Information, Communication & Society

ISSN: 1369-118X (Print) 1468-4462 (Online) Journal homepage: www.tandfonline.com/journals/rics20

The accuracy of survey measures on political activity on Facebook. Evidence from a data donation study

Zoltán Kmetty & Ádám Stefkovics

To cite this article: Zoltán Kmetty & Ádám Stefkovics (25 Aug 2026): The accuracy of survey measures on political activity on Facebook. Evidence from a data donation study, Information, Communication & Society, DOI: 10.1080/1369118X.2026.2713664

To link to this article: https://doi.org/10.1080/1369118X.2026.2713664

View supplementary material

Published online: 25 Aug 2026.

Submit your article to this journal

View related articles

View Crossmark data

INFORMATION, COMMUNICATION & SOCIETY
https://doi.org/10.1080/1369118X.2026.2713664

The accuracy of survey measures on political activity on Facebook. Evidence from a data donation study

Zoltán Kmetty [a,b] and Ádám Stefkovics[a,c]

[a]ELTE Centre for Social Sciences, CSS-Recens Research Group, Budapest, Hungary; [b]ELTE Faculty of Social Sciences, Sociology Institute, Budapest, Hungary; [c]Századvég Foundation, Budapest, Hungary

ABSTRACT

Self-reported survey measures of social media use are commonly used across disciplines, yet the validity of findings in this area depends heavily on the accuracy of these reports. In political communication research, measurement error in reported online political behaviour can distort estimates of media effects and lead to misleading conclusions about online political engagement. To address this concern, our study leveraged donated digital trace data from 755 Hungarian Facebook users and compared it with responses to standard survey questions provided by the same individuals. In addition, in a pre-registered experiment, we varied the ordering of the survey items. Our findings reveal moderate correlations between self-reported and actual digital behaviour. General items showed closer alignment with observed behaviour than self-reports of specific political activities. Participants tended to under-report their most frequent general activities, such as commenting and reacting, whereas less frequent activities showed little systematic directional bias. Importantly, the weaker alignment for political activities may stem not from respondents' inaccuracy but from difficulties in classifying digital actions as politically relevant. We discuss potential avenues for improving survey-based measures and emphasize the need for further research to address the methodological challenges associated with the use of digital trace data.

ARTICLE HISTORY
Received 10 October 2025
Accepted 28 July 2026

KEYWORDS
political behaviour; social media; measurement error; digital traces; data donation

Introduction

As digital media technologies continue to replace traditional media, there is a growing scientific interest in understanding online political behaviour. Digital, and within that, social media use is increasingly being treated as a significant predictor or outcome variable for various topics and research questions (Parry et al., 2021). The validity of these approaches, however, lies in the validity of media use measures. Many studies, so far, have employed self-reported retrospective questions to measure digital media use. Self-reports, however, have long been criticized due to their inability to reliably map media

CONTACT Zoltán Kmetty kmetty.zoltan@tk.hu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/1369118X.2026.2713664.

behaviour (Coromina & Saris, 2009; Greenberg et al., 2005). Concerns about measurement error in digital political behaviour studies have implications for the substantive scientific discourse on the impact of digital media on politics and other fields (Haenschen, 2020). This is particularly crucial in political communication research, where inaccurate measures of media exposure can distort our understanding of how citizens engage with political content and form attitudes.

In contrast to offline media use, individuals' actions on social media generate digital traces that could theoretically offer a more precise measurement of digital behaviour. These new opportunities accelerated scientific attempts that contrast survey data with digital data (see Parry et al., 2021, for a meta-analysis). These studies show a moderate correlation between self-reported and observed social media use. They further highlighted the numerous issues and biases associated with digital data and underscored the limitations regarding incomparability.

Despite these shortcomings, digital traces allow not only the assessment of the overall accuracy of survey measures but also the determination of which types of survey questions yield more precise data (Cardenal et al., 2022). When survey design features, such as question order, systematically influence how accurately respondents report their behaviour, they become a key methodological factor shaping the degree of agreement between survey responses and digital trace data. For instance, the literature on question order effects has produced 70 years of evidence that the order of questions often affects responses (Schuman & Presser, 1981), but the guidance on which design to choose has often been less clear-cut. Studies incorporating digital trace validation, however, remain limited in number and scope, with the vast majority conducted in the United States. These attempts reported low to moderate correlations between self-reports and digital behaviour related to social media use (Cardenal et al., 2022; Ernala et al., 2020; Guess et al., 2019; Haenschen, 2020; Mahalingham et al., 2023; Naab et al., 2019).

This study extends the current body of literature by contrasting survey reports of Facebook use, utilizing digital data donated by 755 participants in Hungary. Survey measures covered general, political, and health-related Facebook activity. By examining correlations across multiple types of activities, we assess whether survey validity systematically differs between general and topic-specific behaviours. Furthermore, in a pre-registered experiment, we varied the ordering of the survey items. One group received questions in an order such that the frequency of each activity was first asked about their activity in general, followed by items on specific topics (general-specific), whilst the other group received the same questions in a specific-general order, with the specific items of each activity first, followed by the general item. Our study is the first to compare survey and digital data across two question orders, enabling us to make recommendations about a longstanding issue of survey methodology (Schwarz et al., 1991). We focus on question order because it can change which considerations come to mind when respondents answer: placing some questions before others can make political content more salient, so the same person may report their behaviour differently depending on the order (Tourangeau & Rasinski, 1988; Zaller, 1992). Because question order in this way affects how accurately people report their behaviour, it also affects how closely their survey answers align with the digital trace data. Finally, we examine political interest as a moderator, as it conditions both the accuracy of self-reports and individuals' susceptibility to survey context effects (Haenschen, 2020). Our study is the first to assess the accuracy of survey measures on political social media use in the Eastern European context. This can be relevant because patterns of social media use and political communication differ markedly in Eastern Europe due to distinct user preferences, lower institutional trust, and unique post-socialist legacies.

Self-reports of social media use vs. digital trace data

This section reviews the literature in three steps: first, we outline general sources of bias in self-reported behavioural measures; second, we summarize empirical comparisons between self-reports and digital trace data; and third, we discuss the implications of these discrepancies for measurement validity. There is considerable prior evidence indicating that the measurement of behaviour via self-reported questions is potentially vulnerable due to cognitive, social, and communicative biases affecting individuals (e.g., Schwarz & Oyserman, 2001; Tourangeau et al., 2000). First, responding to survey questions requires substantial cognitive effort in which motivation and interest are key (Araujo et al., 2017). Drawing on the satisficing theory (Krosnick, 1991), to reduce the cognitive effort that is required to retrieve the necessary information, some people may choose to satisfice instead of optimize and skip some of the steps of the response process (Tourangeau et al., 2000). As a result, people may fail to recall relevant information (Revilla et al., 2017) or inattentively misreport it. With social media use measures, such biases are likely to occur if the questions are difficult or the questions are phrased too vaguely. Furthermore, drawing on the flawed estimation hypothesis (Prior, 2009), respondents may use inappropriate estimation strategies to infer their frequency of use. For instance, basing their answer on particularly memorable or recent sessions rather than typical behaviour. Another possible source of bias is socially desirable responding (Krumpal, 2013), when individuals tailor their true responses in order to present themselves more favourably. Although social media use is not necessarily a sensitive topic, social desirability bias may still occur (Henderson et al., 2021). If one believes posting on Facebook is generally a good thing, an overreport of such behaviour can be expected. Similarly, reacting to political content on social media can be perceived as socially desirable or undesirable depending on the individual's relation to politics and social media, potentially leading to deviations from true responses.

Against this theoretical background, a growing body of research has directly compared self-reported social media use with digital trace data. These studies differ in the types of behaviours they validate, ranging from general measures of platform use to specific political or content-related activities.

Guess et al. (2019) were the first to compare Facebook and Twitter data with self-reports. Focusing on both general use and political activity, they found that self-reports were correlated with observed behaviour at the aggregate level, while substantial discrepancies emerged at the individual level. Moreover, the gap between self-reports and observed data was systematically related to individual characteristics. Subsequent studies examining specific activities reported consistent patterns of directional bias. For example, Haenschen (2020), using a comparable US survey–trace validation design, found that respondents tended to underestimate status posting but overestimate news sharing on Facebook. Henderson et al. (2021) reported comparable discrepancies for Twitter use and further showed that reporting accuracy varies with the time frame specified in survey questions. Using a sample of 397 Google + users, Staddon et al. (2013) showed that self-reported data can be highly inaccurate when respondents are asked to report behaviours that occur infrequently and irregularly, such as editing profile fields.

Other research has focused on general measures of platform use, such as time spent. An early study from Junco (2013) using a sample of US undergraduate students compared self-reported Facebook use with computer-monitored log data and found that although self-reports were positively correlated with actual use, respondents substantially overestimated their time spent on Facebook. Using data from 15 countries, Ernala et al. (2020) found moderate correlations between self-reports and observed Facebook use alongside systematic overestimation of time spent on the platform. Similar overreports of Facebook use were documented among German students by Naab et al. (2019). Extending this approach to scale-based measurement, Mahalingham et al. (2023) assessed the validity of a problematic social media use scale and found only a weak relationship between smartphone data and survey measures.

Beyond measurement error at the individual level, several studies highlight substantive consequences of misreporting. For instance, Shin (2020) showed that partisan differences in media diets appeared substantially larger when measured using self-reports than when based on digital trace data, raising concerns about biased substantive inference.

Building on this literature, we therefore formulate the following hypotheses:

H1a: Self-reported Facebook use is positively and moderately correlated with observed Facebook behaviour as captured by digital trace data.

H1b: Self-reported frequencies of Facebook use systematically differ from observed behaviour, indicating directional over- or underestimation rather than random measurement error.

Earlier research indicates that the accuracy of self-reports differs according to the nature of the activity. For instance, Guess et al. (2019) found higher correlations when comparing general posting or tweeting behaviour than posting/tweeting about politics. Haenschen (2020) reported underreports of status posting and overreports of news sharing behaviour. Henderson et al. (2021) found that survey questions regarding the frequency of posting and retweeting on Twitter were more closely aligned with recent activity, compared to questions about how often individuals share photos, videos, or send direct messages. Drawing on these findings, we pose the following research question.

RQ1: Which types of self-reported activities correlate more strongly with actual behaviour found in the digital data?

The role of question order

Digital trace data allows researchers to assess the quality of different questionnaire designs. Survey methodologists have long been interested in question order effects, i.e., the problem that responses to later items in a questionnaire can be influenced by earlier items (Schuman & Presser, 1981). A specific case of question order effects is expected for so-called part-whole questions, where one general question precedes a set of specific questions or vice versa (Schwarz et al., 1991). When the general question precedes the specific questions, conversational norms discourage repeating information, and respondents tend to interpret the next specific question as if it were worded, 'aside from what you already told us.' This results in contrast effects. In turn, when specific questions precede the general question, the conversational practice of 'summing up' is more likely to occur. In this scenario, the general question is often seen as a summary, leading to assimilation effects. Both contrast and assimilation effects were found for various types of questions and topics, including well-being (Kaplan et al., 2013; Schwarz et al., 1991; Weinberg et al., 2018), health (Garbarski et al., 2015; Izumi et al., 2023) or satisfaction with democracy (Papp et al., 2024).

Part-whole questionnaire designs can be used in the measurement of social media use. For instance, researchers are often interested in both the overall frequency of platform use and the frequency of specific activities performed on the platforms. Platform use frequency differs from metrics like life satisfaction scores, as general platform use represents the maximum frequency, while sub-activities can only match or fall below this level. Drawing on question-order theories, we suspected that in general-specific designs, individuals would use specific questions to *say something different*. In this context, *different* likely refers to a lower frequency. Since we anticipated contrast effects in the general-specific design, this should result in lower specific frequencies compared to the specific-general design. Accordingly, general use frequency should be lower in the specific-general design, as individuals may align their general use frequency with the lower expected specific use frequencies.

These theoretical expectations yield clear, testable implications for reported activity frequencies. We therefore formulate the following hypotheses concerning question order effects in self-reported social media use.

> H2a: General self-reported frequency of each activity will be higher in the general-specific condition.

> H2b: Specific self-reported frequency of each activity will be higher in the specific-general condition.

Some of the literature recommends asking the specific first because general questions are hard to answer and open to many interpretations (McClendon & O'Brien, 1988). Research offering 'ground truth' information on the accuracy of the two approaches, however, is lacking. Therefore, we pose the following research question.

> RQ2: Which of the two questionnaire designs correlates more strongly with actual behaviour?

The moderating role of political interest

Political interest is expected to influence how individuals engage with and report their political activities on social media. Political interest was found to be a strong predictor of political behaviour on social media (Vitak et al., 2011) and exposure to political content on these platforms (Wells & Thorson, 2017).

Beyond levels of participation, political interest is also expected to shape the accuracy of self-reports of political behaviour. Individuals with higher political interest are more likely to monitor, encode, and recall their political activities, resulting in more accurate self-reports, whereas low-interest respondents are more prone to measurement error due to weaker behavioural awareness and recall. Empirical support for this claim is provided by Haenschen (2020), who contrasted self-reports of Facebook activity with behavioural trace data and showed that political interest reduced underreporting of political activities and moderated reporting accuracy, particularly among high-volume users.

Political interest is also likely to shape how strongly question order affects responses. People who are highly interested in politics think about it often and hold firmer, more clearly formed opinions, so they can usually answer questions about their own political behaviour directly from memory (Krosnick, 1991; Zaller, 1992). Because their answers depend less on cues from the surrounding questions, they should be relatively unaffected by how the questionnaire is arranged, including part–whole question ordering. Respondents with little interest in politics, by contrast, have weaker and less stable impressions of their own political behaviour, so they rely more on the immediate context of the questionnaire when answering. This makes them more prone to question-order effects, particularly for routine online activities that are difficult to recall. Taken together, these considerations suggest that political interest systematically conditions the accuracy of self-reported political behaviour.

H3a: Correlations between reported and actual political behaviour (H1) will be stronger among individuals with higher levels of political interest.

H3b: Differences in correlations between reported and actual political behaviour across two question order designs will be smaller among individuals with higher levels of political interest.

Data and methods

Data

We used a data donation approach (Boeschoten et al., 2022; Breuer et al., 2023) to collect the digital traces of the respondents. In the data donation, people share their digital traces through Data Download Packages (DDPs). These DDPs, hypothetically, contain all the digital traces that platforms record about the user. Since passive behaviour data is not recorded by the platforms or has not been added to the DDPs, our paper focuses on active behaviour, such as posting, commenting, sharing, and reacting.[1] Data comes from a study, fielded in Hungary, between February and June 2023. Respondents were recruited from Hungary's largest non-probability-based online panel, NRC. Compared to the general population of Hungary, individuals with a high level of education and from bigger cities are overrepresented in the panel. Respondents were asked to participate in the project in three ways: answer a short questionnaire with eligibility questions and a consent form (1), download and upload their social media data to the project's website (2), and answer a 30-minute-long survey. The invitation letter contained information about the project's aims and incentives and a direct link to the project's dedicated webpage. The request was designed based on a preliminary vignette experiment (see Kmetty et al., 2025).

If the panel member was eligible (had both a Google and a Facebook profile and used both at least monthly), they were presented with a webpage containing detailed guides for each platform (Google, Facebook, Instagram, Twitter, and TikTok). These guides provided step-by-step instructions for exporting and downloading information from the platforms, accompanied by a video tutorial covering the same content. After downloading their data, participants were required to upload the files for each platform to the project's website without opening or extracting them. Uploading data for Google and Facebook was mandatory. Respondents got around 25 euros as an incentive for participating in the study. Those who donated additional datasets (Instagram/TikTok/Twitter) could get additional incentives (maximum 35 euros). Further information is available on the data collection in this OSF project: https://osf.io/7uybr/

After uploading their data, participants were asked to complete a 30-minute questionnaire covering various topics, including social status, leisure activities, news consumption, and political attitudes. By the end of the study, 755 complete questionnaires were collected. Sixteen percent of those who started the data collection process filled out the survey and donated their data. To mitigate potential sampling biases, we applied individual weights to align the sample more closely with population distributions based on socio-demographic factors. The weighting factors include gender, age, education level, type of settlement, and geographical region. Detailed information on the sample distribution and sample bias is available in Kmetty and Stefkovics (2025) and Chen et al. (2025). The socio-demographic distribution of the sample (unweighted/weighted) is available in Table S1 in the supplementary. The data collection project has been approved by the Centre for Social Science Ethical Board (1-FOIG/130-37/2022).

Self-reported measures

We used a standard self-reported measure of Facebook activity (Guess et al., 2019; Henderson et al., 2021) by asking 'How often do you do the following on Facebook?.' The activities were listed in a table on one page. The grid consisted of 14 items. The general questions were related to posting, sharing, reacting and commenting 'in general.' The specific items asked respondents about the same activities, but 'on political content' and 'on health and health-preventive behaviours content.' In this paper, we focus on the politics-related questions. Politics is a field where the impact of social media is extensively studied, and Guess et al. (2019) showed that the accuracy of self-reports can be different for politically related questions. Response options included 'Several times a day,' 'About once a day,' '3–6 days a week,' '1–2 days a week,' 'Every few weeks,' 'Less often' and 'Never.' Political interest was measured by asking 'In general, how interested are you in politics?,' with five response options ranging from 'I don't care at all,' to 'I'm very interested.'

The experimental design

As outlined in the pre-registration,[2] respondents were randomly assigned to one of the two experimental groups. The general-specific group received the general question of each activity first, followed by the two specific items. The specific-general group, in turn, received the specific items of each activity first, followed by the general item. The manipulated questions were in the first 10% of the questionnaire. We did a balance test for gender, age categories, education level, municipality, and political interest. None of them was significantly different between the experimental groups. We note that the sequence of activities was fixed, thus, earlier items may have primed respondents' recall of Facebook use, potentially influencing how they interpreted and reported subsequent activities.

Link to the pre-registration: https://osf.io/gdqj3/

Digital trace measures

In this study, we use the donated Facebook datasets of the study respondents. These datasets include all the reactions and comments provided by the respondents, as well as their sharing and posting activities. All the activities have a timestamp. In the case of comments, posts, and shares, the data includes the textual content of the activity. For example, if a respondent shared a link and added a text to the share, the DDP includes this text. In the case of commenting and reactions, the DDP includes the 'target' of the activity. We can determine if the target is a page, group, or other Facebook member. During the data preprocessing, we masked all non-public names from the texts.[3] Additionally, we masked the names of targets in comments and reactions, except for those of pages and groups. Following the logic of Guess et al. (2019), we calculated the daily average activity for the different activity types. For general activity, we measured the daily number of posts, shares, comments, and reactions. During the explanatory phase of the project, we tested alternative methods to measure respondents' activity, such as the number of days on which the specific activity was performed. The alignment and correlation between digital and survey data were similar to those of this alternative measure. A detailed description of this robustness test is available in the supplementary.

For measuring political activity, the task was more complex, as we had to categorize the content of the activity. We followed different approaches for different activities (see Table 1 for an overview). In the case of posts, shares, and comments, we used an LLM-based text classification method to distinguish between political and non-political content. A detailed description of the text classification is available in the supplementary. Based on manual validation of the classification, our model performs well on predicting political content in comments, but it performs moderately on predicting shares and poorly for posts. It is also important to note that we can apply this classification only when textual content is available. For comments and posts, this was not a problem, as nearly all content contained text. But 86 percent of the sharing did not include any text from the users. This means that only 14 percent of the sharing has been classified by our model. This causes a serious underestimation of political activity in sharing. Given that the classifier performs poorly on posts (and that the large majority of shares cannot be classified at all), we do not consider the digital measures of political posting and political sharing reliable enough for a meaningful comparison with the survey self-reports. We therefore exclude these two activities from the main analysis

Table 1. Processed data, completeness rates and methods.

Type	Processed data	Completeness rates	Method
Posting	Text of the post	0.2% without text	LLM method
Sharing	Text of sharing	86% without text	LLM method
Commenting	Text of commenting	1.8% without text	LLM method
Commenting	type of comment target	75.1% of the targets not page	Target classification
Reaction	type of reaction target	69.5% of the targets not page	Target classification

and report them only in the supplementary material (Figure S5), where they illustrate the limits of DDP-based classification. The analysis of political behaviour in the main text focuses on commenting and reactions, for which content classification is more reliable.

In the case of reactions and comments, we knew the target of the activity. If the target was a page, we could categorize it as either a political or non-political page. Here, we followed two approaches. The first approach was to use Facebook's page category list.[4] All Facebook pages are linked to a category. We manually reviewed the category list and defined a set of categories that could be linked to political content. There are two potential limitations associated with this approach. Facebook page categories are not perfect, and some categories include very different page types. Publishers could be book publishers focusing on literary works, but we also found a fact-checking news portal in this category. Another issue with this approach is that we overlook the context of the activity in the DDPs. We do not see in the data what content the respondent reacted to and commented on. A reaction to a politician is likely a political reaction. However, in a news media site, it could be political content, but it could also be sports content. This is a common problem with all similar approaches that are based on domain-level classification (see for example: Mangold et al., 2024). The list of categories used as a political category is available in the Supplementary Information (Table S2). We tested the alignment of survey and digital data with a narrower set of page categorization, too. This robustness analysis is available in the supplementary. Regarding the second approach, we utilized our own page label categorization. We compiled a list of political pages in a previous project.[5] This approach is more accurate for pages with a bigger audience, but it could miss some smaller pages.

Around 25 percent of the targets of the comments were pages, and 30 percent of the targets of the reactions. In other cases, the targets were groups or other persons, where target classification could not work. Although we can argue that political commentary and reactions take place most often in the political pages section of Facebook, we will surely underestimate political activity with this approach.

For comments, both text-based and page-based classification were available; thus, we included different approaches in the study for robustness checks.

There was one additional aspect we had to consider when we calculated the measures – the period we wanted to include in the digital trace data. A study of Twitter (Henderson et al., 2021) shows that a shorter period, which is closer to the survey data collection, can align better with the survey responses, as people tend to consider general media use questions based on their recent activity. However, a more extended period could smooth the randomness of activities, which could be a strong bias factor in highly skewed digital trace data. We decided to use a half-year approach; thus, we included the last 182 days of data from the DDPs. However, as a robustness check, we also tested the measures with 30, 60, and 365 days. We will refer to these tests in the Results section. The descriptive statistics for general and political digital items are available in Tables S3 and S4 in the supplementary.

Data transformation, used statistics and outlier handling

In the survey, the different activities were measured on an ordinal scale, while in the digital trace data, we measured the daily average activity level. To compare the data from the two sources, we had to align them. For the analysis, we created an ordinal version of the digital trace data. We followed the recoding logic of Guess et al. (2019) here. See Table 2 for the coding scheme.

In the main text, we use Spearman's rank correlation to analyse the relationship between survey and digital data. To assess absolute agreement between self-reported survey data and digital trace data, we calculated the Intraclass Correlation Coefficient (ICC). Specifically, a two-way random-effects model, absolute agreement, single measurement (ICC(2,1)) was utilized (Koo & Li, 2016). For comparison with earlier works, we also tested the Pearson correlation between digital and survey data. Pearson correlation could be heavily affected by outliers in the digital data. In Pearson correlation analyses, we omit outliers from the analysis. We discuss the outlier handling in detail in the supplementary.

Results

Survey vs digital traces

First, we address H1a, H1b, and RQ1 and assess the alignment and correlations between the survey and digital trace data (See Figures 1–4).

As for general activity, the correlation of survey responses and digital traces was between 0.48 (posting) and 0.64 (reaction). This is a medium-strong correlation.[6] The Pearson correlations were somewhat smaller, but still at a medium level, between 0.33 and 0.48 (See Table S9 in the supplementary). We also compared the categorical distribution of the activities coming from the two data sources. We observed slightly higher comment activity, as well as considerably higher reaction activity, in the digital data. The alignment between the survey and the categorized digital data was highest for sharing (ICC = 0.6). For the other three measures, the ICC was around 0.4. To test for systematic directional bias (H1b), we compared the matched survey and digital categories for each general activity using paired Wilcoxon signed-rank tests. The discrepancy was concentrated in the most frequent behaviours: respondents reported both commenting and reacting as less frequent than observed in the digital data, i.e., they under-reported these activities (both $p < 0.001$): for the typical (median) respondent, the self-reported category was one step lower on the seven-point frequency scale than the category derived from the digital traces. For posting the survey and digital categories also differed significantly ($p < 0.001$), although for the typical respondent the two categories coincided; for sharing the difference was not significant ($p = 0.87$). The systematic bias is thus concentrated in the most frequent, habitual activities, which respondents tend to under-report, rather than reflecting a general tendency to over-report less frequent activities.

Table 2. Mapping between survey and digital trace data categories.

Digital trace ordinal variable	Daily activity (average number of activities per day)
Never	0
Less often	0–0.02
Every few weeks	0.02–0.047
1–2 days a week	0.047–0.285
3–6 days a week	0.285–0.857
About once a day	0.857–1.5
Several times a day	>1.5

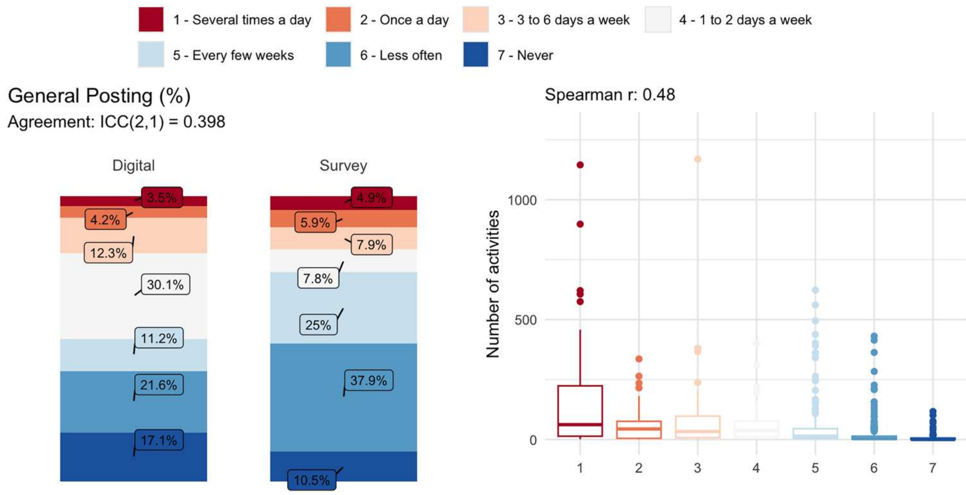

Figure 1. Contrasting survey responses and digital traces – General Posting.

Overall, we found lower correlations between the digital traces and the survey data for political than for general activities. As noted in the Methods, political posting and political sharing cannot be reliably measured from the digital traces, so we report them only in the supplementary material (Figure S5) and focus here on political commenting and reactions.

For political commenting, textual data were available and the political classifier performed well (Figure 4). The correlation between the two data sources was 0.43, which is still lower than in the case of general commenting. We obtained the same level of correlation for the other classification methods, including the manual page classification and the classification based on Facebook's categories. For activity level estimation, the

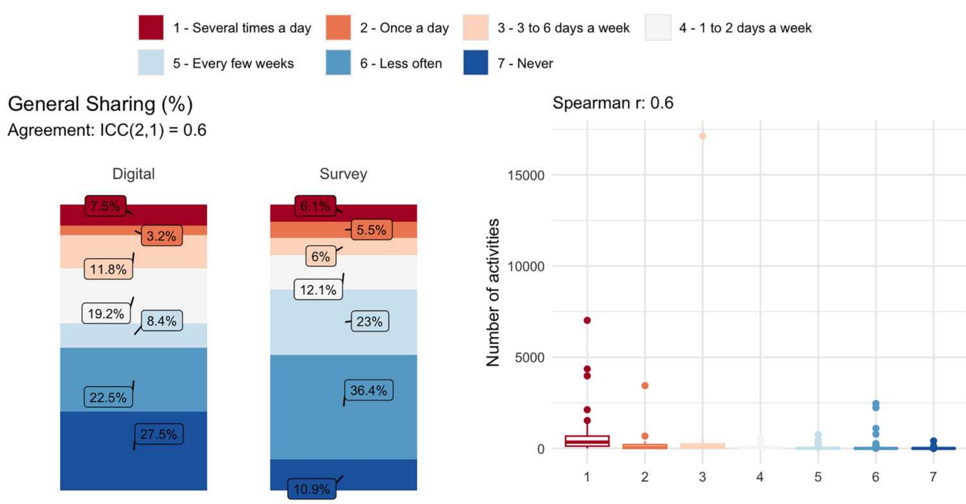

Figure 2. Contrasting survey responses and digital traces – General Sharing.

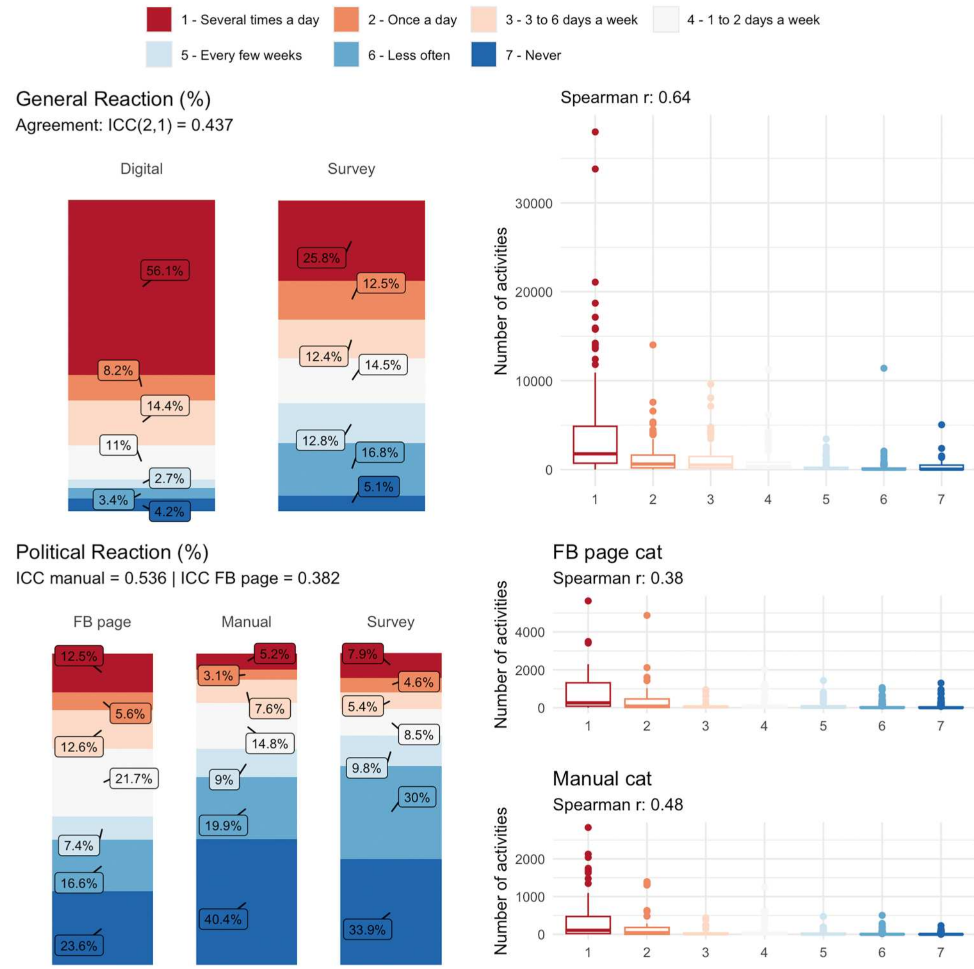

Figure 3. Contrasting survey responses and digital traces – Reaction.

Facebook category-based approach aligned better with the survey; the LLM-based classification and the manual page classification underestimated the activity level compared to the survey.

The last activity we compare is political reactions (Figure 3). The correlation between digital and survey data was lower for political measures (both) than for general activity, but it remained medium-strong. In the case of Pearson correlation, we did not find a difference in the correlation levels between the general and political items and the survey items (see Table S9 in the supplementary). The activity level of political reaction by survey and digital data (both classifications) was also close to each other.

We calculated all correlations for different periods to ensure robustness (see Figure S1). The correlations were broadly stable across window lengths. Point estimates were somewhat lower at 30 days for some activities, but the confidence intervals overlapped across all window lengths, so we do not interpret these differences as statistically meaningful.

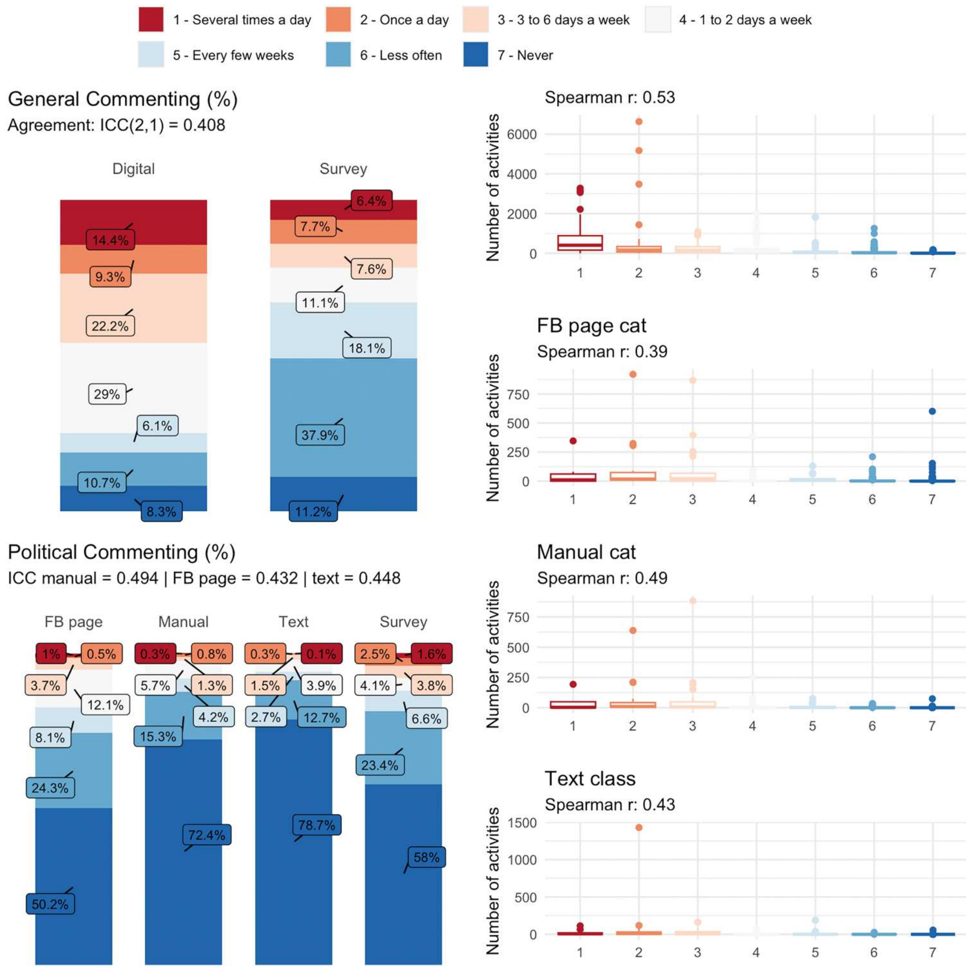

Figure 4. Contrasting survey responses and digital traces – Commenting.

The role of questionnaire design

Our second research question (RQ2) was which questionnaire design aligns better with the digital traces. In order to contextualize our results, we briefly introduced the differences in activity levels by question design in the survey data. We expected that general use frequency would be lower in the specific-general design if individuals aligned their general use frequency with the lower expected specific use frequencies (contrast effect). Our results only partially confirm this expectation; this effect was only present in the case of general reactions. For the other three types of activity, we did not find significant differences (See Figure S2). For the specific self-reported frequency, we expected the opposite (assimilation effect). However, here, we did not find any significant effect in the politically related activities (see Figure S3). To sum up, we did not find strong evidence for either contrast or assimilation effects.

To answer RQ2, we calculated the correlations between digital traces and survey questions separately for the two designs. The correlation was stronger between survey

Table 3. Experimental design correlations and significance.

Category	Activity	Experimental design		
		Gen to Spec	Spec to Gen	Sig. (Fisher Z)
General	Post	0.49	0.47	0.73
	Share	0.6	0.6	0.97
	Comment	0.59	0.47	0.04*
	Reaction	0.66	0.62	0.35
Political	Comment – manual page	0.48	0.51	0.51
	Reaction – manual page	0.51	0.45	0.29

*$p < 0.05$.

responses and the digital data in the case of general commenting in the general-to-specific design (see Table 3).

For political activities, we did not include posting and sharing, as we showed earlier that our approach can only poorly measure them in the digital space. For reactions and comments, we did not find a significant difference between the question designs. We also note that the single significant difference in Table 3 (general commenting, $p = 0.04$) is one of six comparisons reported in the table and would not survive a correction for multiple testing. It should therefore be interpreted with caution, as it may reflect a chance finding.

In the last part of the analysis, we focus on how political interest alters the correlation between survey responses and political digital traces (H3a/b).

We calculated correlations between digital and survey data across different question designs for low-to-moderate (1–3 on the scale) and high (4–5 on the scale) political interest groups (see Figure 5). We saw a strong effect of political interest. Regardless of the activity type (comment, reaction), people with higher political interests have a stronger correlation between their digital and survey data. However, we did not observe a significant question-order design. So we could not confirm that the differences in correlations between reported and actual political behaviour across two question-order designs would be smaller among individuals with higher levels of political interest (H3b). We tested the correlations with alternative cut points of the political interest scale (1-2/3-5). The results were robust regarding our hypothesis (see Figure S4 in the supplementary).

Discussion and conclusion

Self-reported measures of social media use are widely employed across disciplines. However, the validity of findings in this area critically depends on the accuracy of these self-reports. In the context of political communication research, measurement error in reported online political behaviour can bias estimates of media effects and lead to misinformed conclusions about political engagement. Our study used donated digital trace data from Facebook and contrasted it with answers to a standard survey question from the same individuals.

In line with prior evidence (Ernala et al., 2020; Guess et al., 2019; Haenschen, 2020; Mahalingham et al., 2023; Naab et al., 2019), we found moderate correlations between self-reports and digital behaviour (H1a). Correlations varied strongly between the types of activities. Answers to the general items align more closely with actual behaviour compared to specific, political activities. As suggested by the availability heuristic, people

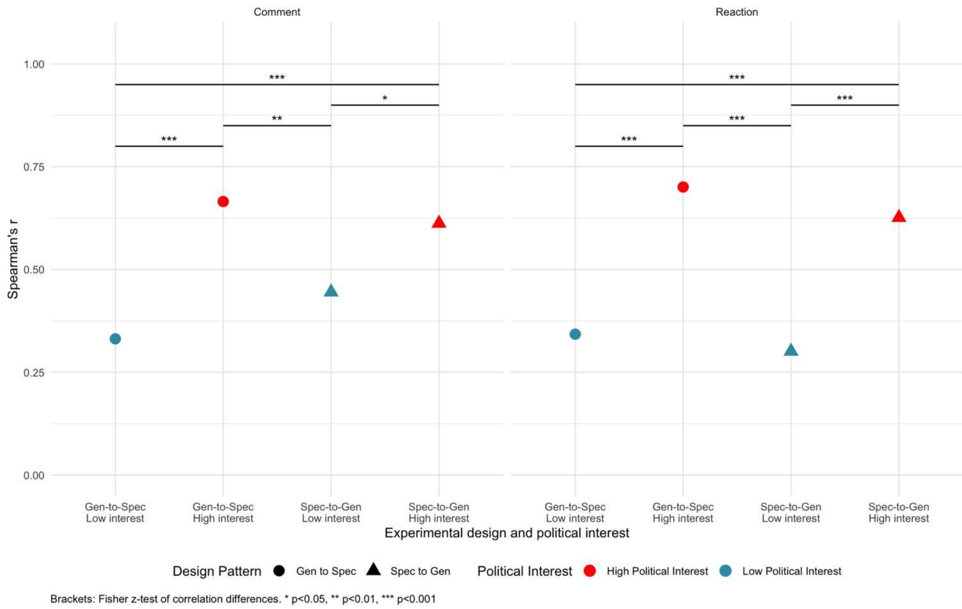

Figure 5. Correlation between survey responses and digital traces (manual page categorization) in the two questionnaire designs by political interest. Significance stars from Fisher z-tests of correlation differences: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

estimate the frequency of an event based on how easily examples come to mind (Tversky & Kahneman, 1973). General social media use is likely to activate more readily accessible memories because it is highly habitual and frequent, whereas topic-specific behaviours are more episodic, occur less often, and are therefore harder to recall accurately (Prior, 2009; Scharkow, 2016). Alternatively, individuals may not necessarily overestimate their specific political activities, but such behaviours are more difficult to detect in digital trace data. This is due to the fragmented nature of the digital data, and the lack of contextual information needed to classify an action as politically relevant. We discuss these limitations later in detail.

Furthermore, the paired comparisons showed that this systematic discrepancy was concentrated in the most frequent, habitual general activities, commenting and reacting, which respondents under-reported relative to the digital traces, whereas posting and sharing showed little or no systematic directional bias (H1b). This echoes the findings of Haenschen (2020) and studies examining internet use (Araujo et al., 2017; Scharkow, 2016; Schwarz & Oyserman, 2001). A possible explanation is that regular, frequently performed behaviours, such as browsing or checking Facebook, may become habitual and automatic, drawing little conscious attention. As a result, they are less likely to be encoded into episodic memory, which can lead to systematic underreporting in self-reports (Bargh, 1994). Less frequent and more deliberate actions might, in principle, stand out as more cognitively accessible and thus be more prone to over-reporting; however, we did not find clear evidence of such over-reporting among the activities we could measure reliably. Nevertheless, the questions about general activities still yielded more accurate responses.

In contrast to our expectations (H2a, H2b), the two question orders produced largely similar self-reports. The absence of strong question order effects may reflect that respondents rely on stable heuristics when reporting their Facebook activities, rather than constructing responses based on the immediate question context. In addition, the clear semantic separation between general and content-specific items, respondents' tendency to recall frequent social media use as a habitual routine rather than to reconstruct it from the immediate survey context (Prior, 2009; Scharkow, 2016), and the fact that our design included only a limited number of specific items rather than an extensive list, as in many prior part–whole studies, may have reduced the scope for classic contrast or assimilation effects to emerge.

The two types of questionnaire design yielded different accuracy in only one case: the correlation for self-reported general commenting was higher in the general-specific design. This was, however, the only significant difference among the six comparisons in Table 3 and would not survive a correction for multiple testing, so we treat it with caution and refrain from drawing strong conclusions from it. Future research should examine different designs, e.g., grids with more specific activities.

The overlap between self-reports and digital behaviour was moderated by political interest. Echoing the findings of Haenschen (2020), higher political interest generally reduced the likelihood of underreporting, suggesting that it increases individuals' awareness of their engagement in political activity. We recommend that researchers at least control for political interest in their analysis to mitigate measurement error.

We would like to highlight the challenges associated with comparing digital trace data to survey results, particularly in the case of political behaviour. One natural issue is the discrepancy between what people consider *political* and what our models coded as political. Digital traces often lack contextual information, making it difficult to reliably classify actions as politically relevant. Two further explanations deserve emphasis. First, what counts as 'political' is itself heterogeneous: respondents may differ from one another, and from our classifier, in where they draw the boundary of political content, so part of the gap between the survey and the digital measures may reflect differing definitions rather than reporting error. Second, a substantial share of political communication takes place in private messages, which are not contained in the DDPs. Such exchanges are invisible in the digital traces but may nonetheless inform respondents' self-reports, widening the apparent discrepancy between the two sources. Moreover, in many cases, it was not possible to accurately assess the political nature of the digital source, due to restrictions in the data. Text-based classification only worked for those activities where textual data was available. Estimating sharing behaviour was especially challenging, as 85 per cent of the sharing activity did not contain any textual data. And the performance of the text classifier differed between comments and posts. We used a model developed from social media comments. Our results highlight that activity-specific models have to be trained for political text classification. More broadly, classifying political content from digital traces is inherently difficult and error-prone: a substantial part of everyday Facebook communication is habitual and non-political (greetings, condolences, weather) yet superficially resembles political language, and the boundary of what counts as political is itself fuzzy. We therefore recommend that future projects map and pre-filter clearly non-political, habitual content before classification, and we stress that the comparison between survey and digital data can only be pushed as far as the quality of the underlying classification allows. The structure of the DDPs, particularly for posting and sharing, also makes the identification of relevant activities difficult. As suggested by Guess et al. (2019), it would be useful to provide anchoring guidance in survey questions in future research, such as referencing posts about policy, government affairs, or societal debates, along with illustrative examples to clarify the intended scope. Decisions on the time frame of the digital data have a significant impact on comparisons (Henderson et al., 2021). In this study, the overlap between self-reports and digital data was, if anything, slightly higher when using 6 months or longer periods than when using a 30-day period, although these differences were small and the confidence intervals overlapped (Figure S1). This tentatively suggests that people may consider relatively long periods when assessing the frequency of their activities, which would be somewhat in contradiction to what Henderson et al. (2021) found on Twitter. However, we also need to consider how we map the digital data to survey data. Mapping digital and survey data is not a straightforward task. In the main part of the paper, we followed the recoding logic of Guess et al. (2019) and mapped digital data to a survey based on users' average number of daily activities. This approach could put some respondents into a more frequent category if they use Facebook on a few days, but very frequently on those days. To assess this potential bias, we tested different coding approaches mainly based on the number of days the respondent did a particular activity (see Table S7 in the supplementary). The difference between the Spearman correlations between the main coding and the alternative coding was remarkably small. The ICC was also very similar in most cases, except for two: it was higher for the alternative coding of general reaction and general comments. These results suggest that for frequent activities more complex coding schemes could add more precise results.

Our study comes with some limitations, many of them related to the digital data. LLM-based classifiers are not perfect. Our model was trained on public Facebook comments, and it worked reasonably well for political comments coming from the DDPs. However, posting is a different activity type, and the model struggled with classifying it. Where targets were available, we can use a page classification approach. But page classification only worked if the target was a Facebook page. We might assume that political comments are more frequent in the case of pages compared to comments on friends. Still, as a large chunk of the comments (and reactions) involve interactions with Facebook friends, the page-based approach would miss some relevant activities. But there is also a potential overestimation in the page-based approach. Many pages include both political and non-political content. News sites publish a wide range of news, including political news, sports news, weather updates, and accident reports. In this page-based approach, we cannot recognize the content of the page on which the respondents reacted or commented. To mitigate this issue, we tested a narrower page classification (see supplementary, Table S10). Nevertheless, this narrower classification did not prove more effective: although its correlation with the survey data was broadly similar to that of the broader approach, it recorded far more respondents as inactive and thereby heavily underestimated the frequency of political reactions and comments. A limitation also arises from aligning the survey answer categories with the digital measures. The survey categories are a mix of daily-based values (posting every week) and variables that contain a daily frequency (posting multiple times per day). It is far from obvious how we translate the activity measured in the digital traces to this scale. Our test results showed that a simple approach performs similarly in terms of ICC, correlation, and distribution to the more complex approach with day-based calculations (see supplementary). This similarity, however, is partly driven by the large number of respondents with no activity, who are coded identically by both schemes; among active users the two coding schemes diverge more, so the choice of coding scheme matters more for frequently performed activities (see supplementary). However, we want to encourage other studies to test various approaches here to find the best solution for this question. Another limitation is that we focus on a single Eastern European country. Given the specificities of post-socialist societies in their behaviour on social media and attitudes towards surveys (Ipsos, 2024), our findings may not easily generalize to other cultures.

While social media use survey measures are routinely used in political communication research, our study has demonstrated that they only moderately reflect actual general or political behaviour. However, part of the lower alignment may reflect classification challenges in the digital trace data rather than true measurement error. Our study also showed that political survey measures perform better among individuals with high political interest. Future research should continue to refine survey instruments by exploring alternative question formats, investigating other platforms and political behaviour.

Notes

1. Even if platforms had to provide complete and comprehensive data to their users, some data recorded by platforms could still be missing in these DDPs (Hase et al., 2024). For example, exposure data is usually missing or available only for a limited time period. This is why we cannot calculate any digital data for the measure on ‘Reading articles, posts without actively engaging with them’.
2. One hypothesis from the pre-registration, concerning the correlation between the general and the specific survey items across the two question orders, falls outside the scope of this paper, which focuses on the comparison of self-reports with digital traces. For completeness, we report its analysis in the supplementary material (Table S11).
3. Differentiating between public and non-public names is not straightforward. But we could use the DDP, which contains the list of pages the respondent followed or liked. This DDP contains only pages. We used this as a reference point and kept the names listed on page DDP.
4. The owners (admins) of the pages define the category when they create the page. They can select the category closest to them from a list provided by Facebook. They can update this category later.
5. To create the initial list, we used a social listening tool called Sentione to collect posts from Hungarian Facebook pages. We used a long list of politicians and politics-related keywords to collect posts from Facebook pages and news sites from 2021 to 2022. We manually checked the pages and coded them as political or non-political by trained annotators. We considered a page political if it posts on actual political issues or on topics related to politics. As we had an ongoing project on analysing the supply side of political content in Hungary, we regularly updated this list after 2022.
6. (Cohen, 2013) defines the correlation below 0.3 as weak, between 0.3 and 0.5 as medium, and above 0.5 as strong.

Acknowledgements

During the preparation of this revised manuscript, the authors used a generative AI assistant (Anthropic Claude, Claude Opus 4) to support language editing and proofreading, to help

restructure and condense parts of the text in response to the reviewers' comments, and to assist in checking and documenting the replication code. The tool was not used to generate or manipulate any research data or results. All AI-assisted output was reviewed and verified by the authors, who made all substantive scientific decisions and take full responsibility for the content of the article.

Author contributions

CRediT: **Zoltán Kmetty:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Writing – original draft; **Ádám Stefkovics:** Conceptualization, Methodology, Writing – original draft.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The data collection was funded by the Eötvös Loránd Research Network within the framework of Flagship Research Projects: KÖ-32/2021. The work of Zoltán Kmetty was supported by the Digital Political Footprint project, funded by NKFIH, with a grant number: K-147329.

Notes on contributors

Zoltán Kmetty, Ph.D., is a research professor at the ELTE Centre for Social Sciences, CSS-RECENS research group; and an associate professor at the Eötvös Loránd University Faculty of Social Sciences, Sociology department. He has diverse research interests, including computational social science, political sociology, and network studies. He is an expert in methodology, survey design, and quantitative analysis [email: kmetty.zoltan@tatk.elte.hu].

Ádám Stefkovics, PhD, is a research fellow at the ELTE Centre for Social Sciences, and Századvég Foundation. His main research interests are survey methodology and political sociology [email: Stefkovics.adam@tk.hu].

Data availability statement

The raw Facebook data (textual content, interaction targets, and event-level timestamps) cannot be shared for privacy reasons. A replication package is available on OSF: https://osf.io/vd42j/. It contains a respondent-level dataset with the survey responses joined to the aggregated digital-trace indicators (counts only, no raw content), the analysis code that reproduces every table and figure in the paper from this dataset, the server-side preprocessing script that built the dataset from the raw data (provided for transparency; not runnable without the restricted raw data), and a README with a full codebook. All statistical results reported in the manuscript can be reproduced from the shared dataset.

Research with human subjects

The data donation study was conducted with the IRB approval of the Centre for Social Science Ethical Board (1-FOIG/130-37/2022) and with the informed consent of the participants. The study adhered to the ethical tenets of the Declaration of Helsinki.

ORCID

Zoltán Kmetty http://orcid.org/0000-0002-6775-8938

References

Araujo, T., Wonneberger, A., Neijens, P., & Vreese, C. d. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, *11*(3), 173–190. https://doi.org/10.1080/19312458.2017.1317337

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Lawrence Erlbaum Associates.

Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, *4*(2), 388–423. https://doi.org/10.5117/CCR2022.2.002.BOES

Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2023). User-centric approaches for collecting Facebook data in the "post-API age": Experiences from two studies and recommendations for future research. *Information, Communication & Society*, *26*(14), 2649–2668. https://doi.org/10.1080/1369118X.2022.2097015

Cardenal, A. S., Victoria-Mas, M., Majó-Vázquez, S., & Lacasa-Mas, I. (2022). Assessing the validity of survey measures for news exposure through digital footprints: Evidence from Spain and the UK. *Political Communication*, *39*(5), 634–651. https://doi.org/10.1080/10584609.2022.2090038

Chen, Y., Kmetty, Z., Iñiguez, G., & Omodei, E. (2025). The public that engages invisibly: What visible engagement fails to capture in online political communication. *Communication Methods and Measures*, *19*(4), 294–312. https://doi.org/10.1080/19312458.2025.2542725

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Coromina, L., & Saris, W. E. (2009). Quality of media use measurement. *International Journal of Public Opinion Research*, *21*(4), 424–450. https://doi.org/10.1093/ijpor/edp014

Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020). How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3313831.3376435

Garbarski, D., Schaeffer, N. C., & Dykema, J. (2015). The effects of response option order and question order on self-rated health. *Quality of Life Research*, *24*(6), 1443–1453. https://doi.org/10.1007/s11136-014-0861-y

Greenberg, B. S., Eastin, M. S., Skalski, P., Cooper, L., Levy, M., & Lachlan, K. (2005). Comparing survey and diary measures of internet and traditional media use. *Communication Reports*, *18*(1-2), 1–8. https://doi.org/10.1080/08934210500084164

Guess, A., Munger, K., Nagler, J., & Tucker, J. (2019). How accurate are survey responses on social media and politics? *Political Communication*, *36*(2), 241–258. https://doi.org/10.1080/10584609.2018.1504840

Haenschen, K. (2020). Self-reported versus digitally recorded: Measuring political activity on facebook. *Social Science Computer Review*, *38*(5), 567–583. https://doi.org/10.1177/0894439318813586

Hase, V., Ausloos, J., Boeschoten, L., Pfiffner, N., Janssen, H., Araujo, T., Carrière, T., Vreese, C. H. d., Haßler, J., Loecherbach, F., Kmetty, Z, Möller, J, Ohme, J, Schmidbauer, E, Struminskaya, B, Trilling, T, Welbers, K, & Haim, H (2024). Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies? *Internet Policy Review*, *13*(3), 1–37. https://doi.org/10.14763/2024.3.1793

Henderson, M., Jiang, K., Johnson, M., & Porter, L. (2021). Measuring Twitter use: Validating survey-based measures. *Social Science Computer Review*, *39*(6), 1121–1141. https://doi.org/10.1177/0894439319896244

Ipsos. (2024). Global trustworthiness index 2024. https://www.ipsos.com/sites/default/files/ct/news/documents/2024-11/ipsos-global-trustworthiness-index-2024.pdf

Izumi, S., Hagiwara, Y., Matsuyama, Y., Shiroiwa, T., Taira, N., Kawahara, T., Konomura, K., Noto, S., Fukuda, T., & Shimozuma, K. (2023). Impacts of the preceding cancer-specific health-related quality of life instruments on the responses to the subsequent EQ-5D-5L. *Health and Quality of Life Outcomes*, *21*(1), 3. https://doi.org/10.1186/s12955-022-02085-8

Junco, R. (2013). Comparing actual and self–reported measures of Facebook use. *Computers in Human Behavior*, *29*(3), 626–631. https://doi.org/10.1016/j.chb.2012.11.007

Kaplan, S. A., Luchman, J. N., & Mock, L. (2013). General and specific question sequence effects in satisfaction surveys: Integrating directional and correlational effects. *Journal of Happiness Studies*, *14*(5), 1443–1458. https://doi.org/10.1007/s10902-012-9388-5

Kmetty, Z., & Stefkovics, Á. (2025). Validating a willingness to share measure of a vignette experiment using real-world behavioral data. *Scientific Reports*, *15*(1), 9319. https://doi.org/10.1038/s41598-025-92349-2

Kmetty, Z., Stefkovics, Á., Számely, J., Deng, D., Kellner, A., Pauló, E., Omodei, E., & Koltai, J. (2025). Determinants of willingness to donate data from social media platforms. *Information, Communication & Society*, *28*(7), 1324–1349. https://doi.org/10.1080/1369118X.2024.2340995

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Mahalingham, T., McEvoy, P. M., & Clarke, P. J. F. (2023). Assessing the validity of self-report social media use: Evidence of no relationship with objective smartphone use. *Computers in Human Behavior*, *140*, 107567. https://doi.org/10.1016/j.chb.2022.107567

Mangold, F., Schoch, D., & Stier, S. (2024). Ideological self-selection in online news exposure: Evidence from Europe and the US. *Science Advances*, *10*(37), eadg9287. https://doi.org/10.1126/sciadv.adg9287

McClendon, M. J., & O'Brien, D. J. (1988). Question-order effects on the determinants of subjective well-being. *Public Opinion Quarterly*, *52*(3), 351–364. https://doi.org/10.1086/269112

Naab, T. K., Karnowski, V., & Schlütz, D. (2019). Reporting mobile social media use: How survey and experience sampling measures differ. *Communication Methods and Measures*, *13*(2), 126–147. https://doi.org/10.1080/19312458.2018.1555799

Papp, Z., Susánszky, P., & Szabó, A. (2024). Question-order effect in the study of satisfaction with democracy: Lessons from three split-ballot experiments. *Sociological Methodology*, *54*(2), 404–417. https://doi.org/10.1177/00811750241254363

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Prior, M. (2009). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, *71*(3), 893–908. https://doi.org/10.1017/S0022381609090781

Revilla, M., Ochoa, C., & Loewe, G. (2017). Using passive data from a meter to complement survey data in order to study online behavior. *Social Science Computer Review*, *35*(4), 521–536. https://doi.org/10.1177/0894439316638457

Scharkow, M. (2016). The accuracy of self-reported internet use—a validation study using client log data. *Communication Methods and Measures*, *10*(1), 13–27. https://doi.org/10.1080/19312458.2015.1118446

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*(2), 127–160. https://doi.org/10.1177/109821400102200202

Schwarz, N., Strack, F., & Mai, H.-P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, *55*(1), 3–23. https://doi.org/10.1086/269239

Shin, J. (2020). How do partisans consume news on social media? A comparison of self-reports with digital trace measures among Twitter users. *Social Media + Society*, *6*(4), 2056305120981039. https://doi.org/10.1177/2056305120981039

Staddon, J., Acquisti, A., & LeFevre, K. (2013). Self-reported social network behavior: Accuracy predictors and implications for the privacy paradox. Proceedings of the 2013 International Conference on Social Computing, 295–302. https://doi.org/10.1109/SocialCom.2013.48.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*(3), 299–314. https://doi.org/10.1037/0033-2909.103.3.299

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Vitak, J., Zube, P., Smock, A., Carr, C. T., Ellison, N., & Lampe, C. (2011). It's complicated: Facebook users' political participation in the 2008 election. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 107–114. https://doi.org/10.1089/cyber.2009.0226

Weinberg, M. K., Seton, C., & Cameron, N. (2018). The measurement of subjective wellbeing: Item-order effects in the personal wellbeing index—adult. *Journal of Happiness Studies*, *19*(1), 315–332. https://doi.org/10.1007/s10902-016-9822-1

Wells, C., & Thorson, K. (2017). Combining big data and survey techniques to model effects of political content flows in facebook. *Social Science Computer Review*, *35*(1), 33–52. https://doi.org/10.1177/0894439315609528

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press.