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]Department of Computational Social Science, GESIS Leibniz Institute
[2]McCourt School of Public Policy, Georgetown University
[3]Department of Politics, New York University
[4]Independent researcher
[5]Department of Communication, UC Davis

August 17, 2024

Abstract

Online panels have become an important resource for research in political science, but the financial compensation involved incentivizes respondents to become "survey professionals", which raises concerns about data quality. We provide evidence on survey professionalism using behavioral web browsing data from three U.S. samples, recruited via Lucid, YouGov, and Facebook (total n = 3,886). Survey professionalism is common but varies across samples: By our most conservative measure, we identify 1.7% of respondents on Facebook, 7.9% of respondents on YouGov, and 34.3% of respondents on Lucid as survey professionals. However, evidence that professionals lower data quality is limited: they do not systematically differ demographically or politically from non-professionals and do not respond more randomly—although they are somewhat more likely to speed, to straight-line, and to take questionnaires repeatedly. While concerns 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

Word count: 8,837

*Corresponding author. Email: bernhard.clemm@gesis.org

1 Introduction

Online panels have become an essential resource for low-cost survey research in political science and other fields. As scholars increasingly use them for data collection, the recruiting methods of many commercial providers raise urgent questions about data quality, such as a lack of representativeness and low attentiveness of participants (Krupnikov, Nam, and Style 2021; Cornesse and Blom 2023). A complicated web of panel companies, who often recruit via third-parties and even from one another, makes it hard for researchers to know where respondents come from (Enns and Rothschild 2022). Further, the goals of commercial providers may be at odds with scientific transparency (Jerit and Barabas 2023).

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 an exemplary advertisement in SM A). As survey professionals are motivated by monetary incentives (Silber, Stadtmüller, and Cernat 2023), they are likely to maximize their revenue by taking as many surveys as possible. Panel providers may turn a blind eye to such behavior, which creates a risk of samples being skewed towards professionals.

Yet, little is known about the extent of survey professionalism. Research into the phenomenon has hitherto relied entirely on self-reports, asking participants how often they do surveys or how many panels they belong to (e.g., Zhang et al. 2020; Matthijsse, De Leeuw, and Hox 2015). However, knowing that survey professionalism is undesirable for researchers, frequent survey participants are not likely to report their behaviors accurately. Hence, self-reports may not be the best way to get an empirical grip on survey professionalism.

Our article provides a novel empirical strategy to tackle this issue. We use *behavioral* measures constructed from web browsing data. This allows us, first, to identify the prevalence of survey professionalism, and then its consequences for researchers. We consider three potential downstream consequences for data quality. The first is that professionals may differ systematically from the general population, in terms of sociodemographics as well as political variables such as ideology (Krupnikov, Nam, and Style 2021). Numerous studies over the last two decades have compared online convenience samples to probability samples, and to the population, with mixed findings (Berinsky, Huber, and Lenz 2012; Malhotra and Krosnick 2007; Valentino et al. 2020). Less is known about whether and how professionals differ from non-professional survey takers in online panels. Second, survey professionalism comes with risks such as inattentive or random responses and speeding (Matthijsse, De Leeuw, and Hox 2015). Third, it is also possible that survey professionals, motivated by financial rewards (Paolacci, Chandler, and Ipeirotis 2010) learn to "play the game" and take the same questionnaire multiple times to increase their earnings (Ahler, Roush, and Sood 2019). All these potential aspects of survey professionalism have important consequences on the quality of data collected, and for the validity of the conclusions researchers make.

To examine the prevalence and the consequences of survey professionalism, we analyze three different U.S. samples from 2018/2019. For each, we have both survey responses and browsing data for multiple waves. The samples were recruited using different methods, through Facebook ads (n = 707), Lucid, now part of Cint (n = 2,222), and YouGov (n = 957), thus allowing us to explore the extent of survey professionalism across different recruitment approaches often used in political science and related fields. The browsing data of these three samples comprise over 95 million web visits in total. We identify survey taking based on existing lists of questionnaire websites, by using regular expressions to identify web addresses likely to constitute survey sites, and by manually coding the most frequent websites in our data.

We first report estimates of the extent of survey professionalism across the three different samples. Our results indicate that survey professionalism is most prevalent in the Lucid sample (54.3% of all visits, or 42.8% of time spent browsing), followed by YouGov (24.5% of all visits, or 18.9% of time spent browsing). We find the lowest prevalence of survey professionals in the sample recruited on Facebook (9.8% of all visits, or 8.4 % of time spent browsing). According to our most conservative dichotomous measure of survey professionalism, we identify 1.7% of Facebook respondents, 7.9% of YouGov respondents, and 34.3% of Lucid respondents as survey professionals. Second, we compare survey professionals and non-professionals on a range of sociodemographics and political characteristics, finding few consistent differences. Third, to shed light on issues of response quality, we compare the two groups in terms of straightlining, speeding, and over-time response stability: Although speeding through surveys is more common among survey professionals, we find only weak evidence that they are more likely to straightline through grid questions. We find no evidence that they change their responses to the same questions over time at a different rate than non-professionals, which would imply more random responding. Last, by analyzing patterns of visits to well-known questionnaire platforms, such as Qualtrics and, SurveyMonkey, we explore whether participants take the same questionnaire multiple times and how common this behavior is, finding that 40.2% for Facebook, 81.3% for Lucid, and 40.4% for YouGov sample percent of subjects took *at least one* questionnaire more than once for the period we collected browsing data. 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. "Survey professionals" do not differ from "on-professionals" consistently in terms of sociodemographic or political variables, and do not show more random responding. They may be more prone to speeding and straightlining, but these are behaviors that researchers can measure and contain by excluding respondents. The higher tendency to take questionnaires repeatedly is arguably the most worrying aspect of survey professionalism, and deserves further study.

2 Extent and consequences of survey professionalism

The literature has mostly understood survey professionalism in terms of the number of panels of which someone is a member, 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 surveys, based on questions such as "How many

Internet surveys have you completed before this one?" and "To how many other online panels do you belong?" (cf. Zhang et al. 2020). It does not easily translate into measurements using fine-grained web browsing data, since concepts such as "a survey" and "a panel" cannot be neatly isolated and counted in the vast and complex web browsing data a subject produces.

Hence, we offer a slightly different definition of survey professionalism, which closely follows the core ideas of previous work, while also allowing us to measure professionalism using web browsing data. Instead of targeting the number of surveys taken by a panelist, 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 answering survey questions (i.e., filling out a questionnaire) and also the steps necessary before and after: signing up on platforms offering survey jobs, selecting surveys on such platforms, and obtaining the rewards, whether monetary or in forms of vouchers. These various activities cannot be cleanly separated from each other, as many online platforms provide several functionalities. For example, some panels have their own questionnaire software, while others send people elsewhere to fill out questionnaires. In sum, we define and measure a subject's survey professionalism as the number (and duration) of survey visits. We offer details on measurement in Section 3.1.

We measure and analyze survey professionalism in two ways. First, we examine it as a continuous variable, i.e., someone can be more or less of a professional by having more or fewer survey visits or by spending 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 to compare panelists in terms of sociodemographics and a number of quality benchmarks, basing this categorization on the number and duration of survey visits. We build three distinct measures of survey professionals based on those 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 also measure whether (d) a subjects meets any of the three criteria above. We see the first three measures as alternative categorizations of survey professionals, and our last measure captures this assumption. For parsimony, we present the first approach in the main paper and report the others in the SM.

2.1 The extent of survey professionalism (RQ1)

From the early days of online survey research, observers have tried to quantify the extent of survey professionalism. A joint project by several survey companies reported substantial numbers of respondents to be members of 20 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. Characterizing the "the life of a panel member", Callegaro et al. (2014, p.48) write "that an 'average' panel member would spend about seven hours filling out questionnaires in a month with high burden panels topping about 16 hours a month [...]. These results are per single panel; if a respondent is a member of multiple panels, then the commitment quickly increases."

The issue with these estimates is that they rely on self-reports. As respondents may surmise that heavy survey participation is not desirable from a researchers' perspective, it is plausible that they under-report their survey taking. A long-standing literature claims that survey respondents generally misreport media-related behaviors (Parry et al. 2021) because of recall error or social desirability bias (Prior 2009). Estimating the number of surveys taken in a certain time frame, the time spent on this activity, or the number of panels to which one belongs may be especially challenging to remember. We measure the extent of survey professionalism across different samples with more reliable online behavioral data and ask: *What is the degree of survey professionalism among online panel members? (RQ1)*

2.2 Sociodemographic peculiarities of survey professionals (RQ2)

A common concern about online survey research is that samples may systematically differ from the general population (Krupnikov, Nam, and Style 2021). Most studies focus on demographic peculiarities of *convenience samples* compared to the population. Amazon Mechanical Turk (MTurk) samples have been found to be younger, more educated and less conservative than the U.S. population (Huff and Tingley 2015) and more likely to be unemployed (Berinsky, Huber, and Lenz 2012). Research on other crowdsourcing platforms comes to similar conclusions (Peer et al. 2017). Relative to the general U.S. population, online panelists tend to be more politically engaged, interested, and knowledgeable (Malhotra and Krosnick 2007; Chang and Krosnick 2009), and differ on personality traits (Valentino et al. 2020).

There is less evidence on the peculiarities of *survey professionals*. Not all members of convenience samples are such professional respondents; but professional respondents are more likely be part of convenience samples. Reviewing a number of studies, Whitsett (2013) concluded that frequent survey respondents in the U.S. were more likely to have lower income, lower education and no full-time job—though no differences emerged in terms of gender and age. A more recent study suggests that professional respondents are more likely to be older, less-well educated and white (Zhang et al. 2020). In the Netherlands, professional respondents are more likely to be female, unemployed, politically interested—but do not differ in terms of ideology or participation (Matthijsse, De Leeuw, and Hox 2015). In sum, the picture is mixed, and we offer new evidence by combining survey questions on basic demographics and political variables with our behavioral data. We ask: *Do survey professionals differ from non-professionals sociodemographically and politically? (RQ2)*

2.3 Low response quality among survey professionals (RQ3)

Concerns that respondents recruited online might deliver lower quality responses—manifesting in failed attention checks, straightlining in grid questions, and speedy responding—have been raised early on (Cape 2008). Again, part of the literature evaluates the quality of responses obtained by convenience samples in general. Comparing a random telephone sample with two Internet samples, Chang and Krosnick (2009) conclude that the latter are less likely to engage in satisfacing[1], and more affected by social desirability bias. In turn, Zhang and Gearhart (2020) compare online convenience samples from Dynata and MTurk and conclude

1. The concept of satisfacing is defined in the survey literature as a lack of careful and thorough thinking to produce meaningful responses. One way to measure it is by checking for straightlining in grid questions—that is, respondents chosing the same scale position on a number of questions even if some of the questions are reverse-coded (Chang and Krosnick 2009). We follow this approach.

that the latter may actually be more attentive and likely to complete surveys. Most recently, Ternovski and Orr (2022) reported low attentiveness among respondents from Lucid.

There is scarce evidence that zooms in on professional respondents, rather than online convenience samples at large. Conrad et al. (2010) examine whether professional respondents are more likely to use shortcuts to provide well-formed but invalid responses. Analyzing a U.S. sample, Zhang et al. (2020) find that professionals were more likely to speed through the questionnaire, but not more likely to engage in satisficing. Matthijsse, De Leeuw, and Hox (2015) report similar results for a Dutch sample. In line with previous research, we first compare professionals to non-professionals on typical markers of response quality: *Are survey professionals more likely to engage in straightlining than non-professionals? (RQ3a) Are they more likely to engage in speeding (RQ3b)?*

Another common concern about quality in survey research regards *insufficient effort responding* (IER), which means that survey participants, because of low intrinsic motivation, do not pay attention to the response items and/or 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 decrease the over-time response stability. Previous studies have explored the test-retest reliability of convenience samples such as MTurk (Shapiro, Chandler, and Mueller 2013; Holden, Dennie, and Hicks 2013), but the issue has not been considered in research on survey professionals. We use the multi-wave structure of each one of our three datasets to test whether professionals' answers to the same question across different waves shows greater variation and ask: *Do survey professionals exhibit higher between-waves 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[2] multiple times to maximize their

2. 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

revenue, especially in the case of well-paid academic surveys compared to lower-costs crowd-sourcing tasks. Only a handful of studies have contemplated measuring this type of behavior, partly due to the methodological challenges of ascertaining multiple survey taking of a single individual. Offering some insight into the problem, Berinsky, Huber, and Lenz (2012) examined whether separate MTurk accounts might belong to the same person by looking for re-occurring IP addresses, finding only few such cases. However, such repeated survey completion may have become more common after 2012, as the opportunities for study participation have proliferated. In a later MTurk study, Ahler, Roush, and Sood (2019) found that about five percent of IP addresses were duplicated. It has also been suggested that survey participants could use virtual private networks to pretend to be in a certain country and qualify for studies rewarded in a stronger currency (Dennis, Goodson, and Pearson 2020).

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 many different 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). What is more, relying on IP addresses to detect repeated participation comes with its own problems, e.g., the possibility to mask IP address. In contrast to the studies cited above, which only check whether participation in *one* questionnaire may be repeated, we explore whether our participants may have engaged in repeated participation of *any* questionnaire we can identify. We do so by analysing the URL structure of a number of questionnaire platforms. We ask: *What is the extent to which participants take the same questionnaire more than once, and do survey professionals engage in more repeated participation than non-professionals? (RQ4)*

3 Data and Methods

We rely on data collected from three distinct U.S. samples between 2018 and 2019. Our first sample was recruited in 2018 for a two-wave survey through ads on Facebook. The ads

taking", which refers to the narrower activity of answering questions.

appeared in the right-hand column of users' pages and led users to a survey website and a website through which participants could submit their browsing data. All survey participants were paid, and those who uploaded browsing data could win one of five $100 Amazon gift cards (see details on the recruitment in SM B.1) Our second sample was acquired in 2019 for a three-wave survey from Lucid, a provider that aggregates respondents from many different sources. Quotas were set on age, gender and education, and respondent were compensated between $10 and $25 (depending on the wave) for providing their browsing data. The third sample was recruited in 2018 from the YouGov Pulse panel (in which panelists install web extensions that record their desktop, mobile and tablet browsing activity) for a two-wave survey, with quotas set on gender, age, race, and education, using a survey frame stratified sampling from the 2016 American Community Survey (ACS). Participants compensation was defined by the YouGov payment standards. All three studies received ethical approval by the authors' respective institutions.

All three data sets combine survey responses with individual-level records of browsing behavior for two months for the YouGov sample[3] and during 90 days before each wave for the Lucid and Facebook samples. In the Facebook and Lucid samples, participants were directed to the open-source tool Web Historian before they started the survey. On the Web Historian site, participants were asked to submit, after informed consent, the 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 to being a panelist. YouGov browsing data are collected on-the-go, i.e., each web visit instantly produces a record, and includes both desktop and mobile data.

Across samples, some 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

3. The YouGov surveys were carried out starting on October 4, 2018 for Wave 1 and December 3, 2018 for Wave 2. The web browsing collection covers the period around the wave 1 survey, starting in September 4 and going up to November 8, 2024

be treated as doing surveys 100 percent of the time—we exclude subjects who submitted data from less than seven days. For $RQ1$, 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).[4] 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).

3.1 Measuring survey professionalism

In all three browsing data sets, each row stands for a web visit by a subject and is represented by 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* (i.e., "https://" or "http://") 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 further details on the differences of hosts and domains and other complexities of URL parsing, see Clemm von Hohenberg et al. (2024).

Identifying survey visits. We identify survey visits at the level of URL hosts, which has two benefits. First, as exemplified above, a URL host sometimes provide more information than its domain, which allows us to minimize false negatives. For example, the host "survey.irbureau.com" signals a survey site while its domain "irbureau.com" does not. Second, it allows us to minimize false positives, as not all hosts with the same domain constitute survey sites. For example, consider the difference between "survey.zoho.com" and "mail.zoho.com". As defined earlier, survey visits comprise both filling out questionnaires and the steps before and after that such as searching for and selecting studies and obtaining the rewards. To

4. In the Facebook and Lucid samples, anyone who did not participate in later waves also did not submit any browsing data. 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.

capture the variety of web sites 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”. We ended up with a list of 229 platforms, with many associated URL hosts in our data.

- Second, assuming that the content of URLs provides clues about their content, we classified all hosts that contained the word “survey” as survey taking. This method identified another 2,714 URL hosts across the three data sets.

- Third, not all survey sites provide a clue in their name, and so an approach based purely on the identification of the word “survey” in the host 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. In practice, survey-taking activities cannot always be separated clearly from other “get-paid-to” activities or click work, such as annotating images. Thus, our coders indicated whether a platform was mostly about survey taking or also about other rewarded activities. Detailed coding instructions can be found in SM B.2.1. 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%). By this method, we identified 291 additional URL hosts as survey sites across all three data sets (of which 231 were coded as primarily survey sites).

The final list of hosts based on these three approaches can be found on [redacted]. To rule out one of the above methods is driving the results, we present results by methods in Supplementary Materials (SM) Figure C.2, and highlight the most dominant survey sites in SM Figure C.3.

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 visits; 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 an approximation using timestamps: We 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, as recommended by Clemm von Hohenberg et al. (2024). We use these continuous individual-level count and duration variables to measure the degree of survey professionalism (*RQ*1), both on the aggregate and the individual level.

To address *RQ*2, *RQ*3 and *RQ*4, we dichotomize these continuous metrics to create three alternative classifications of survey professionals. For our primary analyses we define a survey professional as someone who has on average more than 100 survey visits per active day (i.e., a day on which he or she was online). In the SM, we use three alternative definitions. First, a respondent has more than 50 percent of all visits to survey sites. Second, a respondent spends more than 50 percent of all browsing time on survey sites. We also consider respondents who meet *any* of the three conditions described above.

Measuring repeated participation. To measure how often individuals take one and the same questionnaire multiple times (*RQ*4), we take advantage of the fact that questionnaire platforms assign unique URLs to the same questionnaire. For example, "https://www.survey-monkey.com/r/MT2B22T" will permanently point to one and the same questionnaire. The methodological challenge here is to not count visits to generic URLs from the same platform, e.g., "https://www.surveymonkey.com/login", as a repeated questionnaire participation. To identify patterns in URLs that reliably designate unique questionnaires, we inspected the thirty most frequented questionnaire platforms from the list by Bevec and Vehovar (2021). We could verify such URL patterns by signing up for test accounts of eleven platforms (e.g., Qualtrics, SurveyMonkey, QuestionPro—see SM B.2.2 for a full list).

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 30 minutes later (we vary this parameter to rule out alternative explanations). In sum, for each participant, we count all the questionnaires taken, and how many of those were repeated participations.

3.2 Survey-based measures

Sociodemographics. To compare professionals and non-professionals in term of sociodemographics ($RQ2$), we rely on measures of age (in years), gender (recoded into dummy for female), education (recoded into a dummy for highly educated), and ethnicity (recoded into a dummy for white). We also compare our samples to census statistics on these four variables, using the 2020 U.S. census estimates.

Political outcomes. To explore whether professionals differ from non-professionals on various political characteristics ($RQ2$), we use the following variables: partisanship ("Please select the option that best describes your political party affiliation", 7-point scale); ideology ("Where would you place yourself on this scale, where 0 means liberal and 10 means conservative?" or similar); an out-party feeling thermometer ("Please rate how you feel about the following groups" or similar, scale from 0 to 100, using the respective out-party group and dropping Independents); political interest ("How interested would you say you are in politics?", or similar); political knowledge (measured with different knowledge batteries); following politics in the news ("How closely do you follow politics on TV, radio, newspapers, or the Internet?" or similar). All measurements are reported in greater detail in SM 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. To benchmark our sample statistics against population values, we use the ANES 2020 survey.

Speeding and straightlining. Across all three samples, we use individual duration (in seconds) of taking the first-wave survey to assess speeding (*RQ3b*). 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. In the Facebook sample, we use a 5-item battery about gun control attitudes. In the Lucid sample, we use a 5-item battery measuring attribution of malevolence to the out-party. In the YouGov sample, we use a six-item battery measuring attitudes on abortion. We flag subjects if they chose the same scale value on all items (except if that value was the scale midpoint). Such response patterns, such as consistently strongly agreeing or strongly disagreeing with all the statements regardless of their valence, indicate a lack of attention on behalf of participants.

Over-time response stability. Our final measure of response quality compares professional and non-professionals in terms of their over-time response stability. One way survey professionals might speed through questionnaires is by answering questions randomly. Straightlining behavior would not capture this type of low-quality response and so we additionally compare the variability of responses across waves to identify if survey professionals change their responses to the same item at a higher rate than non-professionals (*RQ3c*). We use all numeric variables that were measured across waves, and with precisely the same question wording and format. We identified 59 such variables in the Facebook data, 151 in the Lucid data and 31 in the YouGov data. These are primarily standard attitudinal and behavioral questions in which researchers are interested in measuring variation over time, such as attitudes towards policy issues, media and social media consumption habits, political preferences and polarization measures, among others.

4 Results

4.1 The extent of survey professionalism (RQ1)

How much do participants in our three samples engage in survey taking? Figure 1 shows the percent of visits out of all visits to web addresses classified as survey sites (in dark blue). These are aggregate-level percentages, i.e., they show how much out of the sample's web visits are visits to survey sites. For comparison, the figure also plots the percent of visits to facebook.com, google.com, amazon.com and youtube.com. The figure illustrates that survey-taking makes up a substantial part of participants' online activity, especially in the professional online panels: 54.3% of all visits in the Lucid sample and 24.5% in the YouGov sample are visits to survey-taking sites. In both, this is much more than visits to, for example, google.com (6.5% and 11.2%, respectively). The prevalence of survey professionalism is the lowest in the sample recruited using Facebook ads, i.e., 9.8% of all visits are to survey-taking sites, which is less than visits to google.com (25.3%).

In SM C.1.1, we report which of our three approaches to identify survey taking accounts for most visits and which sites account for most survey visits. Overall, the distribution of the three methods is similar across datasets, with our approach of manually coding the top 500 hosts representing the largest share of survey visits. SM Figure C.3 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 from a provider, such as YouGov, take surveys across different (and competing) platforms.

Additionally, Figure 2 plots the distribution of survey visits per respondent. While extreme values exist across the three samples, the shapes of the distributions differ drastically across the samples. 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 the individual-level variation, depicted in SM Figure C.4, are slightly smaller than aggregating the total number of visits.

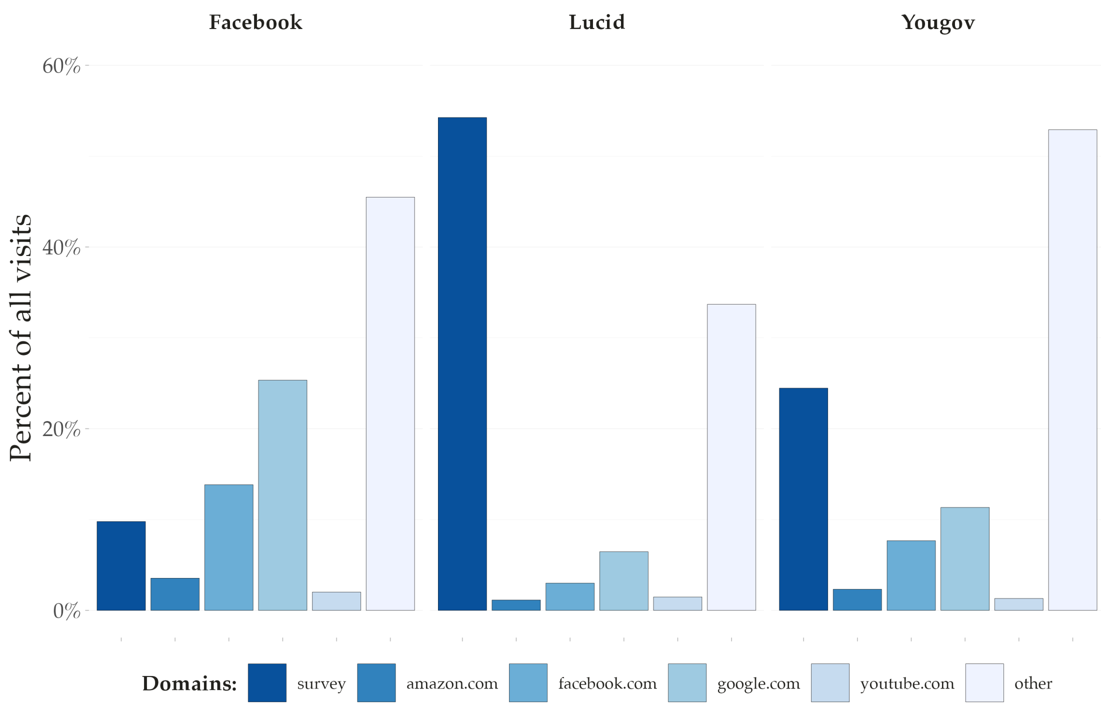

Figure 1: Percent of visits to survey sites (out of all visits) compared to visits to popular web domains.

In SM C.1.2, we also 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 (Lucid: 42.8%; YouGov: 18.9%; Facebook: 8.4%) or as an average individual-level percentage (Lucid: 40.7%; YouGov: 15.7%; Facebook: 6.70%).

To address the next set of research questions ($RQ2$ and $RQ3a, b, c$), we categorize subjects of each sample as either "professionals" or "non-professionals". As described in the methods, we use three different visits-based and time-based categorizations, namely (a) respondents with more than 100 survey visits per active day; (b) respondents with more than 50% of visits to survey sites; (c) respondents with more than 50% of browsing times. Figure 3 shows the percentage of survey professionals in the samples according to the three categorization approaches. We also provide estimates combining all three approaches, considering a survey professional as (d) respondents who fit in any of the three categorizations.

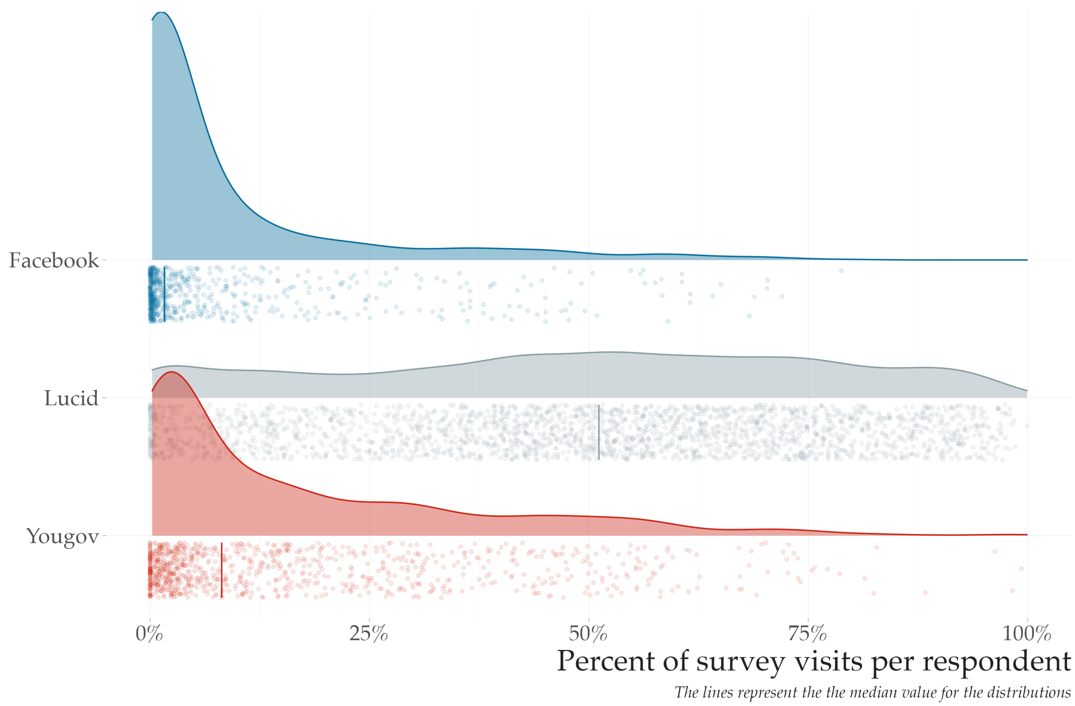

Figure 2: Individual level distribution of survey visits per respondent across the three samples

In the Lucid sample, our estimates of survey professionals vary between roughly 34.3% and 66.3%, and go up to 71.2% when including subjects categorized as professionals by any in of the measures; in the YouGov sample, between 7.9% and 11.4%, with 16.7% appearing in at least one category; and in the Facebook sample, between 1.72% and 10.74%, with 11.53% in at least one category. For the remainder of the paper, we use the first definition of survey professionals (> 100 visits per active day), but report results for the others in the SM.

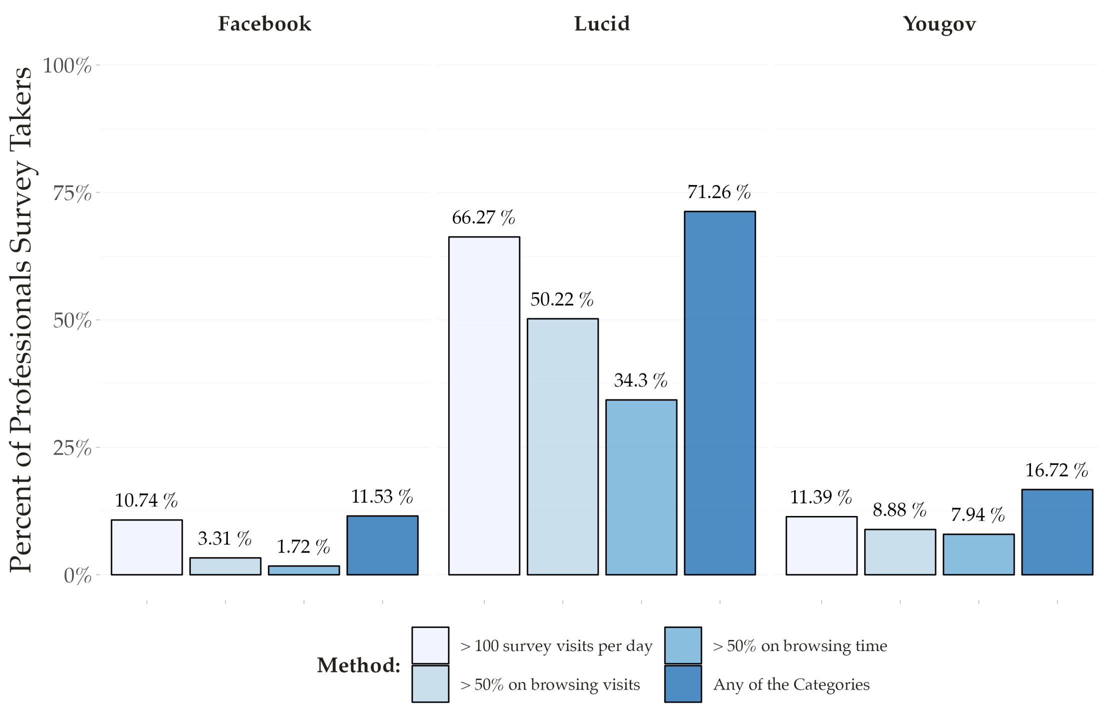

Figure 3: Percent of survey professionals for different definitions of survey professionalism.

In sum, survey taking is prevalent but also varies substantively across samples: Lucid is the panel with the highest proportion of survey professionals and 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.

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 general U.S. population, on a number of key variables. The first set of rows, which compare professionals and non-professional 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 signficance. We see similar patterns in the Facebook sample, but no statistically significant differences between the groups in the YouGov sample. However, both on Facebook and YouGov, sample sizes are small, particularly the number of professional respondents, which affects statistical power. SM Tables C.4, C.5 and C.6 report the same statistics for our alternative approaches to categorizing professionals, 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 differences in terms of self-reported partisanship, political ideology, out-party feeling thermometer, political interest, political knowledge, and following politics in the media. Across the samples, survey professionals tend to be more conservative than non-professionals—although this difference only remains statistically significant across categorization approaches in the Lucid sample (see SM Tables C.4, C.5 and C.6). Professionals also tend to feel more positive towards out-partisans: In the YouGov sample, this difference is statistically significant no matter what the categorization of survey professional we use. 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 in the media than the non-professionals.

Taken together, we do not find consistent differences when comparing professionals and non-professionals. A more consistent picture emerges when we compare professionals with the general population: Whatever the definition and the approach to categorizing a survey professional, compared to the population, professionals are older, more female, better educated and more ethnically white across samples.

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

		Facebook			Lucid			Yougov		
	Population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	40-44	⋆⋆	30-34	42 (0.37)	⋆⋆⋆	35 (0.59)	53 (1.48)		50 (0.54)
Gender (% female)	50.8	75.7 (5.2)		74.8 (1.7)	54.2 (1.3)		53.9 (1.9)	53.8 (5.3)		54.2 (1.8)
Education (% Bachelor or more)	30.4	55.7 (6)		54.8 (2)	50.8 (1.3)	⋆⋆	44.6 (1.8)	37.4 (5.1)		35.8 (1.8)
Ethnicity (% white)	62.6	88.6 (3.8)		84.3 (1.5)	80 (1.1)	∘	76.6 (1.6)	70.3 (4.8)		71.9 (1.7)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.67 (0.27)	∘	3.14 (0.07)	3.62 (0.06)		3.48 (0.07)	3.43 (0.23)		3.49 (0.08)
Ideology (0-1)	0.54 (0.006)	0.49 (0.03)	⋆⋆	0.4 (0.01)	0.5 (0.01)	⋆⋆	0.46 (0.01)	0.55 (0.03)	⋆	0.49 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	29.98 (3.44)		25.32 (0.93)	28.16 (0.72)		28.99 (1)	21.24 (3.2)	⋆	13.97 (0.91)
Political interest (0-1)	0.4 (0.006)	0.64 (0.04)		0.66 (0.01)	0.69 (0.01)	⋆⋆	0.62 (0.02)	0.68 (0.04)		0.79 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.64 (0.01)		0.62 (0.02)	0.55 (0.04)	⋆	0.64 (0.01)
Following politics (0-1)		0.63 (0.04)		0.6 (0.01)				0.56 (0.03)	⋆⋆	0.67 (0.01)

Note: Standard errors 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 for 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

	Facebook			Lucid			Yougov		
	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Straightliner (%)	2.9 (2.1)		1 (0.4)	2 (0.4)		1.3 (0.5)	8.8 (3)	∘	3.8 (0.7)
Survey duration (median seconds)	678.5 (102.46)	∘	833 (1170.086)	1108 (254.527)	⋆⋆	1189 (518.445)	1466 (9650.139)	⋆	1778 (3304.937)

Note: Standard errors 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; personality data from ANES 2016; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

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

How much does response quality vary between professional and non-professional participants in our samples? To answer this question, we focus on three distinct dimensions: straightlining through questionnaires (*RQ3a*), speeding (*RQ3b*), and stability of over-time measurements (*RQ3c*). As in the previous section, we define and categorize subjects of each sample as either professionals or non-professionals using our primary approach, namely, respondents with more than 100 survey visits per active day. In SM C.3, we provide the same results using the other two definition and categorization approaches discussed before.

Table 2 reports the mean comparison across survey professionals and non-professionals for speeding and straightlining. The first row reports behavior on grid questions across survey professionals and non-professionals (*RQ3a*). Overall, across the three samples, professionals show a higher incidence of straightlining through grid questions. However, only for the YouGov sample is this difference is statistically significant (and only at the 90% confidence level).

The more pronounced differences appear in terms of speeding (*RQ3b*). For the Lucid sample, the median professional is 8% faster compared to non-professional, representing a median difference of one minute and twenty seconds. This difference is yet more pronounced for the YouGov and Facebook samples: professionals are 18% and 22% faster in each sample, representing a median difference of five minutes and twenty seconds in the YouGov samples and two minutes and 35 seconds on Facebook. The differences for Lucid and YouGov are statistically significant at conventional levels of statistical significance, while the Facebook difference is statistically significant only at a 90% confidence level.

To examine the stability of responses over time (*RQ3c*), we estimate a simple interactive linear OLS model, as in the equation below:

$$Y_{jw_2} = \alpha + \beta_1 Y_{jw_1} + \beta_2 SP_i + \beta_3 Y_{jw_1} \cdot SP_i + \epsilon \tag{1}$$

We regress the wave-two response for outcome j (Y_{jw_2}) on the wave-one response for the same outcome (Y_{jw_1}), plus the interaction of the wave-one response with a dummy indicating whether the respondent is classified as a survey professional (SP_i). Our quantity of interest is the interactive term—the β_3 parameter—which identifies the difference in changes of responses over time between professional and non-professionals.

Figure 4 presents the distribution of z-scores for β_3 for all survey outcomes measured twice with precisely the same question wording and format. Using this distribution, we also 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, we see that only a few coefficients for the interactive term achieve statistical significance at conventional levels. Only four models recover statistically significant parameters (out of 59) in the Facebook sample, eleven models (out of 149) in the Lucid Sample, and one model (out of 31) in the YouGov sample. The p-values are higher than 0.9 in all three cases, which does not allow us to reject the null hypothesis that the distribution of interactive terms measuring the stability of overtime responses between professional and non-professionals is equal to zero. SM C.3.2 additionally presents the point estimates with confidence intervals for all 239 models estimated across the three samples. We also present the same results using models controlling for standard demographics, such as race, education, gender, partisanship, and self-reported ideology. The results tell the same story as those reported here.

To summarize, even though our results report a high prevalence of survey professionals across the three samples, particularly in the Lucid case, we do not find substantive differences in the quality of the survey responses between professionals and non-professionals. Even though professionals do speed through the questionnaire, we do not find professionals show a higher incidence of straightening through grid questions or more low-effort responding to questions asked repeatedly across waves.

4.4 Repeated questionnaire participation (RQ4)

Do respondents complete the same questionnaires repeatedly? As aforementioned, we identify repeated survey participation if an individual's data show two or more visits to *the same*

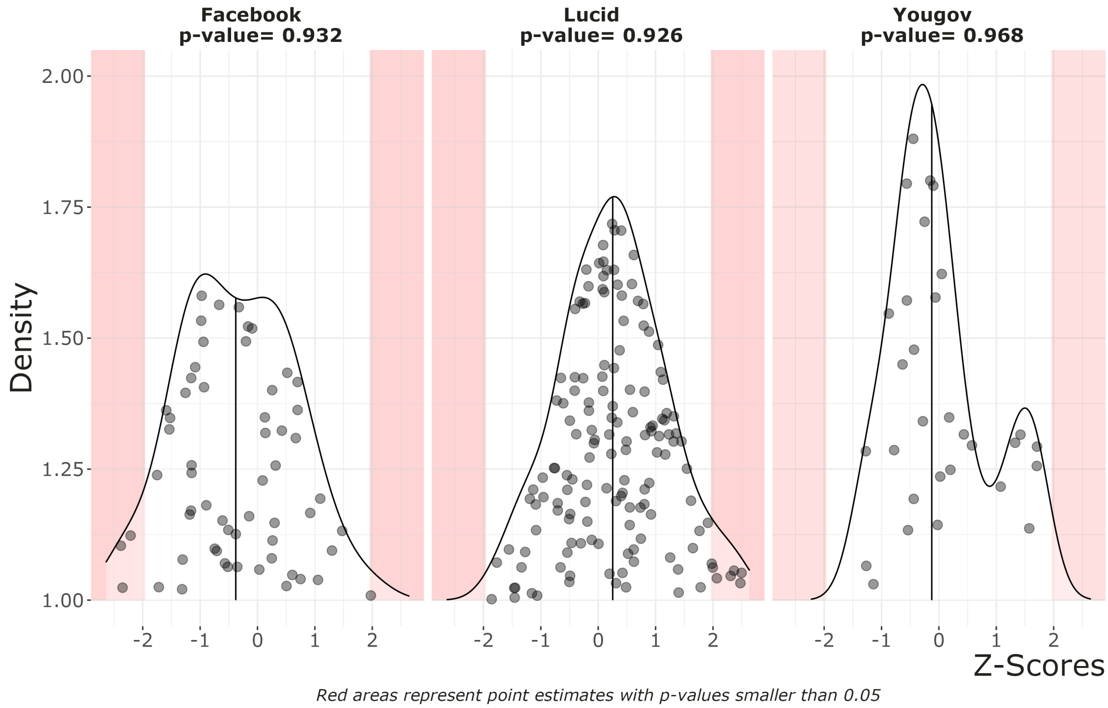

Figure 4: Z-scores for interactive terms between professionalism and wave-1 survey responses

questionnaire URL with a time difference of at least thirty minutes. As described in Section 3, we identified eleven platforms for which we can determine unique questionnaires. For these platforms, our data sets contain 71,579 unique questionnaires, which have been visited by respondents in our samples 325,614 times. Naturally, it is not surprising that questionnaires are taken more than once *across individuals*. However, our research question asked whether multiple completions also happen *within the same individual*, and if this phenomenon is especially pronounced among survey professionals.

Table 3 reports several statistics addressing this question. First, the percent of respondents taking *at least one* questionnaire repeatedly is 40.2% for Facebook, 81.3% for Lucid and 40.5% for YouGov sample. This statistic does not necessarily imply that participants in this set take *many* questionnaires repeatedly—perhaps they just did so for very few questionnaires. Thus, 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 took 5 questionnaires repeatedly, and the other took 15 repeatedly, then their average would be 10 (no matter how many questionnaires they did in total). The table shows an average of 1.96 and 1.37 for the Facebook and YouGov samples, but of almost 12 for the Lucid sample. In other words, on average, a Facebook participant recruited took 1.96 questionnaires more than once, while on YouGov, this number goes down to 1.37, while it goes up to twelve for the average Lucid participant.

Table 3: Repeated questionnaire participation

	Facebook	Lucid	YouGov
Subjects taking at least one questionnaire repeatedly (%)	40.20	81.31	40.45
Number of repeated questionnaires per participant (mean)	1.96	11.69	1.37
Percent of repeated questionnaires per participants (mean)	7.39	8.47	6.20

The third row summarizes, as an average, the share of questionnaires participants take repeatedly out of all questionnaires they take. To again exemplify, if one participant took 50 unique questionnaires in total, and of those took 10 multiple times, her percentage would be 20%; if another took 100 unique questionnaires and 10 of those more than once, his percentage would be 10%; their average would be 15%. The table shows that the averages of repeated participation are very similar across datasets, ranging from 6.20% (YouGov) to 7.39% (Facebook) and 8.47% (Lucid), implying that the overall share of repeated participation is non-negligible.

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* seem to repeat questionnaires on the eleven other platforms we identified, even if they are not sent there by YouGov. This caveat is not necessary for Lucid and Facebook, who do not operate their own platform.

The data suggest that across our samples, taking one and the same questionnaire repeatedly is not uncommon. Is it more prevalent among survey professionals than non-professionals? Table 4 shows the same statistics broken down by professionalism. We find stark contrasts between the two groups. Among professionals, the share of those who have taken at least one questionnaire repeatedly is much higher than among non-professionals across the three datasets. For example, in the Facebook sample, 79.3% of professionals took at least one questionnaire more than once, whereas this is the case for only 38.2% of non-professionals. For Lucid, these numbers are 91.7% versus 58.6%; for YouGov, 78.9% versus 27.1%. In all three samples, professionals also show higher absolute and relative numbers of repeated participation, as reported in the second and third rows.

Table 4: Repeated questionnaire participation, professionals vs. non-professionals

	Facebook		Lucid		YouGov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	84.29	34.24	91.71	58.61	78.85	27.09
Number of repeated questionnaires per participant (mean)	7.87	1.16	15.96	2.36	4.18	0.39
Percent of repeated questionnaires per participants (mean)	8.13	7.29	9.02	7.26	8.10	5.54

Of course, all these results hinge on our cutoff, i.e., counting two visits to the same questionnaire URL as two separate participations when at least 30 minutes have passed. To test the impact of this cutoff, we present results with alternative cutoffs (one hour, six hours, one day), both for the whole sample, and by professionalism, in SM C.4.1. Although the prevalence of repeated participation goes down a little, 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 30.8%, 76.2% and 33.5% for Facebook, Lucid and YouGov respectively. We also disaggregate results by the eleven platforms we identified in SM C.4.2.

5 Discussion

Much of what we know about public opinion and political behavior comes from surveys. In turn, most academic surveys today are collected from online panels maintained by commercial organizations. Many previous studies have focused on the representativeness of and the response quality provided by respondents (e.g., Krupnikov, Nam, and Style 2021; Chang and Krosnick 2009). There is less research on the extent to which online panels attract professional participants and whether this professionalism compromises data quality. We examined three samples, one recruited through the panel aggregator Lucid, another through the reputable polling company YouGov, and a third via Facebook ads. We analysed behavioral browsing data from these samples to offer, first, evidence on the prevalence of survey professionalism, and second, to explore impacts on data quality, i.e., demographic idiosyncrasies, speeding, straightlining, response stability, and repeated questionnaire participation.

We offer three key findings. First, professional survey taking represents a substantial portion of the online activity of the analyzed samples. In particular, Lucid (now Cint), criticized in past work (Ternovski and Orr 2022), shows the highest prevalence of survey professionalism. Using our most conservative measure, 34.3% of Lucid respondents were categorized as survey professionals. These percentages are much lower, though still substantial in the other two samples: 7.9% of YouGov respondents and 1.7% of Facebook respondents were categorized as survey professionals.[5] Visits to survey taking sites are the majority of all visits (54.3%) in the Lucid sample, around a quarter (24.5%) in the YouGov sample, and only one-tenth of all visits (9.8%) in the Facebook sample.

Second, although survey professionals constitute non-trivial parts of our samples, we do not find that they introduce significant problems for online research. Although there are some demographic differences between professionals and nonprofessionals, these largely depend on the approach used to categorize professionalism. Similarly, there are some differences on various political characteristics such as ideology, political interest, knowledge, and affective

5. Using our least conservative measure, that a participant met any of three criteria to be a professional, the percentages were 71.3, 16.7, and 11.5, respectively.

polarization, yet the lack of robust cross-sample difference suggests that survey professionalism does not introduce systematic demographic or political bias. 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 panelists also engage in significantly more straightlining. But to the extent that these behaviors are observable, data providers and researchers themselves can decide whether to remove speeding and straightlining participants from the analyses. Crucially, survey professionals do not offer different responses to the same questions over time, which we tested by taking advantage of the multi-wave nature of our studies. We take this as evidence that they answer questions at least as attentively as non-professionals. Future work should replicate our analysis—potentially based on our list of survey sites—using other indicators of data quality among other samples

Our third core finding does imply at least one problematic consequence of survey professionalism, namely that many participants take one and the same questionnaire repeatedly. In the presumed effort to maximize earnings, repeated questionnaire-taking is a widespread problem in the analyzed panels. Between 40.2% on the Facebook sample, 40.4% for YouGov, and 81.3% in the Lucid sample took at least *one* questionnaire multiple times. Participants in the Lucid sample exhibited a markedly higher frequency of repetition than the other two samples. A Lucid participant on average took twelve questionnaires repeatedly, whereas this average was below two for YouGov and Facebook respondents. Across all three samples, repeated participation is substantially higher among survey professionals.[6] Overall, our findings suggest that bad actors are present in the survey ecosystem, potentially creating multiple identities to take questionnaires several times. This finding has important implications for the design of online surveys, underscoring the importance of building systems to detect repeated participation and the need to account for its impact on data quality and reliability.

In short, our work offers previously unavailable evidence on the extent and implications of survey professionalism in U.S. panels and opens important avenues for future work. Although

6. We emphasize again that this does not mean panelists are taking the same questionnaires repeatedly from the panel we are analyzing—a YouGov or Lucid panelist taking the same questionnaire twice from a different platform would be counted here.

researchers have worried about online panels, survey professionals do not, by and large, seem to have profound implications for the quality of the data they provide for academic studies. We encourage researchers to continue examining these phenomena and their consequences for what the public knows from research. Decades ago, academic lab experimentalists were criticized for their reliance on college students (mostly white, more affluent, and better educated than the general population), which meant that support, or the lack thereof for existing theories was based on highly non-representative samples. Online panels were hoped to offer an antidote, namely access to a diverse group of individuals to complete surveys or partake in experiments. Although the set of people we now use to learn about various political phenomena also tend to be unusual as they contain survey professionals, they may be more diverse and less "problematic" than undergraduates. Given the centrality of online-based survey research to political science, this finding makes us cautiously optimistic.

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.

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.

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.

Conrad, Fredrick G., Roger Tourangeau, Mick P. Couper, and Chan Zhang. 2010. "Professional Web Respondents and Data Quality." Presented at AAPOR Conference Chicago. websm.org/db/12/14322/Bibliography/Professional_Web_Respondents_and_Data_Quality/.

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.

Dennis, Sean A., Brian M. Goodson, and Christopher A. Pearson. 2020. "Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures." *Behavioral Research in Accounting* 32 (1): 119–134. doi:10.2308/bria-18-044.

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

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

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.

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.

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.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5 (5): 411–419. doi:10.1017/S1930297500002205.

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.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. “Beyond the Turk: Alternative platforms for crowdsourcing behavioral research.” *Journal of Experimental Social Psychology* 70:153–163. doi:10.1016/j.jesp.2017.01.006.

Prior, Markus. 2009. “The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure.” *Public Opinion Quarterly* 73 (1): 130–143. doi:10.1093/poq/nfp002.

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.

Silber, Henning, Sven Stadtmüller, and Alexandru Cernat. 2023. “Comparing participation motives of professional and non-professional respondents.” *International Journal of Market Research* 65 (4): 361–372. doi:10.1177/14707853231166882.

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.

Survey Professionalism: New Evidence from Web Browsing Data
ONLINE SUPPLEMENTARY MATERIALS

Author names anonymized

August 17, 2024

Contents

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

B.1 Facebook sample

As participant recruitment via Facebook involves more degrees of freedom than via 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 a website (the landing survey) inviting them to the main survey and provided a link to Web Historian, where respondents where asked to upload their browsing histories. After informed consent, participants could complete the 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.

Of those clicked on the link, 2,760 responded at least some questions of the landing survey. Of those, 805 uploaded their browsing data and completed the main survey. Of those, 707 provided browsing data for at least 7 days (our final sample). 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 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 top 500 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_EXPO, hosts_500_FB and hosts_500_CSMAP 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.2.2 Identifying repeated questionnaire participation

Table B.1 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.1: 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.2 below describes the political characteristics used to address part of RQ3a, in terms of their wording, their response scale and any recoding.

Table B.2: 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.2: 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

C Additional results

C.1 Prevalence of survey professionalism (RQ1)

C.1.1 Survey visits disaggregated by identification approaches and survey sites

Figure C.2 shows the same measures as Figure 1 in the main paper, but splits survey visits according to identification method. Figure C.3 shows the ten most frequent URL hosts among the survey visits and reports how much of the total each accounts for.

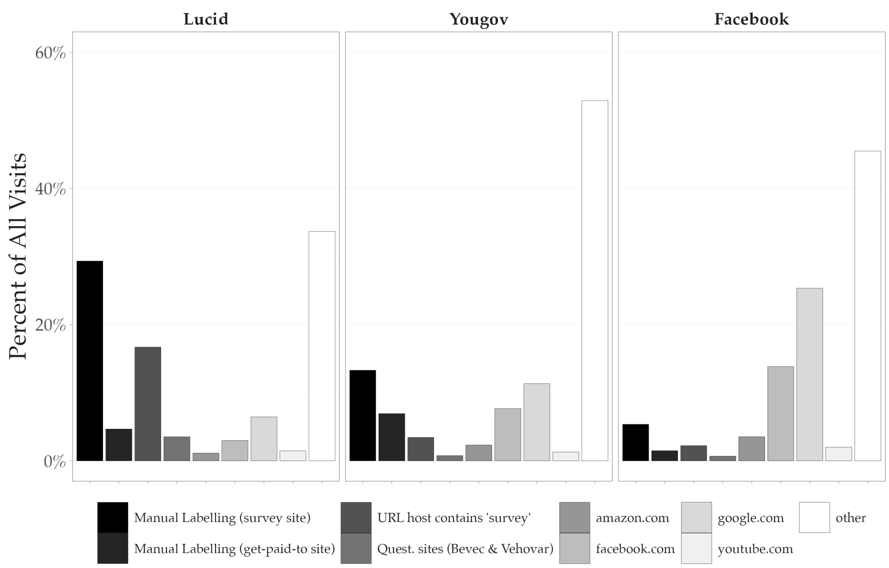

Figure C.2: Percent of visits to survey sites (out of all visits), split by method, and compared to popular web domains.

Figures C.4 takes a slightly different perspective on the data, reporting individual-level survey taking. Figure C.4 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.

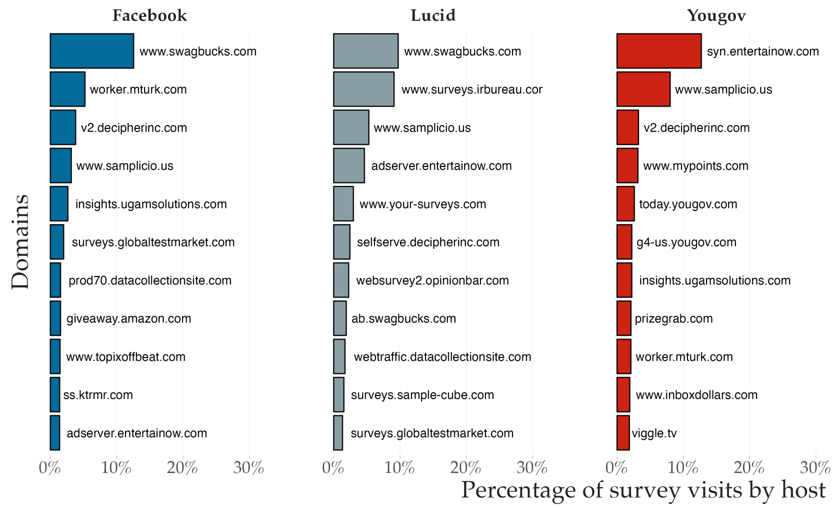

Figure C.3: Ten most frequented survey sites, by sample.

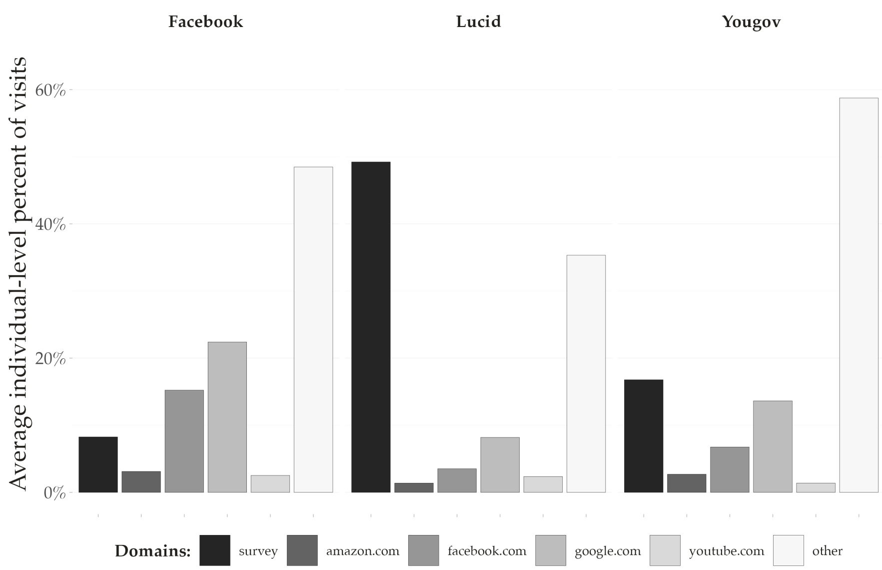

Figure C.4: Average individual-level percent of visits to survey sites out of all visits, compared to popular web domains.

C.1.2 Survey duration

Figure C.5 shows, analogously to Figure 1 in the main paper, the aggregate proportion of survey-taking time out of all browsing time. Figure C.6 shows, again, the average individual-level percentage of survey-taking time out of all browsing, and Figure C.7 the distribution of this individual-level percentage.

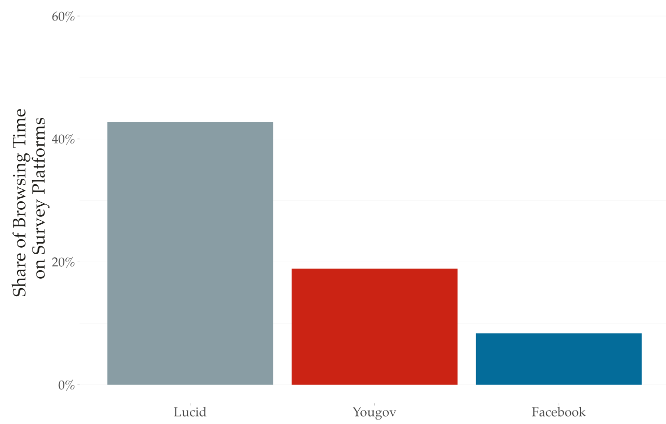

Figure C.5: Percent of visit time to survey sites (out of all browsing time).

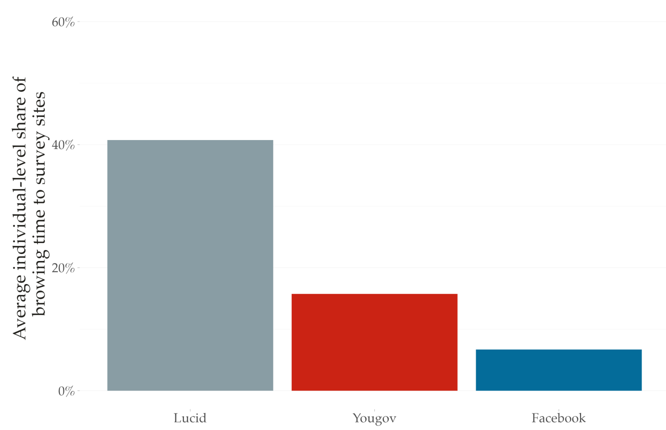

Figure C.6: Distribution of individual-level percent of visit time to survey sites out of all browsing time

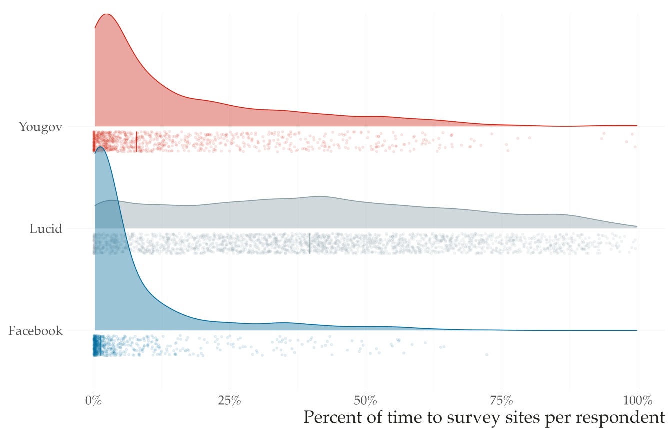

Figure C.7: Distribution of individual-level percent of visit time to survey sites out of total browsing time.

C.2 Sociodemographic and political differences (RQ2)

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 C.3), when defined as anyone with more than 50 percent of browsing time to survey sites (Table C.4), and when meeting any of the three criteria (Table C.5).

Table C.3: Survey professionals vs. non-professionals vs. population (professional = more than 50 percent of visits to survey sites)

		Facebook			Lucid			Yougov		
	Population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	40-44		35-39	42 (0.42)	⋆⋆⋆	37 (0.48)	54 (1.41)	⋆	50 (0.54)
Gender (% female)	50.8	61.9 (10.9)		75.3 (1.7)	53 (1.5)		55.3 (1.5)	59.7 (5.6)		53.6 (1.8)
Education (% Bachelor or more)	30.4	38.1 (10.9)		55.4 (1.9)	49.6 (1.5)		47.9 (1.5)	23.4 (4.9)	⋆	37.3 (1.8)
Ethnicity (% white)	62.6	95.2 (4.8)		84.4 (1.4)	79.1 (1.2)		78.6 (1.3)	67.5 (5.4)		72.1 (1.6)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.65 (0.56)		3.18 (0.07)	3.64 (0.07)		3.5 (0.06)	3.64 (0.27)		3.47 (0.08)
Ideology (0-1)	0.54 (0.006)	0.47 (0.05)		0.41 (0.01)	0.51 (0.01)	⋆⋆⋆	0.46 (0.01)	0.53 (0.03)		0.49 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	29.11 (5.3)		25.7 (0.92)	29.9 (0.86)	⋆⋆	26.84 (0.78)	23.25 (3.47)	⋆	13.89 (0.9)
Political interest (0-1)	0.4 (0.006)	0.64 (0.06)		0.66 (0.01)	0.69 (0.01)	⋆	0.65 (0.01)	0.67 (0.04)		0.79 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.63 (0.01)		0.64 (0.01)	0.54 (0.04)	⋆	0.64 (0.01)
Following politics in the media (0-1)		0.62 (0.06)		0.61 (0.01)				0.52 (0.04)	⋆⋆⋆	0.67 (0.01)

Note: Standard errors 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 for trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

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

		Facebook			Lucid			Yougov		
	Population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	4.5		35-39	41 (0.5)	⋆⋆⋆	39 (0.41)	52 (1.38)	∘	50 (0.54)
Gender (% female)	50.8	50 (15.1)	∘	75.2 (1.7)	52.4 (1.8)		55 (1.3)	63.8 (5.8)		53.3 (1.8)
Education (% Bachelor or more)	30.4	50 (15.1)		55.3 (1.9)	51.3 (1.8)	∘	47.4 (1.3)	24.6 (5.2)	∘	37 (1.7)
Ethnicity (% white)	62.6	91.7 (8.3)		84.3 (1.4)	78 (1.5)		79.2 (1.1)	68.1 (5.7)		72 (1.6)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.09 (0.61)		3.2 (0.07)	3.62 (0.08)		3.55 (0.06)	3.94 (0.29)		3.45 (0.08)
Ideology (0-1)	0.54 (0.006)	0.42 (0.06)		0.41 (0.01)	0.5 (0.01)	⋆	0.47 (0.01)	0.51 (0.04)		0.49 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	35.22 (7.11)		25.78 (0.92)	31.03 (1.07)	⋆⋆	26.9 (0.69)	23.23 (3.48)	⋆	13.99 (0.91)
Political interest (0-1)	0.4 (0.006)	0.74 (0.09)		0.65 (0.01)	0.69 (0.01)	∘	0.66 (0.01)	0.63 (0.04)		0.79 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.62 (0.02)		0.64 (0.01)	0.51 (0.05)	⋆⋆	0.64 (0.01)
Following politics in the media (0-1)		0.67 (0.09)		0.61 (0.01)				0.49 (0.04)	⋆⋆⋆	0.67 (0.01)

Note: Standard errors 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 for trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

Table C.5: Survey professionals vs. non-professionals vs. population (professional = any of the categories)

		Facebook			Lucid			Yougov		
	Population	Professionals		Non-professionals	Professionals		Non-professionals	Professionals		Non-professionals
Sociodemographics										
Age (median years)	38.2	40-44	⋆	30-34	42 (0.36)	⋆⋆⋆	34 (0.65)	53 (1.14)		50 (0.56)
Gender (% female)	50.8	73.7 (5.1)		74.9 (1.8)	53.7 (1.3)		55 (2)	56.4 (4.2)		53.7 (1.9)
Education (% Bachelor or more)	30.4	55.3 (5.7)		55.2 (2)	50.3 (1.3)	⋆	44.9 (2)	30.7 (3.9)		37 (1.8)
Ethnicity (% white)	62.6	89.5 (3.5)		83.8 (1.5)	79.7 (1)		76.5 (1.7)	67.9 (4)		72.5 (1.7)
Political outcomes										
Partisanship (1-7)	4 (0.059)	3.61 (0.26)	∘	3.14 (0.07)	3.62 (0.06)		3.47 (0.08)	3.59 (0.19)		3.46 (0.08)
Ideology (0-1)	0.54 (0.006)	0.47 (0.03)	⋆	0.4 (0.01)	0.5 (0.01)	⋆⋆	0.45 (0.01)	0.53 (0.02)		0.49 (0.01)
Thermometer out-party (1-100)	17.4 (0.425)	29.4 (3.23)		25.48 (0.94)	28.25 (0.7)		28.9 (1.06)	22.03 (2.58)	⋆⋆	13.28 (0.91)
Political interest (0-1)	0.4 (0.006)	0.64 (0.03)		0.66 (0.01)	0.68 (0.01)	⋆	0.63 (0.02)	0.67 (0.03)		0.8 (0.01)
Political knowledge (0-1)	0.5 (0.006)				0.64 (0.01)		0.61 (0.02)	0.55 (0.03)	⋆	0.64 (0.01)
Following politics in the media (0-1)		0.63 (0.04)		0.61 (0.01)				0.54 (0.03)	⋆⋆⋆	0.68 (0.01)

Note: Standard errors 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 for trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

C.3 Response-quality differences (RQ3)

C.3.1 Speeding and straightlining (RQ3a,b)

Table C.6: 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 (%)	4.8 (4.8)	1.1 (0.4)	2.4 (0.5)	∘	1.1 (0.3)	3.9 (2.2)	4.4 (0.7)
Survey duration (median seconds)	833 (143.578)	823 (1084.962)	1116 (295.018)		1140 (387.864)	1773 (11248.516)	1741 (3250.425)

Note:

Standard errors 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; personality data from ANES 2016; 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 C.7: 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)	1.2 (0.4)	2.8 (0.6)	⋆	1.2 (0.3)	2.9 (2)	4.4 (0.7)
Survey duration (median seconds)	894 (140.168)	814.5 (1088.209)	1116 (425.54)		1139.5 (298.746)	1913 (12747.256)	1729 (3209.124)

Note:

Standard errors 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; personality data from ANES 2016; 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 C.8: Response quality of survey professionals vs. non-professionals (professionals = all categories)

	Facebook			Lucid			Yougov	
	Professionals		Non-professionals	Professionals		Non-professionals	Professionals	Non-professionals
Straightliner (%)	4.1 (2.3)	∘	0.9 (0.4)	2.1 (0.4)		1 (0.4)	6.4 (2.1)	3.9 (0.7)
Survey duration (median seconds)	678.5 (95.844)	⋆	829.5 (1202.851)	1106 (284.218)	⋆⋆	1217 (472.576)	1750 (8859.647)	1744 (3306.64)

Note:

Standard errors 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; personality data from ANES 2016; political variables from ANES 2020. Variables trust, political interest, knowledge and partisanship were recoded to a scale from 0 to 1 to ensure comparability.

C.3.2 Stability of Survey Responses (RQ3c)

Below, we present additional results for the models analyzing the stability of survey versions over time, comparing professional and non-professional respondents. Figure C.8 presents the point estimates for the z-scores presented in Figure 4 in the main paper. Figure C.9 and C.10 present similar results but with standard control variables being added to the estimation, such as age, gender, race, education, partisanship, and self-reported ideology. Results are similar to those discussed in the main paper.

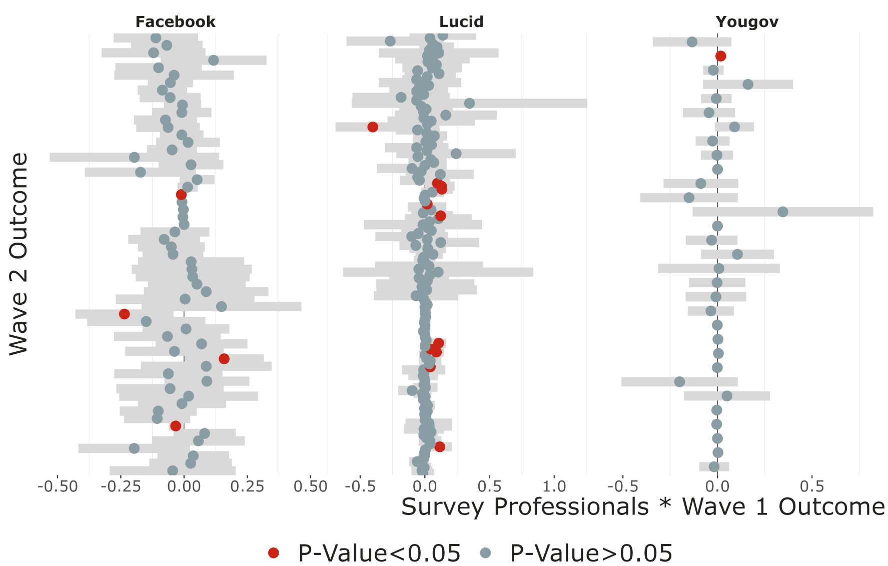

Figure C.8: Point estimates for interactive terms between professionalism and wave 1 survey responses

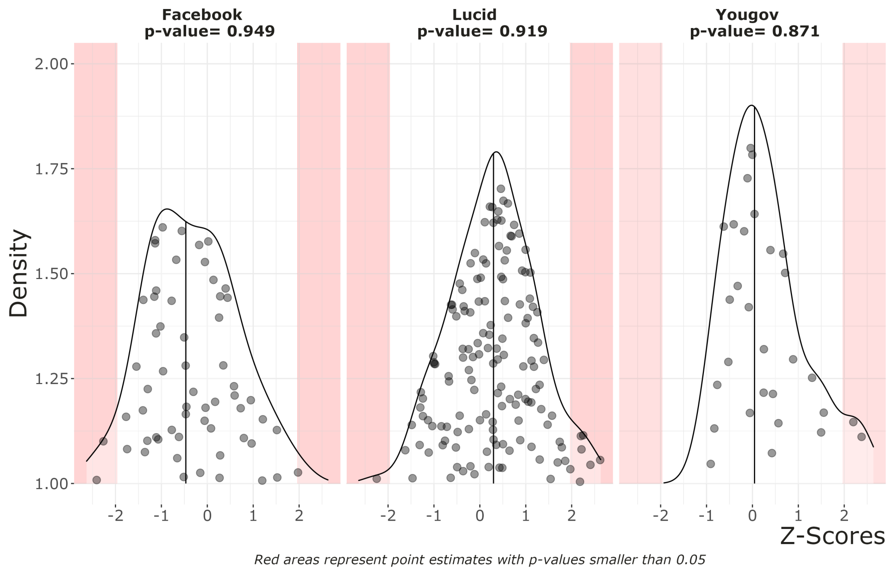

Figure C.9: Z-scores for interactive terms between professionalism and wave 1 survey responses, models estimated with standard demographic controls.

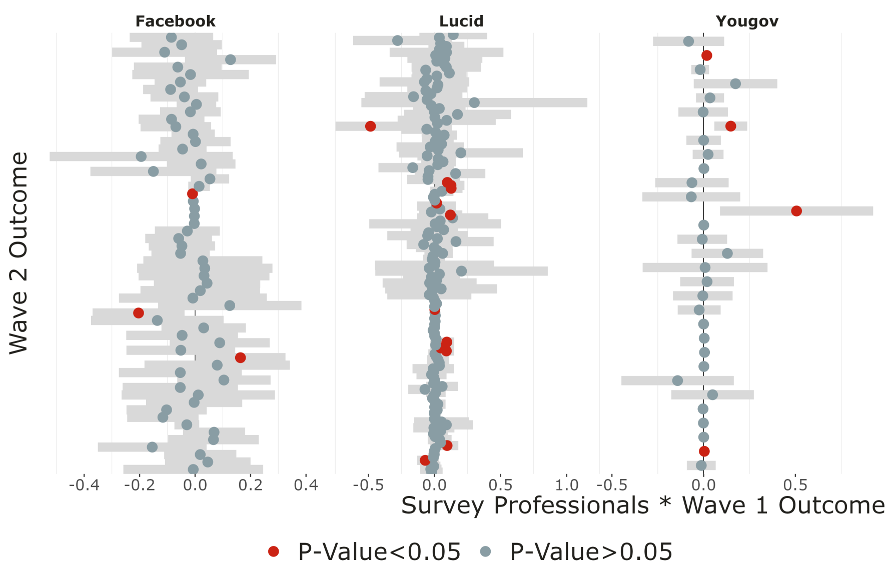

Figure C.10: Point estimates for interactive terms between professionalism and wave 1 survey responses, models estimated with standard demographic controls.

C.4 Repeated questionnaire participation (RQ4)

C.4.1 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 one hour (Tables C.9 and C.10), when greater than six hours (Tables C.11 and C.12) and when greater than 24 hours (Tables C.13 and C.14).

Table C.9: Repeated questionnaire participation (1-hour cutoff)

	Facebook	Lucid	Yougov
Participant taking at least one questionnaire repeatedly (%)	38.67	80.23	39.70
Number of repeated questionnaires per participant (mean)	1.79	10.94	1.30
Percent of repeated questionnaires per participant (mean)	6.76	7.86	5.97

Table C.10: Repeated questionnaire participation (1-hour cutoff), professionals vs. non-professionals

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	84.29	32.50	90.82	57.12	77.88	26.42
Number of repeated questionnaires per participant (mean)	7.21	1.06	14.94	2.20	3.97	0.37
Percent of repeated questionnaires per participants (mean)	7.39	6.68	8.34	6.80	7.68	5.38

Table C.11: Repeated questionnaire participation (1-hour cutoff)

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)	34.41	78.60	35.24
Number of repeated questionnaires per participant (mean)	1.50	9.91	1.17
Percent of repeated questionnaires per participants (mean)	5.43	6.87	4.24

Table C.12: Repeated questionnaire participation (1-hour cutoff), professionals vs. non-professionals

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	78.57	28.43	89.53	54.75	76.92	20.74
Number of repeated questionnaires per participant (mean)	6.17	0.87	13.55	1.96	3.69	0.29
Percent of repeated questionnaires per participants (mean)	6.13	5.34	7.42	5.68	7.08	3.25

Table C.13: Repeated questionnaire participation (1-hour cutoff)

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)	30.83	76.22	33.50
Number of repeated questionnaires per participant (mean)	1.28	8.80	1.01
Percent of repeated questionnaires per participants (mean)	4.12	5.95	3.63

Table C.14: Repeated questionnaire participation (1-hour cutoff), professionals vs. non-professionals

	Facebook		Lucid		Yougov	
	Professionals	Non-professionals	Professionals	Non-professionals	Professionals	Non-professionals
Subjects taking at least one questionnaire repeatedly (%)	77.14	24.56	87.97	50.59	75.96	18.73
Number of repeated questionnaires per participant (mean)	5.44	0.72	12.04	1.72	3.19	0.26
Percent of repeated questionnaires per participants (mean)	5.49	3.93	6.51	4.73	5.64	2.93

C.4.2 Disaggregation by questionnaire platforms

Table C.15: Repeated questionnaire participation, by questionnaire platform

	Facebook	Lucid	Yougov
Subjects taking at least one questionnaire repeatedly (%)			
Confirmit	27.41	60.29	33.33
Dynata	23.88	51.44	50.00
Formsite	22.22	28.07	0.00
Formstack	27.59	16.00	15.38
Qualtrics	28.54	59.59	NA
Questionpro	17.91	28.10	0.00
Surveygizmo	29.48	64.17	0.00
Surveymonkey	23.24	17.31	8.18
Typeform	12.78	9.82	11.00
Unipark	7.14	12.03	0.00
Zoho	38.71	12.96	17.14
Number of repeated questionnaires per participant (mean)			
Confirmit	0.56	3.20	0.69
Dynata	0.66	1.71	1.44
Formsite	0.22	0.32	0.00
Formstack	0.48	0.23	0.15
Qualtrics	0.75	4.20	NA
Questionpro	0.27	0.79	0.00
Surveygizmo	0.64	2.83	0.00
Surveymonkey	0.82	0.36	0.14
Typeform	0.21	0.15	0.17
Unipark	0.07	0.24	0.00
Zoho	1.39	0.24	0.23
Percent of repeated questionnaires per participants (mean)			
Confirmit	5.85	8.02	7.47
Dynata	3.54	5.90	8.99
Formsite	15.74	22.66	0.00
Formstack	22.35	12.99	11.54
Qualtrics	7.67	10.84	NA
Questionpro	9.35	5.84	0.00
Surveygizmo	8.16	11.18	0.00
Surveymonkey	8.23	3.71	2.92
Typeform	4.57	3.82	3.20
Unipark	0.17	4.77	0.00
Zoho	15.69	3.57	7.13