How We Browse: Measurement and Analysis of Browsing Behavior

Yuliia Lut⋆
Unaffiliated researcher
lutyuliia@gmail.com

Michael Wang⋆
Unaffiliated researcher
mzywang@gmail.com

Elissa M. Redmiles
Dept. of Computer Science
Georgetown University
Washington DC, USA
elissa.redmiles@georgetown.edu

Rachel Cummings
Dept. of IEOR
Columbia University
New York, NY USA
rac2239@columbia.edu

***Abstract*—Accurately analyzing and modeling online browsing behavior play a key role in understanding users and technology interactions. In this work, we design and conduct a user study to collect browsing data from 31 participants continuously for 14 days and self-reported browsing patterns. We combine self-reports and observational data to provide an up-to-date measurement study of online browsing behavior. We use these data to empirically address the following questions: (1) Do structural patterns of browsing differ across demographic groups and types of web use?, (2) Do people have correct perceptions of their behavior online?, and (3) Does the length of time that people are observed relate to changes in their browsing behavior? In response to these questions, we find significant differences in level of activity based on user age, but not based on race or gender. We also find that users have significantly different behavior on Security Concerns websites, which may enable new behavioral methods for automatic detection of security concerns online. We find that users significantly overestimate the time they spend online, but have relatively accurate perceptions of how they spend their time online. We find no significant changes in behavior over the course of the study, which may indicate that the observation had no effect on behavior or that users were consciously aware of being observed throughout the study.**

***Index Terms*—digital browsing behavior, user studies, human factors, web use**

I. Introduction

The amount of time users spend online and how that time is spent has been found to relate to digital skills, demographic characteristics [1], [2], social capital [3], academic performance [4], mental health [5], and socioeconomic attributes [6]. Thus, researchers seek to measure browsing behavior to draw inferences about a variety of different digital constructs.

While ideally researchers would be able to directly observe users' browsing behavior, due to difficulties obtaining access to such data, researchers often rely of users' self reports of their online behavior [7]. Potential concerns have been raised about the accuracy of such self-report data [8]–[11].

Observational methods are an alternative approach that have been applied in a small number of prior works on users browsing behavior (e.g., [12]–[14]). However, much of this prior observational work has been exclusively conducted using proprietary industry data; for instance, [13] and [14] examine browsing behavior records of approximately 250,000 users and 89,480 users respectively. The data of such scale is near-impossible to collect in the academic research setting and therefore, the majority of academic researchers do not have access to such data. Additionally, the most recent prior observational work was last conducted a decade ago, despite strong self-report evidence that people's browsing behavior has changed significantly in the past decade [13], [15]. In this work, we design an experimental study where we can directly observe and record users' behavior. Further, observational methods are not without limitations. A broad literature in behavioral economics has shown that people behave differently when they are aware that their actions are observed (e.g., [16]–[20]). However, this literature has focused primarily on behavior in incentivized economic games, not web behavior.

In this work, we seek to (1) provide an up-to-date understanding of browsing behavior of a young, educated (under 45 years old with some college education) sample of internet users and (2) offer insight on studying browsing behavior by comparing two academically-viable methodologies: self-reports and observational study via a browser plug-in rather than proprietary data. To do so, we conducted a user experiment ($n = 31$), surveying participants about their browsing behavior and observing participants' browsing behavior continuously for 14 days in August and September 2019. Using these data, we address the following research questions:

- (RQ1) Does browsing behavior differ across user groups (i.e., demographics) and types of web use?
- (RQ2) Do people have accurate perceptions of their behavior online? Does perception accuracy differ by user group or type of web use?
- (RQ3) Does the length of time people are observed relate to changes in their browsing behavior?

We observe that people spend more time online, relative to prior work conducted in 2010: median of 2.9 hours daily versus one hour in [12] (RQ1). We found little difference across demographic groups by race and gender, but did find significant differences by age, with older participants (aged 35-44) browsing less than younger groups (aged 18-24 and 25-34) across multiple metrics of browsing activity. We find

⋆This work was completed while the first author was at Columbia University and second author was at Georgia Institute of Technology.

Y.L. and R.C. were supported in part by NSF grants CNS-2138834 (CAREER).

few significant differences in browsing behavior across website categories. One notable exception is the Security Concerns category, which had significantly distinct browsing patterns ($p < 10^{-5}$). We suggest ways that this finding can be used to automate detection of security concerns online.

We find that people substantially overestimate their time spent online (80.6% of our participants, by an average of 4.5 hours) when self-reporting vs. when this time is observed via browser plugin (RQ2). This overestimation persists even after controlling for various methodological alternatives and is consistent across demographic groups. However, participants have roughly accurate perceptions of their top-browsed website categories: 50.3% of reported top browsing categories were indeed in the participant's observed top categories.

We hypothesized that participants will have higher awareness of being observed – which prior literature finds leads to changes in behavior [16]–[20] – early in the study, and lower awareness later in the study. However, we do not find changes in browsing activity or in distribution of browsing across website categories throughout the study (RQ3). This could indicate that people do not alter their browsing behavior when aware of being observed, or that a 14-days study is insufficient for them to forget that their data are being collected.

Overall, our findings contribute the following insights:

- People spend significantly more time online, with similar browsing patterns across websites, except for security threat websites which exhibit different browsing patterns.
- Self-report methodologies are appropriate for collecting relative data about people's browsing behavior such as what websites they browse the most.
- Self-report methods are unsuitable for accurately measuring browsing time as people over-report their time spent. This finding confirms and expands prior research about over-reporting time spent on Facebook [9].
- Future work is necessary to (a) confirm our finding of little sociodemographic variance in browsing behavior (except by age) and (b) further explore the lack of observational-bias in our participants' browsing behavior.

II. Related Work

Prior work finds correlations between users' browsing behavior and their demographic or behavioral type. [13] show that internet usage varies more by education than by demographic features. [12] propose website categorization based on log data from Yahoo! users that contains over 50 million page views collected over a week. They show that the top five categories (news, portals, games, verticals, multimedia) account for more than half of all Web activity. However, their analysis does not include demographic data. [14] offer a methodology to predict gender and age from observed browsing behavior, which provides 30.4% and 50.3% improvements respectively compared to baseline algorithms. [21] show that users' browsing patterns can be uniquely identified by the types of websites they access and time of the access with at least 75% accuracy. [2] identifies individuals' daily activity patterns by analyzing their mobile app usage data.

While these works lay a foundation for measuring user behavior online, they do not elicit user perceptions of their own browsing behavior. In this work, we both elicit self-perceptions of browsing and measure browsing behavior, which allows us to evaluate accuracy of users' perceptions.

The accuracy of user perceptions of browsing behavior has been previously studied in limited contexts. [9] show that self-reported data can be often unreliable, and that people tend to overestimate the time they spend on Facebook but underestimate the number of times they visit. [22] find the substantial discrepancy between self-reported and the logged media usage. They discovered similar proportions of under- and over- reporting cases, with 6.12% of self reports within 5% of the log-value estimate. They concluded that self-reported media usage data is rarely accurate and should not be automatically considered a suitable substitute of logged data. Similarly, both [23] and [24] find discrepancies between self-reported and objective smartphone and social media usage data. [4] find that longer times spent on an online educational platform are associated with higher grades. We extend this work in two key ways. First, we examine the accuracy of users' perceptions of their overall browsing behavior rather than on one specific website. Second, we test not only the accuracy of users' perceptions about the time spend online, but also which categories of websites they most frequently browse.

Beyond online browsing behavior studies, prior work in the security domain has examined the accuracy of self-reports. [10] follow a methodology similar to our own, comparing the participants' behavior observed via software for over six weeks to their self-reports about digital security behavior, finding low correlation between self-reported and observed behavior. [8] studied software updating behavior, finding significant correlation between self-reported intentions in response to software update prompts with behavioral responses to the same prompts observed through proprietary industry data. They also find that self-reporting participants reported that they would update faster than observed users did.

III. Methods

To answer our research questions we observed the browsing behavior of 31 participants over a period of 14 days in August and September 2019, and assessed participants' self-reported perceptions of their browsing behavior. In this section we describe our study procedures (Section III-A), data collection (Section III-B), data analysis (Section III-C), and study limitations (Section III-D). All study procedures were approved by our institution's ethics review board.

A. Study Procedures

We recruited participants by flyers on bulletin boards and student gathering spaces on the campus of a large public institution for higher education. The flyer included a link to the online screening survey for eligibility verification and sign-up.

The brief screening survey, hosted on Qualtrics, verified that participants met the study's eligibility criteria: aged 18 or older, native English speakers, and browsing the internet at

TABLE I: Action types collected through the extension

awake	This action indicates that a user is online. Appears every 5 minutes when browser is open and online. Can occur when a user is not actively browsing.
backButton	Clicking on the back button
click	Click that does not cause URL change.
newTab	Opening new tab
omnibox	Typing in omnibox (address bar / search engine)
tabChange	Alternating between existing tabs
type	Typing a single character
urlChange	Click that causes URL change.

least 5 hours per week. Eligible participants were invited to the lab to sign a consent form, complete a pre-study survey, and install the Chrome extension for browsing data collection, to mitigate issues of digital inequity by providing installation support. The pre-study survey gathered participants' demographic information and perceptions of their browsing habits.

Over the next 14 days, the extension tracked participants' web activity, including websites visits, actions within websites, and timestamps of each action. On the 14th day of the study, participants returned to the lab to uninstalled the extension and complete a brief post-study survey on whether their browsing behavior changed over the course of the study. For those unable to return to the lab in-person, the data collection was truncated after 14 days. More details on the surveys and extension-based data collection are provided in Section III-B.

Participants were paid $200 for their full participation in the study, with the option to exit early for prorated payment. No participants exercised this option.

B. Data Collection

Here we describe the browsing extension (Section III-B1) and self-report data collection (Section III-B2).

1) Extension-based Data: We developed a Chrome extension for recording participants' browsing behavior and metadata, which sends the collected data to a server for storage. The anonymized JavaScript code for the extension is available at https://github.com/browsing-experiment/browsing-extension.

A list of the user browsing actions observed through this extension is provided in Table I, which are common browsing actions that generate or affect internet packets from browsing activity. An additional *awake* action was generated every 5 minutes if the browser was online, to ensure the extension was not disabled. The extension recorded the action along with metadata such as the event time, URL (if any), and participant ID, then sent these data to a secure server.

To protect participants' privacy, each was assigned a random ID associated with their browsing actions without linking to their identity. The pre- and post-study surveys were also linked only to these random IDs. Additionally, we truncated the URLs in our data to include only the domain name (e.g., *"facebook.com/UserName"* became *"facebook.com"*).

To analyze patterns of web use, we categorized the websites using the Symantec WebPulse Site Review tool [25], which offers three levels of categorization: categories, subgroups, groups. We focus on subgroups of categories, as they offer the right level of granularity for our analysis (Table II). For simplicity, we refer to the subgroups simply as "categories".

TABLE II: Website categories [25]

Category	Subcategory (examples)
Adult Related	Adult/Mature Content, Gore/Extreme
Liability Concerns	Piracy/Copyright Concerns, Violence/Intolerance
Security Threats	Malicious Outbound Data/Botnets, Phishing
Security Concerns	Compromised Sites, Hacking, Spam
File Transfer	File Storage/Sharing, Peer-to-Peer (P2P)
Society/Government	Charitable/Non-Profit, Government/Legal
Social Interaction	Personal Sites, Social Networking
Multimedia	Audio/Video Clips, Media Sharing
Communication	Email, Internet Telephony, Online Meetings
Health Related	Health, Restaurants/Food, Tobacco
Leisure	Art/Culture, Entertainment, Games
Commerce	Cryptocurrency, Job Search/Careers, Shopping
Technology	Cloud Infrastructure, Computer/Information Security
Information Related	Education, News, Reference, Search Engines/Portals

TABLE III: Participant demographics

Demographic	Group	Number (%)
Gender	Female	12 (38.7%)
	Male	19 (61.3%)
Age	18-24	14 (45.2%)
	25-34	14 (45.2%)
	35-44	3 (9.7%)
Race	Asian	7 (22.6%)
	Black or African American	10 (32.3%)
	White	9 (29.0%)
	Two or more races	3 (9.7%)
Nationality	USA	16 (51.6%)
	Other	15 (48.4%)

2) Self-reported Data: Two surveys were conducted immediately before and after the period of browsing data collection. The pre-study survey asked participants to self-report their demographics, daily time spent online, and most visited websites. In the post-study survey, participants were asked if they altered their behavior in any way during the study. More details about the surveys can be found in the full version of our paper [26].

C. Data Analysis

a) RQ1: Differences in behavior.: We first examine differences in web use between participants of different genders, races, and ages, using two metrics: time spent browsing and number of browsing actions. To compute browsing time, we grouped consecutive actions into *clickstreams*, which represent a period of continuous active browsing. Prior work ended a clickstream after periods of inactivity ranging from 30 seconds on Facebook [9] to 30 minutes across all websites. We chose to use 30 minutes of inactivity as a cutoff because we considered the full range of internet browsing.

We use a one-sided t-test with the null hypothesis of no difference between the mean number of daily browsing actions (or mean number of hours spent browsing) on average across days with observed online behavior, for each pair of demographic groups. For brevity, we will refer to these two metrics respectively as "daily average number of browsing actions" and "daily average browsing time,". We apply bootstrapping to account for smaller sample sizes, and we correct for multiple testing using Bonferonni-Holm correction. We compare only Asian, Black or African American, and white races, since these have sufficient representation in our sample.

To investigate web use patterns, we test whether participants had similar distribution of browsing actions on websites of different categories. Only three types of actions could occur within a specific website: *click*, *type*, and *urlChange*. We measure the empirical distribution of these actions by category and use a Pearson's χ^2-test for homogeneity to test if the distribution of actions were the same across categories.

b) RQ2: Accuracy of perceptions.: To address RQ2, we compare participants' observed browsing behavior with their self-reported daily browsing time and most browsed website categories. To measure browsing time differences, we define δ_i for each participant i as the difference between their daily average and self-reported daily time spent online. That is, if participant i spent S_i total hours of browsing, they were active for n_i out of the 14 days, and they self-report spending t_i hours per day online, then δ_i is defined as $\delta_i = \frac{S_i}{n_i} - t_i$. We use a t-test to determine if the mean of participants' δ_is are significantly different from 0, indicating discrepancy between observed and perceived browsing time. We examine whether these differences vary across demographic group, applying a t-test for mean equality of the δ_is and bootstrap techniques to improve the statistical power.

Next, we assess participants' perceptions of their most commonly browsed websites. We compare their self-reported top browsing categories to the top categories of observed browsing time. We report true/false positive/negative rates for each website category, as well as overall percentage of participants with correct/incorrect perceptions. To determine browsing time in each category, we separate clickstream sessions based on URL category changes. We use categorization from Alexa Top Websites [27] for this analysis, as aligned with our pre-study survey. Although this tool was retired after our study, we used it here for consistency, and we use the Symantec categorization for the remainder of the paper for reproducibility.

c) RQ3: Changes under observation.: Prior work suggests that the effects of observation are amplified by reminders of the observation [19], [20]. If true, we would expect participants' behavior to change over time, moving further from the initial study setup. To test this hypothesis, we compare the distributions of participants' activity during the first (Days 1-7) and the second (Days 8-14) half of the study.

We first test for changes in participants' browsing activity levels, using daily average number of browsing actions and daily average browsing time. We test for changes in the mean and the variance using a t-test and Levene test, respectively. We then test for changes in the distribution of website categories in time, using Wilcoxon signed-rank test to evaluate if the aggregated distribution of both metrics across categories from the first and the second halves of the study are different.

D. Limitations

As with any user study, our findings are subject to multiple practical limitations. First, our sample population was relatively small; to mitigate this we used bootstrapping, a robust and commonly-used technique for calculating estimators for small sample sizes [28]. Our sample is also not fully representative of the internet-using population, and our results should be interpreted in this context. Second, we use self-report data to answer RQ2. Certain self-reported data (e.g., demographics) were unverifiable in the study. Participants could also have misinterpreted the survey questions, or changed their answers due to desirability bias towards a more acceptable behavior [29]. Third, for ethical data collection, users could turn off the extension at any time, allowing them to opt out. Finally, our metric of active browsing time converts a series of instantaneous events into an aggregate measure of time spent browsing, as is the convention in prior work [4], [9], [21]. This approach does not capture passive browsing activities. Additionally, due to the experiment design, our data includes only browsing through a Chrome browser on desktop devices.

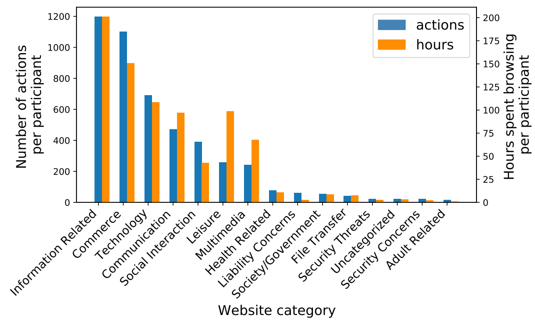

Fig. 1: Distribution of number of browsing actions (blue) and browsing time in hours (orange) on different website categories averaged over all participants in our study.

IV. RQ1: Browsing Behavior Differences

In this section, we measure our participants' browsing behavior in terms of time spent browsing and number of browsing actions. We test whether that behavior differs based on the type of user (i.e., demographics) in Section IV-A and type of web use (i.e., website category) in Section IV-B.

Overall, we observe that our participants spent an average of 146 minutes (SD = 100.5) browsing daily with an average of 968 browsing actions per day (SD = 1529). Figure 1 illustrates the distribution of time spent and number of actions in each category. We see that "Information Related", "Commerce", and "Technology" websites are the most popular according to both metrics of activity. Some categories, such as "Social Interaction," "Leisure," and "Multimedia", are popular under one metric, but not the other, which suggests that browsing such websites as "Leisure" and "Multimedia" does not involve as many click actions as other categories.

Compared to prior work, we observe that our participants visit similar numbers of pages per session (5-151 per session vs. 14-130 in [21]), but spend more time online (median of 2.9 hours per day vs. a median of an hour per day in [12]).

A. Differences across demographic groups

First, we consider the differences in participants' behavior across demographic groups, motivated by prior work linking browsing behavior and demographic features [2], [13], [14], [30]. We explore this by testing for differences in daily average number of browsing actions and daily average browsing time (per person) across demographic groups.

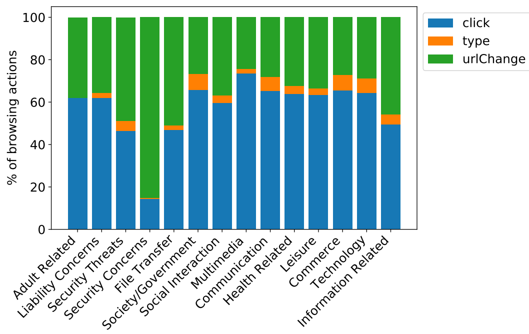

Fig. 2: Empirical distributions of participants' *click*, *type*, and *urlChange* browsing actions within each website category.

With daily average number of browsing actions, we find no significant difference between genders ($t = -0.228$, $p = 0.822$), between Black or African American and Asian participants ($t = 0.688$, $p = 0.502$), between white and Asian participants ($t = -0.321$, $p = 0.754$), and between white and Black or African American participants ($t = -0.760$, $p = 0.461$). We also observe no significant difference in daily activity level between those aged 18-24 and those aged 25-34 ($t = 0.322$, $p = 0.999$). However, the daily activity of older participants (aged 35-44) is significantly lower than that of those aged 18-24 years and 25-34 years ($t = 3.301$, $p = 0.007$ and $t = 2.994$, $p = 0.051$, respectively).

With daily average browsing time, we find that there is no significant difference between genders ($t = -0.073$, $p = 0.950$) and among races (pairwise, $t = -0.842$, $p = 0.381$; $t = -0.642$, $p = 0.501$; $t = 0.114$, $p = 0.918$). We do not observe a significant difference in daily average browsing time between those 18-24 years old and those aged 25-34 ($t = -0.467$, $p = 0.999$). However, participants aged 35-44 on average spend significantly less time online daily than younger participants ($t = 3.297$, $p = 0.007$ in comparison to those aged 18-24, and $t = 3.187$, $p = 0.013$ in comparison to those aged 25-34, respectively).

B. *Behavior differences by web use*

Next we explore how users' behavior varies across different website categories, motivated in part by prior research showing different interaction patterns across websites [2], [12], [21]. We aim to understand whether this behavior varies structurally by website category. Thus we measure behavior by the distribution of browsing actions within each website category.

Figure 2 shows the empirical distribution of click events on different website categories. For most website categories, participant actions were mostly *click*, fewer *urlChange*s, and a small number of *type* actions. A notable exception is Security Concerns websites, which saw significantly different behavior from all categories ($\chi^2 > 27$, $p < 10^{-5}$ for all categories). Behavior on Multimedia websites is significantly different from behavior on File Transfer ($\chi^2 = 15.349$, $p = 0.036$) and Security Threats ($\chi^2 = 15.969$, $p = 0.027$) websites. Differences between other websites pairs were not significant.

a) Security Concerns.: In our study, 10 out of 31 participants visited a total of 62 Security Concerns websites, categorized into four subcategories: Suspicious, Placeholders, Potentially Unwanted Software, and Hacking. The majority of the visits (48 out of 62) were to Suspicious subcategory.

The distinct user behavior on Security Concerns websites aligns with the typical tactics of malicious websites, which aim to redirect users to further malicious pages or capture their credentials [31]. The behavior signals we observe may serve to augment existing approaches to detecting new or unknown Security Concerns websites [32]–[34] and may be useful for developing just-in-time in-browser warnings about a potential security concern based on observed browsing behavior.

V. RQ2: Online Behavior Perceptions

In this section we test whether participants had accurate perceptions of their online browsing behavior. We first evaluate participants' perceptions of their time spent online and how this varies by demographic group in Section V-A, and then we measure participants' perceptions of the website categories that they most frequently browse in Section V-B.

We observe that our participants think they spend on average 6.87 hours (SD = 4.6) per day browsing. From the pre-study survey, the most frequented categories of websites were "Entertainment", "Search", and "Social Network" (respectively 27, 27, and 23 participants out of 31).

A. *Perceptions of time spent browsing*

We find that the majority of participants (26 out of 31, 80.6%) significantly over-reported their daily browsing time. Figure 3 illustrates the relationship between participants' observed time spent browsing and their perceived (self-reported) time spent browsing. Figure 3a shows a scatter plot with one dot corresponding to each participant, where the x-coordinate is their daily average browsing time, and the y-coordinate is the self-reported daily time spent browsing. The red line ($x = y$) corresponds to no error in perceptions, and points further from this line have larger error between perceptions and actual browsing behavior. We observe that most participants substantially overestimated their time spent browsing, as evidenced by the number of points above the red line.

Figure 3b aggregates this information to illustrate the error in participants' browsing behavior perceptions at the population-level. Recall that δ_i is the difference between the observed daily average browsing time of participant i, and their perceived daily browsing time. Since a large fraction of participants had a negative value of δ_i, we see that most participants over-reported their time spent browsing. The average error δ_i among our participants is -4.5 hours (SD=5.24).

Alternative measures of activity. We followed the convention of [21], assuming that a browsing session ends after 30 minutes of inactivity (i.e., no actions aside from the *awake*). Other studies used cutoffs ranging from 30 seconds [9] to 20 minutes [4] of inactivity. Using shorter cutoff times to indicate inactivity would only reduce the recorded time spent browsing, and thus increase overestimation of browsing activity.

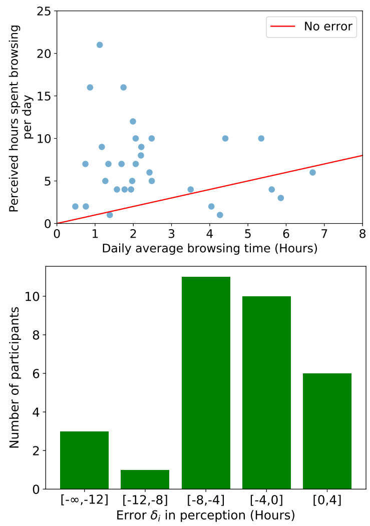

Fig. 3: (a) Scatter plot illustrating actual daily average browsing time vs. perceived (self-reported) number of hours spent browsing per day. Each point corresponds to one participant. (b) Distribution of error values δ_i in the participant population.

Our analysis only included browsing time spent on a laptop or desktop, as measured by our browsing extension, and did not include mobile browsing activity. Recent 2021 data [35] show that 55.9% of browsing time is spent on a desktop/laptop. To account for this, we adjusted participants' self-reported time by scaling down by a factor of 0.559. Even after the adjustment, most participants still overestimate their online time (80.6% of without adjustment vs. 77.4% with adjustment).

Demographic variance. We additionally investigate whether the biases in participants' perceptions differ across demographic groups. For each feature, we test for equality of means for δ_i across groups. We find no significant difference among genders ($t = -0.536$, $p = 0.599$), age groups (pairwise, $t = -0.796$, $p = 0.433$; $t = 0.300$, $p = 0.791$; $t = 0.508$, $p = 0.659$), and races (pairwise, $t = 0.038$, $p = 0.970$; $t = -0.831$, $p = 0.420$; $t = -0.673$, $p = 0.511$).

B. *Perceptions of browsing activity*

Next, we investigate whether participants had correct perceptions about the type of websites they browse. In the pre-study survey, each participant i reported their k_i most frequently visited categories (see Section III-C); we compared this with their observed top k_i website categories. Participants on average chose 4.53 categories (SD = 1.34).

Table IV presents a confusion matrix that summarizes whether participants' self-reported top browsing categories were among their actual most browsed categories. Table IV also shows the percentage of participants with correct and incorrect perceptions for each category. In Table IV, orange cells indicate incorrect perceptions. Specifically, the orange column on the left (observed in top k_i categories of browsing, but not self-reported in top k_i) corresponds to participants who underestimated the amount of time they spent on each category, which indicates false positive rate. The orange column on the right (not observed in top k_i categories of browsing, but self-reported to be in top k_i) corresponds to participants who overestimated the time they spent on each category, which indicates false negative rate.

TABLE IV: Confusion matrix showing accuracy of participants perceptions of their most frequently browsed website categories. Each participant i reported their k_i top categories, which were compared with their observed top k_i categories based on browsing time. Blue cells indicate correct perceptions (true positives/negatives), and orange cells indicate incorrect perceptions (false positives/negatives). Total correct and incorrect perceptions are calculated for each category.

Category	Observed Among Top k_i		Observed Not Among Top k_i		Correct perception	Incorrect perception
	Self-Report Top k_i	Self-Report Not Top k_i	Self-Report Top k_i	Self-Report Not Top k_i		
Blogging	0%	3.1%	12.5%	84.4%	84.4%	15.6%
News	0%	3.1%	18.8%	78.1%	78.1%	21.9%
Search	65.6%	9.4%	18.7%	6.3%	71.9%	28.1%
Social Network	53.1%	9.4%	18.7%	18.8%	71.9%	28.1%
Banking	6.3%	0%	31.2%	62.5%	68.8%	31.2%
References	28.1%	15.6%	21.9%	34.4%	62.5%	37.5%
Entertainment	56.3%	9.4%	28.1%	6.2%	62.5%	37.5%
Shopping	12.5%	6.3%	40.6%	40.6%	53.1%	46.9%
Business	6.3%	18.7%	34.4%	40.6%	46.9%	53.1%

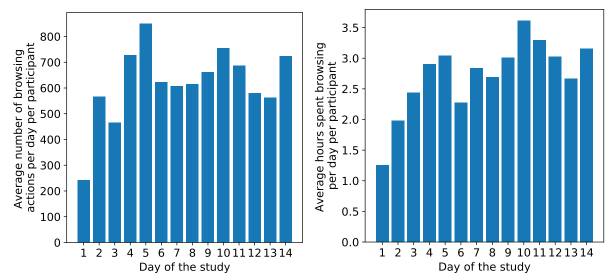

Fig. 4: (a) Average number of browsing actions and (b) time spent browsing, per participant per active browsing day of the study.

While participants over-report their time online, they are relatively accurate in identifying where they spend that time. Of the 145 top categories reported by 31 participants, 50.3% matched their observed behavior. The categories are sorted in Table IV by accuracy of participant perceptions. Most errors were due to overestimating browsing in certain categories (false positives), occurring uniformly across categories.

VI. RQ3: Behavior Changes

In this section, we evaluate whether participants' behavior changed over the course of the experiment. We first test for changes in level of browsing activity in Section VI-A, and then for differences in website categories browsed in Section VI-B.

A. *Changes in level of activity*

We investigate whether participants changed their level of browsing activity during the study. We use as activity metrics both daily average number of browsing actions and daily average browsing time. Figure 4 shows both metrics averaged across all participants. We note that participants installed the browsing extension during the first day of the study, so

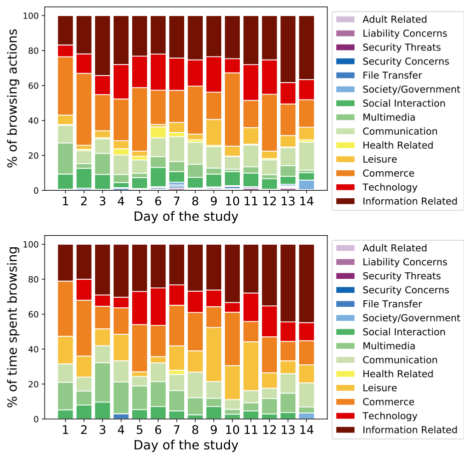

Fig. 5: Proportion of (a) browsing actions and (b) time spent browsing on each website category on each day of the study.

browsing activity is noticeably lower on Day 1.

While we observe variance in average daily activity, there were no significant differences in browsing activity over the study's duration. For the average number of actions per active browsing day, we find that neither the mean ($t = -0.915$, $p = 0.348$) nor variance ($L = 2.009$, $p = 0.182$) differs significantly between the first and second half of the study. Similarly, for daily average browsing time, both mean of the number of hours ($t = -1.23$, $p = 0.208$) and variance ($L = 2.191$, $p = 0.165$) show no significant difference over the time of the study. Post-study survey results support these findings, with only 2 of 31 participants reporting behavior changes.

B. Changes in type of web use

We also investigate whether participants' browsing activity across website categories changed over the course of the study. Figure 5 shows the proportion of browsing activity across website categories as measured by both the number of browsing actions and hours spent browsing. With number of browsing actions, we do not observe a significant change in distribution of web use across website categories between the first half of the study (Days 1-7) and the second half (Days 8-14) ($W = 37.0$, $p = 0.357$); similarly, under time spent browsing, we do not observe a significant difference ($W = 38.0$, $p = 0.390$).

VII. Discussion and Conclusions

In this work we provide an up-to-date picture of browsing behavior in a young, educated (under 45 years old with some college education) sample of internet users (RQ1). We find that people are viewing similar numbers of pages today as in prior work but are spending significantly more time online — an average of three hours a day compared one hour daily in the prior work [12], [21]. Echoing prior work [13], we find little demographic variance in browsing activity, although differently from prior work that leveraged demographic inference data [13], we do observe that the older users (35-44) in our sample browse significantly less than younger users.

Our work adds to the body of knowledge on digital browsing behavior in that we examine not only how much time people spend online and how many pages they view, but *what* they do online. Prior work has studied user behavior in terms of webpage access time [21] and number of page revisits [12]. To the best of our knowledge, this is the first work to study user behavior in terms of the proportion of types of actions performed across different website categories.

a) Implications for System Design.: We find that people spend most of their time on Information Related, Commerce, Technology, Leisure, and Communication websites, with Social Interaction (social media) websites ranking seventh. While time spent on different types of pages varies, user behavior is similar in terms of the actions (clicks, typing, urlChanges) on these pages. There are a few exceptions: Multimedia websites see less typing and urlChanges, as people are primarily clicking on and watching videos; File Transfer websites see a high number of urlChanges characterizing file uploads; and Security Concern websites, which include suspected phishing URLs and misspellings of popular URLs, and can be characterized by a high number of urlChanges, in line with prior findings that malicious websites aim to redirect users to additional malicious pages and capture credential [31]. These findings suggest that one potentially promising direction for augmenting existing approaches [32]–[34] to keeping people safe online is to add common website interaction patterns as signals for detecting malicious websites.

b) Implications for Future Research on Digital Behavior.: Our work addresses a key question for the study of online behavior: the relationship between participants' self-reported online browsing — in terms of time spent online and types of web uses — and their actual behavior observed using our measurement tools (RQ2). We find that participants significantly over-report their daily browsing time, by an average of 4.5 hours per day, with no variation by age, gender, or race. This finding aligns with prior work showing that people are rarely accurate in their estimate of time spent online [9], [22]. This suggests that findings regarding the relationship between various digital constructs [1], [3] and self-reported online time should be interpreted with care: people's perceptions of how much time they spend online may differ from reality.

While participants in our study over-reported their time spent online, they were relatively accurate in their reports about the types of websites where they spent the most time. This suggests, in line with prior work examining the accuracy of people's self reports about the speed with which they update their computers [8], that people may have an accurate *relative* sense of their digital behavior, but inaccurate absolute perceptions (e.g., the exact amount of time they spend online or the precise strength of their passwords [10]). This suggests that observational methods of measurement may be most appropriate for use when precise absolute measurements are necessary, but that self-report measurements may be an appropriate proxy when only relative measurements are required.

Finally, given prior findings from other fields on possible observational biases that may occur when participants are aware that their behavior is being observed [16], we examine whether their behavior changed during our experiment to detect such observational biases in our measurements (RQ3). We find no significant changes in participant behavior over the course of the study. It may be because 14 days is not a sufficiently long period of time for participants to forget that they are being observed. Alternately, people may have such a pervasive sense of being observed online [36] that even installing a browser plugin that they know observes their behavior may not change their activity. Future work is necessary to further explore the question of observation bias in measurements of digital behavior, perhaps through comparison of proprietary industry measurement data – which a user is not actively aware is being collected – with measurement data from a disclosed browser plugin such as ours.

References

[1] E. Hargittai, "Digital natives? Variation in internet skills and uses among members of the "net generation"," *Sociological Inquiry*, vol. 80, no. 1, pp. 92–113, 2010.

[2] T. Li, Y. Li, M. A. Hoque, T. Xia, S. Tarkoma, and P. Hui, "To what extent we repeat ourselves? discovering daily activity patterns across mobile app usage," *IEEE Transactions on Mobile Computing*, vol. 21, no. 4, pp. 1492–1507, 2020.

[3] N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of Facebook "friends:" Social capital and college students' use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, no. 4, pp. 1143–1168, 2007.

[4] P. Calafiore and D. S. Damianov, "The effect of time spent online on student achievement in online economics and finance courses," *The Journal of Economic Education*, vol. 42, no. 3, pp. 209–223, 2011.

[5] I.-H. Chen, C.-Y. Chen, A. H. Pakpour, M. D. Griffiths, C.-Y. Lin, X.-D. Li, and H. W. Tsang, "Problematic internet-related behaviors mediate the associations between levels of internet engagement and distress among schoolchildren during covid-19 lockdown: A longitudinal structural equation modeling study," *Journal of behavioral addictions*, vol. 10, no. 1, pp. 135–148, 2021.

[6] Z. Tu, H. Cao, E. Lagerspetz, Y. Fan, H. Flores, S. Tarkoma, P. Nurmi, and Y. Li, "Demographics of mobile app usage: Long-term analysis of mobile app usage," *CCF Transactions on Pervasive Computing and Interaction*, vol. 3, no. 3, pp. 235–252, 2021.

[7] P. A. Williams, J. Jenkins, J. Valacich, and M. D. Byrd, "Measuring actual behaviors in HCI research–A call to action and an example," *Transactions on Human-Computer Interaction*, vol. 9, no. 4, pp. 339–352, 2017.

[8] E. M. Redmiles, Z. Zhu, S. Kross, D. Kuchhal, T. Dumitras, and M. L. Mazurek, "Asking for a friend: Evaluating response biases in security user studies," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '18, 2018, pp. 1238–1255.

[9] S. K. Ernala, M. Burke, A. Leavitt, and N. B. Ellison, "How well do people report time spent on Facebook? An evaluation of established survey questions with recommendations," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, 2020, pp. 1–14.

[10] R. Wash, E. Rader, and C. Fennell, "Can people self-report security accurately? Agreement between self-report and behavioral measures," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17, 2017, pp. 2228–2232.

[11] H. Mukhtar, F. Seemi, H. Aslam, and S. Khattak, "Browsing behaviour analysis using data mining," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019.

[12] R. Kumar and A. Tomkins, "A characterization of online browsing behavior," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10, 2010, pp. 561–570.

[13] S. Goel, J. Hofman, and M. Sirer, "Who does what on the web: A large-scale study of browsing behavior," in *Proceedings of the International AAAI Conference on Web and Social Media*, ser. ICWSM '12, 2012, pp. 130–137.

[14] J. Hu, H.-J. Zeng, H. Li, C. Niu, and Z. Chen, "Demographic prediction based on user's browsing behavior," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07, 2007, pp. 151–160.

[15] A. Perrin, "Social media usage," *Pew research center*, vol. 125, pp. 52–68, 2015.

[16] J. Andreoni and B. D. Bernheim, "Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, vol. 77, no. 5, pp. 1607–1636, 2009.

[17] D. Ariely, A. Bracha, and S. Meier, "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially," *American Economic Review*, vol. 99, no. 1, pp. 544–555, 2009.

[18] E. Hoffman, K. McCabe, and V. Smith, "Social distance and other-regarding behavior in dictator games," *American Economic Review*, vol. 86, no. 3, pp. 653–660, 1996.

[19] E. Ostrom, J. Walker, and R. Gardner, "Covenants with and without a sword: Self-governance is possible," *The American Political Science Review*, vol. 86, no. 2, pp. 404–417, 1992.

[20] M. Rege and K. Telle, "The impact of social approval and framing on cooperation in public good situations," *Journal of Public Economics*, vol. 88, no. 7, pp. 1625–1644, 2004.

[21] M. Abramson and S. Gore, "Associative patterns of web browsing behavior." in *AAAI Fall Symposia*, 2013.

[22] D. A. Parry, B. I. Davidson, C. J. Sewall, J. T. Fisher, H. Mieczkowski, and D. S. Quintana, "A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use," *Nature Human Behaviour*, vol. 5, no. 11, pp. 1535–1547, 2021.

[23] P. Coyne, J. Voth, and S. J. Woodruff, "A comparison of self-report and objective measurements of smartphone and social media usage," *Telematics and informatics reports*, vol. 10, p. 100061, 2023.

[24] B. Wu-Ouyang and M. Chan, "Overestimating or underestimating communication findings? comparing self-reported with log mobile data by data donation method," *Mobile Media & Communication*, vol. 11, no. 3, pp. 415–434, 2023.

[25] Symantec, "Webpulse site review request," https://sitereview.bluecoat.com/, 2021, accessed: 9/30/2020.

[26] Y. Lut, M. Wang, E. M. Redmiles, and R. Cummings, "How we browse: Measurement and analysis of digital behavior," *arXiv preprint arXiv:2108.06745*, 2021.

[27] Amazon, "Alexa top sites by category," https://www.alexa.com/topsites/categories, 2018, accessed: 6/29/2018.

[28] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical science*, pp. 54–75, 1986.

[29] R. M. Groves, F. J. Fowler Jr, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau, *Survey Methodology*. John Wiley & Sons, 2011, vol. 561.

[30] G. Wu, J. Hong, and P. Thakuriah, "Investigating the temporal changes in the relationships between time spent on the internet and non-mandatory activity-travel time use," *Transportation*, vol. 49, no. 1, pp. 213–235, 2022.

[31] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, ser. S&P '11, 2011, pp. 447–462.

[32] C. Cao and J. Caverlee, "Behavioral detection of spam URL sharing: Posting patterns versus click patterns," in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ser. ASONAM '14, 2014, pp. 138–141.

[33] ——, "Detecting spam urls in social media via behavioral analysis," in *Advances in Information Retrieval: 37th European Conference on IR Research, ECIR 2015, Vienna, Austria, March 29-April 2, 2015. Proceedings 37*. Springer, 2015, pp. 703–714.

[34] X. Dong, J. A. Clark, and J. L. Jacob, "User behaviour based phishing websites detection," in *Proceedings of the 2008 International Multiconference on Computer Science and Information Technology*, 2008, pp. 783–790.

[35] BroadbandSearch, "Mobile vs. desktop internet usage," https://www.broadbandsearch.net/blog/mobile-desktop-internet-usage-statistics, 2021, accessed: 5/4/2021.

[36] B. E. Duffy and N. K. Chan, ""You never really know who's looking": Imagined surveillance across social media platforms," *New Media & Society*, vol. 21, no. 1, pp. 119–138, 2019.