors for all parameters, and truncated at zero (so that the standard deviation is also positive) the prior distributions for the sigma parameters:

$$\alpha \sim \mathcal{N}(0, 1)$$
$$\beta_1 \sim \mathcal{N}(0, 1)$$
$$\beta_2 \sim \mathcal{N}(0, 1)$$
$$\beta_3 \sim \mathcal{N}(0, 1)$$
$$\sigma_{\text{pro}} \sim \mathcal{N}^+(0, 1)$$
$$\sigma_{\text{nonpro}} \sim \mathcal{N}^+(0, 1)$$

We use Stan to perform Bayesian inference for the model. Our quantity of interest is the difference in the posterior distribution between σ_{pro} and σ_{nonpro}. The full model specification in Stan is the following (also contained in our code repository):

```
// heteroscedastic model
data {
  int<lower=1> N;            // Number of observations
  vector[N] t1;              // Predictor (continuous)
  int<lower=0,upper=1> pro[N];  // Group indicator (0 = non-pro, 1 = pro)
  vector[N] t2;              // Outcome
}

parameters {
  real beta_0;               // Intercept
  real beta_1;               // Effect of t1
  real beta_2;               // Effect of pro
  real beta_3;               // Interaction t1 * pro
```

```
  real<lower=0> sigma_nonpro;  // SD for non-professionals
  real<lower=0> sigma_pro;     // SD for professionals
}


model {
  // Priors
  beta_0 ~ normal(0, 1);
  beta_1 ~ normal(0, 1);
  beta_2 ~ normal(0, 1);
  beta_3 ~ normal(0, 1);

  sigma_nonpro ~ normal(0, 1);
  sigma_pro ~ normal(0, 1);

  // Likelihood
  for (i in 1:N) {
    real sigma = pro[i] * sigma_pro + (1 - pro[i]) * sigma_nonpro;
    t2[i] ~ normal(beta_0 + beta_1 * t1[i] + beta_2 * pro[i] + beta_3 * t1[i] * pro[i], sigma);
  }
}
```

In addition to the distribution of z-scores for the estimates of interest reported in the main paper(Figure 4, Figure E.16 presents the median and 95%-credibility intervals from posterior distribution for the z-scores. These results illustrate the difference in the posterior distributions of the parameters for the standard deviation of professionals and non-professionals.

E.2.2 Absolute-difference OLS model

As an alternative modeling approach, we regress the absolute difference between and individual's wave-one and wave-two response on the indicator of professionalism. Figure E.17 and E.18 illustrate the results of these models, which are are similar to those discussed in the main paper.

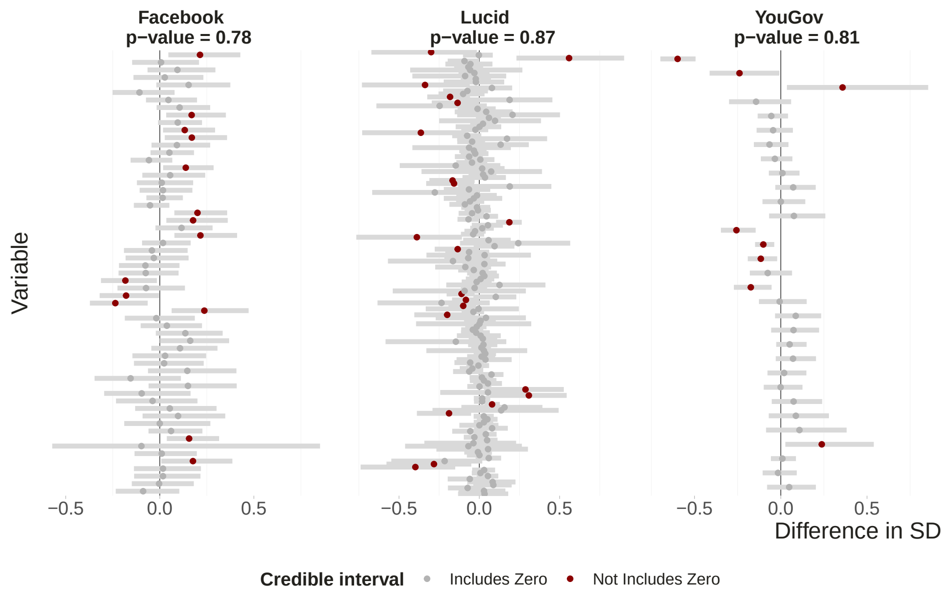

Figure E.16: Median posterior estimate and 95% credible intervals for the difference in standard deviation between professionalism and wave-one survey responses. Positive values indicates higher variance among Professionals. Red dots represents variables for which the interval does not include an estimate of zero.

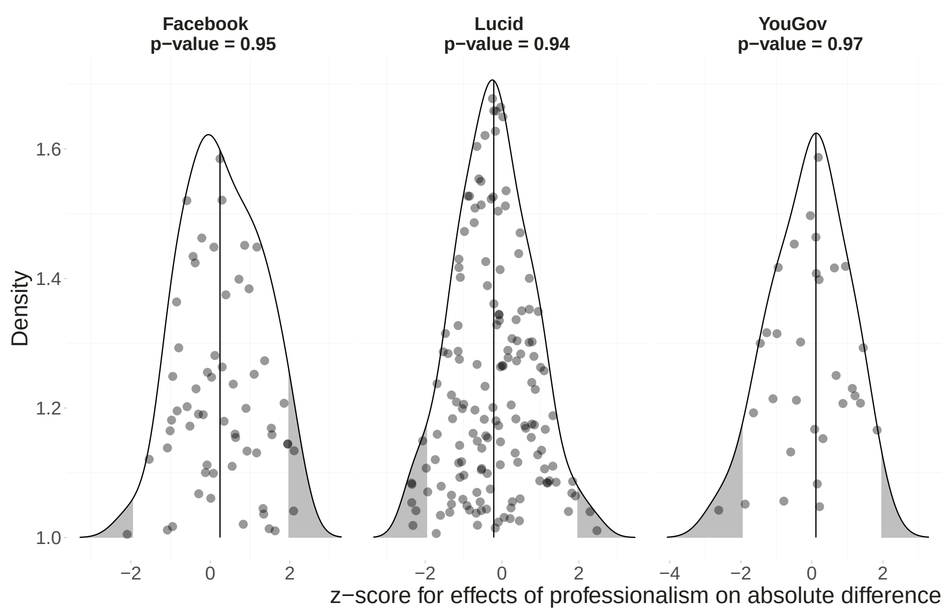

Figure E.17: Z-scores for the effects of professionalism and wave on the absolute difference of wave-two and wave-one responses. Gray areas contain z-scores larger than 1.96.

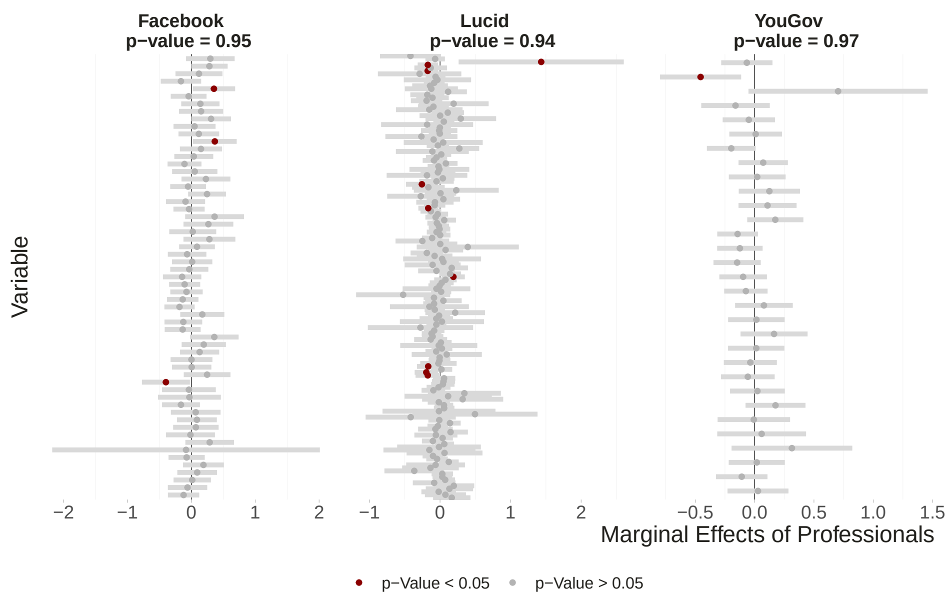

Figure E.18: Point estimates for the effects of professionalism on the absolute difference of wave-two and wave-one responses along with 95% confidence intervals. Red dots represents variables for which the interval does not include an estimate of zero.

E.3 Treatment effect differences

In addition to differences in speeding, straightlining, we also analyzed differences in the effect of two priming experiments embedded in the Lucid survey. In the first experiment, treatment group subjects were exposed to the prime that "[i]t is important to be open to different points of view on political issues", before all subjects responded to a battery capturing the notion of "attribution of malevolence" (five items on a 7-point scale, for example, agreement to statement "I worry that [members of opposite party] are deliberately trying to hurt America. Note that this experiment was not analyzed for Independents. The second experiment primed the treatment group with the statement that "[l]ooking at news sources that present opposing political views is important to good citizenship" before a battery measuring perceived polarization (agreement to four items on a 7-point scale such as "Democrats and Republicans hate each other".).

To test whether these treatments affect professionals and non-professionals differently, we regress the two outcomes (recoded into indices) on a treatment dummy and an interaction between the treatment and a dummy for professionalism. We run separate models for each of the four definitions of professionalism. As Tables E.10 and E.11 show, none of the interactions is significantly different, which implies that professionalism does not matter for these treatment effects.

Table E.10: Effect of openness prime on attribution of malevolence, by professionalism

	Definition of professionalism:			
	> 100 visits/day	> 50% of visits	> 50% of time	Any of the three
Treatment	−0.14	−0.24**	−0.22**	−0.14
	(0.15)	(0.12)	(0.11)	(0.16)
Professional	0.22	−0.02	0.04	0.21
	(0.16)	(0.15)	(0.16)	(0.17)
Treatment * Professional	−0.10	0.07	0.004	−0.08
	(0.18)	(0.17)	(0.18)	(0.19)
Constant	4.55***	4.71***	4.69***	4.54***
	(0.13)	(0.11)	(0.10)	(0.15)
Observations	1,882	1,882	1,874	1,879
R^2	0.01	0.003	0.004	0.01
Adjusted R^2	0.004	0.002	0.002	0.004

Note: *p<0.1; **p<0.05; ***p<0.01

Table E.11: Effect of diversity prime on perception of polarization, by professionalism

	Definition of professionalism:			
	> 100 visits/day	> 50% of visits	> 50% of time	Any of the three
Treatment	−0.09	−0.11	-0.14^{*}	−0.08
	(0.10)	(0.08)	(0.07)	(0.11)
Professional	0.12	0.22^{**}	0.18^{*}	0.13
	(0.11)	(0.11)	(0.11)	(0.12)
Treatment * Professional	−0.08	−0.08	0.01	−0.09
	(0.13)	(0.12)	(0.12)	(0.13)
Constant	4.75^{***}	4.73^{***}	4.76^{***}	4.74^{***}
	(0.09)	(0.07)	(0.07)	(0.10)
Observations	2,151	2,151	2,142	2,147
R^2	0.003	0.01	0.01	0.003
Adjusted R^2	0.002	0.01	0.01	0.002

Note: *p<0.1; **p<0.05; ***p<0.01

F Repeated questionnaire participation (RQ4): additional results

F.1 Alternative time cutoffs

Below, we report statistics on repeated participation with different time cutoffs—that is, when several visits to the same questionnaire URL only count as repeated when the difference is greater than six hours (Tables F.12 and F.13) and when greater than 24 hours (Tables F.14 and F.15).

Table F.12: Repeated questionnaire participation (6-hour cutoff)

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)	24.2 (1.6)	69.9 (1.0)	13.9 (1.2)
Number of repeated questionnaires per participant (mean)	0.9 (0.1)	7.6 (0.4)	0.5 (0.1)
Percent of repeated questionnaires per participants (mean)	4.1 (0.4)	6.2 (0.1)	1.7 (0.3)

Note: Standard errors in parentheses.

Table F.13: Repeated questionnaire participation (6-hour cutoff), professionals vs. non-professionals (professionals = more than 50 of browsing time to survey sites)

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	18.6 (1.5)	72.7 (5.2)	42.7 (1.9)	83.7 (1.0)	6.6 (0.9)	69.2 (4.7)
Number of repeated questionnaires per participant (mean)	0.5 (0.1)	4.5 (0.8)	1.5 (0.2)	10.8 (0.6)	0.1 (0.0)	3.3 (0.5)
Percent of repeated questionnaires per participants (mean)	3.9 (0.5)	5.6 (0.9)	4.9 (0.3)	6.8 (0.1)	1.1 (0.3)	6.4 (0.6)

Note: Standard errors in parentheses.

Table F.14: Repeated questionnaire participation (24-hour cutoff)

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)	22.5 (1.6)	68.2 (1.0)	13.4 (1.2)
Number of repeated questionnaires per participant (mean)	0.8 (0.1)	6.8 (0.4)	0.4 (0.1)
Percent of repeated questionnaires per participants (mean)	3.2 (0.3)	5.5 (0.1)	1.5 (0.3)

Note: Standard errors in parentheses.

Table F.15: Repeated questionnaire participation (24-hour cutoff), professionals vs. non-professionals (professionals = more than 50 of browsing time to survey sites)

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	17.0 (1.5)	69.8 (5.4)	40.2 (1.9)	82.4 (1.1)	6.2 (0.9)	68.6 (4.7)
Number of repeated questionnaires per participant (mean)	0.4 (0.1)	4.0 (0.7)	1.4 (0.1)	9.6 (0.5)	0.1 (0.0)	2.9 (0.5)
Percent of repeated questionnaires per participants (mean)	3.0 (0.4)	4.9 (0.8)	4.4 (0.3)	6.0 (0.1)	1.0 (0.3)	5.4 (0.5)

Note: Standard errors in parentheses.

F.2 Alternative professionalism indicators

Below, we report differences in attempts at repeated questionnaire participation when professionals are defined as those with more 50 percent visits to survey sites (Table F.16), when defined as anyone with more than 50 percent of browsing time to survey sites (Table F.17), and when meeting any of the three criteria (Table F.18).

Table F.16: Repeated questionnaire participation (1-hour cutoff), professionals vs. non-professionals (professionals = more than 50 percent visits to survey sites)

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	25.2 (1.7)	72.4 (9.1)	58.9 (1.5)	83.8 (1.2)	11.6 (1.2)	53.9 (5.7)
Number of repeated questionnaires per participant (mean)	0.9 (0.2)	5.0 (1.6)	3.9 (0.3)	12.5 (0.8)	0.3 (0.1)	2.3 (0.4)
Percent of repeated questionnaires per participants (mean)	4.9 (0.5)	7.7 (2.0)	6.2 (0.3)	7.4 (0.2)	1.9 (0.3)	4.8 (0.6)

Note: Standard errors in parentheses.

Table F.17: Repeated questionnaire participation (1-hour cutoff), professionals vs. non-professionals (professionals = more than 50 of browsing time to survey sites)

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	26.4 (1.7)	69.0 (13.2)	65.9 (1.3)	83.0 (1.4)	13.0 (1.2)	43.7 (6.0)
Number of repeated questionnaires per participant (mean)	1.0 (0.1)	7.8 (3.8)	5.9 (0.4)	12.8 (1.0)	0.4 (0.1)	2.0 (0.4)
Percent of repeated questionnaires per participants (mean)	5.1 (0.5)	4.4 (1.3)	6.5 (0.2)	7.4 (0.2)	2.1 (0.3)	3.7 (0.6)

Note: Standard errors in parentheses.

Table F.18: Repeated questionnaire participation (1-hour cutoff), professionals vs. non-professionals (professionals = any of the measures)

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	21.1 (1.6)	74.0 (4.9)	44.5 (2.0)	82.8 (1.0)	7.1 (1.0)	56.4 (4.1)
Number of repeated questionnaires per participant (mean)	0.6 (0.1)	4.9 (0.9)	1.7 (0.2)	10.9 (0.6)	0.1 (0.0)	2.5 (0.4)
Percent of repeated questionnaires per participants (mean)	4.9 (0.5)	6.2 (0.9)	5.7 (0.4)	7.3 (0.2)	1.5 (0.4)	5.4 (0.5)

Note: Standard errors in parentheses.

F.3 Disaggregation by questionnaire platforms

Table F.19: Repeated questionnaire participation (1-hour cutoff), by questionnaire platform

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)			
Confirmit	15.83	53.59	26.81
Dynata	27.57	60.24	50.17
Formsite	22.01	27.17	0.00
Formstack	23.99	17.18	25.05
Qualtrics	25.74	66.56	NA
Questionpro	18.63	28.06	0.00
Surveygizmo	18.22	56.75	0.00
Surveymonkey	17.88	14.17	5.50
Typeform	12.01	9.79	8.53
Unipark	6.14	11.31	0.00
Zoho	30.31	12.89	14.93
Number of repeated questionnaires per participant (mean)			
Confirmit	0.23	2.55	0.51
Dynata	0.74	2.23	1.46
Formsite	0.22	0.32	0.00
Formstack	0.41	0.21	0.25
Qualtrics	0.64	5.73	NA
Questionpro	0.26	0.92	0.00
Surveygizmo	0.28	1.97	0.00
Surveymonkey	0.61	0.24	0.10
Typeform	0.20	0.23	0.12
Unipark	0.06	0.15	0.00
Zoho	1.29	0.21	0.20
Percent of repeated questionnaires per participants (mean)			
Confirmit	3.57	6.25	5.44
Dynata	4.33	6.27	8.49
Formsite	13.62	23.10	0.00
Formstack	18.78	14.18	17.93
Qualtrics	5.73	10.20	NA
Questionpro	9.02	5.20	0.00
Surveygizmo	5.60	8.90	0.00
Surveymonkey	7.17	3.37	1.91
Typeform	4.81	3.37	2.35
Unipark	3.07	4.16	0.00
Zoho	8.84	3.57	7.45

F.4 Disaggregation by hard-to-reach groups

Table F.20: Repeated questionnaire participation (1-hour cutoff) among hard to reach groups (age)

	Facebook		Lucid		Yougov	
	Age >= 65	Age < 65	Age >= 65	Age < 65	Age >= 65	Age < 65
Subjects taking at least one questionnaire repeatedly (%)	29.2 (7.4)	26.7 (1.7)	72.5 (3.5)	70.4 (1.1)	16.7 (3.2)	14.0 (1.4)
Number of repeated questionnaires per participant (mean)	1.6 (0.7)	1.0 (0.2)	6.2 (0.9)	6.0 (0.3)	0.3 (0.1)	0.5 (0.1)
Percent of repeated questionnaires per participants (mean)	3.7 (1.2)	5.1 (0.5)	6.5 (0.4)	6.6 (0.2)	2.3 (0.7)	2.2 (0.4)

Table F.21: Repeated questionnaire participation (1-hour cutoff) among hard to reach groups (race)

	Facebook		Lucid		Yougov	
	White	Non-White	White	Non-White	White	Non-White
Subjects taking at least one questionnaire repeatedly (%)	24.9 (4.2)	27.8 (1.9)	71.3 (2.3)	70.5 (1.2)	15.1 (2.4)	14.6 (1.5)
Number of repeated questionnaires per participant (mean)	0.7 (0.3)	1.1 (0.2)	6.2 (0.7)	6.0 (0.3)	0.4 (0.1)	0.5 (0.1)
Percent of repeated questionnaires per participants (mean)	7.6 (2.0)	4.6 (0.5)	7.3 (0.4)	6.5 (0.2)	2.8 (0.7)	2.1 (0.4)

Table F.22: Repeated questionnaire participation (1-hour cutoff) among hard to reach groups (partisanship)

	Facebook		Lucid		Yougov	
	Democrat	Republican	Democrat	Republican	Democrat	Republican
Subjects taking at least one questionnaire repeatedly (%)	25.5 (2.2)	31.7 (3.7)	71.2 (1.4)	71.4 (1.8)	15.8 (1.9)	13.9 (2.3)
Number of repeated questionnaires per participant (mean)	0.9 (0.2)	1.7 (0.5)	6.2 (0.4)	6.0 (0.4)	0.5 (0.1)	0.5 (0.2)
Percent of repeated questionnaires per participants (mean)	4.7 (0.7)	5.5 (1.0)	6.5 (0.2)	6.9 (0.3)	2.7 (0.6)	1.7 (0.4)

F.5 Break patterns as an alternative explanation

Below, we explore a potential alternative explanation of repeated questionnaire participation, namely breaks that participants take from answering questionnaires. If a participant takes a particularly long time to complete a question, chances are that he or she did something else in between, such as visiting other web pages. This could generate multiple visits to the same questionnaire URL in our data, even though conceptually, they should not be understood as a repeated participation. Note, however, that browsing data do not record going back to a browser tab that is still open as a new visit, and also do not register refreshing a tab or using the back button as a new visit (see also SM B.2.1), although we are less certain about the exact mechanics in the case of Yougov because of the commercial nature of its collection tool. In other words, the scenario of a long question completion time with some hypothesized visits in between would only lead to separate visits if the participant completely closed the questionnaire or browser and re-opened it later. Hence, even if we saw that long completion times were common, it would not explain many of the cases of multiple questionnaire visits.

That said, we do not find that long question completion times in our Facebook sample as well as the three external data sets. For each of these data sets, we created a dummy for whether an individual ever took more than one hour for a question, as well as the alternative cutoffs of 6 hours and 24 hours. Table F.23 below the revised manuscript, shows that the percentage of people having ever taken more than 60 minutes for a question is 2.61% (Facebook), 0.09% (Lucid from 2021, Kane, Velez, and Barabas (2023)), 0.12% (Dynata from 2020, Clemm von Hohenberg (2023)) and 0.00% (Prolific from 2020, Clemm von Hohenberg (2023)). These percentages decrease, naturally, for longer cutoffs. We conclude that breaks are unlikely to explain a large share of repeated questionnaire participation.

Table F.23: Break patterns in Facebook sample and external studies

	Facebook	Lucid 2021	Dynata 2020	Prolific 2020
% any question > 1h	2.61	0.09	0.12	0
% any question > 6h	0.50	0.00	0.04	0
% any question > 24h	0.12	0.00	0.00	0